Income Tax Reform in a Spatial Equilibrium

Mark Colas
Kevin Hutchinson *

University of Wisconsin-Madison †

October 31, 2015

Abstract

Progressive income taxes provide a disincentive for workers to live in high productivity local labor markets, potentially leading to a misallocation of resources across space. Under certain conditions on preferences, the optimal income tax is flat when there is only one type of worker. However, once the model is extended to include high and low-skill workers, we show that there are cases when the low-skill worker can actually be made worse off by moving from a progressive income tax to a flat tax. To quantitatively evaluate the merits of implementing a flat tax, we augment the empirical spatial equilibrium model in Diamond (2015) to incorporate federal income taxes and estimate it using Census data. Counterfactual simulations show that moving from the current tax schedule to a flat tax would increase the welfare of high-skill workers by 4.2%, while decreasing the welfare of low-skill workers by 3.8%. Our results provide a rationale for progressive taxes in assisting low-skill workers, even *without* redistribution.

---

* **JOB MARKET PAPER:** khutchinson@wisc.edu

† 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 Richard Blundell, Leah Boustan, Christian Dustmann, Chao Fu, David Green, John Haltiwanger, Rasmus Lentz, Robert Miller, Eduardo Morales, Parag Pathak, Ronni Pavan, Nicolas Roys, Holger Sieg and Jim Walker. Finally, we extend our thanks to countless Wisconsin graduate students, as well as participants in the Empirical Micro Lunch and the Student Research Group at UW-Madison. All remaining errors are our own.

1 Introduction

Productivity varies greatly across local labor markets in the United States. For 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. To the extent that these productivity differences increase the marginal product of labor, and hence lead to differences in incomes across local labor markets, a progressive income tax can lead to a 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. While this potential misallocation has been studied by Albouy (2009) and Eeckhout and Guner (2014), what distinguishes our work is the addition of heterogeneously skilled workers. This extension allows us to ....

First, we use a simple model to show that including heterogenous workers substantially enriches the comparative statics associated with the distortionary effects of an income tax in a spatial equilibrium.

2 Motivating Example

The goal of this section is gain some insight into how enriching a standard model of spatial equilibrium to include an extra type of worker changes the standard comparative statics associated with an income tax. To do so, we augment the two-city, two-skill models in Moretti (2011, 2013) with an income tax, as in Albouy (2009). We then utilize the machinery in Busso, Gregory and Kline (2013) to make the optimal tax problem studied by Eeckhout and Guner (2014) tractable in our setting. We show that a flat tax is optimal in the one worker case, where optimal is defined as welfare maximizing subject to a revenue neutral budget constraint. Then we show that the flat tax is no longer optimal in the model with two types of workers. Further, moving from a progessive income tax to a flat tax can actually make low-skill workers worse off. As far as we know, the results for the two worker case are novel.

First, consider the case with a single type of worker. Let the indirect utility of individual i associated with choosing location $j \in \{1,2\}$ be given by

$$v_{ij} = \log([1-\tau_j]w_j) - \alpha \log(r_j) + A_j + \epsilon_{ij} \tag{1}$$

where w_j is the wage associated with location j, τ_j is a city specific tax rate, r_j is the rent associated with location j, α is the budget share of housing, A_j is the amenity value of location j that is common to all workers and ϵ_{ij} is the idiosyncratic component of amenities.[1] It will be convenient to re-write (1) as

$$v_{ij} = \delta_j + \epsilon_{ij}$$

where

$$\delta_j = \log([1 - \tau_j]w_j) - \alpha \log(r_j) + A_j.$$

Further assume that

$$\epsilon_{i2} - \epsilon_{i1} \sim U[-\sigma, \sigma]$$

so that the fraction of workers choosing location 1, N_1 is given by

$$N_1 = \frac{1}{2\sigma}\left[\log\left(\frac{[1-\tau_1]w_1}{[1-\tau_2]w_2}\right) - \alpha \log\left(\frac{r_1}{r_2}\right) + (A_1 - A_2) + \sigma\right].$$

The model is then closed by specifying the housing supply curves, $r_j = \kappa_j N_j$, and the market clearing condition for workers, $N_1 + N_2 = 1$. We use these equations to eliminate rents, which are the only endogenous objects on the right-hand side of (?). In particular,

$$\log\left(\frac{r_1}{r_2}\right) = \log\left(\frac{\kappa_1 N_1}{\kappa_2(1 - N_1)}\right) \approx \log\left(\frac{\kappa_1}{\kappa_2}\right) + 4N_1 - 2$$

and we can now solve for N_1 by substituting equation (?) into equation (?).[2] The equilibrium number of workers in location 1 is then

$$N_1 = \frac{1}{2\sigma + 4\alpha}\left[\log\left(\frac{[1-\tau_1]w_1}{[1-\tau_2]w_2}\right) - \alpha \log\left(\frac{\kappa_1}{\kappa_2}\right) + (A_1 - A_2) + \sigma + 2\alpha\right]. \tag{2}$$

We now define welfare to be the average utility of workers, given by

$$W = \mathbf{E}[\max_j \{v_{ij}\}] \tag{3}$$

and optimal tax rates maximize (3) subject to a revenue neutrality constraint given by

[1] This form for the indirect utility occurs when preferences are Cobb-Douglas in a consumption good and housing, as we assume later in our empirical work.

[2] The approximation here is a Taylor Series expansion around one-half.

$$\tau_1 w_1 N_1 + \tau_2 w_2 N_2 = R \tag{4}$$

where R is the exogenously given amount of revenue the government must raise. This is a complicated problem, but it is greatly simplified by starting with a flat tax, i.e. $\tau_1 = \tau_2$, and then verifying that these tax rates maximize equation (3) subject to the constraint in equation (4). Assuming a flat tax, equation (2) becomes

$$N_1 = \frac{1}{2\sigma + 4\alpha}\left[\log\left(\frac{w_1}{w_2}\right) - \alpha \log\left(\frac{\kappa_1}{\kappa_2}\right) + (A_1 - A_2) + \sigma + 2\alpha\right]. \tag{5}$$

so that the equilibrium populations do not depend on the tax rates, i.e. a flat tax does not distort location choices. This is crucial; to reiterate, N_1 does not depend on the tax rate. We now make use of the result in Busso, Gregory and Kline (2013) that

$$\frac{dW}{d\delta_j} = N_j$$

which implies that

$$\frac{dW}{d\tau} = N_1 \frac{d\delta_1}{d\tau} + N_2 \frac{d\delta_2}{d\tau}$$

and using (4) to substitute out τ yields

$$\delta_j = \log\left(\frac{[w_1 N_1 + w_2 N_2 - R]w_j}{w_1 N_1 + w_2 N_2 - R}\right) - \alpha \log(r_j) + A_j,$$

but because the populations are independent of the tax rates in the flat tax case, $\frac{d\delta_j}{d\tau} = 0$ for $j \in \{1, 2\}$, and welfare is maximized using the flat tax. The actual tax rate is then determined by the level of government spending, G.

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 U.S. Census. We also use the 3%, three-year aggregated American Community Survey (ACS) for the years 2005-2007 (Ruggles et al.

2010).[3]

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 Core Based Statistical Areas (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.[4]

We achieve consistency and fulfill our second requirement by mapping the most disaggregated geographic units available in the IPUMS data, County Groups (CGs) in 1980 and Public Use Microdata Areas (PUMAs) in 1990 and after, into CBSAs.[5] This is easy for the cases when the CGs/PUMAs are completely contained within a CBSA. However, there are a number cases where the boundaries cross. We are able to overcome this issue by constructing allocation factors between CGs/PUMAs and CBSAs. In these cases, we do not know which of the CBSAs the individual resides, but we do know the relevant probabilities. We replicate these observations, so that the same individual may appear multiple times in the data and weight these observations by the relevant overlap probabilities.

One alternative to using CBSAs is the commonly used IPUMS variable *metarea*. This variable does not fulfill our second requirement, because many metropolitan areas are only partially identified. More importantly, the unidentified portions change over time. For example, all residents of the Stamford, CT metro area are identified as living there in 2000, while almost half its residents are coded as "not identifiable in a metro area" in 1980 and 1990.[6] It is also worth mentioning that using CGs and PUMAs directly is inappropriate because they are too small to define a local labor market (between 100,000

[3] We 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.

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

[5] This is the same procedure used by Dorn (2009) and Autor and Dorn (2013) to map CGs/PUMAs into Commuting Zones.

