

Heterogeneous Workers and Federal Income Taxes in a Spatial Equilibrium

Mark Colas
Federal Reserve Bank of Minneapolis and University of Oregon

Kevin Hutchinson
eBay Inc.

Institute Working Paper 3
November 2017

DOI: <https://doi.org/10.21034/iwp.3>

Keywords: Tax incidence; Worker heterogeneity; Local labor markets

JEL Codes: J31, H22, R13

The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Heterogeneous Workers and Federal Income Taxes in a Spatial Equilibrium*

Mark Colas[†]

Kevin Hutchinson[‡]

November 8, 2017

Abstract

This paper studies the incidence and efficiency of a progressive income tax in a spatial equilibrium. We use US census data to estimate an empirical spatial equilibrium with heterogeneous workers, landowners, and firms. The US income tax shifts skilled workers out of high-productivity cities, leading to a deadweight loss of 2% of tax revenue. Flattening the tax schedule significantly increases welfare inequality between skilled and unskilled workers and does not increase overall worker welfare, as the efficiency gains are captured by landowners. This suggests that progressive income taxes reduce welfare inequality without reducing total worker welfare.

JEL Classification: J31, H22, R13

Keywords: Tax incidence, worker heterogeneity, local labor markets.

*A number of people have helped to improve this paper, but we especially wish to thank Chris Taber, John Kennan, and Jesse Gregory for their guidance throughout. We also greatly appreciate advice and suggestions from Naoki Aizawa, Chao Fu, Tom Holmes, Rasmus Lentz, Eduardo Morales, Ronni Pavan, Nicolas Roys, Ananth Seshadri, Jim Walker, and Wes Wilson. Finally, we extend our thanks to countless visiting seminar speakers and graduate students at the University of Wisconsin. All remaining errors are our own.

[†]Opportunity and Inclusive Growth Institute, Federal Reserve Bank of Minneapolis and the University of Oregon, mcolas@uoregon.edu

[‡]eBay Inc., kphutchinson2@gmail.com

1 Introduction

Wages vary greatly across local labor markets in the United States, reflecting, at least partially, underlying heterogeneity in location-specific productivity.¹ This observation implies that a progressive federal income tax can lead to a geographic misallocation of resources. In particular, workers in highly productive locations will be taxed at a higher rate, thus providing a disincentive to live and work in these places. Some policymakers and researchers have argued for a less progressive tax code on precisely these grounds.²

This recommendation does not, however, acknowledge the distributional aspects of such a policy. More educated workers, in addition to earning higher wages, have been shown to be significantly more mobile than less educated workers.³ A spatial analysis of tax progressivity thus needs to recognize that 1) skilled workers are better able to capture the gains from tax changes via migration, and 2) unskilled workers are less able to out-migrate from places that experience rent increases. Allowing for this type of worker heterogeneity means that making taxes less progressive can reduce deadweight loss, but at the cost of widening welfare inequality between workers. The main contribution of this paper is to better understand the spatial version of the classic trade-off between equity and efficiency.

To accomplish this, we develop and estimate a spatial equilibrium model using US census data, similar to those recently used by [Diamond \(2016\)](#) and [Piyapromdee \(2017\)](#). In the model, the decision of where to live is a static discrete choice; workers choose the city that yields the highest utility in terms of *after-tax* wages, rents, and amenities. Absentee landlords own, and may choose to rent, land for housing. Wages and rents are determined in equilibrium by the location and housing consumption choices of workers. Finally, changes to the

¹For example, [Moretti \(2011\)](#) documents that, in the manufacturing industry, total factor productivity in the most productive US counties is three times as large as in the least productive counties.

²[Moody and Hoffman \(2003\)](#), for example, argue that a “flat tax rate on all income with no standard deduction, for example, would not discriminate against taxpayers living and working in high-cost areas,” while [Eeckhout and Guner \(2017\)](#) argue that the optimal federal tax code in a spatial setting should be less progressive than the current tax code.

³Examples of papers documenting the differential mobility of skilled workers include [Bound and Holzer \(2000\)](#), [Wozniak \(2010\)](#), [Malamud and Wozniak \(2012\)](#), and [Notowidigdo \(2013\)](#).

federal tax code will alter the distribution of after-tax wages across cities, thus leading to changes in the spatial distribution of pretax and posttax wages, rents, populations, and welfare in equilibrium.

We focus on how a progressive income tax differentially affects skilled workers, unskilled workers, and landowners. The distribution of wages across cities varies by skill level, implying that a progressive tax will differentially affect workers of different skill groups. We allow workers of different skill groups to vary in the responsiveness of their location choices to after-tax wages, and thus to be differentially responsive to changes in the tax code. If skilled workers are more mobile, as is typically found in the literature, then they may be better able to capture the gains from changes in income tax policy via migration. Furthermore, workers of different skill levels are imperfect substitutes in production, so that the relative supplies of heterogeneous labor in each location determine local *pretax* wages. Thus, taxes affect workers directly via their effect on posttax wages, but also by changing the distribution of workers across cities and therefore changing equilibrium wages and rents. Finally, the cost of developing land varies across cities, leading to differences in the elasticities of the local housing supply curves. We show that the distribution of housing supply elasticities is an important determinant of the incidence of a progressive tax. If taxes lead workers to sort into cities with more elastic housing supplies, total landowner profits will decrease.

The incidence and deadweight loss of the progressive income tax depends crucially on the elasticity of workers' location choices with respect to earnings, the elasticity of substitution across types of labor inputs, and the elasticities of the housing supply curves. If workers' location choices are highly responsive to after-tax wages, changes in the progressivity of the tax code will lead to large distortions in workers' location choices and large efficiency losses. The elasticity of substitution between different types of workers plays an important role in determining the incidence of the tax across different types of workers; if the tax reform induces skilled workers to leave certain cities, the elasticity of substitution determines the degree to which relative wages of unskilled workers in those cities decrease. Finally, as mentioned above, the elasticity of the housing supply curve in each city is a key determinant of landowner incidence.

One of the main benefits of the framework here is the transparency with which these key elasticities are identified. The worker sorting component of the model,

which determines the elasticity of worker location choices with respect to wages, is estimated using a two-step procedure similar to that in [Berry, Levinsohn, and Pakes \(2004\)](#). The first step transforms the nonlinear discrete choice problem into one that can be solved using standard Instrumental Variables. The second step uses changes in national tax policy to generate variation in after-tax wages at the local labor market level. We estimate the elasticity of location choices with respect to after-tax income using this tax instrument to account for the potential correlation of wages with unobserved amenities. The labor demand and housing supply curves are estimated using the immigrant enclave instrument proposed by [Card \(2009\)](#); immigrant supply shocks from sending countries create variation in the relative supplies of labor and variation in housing demand.

Our first counterfactual experiment measures the deadweight loss and incidence of the 2007 tax code. We find that the progressive tax code leads to a re-sorting of workers into lower productivity places, particularly by skilled workers. More precisely, the overall population of skilled workers in the ten highest productivity cities is about 4% lower compared to the case when there are no income taxes. We find that the tax code leads to a deadweight loss caused by the misallocation of workers across locations of 2.1% of tax revenue, with roughly 60% of the tax incidence falling on skilled workers, 25% falling on unskilled workers, and the remainder falling on landowners.

Our second counterfactual experiment evaluates a revenue-neutral flat tax. The flattening of the tax code leads to a moderate reduction in deadweight loss, from 2.1 cents to 1.3 cents per tax dollar raised. Furthermore, workers sort toward cities with more inelastic housing supply curves in response to the policy change, which leads to large increases in rents, increases in landowner profits, and a slight decrease in total worker welfare. The flattening of the tax code also leads to an increase in between-group welfare inequality. Equivalent variation for unskilled workers increases from \$2,600 to \$3,600 dollars per person, per year, while equivalent variation for skilled workers decreases from nearly \$10,000 to \$8,100 per person, per year. This amounts to a large transfer of the tax burden from skilled to unskilled workers, with no increase in total worker welfare compared to the progressive tax. Taken together, our counterfactuals show that a progressive tax reduces between-group welfare inequality and promotes overall

worker welfare by shifting the distribution of workers toward cities with more elastic housing supply curves.

The two papers most closely related to ours are [Albouy \(2009\)](#) and [Eeckhout and Guner \(2017\)](#), who also study the efficiency losses associated with a federal income tax in a spatial equilibrium. In [Albouy \(2009\)](#), as in the models of [Rosen \(1979\)](#) and [Roback \(1982\)](#), workers are identical and perfectly mobile. A fixed supply of land is used in the production of home goods and tradeable goods, thus creating downward-sloping labor demand curves and upward-sloping home good supply curves. Locations may vary in their amenity values; wages and rents adjust so that utility is equal in all inhabited locations. He calculates the deadweight loss of income taxes and presents policies which can alleviate this deadweight loss.

[Eeckhout and Guner \(2017\)](#) also study the spatial effects of an income tax in a model that includes perfectly mobile homogeneous workers, firms that produce a consumption good, and construction firms that combine the consumption good and land in the production of housing. Additionally, the model allows for agglomeration economies and a congestion externality—higher city populations increase productivity and reduce the amenity value of living there. The presence of these externalities implies a role for a corrective income tax to improve welfare; characterizing this optimal tax is the main goal of their paper, which turns out to still be progressive, but less so than the current US tax code. We view these results to be highly complementary to ours. In particular, our work focuses on distributional effects, which is made possible by relaxing the assumption of homogeneous workers. However, it is precisely this assumption that complicates the characterization of optimal taxes, arising from issues of interpersonal utility comparisons.⁴

Our work departs from these two papers in three main ways. The first distinguishing feature of our research is the addition of heterogeneity across workers. In particular, we allow the values of the amenity associated with each location to

⁴In an earlier version of the paper, we calculated the social welfare maximizing tax progressivity with heterogeneous workers. However, we found that our results were very sensitive to our choice of Pareto weights of different types of workers. As we could not find a reasonable way to choose Pareto weights, we did not pursue this approach further.

separately vary across 32 different demographic groups.⁵ All of these groups are also allowed to have differential attachments to their state of birth. Workers are also defined as either skilled or unskilled workers. These skill groups are assumed to be imperfect substitutes in production and vary in the responsiveness of their location choices to wage differentials across cities. This allows us to study the impact of tax reform on between-group welfare inequality. As mentioned earlier, we find that flattening of the tax code leads to increases in between-group welfare inequality, suggesting the equity concerns of a progressive income tax might outweigh the efficiency concerns.

Second, not only do we allow for different skill levels (high and low) and demographic groups, but *within* each group, workers are allowed to have idiosyncratic preferences over locations, so that the labor supply curves in each location are upward sloping. Furthermore, a worker's utility in a location is a function of the distance between the location and the worker's birthplace. These distinctions have important implications for the elasticity of location choice with respect to wages and thus tax incidence and deadweight loss. Given the importance of the elasticity of location choices in determining tax incidence, we believe our relaxation of the perfect elasticity assumption helps us to accurately estimate the incidence associated with the progressive tax.

Finally, rather than use calibration, we estimate a structural spatial equilibrium model. This allows us to consider a much richer setup in terms of observable and unobservable heterogeneity.

To reiterate, the central contributions of this paper are 1) to quantify the effects of a progressive tax code in a model with heterogeneous workers, 2) to measure the incidence of the tax code on heterogeneous workers and landowners, and 3) to analyze the equity-efficiency trade-off associated with tax progressivity.

2 Model

We build a model of spatial equilibrium, similar to those recently used by [Diamond \(2016\)](#) and [Piyapromdee \(2017\)](#). Locations vary along three dimensions: wages, rents, and amenities. The choice of location is modeled as a static discrete

⁵These groups are defined as combinations of education, potential experience, marital status, and number of children.

choice; workers choose the city that yields the highest utility. We assume that a single, tradeable good is produced in each location and workers of different skill levels are imperfect substitutes in its production, so that the relative supplies of heterogeneous labor in each location determine local wages. Landowners in each city own and may choose to rent land for housing. The cost of developing land for housing varies across cities, implying that changes in the distribution of workers across cities can change total landowner profits. Thus, wages, rents, and population in each location are determined endogenously as equilibrium outcomes.

Our model extends the Rosen-Roback framework along four dimensions, all of which are important for answering our question. First, we allow for two imperfectly substitutable groups of workers in production: skilled and unskilled. This is crucial because the substitutability of workers determines how wages react to changes in the local supply of each group. For example, if tax reform induces more skilled workers to choose a certain location, this will decrease the relative wage of the skilled workers in that location.

Second, we allow for rich heterogeneity in workers' preferences. In particular, we allow workers' preferences to include a premium for living close to their state of birth. Bayer, Keohane, and Timmins (2009) demonstrate the importance of doing so; the authors find that ignoring imperfect mobility of this kind can result in substantially biased estimates of the other preference parameters.⁶ Additionally, workers' preferences may vary based on their marital status, number of children, experience level, and education.

Third, we allow for heterogeneity in preferences over locations, *conditional* on worker type, thus relaxing the assumption of perfect mobility. More precisely, we allow for idiosyncratic location-specific preference shocks, where the variance of the shock is allowed to differ by worker skill. As emphasized by Kline and Moretti (2014) and Busso, Gregory, and Kline (2013), this sort of preference heterogeneity is crucial for analyzing policies when workers are mobile. In our case, these variances govern the location choice elasticities, which are essential for quantifying the impact of tax reform on the resulting spatial equilibrium.

Finally, we incorporate the federal income, state income, and payroll taxes, as well as the corresponding deductions for state income taxes, marital status,

⁶In particular, they find that the marginal willingness to pay for air quality is understated by a factor of three when perfect mobility is assumed.

and number of children. We assume the government uses tax revenue to fund a public good which benefits all workers, regardless of their location.⁷

2.1 Workers

Locations are indexed by j and time is indexed by t . Workers, indexed by i , maximize utility by 1) allocating their resources between a nationally traded consumption good, c_{jt} , and housing, h_{jt} , and 2) choosing the location j that yields the highest utility. Workers belong to a narrowly defined demographic group, d , which is defined by marital status, education level, age and number of children. Workers of different demographic groups vary in their tax levels and preferences over locations. Workers are also characterized as either a skilled (S) or an unskilled (U) worker. We index these broad skill groups by e . Workers of different skill groups are imperfect substitutes in production.

We proceed by first solving the workers' maximization problem, conditional on location. Prices of the consumption good and housing are denoted by p_t and r_{jt} , respectively. Notice that the price of the consumption good is constant across all locations, reflecting the law of one price, which applies because c is tradeable. Locations are also distinguished by their amenity value, Γ_{ijt} . Preferences over the consumption good and housing are assumed to be Cobb-Douglas and are written as

$$u_{ijt} = (1 - \alpha^e) \log(c_{jt}) + \alpha^e \log(h_{jt}) + \Gamma_{ijt}, \quad (1)$$

where α^e is a parameter which corresponds to the agent's optimal budget share of housing.⁸

Let I_{jt}^d denote income earned in location j by workers of group d in time t , where income is simply the hourly wage multiplied by hours worked. Further, let $\mathcal{T}_{jt}^d(I_{jt}^d)$ denote the effective (i.e., average) tax rate, which includes federal income, state income, and federal payroll taxes. Note that, in addition to income

⁷As we do not directly model utility from the public good, we hold the amount of tax revenue fixed across counterfactuals.

