



HAL
open science

Structural Demand Estimation with Borrowing Constraints

Amine Ouazad, Romain Rancière

► **To cite this version:**

Amine Ouazad, Romain Rancière. Structural Demand Estimation with Borrowing Constraints. 2015.
halshs-01207997

HAL Id: halshs-01207997

<https://shs.hal.science/halshs-01207997>

Preprint submitted on 1 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

WORKING PAPER N° 2015 – 32

Structural Demand Estimation with Borrowing Constraints

Amine Ouazad
Romain Rancière

JEL Codes:
Keywords:



PARIS-JOURDAN SCIENCES ÉCONOMIQUES

48, Bd JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

Structural Demand Estimation with Borrowing Constraints*

Amine Ouazad[†] Romain Rancière[‡]

September 30, 2015

Abstract

Structural models of housing or product choice use observed demand to estimate household preferences. However, household demand may be partly determined by borrowing constraints, limiting households' choice set. Such borrowing constraints will differ across locations, households, and years. We put forward a model of neighborhood choice with borrowing constraints that accounts for mortgage credit approval rates. We estimate the model's parameters using micro-level data on households, property transactions and mortgage applications for the San Francisco Bay. Approval rates vary significantly both across households and across neighborhoods. The model with borrowing constraints yields significantly higher estimated willingness to pay to live close to good schools and in majority-white neighborhoods. The model provides general equilibrium estimates of the impact of a relaxation of lending standards. Between 2000 and 2006, the model provides two out-of-sample predictions: (i) a compression of the

*We thank Patrick Bayer, Jenna Bednar, Gilles Duranton, Ambika Gandhi, Timothy Van Zandt, and the audience of the INSEAD and University of Southern California seminar series for fruitful comments on previous versions of this paper. Marton Varga provided research assistance. The authors acknowledge financial support from INSEAD, the International Monetary Fund, the Chaire Banque de France. Keywords: Demand Estimation, Housing, Mortgage Credit, House Prices, Segregation . JELs: R21, R23, G21

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

[‡]Paris School of Economics, CEPR and IMF

price distribution and (ii) a decline in black households' exposure to white households. Both predictions are supported by empirical observation.

1 Introduction

The ability to secure mortgage credit is a key determinant of households' decision to purchase housing. Changes in lending standards may 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

¹81% in 2013 according to data provided by RealtyTrac

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 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 an out-of-sample 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 reasonably well the price increase, the dispersion of price changes across neighborhoods, as well as observed patterns of racial exposures that include 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 but also incorporates 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 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, property value, and neighborhood prospects. The screening process leads to binding borrowing constraints for a substantial share of applicants (Ross & Yinger 2002, Dell'Arriccia, Igan & Laeven 2009).

This paper's structural model of location choice features such borrowing constraints. Absent borrowing constraints, 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 is no longer the case when borrowing

constraints 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 prevents 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 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. 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 approved. 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.² 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.³ 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 on approval rates by comparing applicants with dif-

²The contraction mapping is shown to yield a unique vector of utilities, up to a constant, for every observed vector of aggregate product demand.

³The 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.

ferent characteristics within the same neighborhood in the same year using tract–year fixed effects in panel. A final issue is present in prior models of consumer choice. House prices are typically endogenous with respect to local unobservables; 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 the preference of households for 4,416 neighborhoods (i.e. census blockgroups) in the San Francisco Bay area. 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 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 without borrowing constraints yields significantly larger own-price demand elasticities than the model with borrowing constraints. Without borrowing constraints, the estimated price elasticity captures implicitly both household preferences and credit conditions. 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 yields two kinds of demand elasticities: conditional and total (own-price) demand elasticity. Conditional demand elasticity assumes that mortgage approval probabilities and the resulting neighborhood choice set are left unchanged. Total demand elasticity takes into account the effect of a price change *on* the choice set and the effect of a price change on location choice *within* the choice set. 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, out-of-sample 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 at 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. This cross-sectional dispersion in house price change has been documented for a number of U.S. cities. Landvoigt, Piazzesi & Schneider (2015), using a calibrated assignment model, shows how the relaxation of collateral constraints can replicate this empirical finding. However, our paper is, to our knowledge, the first that shows how this pattern emerges within a structurally estimated model of housing with borrowing constraints. We also analyze the effect of the same change in lending standard on spatial segregation. Very much like in the data (Ouazad & Rancière Forthcoming), the model predicts a slight reduction in the isolation of Blacks (-0.62), that results from largely offsetting forces: an reduction in the exposure of Blacks to Whites (-2.04), and an increase in the exposure of Blacks to Hispanics ($+1.9$) and to Asians ($+0.76$).

This paper contributes to the consumer choice literature sparked by Berry et al. (1995) in consumer products and in housing by Bayer & Timmins (2007) and Bayer et al. (2007).⁴ The use of microdata in the estimation of the parameters, in our case on demographics and neighborhood characteristics, builds on the approach of Petrin (2002). This paper's contribution is twofold. The paper's first contribution is an estimated equilibrium model that accounts for the effect of borrowing constraints on households' choice sets. The paper's second contribution is the general equilibrium analysis of the impact of lending standards

⁴See Holmes & Sieg (2015) for a comprehensive survey of structural estimation in urban economics.

on prices and on household segregation, as housing supply elasticity is likely smaller than supply elasticity for consumer products.

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 demand side rather than for the supply side.

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 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 $t = 1, 2, \dots, T$. In year t , the metropolitan area has a number N_t of households indexed by $i = 1, 2, \dots, N_t$. There is a fixed number of $J \geq 2$ neighborhoods and each neighborhood $j = 1, 2, \dots, J$ in year $t = 1, 2, \dots, T$ has a number s_{jt} of housing units. Household i in year t chooses a neighborhood $j(i, t) \in \{1, 2, \dots, J\} \equiv \mathbb{J}$, 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. $j(i, t) = \operatorname{argmax}_{j \in C_{it}} U_{ijt}$. Utility U_{ijt} is allowed to vary across neighborhoods, across individuals, and to vary across neighborhoods according to individual characteristics. Thus:

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

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 $K_2 \geq 1$. 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 $K_1 K_2$ interaction coefficients. Therefore $\mathbf{x}'_{it} \Omega \mathbf{z}_{jt}$ captures how neighborhood utility varies according to households' observable characteristics. In contrast, the term $\tilde{\boldsymbol{\beta}}'_{it} \mathbf{z}_{jt}$ 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 , $k = 1, 2, \dots, K_2$. Normalizing each random coefficient $\tilde{\beta}_{itk}$ by σ_k , with $\tilde{\boldsymbol{\beta}}_{it} = \left(\tilde{\beta}_{itk} \right)_{k=1,2,\dots,K_2}$, 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 $\mathbf{z}'_{jt} \boldsymbol{\beta}$, and by unobservable neighborhood heterogeneity ξ_{jt} :

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

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:

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

where $\mathbf{1}(j \in C_{it}) = 1$ whenever $j \in C_{it}$, and 0 otherwise. Integrating over the vector of random coefficients $\tilde{\boldsymbol{\beta}}_{it}$ obtains the probability conditional on observables and on base utility:

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

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 . 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 probabilities 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:

$$P(C_{it} = C) = \prod_{j \in C} P(j \in C) \prod_{j \notin C} [1 - P(j \in C)] \equiv \prod_{j \in C} \phi_{ijt} \prod_{j \notin C} (1 - \phi_{ijt}) \quad (5)$$

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 .

The probability ϕ_{ijt} of approval is based on banks' benefit-cost analysis of originating a mortgage loan for household i for a house in neighborhood j . Such cost-benefit 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 minus cost ϕ_{ijt}^* depends on household observables \mathbf{x}_{it} and neighborhood characteristics \mathbf{z}_{jt} .

$$\phi_{ijt}^* = \mathbf{x}_{it}'\boldsymbol{\psi} + \mathbf{z}_{jt}'\boldsymbol{\gamma} + \mathbf{x}_{it}'\Psi\mathbf{z}_{jt} + e_{ijt} \quad (6)$$

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 $\Lambda(\cdot)$. The probability of approval in neighborhood j for household i is therefore simply a logistic function of individual and household observables:

$$P(j \in C_{it}) = \Lambda(\mathbf{x}_{it}'\boldsymbol{\psi} + \mathbf{z}_{jt}'\boldsymbol{\gamma} + \mathbf{x}_{it}'\Psi\mathbf{z}_{jt}) \quad (7)$$

Combining (5) and (7), the probability of the choice set depends on the entire vector of neighborhood observables:

$$P(C_{it} = C | \mathbf{x}_{it}, \mathbf{z}_{jt}) = \prod_{j \in C} \Lambda(\mathbf{x}_{it}'\boldsymbol{\psi} + \mathbf{z}_{jt}'\boldsymbol{\gamma} + \mathbf{x}_{it}'\Psi\mathbf{z}_{jt}) \prod_{k \in C} \{1 - \Lambda(\mathbf{x}_{it}'\boldsymbol{\psi} + \mathbf{z}_{kt}'\boldsymbol{\gamma} + \mathbf{x}_{it}'\Psi\mathbf{z}_{kt})\} \quad (8)$$

(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.

$$P(j, t | \boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}) = \sum_{C \in \mathbf{C}} P(j | \boldsymbol{\delta}, \mathbf{z}_t, x_{it}, C_{it}) \cdot P(C | \mathbf{z}_t, x_{it}) \quad (9)$$

As the set of all possible choice sets \mathbf{C} has 2^J , such actual demand is the sum over 2^J terms. Equation (9) 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:

$$D(j, t) \equiv P(j, t | \boldsymbol{\delta}, \mathbf{z}_t, f(\cdot)) = \sum_{C \in \mathbf{C}} \int_{\mathbf{x}} P(j | \boldsymbol{\delta}, \mathbf{z}_t, x_{it}, C_{it}) \cdot P(C | \mathbf{z}_t, x_{it}) \cdot f(\mathbf{x}_{it}) d\mathbf{x}_{it} \quad (10)$$

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 .

