Optimal Spatial Taxation: Are Big Cities too Small?*

Jan Eeckhout† and Nezih Guner‡

August 21, 2017

Abstract

We analyze the role of optimal income taxation across different local labor markets. Should labor in large cities be taxed differently than in small cities? We find that a planner who needs to raise a given level of revenue and is constrained by free mobility of labor across cities does not choose equal taxes for cities of different sizes. The optimal tax schedule is location specific and tax differences between large and small cities depends on the level of government spending, the concentration of housing wealth and the strength of agglomeration economies. Our estimates for the US imply higher optimal marginal rates in big cities than in small cities. Under the current Federal Income tax code with progressive taxes, marginal rates are already higher in big cities which have higher wages, but the optimal difference we estimate is lower than what is currently observed. Simulating the US economy under the optimal tax schedule, there are large effects on population mobility: the fraction of population in the 5 largest cities grows by 7.6% with 3.4% of the country-wide population moving to bigger cities. The welfare gains however are smaller. This is due to the fact that much of the output gains are spent on the increased costs of housing construction in bigger cities. Aggregate goods consumption goes up by 1.51% while aggregate housing consumption goes down by 1.70%.

Keywords. Misallocation. Taxation. Population Mobility. City Size. General equilibrium.
JEL. H21, J61, R12, R13.

---

*We are grateful to seminar audiences and numerous colleagues, and in particular to Morris Davis, John Kennan, Kjetil Storesletten, Aleh Tsyvinski, and Tony Venables for detailed discussion and insightful comments. Andrii Parkhomenko provided excellent research assistance. Eeckhout gratefully acknowledges support by the ERC, Grant 339186. Guner gratefully acknowledges support by the ERC, Grant 263600 and the Spanish Ministry of Economy and Competitiveness Grant ECO2014-54401-P.

†University College London, and Barcelona GSE-UPF-ICREA, `jan.eeckhout@ucl.ac.uk`.

‡CEMFI, `nezih.guner@cemfi.es`.

1 Introduction

What is the role of income taxation for the location choice of agents across different cities? Wages and productivity for identical workers are considerably higher in larger cities. This is known as the Urban Wage Premium. Existing progressive income taxation policies tax earnings of equally skilled workers more in larger cities. Workers in larger cities are more productive and earn higher wages, and as a result, they pay a higher average tax rate. In the US, for example, wages for identically skilled workers living in an urban area like New York (about 9 million workers) are 50% higher than wages of those living in smaller urban areas (say Asheville, NC with a workforce around 130,000). As a result of progressive taxation, the average tax rate of a representative worker is almost 5 percentage points higher in NY than it is in Asheville. At the same time, the size of a local labor market is determined by local prices for labor and housing. Higher wages attract more workers while higher housing prices deter them, until in equilibrium they are indifferent across different locations and utility is equalized across cities. In this General Equilibrium context, we analyze the role of federal income taxation and show that optimal taxation of labor income should depend on the location.

Our main finding is that existing taxation regimes lead to the misallocation of resources across space. Taxation of labor incomes across different locations affects location decisions in general equilibrium. Wages and housing prices are determined endogenously in a world where workers optimally choose consumption and housing, and freely locate where to live and work. Our objective is first to compute the equilibrium allocation of the workforce across cities in the presence of the current tax structure in the US, and then derive the tax schedule that will maximize welfare and collect the same tax revenue. When taxes change, citizens respond by relocating, but that in turn affects equilibrium prices. Those equilibrium effects determine both the optimal tax schedule as well as the quantitative implications.

Within this framework, in which the planner is constrained by free mobility of workers, we find that the optimal income tax rates vary across local labor markets. On the one hand, the planner would like to lower taxes in large, productive cities and attract an even larger workforce to these cities to be able collect a higher tax revenue. On the other hand, a larger workforce increases housing prices in large, productive cities and makes them less attractive places to live. We show that this trade-off depends on three key forces in the model economy: the level of government spending, the concentration of housing wealth, and the strength of agglomeration economies. 1. Taxes in big cities relative to those in small cities decrease as government spending increases. Higher government spending increases all locations, but it is more efficient for the planner to generate the revenue by attracting more workers to the big, more productive cities. This is achieved by setting relatively lower taxes in big cities. 2. Taxes in big cities relative to those in small cities increase as the concentration of housing wealth increases. Since concentrated housing wealth does not benefit the population at large, the utilitarian planner does not put weight on it. A larger fraction of the population in big cities increases the value of housing there, which when concentrated in few hands, is not desirable for the planner. The planner therefore sets relatively high taxes in big cities to make them relatively less attractive. 3. Taxes in big cities relative to those in small cities also decrease as the strength of agglomeration economies increases. The planner finds it optimal to lower taxes in larger cities since agglomeration generates an extra benefit of allocating more workers to productive cities.

Quantifying these forces for the US economy we find that taxes in big cities should be higher than those in smaller cities. Due to existing progressive federal income tax schedules in the US taxes in big cities are already higher than the ones in smaller cities. While there might be many reasons why a progressive tax schedule might be desirable, our spatial general equilibrium model with homogenous workers imply that the optimal tax difference between big and small cities should be lower than what we observe in the US tax code. In the quantitative analysis, we parametrize the relation between after and before-tax wages, $\tilde{w}$ and w, as $\tilde{w} = \lambda w^{1-\tau}$, where $1 - \lambda$ is the level of taxation and τ determines the progressivity. Average tax rate is given by $1-\lambda w^{-\tau}$. Taxes are progressive (regressive) when $\tau > 0$ ($\tau < 0$). For the benchmark economy, $\lambda = 0.85$ (i.e. the average tax rate at $w = 1$ is 15%) and $\tau = 0.12$. We find that the optimal value of τ is quite smaller, $\tau^* = 0.0139$.

For US data, the impact of the optimal tax policy are far reaching. Implementing the optimal tax schedule implies that after tax wages increase in large cities. As a result, there is a first order stochastic dominance shift in the city size distribution. When we move from the current to optimal taxes, the population in the five largest cities grows by 7.6%. About 3.4% of the workforce move from smaller to bigger cities countrywide. The aggregate output increases by 1.42%. The gains in terms of utility are, however, much smaller. The experiment that results in an 1.42% increase in GDP only leads to a 0.07% increase in Utilitarian welfare. The small utility gain is due to the fact that most of the output gain in the more productive cities is eaten up by higher housing prices, which go up by 5.2% on average. As a result, while aggregate consumption goes up by 1.51%, aggregate housing consumption declines by 1.70%. Those moving to the big cities take advantage of the higher after tax incomes, but they end up paying higher housing prices. It is precisely the role of housing prices that implies that the optimal tax schedule has higher taxes in big cities.

The model that we use to quantify the optimal spatial taxation has many features to capture the trade-offs faced by a Ramsey planner. First, the production of housing is endogenous to account for the fact that the value share of land is much higher in big cities than in small cities.[1] And it takes into account that the amount of land available for construction differs across locations. Some coastal cities are constrained by the mountains and the sea, whereas others in the interior have unconstrained capacity for expansion. Second, the model allows for congestion externalities that are increasing in city size. Third, housing is modeled in such a way that the rental price of land is retained in the economy as a transfer, while the construction cost eats up consumption goods. Fourth, we allow for amenities across different locations as the residual of the utility differences. Finally, while government expenditure is distortionary, a share of tax revenues is redistributed to the citizens. Even if we do not explicitly

[1] See Davis and Palumbo (2008), Davis and Heathcote (2007), and Albouy and Ehrlich (2012).

model expenditure on public goods, this accounts for the fact that tax revenues also generate benefits.[2]

We investigate in detail which features matter qualitatively and quantitatively for optimal taxation. We find that the optimal tax schedule is sensitive to the size of government, the concentration of housing ownership and the strength of agglomeration economies. 1. Increasing the size of government, i.e., the aggregate tax rate, from 15% (its benchmark value) to 18.5% implies that the optimal tax schedule is regressive (i.e. taxes are smaller in larger cities). This is because it is more efficient to generate revenue from workers in high productivity – and hence large – cities, which can only be achieved by attracting more workers through a lower progressiveness. 2. In our benchmark economy, the share of homeowners is calibrated to the US economy. If we vary homeownership to 100% – all households own a home – the optimal tax schedule becomes strongly regressive, while it becomes strongly progressive when all homes are owned by absentee landlords. A regressive tax schedule, which attracts workers to larger cities and increases the housing prices there, is less attractive for the planner if the benefits of higher housing prices are enjoyed by absentee landlords. 3. We also introduce agglomeration economies where city Total Factor Productivity is determined endogenously through size of the workforce. This results in a regressive tax schedule.[3]

Instead, we find that the key finding, the fact that the optimal tax schedule is progressive but less than the current US income tax schedule, is not sensitive to the following: whether or not we control for worker characteristics (such as education, gender, and race) in the wage calculations for cities, whether land areas are assumed identical for all cities or whether they are the actual area, whether or not there is housing production, and whether or not taxes are used for transfers and rebated to households.

This paper is related to the work on urban accounting by Desmet and Rossi-Hansberg (2013) who analyze the effects on output from the relocation of productive resources.[4] Instead of analyzing the effect of technological change, we take the technology as exogenous and ask what the role is of the change in an institution, in this case federal income taxation. It is also related to literature that studies inter-state migration in the US using dynamic spatial equilibrium models, e.g. Coen-Pirani (2010), Karahan and Rhee (2014), Kaplan and Schulhofer-Wohl (2017), and Davis, Fisher, and Veracierto (2013). While we study a model with homogenous workers, our paper is also related to literature that study geographical allocation of workers with different skills, e.g. Diamond (2016). Our results on reallocation of labor across cities echoes Klein and Ventura (2009) and Kennan (2013) who study free mobility of workers across countries find larger output gains. In the light of the misallocation debate in macroeconomics on aggregate output differences due to the misallocation of inputs, most notably capital, e.g. Guner,

[2]We exclusively focus on the spatial distortion at the collection side. There could also be a distortion at the benefit side, for example where big cities are more or less generous in federal benefits for the unemployed and the disabled (see for example Glaeser (1998) and Notowidigdo (2011)). In our model, we abstract from this important channel altogether and focus on the role of active, full time workers.

[3]When we set $\lambda = 0.815$ (instead of $\lambda = 0.85$ for the benchmark economy), $\tau^* = -0.0245$ (instead of $\tau^* = 0.0139$ for the benchmark economy). With 100% (0%) home ownership, we find $\tau^* = -0.1378$ ($\tau^* = 0.0776$). Finally, when we introduce agglomeration externalities, $\tau^* = -0.0517$.

[4]See also Topa, Sahin, and Violante (2014) for the role of unemployment frictions on spatial mismatch.

Ventura, and Yi (2008), Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), we add a different insight. Due to existing income taxation schemes, also labor is substantially misallocated across cities within countries. Hsieh and Moretti (2015), Herkenhoff, Ohanian, and Prescott (2017), and Parkhomenko (2016) also study spatial misallocation of labor across cities. They focus, however, on misallocation caused by restrictive housing policies. More closely related to our paper, Fajgelbaum, Morales, Suárez-Serrato, and Zidar (2016) study state taxes as a potential source of spatial misallocation in the United States, and find that tax dispersion across states leads to aggregate losses.