[6] See https://usa.ipums.org/usa/volii/incompmetareas.shtml for more details.

and 200,000 residents) and their definitions change substantially between three of the four cross-sections. Finally, we choose to use CBSAs, rather than the Commuting Zone concept recently used by Dorn (2009) and Autor and Dorn (2013), so we can match the IPUMS data with our housing supply elasticity measure, discussed below.

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, choose each location. Therefore, we also face a tradeoff 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 location comprise approximately 60% of the entire U.S. population. Further, we map individuals that do not live in one of these 70 areas into their corresponding Census division, creating nine additional choices. It is worth emphasizing that these “rest of” locations do not include individuals in the 70 largest CBSAs.

3.2 CBSA Level Data

There are four main variables we use at the CBSA level; mean wages, efficiency units of labor, mean 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. Towards that end, we restrict workers to be full-time, full-year [FTFY] wage earners.[7] Focusing on FTFY workers eliminates concerns that wage differences across labor markets,

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

or over time, could be due to differences in labor force attachment.[8] 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.[9] More specifically, we divide the data into education-experience groups and calculate mean wages for each of these groups by year and labor market.[10] Location specific wages are then weighted averages of the cell means, where the weights are fixed across labor markets and over time.

The most important criteria 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, which are used to create the labor supply index. 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 cell hours and the efficiency weight, which we refer to as efficiency units of labor.

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.[11] 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 (WRLURI), proposed by Gyourko, Saiz and Summers (2008).[12] 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.

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

[9] See, for example, Katz and Murphy (1992), Card and Lemieux (2001), or AKK.

[10] Mean wages are the predicted values for each cell from log wage regressions run separately by labor market and year. The regressions control for a number of observables.

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

[12] This dataset 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, due to 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. In particular, workers receive a payoff for choosing to live in a CBSA that is at least partially contained in their birth state.

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.[13] 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 kids, so we make the restriction 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, i.e. by marital status, number of kids, etc.

4 Model

As mentioned in the introduction, we build a model of spatial equilibrium, similar to those recently used by Diamond (2015) and Piyapromdee (2015). Locations vary along three dimensions; wages, rents and amenities. The choice of location is modeled as

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

a static discrete 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 education levels are imperfect substitutes, so that the relative supplies of hetergeneous labor in each location determine local wages. Finally, an upward sloping housing supply curve, which is allowed to differ by city, maps the local population into a local rent level. 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; high and low skill. 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 high skill workers to choose a certain location, this will decrease the relative wage of the skilled workers in that location.

Second, we relax the assumption of perfect mobility. In particular, we allow workers' preferences to include a premium for living in 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.[14]

Third, we allow for heterogeneity in preferences over locations, *conditional* on worker type. 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 impacts of tax reform on the resulting spatial equilibrium. Finally, we incorporate the federal income tax, as in Albouy (2009), which is our major innovation to the model, relative to Diamond (2015).

4.1 Labor Demand

Locations are indexed by j and time is indexed by t. Each labor market uses the following Constant Elasticity of Substitution (CES) production function to produce an identical tradeable good, using skilled labor, S, and unskilled labor, U, as inputs,

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

$$Y_{jt} = A_{jt}[\lambda_{jt}^{U} U_{jt}^{\rho} + \lambda_{jt}^{S} S_{jt}^{\rho}]^{\frac{1}{\rho}}.$$

Notice that the labor efficiency parameters (i.e. the λ_{jt}'s) are allowed to vary across labor markets and over time, while the elasticity of substitution between skilled and unskilled labor, $\sigma \equiv \frac{1}{1-\rho}$, 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. The efficiency parameters can be standardized to sum to one, with the common multiplying factor being absorbed by TFP. Abusing notation and letting A_{jt} also represent this scaled TFP yields

$$Y_{jt} = A_{jt}[(1-\theta_{jt}^{S}) U_{jt}^{\rho} + \theta_{jt}^{S} S_{jt}^{\rho}]^{\frac{1}{\rho}}, \tag{6}$$

where θ_{jt}^{S} is now the relative productivity of skilled labor. Labor markets are perfectly competitive, so that workers are paid their marginal products, which yields the following expressions for wages

$$\log(w_{jt}^{S}) = \log(A_{jt}) + \log(\theta_{jt}) + (1-\rho)\log(\tilde{Y}) + (\rho-1)\log(S_{jt}) \tag{7}$$

$$\log(w_{jt}^{U}) = \log(A_{jt}) + \log(1-\theta_{jt}) + (1-\rho)\log(\tilde{Y}) + (\rho-1)\log(U_{jt}) \tag{8}$$

where $\tilde{Y}_{jt} = \frac{Y_{jt}}{A_{jt}}$. This production function also admits the familiar CES relative labor demand curve (i.e. Katz and Murphy (1992), Autor, Katz and Kearney (2008), among many others)

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

4.2 Housing Supply

Housing costs, r_{jt}, are determined endogenously.[15] In particular, we specify an upward-sloping housing supply curve of the form

$$\log(r_{jt}) = (\nu_1 + \nu_2 \psi_j^{WRI})\log(N_{jt}) + \zeta_{jt} \tag{10}$$

[15] We use the terms housing costs and rents interchangeably throughout the paper to refer to local price levels.

where N_{jt} is the number of workers in labor market j at time t, the elasticity of rents with respect to population is given by $(\nu_1 + \nu_2 \psi_j^{WRI})$ and ζ_{jt} represents the unobserved component of rents.

Gyourko, Saiz and Summers (2008) used the Wharton Regulation Survey to produce municipality level measures of the strictness of land use regulations. We aggregate up their measures to the 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. See Diamond (2015) or Piyapromdee (2015) for a micro-founded model that generates a supply curve like the one used here.

4.3 Labor Supply

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. 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, A_{ijt}. We index the narrow demographic groups by d and the broad skill groups, U and S, by e.[16] Preferences are assumed to be Cobb-Douglas and are written as

$$u_{ijt}^d = (1 - \alpha^e)\log(c_{jt}) + \alpha^e \log(h_{jt}) + A_{ijt}^d \tag{11}$$

where α represents the budget share of housing.[17]

Let I_{jt}^e denote income earned in location j by workers with education e in time t, where income is simply the hourly wage multiplied by two thousand. Further, let $\tau_{jt}^d(I_j^e)$ denote the effective (i.e. average) tax rate, which includes federal income, state income and federal payroll taxes. Note that, in addition to income 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. We can now write the workers' budget constraint as

[16]Recall that the narrow demographic groups are defined by marital status, age and number of kids.

[17]See Davis and Ortalo-Magné (2011) for evidence that the budget share of housing is indeed constant across metropolitan areas.

$$p_t c_{jt} + r_{jt} h_{jt} = [1 - \tau_{jt}^d (I_j^d)] I_j^d,$$

and solving the workers' problem yields the following indirect utility

$$v_{ijt} = \log \left(\frac{[1 - \tau_{jt}^d (I_j^d)] I_j^d}{p_t} \right) - \alpha^e \log \left(\frac{r_{jt}}{p_t} \right) + A_{ijt}^d. \tag{12}$$

We assume that workers inelastically supply 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 (12).[18] Workers simply choose the location that maximizes indirect utility. In our model, as in the standard Rosen-Roback framework, locations are characterized by a combination of (after-tax) wages, rents and amenities.

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

$$A_{ijt} = \gamma_{hp}^d \mathbb{I} \left(j \in Bstate_i \right) + \xi_{jt}^d + \sigma^e \epsilon_{ijt}$$

where $\mathbb{I} \left(j \in Bstate_i \right)$ is an indicator for location j being in worker i's birth state, $\gamma_{hp}^{d,e}$ measures the value of this premium, 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 measures the dispersion in ϵ_{ijt}.

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

$$v_{ijt} = \delta_{jt}^d + \gamma_{hp}^d \mathbb{I} \left(j \in Bstate_i \right) + \sigma^e \epsilon_{ijt} \tag{13}$$

where

$$\delta_{jt}^d = \log \left(\frac{[1 - \tau_{jt}^d (I_j^d)] I_j^d}{p_t} \right) - \alpha^e \log \left(\frac{r_{jt}}{p_t} \right) + \xi_{jt}^d. \tag{14}$$

[18]The 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).