(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 η_{jS} is equal to the effect of a change in the log price on demand keeping the choice sets C_{it} , $i = 1, 2, \dots, N$, $t = 1, 2, \dots, T$ constant. For a fixed probability distribution $P(C)$ over choice sets, we note

$$\eta_{jt|C} = \left. \frac{\partial \log D(j, t)}{\partial \log(p_{jt})} \right|_{P(C)} \quad (11)$$

the own-price demand elasticity for neighborhood j in year t conditional on $P(C)$.

$$\begin{aligned}\eta_{jt|C} &= \frac{1}{D(j)} \frac{\partial D(j,t)}{\partial \log(p_{jt})} \Big|_{P(C) \text{ fixed}} \\ &= -\frac{\alpha}{D(j,t)} \sum_{C \in \mathbb{C}} \int_X P(C|\mathbf{z}_t, \mathbf{x}_{it}) P(j,t|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C) (1 - P(j,t|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C)) f(\mathbf{x}_{it}) d\mathbf{x}_{it}\end{aligned}\quad (12)$$

where α is the impact of log price on base utility (Equation (2)).⁵ 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 $P(C|\mathbf{z}_t, \mathbf{x}_{it})$ 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,

$$\eta_{jt} = \eta_{jt|C} + \sum_{C \in \mathbb{C}} \int_{\mathbf{x}} \left[\frac{\partial}{\partial \log(p_{jt})} P(C|\mathbf{z}_t, \mathbf{x}_{it}) \right] \cdot P(j|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C_{it}) \cdot f(\mathbf{x}_{it}) d\mathbf{x}_{it} \quad (13)$$

The second term of this expression (13) 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.

$$\begin{aligned}\frac{\partial}{\partial \log(p_j)} P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) &= \frac{\partial}{\partial \log(p_j)} \{ \Pi_{k \in C} \phi_{ijt} \Pi_{k \notin C} (1 - \phi_{ijt}) \} \\ &= \begin{cases} a \cdot (1 - \phi_{ijt}) P(C|\mathbf{z}_t, \mathbf{x}_{it}) & \text{if } j \in C \\ -a \cdot \phi_{ijt} \cdot P(C|\mathbf{z}_t, \mathbf{x}_{it}) & \text{if } j \notin C \end{cases}\end{aligned}\quad (14)$$

where a is the coefficient of *log* price in the approval specification (6). Writing $\Phi_{ijt}(C) =$

⁵Derivations in the case where *log* price affects utility differently according to individual income (or another characteristic) are available on demand.

$\mathbf{1}(j \in C) - \phi_{ijt}$, derivations lead to the following closed-form for total demand elasticity.

$$\eta_j = \eta_{j|C} + \frac{1}{D(j, t)} \cdot \sum_{C \in \mathbf{C}} \int_X \Phi_{ijt}(C) \cdot P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it}; \psi) \cdot P(j|\boldsymbol{\delta}, \mathbf{z}_{\cdot t}, \mathbf{x}_{it}, C) f(\mathbf{x}_{it}) d\mathbf{x}_{it} \quad (15)$$

(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 $S_j(\log(p_{jt}))$ in neighborhood j is an increasing and concave function of $\log(p_{jt})$. In this paper's empirical analysis, Sections 4–5, supply is assumed perfectly inelastic, so that $S_j(\log(p_{jt})) = s_{jt}$ for every j, t .⁶

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 $\mathbf{v}_{jt}(x)$ the fraction of households with characteristic x . Write X^+ the set of neighbors' characteristics that affect household utility.

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

⁶General equilibrium results accounting for the durability of housing, and thus accounting for a larger supply than demand for housing in some neighborhoods, are available from the authors.

$(\mathbf{v}_{jt}^*)_{j=1,2,\dots,J,t=1,2,\dots,T}$ such that:

$$\forall j \in \mathbb{J}, \forall t \in \{1, 2, \dots, T\}, \quad D(j, t; \mathbf{p}^*, \mathbf{v}^*) = S_j (\log(p_{jt}^*)) \quad (16)$$

$$\forall x \in X^+, \quad \mathbf{v}_{jt}(x) = \frac{D(j, t|x; \mathbf{p}^*, \mathbf{v}^*)}{D(j, t; \mathbf{p}^*, \mathbf{v}^*)} \quad (17)$$

Condition (16) expresses the equality of demand and supply. In condition (17), $\mathbf{v}_{jt}(x)$ 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 x 's 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 logit specification (7). This paper identifies such parameters within neighborhoods and within years. The second set of parameters measures

the impact of neighborhood observable and unobservables characteristics on base utility, 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 an interaction 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 in the lack of observability of the full range of time-varying household and house covariates that mortgage originators consider in the underwriting process. Empirical Section (4.2) presents a set of instrumental variables Θ_{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:

$$E[G_0(\psi_0, \gamma_0, \omega_0)] = E[\Lambda(\mathbf{x}'_{it}\psi + \mathbf{z}'_{jt}\gamma + \mathbf{x}'_{it}\Psi\mathbf{z}_{jt})\Theta_{ijt}] = 0 \quad (18)$$

(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 (α, β, ξ) 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 (10), is the sum of a large number of 2^J terms, one for each potential choice set $C_{it} \subset \{1, 2, \dots, J\}$, (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 $\mathbf{D} = (D_{jt})_{j \in \mathbb{J}, t \in \{1, 2, \dots, T\}}$ and the vector of base utilities $\boldsymbol{\delta} = (\delta_{jt})_{j \in \mathbb{J}, t \in \{1, 2, \dots, T\}}$, whenever households are unconstrained, i.e. $\phi_{ijt} = 1$ for all households, all neighborhoods and all years. This paper shows that a similar result of existence and uniqueness of the vector $\boldsymbol{\delta}$ applies when the probabilities ϕ_{ijt} of the choice set are introduced. The unique vector $\boldsymbol{\delta}$ is obtained by iterating the following sequence:

$$\hat{\boldsymbol{\delta}}^k = \hat{\boldsymbol{\delta}}^{k-1} + \log(D_{jt}) - \log\left(D\left(j|\hat{\boldsymbol{\delta}}, \mathbf{z}_t, x_{it}\right)\right) \quad (19)$$

and at the limit $\hat{\boldsymbol{\delta}} = \lim_{k \rightarrow \infty} \hat{\boldsymbol{\delta}}^k$ of such sequence, predicted demand $D(j|\hat{\boldsymbol{\delta}}, \mathbf{z}_t, x_{it})$ is equal to observed demand D_{jt} .

In the contraction mapping (19), demand $D\left(j|\hat{\boldsymbol{\delta}}, \mathbf{z}_t, x_{it}\right)$ is the sum of demand for all possible choice sets $C \subset \{1, 2, \dots, J\}$, 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 small neighborhoods, computing over a sum of 2^{4000} choice sets is computationally unfeasible. We proceed by simulating $D\left(j|\hat{\boldsymbol{\delta}}, \mathbf{z}_t, x_{it}\right)$. A set of \mathcal{S} sets is drawn C_{it}^s , $s = 1, 2, \dots, \mathcal{S}$ for each household i in each year t , where the probability

of drawing the set $C \in \mathbb{C}$ of neighborhoods is $\prod_{j \in S} \hat{\phi}_{ijt} \prod_{j \in S} (1 - \hat{\phi}_{ijt})$. Estimation of demand $D(j|\hat{\boldsymbol{\delta}}, \mathbf{z}_{.t}, x_{it})$ given base utilities, and observables thus is the average of the demands based on the simulated choice sets C_{it}^s :

$$\hat{D}(j|\hat{\boldsymbol{\delta}}, \mathbf{z}_{.t}, x_{it}) = \sum_{i=1}^N \frac{1}{S} \sum_{s=1}^S D(j|\hat{\boldsymbol{\delta}}, \mathbf{z}_{.t}, x_{it}, C_{it}^s) \quad (20)$$

with $D(j|\hat{\boldsymbol{\delta}}, \mathbf{z}_{.t}, x_{it}, C_{it}^s)$ 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 (19).

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 $\boldsymbol{\Xi} = \boldsymbol{\Xi}_{jt}$ orthogonal to the unobservable neighborhood amenities ζ_{jt} , providing a first 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 z_{jt} is first-differenced. The first moment condition expresses the orthogonality of $\Delta\boldsymbol{\Xi}$ with $\Delta\boldsymbol{\zeta}$:

$$E(\Delta\boldsymbol{\Xi}\Delta\boldsymbol{\zeta}) = 0 \quad (21)$$

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 $\Delta\hat{\zeta}_{jt} = \Delta\hat{\delta}_{jt} - \Delta z'_{jt}\beta$ 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 (21) as $E[G_1(\boldsymbol{\theta})] = E[\Delta\Xi\Delta\boldsymbol{\zeta}(\boldsymbol{\theta})] = 0$.

Moment condition (21) together with demand simulation (20) and the contraction mapping (19) 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 $\Sigma = E[\tilde{\boldsymbol{\beta}}_{it}\tilde{\boldsymbol{\beta}}'_{it}]$ measures how unobservable 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 x_{it} of a sample of size N for all households $i = 1, 2, \dots, N$. 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 tract. 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.

$$E \left(\sum_{x_{it} \in \mathcal{X}} D \left(j | \hat{\boldsymbol{\delta}}, \mathbf{z}_{\cdot t}, x_{it}, C_{it} \right) \right) - D_{jt}^{\mathcal{X}} = 0 \quad (22)$$

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

(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 as in Petrin (2002):

$$E \left(G(\boldsymbol{\psi}, \boldsymbol{\gamma}, \Psi, \alpha, \boldsymbol{\beta}, \boldsymbol{\xi}, \Omega, \Sigma) \right) = E \begin{pmatrix} G_0(\boldsymbol{\psi}, \boldsymbol{\gamma}, \Psi) \\ G_1(\alpha, \boldsymbol{\beta}, \boldsymbol{\xi}, \Omega, \Sigma) \\ G_2(\alpha, \boldsymbol{\beta}, \boldsymbol{\xi}, \Omega, \Sigma) \end{pmatrix} = 0 \quad (23)$$

A consistent and asymptotically normal estimator of household preferences and banks' lending standards follows from Hansen (1982):

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} G^*(\boldsymbol{\theta})' G^*(\boldsymbol{\theta}) \quad (24)$$

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 $\hat{\boldsymbol{\psi}}$, $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\Psi}}$ 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 $\Lambda(\mathbf{x}'_{it}\hat{\boldsymbol{\psi}} + \mathbf{z}'_{jt}\hat{\boldsymbol{\gamma}} + \mathbf{x}'_{it}\hat{\boldsymbol{\Psi}}\mathbf{z}_{jt})$.
- (b) Draw a choice set 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 $\tilde{\mathbf{b}}_{it}$.