The idea that taxation affects the equilibrium allocation is of course not new. Tiebout (1956) analyzes the impact of tax competition by local authorities on the optimal allocation of citizens across communities. Wildasin (1980), Helpman and Pines (1980) and Hochman and Pines (1993), among others, explicitly consider federal taxation and argue that it creates distortions. A common result in this literature is that a tax on the immobile factor, land, is necessary to achieve the efficient allocation. This literature, however, often studies highly stylized models that are not amenable to quantitative work. In the legal literature, Kaplow (1995) and Knoll and Griffith (2003) argue for the indexation of taxes to local wages. Albouy (2009) analyzes the question quantitatively. Starting from the Rosen-Roback tradeoff between equalizing differences across locations, he calibrate the model and conclude that any tax other than a lump sum tax is distortionary. He does not, however, attempt to characterize the optimal tax structure. Albouy, Behrens, Robert-Nicoud, and Seegert (2016) study optimal city size in a model with both the city size and the number of cities are allowed to vary and reach a similar conclusion to ours, i.e. big cities might be too small.

What sets our work apart from the existing literature is a comprehensive quantitative framework that fully takes into account the *general equilibrium* effects, the *endogeneity* of housing prices and consumption, which in turn allows us to focus on the *optimality* of taxation. Furthermore, our results highlight two novel forces, the level of government spending and on the concentration of housing wealth, that affect critically the optimal tax structure. The existing literature often assumes that the housing wealth is equally distributed among households, and there has not been any previous attempt to link the optimal tax structure to the size of the government.

2 The Model

Population. The economy is populated by a continuum of identical workers. The country-wide measure of workers is $\mathcal{L}$. There are J locations (cities), $j \in \mathcal{J} = \{1, ..., J\}$. The amount of land in a city is fixed and denoted by T_j. The total workforce in city j denoted by l_j. The country-wide labor force is given by $\mathcal{L} = \sum_j l_j$.

Preferences, Amenities and Congestion. All citizens have Cobb-Douglas preferences over consumption c, and the amount of housing h, with a housing expenditure share $\alpha \in [0, 1]$. This choice is motivated by Davis and Ortalo-Magné (2011), who find that expenditure shares on housing are constant across U.S. metropolitan statistical areas. The consumption good is a tradable numeraire good with price normalized to one. The price for one unit of land is p_j. The real estate market is perfectly competitive so that the flow payment equals the rental price. Workers are perfectly mobile and can relocate instantaneously and at no cost.[5] Thus, in equilibrium, identical workers obtain the same utility level wherever they choose to locate. Therefore for any two cities j, j' it must be the case that the respective consumption bundles for an individual worker satisfy $u(c_j, h_j) = u(c_{j'}, h_{j'})$.

Cities inherently differ in their attractiveness that is not captured in productivity, but rather is valued directly by its citizens. This can be due to geographical features such as bodies of water (rivers, lakes and seas), mountains and temperature, but also due to man-made features such as cultural attractions (opera house, sports teams, etc.). We denote the city-specific amenity by a_j, which is known to the citizens but unobserved to the econometrician. We will interpret the amenities as unobserved heterogeneity that will account for the non-systematic variation between the observed outcomes and the model predictions.[6] It is crucial that for the purpose of the correct identification of the technology, this error term is orthogonal to city size. Albouy (2008) provides evidence that the bundle of observed amenities – both positive and negative – are indeed uncorrelated with city size.

In addition to city-specific amenities, to capture the cost of commuting, we allow for a congestion externality. Unlike the amenity, which is city-specific and fixed, the congestion systematically depends on the city size and is given by l_j^δ, where $\delta < 0$ (as in Eeckhout (2004)).

The utility in city j from consuming the bundle (c, h) is therefore written as:

$$u(c, h) = a_j l_j^\delta c^{1-\alpha} h^\alpha.$$

Technology. Cities differ in their total factor productivity (TFP) which is denoted by A. TFP is exogenously given. In each city, there is a technology operated by a representative firm that has access to a city-specific TFP A_j, given by

$$F(l_j) = A_j l_j. \tag{1}$$

Firms pay wages w_j for workers in city j. Wages depend on the city j because citizens freely locate between cities not based on the highest wage, but, given housing price differences, based on the highest utility. Firms are owned by absentee capitalists.

Housing Supply. The supply of housing in each city j is denoted by H_j. Housing stock is produced by means of capital K_j and the exogenously given land area T_j according to the following CES production

[5] The model could be extended to allow for mobility costs and location-specific preference heterogeneity, as in Fajgelbaum, Morales, Suárez-Serrato, and Zidar (2016).

[6] In Diamond (2016) amenities depend endogenously on the characteristics, such as income, of individuals living in a city.

technology:

$$H_j = B\left[(1-\beta)K_j^\rho + \beta T_j^\rho\right]^{1/\rho}, \tag{2}$$

where $\beta \in [0,1]$ indicates the relative importance of capital and land in housing production, and B indicates the total factor productivity of the construction sector. The elasticity of substitution between K and T is given by $\frac{1}{1-\rho}$. We assume that housing capital is paid for with consumption goods, and hence the marginal rate of substitution between consumption and housing is equal to one and the rental price of capital is equal to the numeraire. The rental price of land is denoted by r_j. Given this constant returns technology, we assume a continuum of competitive construction firms with free entry. A special case where $\beta = B = 1$ is where housing is exogenous and $H_j = T_j$ and $r_j = p_j$.

While the housing capital to build structures is foregone consumption, the land rents are transfers and stay in the economy. We assume that a fraction ψ of land is owned by measure zero landlords and a fraction $1-\psi$ is owned in equal shares by each worker in the economy in the form of a bond that is a diversified portfolio of the country's land supply. As a result, there is a transfer R_j to each agent living in city j:

$$R_j = (1-\psi)\frac{\sum_j r_j T_j}{\sum_j l_j}. \tag{3}$$

The parameter ψ captures the fact that housing ownership is not perfectly diversified.[7] As we will see below, the details of the ownership structure are important for the results.

Market Clearing. The country-wide market for labor clears, $\sum_{j=1}^{J} l_j = \mathcal{L}$, and for housing, there is market clearing within each city $h_j l_j = H_j$, $\forall j$. Under this market clearing specification, only those who work have housing. We interpret the inactive as dependents who live with those who have jobs.

Taxation. The federal government imposes an economy-wide taxation schedule. Its objective is to raise an exogenously given level of revenue G to finance government expenditure. Denote the pre-tax income by w and the post-tax income by $\tilde{w}$. Denote by t_j the specific tax rate that applies to workers in city j. Then $\tilde{w}_j = (1-t_j)w_j$. Often tax schedules are substantially simpler. For example, federal taxes typically do not depend on the location j and there is a systematic degree of progressivity.[8] To that purpose, we assume that the progressive tax schedule can be represented by a two-parameter family

[7] Of course, the ownership structure that equation (3) represents is a shortcut that bypasses the complications that stem from ex post heterogeneity of asset holdings. Ideally we would like to explicitly model the ownership and trade of housing assets in conjunction with the migration decisions. Unfortunately, that portfolio allocation problem is intractable as it leads to high dimensional ex post heterogeneity.

[8] Of course, tax breaks from mortgage interest deductions as in the United States are likely to be higher in big cities since households earn on average higher wages and spend the same share of their income on housing. But there is evidence that such favorable tax treatment does not affect the home ownership rate in comparison with other countries. Ownership rates are similar in Australia, Canada, and the United Kingdom, where there is no such tax deduction for mortgage interest. In fact, the UK gradually abolished mortgage interest deduction between 1975 and 2000, a period in which home ownership rose from 53% to 68%. For a formal treatment, see Glaeser and Shapiro (2003).

that relates after-tax income $\tilde{w}$ to pre-tax income w as:

$$\tilde{w}_j = \lambda w_j^{1-\tau},$$

where λ is the level of taxation and τ indicates the progressivity ($\tau > 0$). This is the tax schedule proposed by Bénabou (2002). Recent papers, e.g. Heathcote, Storesletten, and Violante (2017), Guner, Lopez-Daneri, and Ventura (2016) and Kindermann and Krueger (2014), use the same function to study optimal progressivity of income taxation in the U.S. The average tax rate is given by $1 - \lambda w_j^{-\tau}$ and the marginal tax rate is $1 - \lambda(1-\tau)w_j^{-\tau}$. Taxes are proportional when $\tau = 0$, in which case the average rate is equal to the marginal rate and equal to λ. Under progressive taxes, $\tau > 0$ and the marginal rate exceeds the average rate.

A share of tax revenue is used for transfers. Of the total tax revenue, an amount ϕG is transferred to the households. While there may well be city-specific differences in those federal transfers, we take the agnostic view here that the transfer is lump sum across all agents. Therefore each household receives the transfer $TR = \frac{\phi G}{\mathcal{L}}$.

Equilibrium. We are interested in a competitive equilibrium where workers and firms take wages w_j, housing prices p_j and the rental price of land r_j as given. The price of consumption is normalized to one. Because housing capital is perfectly substitutable with consumption, the rental price of housing capital is also equal to one. All prices satisfy market clearing. Workers optimally choose consumption and housing as well as their location j to satisfy utility equalization. Firms in production and construction maximize profits, which are driven to zero from free entry.

3 The Equilibrium Allocations

Given prices and subject to after tax income, a representative worker in city j solves

$$\max_{\{c_j, h_j\}} u(c_j, h_j) = a_j l_j^{\delta} c_j^{1-\alpha} h_j^{\alpha} \tag{4}$$

subject to

$$c_j + p_j h_j \leq \tilde{w}_j + R_j + TR,$$

for all j. Taking first order conditions, the equilibrium allocations are $c_j = (1-\alpha)(\tilde{w}_j + R_j + TR)$ and $h_j = \alpha \frac{(\tilde{w}_j + R_j + TR)}{p_j}$.[9] The indirect utility for a worker is

$$u_j = a_j l_j^{\delta} \alpha^{\alpha} (1-\alpha)^{1-\alpha} \frac{(\tilde{w}_j + R_j + TR)}{p_j^{\alpha}}. \tag{5}$$

[9]The construction firms buy capital from households. Since the price of capital is one, however, this transaction does not affect the household budget constraint.

Optimality in the location choice of any worker-city pair requires that $u_j = u_{j'}$ for all $j' \neq j$. The optimal production of goods in a competitive market with free entry implies that wages are equal to marginal product: $w_j = A_j$.

Optimality in the production of housing in each city j requires that construction companies solve the following maximization problem:

$$\max_{K_j,\, T_j} p_j B[(1-\beta)K_j^{\rho} + \beta T_j^{\rho}]^{1/\rho} - r_j T_j - K_j.$$

This implies the optimal solution $K_j^{\star} = \left(\frac{1-\beta}{\beta} r_j\right)^{\frac{1}{1-\rho}} T_j$. This, together with the zero profit condition allows us to calculate the housing supply in each city, which in turn predicts a relation between the rental price of land r_j and the housing price p_j.

