Sorting out Housing*

John Finlay †   Trevor Williams ‡

June 24, 2020

Abstract

Housing expenditure shares decline with income. This causes spatial sorting. Unskilled workers avoid expensive cities because their lower incomes make them more sensitive to high housing costs, while skilled workers have a comparative advantage in living in expensive cities because they are less sensitive to costly housing. We estimate preferences which allow for nonhomothetic housing demand and embed them in a standard quantitative spatial model. We find that the increase in skilled workers' wages since 1980 relative to those of unskilled workers strengthened the sorting of skilled versus unskilled workers across cities. Our model explains 14% of the increase in city level sorting observed in the US between 1980 and 2010 through the interaction of the rising skill premium with nonhomothetic housing demand.

1 Introduction

Since 1980 the wages of skilled workers — defined as those with at least a four year college degree — have diverged from those of unskilled workers. Between 1980 and 2010, the skilled-unskilled wage gap grew by 69.4% (Figure 1, left panel). The canonical explanation for the growth in earnings inequality, which we refer to in this paper as the skill premium, is skill-biased technical change (SBTC) (Acemoglu and Autor 2011). Diverging earnings have also tracked diverging choices of where to live. Skilled workers increasingly cluster together in high wage, high rent cities in what Moretti (2012) has called the "Great Divergence."

The right panel of Figure 1 shows the trend in sorting since 1980. As a measure of sorting we propose the variance of the log ratio of skilled to unskilled workers across cities. The variance

*We thank Fabian Eckert, Brian Greaney, Michael Peters, Vítor Possebom, Jaehee Song, Conor Walsh, and seminar participants at Yale for their helpful comments, feedback, and suggestions.

†Yale University. robert.finlay@yale.edu

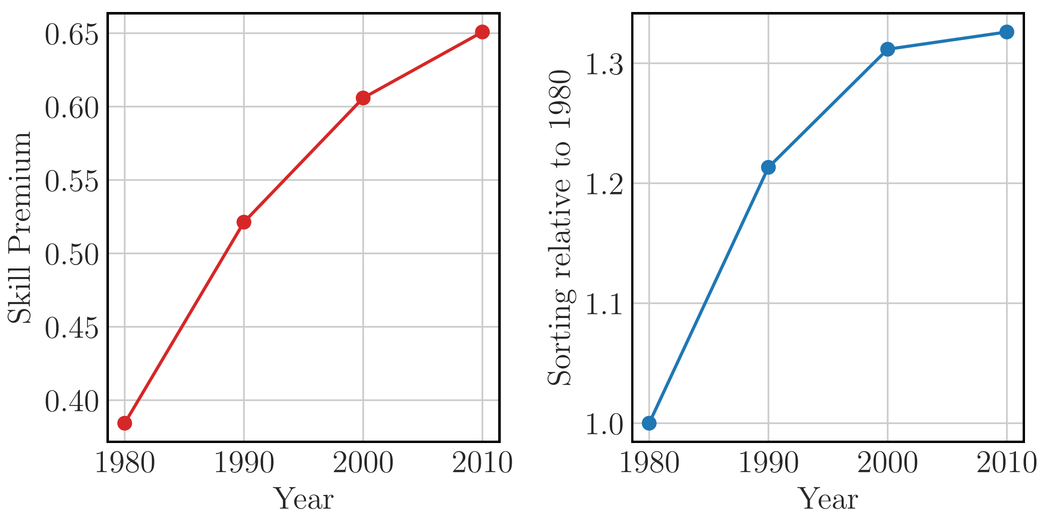
‡Yale University. trevor.c.williams@yale.edu

of the log skill ratio is zero when skilled workers are evenly distributed across space and grows without bound as they become more concentrated. This measure has the advantage that a proportional increase in the number of skilled workers in every location leaves it unchanged. It also maps naturally to the model we develop below. By 2010 the variance of the log skill ratio had increased by 32.6% relative to 1980.[1]

Most papers have explained the increase in sorting either through comparative advantage in production or through different valuations of urban amenities. By contrast, we develop and provide evidence for a novel explanation which operates neither through location-biased wage growth nor through exogenous differences in preferences. Instead, increased sorting is the result of the nationwide rise in the skill premium in Figure 1. Income-inelastic demand for housing links these two facts.[2]

A common view among journalists and policymakers is that expensive cities are increasingly

Figure 1: The Skill Premium and Sorting



Source: Source: Census 1980-2010. Skill premium defined as the log difference between average wages of skilled and unskilled workers. Sorting measured as the variance of the log skill share (blue) and the 80-20 ratio of the log skill share (purple), across cities. Skill share defined as the ratio of workers with at least a four year college degree to those without. See Appendix D for details of sample construction.

[1] Other measures of sorting display similar trends: for example, the 80-20 gap rose by 27.7%. In 1980, the 20th percentile city had two skilled workers for every ten unskilled workers, while the 80th percentile city had four skilled to ten unskilled. In 2010 those figures had risen to 3:10 at the 20th percentile and 8:10 at the 80th.

[2] In this paper, we focus on changes since 1980. However, in Figure A.1 we present data on sorting and the skill premium prior to 1980. The secular increase in sorting began only after 1980 and is robust to alternative choice of geographic units. In addition, sorting tends to track the skill premium over a long time horizon.

only for the rich and that rising housing costs squeeze out working- and middle-class households.[3] However current quantitative spatial models cannot accommodate this force since they typically rely on homothetic preferences. With homothetic preferences, rich and poor alike are equally exposed to housing costs. In this context it makes no sense to speak of one group being 'priced out' of expensive cities. In this paper we depart from the assumption of homothetic housing demand and allow for quantitatively realistic nonhomotheticities estimated from consumption microdata.

We show that the housing share—the fraction of total consumption expenditure dedicated to housing—is declining with total expenditure. Because poor households have high housing shares, they are more sensitive than rich households both to the level of housing prices and to a given change in housing prices. In equilibrium, the level of sorting depends positively on house prices, even conditional on amenities and wages. This nonhomotheticity implies that *any* change in the income distribution will affect the spatial distribution of skilled versus unskilled workers. This is true even when that change is skill and/or location neutral. In this paper we focus on the location neutral component of the rising skill premium.

To determine the quantitative importance of this channel we first estimate the income elasticity of housing demand. We parametrize preferences to be of the Price Independent Generalized Linear (PIGL) form. We find that housing is significantly income inelastic. Increasing a household's income by 10% reduces its housing expenditure share by 2.4%.

We develop a simple model in which the rising skill premium increases sorting whenever housing demand is income inelastic, and has no effect when the income elasticity is one. Then we extend this simple framework into a more realistic quantitative spatial model and use it to isolate the contribution of the rising skill premium to the pattern of increasing sorting shown in Figure 1. Our model is tractable and can be solved by standard 'hat algebra' methods. At the same time it generates rich patterns of sorting based on wages, amenities and housing costs.

We find that sorting rises relative to 1980 in our model, for two reasons. First, the rising skill premium increased the incomes of skilled workers and therefore decreased the share of expenditure they devoted to housing. This lower expenditure share reduced their sensitivity to housing costs and encouraged them to migrate towards expensive cities. In 1980 these cities were already skill intensive, so sorting rose. Second, the rising skill premium was a large positive shock to housing demand in skill intensive cities. Subsequent growth in house prices encouraged out-migration by unskilled workers who were particularly sensitive

[3] For example, Porter and Gates (2019) write in The New York Times that "As the highly educated have flocked to superstar cities, they have pushed housing prices way beyond the reach of people earning less."

to rising housing costs. Quantitatively, our model explains 14% of the observed increase in sorting between 1980 and 2010 through the interaction of income-inelastic housing demand with the rising skill premium.

In a second exercise we investigate the effects of changing the progressivity of the US tax system. We find that increasing the marginal tax rate to a Western European level would reverse nearly all of the increase in sorting since 1980, solely by compressing the income distribution.

Related literature

We contribute to the literature on the causes and consequences of the "Great Divergence."

The sorting of skilled and unskilled workers across space affects aggregate efficiency and equity. Fajgelbaum and Gaubert (2020) find that in a model with productivity and amenity spillovers the US is characterized by excessive spatial sorting, and that the efficient allocation moves skilled workers towards less skill-intensive cities. Chetty et al. (2014) have shown that areas with higher levels of income segregation also have lower levels of intergenerational mobility, suggesting that increased spatial sorting since 1980 has depressed intergenerational mobility. While we do not explicitly model these mechanisms, they motivate our decision to focus on the increasing sorting of skilled versus unskilled workers.

Turning to the causes of sorting, spatial models can generate sorting in three ways: wages, amenities, and prices. Most papers have emphasized the first two channels. Productivity growth which is both skill- and urban-biased motivates sorting based on wages: see Giannone (2019) for the role of skill biased technical change; Eckert (2018) for declining communication costs; and Eckert, Ganapati, and Walsh (2019) for tradable services. Amenities drive sorting in Diamond (2016). An influx of skilled workers causes a city's amenities to endogenously become more desirable. Skilled workers disproportionately enjoy these amenities, prompting even more in-migration of skilled workers, and so forth. We consider endogenous amenities in an extension of our model. Endogenous amenities do not independently link the rising skill premium to sorting, but rather amplify the effects of nonhomothetic housing demand.

We focus on the implications of house prices for sorting, as in Gyourko, Mayer, and Sinai (2013), Ganong and Shoag (2017), and Couture et al. (2019). These papers feature more stylized environments than ours: all three depend on a unit housing requirement which, as we will show, is a quantitatively significant departure from our estimated preferences. Couture et al. (2019) is conceptually the closest to our paper. The key difference is that they study *within*-city rather than *across*-city sorting. In relation to this literature, we present the first spatial equilibrium model which studies sorting across cities with quantitatively realistic nonhomotheticities.

Finally, we follow and contribute to a theoretical and empirical literature on nonhomothetic preferences. We use the PIGL preferences defined by Muellbauer (Muellbauer 1975; 1976) and recently deployed by Boppart (2014) in the context of structural change. Eckert and Peters (2018) are the first to embed these preferences preferences in a spatial model. In a key difference from our work, their paper studies aggregate growth while explicitly shutting down the sorting mechanism on which we focus. Closely related to our paper, Albouy, Ehrlich, and Liu (2016) estimate the income elasticity of housing demand using a spatial equilibrium condition. While we use different data sources and exploit only within-city variation, rather than across-city variation, we reassuringly find similar point estimates. Relative to Albouy, Ehrlich, and Liu (2016), we embed these estimated preferences in a general equilibrium model and use the model to shed light on aggregate trends in sorting.

Outline

In the next section, we estimate the income elasticity—the key parameter of the model. In section 3 we present a simple model and derive comparative statics. This is followed by the full quantitative model, which we calibrate in section 4. In section 5 we show that the estimated preferences are consistent with macro facts like the mean and variance of housing shares. Section 6 measures the effect of the skill premium on sorting and considers counterfactual tax regimes. Section 7 concludes.

2 Estimating the Income Elasticity of Housing Demand

In this section we estimate household preferences. In subsection 2.1 we briefly discuss our sources and take a first pass at the data. In subsection 2.2 we specify preferences for housing. Subsection 2.3 estimates the key parameter of our model.

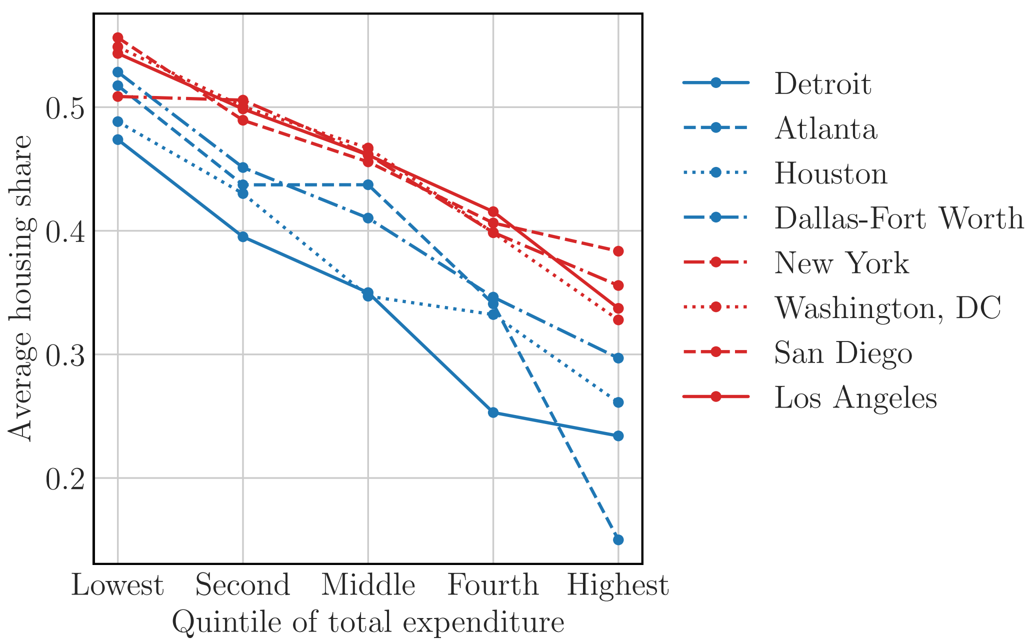
2.1 Data

To estimate the income elasticity, we will use consumption microdata from the Consumer Expenditure Survey (CEX). This is an annual, nationally representative survey that collects detailed information on expenditures on a large number of consumption categories. Particularly important for us is information on rent. Renters report their actual rental payments, while owners are asked “How much do you think your home would rent for monthly, unfurnished and without utilities?" The CEX also provides detailed information on incomes and savings and includes metropolitan statistical area (MSA) identifiers for subset of respondents. More information on the CEX can be found in Appendix D.