2. Then we minimize the objective function as in (24). 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 (19).
- (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 $\hat{D}(j, t | \mathbf{x}_{it}, \mathbf{z}_{jt}, \hat{\boldsymbol{\delta}}_t, \tilde{\boldsymbol{\beta}}_{it}) = \mathbf{1}(C_{ijt} = 1) \cdot \frac{\exp(\delta_{jt} + \mathbf{x}'_{it}\boldsymbol{\Omega}\mathbf{z}_{jt} + \tilde{\boldsymbol{\beta}}'_{it}\mathbf{z}_{jt})}{\sum_{k,s,t.C_{ikt}=1} \exp(\delta_{kt} + \mathbf{x}'_{it}\boldsymbol{\Omega}\mathbf{z}_{kt} + \tilde{\boldsymbol{\beta}}'_{it}\mathbf{z}_{kt})}$.
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 $\hat{\delta}$ 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 equal to the inverse of the variance covariance of the empirical moments in the second step GMM estimation.

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 2010.⁷

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.⁸ Thus, HMDA

⁷As the borders of metropolitan areas change over time, we use constant 2003 borders throughout the analysis and for the three decennial waves of 1990 to 2010.

⁸Specifically, 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.

disclosure requirements apply to more than 90% of all mortgage applications and originations (Dell’Arriccia et al. 2009). We focus on credit standards for first lien mortgages on homeowner occupied houses.⁹ Each mortgage lender report 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 household income distribution 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.,¹⁰ which compiles transaction data based using public records and real estate sales. FNC data not only includes the universe of transaction, but also the specific amount of each transaction and its street address. Street addresses are geocoded using the transaction’s street address and linked to each of our census tracts.¹¹

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 1.¹² 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

⁹HMDA data contains information on the nature of the lien in the 2010 wave, but not in the 1990 and 2000 waves. To eliminate second lien mortgages, also known as piggyback loans, we used the distribution of loan-to-income ratios by lien status for originated loans from Blackbox mortgage data. Below an average LTI threshold of 1.8, a majority of loans are second liens. We eliminate second lien loans by building a ZIP-code-specific LTI threshold, and estimate the approval specification for loans with an LTI above such threshold.

¹⁰FNC Inc., headquartered in Oxford Mississippi, collects data that provides collateral information to the mortgage industry.

¹¹Geocoding was performed using Texas A&M’s geocoding services, with a higher than 90% success rate.

¹²We thank the National Historical Geographic Information System at the University of Minnesota for providing formatted Census files.

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 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.¹³ We verify that the distribution of FNC transaction prices matches very 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 the median transaction price¹⁴ in each block group and for each year of observation.

¹³Census 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.

¹⁴We 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.

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,¹⁵ 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.¹⁶ Household-level data provides an empirical estimate of the distribution $f(x_{it})$ of household characteristics. It thus allows for the integration of demand with respect to household characteristics $\int df(x)$, as preferences for housing, local amenities, and neighbors’ characteristics depend on individual household characteristics (third moments of the model’s identifying conditions (23)). Individual characteristics include race (White, Black, Hispanic, Asian),¹⁷ and income as measure by total money income of all household members age 15+ 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 (23)). In order to do so, we use the fact that the variables used to predict approval probabilities for mortgage applicants can be obtained for *all* households by combining Household data from the Census and transaction data from FNC.

¹⁵Borders of the CBD for the San Francisco bay area are according to the 1982 Census of Retail Trade as in Glaeser, Gottlieb & Tobio (2012).

¹⁶We acknowledge support from the Integrated Public Use Microdata Series center at the University of Missouri for providing comparable micro series across three decades.

¹⁷ 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.”

4.2 Estimation of Approval Probabilities

The objective is to estimate an approval regression model using actual mortgage approval decisions, and then use this estimation to predict approval probabilities for all households across all neighborhoods. HMDA is the only data source that informs approval decisions for the universe mortgage applications in the metropolitan area. By contrast, audit-pair studies, i.e. randomized controlled trials allows to directly estimate the causal impact of household and loan characteristics on approval decisions, but are usually available only for a small subset of applicants (Munnell, Tootell, Browne & McEneaney 1996, Ross & Yinger 2002). HMDA data report, for each application, the approval decision, the income and the race of the applicant, the size of the loan,¹⁸ and the tract-level location of the application. 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 *originated* mortgages loans.

In the baseline logit specification, we predict the decision of mortgage approval, for a panel of mortgage applications filed in 1990, 2000, and 2010, based on a number of borrower’s and loan’s characteristics, including the loan-to-income ratio (LTI), the race of the applicant, the type of loan (Conventional, FHA insured, VA insured, FSA-RHS insured¹⁹), time effects, and census-tract, and issuer (i.e. bank) fixed effects. Results are presented in Table 2 for various combination of fixed effects (no fixed effects (f.e.), census-tract f.e., bank f.e., census-tract and bank f.e.). 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 and the type of loans should be interpreted as relative to the dummy for Whites, and to the dummy for Conventional loan , respectively. Standard errors are clustered at the census tract level.

In all specifications, the coefficients for the LTI Ratio and for race dummies are significant

¹⁸Post 2002 HMDA data also report the interest rate of the loan.

¹⁹FHA - Federal Housing Administration; VA - Veteran Administration; FSA-RHS-(Farm Service Agency or Rural Housing Service

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 3.1 to 4.4 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 more fixed effects are introduced. However, even in the most stringent specification (bank and census-tract fixed effects), Blacks and Hispanics face probabilities of approval which are, respectively, 5.8 and 2.6 percentage points lower than the probability of approval for Whites. Finally, when banks fixed characteristics are controlled for, applications made in 2000, i.e. during the mortgage credit boom, enjoy an approval rate which is 6.6 percentage point higher than applications made in either in 1990 or 2010. Table B in the appendix replicates the baseline regression with tract fixed effects for a number of other metropolitan areas in the U.S. The estimated parameters for these cities are very similar to those estimated for the San Francisco Bay, suggesting that our approval model indeed capture common standards in mortgage origination decisions.

Identifying the causal impact of applicant and neighborhood characteristics on lenders' decision through the specification described above is challenging for at least two reasons. First, households' race, ethnicity and income are typically correlated with unobservable dimensions of borrower creditworthiness. Second, loan characteristics are typically chosen endogenously by banks and borrowers; such loan characteristics depend on the borrower's creditworthiness as well as observable and unobservable collateral (housing) characteristics.

We address the issue of the endogeneity of applicant and loan characteristics in the following way. We augment the specification above by including tract fixed effect by year, to capture time-varying location-specific unobservables, e.g. unobservables related to the potential appreciation of a neighborhood. Introducing census-tract effects interacted with year effects blows the number of fixed effect from 1,608 to about 4,824, and results in a lack

of convergence of the logit estimation procedure. For this reason, we choose to present probit results instead for this specification. Table (A) in the appendix presents the results. The first four columns allow to compare logit and probit results in the the baseline specification with census-tract fixed effect and year fixed effects only, showing that the estimates of the marginal probabilities are very similar in both specifications. The last two columns show the probit regression results with time interacted with census fixed effect, and show that the estimated marginal probabilities are almost the same than those obtained with a specification with tract and year fixed effects only.

The approval regression is then used to predict the probability of approval for all households, and 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 computing using FNC housing price and the total household income of the Census, under the assumption of a loan-to-value of 0.8, that is, the norm for conforming loans used for home purchase (Adelino, Gerardi & Willen 2013).²⁰

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 defines the *base utility* δ), across individuals, and to vary across neighborhoods according to individual characteristics (Equation 1). In turn, the base utility depends in turn on the neighborhood characteristics (Equation 2) which include: log median transaction price $\log(p_j)$, household demographic composition (race, income, and education), school test scores (California’s API), and the quality of housing construction (age structure and number of rooms). As discussion in Section 3.1, the estimation of the relationship between base utility, log price, and neighborhood characteristics requires an IV approach since

²⁰Results are robust to the inclusion or exclusion of loans above the conforming loan limit.

those 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, $\log(\text{median household income})$, % college educated) by neighboring tracts' demographic composition as in Bayer & Timmins (2007).

The selection of the neighborhood characteristics included in the preference regression builds on the existing literature. The racial composition of neighborhoods and its interaction with household race allows to test for heterogeneous racial preferences (Bayer, McMillan & Rueben 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 Districts, relates to the monocentric model (Fujita 1989), and the sorting of individuals according to the relative importance of their taste for either larger houses, or for closer proximity with the city center and shorter commutes. 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 (3) presents the 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 last column regresses the difference in base utility with and without borrowing constraints, and thus assesses the statistical significance of the differences between columns (1) and (2)'s coefficients. 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 3, 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.²¹

The sensitivity of the base utility to a change in the median price of a given neighborhood is significantly stronger in the model without than in the model with borrowing constraints (-1.742 vs. -1.584). In the model without borrowing constraints, the estimated sensitivity of base utility to price reflects implicitly both household preferences and the negative effect of higher prices on approval probability. By contrast, the model with approval constraints disentangles both effects, and reveals a smaller sensitivity of the base utility to an increase in house prices. The model also allows the sensitivity of utility to price to be heterogenous across agents; the estimates reveal a slightly larger dispersion in the price-sensitivity of utility in the model with borrowing constraints.