⁸See [Davis and Ortalo-Magné \(2011\)](#) for evidence that the budget share of housing is indeed constant across metropolitan areas and time. Furthermore, allowing these parameters to vary by skill, rather than by demographic groups, significantly reduces the number of parameters to be estimated.

level, the tax rate also depends on demographics, location, and time. These dependencies account for differences in state income taxes, differences in income tax deductions by demographic group, and changes in the tax code over time. This tax rate implies the following mapping from pretax income, I_{jt}^d , to posttax income, \tilde{I}_{jt}^d :

$$\tilde{I}_{jt}^d (I_{jt}^d) = [1 - \mathcal{T}_{jt}^d (I_{jt}^d)] I_{jt}^d.$$

We can now write the workers' budget constraint as

$$p_t c_{jt} + r_{jt} h_{jt} = \tilde{I}_{jt}^d (I_{jt}^d),$$

and solving the workers' problem yields the following indirect utility associated with choosing location j :

$$v_{ijt} = \log \left(\frac{\tilde{I}_{jt}^d (I_{jt}^d)}{p_t} \right) - \alpha^e \log \left(\frac{r_{jt}}{p_t} \right) + \Gamma_{ijt}. \quad (2)$$

We assume that workers inelastically supply \mathcal{H}^d hours of labor in all locations, so that the only relevant labor supply decision is where to live, which is now a static discrete choice defined by equation (2).⁹ Workers simply choose the location that maximizes indirect utility.

We now decompose the amenity term, Γ_{ijt} , into five distinct components. In particular,

$$\Gamma_{ijt} = \gamma_{hp}^d \mathbb{I}(j \in Bstate_i) + \gamma_{\text{dist}}^d \phi(j, Bstate_i) + \gamma_{\text{dist}^2}^d \phi^2(j, Bstate_i) + \xi_{jt}^d + \sigma^e \epsilon_{ijt}$$

where $\mathbb{I}(j \in Bstate_i)$ is an indicator for location j being in worker i 's birth state, $\phi(j, Bstate_i)$ and $\phi^2(j, Bstate_i)$ are the distance and squared distance, respectively, between agent i 's birth state and location j , ξ_{jt}^d is a common, unobservable component of amenities, ϵ_{ijt} is an idiosyncratic, stochastic term meant to capture the fact that some workers are more or less attached to certain locations, and σ^e measures the dispersion in ϵ_{ijt} .¹⁰ We assume that ϵ_{ijt} follows a Type 1 Extreme Value distribution.

⁹The assumption of inelastic labor supply is typical in this literature. See, for example, [Moretti \(2013\)](#), [Kline and Moretti \(2014\)](#), or [Busso, Gregory, and Kline \(2013\)](#).

¹⁰We do not allow for the possibility of endogenous amenities, as in [Diamond \(2016\)](#).

The term measuring the dispersion of preference shock, σ^e , plays in an important role in determining the elasticity of worker choices with respect to after-tax income. A large value of σ^e implies that idiosyncratic factors play a large role in determining workers' location choices relative to wages and rents, and thus changes in the tax code will not have a large effect on worker location choices. A small value of σ^e implies that workers will be closer to the margin between two cities and thus will be more responsive to changes in the tax code. A number of papers have found that skilled workers are more responsive to wage changes than unskilled workers in their location choices.¹¹ We allow σ^e to vary to skill level to allow for differential responsiveness to income changes across skill groups.

One benefit of the Cobb-Douglas preferences is that, holding prices constant, flat changes in the tax function do not affect worker location choices, thus allowing us to focus on differences in progressivity. To see this, consider the following mapping from pretax to posttax income:

$$\tilde{I}_{jt}^d(I_{jt}^d) = \lambda f(I_{jt}^d),$$

where λ is a constant and f is some arbitrary function. Changes in λ do not affect the elasticity of posttax income with respect to pretax income. This tax function yields the following indirect utility function:

$$v_{ijt} = \log\left(\frac{f(I_{jt}^d)}{p_t}\right) - \log\left(\frac{\lambda}{p_t}\right) - \alpha^e \log\left(\frac{r_{jt}}{p_t}\right) + \Gamma_{ijt}.$$

The second term, $\log\left(\frac{\lambda}{p_t}\right)$, is constant across locations and therefore does not change an agent's preference ordering across locations. Thus, holding rents and wages constant, changes in λ do not affect worker location choices. This allows us to abstract away changes in the *tax level* and focus on the distortionary effect of *tax progressivity*.¹²

Throughout the rest of the paper, it will be useful to separate the indirect utility of each location into a mean level of utility (i.e., the portion of utility that

¹¹See [Bound and Holzer \(2000\)](#), [Wozniak \(2010\)](#), [Malamud and Wozniak \(2012\)](#), and [Notowidigdo \(2013\)](#), for example.

¹²Note that changes in λ will, however, affect location choices in general equilibrium. A higher value of λ implies that agents have less income to spend on housing. This affects equilibrium rental rates.

is identical for all workers in the same demographic group) and an idiosyncratic component. In particular, let

$$v_{ijt} = \delta_{jt}^d + \gamma_{hp}^d \mathbb{I}(j \in Bstate_i) + \gamma_{\text{dist}}^d \phi(j, Bstate_i) + \gamma_{\text{dist}2}^d \phi^2(j, Bstate_i) + \sigma^e \epsilon_{ijt} \quad (3)$$

where

$$\delta_{jt}^d = \log \left(\frac{\tilde{I}_{jt}^d (I_{jt}^d)}{p_t} \right) - \alpha^e \log \left(\frac{r_{jt}}{p_t} \right) + \xi_{jt}^d. \quad (4)$$

2.2 Firms

Perfectly competitive firms in each labor market use the following constant elasticity of substitution (CES) production function to produce an identical tradeable good, using capital, K , skilled labor, S , and unskilled labor, U , as inputs:

$$Y_{jt} = A_{jt} K_{jt}^\eta L_{jt}^{1-\eta}.$$

The labor supply aggregator, L_{jt} , is given by

$$L_{jt} = [(1 - \theta_{jt}^S) U_{jt}^\rho + \theta_{jt}^S S_{jt}^\rho]^{\frac{1}{\rho}}, \quad (5)$$

where U_{jt} and S_{jt} are defined as the total efficiency units of labor supplied by unskilled and skilled workers, respectively.¹³ Notice that the factor intensities (i.e., the θ_{jt}^S 's) are allowed to vary across labor markets and over time, while the elasticity of substitution between skilled and unskilled labor, $\varsigma \equiv \frac{1}{1-\rho}$, is restricted to be the same across locations and time. Total factor productivity (TFP), A_{jt} , is also allowed to vary across labor markets and over time.¹⁴

¹³In particular, a worker of demographic group d supplies $\mathcal{H}^d \Lambda^d$ efficiency units of labor, where \mathcal{H}^d is number of hours worked and Λ^d is a measure of productivity for workers of demographic d . See the Data Appendix for more details.

¹⁴Another option is to allow for capital-skill complementarity, as in [Krusell et al. \(2000\)](#). However, as we do not have data on capital at the local labor market level, we instead use the production function presented here, which can be estimated without data on physical capital.

The production function exhibits constant returns to scale, and thus relative wages are determined by the ratio of skilled to unskilled workers.¹⁵ The extent to which changes in the ratio of skilled to unskilled workers affect relative wages is governed by the parameter ρ . A high value of ρ implies a high elasticity of substitution, ς , and therefore that changes in the ratio of skilled to unskilled workers will lead to small changes in the wage ratio.¹⁶

Labor markets are perfectly competitive, so that workers are paid their marginal products, which yields the following expressions for wages:

$$\begin{aligned} w_{jt}^S &= \frac{\mathcal{P}_t Y_{jt} \eta}{L_{jt}} L_{jt}^{1-\rho} \theta_{jt}^S S_{jt}^{\rho-1} \\ w_{jt}^U &= \frac{\mathcal{P}_t Y_{jt} \eta}{L_{jt}} L_{jt}^{1-\rho} (1 - \theta_{jt}^S) U_{jt}^{\rho-1}, \end{aligned} \tag{6}$$

where \mathcal{P}_t is the price of the output good. We assume that each city is a small open economy and therefore that the price of the output good is exogenous.

We assume that the capital supply is perfectly elastic and has a rental rate of $\bar{\mathcal{R}}_t$. The firm chooses capital such that the marginal revenue product of capital is equal to this rental price of capital. The marginal revenue product of capital is given by

$$\frac{L_{jt}}{K_{jt}} = \left(\frac{\bar{\mathcal{R}}_t}{\mathcal{P}_t A_{jt} (1 - \eta)} \right)^{1/\eta}. \tag{7}$$

Therefore, the ratio of capital to the labor supply aggregate is determined by the price of capital $\bar{\mathcal{R}}_t$, the goods price \mathcal{P}_t , TFP A_{jt} , and the parameter η . Combining capital demand (7) with the equation for the marginal revenue products of capital (6) yields the following labor demand curves:

¹⁵We have assumed away the possibility of agglomeration in the production function, as in [Glaeser and Gottlieb \(2009\)](#) or [Baum-Snow, Freedman, and Pavan \(Forthcoming\)](#). We leave an investigation of the role of progressive income taxes in a model with agglomeration effects for future work.

¹⁶Another option is to allow for imperfect substitution between multiple education groups, instead of the two we have allowed here. [Ottaviano and Peri \(2012\)](#) estimate a high elasticity of substitution between high school dropouts and graduates and between agents with some college and college graduates.

$$\begin{aligned}
w_{jt}^S &= \tilde{A}_{jt} L_{jt}^{1-\rho} \theta_{jt}^S S_{jt}^{\rho-1} \\
w_{jt}^U &= \tilde{A}_{jt} L_{jt}^{1-\rho} (1 - \theta_{jt}^S) U_{jt}^{\rho-1},
\end{aligned} \tag{8}$$

where

$$\tilde{A}_{jt} = \eta \mathcal{P}_t A_{jt} \left(\frac{\bar{\mathcal{R}}_t}{\mathcal{P}_t A_{jt} (1 - \eta)} \right)^{1/\eta}.$$

Note that, because of the assumptions on capital supply and the functional form of the production function, these labor demand curves do not take capital, K_{jt} as an argument. We can therefore estimate the labor demand curves without data on physical capital.

Finally, taking logs of the two marginal revenue products in (8) and differencing yields the relative labor demand curve, which we will use for estimating the model:

$$\log\left(\frac{w_{jt}^S}{w_{jt}^U}\right) = \log\left(\frac{\theta_{jt}^S}{1 - \theta_{jt}^S}\right) - \frac{1}{\varsigma} \log\left(\frac{S_{jt}}{U_{jt}}\right). \tag{9}$$

2.3 Landowners

Perfectly competitive landowners own plots of land, which they may choose to develop and rent to workers for housing. Land varies in how costly it is for landowners to develop for housing, reflecting differences in opportunity and monetary cost. These plots of land form an upward-sloping housing supply curve: for small quantities of housing, the plots with the lowest cost are rented, implying low housing costs. As quantity increases, increasingly costly plots of land must be rented.

In particular, we follow [Kline and Moretti \(2014\)](#) and parameterize the marginal cost curve such that the inverse housing supply curve in city j , time t , is given by

$$r_{jt} = z_{jt} H_{jt}^{k_j},$$

where H_{jt} is quantity of housing, z_{jt} is a parameter, and k_j is a parameter equal to the inverse elasticity of the housing supply curve (i.e., $\frac{\partial \log r_{jt}}{\partial \log H_{jt}} = k_j$). Taking logs we obtain

$$\log(r_{jt}) = k_j \log(H_{jt}) + \zeta_{jt}, \quad (10)$$

where $\zeta_{jt} = \log(z_{jt})$ represents differences in cost levels across cities and over time.

As workers' optimal fraction of after-tax income spent on housing is given by α^e , we can write total housing demand as

$$H_{jt} = \sum_d N_{jt}^d \frac{\alpha^e \tilde{I}_{jt}^d (I_{jt}^d)}{r_{jt}}, \quad (11)$$

where N_{jt}^d is the total number of workers of demographic d living in city j . Plugging this equation for housing demand into the housing supply curve and rearranging yields the following reduced-form relationship:

$$\log(r_{jt}) = (\nu_1 + \nu_2 \psi_j^{WRI}) \log\left(\sum_d N_{jt}^d \alpha^e \tilde{I}_{jt}^d (I_{jt}^d)\right) + \zeta_{jt}, \quad (12)$$

where the inverse elasticity housing supply with respect to rents is parameterized as $\frac{k_j}{1+k_j} = (\nu_1 + \nu_2 \psi_j^{WRI})$. [Gyourko, Saiz, and Summers \(2008\)](#) use the Wharton Regulation Survey to produce municipality-level measures of the strictness of land use regulations. We aggregate their measures up to the core-based statistical area (CBSA) level to obtain our measure of land use regulations, ψ_j^{WRI} . Increasing housing supply is more costly in CBSAs with stricter land use policies, so we expect ν_2 to be positive.

Profits for a given plot of land are given by the difference between rents and the cost of developing the land. Therefore, total landowners profits in city j are given by the area above the housing supply curve and below the rental rate of housing:

$$\Pi_{jt} = \int_0^{H_{jt}} (r_{jt} - z_{jt} x^{k_j}) dx = \frac{k_j}{1+k_j} r_{jt} H_{jt}.$$

Plugging the above equation for housing demand into landowner profits, we obtain the following expression of landowner profits in city j :

$$\Pi_{jt} = \frac{k_j}{1 + k_j} \sum_d N_{jt}^d \alpha^e \tilde{I}_{jt}^d (I_{jt}^d). \quad (13)$$

Finally, total landowner profits are the sum of all landowner profits across cities:

$$\mathbf{\Pi}_t = \sum_{j'} \Pi_{j't}.$$

To get a sense of how changes in the distribution of workers across cities affect landowner profits, consider moving one worker from a city j' to another city j while holding income levels constant. This will lead to a change in total landlord profit of

$$\Delta \mathbf{\Pi}_t = \alpha_e \left(\frac{k_j}{1 + k_j} \tilde{I}_{jt}^d (I_{jt}^d) - \frac{k_{j'}}{1 + k_{j'}} \tilde{I}_{j't}^d (I_{j't}^d) \right).$$

Total landowner profits, $\mathbf{\Pi}_t$, will increase as workers move into areas with higher posttax income and more inelastic housing supply curves (i.e., higher k_j). The effect of a change in taxes on landowner profit will therefore largely depend on the extent to which the tax incentivizes workers to move away from high-paying cities and toward cities with more elastic housing supply curves.¹⁷

2.4 Equilibrium and Uniqueness

An equilibrium is a set of wages, housing rental rates, and distribution of workers that is consistent with optimization by all agents and with the labor and housing markets clearing. We prove the equilibrium is unique in the case in which production is Cobb-Douglas in labor inputs and there are two skill groups, skilled and unskilled workers, and no additional demographic groups. The proof with multiple demographic groups is a straightforward extension. As such, we replace all demographic d superscripts with skill group e superscripts. Specifically, we denote an equilibrium as vectors of wages for both types of workers, \mathbf{w}^{S^*} and \mathbf{w}^{U^*} , both labor supplies, \mathbf{S}^* and \mathbf{U}^* , and rents, \mathbf{r}^* , where bolded variables de-

¹⁷See the Comparative Statics section in the Appendix for a more formal investigation of the effects of tax progressivity on worker and landowner incidence.

note vectors across all cities (e.g. $\mathbf{S}^* = [S_1^*, S_2^*, \dots, S_j^*, \dots, S_j^*]$) and where we drop all t subscripts, for simplicity.