In equilibrium, given amenities a_j, wages w_j, and taxes t_j: i) households choose c_j and h_j optimally; ii) given wages w_j, production firms choose l_j optimally; iii) given p_j and r_j, construction firms choose K_j and T_j optimally; iv) markets clear to pin down prices, w_j, p_j and r_j; and iv) utility equalization across locations pins down l_j. Further details are provided in the Appendix.

4 The Planner's Problem

The objective of our exercise is to evaluate how the efficiency properties of equilibrium allocation vary once we introduce distortions. We focus our attention on the Optimal Ramsey taxation problem where the planner chooses *tax instruments* in order to affect the equilibrium allocation. The planner assumes agents operate in a decentralized economy with equilibrium prices and free choice of consumption and location decisions, albeit affected by a city-specific tax t_j, where $\tilde{w}_j = (1-t_j)w_j$. Consider now a Utilitarian planner who chooses the tax schedule $\{t_j\}$ to maximize the sum of utilities subject to: 1. the revenue neutrality constraint, i.e. she has to raise the same amount of tax revenue; 2. individually optimal choice of goods and housing consumption in a competitive market; and 3. free mobility – utility across local markets is equalized.

As in the case of the equilibrium allocation, the utility given optimal consumption (c, h) in a local labor market is given by (5). Then we can write the Ramsey planner's problem as:

$$\max_{\{t_j\}} \sum_j u_j l_j,$$

subject to $\sum_j A_j t_j l_j = G,\ u_j = u_{j'},\ \forall j' \neq j$, and $\sum_j l_j = \mathcal{L}$.

The solution to this problem involves solving a system of $J + J + 2$ equations (J FOCs and $J + 2$ Lagrangian constraints) in the same number of variables. We cannot derive an analytical solution, so we will characterize the optimal tax schedule from simulating the US economy in the next section. In order to provide intuition for our simulations, we start, however, by showing that the first welfare theorem holds when there is no exogenous government expenditure ($G = 0$), no externalities ($\delta = 0$) and there is no concentration of housing wealth ($\psi = 0$).[10] We then analyze the Ramsey problem for a simple two-city example.

Proposition 1 *Let there be a two city economy with $\beta = 1, \delta = 0, a_j = 1$ and preferences $u(c, h) = c \cdot h$. If there is no government expenditure $G = 0$ and there is no concentration of housing wealth $\psi = 0$, then the decentralized equilibrium allocation and the Ramsey planner's optimal allocation coincide.*

Proof. In Appendix. ■

While this special case provides us with a reference for the case without government expenditure ($G = 0$) and no concentration of housing wealth ($\psi = 0$), it does not give any insights into the role of G and ψ on taxes across locations. For that purpose, we simulate the optimal solution to the Ramsey problem for a two-city example. We obtain two results from this simulation:

1. as government expenditure G increases, relative taxes in big cities decrease (while all taxes increase);
2. as housing wealth concentration ψ increases, relative taxes in big cities increase.

As government expenditure G increases, the planner faces a tradeoff in setting different taxes in big cities relative to small cities: higher taxes in more productive cities generate bigger revenue per person, but attracts fewer workers, and hence leads to a smaller tax base. We find that it is optimal to increase the base in more productive cites: as G increases, the planner taxes those in highly productive city less to make sure that there are enough of them to pay for G.

The result is therefore that relative taxes in big cities decrease as G increases (Figure 1.A). This implies a divergence of the population distribution as the large city becomes larger (Figure 1.B): higher government spending goes together with bigger population differences between small and large difference. That of course implies that output increases in government expenditure since more people live in more productive city, but the output net of government expenditure is decreasing (Figure 1.C).

Taxes in big cities are also affected by the concentration of wealth. As workers locate to big cities, housing prices also increase. As a result, the value of housing that goes to the absentee landlords increases as well. Since the planner does not value the consumption of these absentee landlords, when ψ is high, the optimal taxes in big cities increase relative to those in small cities. The output gains from having more people in productive cities disappear in the pockets of the landowners the planner does not care about. In contrast, for low ψ, the value of housing benefits a larger fraction of households who hold a diversified bond on the economy wide available land. This is illustrated in Figure 2.[11]

[10] We are grateful to John Kennan for pointing us to this equivalence.

[11] In Figure 1, we set the fraction of land owned by measure zero landlords, ψ, to 0.65, the value we use in the quantitative analysis below. Similarly, in Figure 2, G is 16% of output, again close to the value we use in the quantitative analysis.

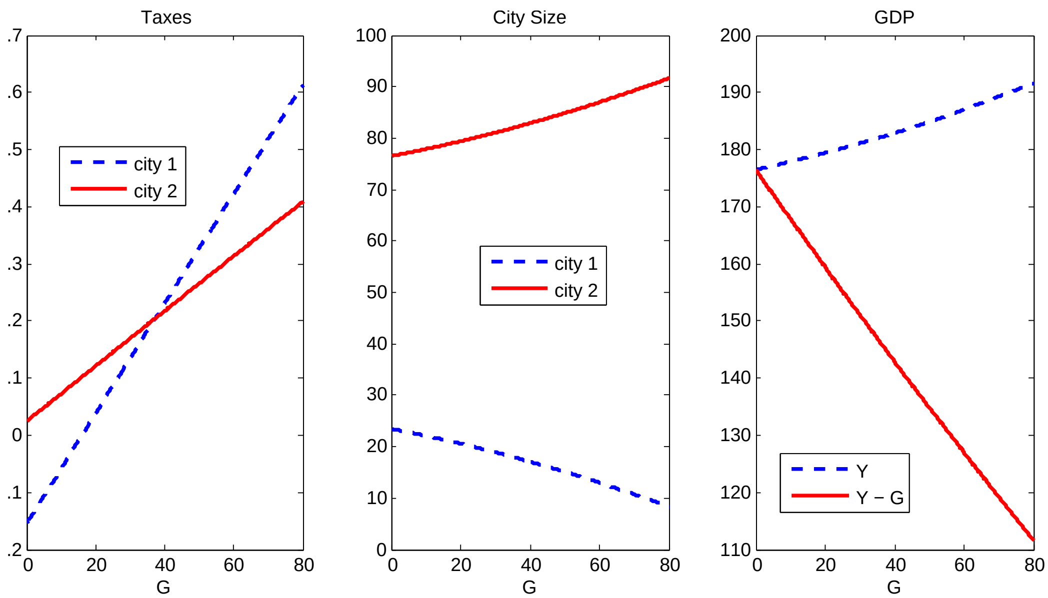

Figure 1: Optimal Ramsey taxes given G in a two city example with a fraction ψ of housing wealth concentration ($\psi = 0.65$): $A_1 = 1, A_2 = 2, \mathcal{L} = 100, \alpha = 0.31$: A. Optimal tax rates t_1, t_2; B. populations l_1, l_2; C. Output Y and output net of government expenditure $Y - G$.

5 Quantifying the Optimal Spatial Tax

We now quantify the magnitude of spatial misallocation. We proceed in following steps: First, given the U.S. data on the distribution of labor force across cities (l_j) and wages in each city (w_j), we back out the productivity parameters A_j. Second, given (l_j, w_j), a representation of current US taxes on labor income, (λ^{US}, τ^{US}), and land area of each city (T_j), we compute a_j values under the assumption that the current allocation of the labor force across cities is an equilibrium, i.e. utility of agents are equalized across cities. Third, for any given $\tau \neq \tau^{US}$, we compute the counterfactual distribution of labor force across cities. In these counterfactuals, we assume revenue neutrality, and for any τ, find the level of λ such that the government collects the same amount of revenue as it does in the benchmark economy. Finally, we find the level of τ that maximizes welfare.

5.1 Labor Force and Wages

The data on the distribution of labor force across cities (l_j) and wages in each city (w_j) are calculated from 2010 American Community Survey (ACS). For 279 Metropolitan Statistical Areas (MSA), we compute l_j as the population above age 16 who are in the labor force. We calculate w_j as weekly wages, i.e. as total annual earnings divided by total number of weeks worked.[12] Figure 12.A and B in the Appendix show the distribution of population and wages across MSAs. The average labor force is

[12]We remove wages that are larger than 5 times the 99th percentile threshold and less than half of the 1st percentile threshold.

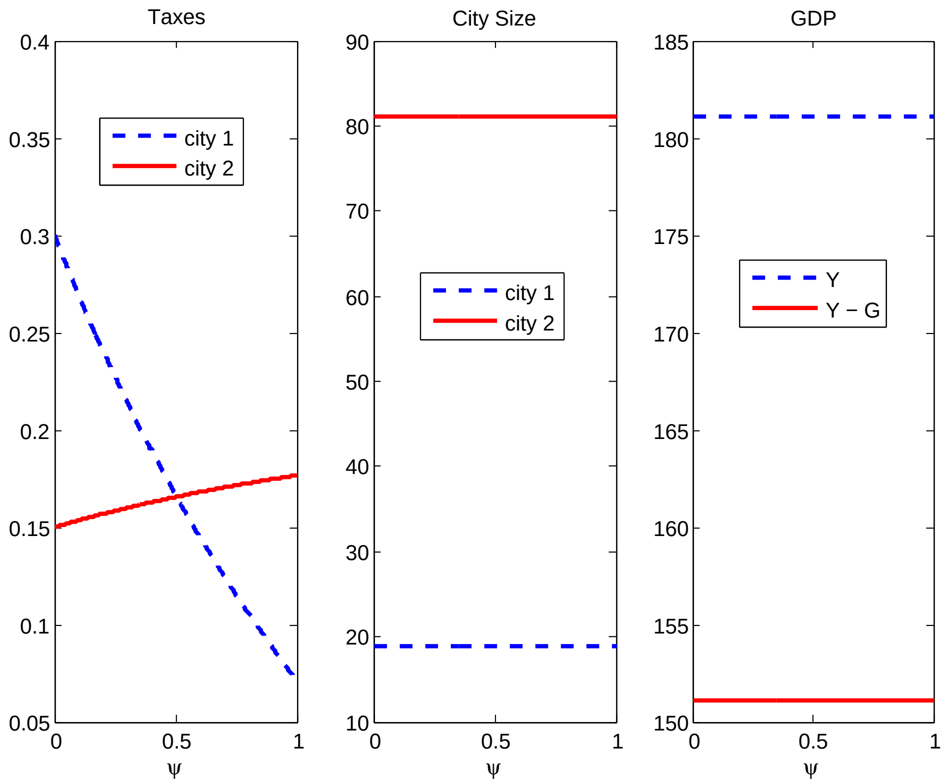

Figure 2: Optimal Ramsey taxes given ψ in a two city example given government expenditure ($G = 30$, 16% of total output): $A_1 = 1, A_2 = 2, \mathcal{L} = 100, \alpha = 0.31$: A. Optimal tax rates t_1, t_2; B. populations l_1, l_2; C. Output Y and output net of government expenditure $Y - G$.