The estimates of the effects of the racial composition of neighborhoods on utility also

²¹As the distribution of the fraction black, Hispanic, or asian 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.

reveals sharp differences between the two models. The willingness to pay for a non-black household to live in a neighborhood with a significant lower fraction of black neighbors (i.e. a variation of one percentage point in the fraction of Blacks) is \$32,871 in the model with borrowing constraints vs. \$25,647 in the model without borrowing constraints. For Blacks, the willingness to pay to live in neighborhoods with a lower fraction of Blacks is, as expected, smaller in both models, but the difference in willingness to pay between the two models is almost as high. Similar differences arise when we consider the fraction of Hispanics and Asians, with the notable differences that Hispanic and Asian households derive a positive utility from living in neighborhoods with more same-race neighbors. As discussed in Ouazad & Ranci ere (Forthcoming) the existence of approval constraints mitigate how preferences for neighborhoods' racial composition gets reflected into demand differences. Therefore it is not surprising that the model with approval constraints reveals sharper underlying differences in preferences with regards to the racial composition of neighborhoods. Overall the model without borrowing constraints tend to underestimate the preference-based willingness-to-pay for more desirable neighborhoods because they are subject to high probability of denial for a large subset of households.

Similarly the model estimated with borrowing constraints exhibits a larger willingness to pay for neighborhoods with a higher median income, a larger fraction of college graduates, better performing public schools²², and more recently constructed dwellings. The differences in willingness to pay between the two models is, however, smaller for differences in these amenities than for differences in the racial composition of neighborhoods. The model without borrowing constraints exhibits a slightly higher preference for larger houses. This latter result suggests that the usual trade-off between a better neighborhood vs. a larger house, appears to be less important when approval constraints are controlled for.

²²The 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.

4.4 Price Elasticity

The model with borrowing constraint allows to estimate separately 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 price* but we shall emphasize that the model provides estimates for the complete set of cross-price and income elasticities.²³ The first component, the conditional elasticity of demand, is computed assuming that approval-based choice sets are not 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 coefficient on neighborhood $\log(\text{price})$. 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. The preference-based demand for relatively expensive neighborhoods is less price-sensitive than those of relatively cheap neighborhoods. The opposite is true for the borrowing elasticity. At a given

²³The full matrix of total demand derivatives is used in the general equilibrium analysis of Section 5.

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 (Figure (3)). 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 (Dell’Arriccia et al. 2009). 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 (ψ, γ, Ψ) 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:

$$\frac{d\mathbf{D}}{d\log(\mathbf{p})} = \underbrace{\frac{\partial\mathbf{D}}{\partial\psi}}_{\text{Direct Impact}} + \underbrace{\frac{\partial\mathbf{D}}{\partial\log(\mathbf{p})} \cdot \frac{d\log(\mathbf{p})}{d\psi}}_{\text{Impact due to price changes}} + \underbrace{\frac{\partial\mathbf{D}}{\partial\mathbf{v}} \cdot \frac{\partial\mathbf{v}}{\partial\psi}}_{\text{Impact due to social preferences}} \quad (25)$$

Evidence suggests that short-run housing supply in San Francisco is inelastic. For the sake of clarity, we start with a perfectly inelastic housing supply.²⁴ The price change that maintains demand equal to supply in each neighborhood ($\frac{d\mathbf{D}}{d\log(\mathbf{p})} = 0$) is such that:

$$d\log(\mathbf{p})/d\psi = \left[\frac{\partial}{\partial\log(\mathbf{p})} \mathbf{D}(\log(\mathbf{p}), \psi) \right]^{-1} \left[-\frac{\partial}{\partial\psi} \mathbf{D}(\log(\mathbf{p}), \psi) - \frac{\partial\mathbf{D}}{\partial\mathbf{v}} \cdot \frac{\partial\mathbf{v}}{\partial\psi} \right] \quad (26)$$

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.

$$\frac{\partial}{\partial\psi} D(j|\boldsymbol{\delta}, \mathbf{z}_t; \psi) = \sum_{C \in \mathbf{C}} \int_X \left[\frac{\partial}{\partial\psi} P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) \right] \cdot P(j|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C_{it}) \cdot f(\mathbf{x}_{it}) d\mathbf{x}_{it} \quad (27)$$

²⁴We extended the analysis with non-perfectly elastic housing supply, with results available on demand.

Empirically the sum $\sum_{C \in \mathcal{C}}$ is taken over the set of simulated choice sets. The derivative of the probability of the choice set $P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi)$ is the derivative of a product of J terms:

$$\frac{\partial}{\partial \psi} P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) = \frac{\partial}{\partial \psi} [\prod_{j \in C} \phi_{jt} \prod_{k \notin C} (1 - \phi_{kt})] \quad (28)$$

Each individual probability ϕ_{jt} has a simple derivative:

$$\frac{\partial}{\partial \psi} \phi_{jt}(\mathbf{z}_{jt}, \mathbf{x}_{it}; \psi) = w_{ijt} \cdot \phi_{jt}(\mathbf{z}_{jt}, \mathbf{x}_{it}; \psi) \cdot \{1 - \phi_{jt}(\mathbf{z}_{jt}, \mathbf{x}_{it}; \psi)\} \quad (29)$$

where w_{ijt} is the value corresponding to the coefficient ψ . Thus the derivative 29 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 ϕ_{jt} for all neighborhoods j of the city.

Aggregating such result over the J_t factors of (28):

$$\begin{aligned} \frac{\partial P(C|\mathbf{z}_t, \mathbf{x}_{it}, \psi)}{\partial \psi} &= P(C|\mathbf{z}_t, \mathbf{x}_{it}, \psi) \\ &\cdot \underbrace{\sum_j w_{ijt} \cdot \{\mathbf{1}(j \in C) - \phi_{jt}(\mathbf{z}_{jt}, \mathbf{x}_{it}; \psi)\}}_{\Upsilon_{ijt}(C)} \end{aligned} \quad (30)$$

The Monte Carlo simulated estimate of (27) is thus the sum:

$$\frac{\partial}{\partial \psi} \widehat{D}(j|\boldsymbol{\delta}, \mathbf{z}_t; \psi) = \frac{1}{\#\hat{\mathcal{C}}} \sum_{C \in \hat{\mathcal{C}}} \{\Upsilon_{ijt}(C) \cdot D(j|C, \boldsymbol{\delta}, \mathbf{z}_t; \psi)\} \quad (31)$$

Each probability is multiplied by the factor $\Upsilon_{ijt}(C)$. Such expression (31) is a column vector.

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

We then turn to the impact of log prices on demand. The matrix $\frac{\partial}{\partial \log(\mathbf{p})} \mathbf{D}(\log(\mathbf{p}), \psi)$ 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.²⁵ 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:

$$\begin{aligned} \frac{\partial D(j|\boldsymbol{\delta}, \mathbf{z}_t; \psi)}{\partial \log(p_k)} &= \underbrace{\sum_{C \in \mathcal{C}} \int_X \frac{\partial}{\partial \log(p_k)} P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) \cdot P(j|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C_{it}) f(\mathbf{x}_{it}) d\mathbf{x}_{it}}_{\text{Shift in choice set}} \\ &+ \underbrace{\sum_{C \in \mathcal{C}} \int_X P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) \cdot \frac{\partial}{\partial \log(p_k)} P(j|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C_{it}) f(\mathbf{x}_{it}) d\mathbf{x}_{it}}_{\text{Shift in utility}} \quad (32) \end{aligned}$$

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.

$$\begin{aligned} \frac{\partial}{\partial \log(p_k)} P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) &= \frac{\partial}{\partial \log(p_k)} [\prod_{j \in C} \phi_{jt} \prod_{l \notin C} (1 - \phi_{lt})] \\ &= \begin{cases} -\alpha_{approval} (1 - \phi_{kt}) P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) & \text{if } k \in C \\ \alpha_{approval} \phi_{kt} P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) & \text{if } k \notin C \end{cases} \end{aligned}$$

²⁵We note “with respect to” as w.r.t. in the remaining parts of this paper.

Hence the first term (“*shift in choice set*”) in equation 32:

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

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 $P(j)$ w.r.t. p_k is $\alpha_{utility} \cdot P(j) \cdot P(k)$. The coefficient $\alpha_{utility} < 0$ is the coefficient of log price in utility. Thus the second term (“*shift in utility*”) in equation 32:

$$\begin{aligned} & \frac{\partial D(j|\boldsymbol{\delta}, \mathbf{z}_t; \psi)}{\partial \log(p_k)} \Big|_{fixed\ choice\ set} \\ &= \alpha_{utility} \sum_{C \in \mathbf{C}} \int_X P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) \cdot P(j|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C_{it}) P(k|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C_{it}) \cdot f(\mathbf{x}_{it}) d\mathbf{x}_{it} \quad (33) \end{aligned}$$

We apply this same logic to calculate the impact of the price of a neighborhood on its own demand.

$$\begin{aligned} \frac{\partial D(j|\boldsymbol{\delta}, \mathbf{z}_t; \psi)}{\partial \log(p_j)} &= \sum_{C \in \mathbf{C}} \int_X \frac{\partial}{\partial \log(p_j)} P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) \cdot P(j|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C_{it}) f(\mathbf{x}_{it}) d\mathbf{x}_{it} \\ &+ \sum_{C \in \mathbf{C}} \int_X P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) \cdot \frac{\partial}{\partial \log(p_j)} P(j|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C_{it}) f(\mathbf{x}_{it}) d\mathbf{x}_{it} \quad (34) \end{aligned}$$

which turns into the following explicit formula:

$$\begin{aligned} \frac{\partial}{\partial \log(p_j)} D(j|\boldsymbol{\delta}, \mathbf{z}_t; \psi) &= \sum_{C \in \mathbf{C}} \int_X P(C|\mathbf{z}_t, \mathbf{x}_{it}; \psi) \\ &\quad \cdot \alpha_{utility} P(j|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C_{it}) \cdot (1 - P(j|\boldsymbol{\delta}, \mathbf{z}_t, \mathbf{x}_{it}, C_{it})) f(\mathbf{x}_{it}) d\mathbf{x}_{it} \end{aligned}$$

(iii) Impact of Lending Standards on Social Preferences

The last step in completing the calculation of general equilibrium effects on prices is to derive the impact of lending standards on neighborhood composition and the subsequent first-order impact on households' demand for such neighborhoods, noted $-\frac{\partial \mathbf{D}}{\partial \mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial \psi}$.