We now formally state the two assumptions that we use for our uniqueness proof.

Assumption 1. *The tax function takes the following form:*

$$\tilde{I}(I_j) = \lambda I_j^{1-\tau}, \quad (14)$$

where $\lambda > 0$ and $\tau \in [0, 1)$.

For the proof we parameterize the tax function using the tax function in [Heathcote, Storesletten, and Violante \(2014\)](#) and [Eeckhout and Guner \(2017\)](#). The parameter τ governs the progressivity of taxes, and the parameter λ determines the overall level of the tax. While we have not specified a functional form of income taxes in the model, we will use this functional form in the counterfactuals section. The restriction that $\tau \in [0, 1)$ allows for flat taxes ($\tau = 0$) and progressive taxes ($\tau \in (0, 1)$) but disallows complete redistribution ($\tau = 1$) and regressive taxes ($\tau < 0$). Notice also that we dropped the dependencies on demographics and location in the tax function; posttax income is only a function of pretax income.

Assumption 2. *The production function is Cobb-Douglas in labor inputs*

$$\left(i.e., Y_j = \hat{A}_j U_j^{(1-\theta^S)} S_j^{\theta^S} \right).$$

As before, we assume that demand for the output good is perfectly elastic with price \mathcal{P} . Therefore, wages for each skill group are decreasing in the relative labor supply of that skill group $\left(\frac{\partial w_j^S}{\partial (S_j/U_j)} \leq 0 \text{ and } \frac{\partial w_j^U}{\partial (U_j/S_j)} \leq 0 \right)$.¹⁸

Given these assumptions, we can prove the following proposition:

Proposition 1 (Uniqueness). *Given assumptions (1)-(2), the sorting equilibrium is unique.*

¹⁸Note that this is a nested case of the general CES production function with capital presented earlier in which $\rho = 0$, $\theta_j^S = \theta^S$ for all j , and $\hat{A}_j = \bar{A}_j$ where

$$\tilde{A}_{jt} = \eta \mathcal{P}_t A_{jt} \left(\frac{\bar{\mathcal{R}}_t}{\mathcal{P}_t A_{jt} (1 - \eta)} \right)^{1/\eta}.$$

The proof is in the Appendix. The basic intuition is that, if there exists a second equilibrium, then this implies a higher population of skill group e in some city j relative to the original equilibrium. This increase in population must lead to either higher rent or lower income for skill group e in city j in the second equilibrium. We then show that this decrease in income or increase in rents lowers utility in city j to an extent that is inconsistent with a higher population in city j relative to the first equilibrium.

3 Data

Large samples are imperative for our analysis, given its local labor market nature. As such, we use the US Integrated Public Use Microdata Series (IPUMS) to draw data from the 5% samples of the 1980, 1990, and 2000 US census. We also use the 3%, three-year aggregated American Community Survey (ACS) for the years 2005-2007 (Ruggles et al., 2010).¹⁹

Throughout the paper, all individual-level calculations are weighted by the product of total hours worked, the census sampling weights, and a set of geographic weights described below. All local labor market level calculations are weighted by population. More details regarding all aspects of the data are available in the Data Appendix.

3.1 Geography

The two most important considerations we face in choosing a local labor market concept are that 1) locations correspond to distinct labor markets, and 2) they can be compared over time. As such, we use CBSAs as our geographic definition. CBSAs naturally satisfy requirement 1), as they are the Office of Management and Budget's (OMB) official definition of a metropolitan area.²⁰ We achieve consistency and fulfill our second requirement by mapping the most disaggregated

¹⁹We do not use ACS data after 2007 because hours worked are only reported in intervals. In principle, one could impute hours in order to extend the analysis. We prefer to use the non-imputed data.

²⁰The OMB replaced the Metropolitan Statistical Area (MSA) concept with CBSAs in 2003.

geographic units available in the IPUMS data, County Groups (CGs) in 1980 and Public Use Microdata Areas (PUMAs) in 1990 and after, into CBSAs.²¹

There is a fundamental trade-off between the size of the choice set and sample size. Many of the 929 CBSAs do not possess large enough samples to be useful. Therefore, our final decision to make regarding geography is the practical definition of the choice set. More specifically, as outlined below, we construct a number of aggregate measures by CBSA. As we include smaller and smaller CBSAs in the choice set, the precision with which we measure these aggregate measures decreases. Further, we also use a model of individual location choice that allows for a rich set of observable heterogeneity. However, our estimation method requires that we observe at least one individual, within each narrowly defined demographic category, choosing each location. Therefore, we also face a trade-off between the dimension of the choice set and the richness of the observable heterogeneity. We choose to use the 70 largest CBSAs, as defined by population in 1980. Although a relatively small subset of the 929 CBSAs, these 70 locations make up approximately 60% of the entire US population. Further, we map individuals that do not live in one of these 70 areas into their corresponding census division, creating nine additional choices.²²

3.2 CBSA-Level Data

We use four main variables at the CBSA level: wages, labor supply, housing rents, and a measure of housing supply elasticity. To maintain comparability with the broader wage inequality literature, we follow [Autor, Katz, and Kearney \(2008\)](#) (AKK) closely in constructing both the wage and labor supply series, though some choices are necessarily different given that we use different data sets and different geographies.

The main concern in constructing average wage levels is comparability. Ultimately, we want to focus on differences in skill prices across labor markets and over time. Toward that end, we restrict workers to be full-time, full-year FTFY)

²¹This is the same procedure used by [Dorn \(2009\)](#) and [Autor and Dorn \(2013\)](#) to map CGs/PUMAs into commuting zones.

²²It is worth noting here that we take the set of 70 cities as a model primitive even though which cities constituted the largest 70 cities in 1980 was itself an equilibrium outcome.

wage earners.²³ Focusing on FTFY workers eliminates concerns that wage differences across labor markets, or over time, could be due to differences in labor force attachment.²⁴ For that same reason, we also focus on male workers only. Second, we perform a composition adjustment to control for differences in demographics by adopting the cell approach commonly used in this literature.²⁵

The most important criterion in selecting a labor supply measure is that it represents the total quantity of labor in a location. Accordingly, we include part-time, self-employed, and female workers, not just FTFY male wage earners. As in AKK, we form two samples, quantity and price. The quantity sample divides total hours worked by all employed workers into gender-education-experience cells. The price sample contains efficiency weights for each gender-education-experience cell, where the weight is the corresponding mean hourly log wage. The final supply measure is the product of the quantity sample and the price sample for each gender-education-experience cell.

The biggest worry in producing a measure of housing costs across CBSAs is that it reflects the user cost of housing. As such, we use data only on renters, as home prices reflect both the current user cost and expected future price changes.²⁶ A second concern is the comparability of housing units across CBSAs. Therefore, we run a hedonic regression of gross rent (which includes utilities) on a set of housing characteristics and a set of CBSA fixed effects, separately by year. The rent index is then generated by the predicted values from the hedonic regressions, holding the set of housing characteristics fixed across CBSAs and time.

Finally, we use the Wharton Residential Land Use Regulation Index (WR-LURI), proposed by [Gyourko, Saiz, and Summers \(2008\)](#).²⁷ The index is based on a nationwide survey of local land use regulations, with the basic idea being that regulation makes it more costly to build, making the local housing supply curve more inelastic.

²³In particular, we drop individuals that worked less than 40 weeks annually and less than 35 hours weekly.

²⁴It also reduces concern over measurement error in wages. See [Baum-Snow and Neal \(2009\)](#).

²⁵See, for example, [Katz and Murphy \(1992\)](#), [Card and Lemieux \(2001\)](#), or AKK. The details of this composition adjustment are in the Data Appendix.

²⁶See [Davis and Ortalo-Magné \(2011\)](#) for a similar argument.

²⁷This data set can be downloaded at <http://real.wharton.upenn.edu/gyourko/landusesurvey.html>.

3.3 Individual Choice Data

We restrict the individual choice sample to FTFY male workers that identify themselves as the head of household. Once again, we make this restriction to minimize concerns about labor force attachment. We also drop immigrants because of concerns about tax compliance. Further, we make use of information on an individual’s state of birth, which we will use to define a home premium in our model.

As mentioned above, we allow for a rich set of observable heterogeneity. First, we split individuals into four education categories: high school equivalents, some college, college graduate, and post college.²⁸ We further split the sample by marital status (single or married) and work experience (those with less than 20 years of potential experience are defined to be “less experienced,” while those with more than 20 years are categorized as “more experienced”). Our final demographic characteristic is number of children. For married workers, we use three categories: zero children, one child, and two or more children. The vast majority of single workers in our sample do not have children, so we make the assumption that all single workers have no children. This gives us 32 distinct demographic groups.

3.4 Tax Calculations

We perform our tax calculations using the NBER’s TAXSIM, a tax calculator that replicates the federal income tax code in a given year, accounting for differences in state income taxes, the deduction of state income taxes in calculating federal taxable income, and the differential deductions afforded to varying demographic groups.

4 Estimation

The parameters to be estimated include the parameters of the production function, the parameters of the housing supply curve, and the worker preference

²⁸High school equivalents are a combination of high school dropouts and high school graduates. Our next revision will split this group out.

parameters. For computational simplicity, we estimate these three sets of parameters separately. As mentioned earlier, precise estimation of the elasticity of location choice, the elasticity of substitution, and the elasticity of the housing supply curve in each city is crucial to accurately measuring the deadweight loss and incidence of the income tax. In what follows, we explain how each set of parameters is estimated and how these key elasticities are identified.

4.1 Labor Supply

Our approach for estimating the labor supply component of the model closely mirrors the procedure commonly used for estimating differentiated product demand systems with microdata (i.e., [Berry, Levinsohn, and Pakes \(2004\)](#), which we refer to as BLP throughout). In particular, we estimate the parameters in two steps, where the first step estimates the home premiums, distance and distance squared parameters, and mean utilities using maximum likelihood, and the second step uncovers the wage and rent preference parameters, using instrumental variables to deal with the endogeneity of wages and rents.

Equations (3) and (4) are the basis for estimating the underlying preference parameters. We proceed by normalizing both the location and scale of these equations and redefining the parameters accordingly. In particular, we normalize the mean utility of location one to zero and divide through by σ^e , which yields

$$v_{ijt} = \delta_{jt}^d + \beta_{hp}^d \mathbb{I}(j \in Bstate_i) + \beta_{dist}^d \phi(j, Bstate_i) + \beta_{dist2}^d \phi^2(j, Bstate_i) + \epsilon_{ijt} \quad (3')$$

and

$$\delta_{jt}^d = \beta_w^e \log \left(\frac{\tilde{I}_{jt}^d (I_{jt}^d)}{p_t} \right) + \beta_r^e \log \left(\frac{r_{jt}}{p_t} \right) + \xi_{jt}^d, \quad (4')$$

where $\beta_w^e \equiv \frac{1}{\sigma^e}$, $\beta_r^e \equiv -\frac{\alpha^e}{\sigma^e}$, $\beta_{hp}^d \equiv \frac{\gamma_{hp}^d}{\sigma^e}$, $\beta_{dist}^d \equiv \frac{\gamma_{dist}^d}{\sigma^e}$, and $\beta_{dist2}^d \equiv \frac{\gamma_{dist2}^d}{\sigma^e}$. Abusing notation, we also have $\delta_{jt}^d \equiv \frac{\delta_{jt}^d}{\sigma^e}$, $v_{ijt} \equiv \frac{v_{ijt}}{\sigma^e}$, and $\xi_{jt}^d \equiv \frac{\xi_{jt}^d}{\sigma^e}$. Our goal is to estimate the vectors $\boldsymbol{\delta}_t^d$ and $\boldsymbol{\beta}^d \equiv [\beta_{hp}^d \ \beta_{dist}^d \ \beta_{dist2}^d]$, which is done in the first step, along with β_w^e and β_r^e , which is done in the second step.

Assuming that ϵ_{ijt} is distributed i.i.d. according to the Type 1 Extreme Value distribution, we can estimate δ_t^d and β^d using maximum likelihood. In particular, given our distributional assumption, the choice probabilities have the following closed-form solution:

$$P_{ijt}^d = \frac{e^{\delta_{jt}^d + \beta_{hp}^d \mathbb{I}(j \in Bstate_i) + \beta_{dist}^d \phi(j, Bstate_i) + \beta_{dist2}^d \phi^2(j, Bstate_i)}}{\sum_{j'=1}^J e^{\delta_{j't}^d + \beta_{hp}^d \mathbb{I}(j' \in Bstate_i) + \beta_{dist}^d \phi(j', Bstate_i) + \beta_{dist2}^d \phi^2(j', Bstate_i)}}, \quad (15)$$

and the corresponding log-likelihood function is

$$\mathcal{L}_t^d(\beta^d, \delta_t^d) = \sum_{i=1}^{N_t^d} \sum_{j=1}^J \mathbb{I}_j^i \log(P_{ijt}^d), \quad (16)$$

where \mathbb{I}_j^i is an indicator equal to one if individual i lives in location j and zero otherwise.²⁹

The second step uses our estimates from the first step to uncover the underlying preference parameters. In particular, we pool all the δ_{jt}^d 's within each skill group and estimate

$$\delta_{jt}^d = \beta_w^e \log\left(\frac{\tilde{I}_{jt}^d(I_{jt}^d)}{p_t}\right) - \beta_r^e \log\left(\frac{r_{jt}}{p_t}\right) + \xi_{jt}^d \quad (17)$$

using IV to address the endogeneity of wages and rents.

First, consider identification of the wage coefficient, β_w^e . The equilibrium nature of the model mechanically induces a correlation between after-tax income, $\tilde{I}_{jt}^d(I_{jt}^d)$, and the unobserved amenity, ξ_{jt}^d . If some place becomes unobservably nicer for only skilled workers, this will induce in-migration of skilled workers, thus driving down their wages.

To address this issue, we take advantage of the fact that our model explicitly accounts for income taxes. Typically, sorting models of this type (Diamond, 2016; Piyapromdee, 2017) need variation in labor demand, but in our case we need either variation in labor demand or variation in income taxes. We use the latter by constructing instruments similar to those developed in Gruber and Saez (2002). In particular, we use changes in the federal income tax code to generate

²⁹Computationally, we invert the choice probabilities using the contraction mapping in Berry (1994) to obtain the unique δ_t^d associated with every β^d .

local labor market variation in after-tax income. We implement this instrument by calculating the change in after-tax income that would have occurred from changes in the tax code, had there been no local changes in labor demand. More formally, we construct our instrument as

$$\Delta Z_j^d(t) = \frac{\tilde{I}_{jt}^d(I_{j,t-1}^d; \mathcal{S}_{t-1}) - \tilde{I}_{jt-1}^d(I_{j,t-1}^d; \mathcal{S}_{t-1})}{\tilde{I}_{jt-1}^d(I_{j,t-1}^d; \mathcal{S}_{t-1})}, \quad (18)$$

where $\tilde{I}_{jt}^d(\cdot; \mathcal{S}_{t-1})$ maps pretax income into posttax income, where the subscript on \tilde{I}_{jt}^d denotes the tax code in year t given demographics d , and the \mathcal{S}_{t-1} denotes the state income tax code. In words, equation (18) holds local income, $I_{j,t-1}^d$ and state income tax codes, \mathcal{S}_{t-1} , constant, while using changes in the national tax code.