We take a first pass at the consumption microdata in Figure 2. For each year, we compute national quintiles of annual expenditure for all households. We then assign renting households to a quintile and compute the average housing share for each quintile × MSA cell, for several large metropolitan areas. Among the twenty or so metropolitan areas identified in the CEX, we select the four least expensive (blue) and four most expensive (red) metropolitan housing markets for Figure 2. We measure price using a metropolitan rental price index constructed by the BEA, which estimates the relative price of housing across cities controlling for the quantity and quality of the housing stock ("Bureau of Economic Analysis" 2020).

Two patterns are apparent. First, within each MSA, the rental share declines with expenditure. This suggests a negative relationship between housing share and total expenditure, holding local prices fixed. Second, there are systematic cross-MSA differences in the rental share for households in the same expenditure quintile. For example, households in San Diego

Figure 2: Housing share by MSA and expenditure



Source: CEX, 2006-2017, renters only. Housing share is defined as housing expenditure divided by total expenditure. The sample is cut into five groups ordered by total expenditure. We rank all cities identified in the CEX according to the BEA's regional price index for housing services. The plot features the four cheapest (blue) and four most expensive (red) cities.

have a higher rental share than households in Atlanta, at every level of expenditure. This suggests a positive relationship between housing share and local prices, holding total expenditure fixed. To accommodate these two patterns, it is necessary to allow for both income and price effects. We now specify the preferences used in the remainder of the paper and describe our estimation strategy.

2.2 Preferences

A household is characterized by its total consumption expenditure e and its location n. There are two goods — housing and a numéraire, freely traded composite consumption good. The price of housing, denoted by p_n, is location specific.

Households have PIGL preferences. A household with expenditure e in location n enjoys indirect utility $v(e, p_n)$ [4], given by

$$v(e, p_n) = \epsilon^{-1} e^{\epsilon} - \nu \psi^{-1} p_n^{\psi}. \tag{1}$$

To understand the role of each parameter it is helpful to write expenditure on housing, denoted by $R(e, p_n)$. Applying Roy's identity to (1) yields

$$R(e, p_n) = \nu e^{1-\epsilon} p_n^{\psi}. \tag{2}$$

$(1 - \epsilon)$ is the elasticity of housing expenditure with respect to total expenditure. To be consistent with the literature, we generally refer to $(1 - \epsilon)$ as the income elasticity, even though this is an abuse of language as, strictly speaking, it is the expenditure elasticity. ϵ is the key parameter of the model. A larger ϵ implies housing is more of a necessity. ψ is the elasticity of housing expenditure with respect to the price of housing and ν is a scaling factor. The housing share, which we define as $\eta \equiv R/e$, satisfies

$$\eta(e, p_n) = \nu e^{-\epsilon} p^{\psi} \tag{3}$$

The housing share is declining in total expenditure with elasticity ϵ and increasing in price with elasticity ψ, consistent with the patterns in Figure 2. When $\epsilon = \psi = 1$ our preferences imply a unit housing requirement. As $\epsilon, \psi \to 0$, we recover Cobb-Douglas preferences with a constant housing share given by ν.

[4]Conditions under which this is a valid indirect utility function are discussed in Appendix B.1.

2.3 Estimation

Taking logs of (2) yields a linear equation

$$\log R(e, p_n) = \log \nu + (1 - \epsilon) \log e + \psi \log p_n. \tag{4}$$

To bring (4) to the data, we allow the intercept term to vary with observable demographics. For household i in MSA n in year t, we denote these by the vector X_{int}, which includes the age, race and gender of the household head; the number of earners in the household; the household size; and a homeownership dummy. The homeownership dummy absorbs any systematic scale differences between self-reported rent by owners and actual rents paid by renters.[5] We replace the price term in (4) with a MSA $\times$ year fixed effect δ_{nt}, as the local housing price index is not observed in the consumption microdata. Allowing for an additive error u_{int}, we obtain an estimating equation

$$\log R_{int} = \alpha + \beta \log e_{int} + \Gamma' X_{int} + \delta_{nt} + u_{int} \tag{5}$$

where $\beta = (1 - \epsilon)$. Homothetic preferences would imply $\beta = 1$.

We can estimate (5) by OLS using log rental expenditure (or self-reported rental equivalent, for homeowners) and log total consumption expenditure from the CEX. Measurement error in consumption is a potential concern. Classical measurement error would bias our estimate of β towards zero, which would suggest nonhomotheticity even if housing consumption were in fact homothetic.

For this reason we instrument for consumption expenditure using the household's log pretax income, as in Aguiar and Bils (2015). The identifying assumptions are that, conditional on X_{int} and location, pretax income is (1) correlated with the true level of consumption and (2) excludable from the structural equation. This holds in all standard models of consumption. The first follows because households consume out of permanent income and current income is mechanically correlated with permanent income. The second holds because current income is independent of rental expenditure conditional on the true level of total consumption.

Table 1 displays the results of estimating (5). Column (1) shows OLS results without MSA fixed effects. The coefficient on total consumption expenditure indicates that housing demand is income inelastic. Column (2) instruments for consumption using pre-tax income. The coefficient rises relative to column (1), which is consistent with classical measurement

[5] See, e.g., Garner and Short (2009), who show that the level of self-reported rents is too high relative to the national income accounts. They also find the level is higher than what would be predicted with a simple model of selection into homeownership.

error in consumption. Columns (3) and (4) add MSA fixed effects and drop households whose MSA was not reported. In particular we focus on column (4), the IV result. The coefficient on total consumption expenditure falls relative to the value in column (2). This is not surprising. Without controlling for MSA, the tendency for higher income households to sort into higher price cities confounds the within-MSA patterns we aim to capture.[6] The coefficient in column (4) is significantly different from 1 at the 1% level, allowing us to reject homothetic housing demand. It is also significantly different from 0, indicating a unit housing requirement does not fit the data well.

Estimation using the CEX has an important drawback. As documented by Aguiar and Bils (2015), the CEX suffers from underreporting of total consumption expenditure. This underreporting is strongest for high income households. At the same time, underreporting is less severe for some expenditure categories. In particular Attanasio and Pistaferri (2016) note that underreporting is less of a problem for rental expenditures. These two facts are likely to bias our estimate of β upwards, since rent will appear to rise quickly with total expenditure simply because the latter suffers from stronger underreporting as we move up the income distribution.

To address this concern, we construct an alternative measure of consumption expenditure. We measure consumption as after tax income minus total savings reported in the CEX. See

Table 1: Income elasticity, CEX

	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) 2SLS
Log expenditure	0.725***	0.881***	0.642***	0.764***	0.532***
	(0.009)	(0.011)	(0.010)	(0.013)	(0.035)
MSA FE			✓	✓	✓
R^2	0.63	0.61	0.71	0.70	0.40
First-stage F-statistic	-	4995.3	-	2344.5	39.7
N	19808	19808	8811	8811	1770
No. of clusters	251	251	239	239	167

Source: CEX, 2006-2017 (2006-2013 for column (5))

Note: Dependent variable is log rent or log rental equivalent. Controls include family size, number of earners, sex, race, age of household head, year, and homeowner dummy. Columns (1)-(4) use reported consumption and instrument with log pre-tax income. Column (5) imputes consumption from reported income and savings for renters only and instruments with years of schooling dummies. Standard errors clustered at year × MSA level. All regressions are weighted by CEX sample weights.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

[6] The model introduced in Section 3 will predict this pattern.

Appendix D.2 for details. Our measure of savings follows Aguiar and Bils (2015), who note this while it is noisy this measure of consumption captures trends in aggregate consumption and consumption inequality better than reported expenditure. When using this measure of consumption, we restrict the sample to renters, as it is challenging to account for savings for mortgage holders.

Column (5) shows results using this measure of consumption expenditure with MSA fixed effects. Consistent with the argument above, our point estimate is lower using this measure than in column (4). We therefore regard the value in column (4) as an upper bound on the income elasticity of housing demand. We use this value going forward, noting that it is conservative relative to the estimate in column (5). This implies $\epsilon = 0.24$.

2.4 Summary

We reject the Cobb-Douglas specification usually employed in the spatial literature, as well as the unit housing requirement sometimes used. Instead we find that housing demand is moderately income inelastic. Our estimate is within the range of estimates generally found in the literature (see Appendix Table A.3 for a comparison). We are particularly close to Lewbel and Pendakur (2009), whose estimates are consistent with ϵ equal to 0.28, as well as Albouy, Ehrlich, and Liu (2016), who estimate ϵ to be approximately 0.35. Other papers (e.g Davis and Ortalo-Magné (2011), Aguiar and Bils (2015)) report estimates closer to the homothetic case of $\epsilon = 0$. The key differences between those studies and ours are (i) we control for household location using MSA fixed effects in order to avoid conflating differences in prices across markets with differences in incomes, and (ii) we allow for the possibility of systematic mismeasurement in homeowner's estimates of the rental values of their homes using a homeownership dummy.

2.5 Robustness and Extensions

To test whether the data support the log-linear specification in (5), we estimate a semi-parametric alternative. We regress log rental expenditure on a set of 50 dummies for bins of total consumption, after partialling out controls for both variables. The data are well-approximated by a log-linear function (see Figure A.2). While we cannot completely rule out the role of supply-side factors in generating consumption patterns that appear nonhomothetic—for example, market power among landlords of low-income renters, or indivisibility of the housing stock—we conclude that a constant elasticity function with income elasticity smaller than one is a good characterization of the data.

In Appendix A we consider numerous alternative specifications. In Table A.1 we consider alternative instruments and samples and find quantitatively similar results. In Table A.2 we repeat the exercise using the PSID; see Appendix D for more details on this dataset. The PSID has included detailed information on most consumption categories since 2005. The PSID does not contain MSA identifiers, but rather state identifiers and a dummy for living in an urban area. This is likely to bias our results: for example, a household in Buffalo and Manhattan would be classified into the same geography. Nevertheless we note that the PSID results are relatively close to the baseline estimates and are consistent with a significant departure from an income elasticity of unity.

3 Model

To study the effect of increases in the aggregate skill premium on local sorting, we need a model. To this end we build a simple quantitative spatial model of location choice.

3.1 Simple Model

Production

There are two types of household, skilled and unskilled, with types denoted by $i = s, u$. Households supply labor to tradable goods producers in the location in which they live, denoted by n. These firms are perfectly competitive and produce using skilled and unskilled labor with the production function

$$F_n(l_{sn}, l_{un}) = z_n(A \cdot l_{sn} + l_{un}). \tag{6}$$

Note that skilled and unskilled labor are perfect substitutes, and that their relative productivities do not vary across locations. This implies that the skill premium is exogenous and equal to A in every location. Households do not save, so wages w_{in} are exactly equal to expenditure e_{in}. Expenditures and wages therefore satisfy

$$e_{sn} = w_{sn} = z_n A \tag{7}$$

$$e_{un} = w_{un} = z_n. \tag{8}$$

Location Choice and Preferences

We use a simple model of location choice based on Fréchet distributed preference shocks (Redding 2016). Household preferences are represented by the indirect utility function introduced in (1), which we restate here for completeness

$$v(e, p_n) = \epsilon^{-1} e^{\epsilon} - \nu \psi^{-1} p_n^{\psi}. \tag{9}$$

Each household ω draws an n-vector of idiosyncratic location preferences $\boldsymbol{\zeta}(\omega)$ from independent Fréchet distributions. The distribution of draws in n has scale B_n and shape θ, so the c.d.f. is

$$F_n(\zeta) = \exp(-B_n \zeta^{-\theta})$$

The preference shocks are multiplicative, so that the utility of household ω as

$$U_{in}(\omega) = \max_n \zeta_n(\omega) v(e_{in}, p_n).$$

Standard properties of the Fréchet distribution (e.g. Redding (2016)) imply that the number of households of type i in location n is given by

$$l_{in} = \frac{v(e_{in}, p_m)^{\theta} B_n}{\sum_m v(e_{im}, p_m)^{\theta} B_m} L_i \tag{10}$$

where L_i is the exogenously given national population of type-i households.

We refer to θ as the migration elasticity. A model with perfect mobility would have $\theta = \infty$, while $\theta = 0$ corresponds to a model where households sort only on the basis of amenities. Note that we impose that amenities do not differ by type i, an assumption we relax later. Equation 9 implies the expenditure of type i in location n on housing, denoted R_{in}, is

$$R_{in} = \nu e_{in}^{1-\epsilon} p_n^{\psi}. \tag{11}$$

Housing Supply

In the simple model, housing is supplied perfectly elastically at an exogenous price p_n in each location.

Equilibrium

Given parameters $\epsilon, \psi, \nu, \theta, A$ and fundamentals $(z_n, B_n, p_n)_n, (L_i)_i$, an equilibrium is a vector of populations l_{in} wages w_{in} and total expenditures e_{in} satisfying (7), (8), (9) and (10).

3.2 Analytical Results

In the simple model described above we can characterize changes in sorting analytically.[7] We do so by considering a small increase in the skill premium, $d\log A \geq 0$. Our interest is in the log skill ratio in each location n, which we denote by $\log s_n$ and define as the log ratio of skilled to unskilled workers in location n

$$\log s_n = \log\left(\frac{l_{sn}}{l_{un}}\right).$$

We measure sorting by the variance of the log skill ratio, S, defined by

$$S = \text{Var}\left(\log s_n\right).$$

which is the statistic plotted in Figure 1. We begin by characterizing the relationship between prices p_n and the skill ratio s_n.