4.4 Equilibrium

Our discussion of the equilibrium properties of the model closely follows Bayer and Timmins (2005). Notice that equations (7), (8), (20) and (12) fully characterize the model. In what follows, we simplify the above notation to facilitate the discussion, but all results go through for the full model. In particular, re-define these four equations as

$$w_j^S = \eta_{0j}^S + \eta_U^S N_j^U + \eta_S^S N_j^S \tag{7'}$$

$$w_j^U = \eta_{0j}^U + \eta_U^U N_j^U + \eta_S^U N_j^S \tag{8'}$$

$$r_j = \gamma(N_j^U + N_j^S) \tag{20'}$$

$$v_{ij}^e = w_j^e - \alpha r_j + \epsilon_{ij}. \tag{12'}$$

Prices, i.e. wages and rents, are endogenously determined as a function of high and low skill populations through (7′), (8′) and (20′), while high and low skill populations are endogenously determined as a function of wages and rents through (12′). Following Bayer and Timmins (2005), we assume that individual i's J-vector of preference shocks, $\bar{\epsilon}_i$, is observed by all other workers. Further, we assume that there is a continuum of both high and low skill workers, ensuring that the preference shocks can be be integrated out, i.e. we can obtain location choice probabilities for each group of workers. This effectively boils down to a Nash equilibrium type of concept where each individual makes their location choice, given all other worker's choices, but in this case each individual worker's choice does not affect N_j^S or N_j^U because they are atomless.

4.4.1 Existence

Let μ^U and μ^S be the measure of high and low skill workers in the overall economy. Then we can write the high and low skill populations of location j as

$$N_j^e = \mu^e P_j^e$$

where P_j^e is the probability a worker of education e chooses location j, which allows us to write equations (7′), (8′) and (20′) in terms of choice probabilities. Further, let

ω denote the parameter vector $\{\eta_U^S, \eta_S^S, \eta_U^U, \eta_S^U, \gamma, \alpha, \mu^U, \mu^S\}$. This allows us to write the location-specific choice probabilities as

$$P_j^e = g_j^e(P_j^U, P_j^S; \omega) \tag{15}$$

given the assumptions made above on $\bar{\epsilon}_i$. Now denote $\mathbf{P}$ as the stacked vector of unskilled and skilled choice probabilities for all J locations. Finally, we can re-write equation (15) as

$$\mathbf{P} = \mathbf{g}(\mathbf{P}; \omega) \tag{16}$$

and proving existence is a straightforward application of Brouwer's fixed-point theorem. See Proposition One in Bayer and Timmins (2005) for details.

4.4.2 Uniqueness

In general, we cannot prove that we have a unique equilibrium, but using arguments from Bayer and Timmins (2005), we can provide some intuition that suggests it is likely. It is instructive to first consider the case where we can prove uniqueness. After that, we describe the complicating factor in our model that prohibits proof; the Q-complementarity of wages, i.e. the positive dependence of one skill group's wage on the quantity of labor supplied by the other skill group.

When is an equilibrium unique? Paraphrasing Proposition Two from Bayer and Timmins (2005), this is true "in the presence of a congestion effect." Here a congestion effect means that increases in the local population can only lower indirect utility (i.e. a housing supply curve). The basic intuition is that an undesirable location will only become less desirable if the population increases; a congestion effect preserves the rank ordering of locations, ruling out the possibility of multiple equilibria.

To see this more clearly, consider the case where wages are downward sloping in the quantity of own-skill population, but do not depend on the population of the other skill group, i.e. $\eta_U^S = \eta_S^U = 0$ and $\eta_S^S < 0, \eta_U^U < 0$. This type of labor demand curve would arise if output is a function of one type of labor and a fixed factor, such as land. Substituting (7'), (8') and (20') into (12')

$$v_{ij}^e = \eta_{0j}^e + (\eta_e^e - \alpha)\mu^e P_j^e - \alpha\mu^{e^-} P_j^{e^-} + \epsilon_{ij}. \tag{17}$$

where e^- denotes the other education group. Notice that the coefficient on both choice probabilities are negative (i.e. there is a congestion effect), so we directly apply the theorem to prove that there is a unique equilibrium. Intuitively, locations are ranked by η^e_{0j}, the intercept of the labor demand curve, and the choice probability for choosing location j is strictly increasing in η^e_{0j}.

The complication arises when wages are allowed to depend on the supply of workers in the other skill group. We assume now that $\eta^S_U > 0, \eta^U_S > 0$, which is consistent with the CES production technology in our full model. In particular, we now get that

$$v_{ij} = \eta^e_0 + (\eta^e_e - \alpha)\mu^e P^e_j + (\eta^e_{e^-} - \alpha)\mu^{e^-} P^{e^-}_j + \epsilon_{ij}. \tag{18}$$

where the term $(\eta^e_{e^-} - \alpha)$ can now be either positive or negative, which from Proposition Three in Bayer and Timmins (2005), implies that we cannot prove there exists a unique equilibrium. The basic intuition is that with a strong enough spillover (i.e. $(\eta^e_{e^-} - \alpha)$ sufficiently greater than zero), a large enough population of one skill group can increase the other group's wage enough to outweigh a low value of η^e_{0j}, changing the rank order of locations and yielding multiple equilibria.

Finally, Bayer and Timmins (2005) also provide some simulation evidence to assess how likely multiple equilibria are in certain models. In particular, two aspects of our model make it unlikely that we will encounter multiple equilibria. The first is the relatively large number of choices in the model. The second is heterogeneity in household preferences, which comes through the home premium in our model. Further, they propose a test for the presence of multiple equilibria, which we can use on any of the equilibria we calculate in this paper.[19]

5 Estimation

5.1 Labor Demand

We estimate the labor demand parameters of our model using the relative labor demand curve defined in (9), which is typical in the literature.[20] First, we parameterize the relative demand for skilled workers as

[19]These tests will be applied in the next revision.

[20]Examples include Acemoglu and Autor (2011), Autor (2014), Autor, Katz and Kearney (2008), Card (2009), Card and Lemieux (2001), Katz and Murphy (1992), Ottaviano and Peri (2012) and Piyapromdee (2015), among many others.

$$\log(\frac{\theta^S_{jt}}{1-\theta^S_{jt}}) = \alpha^t_0 + \alpha^t_1 * logPop80_j + \phi_j + \epsilon_{jt}$$

and taking differences yields the following estimating equation

$$\Delta\log(\frac{w^S_{jt}}{w^U_{jt}}) = \Delta\alpha^t_0 + \Delta\alpha^t_1 * logPop80_j - \frac{1}{\sigma}\Delta\log(\frac{S_{jt}}{U_{jt}}) + \Delta\epsilon_{jt}. \tag{19}$$

The concern in estimating equation (19) is that unobserved changes in skill-biased labor demand ($\Delta\epsilon_{jt}$) induce changes in the quantities of skilled labor (S_{jt}) across labor markets. Therefore, we use Instrumental Variables (IV) to estimate equation (19) using the instruments proposed in Card (2009) and Moretti (2004). In particular, the Card instrument is constructed as

$$c_{jt}(e) = \sum_{p=1}^{P} \nu^{80}_{pj}(e) I_{pt}(e)$$

where $I_{pt}(e)$ is the national inflow rate of immigrants with education level e from country p in time t and $\nu^{80}_{pj}(e)$is the share of overall immigrants from country p of education level e living in labor market j in 1980. The identifying assumption is that historical settlement patterns of immigrants are uncorrelated with current labor market conditions.

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

$$m_{jt}(e) = \sum_{l=1}^{L} \omega^{80}_{lj}(e) H_{lt}(e)$$

where $\omega^{80}_{lj}(e)$ is the share of hours worked by group l in labor market j with education level e and $H_{lt}(e)$ is national hours worked by group l with education level e in time t. To be clear, the denominator of $\omega^{80}_{lj}(e)$ is total hours worked by education group e in labor market j. The relevant predicted hours measures, $\hat{h}_{jt}(e)$, are then used to predict changes in relative quantities.

5.2 Housing Supply

Taking first differences of the equation (20), we obtain our estimating equation for housing supply:

$$\Delta \log(r_{jt}) = (\nu_1 + \nu_2 \psi_j^{WRI}) \Delta \log(N_{jt}) + \Delta \zeta_{jt} \tag{20}$$

As with the wage equations, the concern with estimating equation (20) via least squares is that $\Delta\zeta_{jt}$ will be correlated with $\Delta \log(N_{jt})$ because agents move towards locations with lower changes in rents. Therefore we utilize the Card instrument to instrument for changes in population, $\Delta \log(N_{jt})$. The identifying assumption is that historical immigrant settlement patterns are uncorrelated with current changes in housing supply shifters.