Figure (1) provides some suggestive evidence that changes in the national tax code in the data have led to significant changes in differences in after-tax income. Panel A summarizes the evolution of federal income taxes in the US between 1980 and 2007. In particular, panel A displays effective income tax schedules, for single individuals with no children, in all four of our sample years, where effective tax rates are defined as the fraction of income paid in taxes. For example, someone making \$40,000 in 1980 (in year 2000 dollars) paid approximately 18% of their income in taxes, while someone making \$40,000 in 1990 (again, in year 2000 dollars) paid about 14% of their income in taxes.³⁰

Panel B of Figure (1) shows an example of the local labor market variation generated by these tax changes. In particular, this figure plots $\Delta Z_j^S(1990)$ for workers with at least a college degree, by CBSA. To reiterate, we hold income levels constant at their 1980 pretax levels and calculate the corresponding after-tax income under both the 1980 and 1990 tax codes, holding both state income taxes and deductions constant. Changes are then constructed as the within-CBSA percent change in after-tax income that results from moving to the 1990 tax code from the 1980 tax code. We can see that there is substantial heterogeneity in these changes, ranging from 4% to 7%. This heterogeneity suggests that these changes to the tax code provided skilled workers with an incentive to live in the higher-wage CBSAs as a direct implication of a less progressive tax schedule.

³⁰Note that these rates do not include state or payroll taxes.

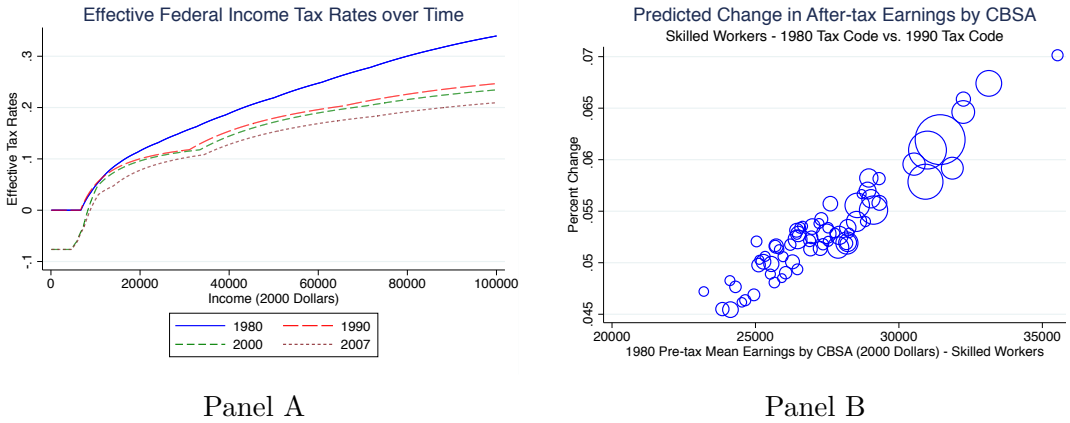


Figure 1: Panel A displays effective federal income tax schedules over time, which are calculated using TAXSIM for single individuals with no children. Panel B uses data from the 5% samples of the 1980 census (Ruggles et al. 2010) to construct composition-adjusted skilled wage levels by CBSA.

The validity of the instrument relies on two identifying assumptions. The first is that changes in the federal income tax code are not related to local changes in unobserved amenities. Second, we need that lagged income levels are uncorrelated with changes in unobserved amenities, that is,

$$E[I_{j,t-1}^d \cdot \Delta \xi_{jt}^d] = 0.$$

To identify β_r^e , we follow Diamond (2016) and interact $\Delta Z_j^d(t)$ with our measure of housing supply elasticity. The idea is that as workers in-migrate to take advantage of higher after-tax income, cities with more inelastic housing supply curves will see rents bid up faster. This creates variation in rent levels across cities that is assumed to be exogenous to changes in unobserved amenities. Workers' responsiveness to these rent changes is used to identify β_r^e . Formally, we assume

$$E[(\Delta Z_{jt}^d \cdot \psi_j^{WRI}) \cdot \Delta \xi_{jt}^d] = 0.$$

In our preferred specification, we also use a Bartik instrument when estimating labor supply (Bartik, 1991). The instrument interacts the lagged industry composition of each city with current national changes in total hours worked in each industry. Conceptually, cities that historically have large concentrations of industries that are currently growing will see faster wage growth than cities which

historically have large concentrations of industries which are currently declining. Formally, we can write the instrument as follows:

$$\text{Bartik}_{jt}(e) = \sum_{m=1}^M \kappa_{mj}^{80}(e) \text{NatHoursInd}_{mt}(e),$$

where m indexes industries, $\kappa_{mj}^{80}(e)$ is total hours of labor by workers of skill level e in city j in 1980 as a fraction of the total hours worked in all industries in city j in 1980, and NatHoursInd_{mt} is the total hours worked nationally in industry m in time t . The instrument creates variation in city-level wages that are assumed to be orthogonal to changes in unobserved amenities.

We include both the Bartik instrument and the Bartik instrument interacted with our measure of housing supply elasticity as instruments. Our estimates of β_w^e and β_r^e when we do not use the Bartik instrument are qualitatively similar but are larger in magnitude and less precisely estimated.

4.2 Labor Demand

We estimate the labor demand parameters of our model using the relative labor demand curve defined in (9). First, we parameterize the factor intensity parameters as

$$\log\left(\frac{\theta_{jt}^S}{1 - \theta_{jt}^S}\right) = \alpha_0^t + \alpha_1 (t \times \log \text{Pop}80_j) + \alpha_2 (t \times \log \text{Pop}80_j^2) + \mu_j + \varepsilon_{jt},$$

where $\log \text{Pop}80_j$ is the log population of city j , measured in 1980. Taking differences of the relative labor demand curve yields the following estimating equation:

$$\Delta \log\left(\frac{w_{jt}^S}{w_{jt}^U}\right) = \Delta \alpha_0^t + \alpha_1 \log \text{Pop}80_j + \alpha_2 \log \text{Pop}80_j^2 - \frac{1}{\varsigma} \Delta \log\left(\frac{S_{jt}}{U_{jt}}\right) + \Delta \varepsilon_{jt}. \quad (19)$$

The concern in estimating equation (19) is that unobserved changes in skill-biased labor demand ($\Delta \varepsilon_{jt}$) induce changes in the quantities of skilled labor (S_{jt}) across labor markets. This will lead to a correlation between unobserved changes

in skill-biased labor demand and skilled labor, and, therefore, biased estimates of the elasticity of substitution ς .

Therefore, we estimate equation (19) using the instruments proposed in Card (2009) and Moretti (2004) to instrument for changes in the labor supply ratio $\left(\Delta \log\left(\frac{S_{jt}}{U_{jt}}\right)\right)$. In particular, the Card instrument is constructed as

$$\text{Card}_{jt}(e) = \sum_{g=1}^G \mu_{gj}^{80}(e) \text{Imm}_{gt}(e),$$

where $\text{Imm}_{gt}(e)$ is the national inflow rate of immigrants with education level e from country g in time t and $\mu_{gj}^{80}(e)$ is the share of total immigrants from country g of education level e living in labor market j in 1980. The intuition is that the historical distribution of immigrants from a given country g is correlated with the distribution of new immigrants from country g but is uncorrelated with current labor market conditions. For example, a large fraction of unskilled Mexican immigrants have traditionally settled in Los Angeles, but only a small fraction have settled in New York. Therefore, if a shock in Mexico leads to a large inflow of unskilled Mexican immigrants to the United States, a large number of these immigrants will settle in Los Angeles relative to New York. This creates variation in labor supplies across the two cities that is assumed to be uncorrelated with labor demand shocks in the two cities.

We also use the Moretti instrument to predict changes in the quantities of skilled and unskilled labor. In particular, the instrument interacts the long-term trend of increasing educational attainment with the lagged age structure of labor market. For instance, labor markets that are disproportionately young or old are predicted to have larger increases in skilled labor. This is because the young are more likely to obtain education and the old are less likely to be educated and will be leaving the labor force. More formally, we predict hours as

$$\text{Moretti}_{jt}(e) = \sum_{l=1}^L \omega_{lj}^{80}(e) \text{NatHoursAge}_{lt}(e),$$

where $\omega_{lj}^{80}(e)$ is the share of hours worked by age group l in labor market j with education level e and $\text{NatHoursAge}_{lt}(e)$ is national hours worked by group l with education level e in time t . To be clear, the denominator of $\omega_{lj}^{80}(e)$ is total hours

worked by skill group e in labor market j . The predicted hours measures are then used to predict changes in relative labor supply across cities.

4.3 Housing Supply

Now we turn to the estimation of the housing supply curve in equation (10), which maps total housing consumption, H_{jt} , into a local rent level, r_{jt} . Taking first differences of equation (12), the reduced-form equation for housing rents, we obtain our estimating equation for housing supply:

$$\Delta \log(r_{jt}) = (\nu_1 + \nu_2 \psi_j^{WRI}) \Delta \log\left(\sum_d N_{jt}^d \hat{\alpha}^e \tilde{I}_{jt}^d(I_{jt}^d)\right) + \Delta \tilde{\zeta}_{jt} \quad (20)$$

where $\hat{\alpha}^e$ is the estimate of the budget share of housing (described above), ν_1 and ν_2 are parameters to be estimated, and r_{jt} , N_{jt}^d , and $\tilde{I}_{jt}^d(I_{jt}^d)$ are observed in the data.

The estimation strategy is to use variation in housing demand to identify housing supply. However, as with the wage equations, the concern with estimating equation (10) via least squares is that $\Delta \tilde{\zeta}_{jt}$ will be correlated with $\Delta \log(N_{jt}^d)$ because workers prefer locations with lower rents. Therefore, we again utilize the Card instrument and Bartik instrument to instrument for changes in population. The identifying assumptions are that historical immigrant settlement patterns and historical industry compositions are uncorrelated with current changes in housing supply shifters.

5 Results

5.1 Parameter Estimates

Estimates for all the key parameters are displayed in Table (1). This section discusses each of the estimates related to the labor demand, housing supply, and labor supply components of the model. First, we estimate an elasticity of substitution, ς , of 3.71, which is on the higher end of estimates in the literature, but within reason. In particular, studies that use local labor market variation (e.g., Card (2009)) tend to estimate higher elasticities of substitution than those

using time-series variation at the national level (e.g., [Katz and Murphy \(1992\)](#) or [Heckman, Lochner, and Taber \(1998\)](#)).

The next panel shows the parameters governing the housing supply curves. We estimate ν_1 to be 0.31 and ν_2 , the interaction between housing consumed and the Wharton Regulation Index, to be 0.12. The signs of these estimates imply that more housing consumed in a local labor market leads to higher rents in that location and this rent increase is larger in labor markets with more inelastic housing supply curves. To facilitate interpretation, [Table \(2\)](#) displays estimates of the inverse housing supply elasticity (k_j) for selected CBSAs. We can see that there is substantial variation in the inverse housing supply elasticity across cities, ranging from 0.27 to 1.13. The mean inverse housing supply elasticity is 0.57, similar to [Saiz \(2010\)](#) and [Piyapromdee \(2017\)](#), and higher than the mean housing supply elasticity in [Diamond \(2016\)](#).

The bottom panel of [Table \(1\)](#) contains estimates for the labor supply component of the model. We estimate β_w^e to be 12.05 and 10.36, respectively, for skilled and unskilled workers. To get a better sense of what our estimates imply for mobility, we simulate the equilibrium changes in local population associated with a local increase in income. We do this by separately giving a 1% income subsidy for workers of a given education level in each city and calculating the resulting equilibrium.³¹ We find that these general equilibrium elasticities are considerably lower than the partial equilibrium elasticities, with a mean general equilibrium elasticity of 2.6 for unskilled workers and 3.5 for skilled workers.

Our estimates of β_w^e are larger than the parameters on pretax wages in [Diamond \(2016\)](#) and [Piyapromdee \(2017\)](#). As the tax is progressive, changes in log pretax wages translate to smaller changes in log posttax income. This should lead to larger coefficient estimates.

Next, we estimate β_r^e , the partial equilibrium elasticity of the location choice probability with respect to rents, to be -5.45 and -8.62, respectively, for skilled and unskilled workers. Together, our estimates of β_w and β_r imply the budget shares of housing, α^e , to be 0.44 and 0.64, respectively, for skilled and unskilled workers.

³¹Specifically, we assume that wages for skill group e in city j' are given by $\hat{I}_{j't}^e = 1.01I_{j't}^e$, where $1.01I_{j't}^e$ is the equilibrium wage without the subsidy.

I. Labor Demand			
ς : Elasticity of Sub.	3.71		
	(1.52)		
II. Housing Supply			
ν_1 : Baseline	.32	ν_2 : Regulation	.09
	(.07)		(.04)
III. Labor Supply			
	Low Skill	High Skill	
β_w^e : Wage	10.36	12.05	
	(4.77)	(2.47)	
β_r^e : Rent	-6.65	-5.28	
	(3.07)	(1.00)	
α^e : Share Housing	0.64	0.44	
	(0.03)	(0.03)	

Table 1: Parameter Estimates. Standard errors in parentheses.

The estimates of the birthplace premium, distance, and distance squared estimates for each demographic group are available on request. Generally speaking, the birthplace premium and the distance cost are decreasing in education, reflecting the greater propensity of skilled agents to live away from their birth state. The estimates of the distance squared parameters suggest that cost of living far from one’s birth is concave in distance.

Rank	City	Inv. Housing Supply Elasticity
1	Honolulu, HI	1.13
2	Worcester, MA	1.03
3	Boston-Cambridge-Quincy, MA-NH	0.97
...		
68	Kansas City, MO-KS	0.34
69	Baton Rouge, LA	0.34
70	New Orleans-Metairie-Kenner, LA	0.27

Table 2: Estimates of inverse housing supply elasticity for selected CBSAs. The inverse housing supply elasticity is calculated as $k_j = \frac{\nu_1 + \nu_2 \psi_j^{WRI}}{1 - (\nu_1 + \nu_2 \psi_j^{WRI})}$.

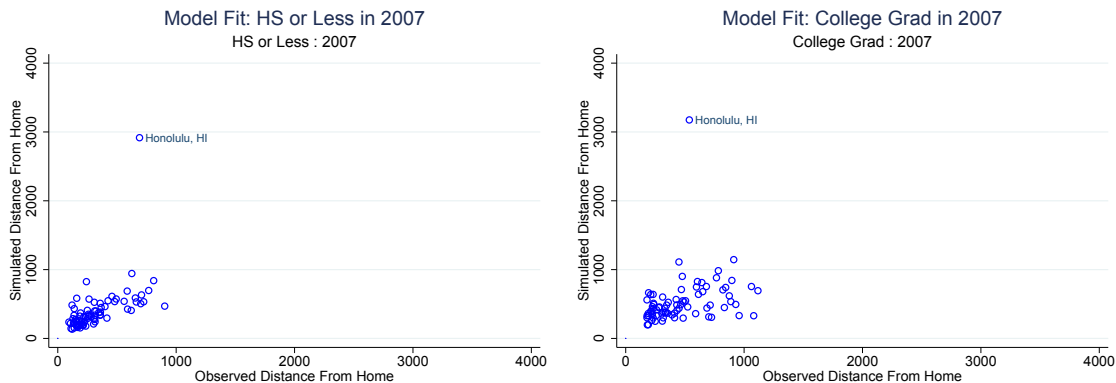


Figure 2: This figure shows the average distance traveled from birth state in each city in the data and predicted by the model for two of the four education groups. Each bubble represents a CBSA, and the size is proportional to the city’s total population. The horizontal axis is the average distance from birth state in the data; the vertical axis is the model’s prediction. The results for other years and other education groups are similar and are available on request.