432,523, with a maximum (New York-Northern New Jersey-Long Island) of more than 9 million and a minimum (Gadsden, AL) of about 44,195. The population distribution is highly skewed, close to log-normal, where the top 5 MSAs account for 22.4% of total labor force.[13] Average weekly wages are \$831. The highest weekly wage is more than twice as high as the mean level (Stamford, CT) and the lowest is 64% of the mean level (Laredo, TX). Figure 3 shows the positive relation between population size and wages, the well-known urban wage premium in the data. We take both population and wage date as inputs to simulate the benchmark economy. The elasticity of wages with respect to population size is about 0.07. In Figure 3, as well as in all other figures below, we indicate the ten largest MSA's together with the MSA's with the highest and lowest average wages.

5.2 Taxes

As we mentioned above, we assume that the relation between after and before tax wages are given by $\tilde{w} = \lambda w^{1-\tau}$, where λ is the level of taxation and τ indicates the progressivity ($\tau > 0$). In order to estimate λ and τ for the US economy, we use the OECD tax-benefit calculator that gives the gross

[13]The five largest MSAs are New York, NY-Northeastern NJ; Los Angeles-Long Beach, CA; Chicago, IL; Dallas-Fort Worth, TX; and Washington, DC/MD/VA.

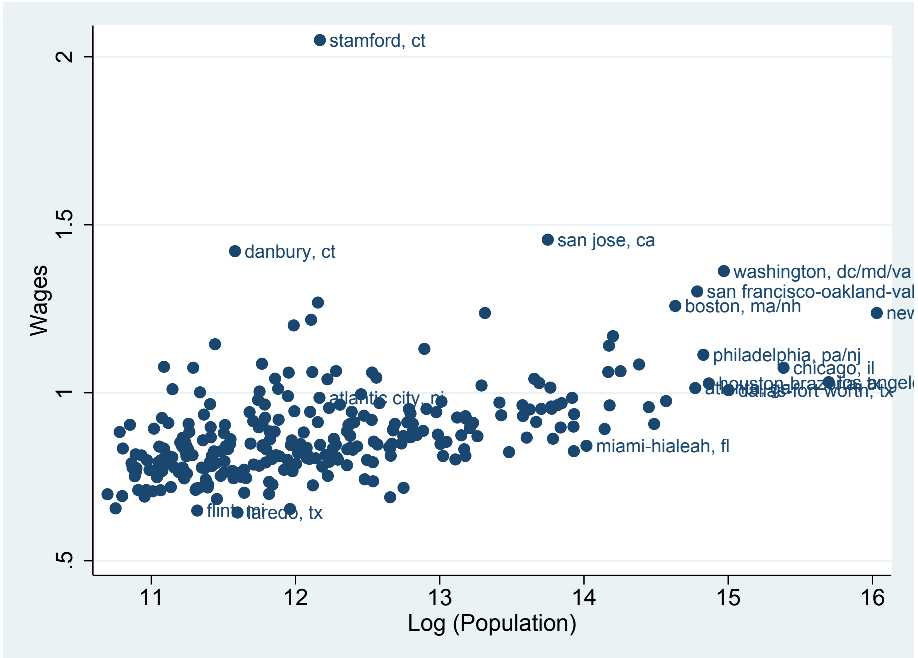

Figure 3: The Urban Wage premium.

and net (after taxes and benefits) labor income at every percentage of average labor income on a range between 50% and 200% of average labor income, by year and family type.[14] The calculation takes into account different types of taxes (central government, local and state, social security contributions made by the employee, and so on), as well as many types of deductions and cash benefits (dependent exemptions, deductions for taxes paid, social assistance, housing assistance, in-work benefits, etc.). Non-wage income taxes (e.g., dividend income, property income, capital gains, interest earnings) and non-cash benefits (free school meals or free health care) are not included in this calculation.

We simulate values for after and before taxes for increments of 25% of average labor income. As the OECD tax-benefit calculator only allows us to calculate wages up to 200% of average labor income, we use the procedure proposed by Guvenen, Burhan, and Ozkan (2014) and detailed in Appendix, to calculate wages up to 800% of average labor income. As a benchmark specification, we calculate taxes for a single person with no dependents. Given simulated values for wages, we estimate a simple OLS regression

$$\ln(\tilde{w}) = \ln(\lambda) + (1 - \tau)\ln(w).$$

The estimated value of τ^{US} is 0.120. Estimating the same tax function with the U.S. micro data on taxes from the Internal Revenue Services (IRS), Guner, Kaygusuz, and Ventura (2014) estimate lower values for τ, around 0.05. Their estimates, however, are for total income while the estimates here are for labor income. One advantage of the OECD tax-benefit calculator, compared to the micro data is that it takes into account social security taxes, which is not possible with the IRS data. Our estimates are

[14] http://www.oecd.org/social/soc/benefitsandwagestax-benefitcalculator.htm, accessed on March 15, 2013.

closer to the ones provided by Guvenen, Burhan, and Ozkan (2014) who also use the OECD tax-benefit calculator to estimate tax rates using a more flexible functional form. Below we report results with Guner, Kaygusuz, and Ventura (2014) estimates for τ as a robustness check.

The parameter λ determines the average level of taxes. We set $\lambda^{US} = 0.85$, i.e. on average taxes are about 15% of GDP in the benchmark economy. This is the average value for sum of personal taxes and contributions to government social insurance program as a percentage of GDP for 1990-2010 period.[15] Hence at mean wages ($w = 1$), tax rate is 15%. Tax rates at $w = 0.5$, $w = 2$ and $w = 5$ are 7.6%, 21.8% and 30.0%, respectively. With $w_2 = 2.5$ and $w_1 = 0.5$, our estimates imply a progressivity wedge of 0.176, defined as $1 - \frac{1-t(w_2)}{1-t(w_1)}$ where $t(w_i)$ is the tax rate at income level w_i.[16]

Figure 4 shows what our representation of the effective Federal Taxes in the US implies for how tax rates differ across cities. In the benchmark economy, each wage level, and as a result each tax rate, corresponds to a city. The average tax rate in San Jose, CA, for example, is almost 10% points higher than it is in Laredo, TX.

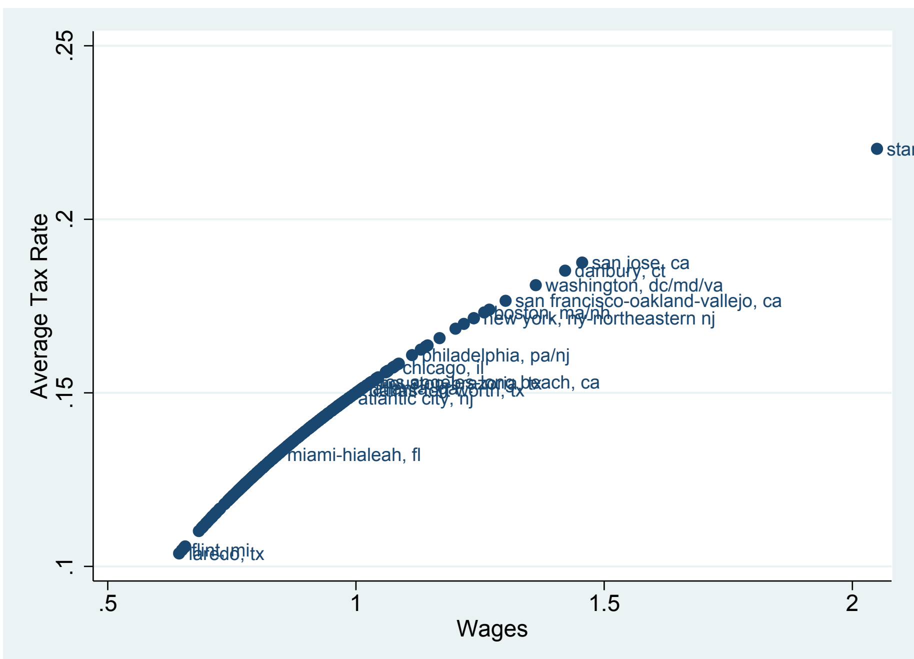

Figure 4: Taxes across cities

Finally, since the share of defense expenditure in the Federal Government's budget is 18% in the US, we assume that the rest, 82% of taxes, is rebated back to households, i.e. $TR = 0.82\frac{G}{T}$.[17]

[15] National Income and Product Accounts, Bureau of Economic Analysis, Table 3.2. Federal Government Current Receipts and Expenditures, http://www.bea.gov/iTable/index_nipa.cfm

[16] Guvenen, Burhan, and Ozkan (2014) estimate a progressivity wedge of 0.15. Given the particular tax function we are using, the progressivity only depends on τ.

[17] National Income and Product Accounts, Bureau of Economic Analysis, Table 3.16. Government Current Expenditures by Function, http://www.bea.gov/iTable/index_nipa.cfm

5.3 Housing Production

The CES housing supply technology basically stipulates that the cost of construction of housing is increasing in the size of the house, but at a (weakly) decreasing rate. If housing capital and land are complements (the elasticity of substitution is less than one), then the housing cost is decreasing in the size of the house. For example, small apartments still need a bathroom and a kitchen, so the unit cost per square meter is higher, or, it is more expensive per unit of housing to build a high-rise than a stand alone home. The implication of this is that the share of land in the value of housing is increasing in the population density, as transpires from the data.

The data on land areas of cities (MSAs), T_j, is taken from the Census Bureau.[18] Average land area of MSAs is about 5254 km^2 and there is very large variation in land areas across MSA.[19] The largest MSA in terms of land areas is huge with 70630 km^2 (Riverside-San Bernardino,CA) while the smallest one has and area of only 312 km^2 (Stamford, CT). Albouy and Ehrlich (2012) document that the share of land in housing is about one-third on average across MSAs and it ranges from 11% to 48%. We set $\beta = 0.235$ and $\rho = -0.2$ to match these two targets in the benchmark economy. Finally, we set $B = 0.028$ such that on average housing consumption is about 200m^2.

5.4 Land Ownership

To determine the share of total land owned by the absentee landlords, ψ, we use the following information on the concentration of housing wealth. First, according to Mishel, Bivens, Gould, and Shierholz (2012), about 12.6% of the housing equity is owned by the top 1% of the wealthy individuals in the US in 2010. Furthermore, Mishel, Bivens, Gould, and Shierholz (2012) also report that in 2006, just before the recent financial crisis, the homeowner equity as a share of total home values was about 60%. We assume that the ownership of the remaining 40%, i.e. debt, is also concentrated. Hence, about 52% of total housing value, 40% of 87.4%, enters into planner's objective function. Finally, only 67% of households own a house in the US between 2000 and 2010.[20] Therefore, we set $1 - \psi$ to be 35% (67% of 52.4%). Hence in the model economy about 65% of land is owned by measure zero landlords and 35% is owned in equal shares by each worker in the economy in the form of a bond that is a diversified portfolio of the country's land supply.

[18] We use data available at https://www.census.gov/population/www/censusdata/files/90den_ma.txt and published in U.S.Census.Bureau (2004).

[19] Figure 13 in the Appendix shows the distribution of land across MSAs.