5.3 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 and mean utilities using maximum likelihood and the second step uncovers the wage and rent preference parameters, using IV to deal with the endogeneity of wages and rents.[21]

Equations (13) and (14) 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 subtract the mean utility of location one from all other locations and divide through by σ^e, which yields

[21] To be clear, we do not estimate random coefficients for the wage and rent preferences, as is also typically done when estimating differentiated product demand systems. The goal of the random coefficients is to relax the Independence from Irrelevant Alternatives (IIA) structure imposed by the logit model. We abstract from this complication because the IIA does not hold in our model. In particular, differential choice probabilities for different demographic groups breaks the IIA in terms of aggregate choice probabilities (i.e. the combination of all demographic group choice probabilities). Further, the presence of the home premium breaks the IIA *within* demographic groups.

$$v_{ijt} = \delta^d_{jt} + \beta^d_{hp}\mathbb{I}\left(j \in Bstate_i\right) + \epsilon_{ijt} \tag{13'}$$

and

$$\delta^d_{jt} = \beta^e_w \log\left(\frac{[1-\tau^d_{jt}(I^d_j)]I^d_j}{p_t}\right) - \beta^e_r \log\left(\frac{r_{jt}}{p_t}\right) + \xi^d_{jt} - \delta^d_{1t} \tag{14'}$$

where $\beta^e_w \equiv \frac{1}{\sigma^e}$, $\beta^e_r \equiv \frac{\alpha^e}{\sigma^e}$ and $\beta^d_{hp} \equiv \frac{\gamma^d_{hp}}{\sigma^e}$. Abusing notation, we also have $\delta^d_{jt} \equiv \frac{\delta^d_{jt}}{\sigma^e}$, $v_{ijt} \equiv \frac{v_{ijt}}{\sigma^e}$, and $\xi^d_{jt} \equiv \frac{\xi^d_{jt}}{\sigma^e}$. Our goal is to estimate $\boldsymbol{\delta^d_t}$ and β^d_{hp}, which is done in the first step, along with β^e_w and β^e_r, 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 now estimate $\boldsymbol{\delta^d_t}$ and β^d_{hp} using maximum likelihood. In particular, given our distributional assumption, the choice probabilities have the following closed form solution

$$P^d_{ijt} = \frac{e^{\delta^d_{jt}+\beta^d_{hp}\mathbb{I}(j\in Bstate_i)}}{\sum_{j=1}^J e^{\delta^d_{jt}+\beta^d_{hp}\mathbb{I}(j\in Bstate_i)}} \tag{21}$$

and the corresponding log-likelihood function is

$$\mathcal{L}^d_t(\beta^d_{hp}, \boldsymbol{\delta^d_t}) = \sum_{i=1}^{N^d_t}\sum_{j=1}^{J}\mathbb{I}^i_j \log(P^d_{ijt}) \tag{22}$$

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

Directly estimating (**??**′) is complicated by the fact that, for each demographic group, there are 79 parameters that need to be estimated (the home premium and 78 mean utilities), which is computationally difficult.[22] However, Berry (1994) proves that for any β^d_{hp} there exists a unique vector $\boldsymbol{\delta^d_t}$ such that the choice probabilities implied by the model are equal to those in the data. This allows us to concentrate the likelihood function, so that the maximization is now only over a single parameter. In particular, the choice probabilities are now written as

[22]Recall that utility is normalized, so there are 78 mean utilities even though there are 79 choices.

$$P_{ijt}^d = \frac{e^{\delta_{jt}^d(\beta_{hp}^d)+\beta_{hp}^d \mathbb{I}(j\in Bstate_i)}}{\sum_{j=1}^J e^{\delta_{jt}^d(\beta_{hp}^d)+\beta_{hp}^d \mathbb{I}(j\in Bstate_i)}} \tag{23}$$

and the new log-likelihood function is

$$\mathcal{L}_t^d(\beta_{hp}^d) = \sum_{i=1}^{N_t^d} \sum_{j=1}^{J} \mathbb{I}_j^i \log(P_{ijt}^d) \tag{22'}$$

where computationally we invert the choice probabilities using the contraction mapping in Berry, Levinsohn and Pakes (1995) to obtain the unique $\boldsymbol{\delta_t^d}$ associated with every β_{hp}^d.[23] See the Estimation Appendix for more details on our computational approach. It is worth emphasizing that we estimate $\boldsymbol{\delta_t^d}$ and β_{hp}^d separately for each demographic group, so that the home premiums and the value of unobserved amenities are allowed to be different across groups.

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

$$\hat{\delta}_{jt}^d = \beta_w^e \log\left(\frac{[1-\tau_{jt}^d(I_j^d)]I_j^d}{p_t}\right) - \beta_r^e \log\left(\frac{r_{jt}}{p_t}\right) + \xi_{jt}^d - \delta_{1t}^d \tag{24}$$

using IV to address the endogeneity of wages and rents. Recall that the mean utilities were estimated by normalizing the mean utility of location one to zero. Hence, the δ_{1t}^d's are estimated as demographic-year fixed effects. β_w^e and β_r^e are estimated using separate regressions and are restricted to be the same across different demographic groups within a skill group to help obtain more precise estimates.

[INSERT: Discussion of Instruments]

[23] Recall that the logit model requires the implied choice probabilities from the model equal the observed choice probabilities. Using the Berry inversion to estimate (22') is equivalent to imposing the first order conditions for $\boldsymbol{\delta_t^d}$ from (22). To see this note that

$$\frac{\partial \mathcal{L}_t^d(\beta_{hp}^d, \boldsymbol{\delta_t^d})}{\partial \delta_{jt}^d} = \sum_{\{i:\mathbb{I}_j^i=1\}} (1-P_{ijt}^d) + \sum_{\{i:\mathbb{I}_j^i=0\}} -P_{ijt}^d = 0 \implies \frac{N_{jt}^d}{N_t^d} = \frac{1}{N_t^d}\sum_{i=1}^{N} P_{ijt}^d.$$

6 Results

6.1 Labor Demand

The parameter estimates for labor demand are displayed in the first panel of table 1. The elasticity of substitution between high and low skilled workers is estimated to be 3.71. The average relative factor productivity of high skilled workers increases from .52 in 1980 to .61 in 2007.

6.2 Housing Supply

The inverse housing supply estimates are shown in the second panel of table 1. As expected, CBSAs with greater regulatory restriction have more inelastic housing supplies. The average inverse housing supply elasticities range from .16 to to .59 with a mean of 3.2.

6.3 Labor Supply

The estimates of the worker preference parameters are displayed in the third panel of table 1. We estimate the coefficient on log post tax income, β_w^e, as 13.0 for low skilled agents and 8.5 for high skilled agents. The coefficient on log rents, β_w^r is estimated as -8.67 for low skilled agents and -2.26, implying budget shares on local goods, α, of .65 and .27 for low and high skilled agents, respectively.

The large difference in budget shares on local goods between the two skill groups does not seem to be supported by the data. Using micro data from the 2000 Consumer Expenditure Survey, Diamond (2015) finds housing expenditure shares of .39 for non-college households and .43 for college households. Piyapromdee (2015), using data from the American Community Survey, finds housing expenditure shares of .3 for high skilled natives and .32 for low skilled natives. Moretti (2013), using data from the BLS, estimates an expenditure share on local goods of .61. We are currently trying to understand why there is such a large difference in our estimates for α between the two skill groups.

Diamond and Piyapromdee also estimate indirect utility functions in a static spatial equilibrium model. Both assume agents value pretax wages instead of post tax income. Diamond estimates one specification in which α is estimated directly and another in which α is calibrated to .62. When she estimates α directly, she obtains wage coefficient estimates of 3.3 and 4.9 with α of .9 and .4 for low and high skilled agents, respectively.

With calibrated α, she estimates a coefficient on pretax wages of 4.0 for low skilled agents and 2.1 for high skilled agents. Piyapromdee calibrates α and estimates coefficients on log wages of 1.6 and 5.3 for low skilled and high skilled natives, respectively.

Our estimates of β_w^e are larger than the parameters on pretax wages in Diamond and Piyapromdee for a number of reasons. First, as the tax is progressive, changes in log pretax wages translate to smaller changes in log post tax income. This should lead to larger coefficient estimates. Second, we use different instrument variables to estimate β_w^e; Diamond and Piyapromdee both use Bartik instruments to instrument for wage changes while we use exogenous changes in the federal income tax rate.

Finally, table 2 shows the birth place premium estimates from 1980. The birth place premium is decreasing in education, reflecting the greater propensity of high skilled agents to live away from their birth state. The results from 1990, 2000 and 2007 are qualitatively similar and are available form the others on request.