5.2 Model Fit

In this section, we analyze how well our model can replicate the data. Recall that we estimate separate logits for each city-demographic group combination, which implies that we will exactly match the population of each demographic group in each city. Therefore, we plot the simulated and observed average distance between an agent’s birth state and chosen city for each city. Figure (2) plots these average distances for each education level for the year 2007. Each dot represents a CBSA, and the size of the dot is proportional to population. Honolulu is an obvious outlier: the simulated average distance from the birth state is nearly 3,000 miles compared to under 1,000 miles in the data. Otherwise, the model appears to fit this aspect of the data fairly well.

6 Counterfactuals

Next we analyze the effects of progressivity of the federal income tax code on wages, rent, welfare, and location choices. Before proceeding to the counterfactual results, we first explain the tax function we use to simulate changes in

the progressivity of the tax code and the calculation of deadweight loss and tax incidence.

6.1 The Tax Function

When estimating the model, we used TAXSIM to calculate the federal income, state income, and payroll taxes based on an agent’s location and demographics. For our counterfactuals, we seek a parameterization of the tax code which allows us to change the progressivity of the tax code while holding total tax revenue constant and allows us to abstract away from the distortions caused by state income taxes and payroll taxes. We therefore set income and payroll taxes to zero and parameterize the federal income tax function using the tax function in [Heathcote, Storesletten, and Violante \(2014\)](#) and also used by [Eeckhout and Guner \(2017\)](#). In particular, this family of tax functions is given by

$$\tilde{I}(I) = \lambda I^{1-\tau}, \tag{21}$$

where I is pretax income, \tilde{I} is posttax income, τ is a parameter which governs the progressivity of taxes, and λ is a parameter which determines the overall level of taxes. The effective tax rate is given by

$$\mathcal{T}(I) = 1 - \lambda I^{-\tau}.$$

This functional form allows for a variety of different tax codes. For instance, $\tau = 0$ is a flat tax with marginal/average rate equal to $1 - \lambda$, while $\tau = 1$ is complete redistribution with workers receiving an after-tax income of λ . Taxes are progressive when $\tau \in (0, 1)$ and regressive for $\tau < 0$.

In the following counterfactuals, we change τ to reflect differences in progressivity across tax regimes and adjust λ to keep tax revenue constant across counterfactuals. As such, we can compare the effects of changes in progressivity of the tax code while abstracting from changes in tax revenue.

6.2 Deadweight Loss Calculations

We calculate deadweight loss as the total equivalent variation of workers and landowners minus revenue raised from a tax.³² Conceptually, we can think of the equivalent variation as the dollar amount workers and landowners in an equilibrium without taxes would be willing to pay to avoid switching to the taxed equilibrium. Starting in an equilibrium with no taxes, the equivalent variation for a given worker is the lump-sum payment that would give the worker the same utility as the taxed equilibrium. As landowners are profit maximizers, the equivalent variation for landowners is simply the difference in landowner profits between the no-tax and taxed equilibrium.

Specifically, denote a given tax regime as (λ, τ) . Abusing notation and dropping time subscripts, a tax regime raises tax revenue $\text{TaxRev}(\lambda, \tau)$ and leads to total landowner profits $\Pi(\lambda, \tau)$. The change in landowner profits as a result of the tax is given by $(\Pi(\lambda, \tau) - \Pi(1, 0))$, where $\Pi(1, 0)$ is landowner profits in the equilibrium with no income taxes. As noted above, the equivalent variation for a worker measures the lump-sum transfers that, starting in the no-tax equilibrium, bring a worker to the utility level in the tax equilibrium. We calculate this equivalent variation numerically. Given the wages and rents implied by the no-tax equilibrium, we search for the lump-sum tax for each individual agent that makes their realized utility equal to their realized utility level in the taxed equilibrium.

The deadweight loss is then the sum of workers' equivalent variation and the difference in landowner profits minus tax revenue, measured as a fraction of total tax revenue:

$$DWL(\lambda, \tau) = \frac{\sum_i (EV_i) + \Pi(1, 0) - \Pi(\lambda, \tau) - \text{TaxRev}(\lambda, \tau)}{\text{TaxRev}(\lambda, \tau)} \quad (22)$$

6.3 The Effect of the 2007 Tax Progressivity

In this section, we simulate changing from an equilibrium with no taxes to the 2007 tax progressivity. In order to compute the counterfactual distribution of populations, wages, rents, and the corresponding expected utilities of workers,

³²Firms earn zero profit in any equilibrium and thus do not factor into deadweight loss calculations.

	Equivalent Variation		Δ Landlord Profits	DWL
	High	Low		
2007 Tax Code	59.5%	26.6%	16.0%	2.1%
Flat Tax	49.1%	37.3%	14.9%	1.3%

Table 3: All values are displayed as a percentage of tax revenue. For all counterfactuals, unobserved amenities, tax revenue, and agent demographics are fixed at their 2007 levels.

we fix workers' characteristics (demographics, education, birth states, etc.), the value of unobserved amenities for each demographic group, and labor demand parameters to their 2007 values. We estimate the progressivity level for 2007 as $\tau_{2007} = .08$.

The effects of the tax progressivity relative to an equilibrium with no taxes are displayed in Figure (3). The tax decreases the take-home pay of skilled workers in high-productivity cities. As a result, skilled workers move away from high-productivity cities, decreasing the college share, decreasing pretax wages for unskilled workers, decreasing rents, and increasing pretax wages for skilled workers in high-productivity cities. Specifically, the population of skilled workers in the 10 highest-paying cities decreases by nearly 4%. This results in a wage increase of 0.07% for skilled workers and a wage decrease of 0.7% for unskilled workers. Rents in these cities are 7% lower as a result of the tax.

The tax progressivity also decreases the take-home pay of unskilled workers in high-productivity cities relative to less productive cities. However, as unskilled workers spend a higher fraction of their income on local goods, the increase in rents induced by the sorting of skilled workers into high-productivity cities disincentivizes unskilled workers from living in productive cities. As a result, the change in the tax code leads to a small decrease in the proportion of unskilled workers living in productive cities.

The first row of Table (3) shows the tax incidence and deadweight loss of the 2007 tax code. The first two entries show the worker incidence of the tax code, measured as equivalent variation divided by tax revenue. Skilled workers face the largest incidence of the tax: total equivalent variation for skilled workers is equal to 60% of tax revenue, equal to nearly \$10,000 per worker, per year. Unskilled workers incur a smaller fraction of the tax incidence, at 27% of tax revenue, or roughly \$2,600 a person, per year. Landowners also suffer significant losses as

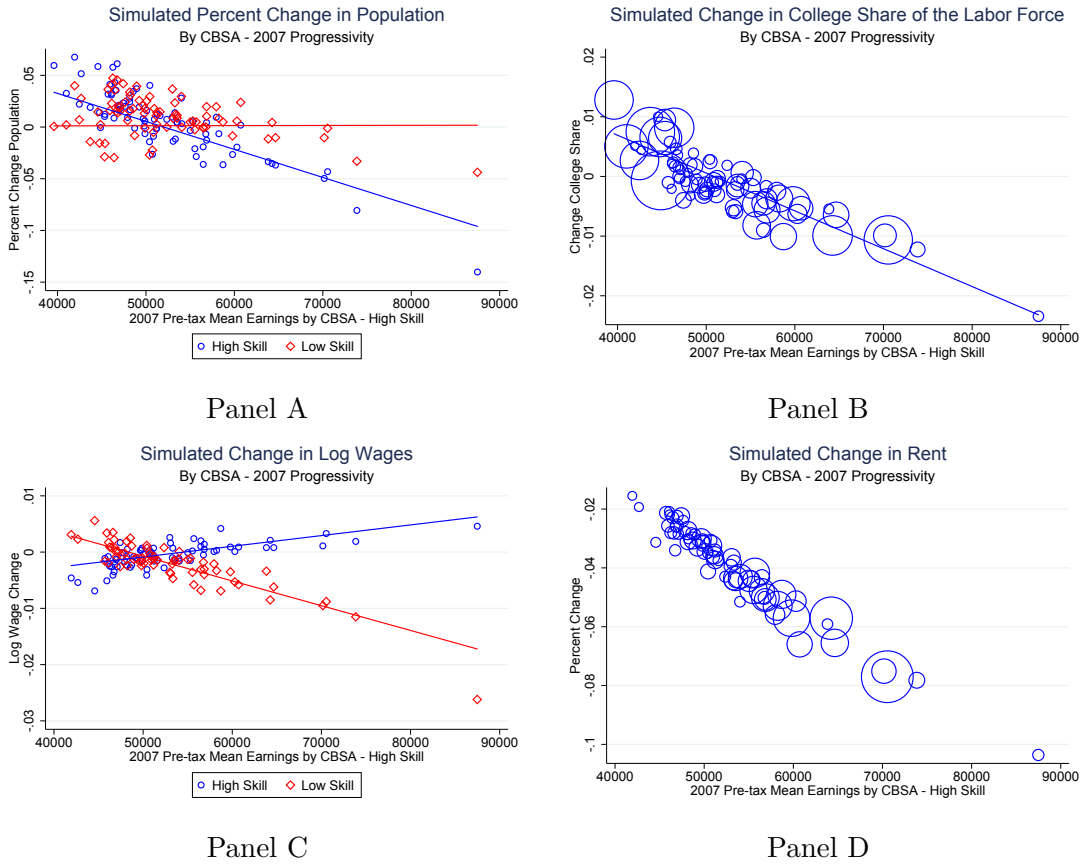


Figure 3: This figure shows the counterfactual changes in population levels (A), college shares (B), log wages (C), and rent (D) as a result of changing from a no-tax equilibrium to the 2007 tax progressivity. Each dot represents a CBSA. The horizontal axis is the 2007 pretax mean earnings for skilled workers. Unobserved amenities, agent demographics, and labor demand parameters are held fixed at their 2007 levels. Bubbles are proportional to the 1980 CBSA population in panels B and D.

a result of the tax, as workers have lower earnings overall and as workers move away from more productive cities. Total landowner profit decreases by an amount equal to 16% of tax revenue. Overall, this implies a deadweight loss of 2.1 cents for each dollar of tax revenue raised.

6.4 A Flat Tax

To isolate the effects of the tax progressivity, we simulate changing from the 2007 tax progressivity to a flat tax and choose λ to keep tax revenue constant

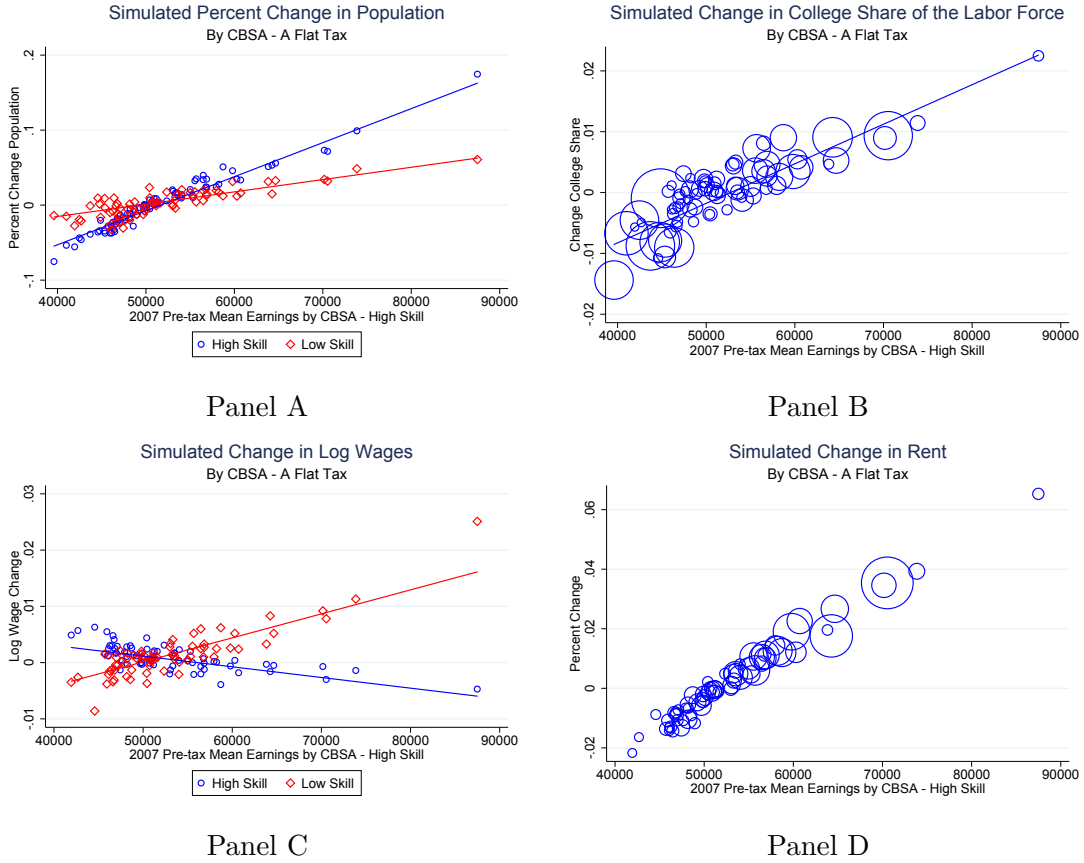


Figure 4: This figure shows the counterfactual changes in population levels (A), college shares (B), log wages (C), and rent (D) as a result of changing from the 2007 tax schedule to a revenue-neutral flat tax. Each dot represents a CBSA. The horizontal axis is the 2007 pretax mean earnings for skilled workers. Unobserved amenities, agent demographics, and labor demand parameters are held fixed at their 2007 levels. Bubbles are proportional to the 1980 CBSA population in panels B and D.

across counterfactuals. The effects of moving from the 2007 tax progressivity to a revenue-neutral flat tax on equilibrium wages, rents, and the distribution of workers are displayed in Figure (4).³³

The flat tax generally has the opposite effect of the progressive tax. Relative to the progressive tax, skilled workers move to productive cities, leading to a 6% increase in the number of skilled workers in the top ten highest-paying cities. This increases the college share in those cities, which leads to a 0.03% wage decrease

³³To be clear, a flat tax is not efficient in this setting; only lump-sum taxes are efficient. However, in our utility function, flat taxes do not affect workers decisions in partial equilibrium. Therefore, flat tax is a natural choice to partially alleviate the distortions caused by taxes.

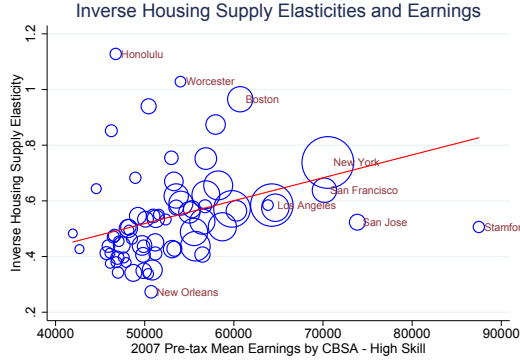


Figure 5: This figure shows the correlation between estimated inverse housing supply elasticities and earnings. Bubbles are proportional to the 1980 CBSA population.

for skilled workers and a 0.7% wage increase for unskilled workers. As a result of the increase in workers and earnings in high-productivity cities, rents in the ten highest-paying cities increase by nearly 3% compared to the progressive tax.