The second factor $\frac{\partial \mathbf{v}}{\partial \psi}$ is simply the vector of demands scaled by the total demand for the neighborhood.

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

where n^k/n is the fraction of population k in the overall population of the metropolitan area.

The first factor $\frac{\partial \mathbf{D}}{\partial \mathbf{v}}$ is also straightforwardly derived. Noting $\gamma^k(\mathbf{x}_{it})$ the coefficient of social interactions for an individual with characteristics \mathbf{x}_{it} ,

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

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

(iv) 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 (26) 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 4 estimates

the change in lending standards between 2000 and 2006, and suggests that there were changes in underwriters’ sensitivity to the $\log(\text{price})$ of the house, the $\log(\text{income})$ 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 2), with added interactions between the 2006 year dummy and respectively: the $\log(\text{price})$, the $\log(\text{income})$ 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 4 illustrates the effect of such change in lending standards on mortgage approval, housing demand, and the distribution of housing prices. Figure 4 (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 ($\log(\text{prices})$ above 14, i.e. prices above \$1.2M). The rise in the approval probability is higher in neighborhoods with lower prices. Figure 4 (c) plots the implied partial equilibrium effect, where the increase in demand (vertical axis) at constant prices is plotted against the initial $\log(\text{price})$ 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 4 (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 4 (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 4(c) combined with the social interaction impacts of Figure 4(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 4(a)). The general equilibrium impact on prices is obtained following equation 26, and rescaled to obtain the equivalent annual change. As the change in lending standards is calculated over 2000-2006, the vertical axis of Figure 4(a) is the annual Compound Annual Growth Rate (CAGR).

The general equilibrium comparative statics of the estimated model is thus able to reproduce the negative correlation between the increase in neighborhood prices and their initial values that has been observed in the San Francisco Bay and in several other major cities during the boom (Landvoigt et al. 2015). The compression of the price distribution between 2000 and 2006, which is partly driven by differences in the intensity of borrowing constraints across households and neighborhoods, arises even if the estimation of the model did not use any information for the boom years beyond the year 2000.

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:

$$\frac{dD_{rjt}}{d\psi} = \underbrace{\frac{\partial D_{rjt}}{\partial \psi}}_{\text{partial equilibrium}} + \underbrace{\frac{\partial D_{rjt}}{\partial \log(\mathbf{p})} \cdot \frac{d \log(\mathbf{p})}{d\psi}}_{\text{general equilibrium}} + \underbrace{\frac{\partial D_{rjt}}{\partial \mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial \psi}}_{\text{preferences for same race neighbor}} \quad (35)$$

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 isolation and exposure indices,

as in Massey & Denton (1988) and Cutler, Glaeser & Vigdor (1999). The isolation index for a given racial subgroup (say, Hispanics of any race) measures the average fraction of same-race neighbors for an average member of that racial subgroup (e.g. the average Hispanic resident). Thus the

$$Isolation(Blacks)_t = \sum_{j=1}^J \frac{Black_{jt}}{Black} \cdot \frac{Black_{jt}}{Population_{jt}} \quad (36)$$

where the first factor $Black_{jt}/Black$ is the share of black population living in neighborhood j in year t ; and the second factor $Black_{jt}/Population_{jt}$ is the fraction black in neighborhood j in year t .

An increase in black isolation is due to a decline in black exposure to other races. Black exposure to Whites, for instance, is the average fraction of white neighbors for an average black household.

$$Exposure(Whites|Blacks)_t = \sum_{j=1}^J \frac{Black_{jt}}{Black} \cdot \frac{White_{jt}}{Population_{jt}} \quad (37)$$

In both definitions (36) and (37) black population in neighborhood j in year t 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 (36) with respect to the lending standard parameter ψ as in the previous section:

$$\begin{aligned} \frac{d}{d\psi} Isolation(Blacks)_t &= \frac{d}{d\psi} \left[\sum_{j=1}^J \frac{Black_{jt}}{Black} \cdot \frac{Black_{jt}}{Population_{jt}} \right] \\ &= \frac{Black}{Population} \cdot \frac{d}{d\psi} \left[\sum_{j=1}^J \frac{D_{black,jt}}{D_{jt}} \right] \end{aligned} \quad (38)$$

And similarly for the exposure of, say, Blacks to Whites. The change in isolation is proportional to the average ratio of the change in black demand to the change in total demand.

The resultant analysis of the impact of lending standards on neighborhoods' racial composition and on racial segregation are presented in Figure 5 and in Table 5. The starting point of such comparative statics is the spatial distribution of racial groups in the Bay area in 2000.

Figure 5(a) is a scatter plot where each black dot is a neighborhood. The vertical axis is the predicted change in white demand for the neighborhood *at given prices*, as a percentage of initial demand, while the horizontal axis is the fraction black in 2000. 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 predicted by the partial equilibrium changes, but remains substantially higher in majority non-black neighborhoods. Thus Figure 5(a) suggests that the exposure of Blacks to White should decline when lending standards are relaxed.

Figure 5(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 5(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 5(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 4 and the negative correlation between the fraction black and the $\log(\text{price})$. However, the relationship between the increase in $\log(\text{price})$ and the initial fraction black, displayed in Figure 5(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 5, 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 not predicted by partial equilibrium analysis: Table 5(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.

Overall Table 5(b) predicts that the relaxation of lending standards will lead to a decline in black isolation only through the increased black exposure to Hispanics and Asians.

Such segregation results, provided by the comparative statics of the estimated model, confirm the empirical finding of Ouazad & Ranci ere (Forthcoming). Using data for all U.S. metropolitan areas, Ouazad & Ranci ere (Forthcoming) shows that the relaxation of lending standards during the credit boom has led to a reduction in the exposure of Whites to Blacks. The estimated model provides a direct quantification of the relative importance of general vs. partial equilibrium effects, and shows how the compression of the price distribution 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 5 (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 5 (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 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.

Finally, the model focuses on household mobility within a city. A natural (but far from trivial) extension would be to incorporate mobility between cities as well.

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.
- Alonso, W. et al. (1964), ‘Location and land use. toward a general theory of land rent.’, *Location and land use. Toward a general theory of land rent.* .
- 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.

- Fujita, M. (1989), *Urban economic theory: land use and city size*, Cambridge university press.
- Geweke, J. (1996), ‘Monte carlo simulation and numerical integration’, *Handbook of computational economics* **1**, 731–800.
- 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 microeconomic 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.
- 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.
- Munnell, A. H., Tootell, G. M. B., Browne, L. E. & McEneaney, J. (1996), ‘Mortgage lending in boston: Interpreting hmda data’, *The American Economic Review* **86**(1), 25–53.
- Nevo, A. (2000), ‘Mergers with differentiated products: The case of the ready-to-eat cereal industry’, *The RAND Journal of Economics* pp. 395–421.
- Ouazad, A. & Ranci ere, 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.
- 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.
- Ross, S. L. & Yinger, J. (2002), *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement*, Vol. 1 of *MIT Press Books*, The MIT Press.
- 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: Predicted Mortgage Approval Rates in the Bay Area

This map presents, at the blockgroup level, predicted approval rate $E(\hat{\phi}_{ijt})$ for the mortgage application of an average borrower in 2000, based on the identification strategy of section 4.2 with the results of Table 2.