Lemma 1 *Define the ratio of prices to local productivity as*

$$\eta_n = p_n^{\psi} z_n^{-\epsilon}.$$

The log skill ratio $\log s_n$ *is an increasing function of* η_n*, given by*

$$\log s_n = \log\xi + \theta\log\left(\frac{\psi A^{\epsilon} - \epsilon\eta_n}{\psi - \epsilon\eta_n}\right)$$

where ξ *is a function only of aggregate terms.*

The intuition for this result is that as prices rise relative to income, they increase the gap between skilled and unskilled utilities and so encourage sorting. In this model with the skilled-unskilled wage ratio constant across locations and no differences in amenities between types, differential sensitivity to housing prices is the only motive for sorting. In a richer environment the level of sorting would still be positively correlated with house prices, conditional on wages and amenities. The strength of this force increases with ϵ, which indexes the extent of nonhomotheticity in preferences. In the limit as $\epsilon \to 0$, preferences are homothetic, sorting on prices vanishes, and the log skill ratio does not vary across locations.

The next proposition characterizes how changes in the log skill ratio, $d\log s_n$, vary with the level of the skill ratio, s_n.

[7] Proofs of all the statements in this section can be found in Appendix B.2.

Proposition 1 *The change in the log skill ratio $d\log s_n$ is an increasing function of s_n given by*

$$d\log s_n = \beta + \theta\epsilon \left(1 - \left(\frac{\xi}{s_n}\right)^{\frac{1}{\theta}}\right)\left(1 - A^{-\epsilon}\right)^{-1} d\log A$$

where β and ξ depend only on aggregate terms and $\xi < s_n$ for all n. Therefore

$$dS \geq 0.$$

When preferences are homothetic, there is no change in sorting. That is, as $\epsilon \to 0$ $dS \to 0$.

Proposition 1 tells us that increases in the skill premium amplify pre-existing patterns of sorting. The reason is that as the income of skilled workers rises, they decrease their share of expenditure on housing. This makes them less sensitive to housing costs and encourages movement towards expensive cities. Because expensive cities are initially more skill intensive by Lemma 1, sorting rises.

3.3 Quantitative Model

To take the simple model above to the data, we enrich it on several dimensions: imperfect substitution between skilled and unskilled workers; inelastic housing supply; progressive taxation; and type-specific amenities.

We replace the production function (6) with a CES production function

$$F_n(l_s, l_u) = \left((A_{sn}l_s)^{\frac{\sigma-1}{\sigma}} + (A_{un}l_u)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$$

where σ is the elasticity of substitution between skilled and unskilled workers. Note we allow the productivities A_{in} to vary freely. This implies wages

$$w_{in} = A_{in} l_{in}^{\frac{-1}{\sigma}} \left((A_{sn}l_{sn})^{\frac{\sigma-1}{\sigma}} + (A_{un}l_{un})^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}}. \tag{12}$$

We endogenize housing prices with a reduced form model of housing supply similar to Hsieh and Moretti (2019). The price of housing in location n is given by

$$p_n = \Pi_n (HD_n)^{\gamma_n} \tag{13}$$

where HD_n is (physical) housing demand in n and Π_n is an exogenous price shifter. γ_n is the (inverse) elasticity of housing supply, which is allowed to vary by location. Housing demand is the sum of housing consumption by both types of household

$$HD_n = \sum_i R_{in} l_{in} p_n^{-1}. \tag{14}$$

We allow for progressive taxation following Heathcote, Storesletten, and Violante (2017) and define the expenditure (which equals after tax income) of a worker of type i in location n as

$$e_{in} = \lambda w_{in}^{1-\tau}. \tag{15}$$

τ determines the progressivity of the tax system and λ is chosen so that the government budget balances.

Finally, we leave workers' preferences within a location unchanged relative to the simple model, but allow amenities B_{in} to vary with worker type. Equation 10 then becomes

$$l_{in} = \frac{v(e_{in}, p_m)^\theta B_{in}}{\sum_m v(e_{im}, p_m)^\theta B_{im}} L_i. \tag{16}$$

Equilibrium

Given parameters and fundamentals $(A_{in}, B_{in})_{i,n}, (\Pi_n)_n, (L_i)_i$, an equilibrium is a vector of populations l_{in}, wages w_{in}, total expenditures e_{in}, housing expenditures R_{in}, housing demands HD_n, and prices p_n satisfying equations (12), (15), (16), (9), (11), (13) and (14). This system can be solved using the 'hat algebra' method used in the international trade literature (Dekle, Eaton, and Kortum 2007). For a formal statement please see Appendix B.3.

Finally, we provide a neutrality result for this quantitative model.

Proposition 2 *Suppose preferences are homothetic and represented by an indirect utility function of the form*

$$v(e, p) = eu(p)$$

for some function u.[8] *Then a proportional increase in A_{sn} in all locations has no effect on sorting, i.e $dS = 0$.*

Proof. Please see Appendix B.3.1 ∎

[8] Any within-location preferences over housing and non-housing consumption which are homothetic can be represented by such an indirect utility function. For example Cobb-Douglas preferences imply $u(p) = p^{-\eta}$ for some constant η.

It is easy to see that inflating the productivity of skilled workers by a constant factor in every location—i.e. location-neutral, skilled biased technological change—will not affect sorting when preferences are homothetic. Then prices do not determine the relative attractiveness of a location to skilled versus unskilled workers and so do not cause sorting. Proposition 2 is useful because it implies that any increases in sorting observed in our quantitative model are ultimately the result of nonhomothetic housing demand.

4 Calibration

We set ϵ at the value obtained in Section 2. The remaining parameters to be calibrated are the price elasticity, ψ; the migration elasticity, θ; the location specific inverse supply elasticities, γ_n; the elasticity of substitution between skilled and unskilled labor in production, σ; and the tax progressivity, τ. We choose units such that the scaling factor ν is normalized to one. Before describing the calibration of the model, we give a brief overview of the data sources used. Because the data are standard, we refer the interested reader to Appendix D for more details on the data, variable definitions, and sample construction.

4.1 Data

Location-level information on wages, rents and employment are computed using 5% samples from the decennial Census, 1980-2000, plus the 3% sample from the 2009-2011 ACS (Ruggles et al. 2020). We work at the Metropolitan Statistical Area (MSA) level of geography. An MSA is defined as a set of counties with strong economic ties and a dense urban core, which corresponds closely to the notion of a region in our model. After ensuring locations are consistently defined across years we are left with a panel of 219 MSAs plus 50 residual state locations. Our Census sample constitutes prime-age adults who report strong labor force attachment.

We adjust Census rents so that the average ratio of rents to post-tax income in the Census matches the average ratio of rents to expenditure in the CEX in each year. The adjustment makes the Census income data, which is estimated from an individual-level sample, consistent with the Census rent data, which is estimated from a household-level sample.

4.2 Parameters

Estimating the price elasticity To estimate the price elasticity of housing demand ψ we use Census data on incomes and rents. This is preferable to using the CEX as our Census data covers 269 locations and therefore provides substantial variation in price across housing markets. By contrast the CEX reports information for only a handful of large cities. As our measure of price p_n we use the BEA Metro Area Regional Price Parities, discussed in Section 2.

We start by taking logs of equation 11 and replacing expenditures with wages using (15)

$$\log R_{in} = \log \nu + \log \lambda + (1-\tau)(1-\epsilon) \log w_{in} + \psi \log p_n.$$

We define 'excess rent' $\widetilde{R}_{in}$ by

$$\log \widetilde{R}_{in} = \log R_{in} - (1-\tau)(1-\epsilon) \log w_{in}$$

so that we obtain

$$\log \widetilde{R}_{in} = \log \nu + \log \lambda + \psi \log p_n.$$

Intuitively, rents can be high in an MSA because its residents have high total expenditures or because prices are high (or both). Excess rent purges actual rents of total expenditure, leaving just the price term on the right hand side of the estimating equation. Replacing the constants with an intercept and allowing for an additive error term u_{in} gives us an estimating equation

$$\log \widetilde{R}_{in} = \alpha + \psi \log p_n + u_{in}. \tag{17}$$

We focus on the year 2010 as the BEA price data is only available in that year and is not comparable across time. The log of the BEA index is equal to $\log p_n$ (up to an additive constant). We construct 'excess rent' using Census data on rental expenditures and the wages of renters. We also use the value of ϵ estimated in Section 2 and the value of τ calibrated below.

Column (1) of Table 2 shows the results of estimating (17) by OLS. OLS estimation of (17) is likely to be biased for several reasons. Demand shocks which raise rental expenditures and simultaneously raise prices via the supply side (equation 13) will tend to bias OLS estimates of ψ upwards. On the other hand measurement error in prices will bias ψ towards zero. To avoid these problems, column (2) instruments for p_n using measures of land unavailability and land use regulation from Saiz (2010). This procedure yields an unbiased estimate of ψ if these instruments are uncorrelated with unobserved shifters of rental expenditure. Note that these instruments remain valid even if they are correlated with local amenities; although these amenities are likely to result in a positive demand shock as the location draws in more workers, they only impact expenditure via the price term in 17. The point estimate in column

Table 2: Price elasticity of demand

	(1)	(2)
Log price	0.482***	0.549***
	(0.018)	(0.040)
R^2	0.60	0.58
First-stage F-statistic	-	24.5
N	420	386
No. of clusters	43	42

Note: All regressions weighted by 2010 employment. Standard errors clustered at the state level. Instruments in (2) are share of land unavailable and Wharton Land Use Regulation Index.

Source: Census and BEA Metro Area Regional Price Parities

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

(2) is 0.55, which is the value we choose for ψ. The price elasticity is significantly different from both 0 and 1, indicating that neither Cobb-Douglas nor a unit housing requirement fit the data well.

In Appendix C.1, we repeat the exercise using price data from Zillow, a real estate listings website. The results are very similar.

Calibrating the migration elasticity We calibrate the migration elasticity, θ, by matching the long-run employment elasticity estimated by Hornbeck and Moretti (2019).[9] They run IV regressions

$$\log Y_{n,2000} - \log Y_{n,1980} = \pi^Y (\log A_{n,1990} - \log A_{n,1980}) + \Gamma' X_n + \upsilon_n$$

where Y is an outcome, A is a measure of TFP recovered from plant-level data, and X is a vector of controls.

We consider the outcomes for earnings and employment.[10] The results with the full set of instruments (see Hornbeck and Moretti (2019)'s Table 4, column 8, panels A and B) imply an employment elasticity of

$$\frac{\hat{\pi}^{\text{employment}}}{\hat{\pi}^{\text{earnings}}} = \frac{3.35}{1.54} = 2.18$$

[9]This strategy is borrowed from Greaney (2020).

[10]In principle, this employment elasticity conflates migration with labor supply. Households in our model supply labor inelastically, so that we are interested in the migration margin only. Hornbeck and Moretti (2019) find that most of their estimated employment response is from migration rather than increased labor supply of incumbent workers.

We solve for the value of θ that produces the same response in our model.[11] This procedure implies $\theta = 11.88$. This value is not directly comparable to those reported in the literature because of our use of PIGL preferences. To make a rough comparison, we calculate the elasticity of location choice with respect to nominal wages for a worker with a housing expenditure share of 0.30, the average in 1980, using (9) and(16). We obtain a value of 3.28, similar to the values usually reported. For example, Hsieh and Moretti (2019) calibrate a value of 3.33. For more details of the calibration of θ, see Appendix C.2

Calibrating the housing supply elasticities Next we calibrate the set of housing supply elasticities, γ_n. The housing supply equation (13) is specified in terms of physical quantity of housing, HD_n, which is not observed. To obtain an estimating equation, rewrite (13) with HD_n expressed in terms of price and expenditure

$$p_n = \tilde{\Pi}_n \left(\sum_i R_{in} l_{in} \right)^{\chi_n}$$

where $\chi_n = \gamma_n/(1+\gamma_n)$ and $\tilde{\Pi}_n = \Pi_n^{1/(1+\gamma_n)}$. Taking logs and differencing over time yields an equation which is linear in parameters

$$\Delta \log p_n = \Delta \log \tilde{\Pi}_n + \chi_n \Delta \log \left(\sum_i R_{in} l_{in} \right).$$

Similar to Saiz (2010), we parametrize χ_n as a function of geographical and regulatory constraints

$$\chi_n = \chi + \chi_L \Lambda_n + \chi_R WRLURI_n$$

where Λ_n is Saiz (2010)'s measure of the fraction of land unavailable for development and $WRLURI_n$ is the Wharton Residential Land Use Regulation Index developed by Gyourko, Saiz, and Summers (2008). This yields an estimating equation for χ, χ_L and χ_R

$$\Delta \log p_n = \Delta \log \tilde{\Pi}_n + (\chi + \chi_L \Lambda_n + \chi_R WRLURI_n) \, \Delta \log \left(\sum_i R_{in} l_{in} \right). \tag{18}$$

We use data on rents and wages from the 1980 and 2010 Censuses to construct $\Delta \log p_n$ using (11). To circumvent the endogeneity of rents and population we follow Diamond (2016) and

[11] In the implementation, we only simulate skill-neutral TFP shocks to remain consistent with Hornbeck and Moretti (2019). We draw these shocks from a lognormal distribution whose mean and variance matches the mean and variance of TFP shocks we recover from the data.

instrument using Bartik instruments for labor demand. We construct these instruments from the Census. These IVs predict regional wage changes using a region's sectoral composition in 1980 interacted with national wage trends. Denoting skill levels by i, regions by n, years by t, and industries by k, we define the Bartik IV as

$$Z_{int} = \sum_k w_{i,k,-n,t} \cdot \frac{l_{i,k,n,1980}}{l_{i,n,1980}} \tag{19}$$

where $w_{i,k,-n,t}$ is the log average wage of workers in group i and industry k and year t, in all regions *except* n. $l_{i,k,n,1980}$ is employment of group i in industry k and region n in 1980, and $l_{i,n,1980} = \sum_k l_{i,k,n,1980}$ is total employment of group i in region n in 1980. The identifying assumption is that a location's sectoral employment shares in 1980 are not correlated with future shocks to the housing supply shifter $\Delta\Pi_n$, but do impact employment and rental expenditures. To identify χ_L and χ_R we use the interaction of these Bartik shocks with Λ_n and $WRLURI_n$. We then set $\gamma_n = \chi_n/(1-\chi_n)$.