[20] US Census Bureau Table 5. Homeownership Rates for the United States: 1968 to 2014, available at http://www.census.gov/housing/hvs/data/q314ind.html

5.5 Preferences and Productivity

We calculate productivity level in each city as

$$A_j = w_j, \forall j.$$

Then, we calculate amenities a_j from utility equalization condition across cities. Given the indirect utility function in equation (5), for any two locations j and j', the following equality must hold:

$$\begin{aligned} u_j &= a_j[(1-\alpha)^{1-\alpha}](\tilde{w}_j + R_j + TR)^{1-\alpha} l_j^{\delta-\alpha} H_j^{\alpha} \\ &= a_{j'}[(1-\alpha)^{1-\alpha}](\tilde{w}_{j'} + R_{j'} + TR)^{1-\alpha} l_{j'}^{\delta-\alpha} H_{j'}^{\alpha} \\ &= u_{j'} \end{aligned}$$

Let $a_1 = 1$. Then,

$$\begin{aligned} a_j &= \frac{(\widetilde{w}_1 + R_1 + TR)^{1-\alpha} l_j^{\alpha-\delta} H_1^{\alpha}}{(\widetilde{w}_j + R_j + TR)^{1-\alpha} l_1^{\alpha-\delta} H_j^{\alpha}} \\ &= \frac{(\widetilde{w}_1 + R_1 + TR)^{1-\alpha} l_j^{\alpha-\delta} \left[(1-\beta)\left(\frac{1-\beta}{\beta} r_1\right)^{\frac{\rho}{1-\rho}} + \beta\right]^{\alpha/\rho}}{(\widetilde{w}_j + R_j + TR)^{1-\alpha} l_1^{\alpha-\delta} \left[(1-\beta)\left(\frac{1-\beta}{\beta} r_j\right)^{\frac{\rho}{1-\rho}} + \beta\right]^{\alpha/\rho}} \end{aligned} \tag{6}$$

Calculations for a_j obviously depend, among other parameters, on the values we assume for α and δ. We set $\alpha = 0.319$. Davis and Ortalo-Magné (2011) estimate that households on average spend about 24% of their before-tax income on housing. This would translate to a spending share of $\alpha/\lambda = \frac{0.24}{0.85} = 0.2824$ from after-tax income at mean income ($w = 1$).

We interpret the congestion term $l^{-\delta}$ in the utility as commuting costs and calibrate it using the available evidence on the relationship between city size and commuting costs. The elasticity of commuting time with respect to city size is estimated to be 0.13 by Gordon and Lee (2011). Average commuting time in the US is about 50 minutes (McKenzie and Rapino (2011)). Assuming a 20$ hourly wage, this 50 minutes costs about 17$ for households, which is about 11% of their daily income (17/160). Commuting also has a monetary cost. Roberto (2008) reports that households on average spend about 5% of their income on transportation expenditures, while Lipman et al. (2006) find these costs to be higher, close to 20%. If we take 10% as an intermediate value, then the total, time and money, cost of travel for households is about 20% of their income, which is simply the elasticity of the total commuting costs with respect to the commuting time. As a result, the elasticity of total commuting costs with respect to city size, which is the elasticity of the total commuting costs with respect to the commuting time times the elasticity of commuting time with respect to the city size is $(0.13)(0.2) = 0.026$.[21]

5.6 Benchmark Economy

In Figure 5.A we report the computed values of a_j across metropolitan statistical areas. We set $a_1 = 1$ for New York-Northeastern NJ MSA. The mean value of a_j across MSAs is also about 0.9. The highest levels of a_j, above 1.1, are calculated for Chicago (IL), Los Angeles-Long Beach (CA) and Miami-Hialeah (FL). The calibration procedure assigns a high value of a for Chicago (IL) and Los Angeles-Long Beach (CA) to account for their large size. On the other hand, a relatively low productivity city like Miami-Hialeah (FL), with averages wages that are about 85% of the national average, also requires a high a to justify its size, which possibly reflects better weather conditions.

The lowest values are below 0.7, for Stamford (CT), Anchorage (AK) and Flagstaff (AZ/UT). Stamfod (CT) and Anchorage (AK) are MSAs with high wages but with small populations and low values of a are assigned to justify why more people are not living there. The figure shows the relation between population and amenities adjusted for congestions, i.e. $al^{-\delta}$, across MSAs in the benchmark economy. The correlation between amenities and population size is about 0.14, which is in line with the findings of Albouy (2008) who finds no correlation between amenities and population size.

Panel B in Figure 5 shows the relation between population size and the share of land values in housing prices, which we use as a target to calibrate housing production technology.

The benchmark economy generates a distribution of equilibrium housing prices across MSAs. Estimated housing prices are about 405 per km^2 in San Francisco-Oakland-Vallejo (CA), followed by Stamford (CT) and Chicago (IL) where housing prices are 379 and 372, respectively. The lowest housing prices are computed for Flagstaff (AZ-UT), 31, and Yuma (AZ), 46. While housing consumption is about 200m^2 across MSA, those in Chicago live in houses that are about 86m^2 and about 9 times smaller than houses in Flagstaff (AZ-UT). Panel C in Figure 5 shows the relation between population size and housing prices across MSAs in the benchmark economy. The figure implies an elasticity of housing prices with respect to population size that is about 0.23.

Finally, Figure 6 compares housing prices from the benchmark economy with actual housing prices. It is important to note that we do not target directly actual housing prices in our calibration. In the model economy, housing is a homogenous good with a location specific per unit price p_j. In the data, on the other hand, housing differs in many observable dimensions, and as a result, observed housing prices reflect both the location and the physical characteristics of the unit. We follow Eeckhout, Pinheiro, and Schmidheiny (2014), and estimate the city specific price level as a location-specific fixed effect in a simple hedonic regression of log rental prices on the physical characteristics, such age number of rooms,

[21] In this paper, we assume each city has a different, exogenously given, land area and there is congestion. An alternative strategy would be to endogenize land area by capturing the cost of commuting, for example as in Combes, Duranton, and Gobillon (2013), in the presence of a central business district. However, in our model there is no within city heterogeneity, and commuting costs are captured by the congestion externalities in utility, rather than in housing production. As we show in section 6, incorporating the exact land area in the model is an important ingredient to fit the data.

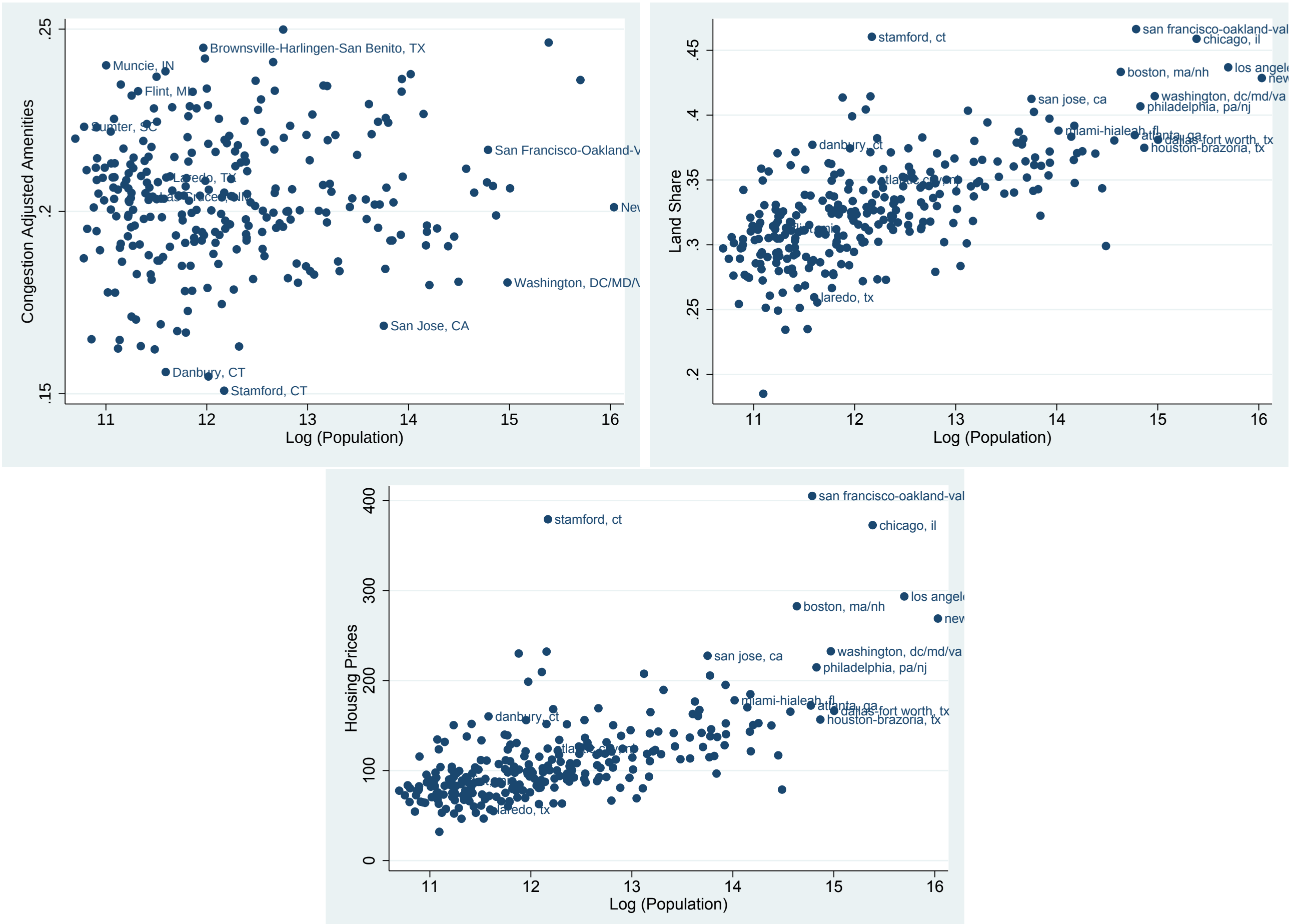

Figure 5: Benchmark Economy. A. Amenities and Population; B. Land Share in the Value of Housing and Population; C. Housing Prices and Population.

age of the unit, and the units structure (one family detached unit vs. one family attached unit etc.).[22] For both the model and the data, we report prices in each city as a fraction of average prices across all cities. The model does a very good job capturing variation in housing prices in the data. The correlation between the model-implied and actual prices is about 60%. The variance of housing prices in the model economy is higher than it is in the data.

5.7 Optimal Allocations

Given values for A_j and a_j, the next step is to find counterfactual allocations for any level of $\tau \neq \tau^{US}$. This is done simply by first writing equation (6) as

[22]We use 2010 American Community Survey (ACS) data on housing rentals and housing characteristics.