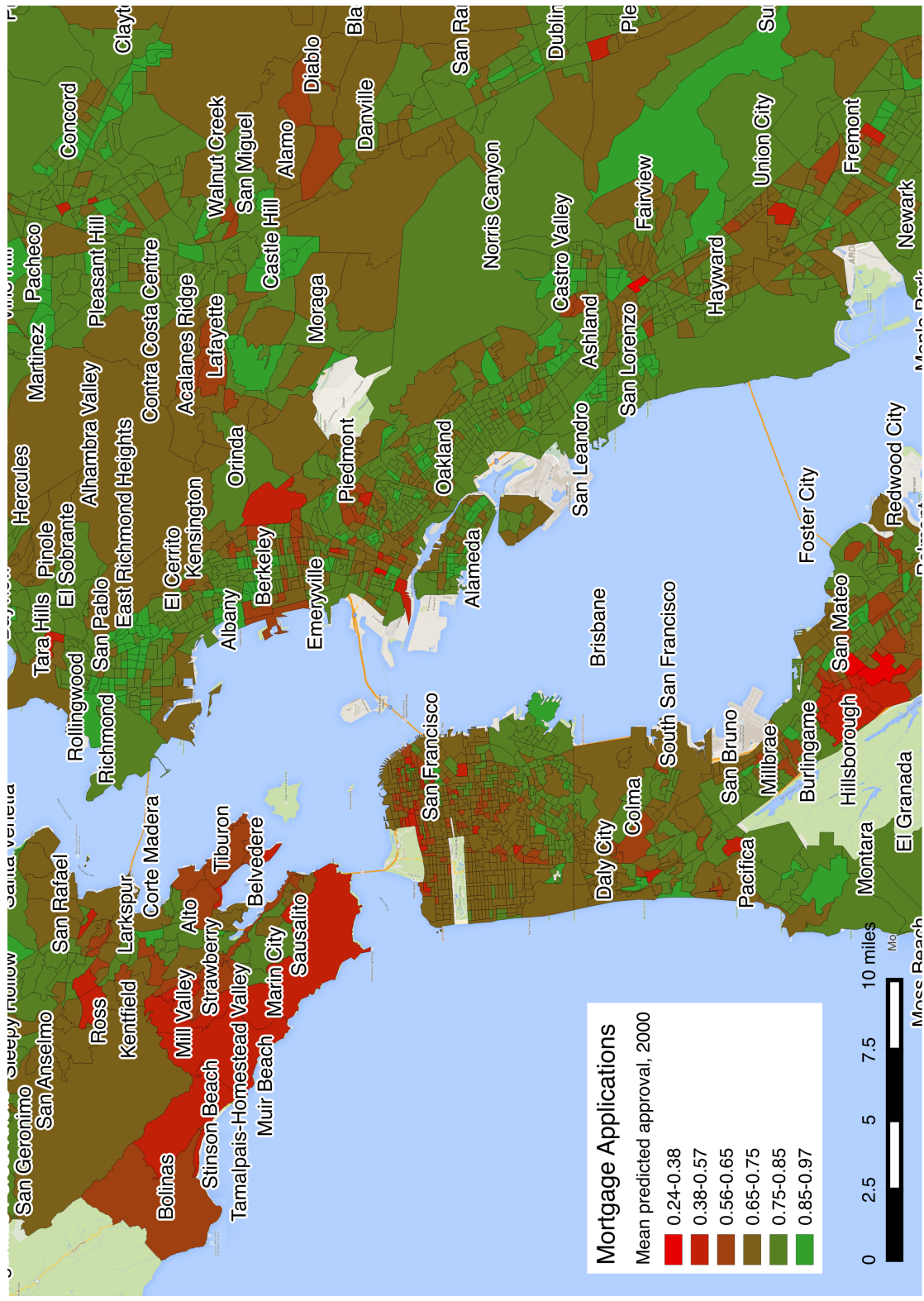


Figure 2: Demand Elasticity Distributions – Model with Borrowing Constraints

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.

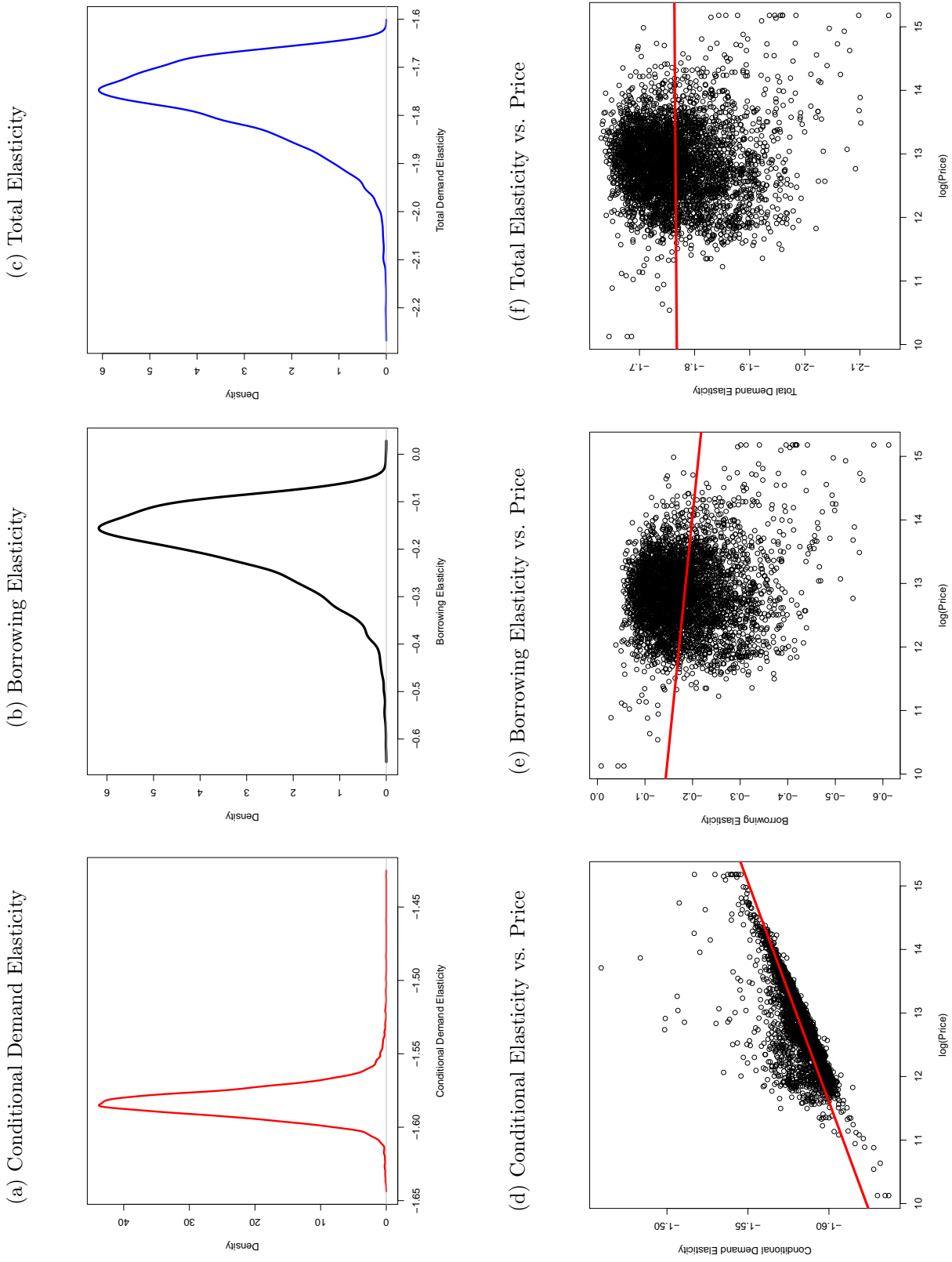
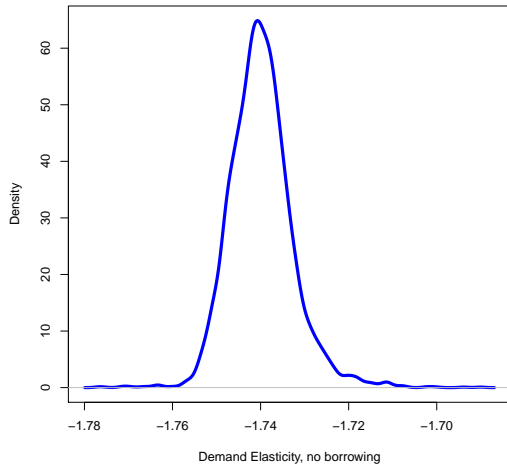


Figure 3: Demand Elasticity Distributions – Model without Borrowing Constraints

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)

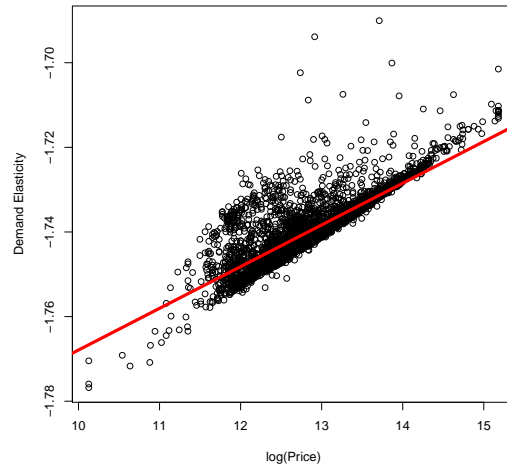


Figure 4: General Equilibrium Effects of Lending Standards on the Price Distribution

These figures illustrate the compression of the price distribution caused by the relaxation of lending standards. Figure (a) shows the relation between initial $\log(\text{price})$ in 2000 and $\log(\text{price})$ 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.