The population-weighted average of the γ_n obtained using this procedure is 0.37. This is squarely within the range of other estimates of this parameter in the literature: Saiz (2010) finds a higher value of 0.78, while Diamond (2016) finds a lower value of 0.21.

Calibrating the elasticity of substitution The production side of the model is standard and we externally calibrate σ, the elasticity of substitution between skill groups in the production function. We set $\sigma = 3.85$ to match the estimate from Card (2009).[12] This is somewhat higher than many estimates, which are typically in the range of 1.5-2.5 (cf. Katz and Murphy (1992)). We prefer the estimate in Card (2009) for two reasons. One is that Card exploits cross-sectional rather than time-series variation, which is conceptually more appropriate in our setting. Second is that estimates of the elasticity of substitution is sensitive to changes in worker quality (Carneiro and Lee 2011; Bowlus and Robinson 2012; Hendricks and Schoellman 2014). Bowlus et al. (2017) argue that correct measurement of worker quality implies higher substitutability between worker types. In particular, the quality of college workers has risen over time relative to non-college workers, so that relative skill prices move less than relative wages for a given shift in relative supply. Bowlus et al. (2017) report estimates of σ in the range of 4-5.5.

Calibrating the tax system To calibrate the progressivity parameter τ we follow the strategy of Heathcote, Storesletten, and Violante (2017). From (15), log post-tax income for

[12]See Table 5, column 7 in Card (2009) for the negative inverse elasticity of -0.26.

individual i in year t is equal to

$$\log y_{it} = \log \lambda_t + (1 - \tau) \log w_{it} \tag{20}$$

The tax system is not location specific, so we drop the location subscript. Allowing for an idiosyncratic error, (20) suggests that regressing log post-tax income on log pre-tax income and a year fixed effect yields a consistent estimate of $1 - \tau$. We implement this using the PSID for 1980, 1990, 2000, and 2010. Pooled OLS yields a point estimate of 0.174 (robust s.e. 0.003), close to the value of 0.181 reported by Heathcote, Storesletten, and Violante (2017) for 1978-2006. For more details, see Appendix C.3.

5 Empirical Implications

Before describing the counterfactual exercise, we show that the estimated model is qualitatively consistent with aggregate facts about housing demand and local employment.

Cross-sectional dispersion of housing expenditure

A prediction of the model is that the dispersion of housing expenditure is smaller than the dispersion of post-tax incomes within an MSA. To see this, take the variance of (4) within MSA n to get

$$\text{var}(\log R_{in}) = (1 - \epsilon)^2 \text{var}(\log y_{in}) \tag{21}$$

If the income elasticity is less than 1, then $\text{var}(\log y_{in}) > \text{var}(\log R_{in})$. In a homothetic environment, these moments are the same because log rents scale with log income one-to-one.

We test (21) using Census data. Both the variance of log rents and the variance of log post-tax incomes are computed only for renting households using the same taxation rule described in the model. The results are in Table 3. We reject a homothetic model, which would imply a coefficient of unity and an implied ϵ of zero. By contrast, the coefficient is significantly below unity. The ϵ implied by column (2), which includes year fixed effects, is almost identical to our preferred estimate.

MSA-level expenditure shares

Our estimated preferences predict that housing expenditure shares in an MSA will depend on that MSA's income distribution and the price of housing there. By contrast, Cobb-Douglas preferences would predict that variation in housing expenditure shares should be unrelated to

Table 3: Dispersion of rents and incomes

	(1)	(2)	(3)	(4)
Var. log income	0.378***	0.560***	0.261***	0.413***
	(0.055)	(0.056)	(0.054)	(0.108)
Year FE		✓		✓
MSA FE			✓	✓
Implied epsilon	0.386	0.251	0.490	0.357
Std. error	0.045	0.037	0.053	0.084
R^2	0.38	0.50	0.76	0.82
N	1076	1076	1076	1076

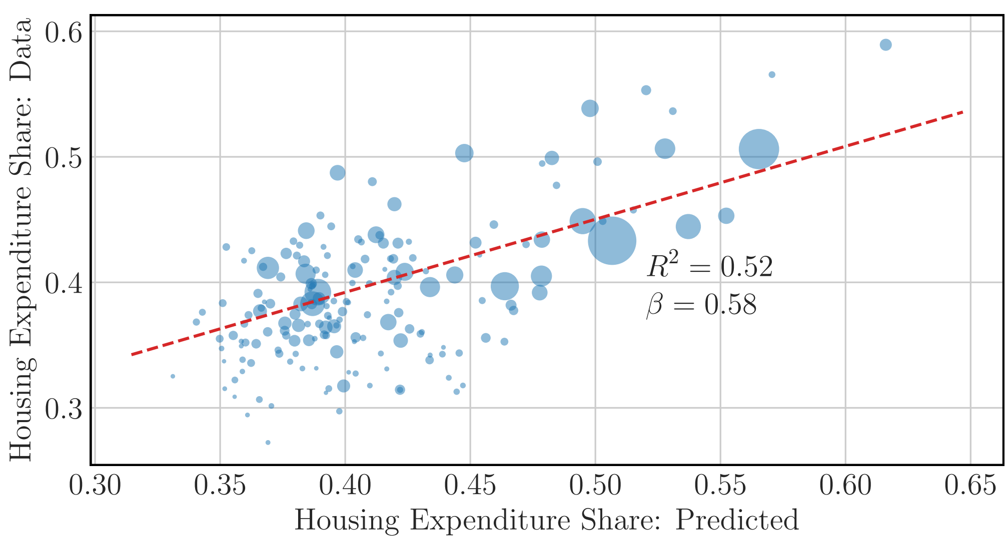
Source: Census

Note: We regress of the variance of log rents on the variance of log wages at the year × MSA level (see (21)). The values are computed for renting households only. Regressions weighted by 1980 employment. Standard errors are heteroskedasticity-robust.

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

incomes and prices. Davis and Ortalo-Magné (2011) find that there is little variation across MSAs in the ratio of wages to rents and interpret this as evidence in favor of Cobb-Douglas

Figure 3: MSA-level expenditure shares



Source: Census, CEX and Zillow. 'Housing Expenditure Share: Data' defined as the ratio of rental expenditure to after-tax income of renters, where after-tax income is calculated using Census data on wages and (15). 'Housing Expenditure Share: Predicted' calculated using (3) and Zillow data on rental prices. Both predicted and actual expenditure shares are scaled to match average expenditure shares of renters in the CEX in 2010. Dashed line shows regression of actual on predicted expenditure shares. Size of marker indicates number of renters by MSA.

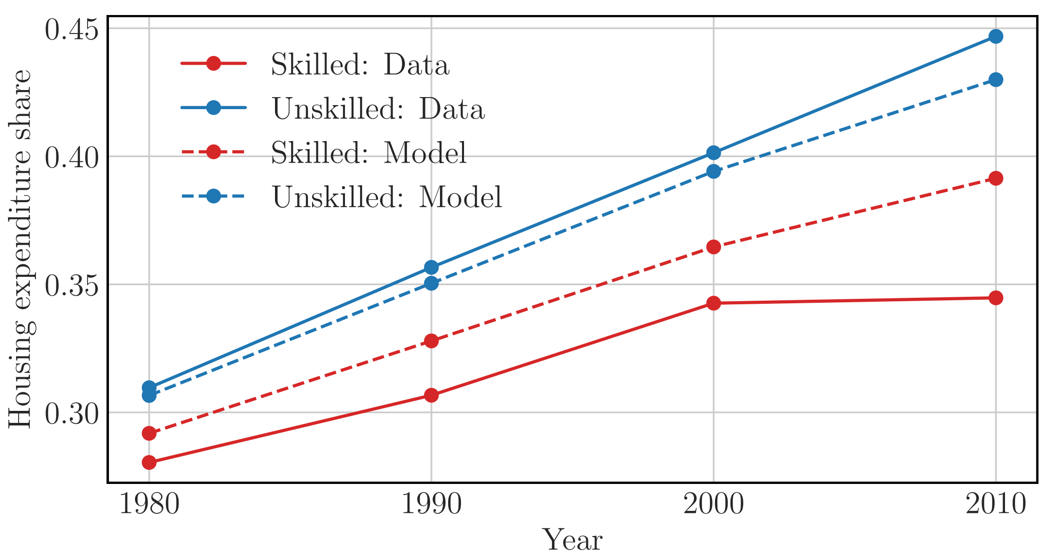
preferences. This is not inconsistent with our PIGL preferences: if high income cities also have high prices, then the price and income effects may offset each other so that we observe little variation in expenditure shares. In Figure 3 we calculate our model's predictions for housing expenditure shares in 2010 using data on rental prices per square foot from Zillow and Census wages. We find that our model's predictions align well with the data. We explain roughly 52% of cross-MSA variation in expenditure shares.

Time series patterns in housing expenditure

Only cross-sectional variation was used to estimate the nonhomotheticity in our preferences. Here we compare the predictions of our model with data on time series patterns in housing expenditure across skill groups. Figure 4 shows the evolution of skilled and unskilled housing expenditure shares between 1980 and 2010. In every period, the skilled expenditure share lies below the unskilled expenditure share. Furthermore, the gap between the two grows over time. The expenditure share of unskilled workers was 2.9 percentage points higher in 1980; by 2010 this figure was 10.2 percentage points.

The dashed lines in Figure 4 show the same objects in our model. Specifically, we match average expenditure shares from the CEX by construction, but allow equation 11 to determine

Figure 4: Diverging Expenditure Shares



Source: Census 1980-2010 and CEX 1980-2010. CEX was used the calculate the mean expenditure share in 1980-2010. Census expenditure shares are then rescaled to match CEX expenditure shares. Skilled and unskilled expenditure shares were then calculated using year-by-year regressions of individual expenditure shares on MSA dummies and an indicator for having attended a four year college.

the difference between expenditure shares between skilled and unskilled workers. The model qualitatively matches the data in that the gap grows over time; this is a natural consequence of the rising skill premium shown in Figure 1. Quantitatively the model understates the size of the gap. In 1980 the model implied gap was 1.5 percentage points, rising to 3.9 percentage points in 2010.

6 The Skill Premium and Sorting

In this section we use the quantitative model developed in Section 3.3 to investigate the effect of the rising skill premium on spatial sorting. We feed into the model (i) the increase in the productivity of skilled workers relative to unskilled workers since 1980, and (ii) the increase in the number of skilled workers since 1980. Together these two shocks determine the evolution of the skill premium in our model. Our focus is on how these location-neutral shocks can nevertheless lead to very different outcomes by location. Unskilled productivities, amenities, and house price shifters are held constant at their 1980 values.[13]

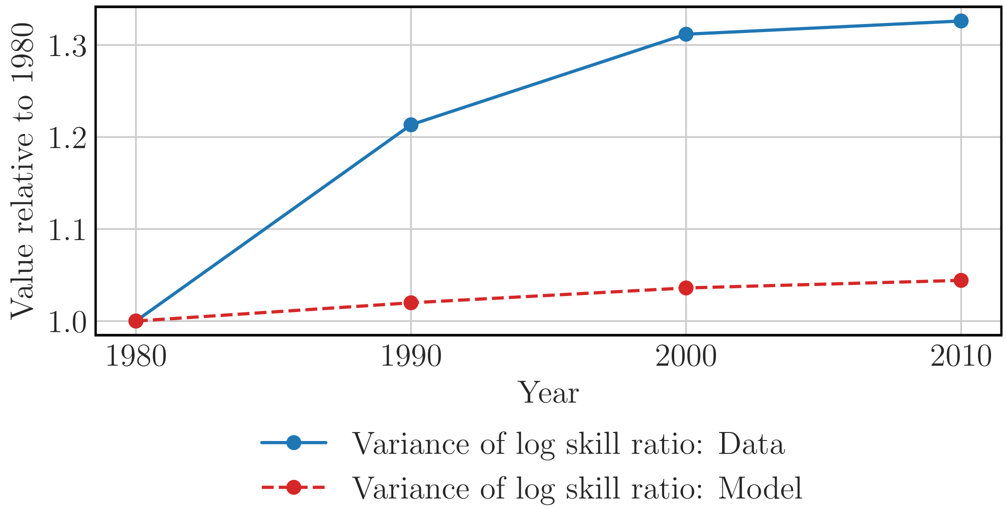
Results

The dashed line in Figure 5 shows the evolution of sorting under the counterfactual path for fundamentals. As in the data it rises monotonically, reaching a level 3.9% above its 1980 value in 2010. From these results we conclude that the rising skill premium at the national level explains 13.55% of the increase in sorting since 1980. Sorting rises in the counterfactual for two reasons. First, as the income of skilled workers rises they lower their housing expenditure shares and therefore become less sensitive to housing costs. This encourages skilled workers to move towards expensive cities, which in 1980 were also skill intensive cities. Second, the influx of skilled workers into these cities, as well as the higher incomes of skilled workers in general, force housing costs up in skill intensive cities. This encourages out-migration, particularly by unskilled workers. These two mechanisms reinforce each other so that sorting rises.