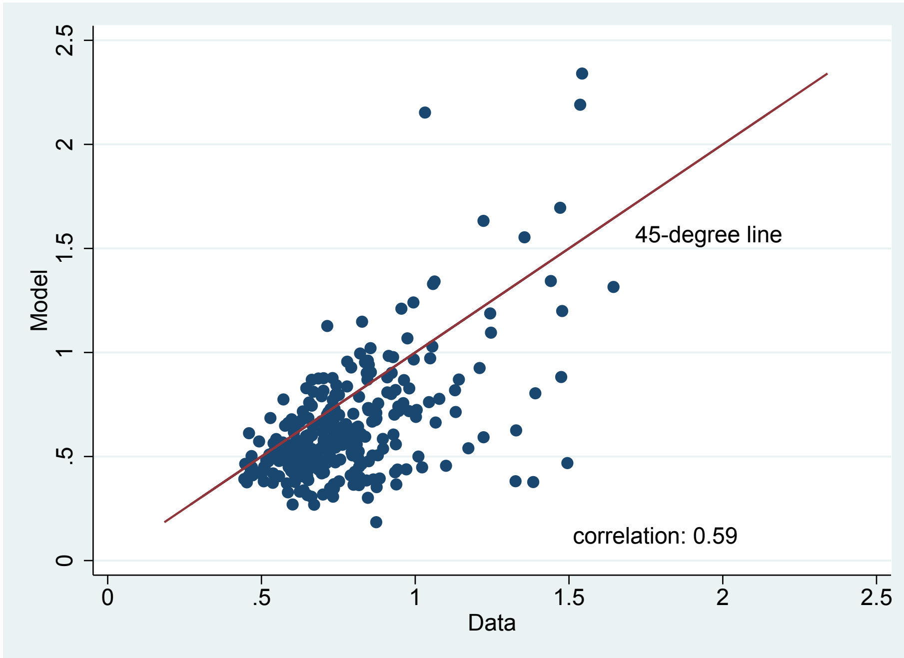

Figure 6: Housing prices: Data versus Model.

$$a_j = \frac{(\lambda w_1^{1-\tau} + R_1 + TR)^{1-\alpha} l_j^{\alpha-\delta} \left[(1-\beta)\left(\frac{1-\beta}{\beta} r_1\right)^{\frac{\rho}{1-\rho}} + \beta\right]^{\alpha/\rho}}{(\lambda w_j^{1-\tau} + R_j + TR)^{1-\alpha} l_1^{\alpha-\delta} \left[(1-\beta)\left(\frac{1-\beta}{\beta} r_j\right)^{\frac{\rho}{1-\rho}} + \beta\right]^{\alpha/\rho}},$$

which can be used to calculate new allocations for any τ

$$l_j(\tau) = l_1(\tau)[a_j^{\frac{1}{\alpha-\delta}} (\frac{\lambda w_j^{1-\tau} + R_j + TR}{\lambda w_1^{1-\tau} + R_1 + TR})^{\frac{1-\alpha}{\alpha-\delta}} (\frac{(1-\beta)\left(\frac{1-\beta}{\beta} r_j\right)^{\frac{\rho}{1-\rho}} + \beta}{(1-\beta)\left(\frac{1-\beta}{\beta} r_1\right)^{\frac{\rho}{1-\rho}} + \beta})^{\left(\frac{\alpha}{\rho}\right)\frac{1}{\alpha-\delta}}].$$

where $l_j(\tau)$ is the counterfactual allocation for tax schedule τ.

We want the counterfactual to be revenue neutral, so for each τ we find a value of λ such that the government collects the same tax revenue as it does in the benchmark economy, i.e.

$$\sum_j l_j(\tau) w_j(\tau)(1 - \lambda w_j^{-\tau}) = \sum_j l_j w_j (1 - \lambda^{US} w_j^{-\tau^{US}}).$$

Finally, we find the value of τ that maximizes the welfare. Figure 7 shows the percentage change in utility from the benchmark economy for different values of τ. The optimal value $\tau^\star$, is 0.0139. The optimal $\tau^\star$ is less than τ^{US}, i.e. taxes in big cities should be lower than those implied by the progressiveness of observed income taxes. However, the optimal τ is not zero. While $\tau = 0$ results in larger movements of population to more productive cities and results in larger output gains, it does not necessarily maximize consumer's utility as consumers are hurt by higher housing prices in larger cities. Figure 7 shows the implied tax schedule under $(\lambda^{US}, \tau^{US})$ and $(\lambda^{\star}, \tau^{\star})$. While, given the particular tax function we use, tax rates for $w = 1$ are identical under two sets of parameters, tax function is more flat with $(\lambda^{\star}, \tau^{\star})$. As a result, for $w = 0.5, w = 2$ and $w = 5$, the tax rates are 14.2%, 15.8% and 16.9%, respectively.

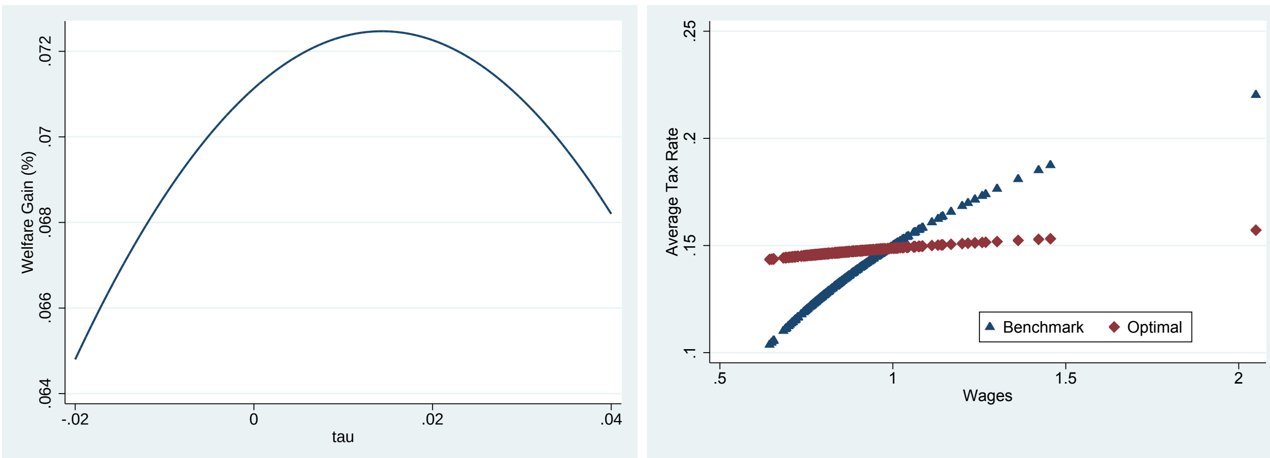

Figure 7: A. Welfare gain for different values of τ; B. The optimal tax schedule $\tau^{\star}$ compared to that in the benchmark economy τ^{US}.

Now we can evaluate the implications of a tax change in the tax schedule from τ^{US} to $\tau^{\star}$, both for individual cities and in the aggregate. Consider first the impact on individual cities which is summarized in Figure 8 and Table 1. The table gives the numerical values for those cities with extreme values either for TFP A or for amenities a.[23]

Since the optimal degree of tax difference $\tau^{\star}$ is below existing τ^{US}, the optimal policy lowers tax payments in high productivity cities. Figure 8.A. shows that the high A cities grow in size while the low productivity A cities loose population. The largest population growth rate, for Stamford (CT), is more than 40% whereas Laredo (TX) looses 25% of its population. As is apparent in Figure 8.B., in contrast with productivity, there is no systematic relation between amenities and population change.

The economic mechanism that drives the population mobility is the following. Due to lower marginal taxes, more productive cities pay higher after tax wages (Figure 8.C). This in turn attracts more workers relative to the benchmark equilibrium with τ^{US}. The new equilibrium is attained when utility across locations equalizes. The main countervailing force that stops further population mobility against the attractiveness of higher after tax wages is housing prices. Figure 8.D shows the change in housing prices. High productivity cities are up to 24% more expensive while low productivity cities face housing price drops of up to 9%.

[23] All our results are robust and change very little if we remove Stamford CT from the sample.

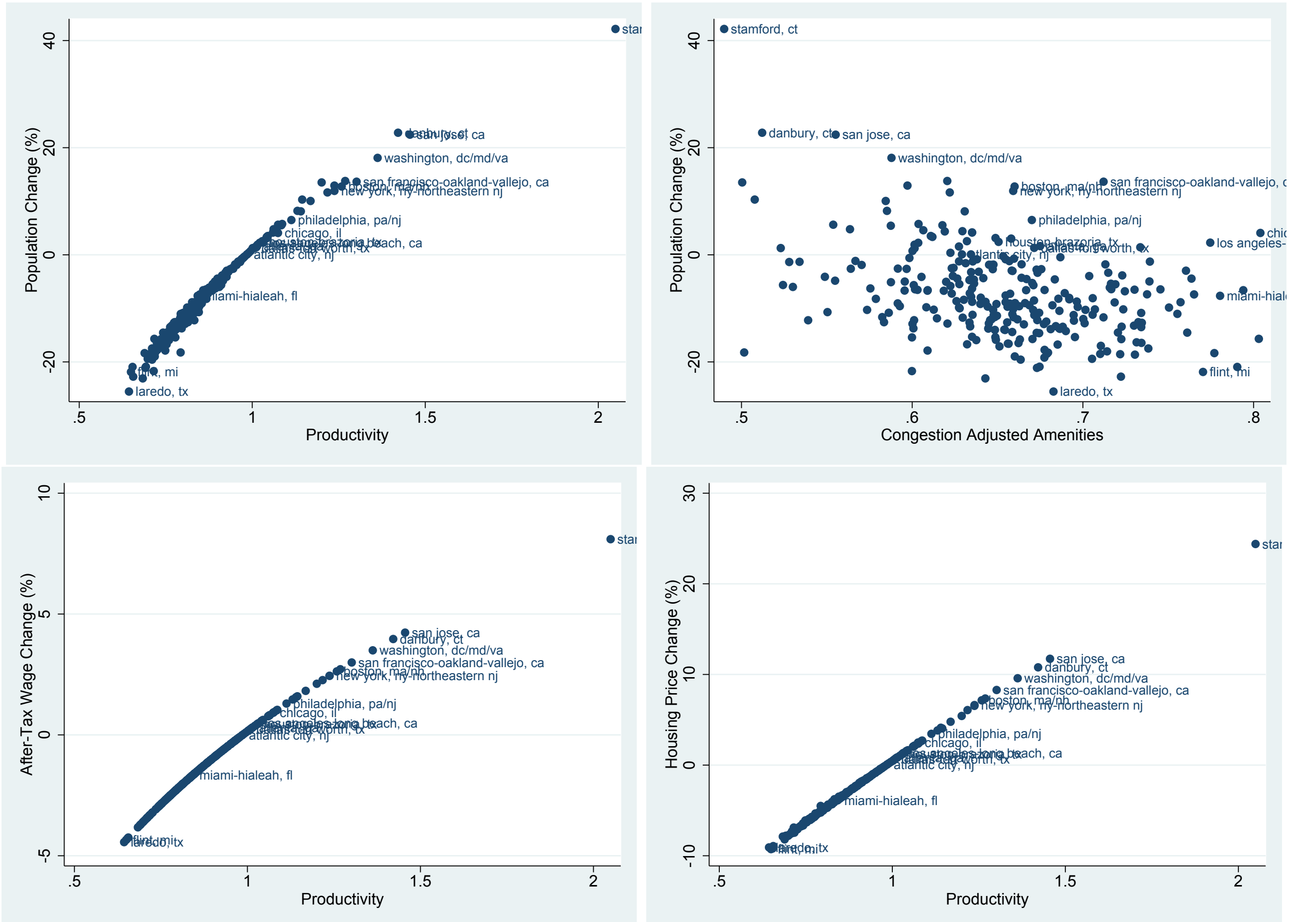

Figure 8: Implied changes of implementing the optimal policy $\tau^{\star}$. A. Change in population by TFP; B. Change in population by a; C. Change in $\tilde{w}$ by TFP; D. Change in housing prices p by TFP.