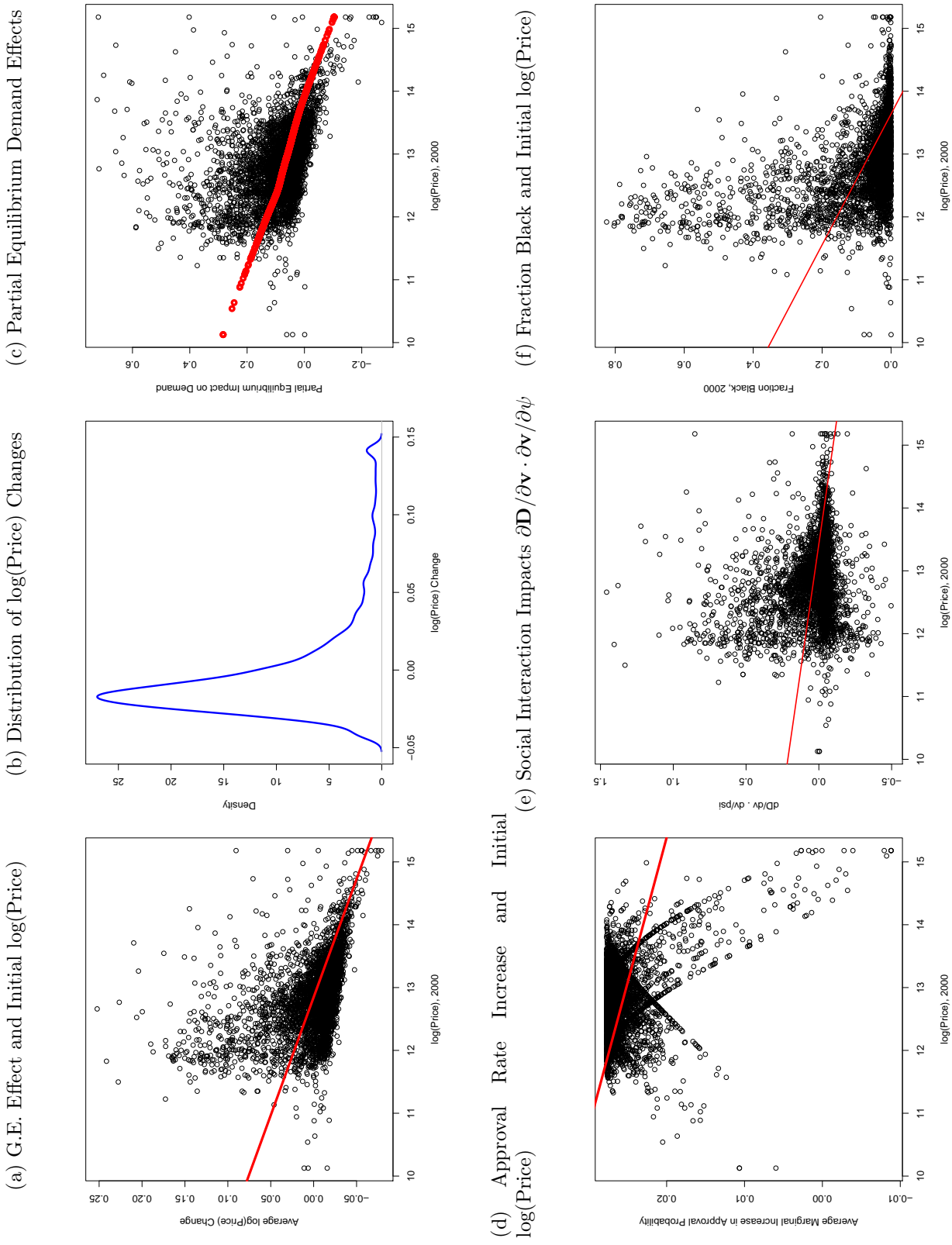
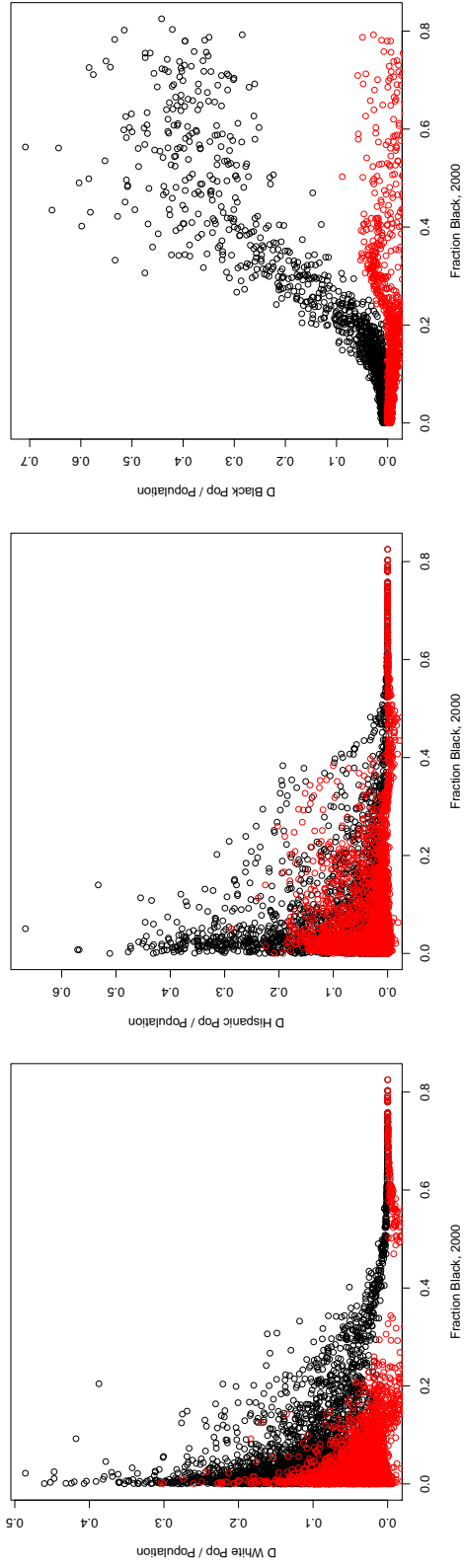


Figure 5: General Equilibrium Effects of Lending Standards on Segregation

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

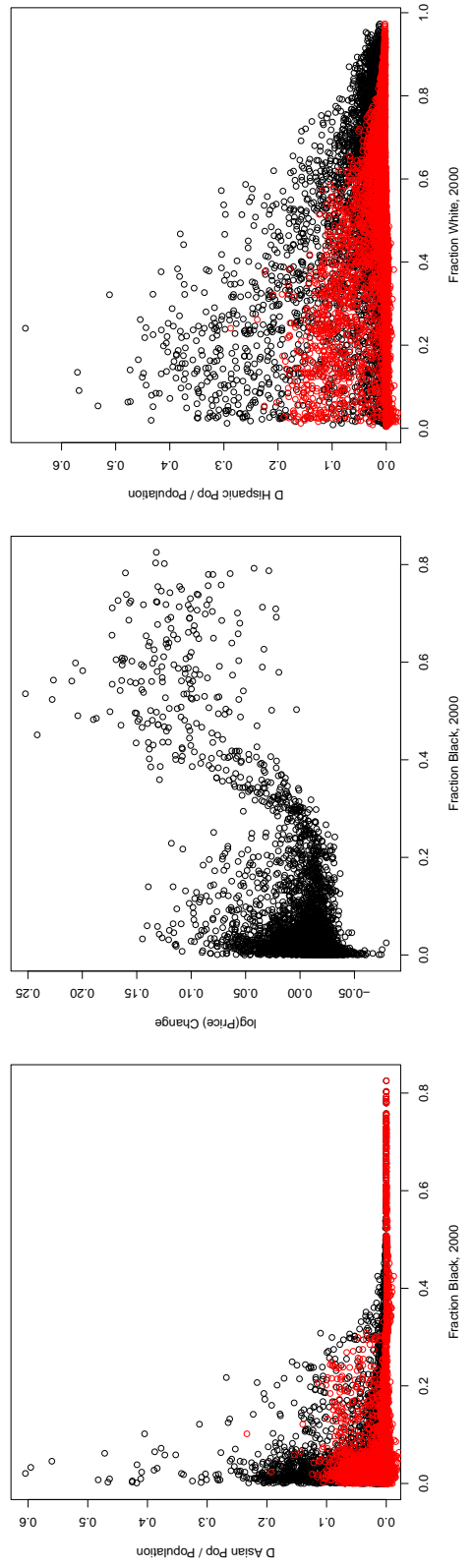


Table 1: Summary Statistics and Data Sources

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.
<i>Blockgroup Data</i>						
log(Median price)	12.757	12.737	0.688	7.601	18.258	11,611
Median number of rooms	5.156	5.062	1.254	0.074	9.100	11,611
Median age of structure	33.562	35.000	17.052	0.000	61.000	11,611
Frac. Black	0.078	0.025	0.140	0.000	0.948	11,611
Frac. Hispanic	0.278	0.170	0.289	0.002	1.818	11,611
Frac. Asian	0.175	0.116	0.163	0.000	0.983	11,611
log(Median household income)	11.050	11.053	0.542	6.787	12.429	11,611
Frac. college educated	0.439	0.426	0.211	0.000	0.972	11,611
Frac. more than high school	0.582	0.517	0.251	0.033	1.000	11,611
Frac. denied	0.132	0.141	0.052	0.000	0.615	11,611
<i>Micro Census Data</i>						
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
<i>Mortgage Application Data</i>						
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

Table 2: Mortgage Approval Equation – 1990–2010

The table presents estimation of the approval model (Specification 6). Columns (1)–(2) present the specification without neighborhood-specific unobservable controls. Columns (3)–(4) include neighborhood-specific unobservable controls. Columns (5)–(6) include bank-specific and other mortgage originator specific lending standard controls.

	Logit Coefficients	Marginal Probabilities	Logit Coefficients	Marginal Probabilities
Specification:				
	No Fixed Effect		Tract FE	
LTI Ratio	-0.239*** (0.008)	-0.028*** (0.001)	-0.252*** (0.008)	-0.028*** (0.001)
Black	-0.939*** (0.031)	-0.110*** (0.004)	-0.718*** (0.031)	-0.080*** (0.003)
Asian	-0.121*** (0.020)	-0.014*** (0.002)	-0.066*** (0.021)	-0.007*** (0.002)
Hispanic	-0.555*** (0.025)	-0.065*** (0.003)	-0.331*** (0.029)	-0.037*** (0.003)
Other Race	-0.467*** (0.025)	-0.055*** (0.003)	-0.410*** (0.024)	-0.046*** (0.003)
2000	-0.019 (0.023)	-0.002 (0.003)	0.026 (0.024)	0.003 (0.003)
2010	0.005 (0.027)	0.001 (0.003)	0.031 (0.031)	0.003 (0.003)
FHA Insured Loan	0.011 (0.026)	0.001 (0.003)	0.150*** (0.029)	0.017*** (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
Specification:				
	Bank FE		Tract and Bank FE	
LTI Ratio	-0.331*** (0.027)	-0.032*** (0.003)	-0.343*** (0.009)	-0.031*** (0.008)
Black	-0.819*** (0.034)	-0.078*** (0.003)	-0.630*** (0.036)	-0.058*** (0.015)
Asian	-0.157*** (0.034)	-0.015*** (0.003)	-0.123*** (0.022)	-0.011*** (0.003)
Hispanic	-0.446*** (0.049)	-0.042*** (0.005)	-0.285*** (0.031)	-0.026*** (0.007)
Other Race	-0.388*** (0.042)	-0.037*** (0.004)	-0.353*** (0.025)	-0.032*** (0.009)
2000	0.691*** (0.264)	0.066*** (0.025)	0.742*** (0.045)	0.068*** (0.018)
2010	0.104 (0.186)	0.01 (0.018)	0.133** (0.055)	0.012** (0.006)
FHA Insured Loan	-0.207 (0.156)	-0.020 (0.015)	-0.093*** (0.033)	-0.008*** (0.003)
Observations	189,470	189,470	188,578	188,578
Census Tracts	472	472	1,595	1,595
Pseudo R Squared	0.143	0.143	0.162	0.162