The final row of Table (3) shows the incidence and deadweight loss of a revenue neutral flat tax. The flat tax is more efficient than the progressive tax, with a deadweight loss of 1.3 cents per dollar of tax revenue, compared to 2.1 cents for the progressive tax. However, total worker equivalent variation is nearly identical under both tax codes, at roughly 86% of tax revenue.³⁴ Instead, landowners enjoy the efficiency gains from the new tax code, with their total profits increasing from 14.9% to 16% of tax revenue.

In order to better understand how the flattening of the tax code leads to an increase in landowner profits, Figure (5) shows the correlation between earnings and inverse housing supply elasticities. We can see that the two are positively correlated: cities with higher income levels generally have more inelastic housing supply curves. As shown in panel A of Figure (4), both skilled and unskilled workers move toward high-earning cities, and therefore, toward cities with more inelastic housing supply curves. As shown in Section (2.2) and in the Comparative Statics section in the Appendix, this sorting toward cities with higher earnings and more inelastic housing supply curves leads to an increase in landowner profits. Effectively, this implies that a progressive income tax subsidizes workers to live in

³⁴Recall that tax revenue is held constant across both counterfactuals.

areas with elastic housing supply curves, which lowers landowner profits relative to worker welfare.

Furthermore, the flattening of the tax code leads to an increase in between-group welfare inequality. Equivalent variation for unskilled workers increases from 27% to 37% of tax revenue, or from \$2,600 to \$3,600 per person, per year. Equivalent variation for skilled workers decreases from 60% of tax revenue to 49% of tax revenue, or from nearly \$10,000 to \$8,100 per person, per year. Thus, reducing the progressivity of the income tax leads to a large increase in between-group welfare inequality with no increase in total worker welfare.

7 Conclusion

This paper uses a spatial equilibrium model to measure the deadweight loss and incidence of a progressive federal income tax. Our quantitative model builds upon previous work that evaluates optimal spatial taxes in two key respects. First, we relax the assumption that workers are perfectly mobile, which is crucial for measuring the distortionary effects of the tax; if workers' location choices do not respond strongly to taxes, then the adverse effects of the tax are mitigated. Second, we relax the assumption that workers are completely homogeneous, which effectively removes any redistribution incentive from the optimal tax problem. Put simply, our contribution is to measure the incidence of income tax across heterogeneous workers and landowners and to incorporate the classic equity-efficiency trade-off into the literature studying the spatial distortions of the federal income tax.

We reach two main conclusions. First, we find that the current tax code leads to a moderate deadweight loss, with skilled workers incurring the majority of the incidence. Second, we find that reducing the progressivity of the tax code would result in a large increase in between-group welfare inequality, with skilled workers benefiting and no increase in total worker welfare. Instead, landowner profits increase drastically as workers move toward more cities with more inelastic housing supply curves. Overall, our results suggest that a progressive income tax code can reduce between-group welfare inequality without decreasing total worker welfare. In future work, researchers could calculate the city-specific taxes that minimize deadweight loss while holding inequality constant.

References

- Albouy, David.** 2009. “The unequal geographic burden of federal taxation.” *Journal of Political Economy*, 117(4): 635–667.
- Autor, David H., and David Dorn.** 2013. “The growth of low-skill service jobs and the polarization of the US labor market.” *American Economic Review*, 103(5): 1553–1597.
- Autor, David H., David Dorn, and Gordon H. Hanson.** 2013. “The China syndrome: Local labor market effects of import competition in the United States.” *American Economic Review*, 103(6): 2121–2168.
- Autor, David H., Lawrence F. Katz, and Melissa S. Kearney.** 2008. “Trends in US wage inequality: Revising the revisionists.” *Review of Economics and Statistics*, 90(2): 300–323.
- Bartik, Timothy J.** 1991. *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research.
- Baum-Snow, Nathaniel, and Derek Neal.** 2009. “Mismeasurement of Usual Hours Worked in the Census and ACS.” *Economics Letters*, 102(1): 39–41.
- Baum-Snow, Nathaniel, Matthew Freedman, and Ronni Pavan.** Forthcoming. “Why has urban inequality increased?” *American Economic Journal: Applied Economics*.
- Bayer, Patrick, Nathaniel Keohane, and Christopher Timmins.** 2009. “Migration and hedonic valuation: The case of air quality.” *Journal of Environmental Economics and Management*, 58(1): 1–14.
- Berry, Steven.** 1994. “Estimating Discrete-Choice Models of Product Differentiation.” *RAND Journal of Economics*, 25(2): 242–262.
- Berry, Steven, James Levinsohn, and Ariel Pakes.** 2004. “Differentiated Product Demand Systems from a Combination of Micro and Macro Data: The New Car Market.” *Journal of Political Economy*, 112(1): 68–105.
- Bound, John, and Harry J. Holzer.** 2000. “Demand Shifts, Population Adjustments, and Labor Market Outcomes During the 1980s.” *Journal of Labor Economics*, 18(1): 20–54.
- Busso, Matias, Jesse Gregory, and Patrick Kline.** 2013. “Assessing the Incidence and Efficiency of a Prominent Place Based Policy.” *American Economic Review*, 103(2): 897–947.
- Card, David.** 2009. “Immigration and Inequality.” *American Economic Review: Papers and Proceedings*, 99(2): 1–21.
- Card, David, and Thomas Lemieux.** 2001. “Can Falling Supply Explain the Rising Return to College for Younger Men?” *Quarterly Journal of Economics*, 116(2): 705–746.
- Davis, Morris A., and François Ortalo-Magné.** 2011. “Household expenditures,

- wages, rents.” *Review of Economic Dynamics*, 14(2): 248–261.
- 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.
- Dorn, David.** 2009. “Essays on Inequality, Spatial Interaction, and the Demand for Skills.” Dissertation, University of St. Gallen no. 3613.
- Eeckhout, Jan, and Nezh Guner.** 2017. “Optimal Spatial Taxation: Are Big Cities Too Small?” CEMFI Working Paper 1705.
- Fajgelbaum, Pablo D, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen M. Zidar.** 2015. “State taxes and spatial misallocation.” National Bureau of Economic Research Working Paper 21760.
- Glaeser, Edward L., and Joshua D. Gottlieb.** 2009. “The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States.” *Journal of Economic Literature*, 47(4): 983–1028.
- Gruber, Jon, and Emmanuel Saez.** 2002. “The elasticity of taxable income: Evidence and implications.” *Journal of Public Economics*, 84(1): 1–32.
- Gyourko, Joseph, Albert Saiz, and Anita A. 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.
- Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante.** 2014. “Optimal tax progressivity: An analytical framework.” National Bureau of Economic Research Working Paper 19899.
- Heckman, James J., Lance Lochner, and Christopher Taber.** 1998. “Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents.” *Review of Economic Dynamics*, 1(1): 1–58.
- Herkenhoff, Kyle F., Lee E. Ohanian, and Edward C. Prescott.** 2014. “Tarnishing the Golden and Empire States: Land-Use Restrictions and the U.S. Economic Slowdown.” Working paper.
- Jaeger, David A.** 1997. “Reconciling the old and new Census Bureau education questions: Recommendations for researchers.” *Journal of Business & Economic Statistics*, 15(3): 300–309.
- Katz, Lawrence F., and Kevin M. Murphy.** 1992. “Changes in Relative Wages, 1963–1987: Supply and Demand Factors.” *Quarterly Journal of Economics*, 107(1): 35–78.
- Kline, Patrick, and Enrico Moretti.** 2014. “People, places, and public policy: Some simple welfare economics of local economic development programs.” *Annual Review of Economics*, 6(1): 629–662.
- Krusell, Per, Lee E. Ohanian, José-Víctor Ríos-Rull, and Giovanni L. Violante.** 2000. “Capital-skill complementarity and inequality: A macroeconomic anal-

- ysis.” *Econometrica*, 68(5): 1029–1053.
- Malamud, Ofer, and Abigail K. Wozniak.** 2012. “The Impact of College Education on Geographic Mobility.” *Journal of Human Resources*, 47(4): 913–950.
- Meyer, Peter B., and Anastasiya M. Osborne.** 2005. *Proposed category system for 1960-2000 census occupations*. US Department of Labor, Bureau of Labor Statistics, Office of Productivity and Technology.
- Moody, Scott, and David K. Hoffman.** 2003. *Federal Income Taxes and the Cost of Living: States with Lower Cost of Living Enjoy Greater Federal Tax Advantage*. Tax Foundation.
- Moretti, Enrico.** 2004. “Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data.” *Journal of Econometrics*, 121(1): 175–212.
- Moretti, Enrico.** 2011. “Local labor markets.” In *Handbook of Labor Economics*. Vol. 4, , ed. Orley Ashenfelter, and D. Card, 1237–1313. Amsterdam:Elsevier.
- Moretti, Enrico.** 2013. “Real Wage Inequality.” *American Economic Journal: Applied Economics*, 5(1): 65–103.
- Notowidigdo, Matthew J.** 2013. “The Incidence of Local Labor Demand Shocks.” Unpublished, University of Chicago Booth School of Business.
- Ottaviano, Gianmarco I.P., and Giovanni Peri.** 2012. “Rethinking the effect of immigration on wages.” *Journal of the European Economic Association*, 10(1): 152–197.
- Park, Jin Heum.** 1994. “Estimation of sheepskin effects and returns to schooling using the old and the new CPS measures of educational attainment.” Working Paper 338.
- Piyapromdee, Suphanit.** 2017. “The impact of immigration on wages, internal migration and welfare.” Working paper.
- Roback, Jennifer.** 1982. “Wages, Rents, and the Quality of Life.” *Journal of Political Economy*, 90(6): 1257–1278.
- Rosen, Sherwin.** 1979. “Wage-based Indexes of Urban Quality of Life.” In *Current Issues in Urban Economics*. , ed. Peter Mieszkowski, and Mahlon Straszheim. Baltimore, MD:John Hopkins Univ. Press.
- Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek.** 2010. “Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database].” Minneapolis: University of Minnesota.
- Saiz, Albert.** 2010. “The geographic determinants of housing supply.” *Quarterly Journal of Economics*, 125(3): 1253–1296.
- Wozniak, Abigail K.** 2010. “Are College Graduates More Responsive to Distant Labor Market Opportunities?” *Journal of Human Resources*, 45(4): 944–970.

Appendix

Comparative Statics

To better understand how changes in tax progressivity affect worker and landowner welfare, in this section we perform comparative statics with a simplified version of the model. Throughout the section, we will assume that there is only one demographic group and that pretax income is given exogenously (i.e., $\rho = 1$). As such, we will drop all demographic d and skill e superscripts and all time t subscripts. For simplicity, we will also ignore the birth state premium parameters (i.e., $\beta^d = 0$). Further, again assume that taxes can be parameterized by the function

$$\tilde{I}(I_j) = \lambda I_j^{1-\tau}, \quad (23)$$

where I is pretax income, \tilde{I} is posttax income, τ governs the progressivity of taxes, and λ determines the overall level of taxes.

First we will examine the effects of tax progressivity on worker welfare. Given the Type 1 Extreme Value shocks, total utility across all workers is given by

$$\mathcal{W} = N\mathbb{E} \left[\max_j v_{ij} \right] = N\bar{\gamma} + N \log \left[\sum_{j \in J} \exp \left(\beta_w \log \left(\frac{\tilde{I}(I_j)}{p} \right) + \beta_r \log \left(\frac{r_j}{p} \right) + \xi_j \right) \right],$$

where $\bar{\gamma}$ is Euler's constant and N is the total number of workers.

The total derivative of worker utility with respect to τ is

$$\frac{d\mathcal{W}}{d\tau} = \sum_j N_j \left(-\beta_w \log(I_j) + \beta_r \frac{k_j}{1+k_j} \left(\frac{dN_j}{d\tau} N_j^{-1} - \log(I_j) \right) \right) \quad (24)$$

where we have made use of the relationships

$$N_j = N \frac{\exp \left(\beta_w \log \left(\frac{\tilde{I}(I_j)}{p} \right) + \beta_r \log \left(\frac{r_j}{p} \right) + \xi_j \right)}{\sum_{j'} \exp \left(\beta_w \log \left(\frac{\tilde{I}(I_{j'})}{p} \right) + \beta_r \log \left(\frac{r_{j'}}{p} \right) + \xi_{j'} \right)}$$

and

$$\log(r_j) = \frac{k_j}{1+k_j} \log(N_j \alpha \tilde{I}(I_j)) + \tilde{\zeta}_j.$$

We can rearrange equation (24) to obtain the following expression for the effects of tax progressivity on worker welfare:

$$\frac{dW}{d\tau} = - \underbrace{\beta_w \sum_j N_j \log(I_j)}_{\text{Direct Effect on Income}} - \underbrace{\beta_r \sum_j N_j \frac{k_j}{1+k_j} \log(I_j)}_{\text{Effect of Income on Rents}} + \underbrace{\beta_r \sum_j \frac{k_j}{1+k_j} \frac{dN_j}{d\tau}}_{\text{Effect of Mobility on Rents}}.$$

The first term gives the welfare effect of the change in τ on total weighted worker income. The tax change leads to a change of $-\log(I_j)$ in posttax income in each location. The change in posttax income is weighted by the number of workers in each location and multiplied by β_w to convert to utility terms. As β_w is positive, the direct effect of tax progressivity on posttax income will lead to a decrease in total worker welfare.

The next two terms give the effects of equilibrium changes in rents. The second term gives the effect of the change in posttax income on rents, holding the distribution of workers constant. Posttax income in city j decreases by $\log(I_j)$, which leads to a decrease of log rents of $\frac{k_j}{1+k_j}$. This change in rent is also weighted by the number of workers in each city and multiplied by β_r to convert to utility terms. The term β_r is negative, so this effect is positive: a decrease in income overall leads to a decrease in rents, which increases total worker welfare.

The final term gives the effect of worker mobility on rents. The tax leads to a change of $\frac{dN_j}{d\tau}$ in the population of location j , which leads to a change of $\frac{dN_j}{d\tau} N_j^{-1}$ in the log housing demand, and thus a $\frac{k_j}{1+k_j} \frac{dN_j}{d\tau} N_j^{-1}$ increase in log rents. The sign of this third term will depend on the covariance between worker moves and the elasticity of housing supply across cities. If workers are moving toward cities with more inelastic housing supplies, this third term will be negative, as $\beta_r < 0$.³⁵

³⁵To see this, note that

$$\sum_j \frac{k_j}{1+k_j} \frac{dN_j}{d\tau} = J \times \text{Cov} \left[\frac{k_j}{1+k_j}, \frac{dN_j}{d\tau} \right],$$

where we have used the fact that $\sum_j \frac{dN_j}{d\tau} = 0$.