Of course, the 'Great Divergence' refers not only to the increasing tendency of skilled workers to cluster together, but also to the faster growth of wages and rents in locations gaining skilled workers. The solid lines in Figure 6 show the covariance of the log skill ratio with the log of skilled and unskilled wages and rents across MSAs since 1980. All three have risen since

[13]Formally, for each decade 1980-2010 we use equation 12 and Census data on wages and employment by MSA to back out productivities A^t_{is}. We define Δ^t as the average of $\frac{A^t_{sn}}{A^t_{un}}$, weighted by 1980 employment. We then solve for counterfactual outcomes when $\hat{A}^t_{sn} = \frac{\Delta^t}{\Delta^0}$ and $\hat{L}^t_s = \frac{L^t_s}{L^0_s}$, where we take year 0 to be 1980.

Figure 5: Sorting since 1980



Source: 'Data' calculated using Census data on employment by education level, 1980-2010. Starting with the 1980 equilibrium, the 'Model' exercise (i) increases the productivity of skilled workers in every location and (ii) increases the number of skilled workers nationally; these shocks are calibrated to match the actual increase in mean skilled labor productivity and actual share of skilled workers in the data.

1980, indicating that skilled workers increasingly choose to live in high wage, high rent cities. The dashed lines in Figure 6 show the evolution of these measures in our model. In each case the model displays the same increasing trend as in the data. In particular panel (c) shows that the rising national skill premium explains essentially all of the rising covariance between the skill ratio and rents.

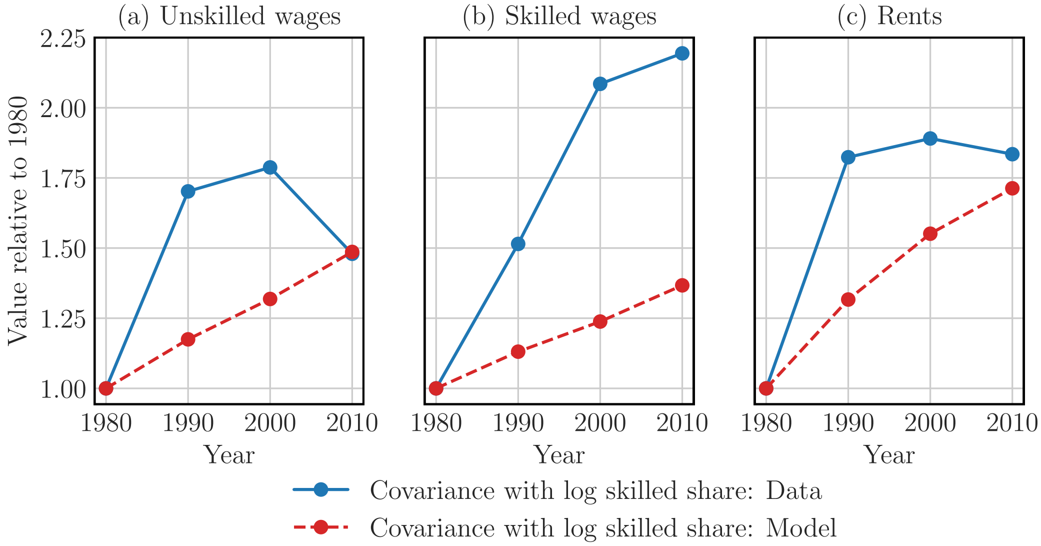
Alternative housing demand specifications

In this section we contrast our results with those obtained under alternative assumptions about housing demand. We consider two extreme cases: Cobb-Douglas preferences over housing and the tradable good; and a unit housing requirement. We focus on these because they have been used extensively in the spatial and urban literatures. [14] We introduce Cobb-Douglas preferences by replacing (9) with $v(e,p) = ep^{-\eta}$, where η is constant across time and across workers. We choose η to match the average expenditure share on housing in 1980 in the CEX, 0.30. We introduce a unit housing requirement by replacing (9) with $v(e,p) = e - p$, which corresponds to setting $\epsilon = \psi = 1$.

Figure 7 shows how sorting evolves in our counterfactual under these alternative preferences,

[14] For Cobb-Douglas, see for example Davis and Ortalo-Magné (2011) or Diamond (2016); for a unit housing requirement, see for example Saiz (2010) or Couture et al. (2019).

Figure 6: The Skill Share and Wages and Rents

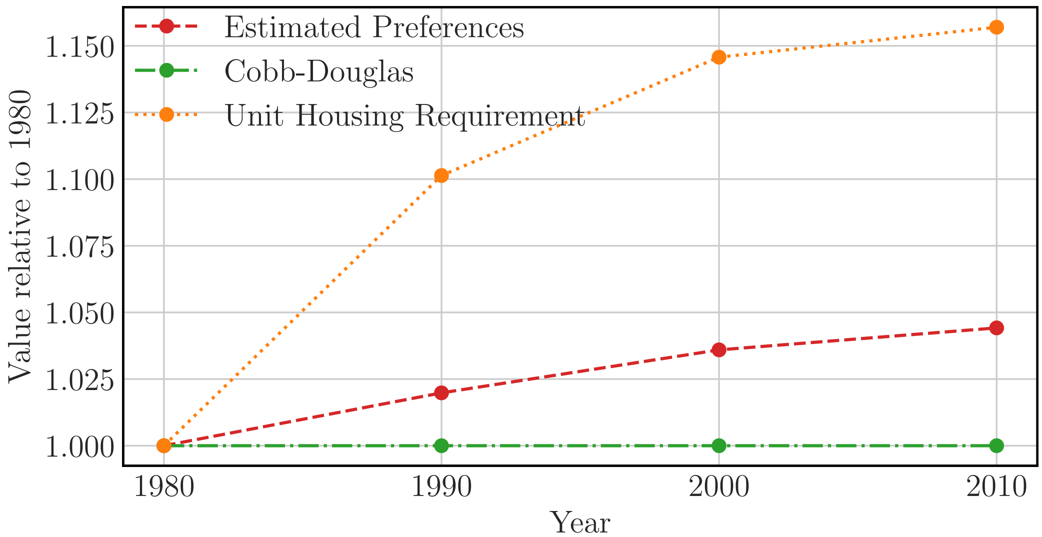


Source: 'Data' calculated using Census data on employment by education level, 1980-2010. Starting with the 1980 equilibrium, the 'Model' exercise (i) increases the productivity of skilled workers in every location and (ii) increases the number of skilled workers nationally; these shocks are calibrated to match the actual increase in mean skilled labor productivity and actual share of skilled workers in the data.

as well as under our estimated preferences. We can see that Cobb-Douglas preferences imply exactly zero change in sorting as a result of the rising national skill premium. In fact this is a direct result of Proposition 2, which continues to apply in this richer environment. In contrast a unit housing requirement implies housing demand is strongly nonhomothetic and perfectly price inelastic. The strong nonhomotheticity implies very different sensitivities to housing costs for skilled versus unskilled workers. The low price elasticity implies workers facing high prices have little scope to reduce their exposure to said prices, which makes moving to a cheaper location more attractive. Both these forces help to explain why a unit housing requirement (the dotted line in Figure 7) results in an increase in sorting over twice as large as under our estimated preferences.

The results in Figure 7 show that estimating housing demand, rather than imposing knife-edge cases as is usually done in the literature, is crucial for evaluating the effects of the rising skill premium on sorting. Our estimated parameters imply housing demand is moderately nonhomothetic and price inelastic - therefore our quantitative results fall between the extremes implied by Cobb-Douglas preferences and a unit housing requirement.

Figure 7: Alternative Demand Specifcations



Source: 'Data' calculated using Census data on employment by education level, 1980-2010. Starting with the 1980 equilibrium, the 'Model' exercise (i) increases the productivity of skilled workers in every location and (ii) increases the number of skilled workers nationally; these shocks are calibrated to match the actual increase in mean skilled labor productivity and actual share of skilled workers in the data. Each line measures the change in the variance of the log skill share since 1980 under a particular specification of preferences: estimated PIGL (red), Cobb-Douglas (green), or unit housing requirement (orange).

Taxation

We now consider how changes in taxes and transfers might reverse some of the increase in sorting shown in Figure 5. Recall that the progressivity of the tax system is determined by τ as in equation 15. A higher τ implies a more progressive tax system. Increasing τ compresses the income distribution and reduces the relative attractiveness of expensive housing markets for skilled workers; decreasing τ would dilate the income distribution and have the opposite effect.

We compute counterfactual 2010 equilibria under two different values of τ, keeping all other fundamentals constant. The results are in Table 4. In the first counterfactual, we decrease τ by 4 percentage points, from 0.174 to 0.134. Relative to the baseline, this lower τ approximately reflects the least progressive tax scheme observed in our sample period, which was in 1990 following the Reagan administration. A modest reduction in the progressivity of the tax system would further increase sorting by nearly ten percent in 2010, relative to the actual increase in the data from 1980-2010.

In a second tax counterfactual, we consider what would happen if the US were to adopt a Western-European style tax system. We use the results from Chang, Chang, and Kim

Table 4: Taxation and sorting

Scenario	τ	Change in sorting, as % of 1980-2010 increase	Gini (pretax)	Gini (posttax)
2010 data	0.174	-	0.193	0.159
2010 data, 1990 taxes	0.134	11.7%	0.194	0.167
2010 data, German taxes	0.435	-75.9%	0.190	0.105

All rows describe sorting in 2010 under various scenarios. The first row is the data; the second decreases τ to its 1990 level, which is the lowest in the 1980-2010 period; the third increases τ to the level in Germany. The third column shows the percentage change in sorting under each counterfactual in 2010, as a fraction of the actual increase from 1980 to 2010. The fourth and fifth columns show the Gini coefficients on pre- and posttax income.

(2018), who estimate (15) for a number of OECD countries. They find that Germany has the most progressive tax system, estimating a τ of 0.509 for Germnay compared to 0.248 for the US.[15] To simulate a German level of taxation in our model, we increase our τ by the difference, 0.261, and solve for the new equilibrium under 2010 fundamentals. Table 4 shows that a tax system as progressive as Germany's undoes nearly *all* of the observed increase in sorting since 1980. This is hardly surprising. The policy represents an extraordinary level of redistribution. The Gini coefficient on post-tax income in this counterfactual falls by nearly fifty percent relative to the 2010 baseline.

Since we do not model labor supply, the results here are intended to be suggestive. If skilled workers reduced their labor supply relatively more than unskilled workers in response to higher marginal tax rates, then a more progressive tax system would compress the income distribution even further. As a consequence sorting would fall even more. However, even without a labor supply response, we can conclude from these two counterfactuals that income inequality motivates income segregation. More broadly these results highlight the tight connection between the income distribution and the spatial distribution of economic activity implied by our model.

Extensions and Robustness

In Appendix E we extend our model to incorporate agglomeration externalites and endogenous amenities following Diamond (2016) and repeat the counterfactual described above. Here we highlight three results. First, Proposition 2 continues to apply. Even in this richer

[15]The level of $\hat{\tau}_{US}$ is somewhat higher than our estimate or the estimate from Heathcote, Storesletten, and Violante (2017), which Chang, Chang, and Kim (2018) attribute to different data sources.

environment, if preferences are homothetic then agglomeration and endogenous amenities *per se* do not link the rising skill premium to sorting. Second, with PIGL preferences these mechanisms tend to amplify the magnitude of our results. Third, the welfare implications of a nonhomothetic model and an endogenous amenity model are different. In our paper, sorting reflects but does not cause inequality by skill. In the model of Diamond (2016), sorting is associated with an externality which further exacerbates inequality: skilled workers benefit from being close to other skilled workers.

7 Conclusion

The United States is increasingly stratified by skill. Sorting across cities grew by more than 30% between 1980 and 2010. An emerging body of academic work, by Chetty et al. (2014), Fajgelbaum and Gaubert (2020), and others, has begun to unpack the *consequences* of this segregation by skill. Individuals' residences determine their access to labor markets, housing, amenities, and educational opportunities for themselves and their families. In this paper, we take a step back, proposing and quantifying a novel mechanism by which secular changes in the economy *cause* segregation by skill. One of the first principles of the spatial literature is that desirable locations—with high wages, high amenities, or both—tend to have high house prices in equilibrium, and that high house prices regulate the level of population. We show in this paper that high house prices also regulate the *composition* of population.

We construct a new model featuring skilled and unskilled workers and PIGL preferences over housing and nonhousing consumption. We estimate that housing expenditure within a location is income inelastic, with an elasticity of 0.76. Skilled workers have a comparative advantage in consumption in expensive cities because they are relatively less sensitive to costly housing. Put another way, expensive cities are not equally expensive for everyone. The wedge in cost-of-living between skilled and unskilled drives sorting.

A novel implication of our model is that changes in the income distribution are non-neutral over space, even if the shocks themselves are inherently location-neutral. In our main application we find that the rising aggregate skill premium explains 14% of the increase in sorting since 1980. In a second application, we show that the tax system, to the extent that it compresses or dilates the income distribution, is a policy instrument that can change sorting. Our results differ sharply from those that would be obtained under the preferences typically employed in the spatial literature; these either shut down the mechanism we discuss entirely, or significantly overstate its importance. We do not explicitly model the consequences of sorting. Our parsimonious model is static and efficient. However, to the extent that sorting is a matter of social or economic concern, our results suggest that addressing income inequality at a national level is likely to undo some of the geographical polarization that has occurred over the past 30 years.

References

"Bureau of Economic Analysis". 2020. *Metropolitan Area Regional Price Parities, Services: Rent*. https://apps.bea.gov/. Dataset.

Acemoglu, Daron, and David Autor. 2011. "Skills, tasks and technologies: Implications for employment and earnings". In *Handbook of labor economics*, 4:1043–1171. Elsevier.

Aguiar, Mark, and Mark Bils. 2015. "Has Consumption Inequality Mirrored Income Inequality?" *American Economic Review* 105, no. 9 (): 2725–56.