6.4 Partial and General Wage Elasticities

From equation (21), we can rewrite the probability of agent i choosing location j as:

$$P_{ijt}^d = \frac{e^{\delta_{jt}^d + \beta_{hp}^d \mathbb{I}(j \in Bstate_i)}}{\sum_{j=1}^J e^{\delta_{jt}^d + \beta_{hp}^d \mathbb{I}(j \in Bstate_i)}}$$

where

$$\delta_{jt}^d = \beta_w^e \log\left(\frac{[1-\tau_{jt}^d(I_j^d)]I_j^d}{p_t}\right) - \beta_r^e \log\left(\frac{r_{jt}}{p_t}\right) + \xi_{jt}^d - \delta_{1t}^d$$

Differentiating with respect to income in location j, we obtain:

$$\left(\frac{\partial P_{ij}^d}{\partial I_j^d}\right)\frac{I_j^d}{P_{ij}^d} = \beta_w^e\left(1+P_{ij}^d\right)\left(1-\frac{\partial \tau_{jt}^d}{\partial I_j^d}I_j^d - \tau_{jt}^d\right) \tag{25}$$

As the probability of an agent choosing any individual location and the derivative of the tax function are small, we can approximate:

$$\left(\frac{\partial P_{ij}^d}{\partial I_j^d}\right)\frac{I_j^d}{P_{ij}^d} \approx \beta_w^e\left(1-\tau_{jt}^d\right) \tag{26}$$

Assuming 17% and 20% average tax rates for low and high skilled agents in 1980, this implies a partial labor supply elasticity of 6.8 for high skilled agents and 10.8 for low skilled agents.

However, these partial elasticities do not account for general equilibrium responses of wages and rents and therefore will likely overestimate migration responses observed in the data. To get a better sense of what our parameters estimate imply for mobility location choices, we separately raise the total factor productivity of each city by 1% and calculate the new equilibrium[24]. The distribution of 1980 implied general equilibrium wage elastities across cities are displayed in figure 2. The average wage elasticity across cities is 3.7 for high skilled agents and 3.1 for low skilled agents.

It is difficult to find a direct comparison in the literature for our implied general elasticities. Bound and Holzer (2000) regress changes in metropolitan area population on changes in labor demand, which they measure as total hours worked. They find population elasticities of .41 and .09 for high and low skilled agents, respectively. However our results are not easily comparable, as their independent variable is hours worked while ours is income. Kennan and Walker (2011) estimate a dynamic model of state-to-state migration. To analyze labor supply elasticies, they raise the wage in a given state and analyze the population change in that state compared to the baseline simulation without the wage change. From this exercise they find a labor supply elasticities of about .5 after 10 years. We should expect that measures of state-to-state labor supply elasticities should be considerably lower than our city-to-city elasticities.

6.5 Model Fit

In this section we analyze how well our model can replicate the data. As we estimate a separate unobserved amenity level for each city for each demographic group, we will exactly match the population of each demographic group in each city. Similarly, as we estimate a separate level of total factor productivity and relative factor productivity parameter for each city, we will match the wage levels for high and low skilled agents in each city[25].

[24] If labor supplies are constant, increasing TFP by 1% is equivalent to increasing wages by all agents in the city by 1%.

[25] The wages for each education group will not perfectly match the data for two reasons: 1) We use two different samples for labor demand estimation and for simulation. Most importantly, we include immigrants in the sample used for labor demand estimation. 2) We assume the relative productivity levels of high school dropouts and some college students and of college grads and post college grads are constant across locations. In practice, the wage levels across cities fit very well.

Therefore we follow Piyapromdee (2015) and plot the simulated and observed proportion of the population in each city living outside their birthplace. Figures 1 plot the observed and simulate proportion of agents living out of their birthplaces for each of the four education groups. Each dot represents and city and the size of the dot is proportional to each city's population. Overall, the model seems to fit this facet of the data quite well.

7 Counterfactuals

7.1 The Effects of Federal Income Tax Changes

Next we analyze the effects of changes in the federal income tax code on wages, rent, welfare and location choices.

To facilitate comparison across counterfactuals, we fix worker characteristics and unobserved city amenities at their 1980 values. Workers' utility is measured in log dollar equivalent and is calculated using the formula for expected value of an extreme value type I random variable:

$$\mathbb{E}\left(v_{ijt}^{C}\right) = \bar{\gamma} + \log\left[\sum_{j\in J}\exp\left(\delta_{jt}^{d,C} + \gamma_{hp}^{d}\mathbb{I}\left(j \in Bstate_i\right)\right)\right] \tag{27}$$

where

$$\delta_{jt}^{d,C} = \log\left(\frac{[1-\tau_{jt}^{dC}(I_j^{d,C})]I_j^{d,C}}{p_t}\right) - \alpha^e \log\left(\frac{r_{jt}^{C}}{p_t}\right) + \xi_{jt}^{d,80}. \tag{28}$$

and C indexes counterfactual simulations and $\bar{\gamma}$ is Euler's constant.

Welfare differences across counterfactuals are calculated in differences in average expected utility. Locations choices are calculated as conditional choice probabilities.

For each counterfactual we calculate a "mobility" outcome and a "no mobility" outcome. For the "mobility" outcome, we solve the new equilibrium in which agents reoptimize. For the "no mobility" outcome, we assume locations are fixed at their 1980 values. In the "no mobility" case, welfare changes are equal to the change in log earnings.

7.1.1 Effect of Tax Change

To isolate the effects of the tax change we apply the 2007 tax schedule but keep labor demand parameters fixed at their 1980 levels.

The effects of the tax change in the mobility case are displayed in figures 3 through 6. The tax change increases the take home pay of high skilled workers in high productivity cities. As a result, high skilled workers sort to higher productivity cities, increasing wages for low skilled workers, increasing rents, and decreasing wages for high skilled workers.

The tax change also increases the take home pay of low skilled workers in high productivity cities. However, as low skilled workers spend a higher fraction of their income on local goods, the increase in rents induced by the sorting of high skilled workers to high productivity cities outweighs the increased wages and pushes low skilled workers towards lower productivity cities.

The first row of table 3 compares welfare changes in the mobility and no mobility cases. Compared to the no mobility case, changes in welfare in the mobility case are lower for high skilled and higher for low skilled workers. For high skilled workers, the partial equilibrium effect of more workers sorting to high productivity cities in the mobility case is outweighed by the general equilibrium effect of lower wages and higher rents in high productivity cities. This is not true for low skilled workers.

In the no-mobility case, the tax change leads to a 2.3 percentage point increase in welfare inequality between high and low skilled workers as high skilled workers received a larger average tax cut. However, when allowing for mobility, welfare inequality increases by only 1.8% as low skilled workers are able to obtain higher welfare while high skilled workers are hurt by general equilibrium effects. Therefore, not accounting for mobility would lead us to overestimate the change in welfare inequality by about one forth.

7.1.2 Effect of Changes in Labor Demand

Next, we attempt to isolate the effects of the labor demand change on utility by simulating an equilibrium with the 2007 labor demand parameters while keeping the 1980 tax schedule. The welfare results are displayed in the second row of table 3.

In the no-movement case, the change in labor demand leads to a 21.2% increase in welfare for high skilled workers and a 9.4% decrease in utility for low skilled agents, implying a 30.5 percentage point increase in welfare inequality. However, in the mobility case, high skilled welfare increases by only 16.3% while low skilled utility decreases by 5.6%, implying an increase in income inequality of 21.9%. As in the previous counterfactual, the general equilibrium effects of lower wages and higher rents lead to a decrease in average welfare for high skilled workers, while low skilled workers experience an increase in average utility compared to the no-mobility case.

Comparing the second and third rows, we see that the welfare inequality of the tax change is equal to about 10% of the welfare inequality resulting from the change in labor demand.

7.1.3 Combined Effect of Tax Change and Labor Demand

Finally, we simulate the combined effect of the changes in the tax schedule and labor demand. The welfare results are displayed in the third row of table 3. In the no-mobility case, welfare for high skilled agents increases by 35.3% while welfare for low-skilled agents does not change compared to the baseline case.

Again, when agents are allowed to move, welfare inequality is lower compared to the no-mobility case. Not accounting for mobility would lead us to overestimate the increase in welfare inequality by about one third.

References

[1] Acemoglu, D., and Autor, D. H. Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics 4* (2011), 1043–1171.

[2] Albouy, D. The unequal geographic burden of federal taxation. *Journal of Political Economy 117*, 4 (2009), 635–667.

[3] Autor, D. H. Skills, education, and the rise of earnings inequality among the" other 99 percent". *Science 344*, 6186 (2014), 843–851.