Finally, the above equation can again be rearranged to obtain the following:

$$J^{-1} \frac{d\mathcal{W}}{d\tau} = -\beta_w \text{Cov} [N_j, \log(I_j)] - \beta_r \text{Cov} \left[\frac{k_j}{1+k_j}, N_j \log(I_j) \right] + \beta_r \text{Cov} \left[\frac{k_j}{1+k_j}, \frac{dN_j}{d\tau} \right] \\ - \beta_w \frac{N}{J} \mathbb{E} [\log(I_j)] - \beta_r \mathbb{E} \left[\frac{k_j}{1+k_j} \right] \mathbb{E} [N_j \log(I_j)].$$

Consider holding the population and income constant in each city j and changing the distribution of inverse housing supply elasticities (i.e., k_j 's) across cities. Therefore, the second line of the equation is held constant. Then, we can make the following two observations. First, the effect of increasing tax progressivity on welfare is increasing in $\text{Cov} \left[\frac{k_j}{1+k_j}, N_j \log(I_j) \right]$. This implies that a positive relationship between inelastic housing supply curves and population and income decreases the worker incidence of a progressive tax. Next, the effect of increasing tax progressivity on worker welfare is decreasing in $\text{Cov} \left[\frac{k_j}{1+k_j}, \frac{dN_j}{d\tau} \right]$. This implies that the effect of an increase in tax progressivity is decreasing in the extent to which workers sort toward cities with more inelastic housing supply curves. Thus, if a tax encourages workers to sort toward cities with more elastic housing supply curves, this will decrease the worker incidence of the progressive income tax.

Now consider the effects of tax progressivity on landowner profits. In this case, total landowner profits are given by

$$\Pi = \alpha \lambda \sum_j \frac{k_j}{1+k_j} N_j (I_j)^{1-\tau}.$$

The total derivative of total landowner profits with respect to tax progressivity, τ , is given by

$$\frac{d\Pi}{d\tau} = \underbrace{\alpha \sum_j \frac{k_j}{1+k_j} \frac{dN_j}{d\tau} \lambda I_j^{1-\tau}}_{\text{Effect of Mobility}} - \underbrace{\alpha \sum_j \frac{k_j}{1+k_j} N_j \lambda I_j^{1-\tau} \log(I_j)}_{\text{Effect of Income}}. \quad (25)$$

The first term is the effect of worker mobility on landowner profits, and the second term shows the effect of the direct change in posttax income on landowner profits. We can rewrite this equation as

$$\begin{aligned} \frac{d\Pi}{d\tau} (\alpha\lambda J)^{-1} = & \text{Cov} \left[\frac{dN_j}{d\tau}, \frac{k_j}{1+k_j} I_j^{1-\tau} \right] - \text{Cov} \left[\frac{k_j}{1+k_j}, N_j I_j^{1-\tau} \log(I_j) \right] \\ & - \mathbb{E} \left(\frac{k_j}{1+k_j} \right) \mathbb{E} (N_j I_j^{1-\tau} \log(I_j)). \end{aligned}$$

Therefore, if tax progressivity leads workers to sort into cities with lower income and more elastic housing supply curves, tax progressivity will have a higher incidence on landowners. Furthermore, if cities with more inelastic housing supplies generally have higher population and higher income, tax progressivity will have a higher landowner incidence.

Uniqueness Proof

This Appendix contains the proof of a unique equilibrium under the assumptions in Section 2.4. We first prove two lemmas which we use in the uniqueness proof.

Lemma 1 (Population Increase \implies Rent Increase). *If $S_j^* \geq S'_j$ and $U_j^* \geq U'_j$, then $r_j^* \geq r'_j$.*

Proof. Combining the housing supply and housing demand equations and rearranging, we obtain

$$r_j^{1+k_j} = z_j \left(S_j \alpha^S \tilde{I}_j^S + U_j \alpha^U \tilde{I}_j^U \right)^{k_j}$$

where α^e is the budget share of housing for workers of skill group e . We know that wages for a given skill group are increasing in labor supply of the other skill group. Therefore, we can prove the lemma by showing that the total wage bill received by all agents for each skill group is increasing in labor supply:

$$\begin{aligned} \frac{\partial S_j \tilde{I}_j^S}{\partial S_j} &> 0 \\ \frac{\partial U_j \tilde{I}_j^U}{\partial U_j} &> 0. \end{aligned}$$

Given the Cobb-Douglas production function and the tax function, we can write the total income received by both skill groups as

$$\begin{aligned}\tilde{I}_j^S S_j &= \lambda \left(H^S \hat{A}_j \theta \left(\frac{U_j}{S_j} \right)^{1-\theta} \right)^{1-\tau} S_j = \lambda \left(H^S \hat{A}_j \theta U_j^{1-\theta} \right)^{1-\tau} S_j^{(\theta-1)(1-\tau)+1} \\ \tilde{I}_j^U U_j &= \lambda \left(H^U \hat{A}_j (1-\theta) \left(\frac{S_j}{U_j} \right)^\theta \right)^{1-\tau} U_j = \lambda \left(H^U \hat{A}_j (1-\theta) S_j^\theta \right)^{1-\tau} U_j^{(-\theta)(1-\tau)+1}.\end{aligned}$$

We know that $\theta \in [0, 1]$ and $\tau \in [0, 1)$. Therefore, both income bills are increasing in labor supply. □

Lemma 2 (Relative Populations as Function of Mean Utility Differences). *Let $\delta_j^e = \log \left(\frac{\tilde{I}(I_e^j)}{p} \right) - \alpha^e \log \left(\frac{r_j}{p} \right) + \xi_j^e$ represent the portion of indirect utility that is common across all agents of skill group e . Then $\frac{S_j}{S_l}$ is an increasing function of $(\delta_j^S - \delta_l^S)$ and $\frac{U_j}{U_l}$ is an increasing function of $(\delta_j^U - \delta_l^U)$.*

The proof follows directly from the Type 1 Extreme Value choice probabilities. This implies that the population of a given city relative to another city can be determined from differences in the portion of utility that is common across workers. This allows us to map changes in wages and rents into changes in relative labor supply.

Lemma 3 (Population Differences Across Cities). *Suppose there exist the labor supply allocations $\mathbf{S}^*, \mathbf{U}^*, \mathbf{S}', \mathbf{U}'$ such that $\sum_{j' \in J} S_j^* = \sum_{j' \in J} S_j'$ and $\sum_{j' \in J} U_j^* = \sum_{j' \in J} U_j'$ and $S_j^* \neq S_j'$ or $U_j^* \neq U_j'$ for some j . Then at least one of the following two conditions must hold:*

1. *There exist j and l such that:*

$$\begin{aligned}S_j^* &\geq S_j', & U_j^* &\leq U_j' \\ S_l^* &\leq S_l', & U_l^* &\geq U_l',\end{aligned}$$

where at least one of the inequalities must be strict.

2. There exist j and l such that:

$$\begin{aligned} S_j^* &> S'_j, & U_j^* &> U'_j \\ S_l^* &< S'_l, & U_l^* &< U'_l. \end{aligned}$$

Proof. We have two equilibria, $(\mathbf{w}^{S^*}, \mathbf{w}^{U^*}, \mathbf{S}^*, \mathbf{U}^*, \mathbf{r}^*)$ and $(\mathbf{w}^{S'}, \mathbf{w}^{U'}, \mathbf{S}', \mathbf{U}', \mathbf{r}')$, such that $S_j^* \neq S'_j$ for some j .

If there are only two cities (i.e., $J = 2$), the proof is trivial. Suppose $J > 2$. Then for some j , one of the following two conditions must hold:

1. $S_j^* > S'_j$ and $U_j^* \leq U'_j$
2. $S_j^* > S'_j$ and $U_j^* > U'_j$

Suppose the first condition holds. As the total number of unskilled and skilled workers is the same across the two equilibria, there must exist a city l such that $S_l^* < S'_l$. If $U_l^* \geq U'_l$, then we are in the first case. If $U_l^* < U'_l$, then there must be a third city m such that $U_m^* > U'_m$. Then we can have either $S_m^* > S'_m$ or $S_m^* \leq S'_m$. If $S_m^* > S'_m$, then cities l and m constitute case 2. If $S_m^* \leq S'_m$, then cities j and m constitute case 1.

Suppose the second condition holds. There must exist a city l such that $S_l^* < S'_l$. If $U_l^* < U'_l$, then we are in the second case. If $U_l^* \geq U'_l$, then there must be a third city m such that $U_m^* < U'_m$. Then we can have either $S_m^* \geq S'_m$ or $S_m^* < S'_m$. If $S_m^* \geq S'_m$, then cities l and m constitute case 1. If $S_m^* < S'_m$, then cities j and m constitute case 2. \square

Proof of Proposition (1) (Uniqueness)

Proof. The proof is by contradiction. Suppose there exist two equilibria: $(\mathbf{w}^{S^*}, \mathbf{w}^{U^*}, \mathbf{S}^*, \mathbf{U}^*, \mathbf{r}^*)$ and $(\mathbf{w}^{S'}, \mathbf{w}^{U'}, \mathbf{S}', \mathbf{U}', \mathbf{r}')$ such that $S_j^* \neq S'_j$ or $U_j^* \neq U'_j$ for some j . Then one of the two cases in Lemma (3) must hold. In what follows, we assume that the first inequality of case 1 is strict. Proving the cases when one of the other inequalities is strict is equivalent. We show that both of the above cases result in a contradiction, thus proving that there can only exist one equilibrium.

CASE 1: $S_j^* > S_j'$, $U_j^* \leq U_j'$, $S_l^* \leq S_l'$ and $U_l^* \geq U_l'$

This implies that the relative labor supply of skilled workers in j relative to l is higher and the relative labor supply of unskilled workers in j relative to l is lower in the first equilibrium (\star) versus the second equilibrium (l):

$$\frac{S_j^*}{U_j^*} > \frac{S_j'}{U_j'}$$

$$\frac{S_l^*}{U_l^*} \leq \frac{S_l'}{U_l'}$$

The Cobb-Douglas production function and assumption that demand for output is perfectly elastic imply that wages are decreasing in relative labor supply. Therefore, we have

$$w_j^{S^*} \leq w_j^{S'}, \quad w_l^{S^*} \geq w_l^{S'}$$

$$w_j^{U^*} \leq w_j^{U'}, \quad w_l^{U^*} \geq w_l^{U'}$$

For simplicity, let $\tilde{I}_j^e = \tilde{I}(I_j^e)$. Posttax income is an increasing function of wages. Therefore, we can write

$$\tilde{I}_j^{S^*} \leq \tilde{I}_j^{S'}, \quad \tilde{I}_l^{S^*} \geq \tilde{I}_l^{S'}$$

$$\tilde{I}_j^{U^*} \leq \tilde{I}_j^{U'}, \quad \tilde{I}_l^{U^*} \geq \tilde{I}_l^{U'}$$

We also have more skilled labor in location j relative to location l and more unskilled labor in location l relative to j in the first equilibrium:

$$\frac{S_j^*}{S_l^*} > \frac{S_j'}{S_l'}$$

$$\frac{U_j^*}{U_l^*} \leq \frac{U_j'}{U_l'}$$

Lemma (2) tells us that the ratio of population between two locations is an increasing function of the difference in the mean utility associated with each location choice. Therefore, we must have the following relationships:

$$\begin{aligned}\delta_j^{S^*} - \delta_l^{S^*} &> \delta_j^{S'} - \delta_l^{S'} \\ \delta_j^{U^*} - \delta_l^{U^*} &\leq \delta_j^{U'} - \delta_l^{U'}.\end{aligned}$$

Plugging in the indirect utility formulas, we obtain

$$\begin{aligned}\left[\log \tilde{I}_j^{S^*} - \log \tilde{I}_l^{S^*} - \left(\log \tilde{I}_j^{S'} - \log \tilde{I}_l^{S'} \right) \right] - \alpha^S (\log r_j^* - \log r_l^* - (\log r_j' - \log r_l')) &> 0 \\ \left[\log \tilde{I}_j^{U^*} - \log \tilde{I}_l^{U^*} - \left(\log \tilde{I}_j^{U'} - \log \tilde{I}_l^{U'} \right) \right] - \alpha^U (\log r_j^* - \log r_l^* - (\log r_j' - \log r_l')) &\leq 0.\end{aligned}$$

The first term in the first inequality is weakly negative, and the first term in the second inequality is weakly positive. Therefore, we can rearrange the equations to obtain

$$\begin{aligned}(\log r_j^* - \log r_l^* - (\log r_j' - \log r_l')) &< 0 \\ (\log r_j^* - \log r_l^* - (\log r_j' - \log r_l')) &\geq 0,\end{aligned}$$

which is a contradiction. The intuition here is that the differences in factor ratios have led to lower skilled wages in j relative to l and lower unskilled wages in l relative to j in the first equilibrium (\star) versus the second equilibrium (\prime). Therefore, in order to induce higher skilled populations in j and higher unskilled populations in l in the first equilibrium, rents must be relatively lower in both locations, which is impossible.

CASE 2: $S_j^* > S_j'$, $U_j^* > U_j'$, $S_l^* < S_l'$ and $U_l^* < U_l'$

Population of both groups is higher in j in the first equilibrium. By Lemma (1), this implies that

$$\begin{aligned}
r_j^* &\geq r'_j \\
r_l^* &\leq r'_l.
\end{aligned}$$

We also have that labor supplies of both groups are higher in location j relative to location l in the first equilibrium:

$$\begin{aligned}
\frac{S_j^*}{S_l^*} &> \frac{S'_j}{S'_l} \\
\frac{U_j^*}{U_l^*} &> \frac{U'_j}{U'_l}.
\end{aligned}$$

By Lemma (2), we must have the following relationships:

$$\begin{aligned}
\delta_j^{S^*} - \delta_l^{S^*} &> \delta_j^{S'} - \delta_l^{S'} \\
\delta_j^{U^*} - \delta_l^{U^*} &> \delta_j^{U'} - \delta_l^{U'}.
\end{aligned}$$

Plugging in the formulas for indirect utility, we have

$$\begin{aligned}
\left[\log \tilde{I}_j^{S^*} - \log \tilde{I}_l^{S^*} - \left(\log \tilde{I}_j^{S'} - \log \tilde{I}_l^{S'} \right) \right] - \alpha^S (\log r_j^* - \log r_l^* - (\log r'_j - \log r'_l)) &> 0 \\
\left[\log \tilde{I}_j^{U^*} - \log \tilde{I}_l^{U^*} - \left(\log \tilde{I}_j^{U'} - \log \tilde{I}_l^{U'} \right) \right] - \alpha^U (\log r_j^* - \log r_l^* - (\log r'_j - \log r'_l)) &> 0.
\end{aligned}$$

From the above conditions on rents, we know that $(\log r_j^* - \log r_l^* - (\log r'_j - \log r'_l)) \geq 0$. We can rearrange the above inequalities to obtain

$$\begin{aligned}
\left[\log \tilde{I}_j^{S^*} - \log \tilde{I}_l^{S^*} - \left(\log \tilde{I}_j^{S'} - \log \tilde{I}_l^{S'} \right) \right] &> 0 \\
\left[\log \tilde{I}_j^{U^*} - \log \tilde{I}_l^{U^*} - \left(\log \tilde{I}_j^{U'} - \log \tilde{I}_l^{U'} \right) \right] &> 0.
\end{aligned}$$

The Cobb-Douglas production function, the tax function, and the assumption that output demand is perfectly elastic imply the following log posttax income functions:

$$\begin{aligned}\log \tilde{I}_j^S &= \log \lambda + (1 - \tau) \left[\log \mathcal{H}^S + \log \hat{A}_j + \log \theta^S + (1 - \theta^S) \log \left(\frac{U_j}{S_j} \right) \right] \\ \log \tilde{I}_j^U &= \log \lambda + (1 - \tau) \left[\log \mathcal{H}^U + \log \hat{A}_j + \log (1 - \theta^S) + \theta^S \log \left(\frac{S_j}{U_j} \right) \right],\end{aligned}$$

where \mathcal{H}^e is the number of hours worked by workers of skill e .