Figure 9 shows the distribution of output and price changes across MSAs. Output in some MSAs grows as much as 40% while in others it declines by 25%. Output declines in the majority of MSAs, as many small MSAs loose population. Few productive, and large, MSAs on the other hand gain population. The distribution of changes in prices reflects the same forces. Prices decline in many small MSAs, and increase in few large ones.

Of course with higher housing prices goes substitution of housing for consumption. In the high productivity cities, workers live in even smaller housing while increasing goods consumption. Housing consumption decreases by more than 5% in the high productivity cities in substitution for nearly 4% higher goods consumption. In the less productive cities housing consumption increases by up to 6% at the cost of decreased goods consumption by 3%. Given homothetic preferences, the marginal rate of substitution is constant (see Figure 10.A).

Table 2 shows the aggregate outcomes from moving the benchmark allocation to the optimal. On average output and consumption go up by about 1.42% and 1.51%, respectively. This is driven by the population moving to the more productive cities. The population in the 5 largest cities grows by

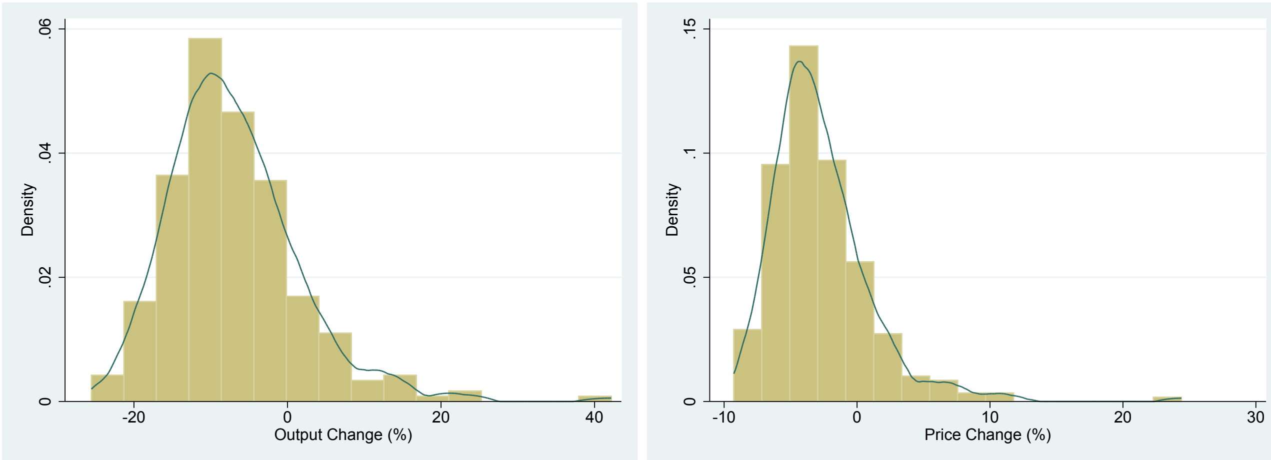

Figure 9: Distribution of changes: A. Ouput; B. Housing Prices.

Table 1: Benchmark Economy, move from τ^{USA} to $\tau^{\star}$. Outcomes for Selected Cities.

MSA	A	a	$\%\Delta l$	$\%\Delta p$	$\%\Delta c$	$\%\Delta h$
Highest A						
Stamford, CT	2.05	0.67	42.17	24.39	7.41	-13.65
San Jose, CA	1.46	0.79	22.43	11.73	3.80	-7.10
Danbury, CT	1.42	0.69	22.79	10.79	3.56	-6.52
Lowest A						
Brownsville, TX	0.65	1.08	-20.94	-9.20	-3.21	6.59
Flint, MI	0.65	1.03	-21.87	-9.27	-3.26	6.62
Laredo, TX	0.64	0.92	-25.53	-9.08	-3.33	6.33
Highest a						
Chicago, IL	1.07	1.20	4.08	2.49	0.87	-1.58
Los Angeles-Long Beach, CA	1.03	1.16	2.27	1.34	0.51	-0.82
Miami-Hialeah, FL	0.84	1.12	-7.66	-3.89	-1.25	2.75
Lowest a						
Anchorage, AK	1.20	0.68	13.51	5.42	1.91	-3.33
Stamford, CT	2.05	0.67	42.17	24.39	7.41	-13.65
Flalgstaff, AZ/UT	0.79	0.67	-18.24	-4.51	-1.74	2.90

7.62%, despite the fact that the top three are large in part because they also offer high amenities a. Most importantly, in the aggregate there is a reallocation of population from less productive, smaller cities to the more productive, larger cities. As a result there is first-order stochastic dominance in the population distribution, as is evident from Figure 10.B. Not surprisingly, aggregate housing prices go up by 5.20%. Due to higher prices, aggregate housing consumption declines by 1.70%.

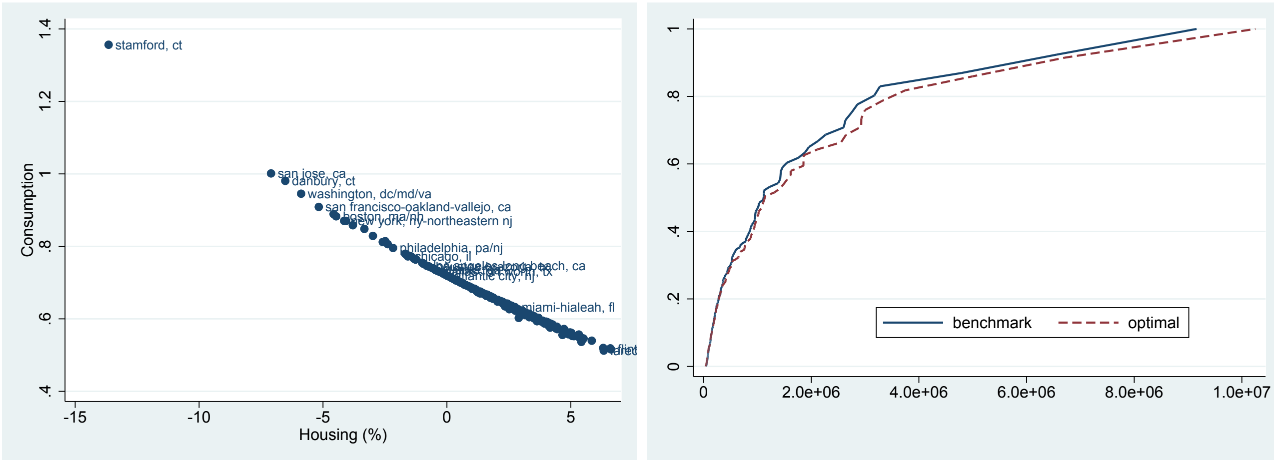

Figure 10: Implied changes of implementing the optimal policy $\tau^{\star}$: A. Substitution between c and h; B. Cumulative Distribution of city sizes.

Despite relatively large output gains, welfare gains are tiny. Given free mobility and a representative agent economy, all agents have the same utility level. After implementing the optimal policy, utility increases by 0.072%, almost nothing. The reason for such tiny welfare gains is quite simple. Under the optimal taxes, after tax wages in cities that have initially high productivity increases. These cities, however, also get more crowded and housing prices goes up. With higher prices, housing consumption in these cities declines. True, from substitution of goods for housing, this generates higher goods consumption. However, the welfare gains associated with higher goods consumption get almost completely offset by lower housing consumption.

Table 2: Benchmark Economy, move from τ to $\tau^{\star}$

Outcomes	
Welfare gain (%)	0.072 0
Output gain (%)	1.42
Consumption (%)	1.51
Housing Consumption (%)	-1.70
Population change top 5 cities (%)	7.62
Fraction of Population that Moves (%)	3.44
Change in average prices (%)	5.20

6 Understanding the Mechanism

In this section, we discuss how sensitive our results are to different aspects of our modeling and calibration choices. First, we focus on two model parameters that play a critical role in determining the optimal level of tax differences across cities: the initial level of government spending (λ) and the concentration of housing ownership (ψ). Next, we show how our results change if we introduce agglomeration economies. Finally, we discuss several model features that do not matter qualitatively but play some role quantitatively. In particular, we show how our results change if in the calculation of wages we control for observable worker characteristics (such as education, age, gender and race) and use a measure of residual wages, if all cities have an equal amount of land (i.e. $T_j = T$ for all j), if there is no housing production and households simply consume land (i.e. $B = \beta = 1$), and if the amount of tax revenue that is transferred back to individuals is zero (i.e. $\phi = 0$).

6.1 What Matters Qualitatively and Quantitatively

6.1.1 The Initial Level of Government Spending

Based on the evidence for the US economy, we have chosen parameter values for λ and τ that are most plausible. The total tax revenue is given by $1 - \lambda$. Our value for the tax revenue of 15% ($\lambda = 0.85$) includes income tax as well as social security taxes. Instead, we could exclude social security contributions in which case tax revenues would be around 10% ($\lambda = 0.9$). Or we could instead allow for the whole tax revenue including corporate and other taxes not related to labor income, in which case tax revenue is 18.5%.[24]

Similarly, to calculate the progressivity, our preferred value of $\tau = 0.12$ reflects taxes on labor income based on the OECD tax calculator. Instead, we could have focused on total household income from the IRS micro data that includes income on assets. Considering both taxes paid and Earned Income Tax Credits (EITC) refunds received by the households, Guner, Kaygusuz, and Ventura (2014) estimate a lower progressivity, $\tau = 0.053$, for all households. Their estimates for married households with children, who are much more likely to benefit EITC, imply a higher progressivity, $\tau = 0.2$.

For these nine parameter combinations, we repeat the same exercise, the results of which are reported in Table 3. For each of the nine combinations we report: the optimal progressively $\tau^\star$, the change in welfare, the changes in output, consumption and housing consumption, the change in population of the top 5 cities, the fraction of movers in the entire economy, and the average housing price change.

The most important thing to observe is that as government spending increases (λ decreases), τ^* decreases and optimal taxes becomes relatively lower in bigger cities. As a result, a smaller government ($\lambda = 0.9$) implies relatively higher taxes in larger, more productive cities (τ^* ranges from 0.06 to almost 0.07 depending on the initial level of τ), while a larger government ($\lambda = 0.815$) results in taxes that are indeed lower in larger, more productive cities (τ^* ranges from -0.02 to almost -0.03 depending on the initial level of τ).