Albouy, David, Gabriel Ehrlich, and Yingyi Liu. 2016. "Housing demand, cost-of-living inequality, and the affordability crisis". Working Paper.

Attanasio, Orazio P., and Luigi Pistaferri. 2016. "Consumption Inequality". *Journal of Economic Perspectives* 30, no. 2 (): 3–28.

Attanasio, Orazio P, et al. 2012. "Modelling the demand for housing over the life cycle". *Review of Economic Dynamics* 15 (1): 1–18.

Boppart, Timo. 2014. "Structural change and the Kaldor facts in a growth model with relative price effects and non-Gorman preferences". *Econometrica* 82 (6): 2167–2196.

Bowlus, Audra J, and Chris Robinson. 2012. "Human capital prices, productivity, and growth". *American Economic Review* 102 (7): 3483–3515.

Bowlus, Audra, et al. 2017. "Wages and employment: The canonical model revisited". Working Paper.

Card, David. 2009. "Immigration and inequality". *American Economic Review* 99 (2): 1–21.

Carneiro, Pedro, and Sokbae Lee. 2011. "Trends in quality-adjusted skill premia in the United States, 1960-2000". *American Economic Review* 101 (6): 2309–49.

Chang, Bo Hyun, Yongsung Chang, and Sun-Bin Kim. 2018. "Pareto weights in practice: A quantitative analysis across 32 OECD countries". *Review of Economic Dynamics* 28:181–204.

Chetty, Raj, et al. 2014. " Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *". *The Quarterly Journal of Economics* 129, no. 4 (): 1553–1623. ISSN: 0033-5533. eprint: https://academic.oup.com/qje/article-pdf/129/4/1553/30631636/qju022.pdf.

Couture, Victor, et al. 2019. "Income growth and the distributional effects of urban spatial sorting". Working paper.

Davis, Morris A, and François Ortalo-Magné. 2011. "Household expenditures, wages, rents". *Review of Economic Dynamics* 14 (2): 248–261.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum. 2007. "Unbalanced Trade". *American Economic Review* 97, no. 2 (): 351–355.

Diamond, Rebecca. 2016. "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000". *American Economic Review* 106 (3): 479–524.

Eckert, Fabian. 2018. "Growing Apart: Tradable Services and the Fragmentation of the U.S. Economy". Working Paper.

Eckert, Fabian, Sharat Ganapati, and Conor Walsh. 2019. "Skilled Tradable Services: The Transformation of U.S. High-Skill Labor Markets". Working Paper.

Eckert, Fabian, and Michael Peters. 2018. "Skilled-biased technical change and regional convergence". Working Paper.

Eckert, Fabian, et al. 2020. "A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790". Working Paper.

Fajgelbaum, Pablo D, and Cecile Gaubert. 2020. "Optimal spatial policies, geography, and sorting". *The Quarterly Journal of Economics* 135 (2): 959–1036.

Ganong, Peter, and Daniel Shoag. 2017. "Why has regional income convergence in the US declined?" *Journal of Urban Economics* 102:76–90.

Garner, Thesia I, and Kathleen Short. 2009. "Accounting for owner-occupied dwelling services: Aggregates and distributions". *Journal of Housing Economics* 18 (3): 233–248.

Giannone, Elisa. 2019. "Skilled-biased technical change and regional convergence". Working Paper.

Goldin, Claudia D., and Lawrence F. Katz. 2009. *The race between Education and Technology.* Cambridge, MA: Harvard University Press.

Greaney, Brian. 2020. "The Distributional Effects of Uneven Regional Growth". Working Paper.

Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. "Superstar cities". *American Economic Journal: Economic Policy* 5 (4): 167–99.

Gyourko, Joseph, Albert Saiz, and Anita Summers. 2008. "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index". *Urban Studies* 45 (3): 693–729.

Hansen, Julia L, John P Formby, W James Smith, et al. 1996. "The income elasticity of demand for housing: evidence from concentration curves". *Journal of Urban Economics* 39 (2): 173–192.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante. 2017. "Optimal tax progressivity: An analytical framework". *The Quarterly Journal of Economics* 132 (4): 1693–1754.

Hendricks, Lutz, and Todd Schoellman. 2014. "Student abilities during the expansion of US education". *Journal of Monetary Economics* 63:19–36.

Hornbeck, Richard, and Enrico Moretti. 2019. "Who benefits from productivity growth? Direct and indirect effects of local TFP growth on wages, rents, and inequality". Working Paper.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation". *American Economic Journal: Macroeconomics* 11, no. 2 (): 1–39.

Ioannides, Yannis M, Jeffrey E Zabel, et al. 2008. "Interactions, neighborhood selection and housing demand". *Journal of Urban Economics* 63 (1): 229.

Katz, Lawrence F, and Kevin M Murphy. 1992. "Changes in relative wages, 1963–1987: supply and demand factors". *The quarterly journal of economics* 107 (1): 35–78.

Larsen, Erling Røed. 2014. "The Engel curve of owner-occupied housing consumption". *Journal of Applied Economics* 17 (2): 325–352.

Lewbel, Arthur, and Krishna Pendakur. 2009. "Tricks with Hicks: The EASI demand system". *American Economic Review* 99 (3): 827–63.

Moretti, Enrico. 2013. "Real Wage Inequality". *American Economic Journal: Applied Economics* 5, no. 1 (): 65–103.

— . 2012. *The New Geography of Jobs.* Boston: Houghton Mifflin Harcourt.

Muellbauer, John. 1975. "Aggregation, income distribution and consumer demand". *The Review of Economic Studies* 42 (4): 525–543.

— . 1976. "Community preferences and the representative consumer". *Econometrica*: 979–999.

Porter, Eduardo, and Guilbert Gates. 2019. "Why Workers Without College Degrees Are Fleeing Big Cities". *New York Times* (). https://www.nytimes.com/interactive/2019/05/21/business/economy/migration-big-cities.html.

Redding, Stephen J. 2016. "Goods trade, factor mobility and welfare". *Journal of International Economics* 101:148–167. ISSN: 0022-1996.

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

Rosenthal, Stuart S, and William C Strange. 2004. “Evidence on the nature and sources of agglomeration economies”, 4:2119–2171.

Ruggles, Steven, et al. 2020. *IPUMS USA: Version 10.0*. Minneapolis, MN. doi:https://doi.org/10.18128/D010.V10.0.

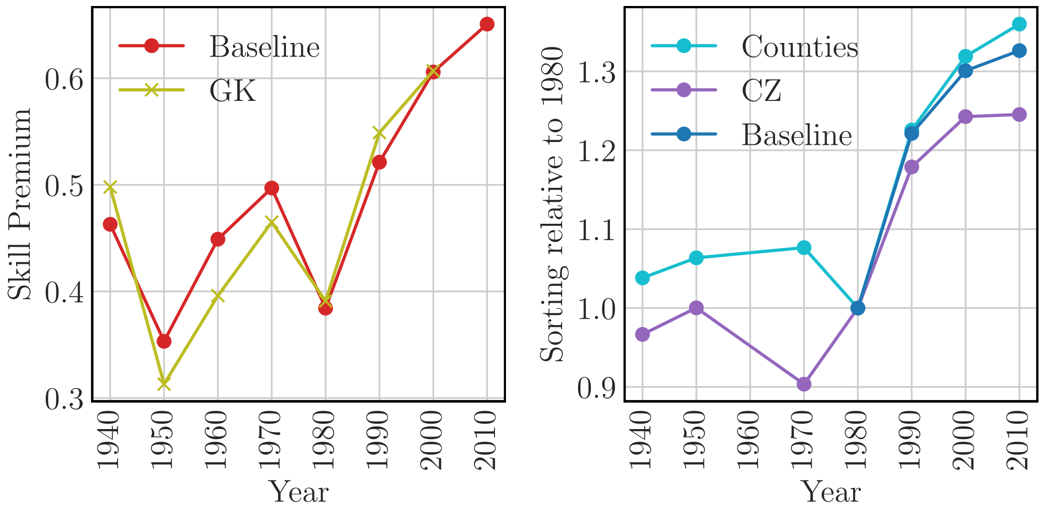
Saiz, Albert. 2010. “The geographic determinants of housing supply”. *The Quarterly Journal of Economics* 125 (3): 1253–1296.

Zabel, Jeffrey E. 2004. “The demand for housing services”. *Journal of Housing Economics* 13 (1): 16–35.

Zillow. 2017. *Zillow Rent Index (ZRI), Metro, Per Sq. Ft., All Homes*. https://data.world/zillow-data/all-zillow-metrics-metro. Dataset.

A Figures

Figure A.1: Sorting and the skill premium over time



Source: Red series (left) and blue series (right) are the baseline measures of the skill premium and sorting used in the paper. GK refers to the skill premium computed in Goldin and Katz (2009). The alternative historical sorting measures are computed from the Census, 1940-2010. Sorting is measured as the variance of the log skill ratio. For the historical series, we define the skill ratio as the ratio of adults aged 25 and older with at least a four year college degree to adults aged 25 and older without a college degree; for the baseline series, the skill ratio is computed in terms of the labor force rather than the adult population. Skill premium is defined as the log difference between average wages of skilled and unskilled workers. CZ refers to 1990 commuting zones; we use a crosswalk from Eckert et al. (2020). Data on 1960 educational attainment at the county level are not digitally available.

Table A.1: Income elasticity, Alternative specifications

	(1) Renters	(2) Owners	(3) Both	(4) Both	(5) Both, OOP	(6) Owners, OOP	(7) Both, OOP
Log expenditure	0.798***	0.753***	0.800***	0.863***	0.832***	0.855***	0.823***
	(0.021)	(0.017)	(0.019)	(0.017)	(0.016)	(0.022)	(0.018)
MSA FE	✓	✓	✓	✓	✓	✓	✓
Tenure dummy	-	-	✓		✓	-	
IV	Income	Income	Educ.	Educ.	Income	Income	Education
R^2	0.62	0.62	0.70	0.67	0.56	0.56	0.56
First-stage F-statistic	937.8	1655.1	124.6	153.0	2469.3	1776.0	140.3
N	2956	5855	8811	8811	8811	5855	8811
No. of clusters	237	237	239	239	239	237	239

Source: CEX, 2006-2017

Note: Dependent variable is log rent or rental equivalent. 'OOP' denotes out-of-pocket expenses for homeowners, excluding mortgage principal. Controls include family size, number of earners, sex, race, age of household head, and year. Income instrument is log pre-tax income and education instrument is a set of dummies for years of schooling. Standard errors clustered at year × MSA level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table A.2: Income elasticity, PSID

	(1) Renters	(2) Renters	(3) Renters	(4) Renters	(5) Owners	(6) Both	(7) Both
Log expenditure	0.746***	0.800***	0.744***	0.840***	0.857***	0.904***	0.829***
	(0.025)	(0.034)	(0.024)	(0.034)	(0.054)	(0.031)	(0.039)
State × metro FE	✓	✓	✓	✓	✓	✓	✓
Tenure dummy	-	-	-	-	-		✓
IV	Income	Lagged expenditure	Mean expenditure	Educ.	Income	Income	Income
R^2	0.51	0.53	0.52	0.50	0.62	0.64	0.66
First-stage F-statistic							
N	13582	6779	11601	13267	2298	4768	4768
No. of clusters	342	270	336	340	50	51	51

Source: PSID, 2005-2017

Note: Dependent variable is log rent or rental equivalent. Rent equivalent is only available in the 2017 wave. Controls include family size, number of earners, sex, race, age of household head, and year. Instruments: 'income' is log family income; 'lagged expenditure' is lagged log total expenditure; 'mean expenditure' is mean log total household expenditure in all other waves of the survey in which the household participated; 'educ.' is a set of dummies for years of schooling. Standard errors are clustered at state × year level.

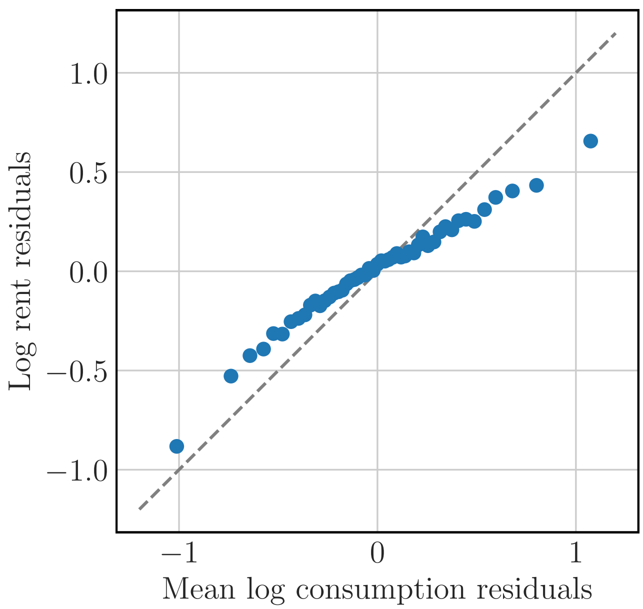
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Table A.3: Price and income elasticities in the literature

Paper	Income elasticity (renters)	Income elasticity (homeowners)	Price elasticity	Notes
Lewbel and Pendakur (2009)[a]	0.72			Authors' calculation
Albouy, Ehrlich, and Liu (2016)[b]	0.68		0.35	Table 3, column 2
Larsen (2014)[c]		0.33		Table 3, median
Ioannides, Zabel, et al. (2008)[d]		0.21	0.18	Table 5, column 1
Hansen, Formby, Smith, et al. (1996)[e]	0.37	0.63		Table 5, columns 1-2, mean
Aguiar and Bils (2015)[f]	0.92			Table 2, column 1
Zabel (2004)[g]		0.48	0.93	Table 3, median
Attanasio et al. (2012)[h]	0.78		0.60	Table 4 panel B, Table 7 panel B
Rosenthal (2014)[i]	0.12	0.41		Table 5
Paper benchmark	0.80	0.75	0.55	CEX; BEA; Census

[a] Canadian Family Expenditure Surveys, 1969-1996. Uncompensated elasticity is median value computed using Appendix VII.1.
[b] US Census, 1970-2014.
[c] Norweigian Rental Survey and Consumer Expenditure Survey, 2007.
[d] American Housing Survey, 1985-1993.
[e] American Housing Survey, 1989.
[f] US CEX, 1980-2010.
[g] American Housing Survey, 2001.
[h] British Household Panel Survey, 1991-2002. Estimates for high- and low-education groups are averaged with weights one-third and two-thirds, respectively.
[i] American Housing Survey, 1985-2011.