Plugging these expressions for log income into the above inequalities and simplifying yields

$$\begin{aligned}(1 - \theta^S) \left[\log \left(\frac{U_j^*}{S_j^*} \right) - \log \left(\frac{U_l^*}{S_l^*} \right) - \left(\log \left(\frac{U_j'}{S_j'} \right) - \log \left(\frac{U_l'}{S_l'} \right) \right) \right] &> 0 \\ \theta^S \left[\log \left(\frac{S_j^*}{U_j^*} \right) - \log \left(\frac{S_l^*}{U_l^*} \right) - \left(\log \left(\frac{S_j'}{U_j'} \right) - \log \left(\frac{S_l'}{U_l'} \right) \right) \right] &> 0.\end{aligned}$$

After some algebra, this yields

$$\begin{aligned}\left[\log \left(\frac{S_j^*}{U_j^*} \right) - \log \left(\frac{S_l^*}{U_l^*} \right) - \left(\log \left(\frac{S_j'}{U_j'} \right) - \log \left(\frac{S_l'}{U_l'} \right) \right) \right] &> 0 \\ \left[\log \left(\frac{S_j^*}{U_j^*} \right) - \log \left(\frac{S_l^*}{U_l^*} \right) - \left(\log \left(\frac{S_j'}{U_j'} \right) - \log \left(\frac{S_l'}{U_l'} \right) \right) \right] &< 0,\end{aligned}$$

which is a contradiction. The intuition is that, if the population of both groups is higher in j in the first equilibrium (\star), rents must be higher in j in the first equilibrium. Therefore, in order to induce higher populations of both groups of workers to live in j in the first equilibrium, wages in j relative to l must be higher for both groups of workers in the first equilibrium. This is impossible, given the Cobb-Douglas production function.

We have shown that both cases result in a contradiction. Therefore, a second equilibrium cannot exist. \square

Data Appendix: For Online Publication Only

We construct our data using the 5% samples of the 1980, 1990, and 2000 US census. We also use the 3%, three-year aggregated American Community Survey (ACS) for the years 2005-2007.³⁶ The data are downloaded from the US Integrated Public Use Microdata Series (IPUMS) website ([Ruggles et al., 2010](#)). To maintain comparability with the broader wage inequality literature, we follow [Autor, Katz, and Kearney \(2008\)](#) (AKK) as closely as possible in constructing the samples. However, because our analysis is at the local labor market level, there are necessarily some differences. We try to be explicit about these differences throughout the Data Appendix.

Before describing the distinct procedures used to construct each series, we first highlight a few definitions and sample selection rules that are consistent throughout the analysis. Individuals residing in group quarters such as prisons and psychiatric institutions are always dropped. Wages are deflated using the PCE deflator, with 1999 as the baseline.³⁷ All individual-level calculations are weighted by the product of total hours worked, the census sampling weights, and the geographic weights described below. All local labor market level calculations are weighted by the corresponding population in 1980.

Some of the series below use industry and occupation in their construction. Creating a balanced panel of occupations and industries over time is complicated by the fact that the Census Bureau redefines the classification systems for each decennial census. Although [Meyer and Osborne \(2005\)](#) provide a crosswalk between the different census years, there are still instances where some occupations and industries are available in one year but not another. Therefore, we use David Dorn's crosswalks to aggregate occupations and industries into a balanced

³⁶We do not use ACS data after 2007 because hours worked are only reported in intervals.

³⁷We use the PCE in the year preceding the decennial census surveys (i.e., 1979, 1989, and 1999) because the questionnaire asks about income earned in the previous year. The procedure for deflating the ACS data is slightly different. All three years are reported in real terms, where 2007 is the baseline. However, the ACS questionnaire asks about income earned in the previous twelve months rather than the previous calendar year. Therefore, we deflate wages in the ACS using the average value of the PCE in 2006 and 2007 to reflect the change in the question.

panel.³⁸ See [Dorn \(2009\)](#), [Autor and Dorn \(2013\)](#), and [Autor, Dorn, and Hanson \(2013\)](#) for more details.

The same methods for assigning individuals to education levels and constructing potential experience are used throughout. In particular, we create five different education categories: dropout, high school graduate, some college, college graduate and post college. Beginning with the 1990 census, the educational attainment question changed its focus from years of education to degree receipt. We use the method proposed by [Jaeger \(1997\)](#) to make the categories listed above comparable across surveys. In the 1980 sample, individuals with less than twelve years of schooling completed are defined as high school dropouts; those with exactly twelve years as high school graduates; those with some college but less than one year and those with between one and three years of college completed as some college; those with either four or five years of college as college graduates; and those with six or more years of college as post college. In the later samples, individuals whose highest grade completed is grade 11 or less are defined as high school dropouts; those with a high school degree, a GED, or those who completed grade 12 but did not receive a diploma as high school graduates; those with an associate degree or that attended college but did not receive a degree as some college; those with a bachelor's degree as college graduates, and those with a master's degree, professional degree, or doctorate as post college. Broader education definitions, such as high school and college equivalents, are weighted averages of these five education levels, where the weights depend on the particular definition and will be defined when necessary.

Potential experience is defined as age less assigned years of schooling less six. We assign zero years of schooling to observations coded as no schooling, nursery school, preschool or kindergarten in all samples. In 1980, assigned years of schooling simply corresponds to the educational attainment question.³⁹ In later samples, we follow [Park \(1994\)](#) to assign years of schooling to each degree category.⁴⁰ Both the [Jaeger \(1997\)](#) method described in the above paragraph and

³⁸These crosswalks are available for download on Dorn's website.

³⁹Recall that the education attainment question explicitly asked about years of schooling in 1980.

⁴⁰Note that we round up all assignment values in [Park \(1994\)](#) that are non-integers to keep the number of experience cells manageable.

the [Park \(1994\)](#) method described here capitalize on the sampling structure of the Current Population Survey, which implemented the same question change as the census, to create their rules. In particular, they match individuals that were asked the old education question in 1991 and the new education question in 1992.

Local Labor Market Geography: PUMA/County Group to CBSA Crosswalks

After 1990, the Public Use Microdata Area (PUMA) is the smallest geographic unit available in the IPUMS microdata. PUMAs are defined to have between 100,000 and 200,000 residents, are an aggregate of both counties and census tracts, and are contained entirely within states. Defining local labor markets as PUMAs has two shortcomings. First, they are too small; for example, there are upwards of 50 PUMAs in Los Angeles County alone. Second, the PUMA definitions, and their corresponding boundaries, changed drastically between 1990 and 2000, which complicates making comparisons over time. The corresponding concept in 1980 is the county group (CG), which is an aggregation of counties only, whose boundaries are also different from those of the 1990 and 2000 PUMAs.

To overcome these problems, we use core based statistical areas (CBSAs), defined by the Office of Management and Budget (OMB), as our geographic concept of local labor markets. The OMB replaced the old concept of metropolitan statistical areas (MSAs) with CBSAs in 2003. CBSAs include both “micropolitan” and “metropolitan” areas, where the former is based on Census Bureau-defined urban clusters of between 10,000 and 50,000 people and the latter is based on Census Bureau-defined urbanized areas of at least 50,000 people. CBSAs provide us with a more natural concept of a local labor market, and we are able to hold their boundaries fixed over time.⁴¹

The primary challenge with using CBSAs is that their definitions do not line up with the geographic information contained in the census. In particular, the key

⁴¹ Note that the metropolitan area variable, *metarea*, in the IPUMS data is essentially unusable. For reasons of confidentiality, any persons living in a PUMA whose border overlaps with a metropolitan area is counted as not living in that metropolitan area. As noted by IPUMS, these omissions are not necessarily representative. See <https://usa.ipums.org/usa/volii/incompmetareas.shtml> for more details.

complication is that sometimes PUMAs (and CGs) are not completely contained in a particular CBSA. We solve this problem by following a strategy similar to the one used by [Autor and Dorn \(2013\)](#) and [Dorn \(2009\)](#), who define local labor markets as commuting zones (CZs).⁴² In particular, we relate PUMAs (and county groups) to CBSAs by utilizing the county-PUMA overlap files constructed by the Census Bureau.⁴³ Specifically, we construct weights that correspond to the fraction of the overall PUMA population contained in a CBSA. For example, suppose that PUMA A is completely contained in CBSA 1 and PUMA C is completely contained in CBSA 2. Suppose further that PUMA B overlaps with CBSAs 1 and 2, where the fraction of PUMA B's total population contained in CBSA 1 is 50% and the fraction contained in CBSA 2 is 50%. To calculate CBSA-level aggregates using individual-level data, we replicate the observations in PUMA B so that one observation is labeled as CBSA 1 and one is labeled CBSA 2. Calculations are then weighted according to the population overlap.

Wage Series

The sample used to construct the relative wage series includes nonfarm, non-military workers, between the ages of 16 and 64, that were not participating in unpaid family work. Workers with positive business income are dropped. Given concerns about measurement error in wages (see [Baum-Snow and Neal \(2009\)](#)), we drop individuals that worked less than 40 weeks annually and less than 35 hours weekly (i.e., we use only full-time, full-year [FTFY] workers). Respondents with missing or imputed values for education are dropped. Observations with values of zero for wage income, usual hours worked or weeks worked, as well as those with imputed values for any of these variables, are also dropped. Finally, immigrants with missing or imputed birthplaces are dropped.

⁴²Note that we use CBSAs, rather than commuting zones, so we can use the [Gyourko, Saiz, and Summers \(2008\)](#) housing supply elasticity measures.

⁴³The PUMA files can be downloaded at <http://mcdc.missouri.edu/websas/geocorr2k.html>. The county group files are not available in a downloadable form from the Census Bureau, but the information can be found at <https://usa.ipums.org/usa/resources/volii/cg98stat.txt>. Please email Kevin Hutchinson if you would like a copy of the Stata file we built containing this data.

Hourly wages are constructed by dividing total wage income by the product of usual hours worked (per week) and weeks worked (per year). We drop observations where the hourly wage is less than 80% of the nominal minimum wage in that year. Following [Autor and Dorn \(2013\)](#), top coded wage incomes are multiplied by a factor of 1.5 and hourly wages are set not to exceed this value divided by 50 weeks times 35 hours. A few issues related to top coding complicate comparisons over time. First, in 1980, 1990, and 2000, the nominal thresholds for top coding are \$75,000, \$140,000 and \$175,000, respectively. The corresponding values in real terms are \$153,178, \$175,667, and \$175,000, implying a more severe right truncation in 1980. Second, in 1980, all values above \$75,000 are coded as \$75,000. In contrast, values above the threshold are expressed as state medians in 1990 and state means in 2000, again implying a more restrictive right truncation in 1980. Because wages tend to be higher in large cities, relaxing the right censoring disproportionately raises mean wages in large cities, even if the underlying city-level wage distributions are unchanged. Although only a small fraction of the sample is top coded, it is concerning that right censoring might be driving changes in wages.

The problem is even more pronounced in the ACS, where top codes are state specific, equal to the 99.5th percentile of the state income distribution, and values above the top code are equal to the state mean of all observations above the cutoff. We address this issue by imposing a comparable top code on the 1990, 2000, and 2007 data. Specifically, we set the nominal top codes in each year so the real value of the top code is \$153,178 across all three samples.⁴⁴ We then set all values of wage income above the top code to equal the top code.

We follow AKK and create composition-adjusted wages by using the predicted values from a series of log wage regressions. More specifically, we run separate log wage regressions by gender, CBSA, and year on the following covariates:

- Five indicators for race (White, Black, Asian, Native American, or Other);
- Five indicators for marital status (Married, Separated, Divorced, Widowed, or Single);

⁴⁴The respective nominal values for 1980, 1990, 2000, and 2007 are \$75,000, \$122,078, \$153,178, and \$181,138.

- An indicator for veteran status;
- Five education categories (H.S. Dropout, H.S. Graduate, Some College, College Graduate, or Post College);
- A quartic in experience;
- Interactions between the experience quartic and a broader education indicator, called College Plus, that includes College Graduates and Post College;
- An immigrant indicator (Native or Immigrant);
- An interaction between immigrant status and three indicators for English proficiency (Speaks English, Poor English, or None).
- An interaction between immigrant status and three indicators for years in the United States (0-10 years, 11-20 years, or 21+ years);
- A full set of interactions between immigrant status and education categories;
- Time effects in the ACS regressions.⁴⁵

We then use the estimated coefficients to predict log wages by gender-education-experience-CBSA cells in each year.⁴⁶ Again, as in AKK, we use four different experience groups; 5 years, 15 years, 25 years, and 35 years, which yields 40 cells per commuting zone. The key difference between our procedure and the one used by AKK is that we run separate regressions for *each* local labor market. Mean log wages for each CBSA, in each year, are weighted averages of the corresponding cells, where the weights are the share of total hours worked in 1980. This holds the composition of the labor force constant across locations and over time.

⁴⁵Recall that the ACS data are an aggregate of 2005, 2006, and 2007 data.

⁴⁶Predictions are evaluated for white, married, nonveteran natives. ACS predictions are evaluated using the estimated 2007 time effect.

Labor Supply Series

The relative supply series, again following AKK, is constructed by forming two samples: “quantity” and “price.” The quantity sample includes nonfarm, non-military workers, between the ages of 16 and 64, that were not participating in unpaid family work. However, in contrast to the relative wage series, the quantity sample includes all employed workers (i.e., including part-time and self-employed workers). Respondents with missing or imputed values for education are dropped. Observations with values of zero for wage income *and* business income are dropped. Individuals with values of zero for usual hours worked or weeks worked are dropped. Finally, observations with imputed values for any of the preceding variables are also dropped.

The quantity sample divides total hours worked by all employed workers into gender-education-experience cells. In particular, the experience cells are single-year categories of 0-39 years of potential experience. Workers with greater than 39 years of potential experience are included in the 39-year cell. The education cells are the five categories described above. This yields 400 gender-education-experience cells.

The price sample is created using full-time, full-year wage earners (i.e., the same workers used to construct relative wages). More specifically, each cell in the price sample is the mean FTFY real hourly wage for that gender-education-experience combination. Wages in each of the cells, in each year, are normalized by dividing by the wage of male high school graduates with 10 years of potential experience. An efficiency unit is computed for each gender-education-experience combination by averaging price samples across 1980, 1990, 2000, and 2007. The price and quantity samples are then merged to create the final supply measure for each cell, which is the efficiency unit multiplied by the total hours worked in that cell. Aggregated quantities, such as high school and college equivalents, are simply sums of the relevant cells.

Tax Calculations

All tax calculations are performed using TAXSIM, a tax calculator housed at the NBER. We use the Stata interface, which returns federal, state, and payroll tax

liabilities, given a set of 21 inputs, by year. The relevant inputs for our exercise are year, state, marital status, number of dependents, and wage income of the primary taxpayer. We also utilize the itemized deduction input to construct our counterfactual changes in wages. To calculate after-tax incomes, we simply convert incomes back to their nominal values, run TAXSIM, subtract federal, state, and payroll taxes from pretax income, and convert these after-tax incomes back to 2000 dollars.