Table 3: 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	$\delta_{credit} - \delta_{no\ credit}$
log(Median price)	-1.584 (0.101)	-	-1.742 (0.095)	-	0.157 (0.010)
$\sigma(\log(\text{Median price}))$	0.138 (0.025)	-	0.113 (0.028)	-	-
Frac. Black	-11.928 (0.198)	-7.53% -\$ 32,871	-10.235 (0.187)	-5.88% -\$ 25,647	-1.693 (0.020)
× Black	15.758 (0.089)	2.42% \$ 10,555	13.754 (0.080)	2.22% \$ 9,698	-
Frac. Hispanic	-1.993 (0.201)	-1.26% -\$ 5,492	-2.066 (0.189)	-1.19% -\$ 5,177	0.072 (0.020)
× Hispanic	7.627 (0.077)	3.56% \$ 15,526	7.412 (0.073)	3.38% \$ 14,732	-
Frac. Asian	-2.936 (0.172)	-1.85% -\$ 8,091	-2.717 (0.162)	-1.56% -\$ 6,808	-0.218 (0.018)
× Asian	7.643 (0.087)	2.94% \$ 12,971	7.295 (0.081)	2.63% \$ 11,472	-
log(Median hh income)	0.636 (0.046)	2.18% \$ 9,499	0.622 (0.043)	1.94% \$ 8,448	0.014 (0.005)
× (log(Individual Household Income) - mean log(income))	0.954 (0.004)	6.02% \$ 26,290	0.934 (0.004)	5.90% \$ 25,739	-
Frac. College Graduate	0.562 (0.121)	0.35% \$ 1,549	0.529 (0.114)	0.30% \$ 1,326	0.033 (0.012)
Median Age of Structure	-0.018 (0.001)	-1.84% -\$ 8,026	-0.018 (0.001)	-1.67% -\$ 7,298	0.000 (0.000)
Median Number of Rooms	0.129 (0.182)	1.04% \$ 4,550	0.149 (0.172)	1.09% \$ 4,779	-0.02 (0.019)
Median Number of Rooms × log(Distance to CBD)	-0.019 (0.017)	0.53% \$ 2,306	-0.021 (0.016)	0.48% \$ 2,111	-
Academic Performance Index	0.061 (0.009)	0.39% \$ 1,681	0.025 (0.001)	0.14% \$ 626	0.036 (0.001)
× (log(Individual Household Income) - mean log(income))	-0.227 (0.003)	-1.43% -\$ 6,256	-0.1544 (0.003)	-0.97% -\$ 4,255	-
Observations	13,154	-	13,154	-	13,154
Neighborhoods (Blockgroups)	4,416	-	4,416	-	4,416
Years	3	-	3	-	3
R Squared	0.336	-	0.293	-	0.983

** : significant at 1%, * : significant at 5%, + : significant at 10%.

Standard errors clustered by block group in parenthesis.

Table 4: 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* (0.127)	+0.159 (0.140)
log(Income)	0.759* (0.110)	-0.503* (0.115)
Constant	-2.111 (0.962)	+2.967* (1.113)
Black or African American, Nonhispanic	-1.154* (0.086)	+0.345* (0.093)
Hispanic, Any Race	-0.716* (0.103)	+0.294** (0.117)
Asian, Nonhispanic	-0.286* (0.069)	+0.404* (0.075)
Observations		103,176
Census Tracts		1,315
Pseudo R Squared		0.0407

Table 5: 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	+3.99 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	-2.89 ppt

Appendix Table A: Mortgage Approval Equation – Additional Identification Strategies
– 1990–2010

The table presents estimation of the approval model (Specification 6) using a probit with Tract \times Year fixed effects. Columns (1)–(2) present the coefficients and marginal probabilities for the probit model with fixed effect. Columns (3)–(4) present results with the Tract \times Year fixed effects.

	Probit Tract F.e.		Probit Tract \times Year F.e.	
	(1) Coeff.	(2) M.P.	(3) Coeff.	(4) M.P.
Loan to Income Ratio	-0.136*** (0.004)	-0.029*** (0.001)	-0.139*** (0.004)	-0.029*** (0.001)
Black	-0.403*** (0.018)	-0.086*** (0.004)	-0.413*** (0.018)	-0.085*** (0.004)
Asian	-0.037*** (0.011)	-0.008*** (0.002)	-0.043*** (0.011)	-0.009*** (0.002)
Hispanic	-0.184*** (0.016)	-0.039*** (0.003)	-0.192*** (0.016)	-0.040*** (0.003)
Other Race	-0.224*** (0.013)	-0.048*** (0.003)	-0.228*** (0.013)	-0.047*** (0.003)
FHA Insured Loand	0.077*** (0.016)	0.016*** (0.003)	0.091*** (0.016)	0.019*** (0.003)
VA-guaranteed	0.098*** (0.038)	0.021*** (0.008)	0.108*** (0.039)	0.022*** (0.008)
FSA-RHS	-0.242 (0.168)	-0.051 (0.036)	-0.285* (0.172)	-0.059* (0.035)
Year 2000	0.020 (0.013)	0.004 (0.003)	0.012*** (0.003)	
Year 2010	0.020 (0.016)	0.004 (0.003)	0.524*** (0.020)	
Observations	208,206	208,206	206,277	206,277
Census Tracts	1,606		1,603	
Pseudo R Squared	.0434		.0587	

Appendix Table B: Mortgage Approval Equation for other Metropolitan Areas – 1990–2010

The table presents estimation of the approval model (Specification 6) in other major metropolitan areas between 1990 and 2010: San Francisco, Boston, Los Angeles, Houston, Miami metropolitan statistical areas.

	San Francisco MSA	Boston MSA	Los Angeles MSA	Houston MSA	Miami MSA
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
	M.P.	M.P.	M.P.	M.P.	M.P.
Loan to Income Ratio	-0.244*** (0.009)	-0.307*** (0.012)	-0.276*** (0.007)	-0.473*** (0.019)	-0.396*** (0.012)
Black	-0.028*** (0.001)	-0.626*** (0.049)	-0.443*** (0.031)	-0.721*** (0.048)	-0.446*** (0.038)
Asian	-0.081*** (0.003)	-0.120** (0.050)	0.174*** (0.023)	0.039 (0.059)	0.017 (0.062)
Hispanic	-0.008*** (0.002)	-0.335*** (0.065)	-0.011** (0.006)	0.005 (0.007)	0.002 (0.009)
Other Race	-0.345*** (0.029)	-0.031*** (0.006)	-0.018 (0.023)	-0.091 (0.060)	-0.244*** (0.029)
Year 2000	-0.411*** (0.024)	-0.519*** (0.034)	-0.334*** (0.021)	-0.558*** (0.047)	-0.369*** (0.026)
Year 2010	0.004 (0.003)	0.597*** (0.037)	-0.129*** (0.021)	0.455*** (0.131)	0.349*** (0.054)
FHA Insured Loan	0.003 (0.030)	0.313*** (0.039)	0.029*** (0.004)	-0.017*** (0.056)	-0.316*** (0.061)
VA-guaranteed	0.155*** (0.029)	0.003 (0.034)	0.000 (0.023)	0.547*** (0.056)	0.624*** (0.029)
FSA-RHS	0.018*** (0.071)	0.003 (0.076)	0.550*** (0.061)	0.073*** (0.076)	0.792*** (0.064)
Observations	-0.035 (0.311)	-0.428* (0.245)	0.007 (0.263)	0.040 (0.586)	0.261 (0.874)
Census Tracts	210,527	94,796	228,614	58,378	118,181
Pseudo R Squared	1,624	935	2,445	1,100	1,063
	.0465	.0713	.0429	.101	.0676

A Appendix: Equilibrium Properties

Proposition 1. (*Equilibrium Existence*) *There exists an equilibrium vector of prices and neighborhood demographics $(\mathbf{p}^* = (p_{jt}^*), \mathbf{v}^* = (\mathbf{v}_{jt}^*))_{j=1,2,\dots,J,t=1,2,\dots,T}$ that satisfies conditions (16) and (17).*

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:

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

and define the mapping: $\phi : \mathbb{R}^J \times [0, 1]^J \rightarrow \mathbb{R}^J \times [0, 1]^J$. 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 $(p_1^*, \dots, p_J^*, v_1^*, \dots, v_J^*)$ that satisfies the $2J$ equations that define the equilibrium.

We then consider the case of a perfectly inelastic housing supply. Set $s_j(p) = s_j \cdot p^{\eta}$

with η^s the elasticity of housing supply. For a given η^s consider the set of equilibrium vectors $E(\eta^s) = \{(p_1^*, \dots, p_J^*, v_1^*, \dots, v_J^*)\}$. We just showed that $E(\eta^s) \neq \emptyset$ for any $\eta^s > 0$. Consider a sequence of equilibrium vectors for a sequence of $\eta^s \rightarrow 0$. Such sequence of equilibrium vectors converges to a vector $(\mathbf{p}^* = (p_{jt}^*), \mathbf{v}^* = (\mathbf{v}_{jt}^*))_{j=1,2,\dots,J, t=1,2,\dots,T}$ 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 $D_j(p_1, \dots, p_J, p)$, 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

B.1 Estimation Algorithm

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 global nature of the optimum $\hat{\boldsymbol{\theta}}$ is assessed using profile GMM objective functions. Specifically, for a given estimate $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_K)$ and for each k , we plot the function $G(\hat{\theta}_1, \dots, \hat{\theta}_{k-1}, \theta_k, \hat{\theta}_{k+1}, \hat{\theta}_K)$ with θ_k going from $\underline{\theta}$ to $\bar{\theta}$.