Figure A.2: Housing expenditure, non-parametric regression



Source: CEX, baseline sample. Log rental expenditure is regressed on a set of dummies for fifty bins of log total expenditure, after partialling out demographic controls for both variables. The grey line is the 45 degree line, which corresponds to the homothetic case.

B Theory

B.1 Section 2 Proofs

Validity of $v(e,p)$ as an indirect utility function

This section draws directly on Boppart (2014). We require $0 < \epsilon \leq \psi < 1$ and $\nu > 0$. These restriction are satisfied by the estimates of ϵ and ψ we use from section 2 onwards. We further require that

$$e^{\epsilon} \geq \left(\frac{1-\epsilon}{1-\psi}\right) \nu p^{\psi},$$

which corresponds to Lemma 1 in Boppart (2014). This restriction corresponds to imposing that all housing expenditure shares η_{in} satisfy

$$\eta_{in} \leq \frac{1-\psi}{1-\epsilon}$$

which for the parameter values we estimate is $\eta_{in} \leq 0.9$. We find that this restriction is satisfied in our dataset and all our counterfactual experiments.

B.2 Section 3.2 Proofs

Proof of Lemma 1

Take the ratio of equation 10 for $i = s$ and $i = u$ to obtain

$$s_n = \left(\frac{v_{sn}}{v_{un}}\right)^{\theta} \left(\frac{L_u \sum_m B_m v_{sm}^{\theta}}{L_s \sum_m B_m v_{um}^{\theta}}\right)^{-1}.$$

Label the final term ξ and use the definition of v_{in} to obtain

$$s_n = \left(\frac{\epsilon^{-1} e_{sn}^{\epsilon} - \psi^{-1} p_n^{\psi}}{\epsilon^{-1} e_{un}^{\epsilon} - \psi^{-1} p_n^{\psi}}\right)^{\theta} \xi.$$

Now note that $e_{sn} = A z_n$ from (7) and $e_{un} = z_n$ from (8). After some rearrangement we obtain

$$s_n = \xi \left(\frac{\psi A_s^{\epsilon} - \epsilon \eta_n}{\psi A_u^{\epsilon} - \epsilon \eta_n}\right)^{\theta}$$

as desired. Note that ξ must be smaller than 1 because it is the ratio of the utility of unskilled workers to skilled workers. Differentiating with respect to η_n shows this function is strictly increasing in η_n as long as $A_s > A_u$.

Proof of Proposition 1

Differentiating the expression in Proposition 1 yields

$$d\log s_n = d\log\xi + \theta\epsilon\left(\frac{\psi}{\psi A_s^{\epsilon} - \epsilon\eta_n}\right) d\log A_s.$$

Next, invert the expression for s_n in Proposition 1 to obtain an expression for η_n. The substitute this into the expression for $d\log s_n$ above to obtain

$$d\log s_n = \beta + \theta\epsilon\left(1 - \left(\frac{\xi}{s_n}\right)^{\frac{1}{\theta}}\right)\left(1 - \left(\frac{A_u}{A_s}\right)^{\epsilon}\right)^{-1} d\log A_s$$

with $\beta = d\log\xi$. It is clear this is increasing in s_n as long as $A_u < A_s$.

Next we show $dS > 0$. By definition

$$S = \sum_n \omega_n(\log s_n - \bar{s})^2$$

where we allow for exogenous weights ω_n and $\bar{s}$ is the (weighted) average of $\log s_n$. Differentiating, we obtain

$$dS = 2\sum_n \omega_n(d\log s_n - d\bar{s})(s_n - \bar{s}) = 2\text{Cov}(d\log s_n, s_n).$$

We have shown above that $d\log s_n$ is an increasing function of s_n. Therefore this covariance is positive and $dS > 0$.

B.3 Section 3.3 Proofs

Quantitative Model Hat Algebra

Statement of result

Suppose we have observations on an equilibrium $\left[\left(\bar{l}_{in}, \bar{w}_{in}\right)_i, \bar{p}_n\right]_n$ generated by fundamentals $\left[\left(\bar{A}_{in}, \bar{B}_{in}\right)_i, \bar{\Pi}_n\right]_n, (\bar{L}_i)$. Denote a counterfactual set of fundamentals by $\left[\left(\tilde{A}_{in}, \tilde{B}_{in}\right)_i, \tilde{\Pi}_n\right]_n$, $(\tilde{L}_i)$ and the associated equilibrium by $\left[\left(\tilde{l}_{in}, \tilde{w}_{in}\right)_i, \tilde{p}_n\right]_n$. Finally for any variable x let $\hat{x} = \frac{\tilde{x}}{\bar{x}}$. Then $\left[\left(\hat{l}_{in}, \hat{w}_{in}\right)_i, \hat{p}_n\right]_n$ solves the following system of equations:

Production and Taxation

$$\hat{w}_{in} = \hat{A}_{in}^{\frac{\sigma-1}{\sigma}} \hat{l}_{in}^{\frac{-1}{\sigma}} \left(\sum_j \left(\hat{A}_{in} \hat{l}_{in} \right)^{\frac{\sigma-1}{\sigma}} \left(\frac{\bar{w}_{jn} \bar{l}_{jn}}{\sum_h \bar{w}_{hn} \bar{l}_{hn}} \right) \right)^{\frac{1}{\sigma-1}} \tag{22}$$

$$\hat{y}_{in} = \hat{\lambda} \hat{w}_{in}^{1-\tau} \tag{23}$$

$$\hat{\lambda} = \frac{\left(\sum_{i,n} \hat{l}_{in} \hat{w}_{in} \left(\frac{\bar{l}_{in} \bar{w}_{in}}{\sum_{j,m} \bar{l}_{jm} \bar{w}_{jm}} \right) \right)}{\left(\sum_{i,n} \hat{l}_{in} \hat{w}_{in}^{1-\tau} \left(\frac{\bar{l}_{in} \bar{w}_{in}^{1-\tau}}{\sum_{j,m} \bar{l}_{jm} \bar{w}_{jm}^{1-\tau}} \right) \right)} \tag{24}$$

Preferences and Location Choice

$$\hat{l}_{in} = \frac{\hat{B}_{in} \hat{v}_{in}^{\theta}}{\sum_m \bar{l}_{im} \hat{B}_{in} \hat{v}_{in}^{\theta}} \hat{L}_i \tag{25}$$

$$\hat{v}_{in} = \hat{y}_{in} \left(\frac{\psi - \epsilon \bar{\eta}_{in} \hat{\eta}_{in}}{\psi - \epsilon \bar{\eta}_{in}} \right)^{\frac{1}{\epsilon}} \tag{26}$$

$$\hat{\eta}_{in} = \hat{p}_n^{\psi} \hat{y}_{in}^{-\epsilon} \tag{27}$$

Housing Supply

$$\hat{p}_n = \hat{\Pi}_n \hat{HD}_n^{\gamma_n} \tag{28}$$

$$\hat{HD}_n = \hat{p}_n^{-1} \sum_j \hat{l}_{jn} \hat{y}_{jn} \hat{\eta}_{jn} \left(\frac{\bar{l}_{jn} \bar{y}_{jn} \bar{\eta}_{jn}}{\sum_h \bar{l}_{hn} \bar{y}_{hn} \bar{\eta}_{hn}} \right) \tag{29}$$

where $\bar{\eta}_{in}$ is defined by

$$\bar{\eta}_{in} = \bar{\eta}_n \left(\frac{\bar{y}_{in}}{\bar{Y}_n} \right)^{-\epsilon}$$

$$\bar{Y}_n = \frac{\sum_i \bar{l}_{in} \bar{y}_{in}}{\sum_i \bar{l}_{in}}$$

and likewise for $\hat{\eta}_{in}$.

Production and Taxation

(22) is obtained by taking the ratio of (12) between counterfactual and baseline. (23) is obtained by taking the ratio of (15) between counterfactual and baseline. To obtain (24) we start by summing (15) over all workers

$$\sum_{i,n} y_{in} l_{in} = \lambda \sum_{i,n} w_{in}^{1-\tau} l_{in}.$$

Now we impose government budget balance, which is equivalent to imposing that total GDP is unchanged by the tax regime. That is

$$\sum_{i,n} y_{in} l_{in} = \sum_{i,n} w_{in} l_{in}.$$

The two equations above imply

$$\lambda = \frac{\sum_{i,n} w_{in} l_{in}}{\sum_{i,n} w_{in}^{1-\tau} l_{in}}.$$

Taking the ratio of this equation between counterfactual and baseline yields (24).

Preferences and Location Choice

(25) is obtained by taking the ratio of (16) between counterfactual and baseline. (27) is obtained by taking the ratio of (3) between counterfactual and baseline. Note that we have replace $v(y_{in}, p_n)$ with v_{in} for notational convenience. To obtain (26) we start by taking the ratio of (9) between counterfactual and baseline. This yields

$$\hat{v}_{in} = \left(\left(\frac{\epsilon^{-1} \bar{y}_{in}^{\epsilon}}{\epsilon^{-1} \bar{y}_{in}^{\epsilon} - \frac{\nu}{\psi} \bar{p}_n^{\psi}} \right) \hat{y}_{in}^{\epsilon} - \left(\frac{\frac{\nu}{\psi} \bar{p}_n^{\psi}}{\epsilon^{-1} \bar{y}_{in}^{\epsilon} - \frac{\nu}{\psi} \bar{p}_n^{\psi}} \right) \hat{p}_n^{\psi} \right)^{\frac{1}{\epsilon}}$$

Now substitute (3) evaluated in the baseline equilibrium

$$\hat{v}_{in} = \left(\left(\frac{1}{1 - \frac{\epsilon}{\psi} \bar{\eta}_{in}^{\psi}} \right) \hat{y}_{in}^{\epsilon} - \left(\frac{\frac{\epsilon}{\psi} \bar{\eta}_{in}}{1 - \frac{\epsilon}{\psi} \bar{\eta}_{in}^{\psi}} \right) \hat{p}_n^{\psi} \right)^{\frac{1}{\epsilon}}.$$

Now substitute (27) to remove $\hat{p}_n$ and rearrange to obtain (26).

Housing Supply

(28) is obtained by taking the ratio of (13) between counterfactual and baseline. (29) is obtained by taking the ratio of (14) between counterfactual and baseline.

B.3.1 Proof of Proposition 2

Start by evaluating (16) for the skilled and unskilled type in location n. Take the log difference to obtain

$$\log\left(\frac{l_{sn}}{l_{un}}\right) = \log\left(\frac{\xi_s}{\xi_u}\right) + \theta\log\left(\frac{y_{sn}}{y_{un}}\right) + \log\left(\frac{B_{sn}}{B_{un}}\right)$$

Differencing this between the baseline allocation and the counterfactual yields

$$\log\frac{\hat{l}_{sn}}{\hat{l}_{un}} = \log\frac{\hat{\xi}_s}{\hat{\xi}_u} + \theta\log\frac{\hat{y}_{sn}}{\hat{y}_{un}} + \log\frac{\hat{B}_{sn}}{\hat{B}_{un}}.$$

Next use (23) to replace incomes with wages

$$\log\frac{\hat{l}_{sn}}{\hat{l}_{un}} = \log\frac{\hat{\xi}_s}{\hat{\xi}_u} + \theta(1-\tau)\log\frac{\hat{w}_{sn}}{\hat{w}_{un}} + \log\frac{\hat{B}_{sn}}{\hat{B}_{un}}.$$

Next use (22) to replace wages with productivities and employment shares

$$\log\frac{\hat{l}_{sn}}{\hat{l}_{un}} = \log\frac{\hat{\xi}_s}{\hat{\xi}_u} + \theta(1-\tau)\left(\left(\frac{\sigma-1}{\sigma}\right)\log\frac{\hat{A}_{sn}}{\hat{A}_{un}} - \left(\frac{1}{\sigma}\right)\log\frac{\hat{l}_{sn}}{\hat{l}_{un}}\right) + \log\frac{\hat{B}_{sn}}{\hat{B}_{un}}.$$

Rearranging and exponentiating gives the desired result.

Table C.1: Price elasticity of demand

	(1) OLS	(2) 2SLS
Price	0.399***	0.493***
	(0.0241)	(0.0509)
R^2	0.57	0.44
First-stage F-statistic	-	274
N	318	318
No. of clusters	39	39

Note: All regressions weighted by 2010 employment. Standard errors clustered at the state level. Instruments in (2) are share of land unavailable and Wharton Land Use Regulation Index.

Source: Census and Zillow Rent Index

* p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01. Standard errors in parentheses.

C Estimation

C.1 Estimating the price elasticity with an alternative data source

In this section we repeat the estimation of (17) using data from Zillow, a real estate listings website (Zillow 2017). Zillow estimates the price per square foot of the rental stock for major metropolitan areas since 2010. The advantage of this data is that can be compared over time, unlike the BEA price index which is defined relative to an average value at each point in time. In addition, the Zillow data considers the square footage of housing units at a more granular level than the BEA, which considers only the number of rooms and bedrooms. However, the Zillow data does not attempt to control for other dimensions of housing quality, such as the age or type of structure. Nevertheless, Figure C.1 shows that the point estimate on log price is virtually unchanged despite the data difference.