[4] Autor, D. H., and Dorn, D. The growth of low-skill service jobs and the polarization of the us labor market. *The American Economic Review 103*, 5 (2013), 1553–1597.

[5] Autor, D. H., Dorn, D., and Hanson, G. H. The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review 103*, 6 (2013), 2121–2168.

[6] Autor, D. H., Katz, L. F., and Kearney, M. S. Trends in us wage inequality: Revising the revisionists. *The Review of Economics and Statistics 90*, 2 (2008), 300–323.

[7] Bartik, T. J. *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute for Employment Research, 1991.

[8] Baum-Snow, N., and Neal, D. Mismeasurement of usual hours worked in the census and acs. *Economics Letters 102*, 1 (2009), 39–41.

[9] Baum-Snow, N., and Pavan, R. Understanding the city size wage gap. *The Review of economic studies 79*, 1 (2012), 88–127.

[10] Baum-Snow, N., and Pavan, R. Inequality and city size. *Review of Economics and Statistics 95*, 5 (2013), 1535–1548.

[11] Baum-Snow, N., Pavan, R., and Freedman, M. Why has urban inequality increased? 2014.

[12] Bayer, P., Keohane, N., and Timmins, C. Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management 58*, 1 (2009), 1–14.

[13] Berry, S. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics 25*, 2 (1994), 242–262.

[14] Berry, S., Levinsohn, J., and Pakes, A. Automobile prices in market equilibrium. *Econometrica 63*, 4 (1995), 841–890.

[15] Berry, S., Levinsohn, J., and Pakes, A. Differentiated product demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy 112*, 1 (2004), 68–105.

[16] Bound, J., and Holzer, H. J. Demand shifts, population adjustments, and labor market outcomes during the 1980s. *Journal of Labor Economics 18*, 1 (2000), 20–54.

[17] Busso, M., Gregory, J., and Kline, P. Assessing the incidence and efficiency of a prominent place based policy. *American Economic Review 103*, 2 (2013), 897–947.

[18] Card, D. Immigration and inequality. *American Economic Review: Papers and Proceedings 99*, 2 (2009), 1–21.

[19] Card, D., and Lemieux, T. Can falling supply explain the rising return to college for younger men? *Quarterly Journal of Economics 116*, 2 (2001), 705–746.

[20] Diamond, R. The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. Unpublished, Stanford University, 2015.

[21] Dorn, D. Essays on inequality, spatial interaction, and the demand for skills. Dissertation University of St. Gallen no. 3613, Septemeber 2009.

[22] Dube, J.-P., Fox, J. T., and Su, C.-L. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica 80*, 5 (Septemeber 2012), 2231–2267.

[23] Glaeser, E. L., and Maré, D. C. Cities and skills. *Journal of Labor Economics 19*, 2 (2001), 316–342.

[24] Gyourko, J., Saiz, A., and Summers, A. A. A new measure of the local regulatory environment for housing markets; the wharton residential land use regulatory index. *Urban Studies 45*, 3 (2008), 693–729.

[25] Katz, L. F., and Murphy, K. M. Changes in relative wages, 1963-1987: Supply and demand factors. *Quarterly Journal of Economics 107*, 1 (1992), 35–78.

[26] Kennan, J. Spatial variation in higher education financing and the supply of college graduates. Unpublished, University of Wisconsin, 2015.

[27] Kennan, J., and Walker, J. R. The effect of expected income on individual migration decisions. *Econometrica 79*, 1 (2011), 211–251.

[28] Meyer, P. B., and Osborne, A. M. *Proposed category system for 1960-2000 census occupations*. US Department of Labor, Bureau of Labor Statistics, Office of Productivity and Technology, 2005.

[29] Moretti, E. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of econometrics 121*, 1 (2004), 175–212.

[30] Moretti, E. Real wage inequality. *American Economic Journal: Applied Economics 5*, 1 (2013), 65–103.

[31] Ottaviano, G. I., and Peri, G. Rethinking the effect of immigration on wages. *Journal of the European Economic Association 10*, 1 (2012), 152–197.

[32] Park, J. H. *Estimation of sheepskin effects and returns to schooling using the old and the new CPS measures of educational attainment*. No. 338. Industrial Relations Section, Princeton University, 1994.

[33] Piyapromdee, S. The impact of immigration on wages, internal migration and welfare. 2015.

[34] Roback, J. Wages, rents, and the quality of life. *Journal of Political Economy 90*, 6 (1982), 1257–1278.

[35] Rosen, S. Wages-based indexes of urban quality of life. *Current Issues in Urban Economics* (1979).

[36] Ruggles, S., Alexander, J. T., Genadek, K., Goeken, R., Schroeder, M. B., and Sobek, M. Integrated public use microdata series: Version 5.0 [machine-readable database]. Minneapolis: University of Minnesota, 2010.

I. Labor Demand			
σ: Elasticity of Sub.	3.71		
	(1.52)		
II. Housing Supply			
ν_1: Baseline	.31	ν_2: Regulation	.12
	(.07)		(.09)
III. Labor Supply			
	Low Skill	High Skill	
β_w^e: Wage	13.04	8.53	
	(4.47)	(1.29)	
β_r^e: Rent	-8.67	-2.26	
	(1.62)	(0.36)	
α^e: Share Housing	0.65	0.27	
	(0.03)	(0.03)	

Table 1: Parameter Estimates

[37] Saiz, A. The geographic determinants of housing supply. *The Quarterly Journal of Economics 125*, 3 (2010), 1253–1296.

HS Grad							
Young				Old			
Single	Marr. 0 child	Marr. 1 child	Marr. $\geq$ 2 child	Single	Marr. 0 child	Marr. 1 child	Marr. $\geq$ 2 child
3.97	4.15	4.14	4.12	3.80	3.83	3.97	4.05
(0.014)	(0.013)	(0.011)	(0.008)	(0.013)	(0.008)	(0.009)	(0.007)
Some College							
Young				Old			
Single	Marr. 0 child	Marr. 1 child	Marr. $\geq$ 2 child	Singles	Marr. 0 child	Marr. 1 child	Marr. $\geq$ 2 child
3.59	3.80	3.82	3.78	3.32	3.32	3.47	3.62
(0.013)	(0.014)	(0.014)	(0.010)	(0.025)	(0.015)	(0.018)	(0.013)
College Grad							
Young				Old			
Single	Marr. 0 child	Marr. 1 child	Marr. $\geq$ 2 child	Singles	Marr. 0 child	Marr. 1 child	Marr. $\geq$ 2 child
3.16	3.39	3.41	3.34	3.00	2.95	3.08	3.22
(0.014)	(0.014)	(0.015)	(0.010)	(0.030)	(0.017)	(0.019)	(0.015)
Post College							
Young				Old			
Single	Marr. 0 child	Marr. 1 child	Marr. $\geq$ 2 child	Singles	Marr. 0 child	Marr. 1 child	Marr. $\geq$ 2 child
2.60	2.72	2.82	2.80	2.62	2.66	2.74	2.87
(0.022)	(0.023)	(0.023)	(0.014)	(0.044)	(0.026)	(0.029)	(0.024)

Table 2: Birth State Premium 1980

	Movement			No Movement		
	High	Low	Difference	High	Low	Difference
2007 Tax	11.1%	9.3%	1.8	11.4%	9.1%	2.3
2007 LD Params	16.3%	-5.6%	21.9	21.2%	-9.4%	30.5
2007 Tax + LD Params	29.3%	4.1%	25.3	35.3%	0.0%	35.2

Table 3: Utility changes are calculated as the difference in utility measured in log dollar equivalent across counterfactual simulations. Expected utility for each agent is calculated using the formula for expected value of a extreme value type 1 random variable. For all counterfactuals, unobserved amenities and agent demographics are fixed at their 1980 levels.

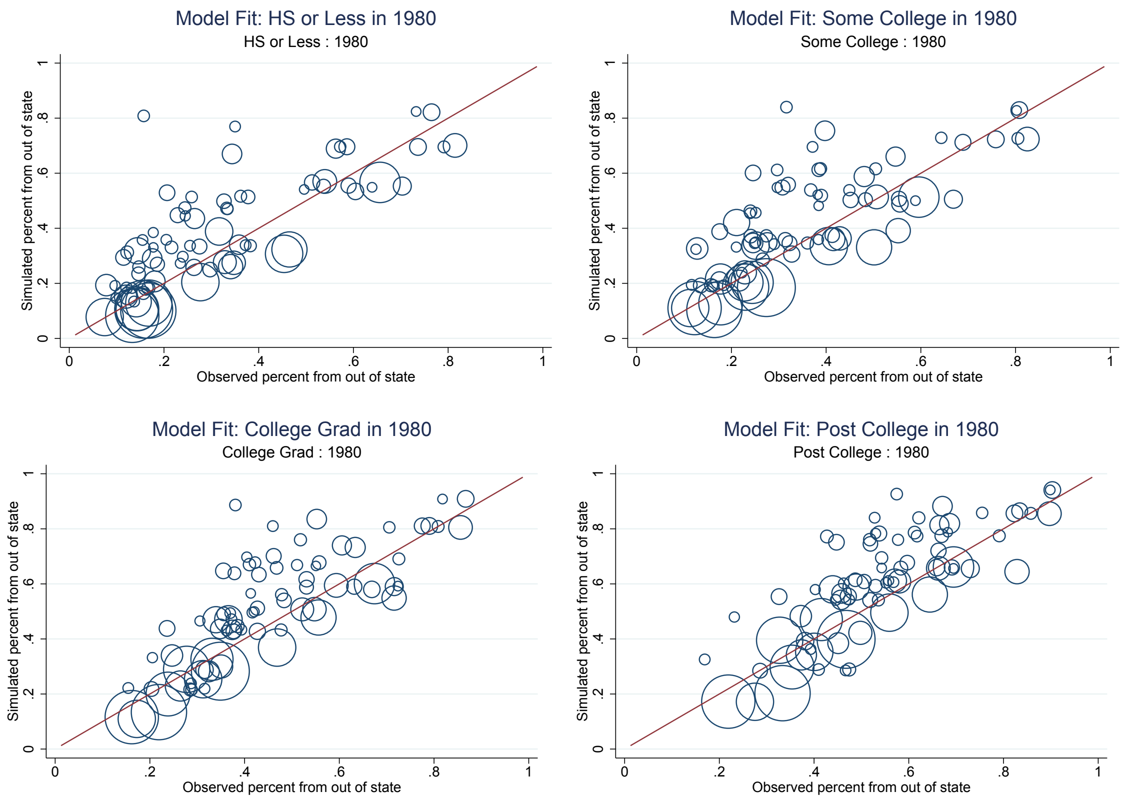

Figure 1: This figure shows the proportion of the population of each city that is born out of state in the data and predicted by the model. Each bubble represents a CBSA, the size is proportional to the city's total population. The horizontal access is the proportion from out of state in the data; the vertical access is the model's prediction.

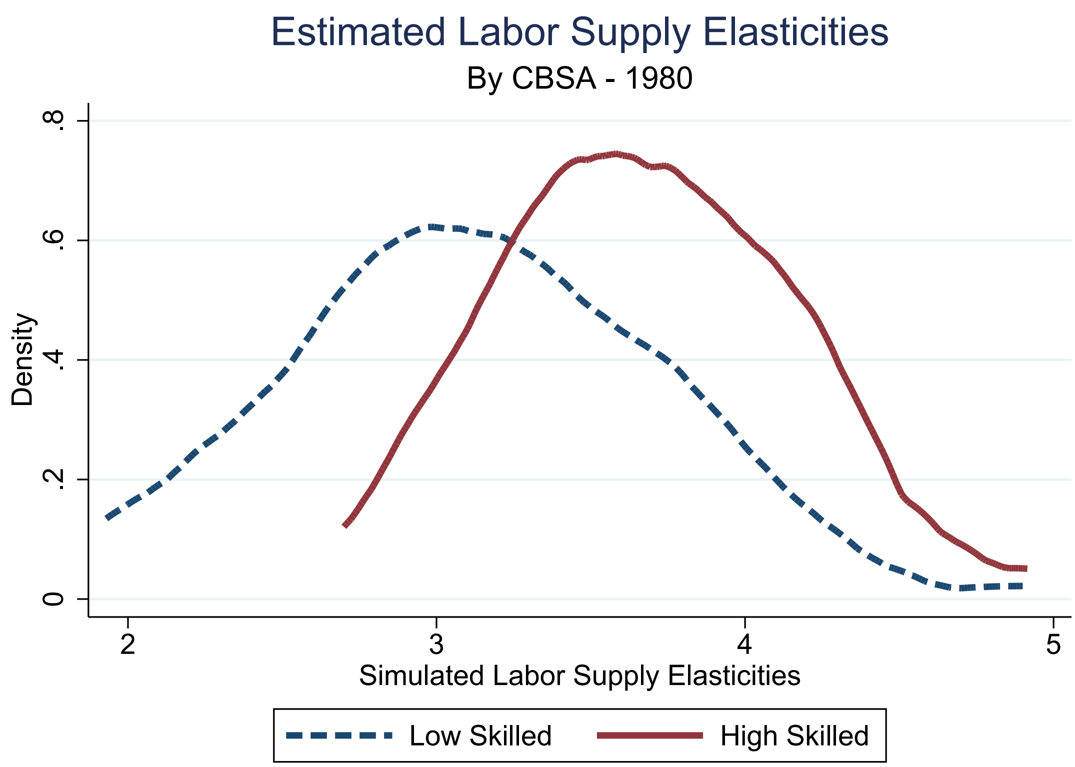

Figure 2: This figure displays simulated labor supply elasticities.

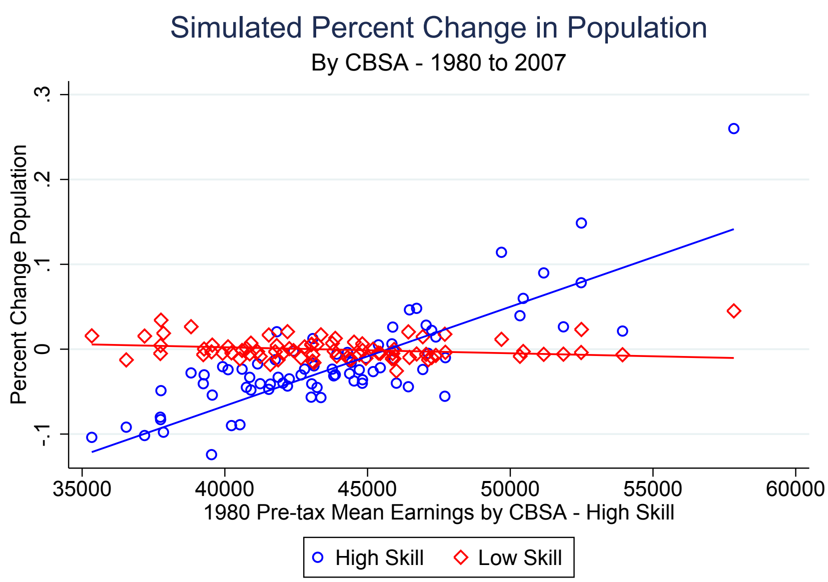

Figure 3: To be written...

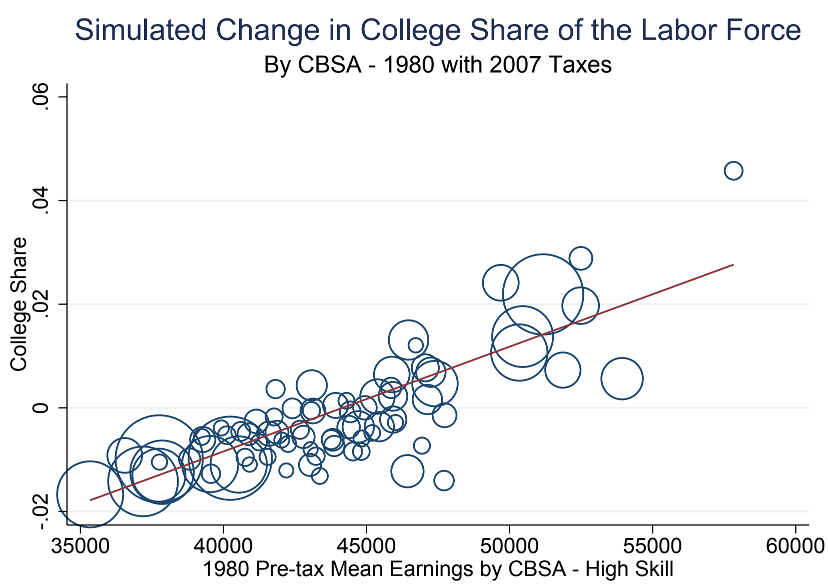

Figure 4: To be written...