C.2 Calibration of the migration elasticity

To calibrate θ, the elasticity of migration to real wages, we implement an indirect inference procedure. Conditional on all other parameters and data, we can choose a unique value of θ such that the reduced-form migration elasticity in our model exactly matches the corresponding empirical moment in Hornbeck and Moretti (2019). The procedure is as follows:

1. Guess a value $\widetilde{\theta}$.
2. Invert the model at this guess. In particular, recover 1980 TFP, $\widetilde{A}_{in}^{1980}$
3. Draw a vector of MSA-level TFP shocks $\widehat{s}_n$.

4. Solve the model at the new levels of TFP, holding fixed all other fundamentals.

5. Compute the migration elasticity as a ratio of OLS coefficients. The numerator is the elasticity of employment growth to TFP growth and the denominator is the elasticity of wage growth to TFP growth.

6. Update $\widetilde{\theta}$ until convergence.

To be consistent with Hornbeck and Moretti (2019), who do not recover type-specific labor productivity, we enforce the shocks in our simulation to be skill-neutral. We draw independent shocks from a lognormal distribution with mean and variance determined by the data. To estimate the mean and variance, we compute the empirical distribution $\mathcal{S}$ of realized shocks $\widehat{s}_n$. The set of realized shocks is defined such that counterfactual output (given fundamental productivity $\widehat{s}_n \cdot A_{in}^{1980}$ and all other 1990 fundamentals) matches observed output for each MSA in 1990.

We find that the mean and standard deviation of $\mathcal{S}$ are 1.078 and 0.072, respectively. The 90/10 ratio (in logs) is 0.159. By way of comparison, Hornbeck and Moretti (2019) report that the mean of actual TFP shocks in US manufacturing was 1.053 in the same period, with a log 90/10 ratio of 0.151. The comparison suggests that $\mathcal{S}$ is a reasonable approximation to the actual distribution of shocks in the data.

C.3 Calibration of the tax system

We use data from the 1981/91/2001/11 waves of the PSID (each containing summary information on the *prior* year's income). Using the same sample restrictions as in section 2, we run the PSID data through the NBER's TAXSIM program. For each household, pre-tax income is computed as adjusted gross income minus Social Security transfers. Post-tax income is computed as pre-tax income minus federal and state taxes (including payroll taxes) plus Social Security transfers.

We estimate (20) by pooled OLS over the four periods. Our estimated $\hat{\tau}$ is 0.174 (robust s.e. 0.003). The R^2 of the regression is 0.98, suggesting that, despite its parsimony, a log-linear tax equation is a good approximation to the actual tax system in the United States. Our estimate is quite close to Heathcote, Storesletten, and Violante (2017), who estimate $\hat{\tau} = 0.181$ despite constructing a household sample with slightly different inclusion criteria.

In alternative specification, we permit τ to vary over time. We estimate (20) year-by-year. Table C.2 contains the estimated coefficients.

Table C.2: Tax estimates

Year	$\hat{\tau}$
1980	0.187
	(0.003)
1990	0.145
	(0.009)
2000	0.182
	(0.003)
2010	0.190
	(0.006)

Source: PSID. This is the estimated coefficient in a regression of log post-tax income on log pre-tax income. Regression is done year-by-year. Robust standard errors in parentheses.

D Data

D.1 Census variable definitions

We use the 5% public use samples from the 1980, 1990, and 2000 Censuses. For the final period of data, we use the 2009-2011 American Community Survey, a 3% sample. For convenience we refer to this as the "2010 data." We do not use a 5% sample in the final period for two reasons: first, we want to abstract as much as possible from the recession in 2008, and second, the MSA definitions changed slightly in 2012.

IPUMS attempts to concord geographic units across years, although complete concordance is not possible because of data availability and disclosure rules. We classify MSAs according to the variable `metarea`. We produce a balanced panel using the following rule: if an MSA appears in all four years, then it is kept. If an MSA does not appear in all four years, then we assign all individuals in that MSA across all years to a residual state category. For example, Charlottesville, VA appears in 1980, 2000, and 2010, but not in 1990. Therefore we assign all individuals in Charlottesville in every year to "Virginia." This procedure gives us 219 MSAs (including Washington, D.C.) and 50 residual state categories, for a total of 269 regions.

A worker is considered skilled if she or he has completed at least a four year college degree according to the variable `educ`. By this metric, the national fraction of workers who are skilled is 22.5% in 1980, 26.5% in 1990, 30.2% in 2000, and 35.7% in 2010.

We compute wages and employment for each region, skill level, and year. Wages are computed from the variable `incwage`. To be included in the wage and employment sample, workers must be between 25 and 55 years old, inclusive; not have any business or farm income; work at least 40 weeks per year and 35 hours per week; and earn at least one-half the federal minimum wage. For each region and year, rents are averaged among all renting households whose head is included in the wage/employment sample. The rent variable used is `rentgrs`. Rents are not weighted by household size or computed separately by skill. Because the ratio of rents to wages in the data may differ from the analogous model object—for example, due to differences in household size and composition, intra-household bargaining, savings behavior, or non-wage income—we multiply all rents in each year by a scalar to exactly match the aggregate rental share computed from the CEX. Wages and rents are adjusted to 2000 real values using Non-Shelter CPI from the BLS.

D.1.1 Bartik instruments

In order to obtain instrumental variables for labor demand, we construct Bartik shift-share variables. The share is a region's industrial composition in 1980, and the shift is change in average wages nationwide (excluding the region itself).

We use the industry categories in the Census variable `ind1990`. We create a balanced panel of industries using our own crosswalk. The result is 208 industries which are consistently defined over all four periods. We drop individuals who cannot be classified into any industry ($\approx 0.3\%$ of workers) or who are in the military ($\approx 0.9\%$ of workers).

D.2 CEX variable definitions

We append the 2006-2017 Consumer Expenditure Surveys (CEX) together and annualize at the household level. We define rental expenditure as actual rent paid for renters (`rendwe`) and self reported rental-equivalent (`renteqvx`) for owners. We add utilities `util` to be consistent with the data available in the Census. To solicit rental equivalent, homeowners are asked "If someone were to rent your home today, how much do you think it would rent for monthly, unfurnished and without utilities?" We define total consumption expenditure as equal to total reported expenditure `totexp` *less* retirement and pension savings `retpen`, cash contributions `cashco`, miscellaneous outlays `misc` (which includes mortgage principal), and life and personal insurance `lifins`.

Pre-tax income is taken from the variable `fincbtxm`. We cut respondents in the top and bottom 1% of the distribution of pre-tax income in each year to guard against serious misreporting errors. We also consider the same sample criteria as in our Census data: prime-age and strong labor force attachment. The controls included in the regressions are dummies for family size bins, number of earners, age bins, sex of household head, race of household head, year, and a dummy for home ownership. We use the CEX sample weights in all regressions.

It bears mention that in theory, rental equivalent should be added to pre-tax income for homeowners—it is capital income, after all. In practice, we do not do this because we use pre-tax income as an instrument. Hence, any measurement error in rental equivalent would be mechanically correlated with the instrument. Including a dummy for home ownership in the first stage addresses this problem, as predicted consumption is allowed to be higher for homeowners at the same level of measured income (i.e., income excluding rental equivalent.)

D.3 PSID variable definitions

In a robustness exercise, we re-estimate the income elasticity using the PSID. The results are consistent with an income elasticity of around three-quarters. The PSID included waves from 2005-2017, biannually. Total expenditure is computed as the sum of all reported consumption categories: rent, food, utilities, automobile expenses (including car loans, down payments, lease payments, insurance, repairs, gas, and parking), other transportation expenses, education, childcare, healthcare, home repairs, furniture, computers (2017 only), clothing, travel, and recreation. Rental expenditure is equal to rent plus utilities. Homeowners were not asked to estimate the rental value of their home until 2017, so we consider renters only until 2017 and include owners thereafter. In the PSID we consider the exact same sample selection critera and controls as in the CEX, described in Section D.2. We use the PSID sample weights in all regressions.

E Extended Model

In this section we extend the model in Section 3 to include forces highlighted by the spatial literature; in particular we consider agglomeration externalities and endogenous amenities.

E.1 A General Model

We extend the model in Section 3 in two ways. We replace the exogenous productivities A_{in} with

$$A_{in} = a_{in} h_n(l_{sn}, l_{un}) \tag{30}$$

where a_{in} is an exogenous shifter and we do not restrict the function h_n. We replace the exogenous amenities B_{in} with

$$B_{in} = b_{in} \left(\frac{l_{sn}}{l_{un}} \right)^{\delta_i} \tag{31}$$

where b_{in} is an exogenous shifter. It is important to note what we are excluding here. First, while we allow different skill-groups to make different *contributions* to the agglomeration externality embedded in A_{in}, we are not allowing them to *benefit* differentially from the externality. Second, we are forcing the endogenous amenity to depend only on the ratio of skilled to unskilled workers, although the weight placed on this amenity may differ between skill groups.

E.2 A Neutrality Result

With the model above we have the following result.

Proposition 3 *Consider the counterfactual described in Section 6 and suppose the function $v(y, p)$ represents homothetic preferences. That is, $\hat{a}_{sn} = \kappa$ and $\hat{a}_{un} = 1$ for all n, and $\hat{b}_{in} = 1$ for all i and n. Then for any locations n and m*

$$\frac{\hat{l}_{sn}}{\hat{l}_{un}} = \frac{\hat{l}_{sm}}{\hat{l}_{um}}.$$

A direct implication is that the variance of the log skill ratio does not change between baseline and counterfactual.

Proof. The result from Proposition 2 continues to apply

$$\left(\frac{\hat{l}_{sn}}{\hat{l}_{un}} \right) \propto \left(\frac{\hat{A}_{sn}}{\hat{A}_{un}} \right)^{\frac{(\sigma-1)\theta(1-\tau)}{\sigma+\theta(1-\tau)}} \left(\frac{\hat{B}_{sn}}{\hat{B}_{un}} \right)^{\frac{\sigma}{\sigma+\theta(1-\tau)}}.$$

Using the definitions of productivities and amenities given above, we obtain

$$\left(\frac{\hat{l}_{sn}}{\hat{l}_{un}}\right) \propto \left(\frac{\hat{a}_{sn}}{\hat{a}_{un}}\right)^{\frac{(\sigma-1)\theta(1-\tau)}{\sigma+\theta(1-\tau)}} \left(\frac{\hat{b}_{sn}}{\hat{b}_{un}}\left(\frac{\hat{l}_{sn}}{\hat{l}_{un}}\right)^{\delta_s-\delta_u}\right)^{\frac{\sigma}{\sigma+\theta(1-\tau)}}.$$

Using the values we specified for $\hat{a}$ and $\hat{b}$ we obtain the desired result. ■

Proposition 3 is useful because it tells us that we can regard agglomeration externalities and endogenous amenities as mechanisms which amplify or diminish the strength of the mechanism we emphasize in the body of the paper, i.e the interaction of the rising national skill premium with nonhomothetic housing demand. They do not independently link the skill premium to sorting.

E.3 Parameterization and Counterfactual

We parameterize the agglomeration externality h_n as

$$h_n(l_{sn}, l_{un}) = (l_{sn} + l_{un})^{\zeta}$$

and we set $\zeta = 0.06$, consistent with Rosenthal and Strange (2004).

We parameterize the endogenous amenity using the results of Diamond (2016). She estimates the parameters of an externality of the form

$$B_{in} = b_{in}\left(\frac{l_{sn}}{l_{un}}\right)^{\gamma_a \tilde{\delta}_i}$$

Diamond (2016) reports point estimates for $(\gamma_a, \tilde{\delta}_s, \tilde{\delta}_u)$ of $(2.6, 1.012, 0.274)$.

However, using these directly to calculate our δ_i does not work. The reason is that if we do so, we obtain an exponent on $\frac{\hat{l}_{sn}}{\hat{l}_{un}}$ on the right hand side of the expression above which is greater than 1, implying that (at least for the homothetic case) the interior equilibrium is not stable. The intuition is that these parameters imply a very strong desire for skilled workers to colocate with other skilled workers. This force is so strong that it overcomes the congestion forces on the production side and results in perfect sorting.

For this reason use a lower estimate of γ_a. In particular we choose a value one standard error below Diamond (2016)'s point estimate, which gives $\gamma_a = 1.47$. We therefore emphasize that the results below should only be interpreted as illustrative. They are intended to show how our main results might change in a richer environment, rather than to be precise estimates.

We repeat the counterfactual described in Section 6. We focus on the change in sorting. We find that in this richer environment, sorting rises by 13.4% by 2010 relative to 1980. For reference, our baseline model found an increase of 3.9% and an increase of 32.6% was observed in the data. In results not reported here, we separately investigate the roles of agglomeration versus endogenous amenities. Agglomeration has essentially zero effect; these results are entirely driven by endogenous amenities. It is not surprising that endogenous amenities amplify the mechanism discussed in the paper. As skilled workers concentrate in cities with high housing costs, they raise the amenity value of these cities. The parameterization taken from Diamond (2016) implies skilled workers place a higher value on these amenities, and so this encourages further sorting.

We conclude this section with two points. First, Proposition 3 shows that agglomeration and endogenous amenities *per se* do not introduce new mechanisms linking the rising national skill premium to sorting. Rather, they only matter in the presence of nonhomothetic housing demand. Second, the results of repeating our counterfactual indicate adding these features to our model is likely to amplify rather than weaken the forces discussed in the paper.