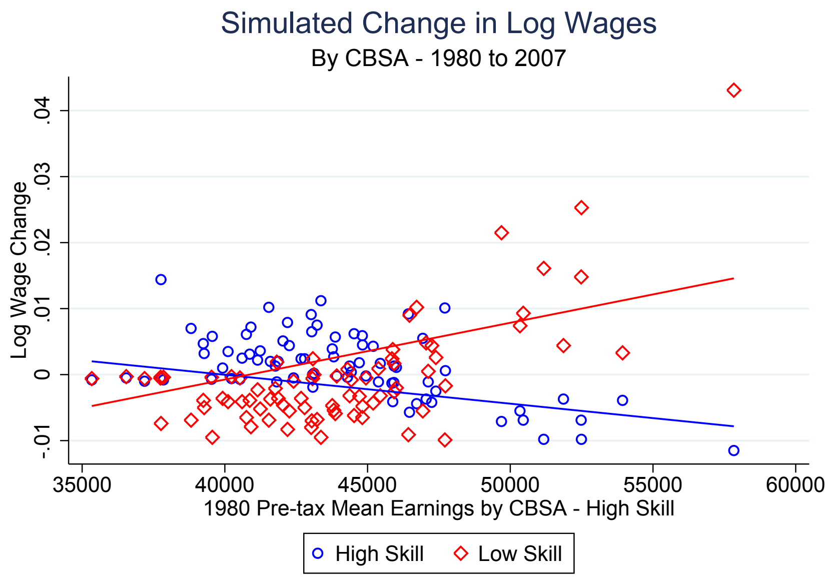

Figure 5: To be written...

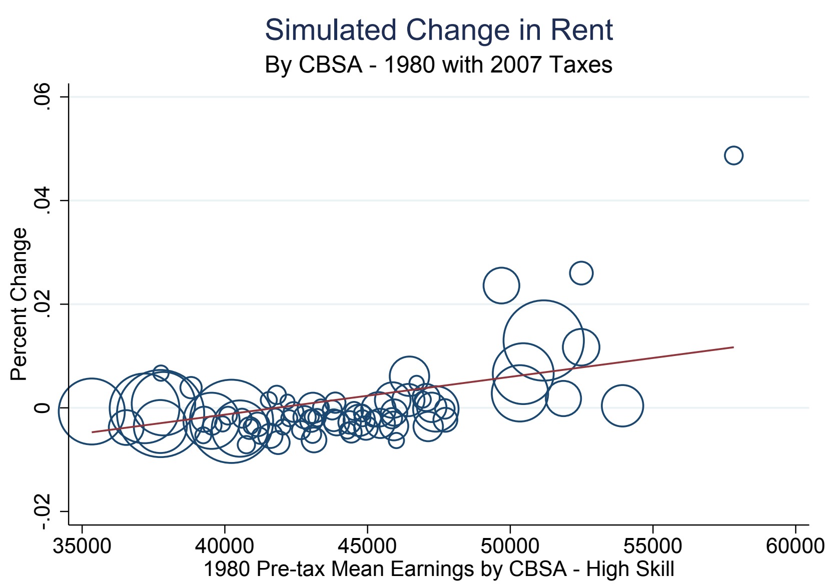

Figure 6: To be written...

Data Appendix

We construct our data using the 5% samples of the 1980, 1990, and 2000 U.S. Census. We also use the 3%, three-year aggregated American Community Survey (ACS) for the years 2005-2007.[26] The data is 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 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.[27] 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 panel.[28] See Dorn (2009), Autor and Dorn (2013), and Autor, Dorn, and Hanson (2013) for more details.

The same methods for assigning individuals to education groups and constructing potential experience are used throughout. In particular, we create five different educa-

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

[27]We 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.

[28]These crosswalks are available for download on Dorn's website.

tion 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 Eleven or less are defined as high school dropouts; those with a high school degree, a GED, or those who completed Grade Twelve, but did not receive a diploma as high school graduates; those with an associate's degree or that attended college, but did not receive a degree as some college; those with a bachelor's 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 groups, 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.[29] In later samples, we follow Park (1994) to assign years of schooling to each degree category. Table A1 displays these assigned years.[30] Both the Jaeger (1997) method described in the above paragraph and 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.

[29]Recall that the education attainment question explicitly asked about years of schooling in 1980.

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

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

After 1990, the Public Use Microdata Area (PUMAs) 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. There are two shortcomings with defining local labor markets as PUMAs. 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 are aggregations 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 with a more natural concept of a local labor market and we are able to hold their boundaries fixed over time.[31]

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 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).[32] In particular, we relate PUMAs (and County Groups) to CBSAs by utilizing the county-PUMA overlap files constructed by the Census Bureau.[33] Specifically, we con-

[31] 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.

[32] Note that we use CBSAs, rather than Commuting Zones, so we can use the Gyourko, Saiz and Summers (2008) housing supply elasticity measures.

[33] The 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, but the information can be found at https://usa.ipums.org/usa/resources/volii/cg98stat.txt. Please email Kevin if you would like a copy of the Stata file we built containing this data.

struct 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. Table A2 illustrates the procedure.

Wage Series

The sample used to construct the relative wage series includes non-farm, 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 birth places are dropped.

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. There are a few issues related to top coding that 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.[34] 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);
- 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;

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

- Time effects in the ACS regressions.[35]

We then use the estimated coefficients to predict log wages by gender-education-experience-CBSA cells in each year.[36] Again, as in AKK, we use four different experience groups; five years, fifteen years, twenty-five years, and thirty-five 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.

Dahl (2002) Selection Correction

To be written...

Labor Supply Series

The relative supply series, again following AKK, is constructed by forming two samples; "quantity" and “price." The quantity sample includes non-farm, 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 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

[35] Recall the ACS data is an aggregate of 2005, 2006, and 2007 data.

[36] Predictions are evaluated for white, married, non-veteran natives. ACS predictions are evaluated using the estimated 2007 time effect.

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 ten years of potential experience. An efficiency unit is computed for each gender-education-experience 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.

Hedonic Rent Index

To be written...

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 pre-tax income and convert these after-tax incomes back to 2000 dollars.

In the Motivation section, we consider a series counterfactual changes in after-tax earnings, by CBSA. We now outline the specifics of the procedure used to construct these measures.

1. Convert 1980 incomes to nominal values.

2. Run TAXSIM to calculate federal, state and payroll tax liabilities. Save these outputs for later.

3. Convert 1980 incomes and state income tax liabilities to 2007 dollars.

4. Set state income taxes to zero and label the 1980 state income tax liability (now in 2007) as an itemized deduction. This holds constant 1980 state income taxes, which

can be deducted from adjusted gross income when calculating federal income tax liabilities.

5. Run TAXSIM to get counterfactual federal income tax liabilities, i.e. taxes using the 2007 tax system and the 1980 wage structure.

6. Convert 2007 counterfactual tax liabilities to 1980 dollars.

7. Construct actual 1980 after-tax earnings and counterfactual 1980 after-tax earnings:

   (a) 1980 after-tax earnings = 1980 earnings - 1980 federal income tax liability - 1980 state income tax liability - 1980 payroll tax liability

   (b) 1980 counterfactual after-tax earnings = 1980 earnings - 1980 counterfactual federal income tax liability - 1980 state income tax liability - 1980 payroll tax liability

Finally, for our counterfactual simulations, we need to estimate an income tax function for each state-year-demographic type, as outlined in the text. We simulate data using TAXSIM to estimate equation (?). In particular, we calculate effective tax rates for incomes between $0 and $100,000, in increments of $100. This yields a sample size of 1,000 observations for each state-year-demographic combination. These functions are in nominal terms, so we convert between nominal and real terms as necessary.

Table A1: Park (1994) Assignment Rules

Education Category	**Assigned Years of Education**
Grades 1, 2, 3, or 4	3
Grades 5, 6, 7, or 8	7
Grade 9	9
Grade 10	10
Grade 11	11
12th Grade - No Diploma, High School Graduate, or GED	12
College Credit - No Degree, or Associate's Degree	14
Bachelor's Degree	16
Master's Degree, Professional Degree, or Doctoral Degree	18

Rules for assigning years of education to degree attained. See Park (1994) for more details.

Table A2: CBSA Aggregation Example

CBSA	PUMA	Wage	Weight	Weighted Mean CBSA Wage
1	A	6	1	
1	B	3	.5	**5**
2	B	3	.5	
2	C	9	1	**7**

Hypothetical CBSA aggregation.

Estimation Appendix

BLP (1995) Contraction Mapping

To be written...

Initial Guesses

To be written...

Solution Algorithm

To be written...

Log-likelihood Analytic Derivatives

To be written...

Inference for Two-Step Estimation

To be written...