The results in Table 3 indicate that while the initial level of τ does not change the level of optimal dramatically, it does have an important effect on output and as a result welfare. Observe also the

[24]Source: National Income and Product Accounts (NIPA) Table 3.2. - Federal Government Current Receipts and Expenditures.

Table 3: Different initial levels of average taxes and progressivity.

	$\lambda = 0.9$			$\lambda = 0.85$			$\lambda = 0.815$		
τ	0.053	0.12	0.2	0.053	0.12	0.2	0.053	0.12	0.2
Optimal τ^*	0.0604	0.0641	0.0690	0.0097	0.0139	0.0189	-0.0286	-0.0245	-0.0188
Output gain (%)	-0.10	0.78	1.84	0.57	1.42	2.44	1.05	1.88	2.87
Consumption (%)	-0.11	0.82	1.93	0.61	1.51	2.60	1.12	2.01	3.07
Housing Consumption (%)	0.12	-0.92	-2.24	-0.67	-1.70	-3.00	-1.24	-2.27	-3.55
Pop top 5 (%)	-0.55	4.20	9.84	3.08	7.62	13.03	5.63	10.04	15.40
Pop moves (%)	0.25	2.78	6.82	1.40	3.44	5.87	2.55	4.53	6.86
Avg. prices (%)	-0.35	2.78	6.81	2.03	5.20	9.29	3.79	7.00	11.10
Welfare gain (%)	0.0004	0.0217	0.1186	0.0119	0.0720	0.2086	0.0401	0.1254	0.2878

impact on output. If progressivity is low to start with, and government spending is low (the combination $\tau = 0.053, \lambda = 0.9$), then optimality demands more progressivity and as a result, a decrease in output. But output changes are of course biggest when the actual progressivity is high ($\tau = 0.2$). The optimal progressivity is much lower which leads to huge gains in output, up to 3% of GDP. And this goes together with enormous changes in population (up to 11% increase in the top 5 cities) as well as big increases in average housing prices. Even in these extreme cases, the effect on welfare remains mild, precisely because the output gains go hand in hand with increases housing prices.

6.1.2 The Concentration of Housing Ownership

In the benchmark economy, we have modeled ownership of housing as a mixture between a fraction of housing held by households in the form of a perfectly diversified bond and the remainder (ψ) held by a zero measure of landlords. The landlords are supposed to capture the degree of concentration of housing ownership. Those in small cities typically own less valuable housing, thus violating the notion that all households hold an equal (and diversified) portfolio of housing. Now suppose instead that this were nonetheless the case and there are no zero measure landlords ($\psi = 0$). We know then that the outcome with zero government expenditure will generate no tax difference and with positive G taxes will be lower in big cities. This is indeed the case as can be seen in Table 4. The optimal tax has $\tau = -0.1378$ which is more than 10 percentage points lower than the benchmark and 25 percentage points below the current US tax. Not surprisingly, the impact on housing, population and prices is larger: output goes up by nearly 3.1%, the population in the top 5 cities grows by 17% and average housing prices increase by 12.7%. While the welfare effects are substantially higher, they are still relatively small.

Instead at the polar opposite, if housing is fully concentration in the hands of zero measure landlords ($\psi = 1$), then taxes in big cities will be substantially higher. The concentrated ownership of housing does not affect labor productivity and raises no tax revenue. Now there is a tradeoff between attracting more workers to productive cities and loosing revenue due to the fact that housing in big cities is more expensive. The planner resolves this with a higher tax on big cities and as a result the optimal τ is

Table 4: Different Fractions of Zero Measure Landlords (ψ).

Outcomes	Benchmark $\psi = 0.65$	All Bond $\psi = 0$	All Landlords $\psi = 1$
Optimal τ	0.0139	-0.1378	0.0776
Output gain (%)	1.42	3.17	0.59
Consumption (%)	1.51	3.61	0.61
Housing Consumption (%)	-1.70	-3.90	-0.70
Pop. change top 5 cities (%)	7.62	17.22	3.20
Frac. of Pop. that Moves (%)	3.44	7.52	1.45
Change in average prices (%)	5.20	12.67	2.08
Welfare gain (%)	0.0720	0.3386	0.0132

positive. For our economy, it is equal to $\tau = 0.0776$, which is still lower than than τ^{US}. The effects on output and migration is now much more muted.

6.1.3 Agglomeration Economies

There is a large empirical literature in urban economics that documents the extent of agglomeration economies in cities. Rosenthal and Strange (2004), Duranton and Puga (2004), Combes, Duranton, and Gobillon (2013) and Combes and Gobillon (2015) provide reviews of the recent papers that find elasticities of city level productivity with respect to the city size on the order of 0.03 to 0.08.

In this section, we introduce agglomeration economies as an externality in the production function. We assume that the production is given by $F(l_j) = (A_j l_j^\eta) l_j$, where the l_j^η term captures the level of agglomeration economies. Competitive firms still choose l_j to maximize profits, taking as given the externality $(A_j l_j^\eta)$. The resulting wage rate is now given by $w_j = A_j l_j^\eta$, where η is the elasticity of wages with respect to the city size. As in Section 5, we use data on wages and the size of the work force across MSAs to estimate A_j and η, and then repeat our main quantitative exercise.[25] Since given l_j, we estimate A_j and η to fit the observed wages, w_j, in each city, the allocations in the economy with agglomeration externalities are identical to ones in the benchmark economy. The planner problem, on the other hand, now takes into account the fact that a larger workforce in a given city has a positive effect on average wages there.

Table 5 shows the aggregate outcomes for an economy with agglomeration externalities. The planner chooses a regressive tax schedule, i.e. taxes decline in city size, since there is now an extra external benefit of allocating workers to productive cities. As a result, the share of population in the largest five MSAs grows by more than 20% and the resulting reallocation of labor generates a significant output gain that is higher than 4%.[26]

[25] The estimated value of η is about 0.07.

[26] Changing the congestion externality parameter has qualitatively similar effects as the introduction of agglomeration externalities since the functional form is the same. Congestion externalities however are tiny compared to the agglomeration

Table 5: Agglomeration Economies

Outcomes	Benchmark	Aggl. Econ.
Optimal τ	0.0139	-0.0517
Output (%)	1.42	4.31
Consumption (%)	1.51	4.58
Housing Consumption (%)	-1.70	-5.00
Pop. change top 5 cities (%)	7.92	20.65
Frac. of Population that Moves (%)	3.44	9.14
Change in average prices (%)	5.20	15.47
Welfare gain (%)	0.0720	0.3129

6.2 What Does Not Matter Qualitatively

For the next set of variations we show that there is no qualitative change in our conclusions, even if there are quantitative differences worth pointing out.

6.2.1 Controlling for Worker Heterogeneity in Wage Calculations

While the model economy is populated by identical workers, average wages in each MSA in the data reflects several permanent worker characteristics, such as education. If more educated workers sort themselves into more productive cities, then higher average wages in more productive cities would partly be due to higher average human capital of workers in these cities. In order to mitigate this problem, we redo our quantitative exercise using residual average wages for each MSA that controls for workers' education, age, gender and race.[27] As Table 6 shows, with residual wages the optimal level of τ is 0.0237, slightly higher than the optimal level of τ for the benchmark economy. Hence the planner chooses a slightly higher level of progressivity when we use residual wages. This happens as wage differences across MSA is more muted and the potential gains from reallocating workers across MSAs is smaller when wages are calculated net of worker characteristics.[28]

6.2.2 An Economy with Identical Land Areas

In the benchmark economy land areas across cities differ as they do in the data. Consider now an economy, in which each city has the average land area in the data, about 5000 km^2. The economy with identical land areas looks very similar to the benchmark economy with different land areas, with one

externalities we find here.

[27] We estimate the following regression $\log(w_{ij}) = \alpha + \beta_1 \text{Education}_i + \beta_2 \text{Hispanic}_i + \beta_3 \text{White}_i + \beta_4 \text{Age} + \beta_5 \text{Age}^2 + \gamma_j + \varepsilon_{ij}$, where w_{ij} is wage of worker i in MSA j, Education is a dummy variable for educational attinement (with high-school dropouts, high school graduates, college graduate categories), Hispanic and White are dummies for race, and γ_j is a MSA fixed-effect. The residual wage for each worker is calculated as $\alpha + \gamma_j + \varepsilon_{ij}$.

[28] The wage gap between the MSAs with the highest and lowest weekly wages is about $1,300 when we do not control for worker characteristics, while it declines to about $900 for residual wages.

Table 6: Robustness

Outcomes	Benchmark	Residual Wages	Fixed Land	No Housing Production	No Tax Rebates
Optimal τ	0.0139	0.0237	0.0229	0.0833	0.0069
Output (%)	1.42	0.72	1.31	0.17	1.75
Consumption (%)	1.51	0.77	1.39	0.17	2.12
Housing Consumption (%)	-1.70	-1.01	-1.99	0	-2.02
Pop. change top 5 cities (%)	7.92	5.97	6.43	0.97	9.39
Frac. of Population that Moves (%)	3.44	2.47	3.09	0.43	4.25
Change in average prices (%)	5.20	3.22	5.51	1.85	6.52
Welfare gain (%)	0.0720	0.0328	0.0581	0.0027	0.108

important exception. An economy with equal land areas across cities is not able to generate housing price differences across MSAs that are consistent with the data. Figure 11.B shows the housing prices in the model when each city has the same amount of land. In addition and in contrast to the benchmark economy (see Panel C in Figure 5), the relation between city size and housing prices is nearly perfectly correlated.

The optimal level of τ is 0.0247 for an economy with equal sized cities, while it was 0.0145 for the benchmark economy. Hence, the optimal level of tax in big cities is lower under the benchmark with the actual land sizes. This is not surprising. When the all cities have the same land area, it is more costly to move people to some of the productive cities that have indeed quite large land areas in the data, such as New York-Northeastern NJ MSA. Table 6 shows the aggregate outcomes for an economy with equal sized cities. Given the lower values for τ^*, the changes are muted compared to the benchmark economy in which land areas differ across MSAs.

6.2.3 No Housing Production

We next shut down the housing production by setting $B = \beta = 1$. Hence, per-capita housing consumption in a city is simply given by T_j/l_j. Figure 11.C shows the housing prices in the model when there is no production of housing. While the relation between population size and prices are now more in line with what we observe in the data, a model without housing production implies a much larger dispersion of prices across cities. In the benchmark economy, the ratio of the maximum to the minimum housing price is about 13 (407/32). This ratio is much higher, about 1200, when we do not allow for production. Hence, housing production allows us to generate housing prices that are in line with the data.

In an economy without housing production, the optimal level of τ is lower than its benchmark value, i.e. the planner again finds it optimal to lower taxes in larger, more productive cities. The decline in τ, however, is quite muted (τ declines from 0.12 to 0.08 while the decline was from 0.12 to 0.0145 in the benchmark economy). As a result, both output and welfare gains are smaller than they were in

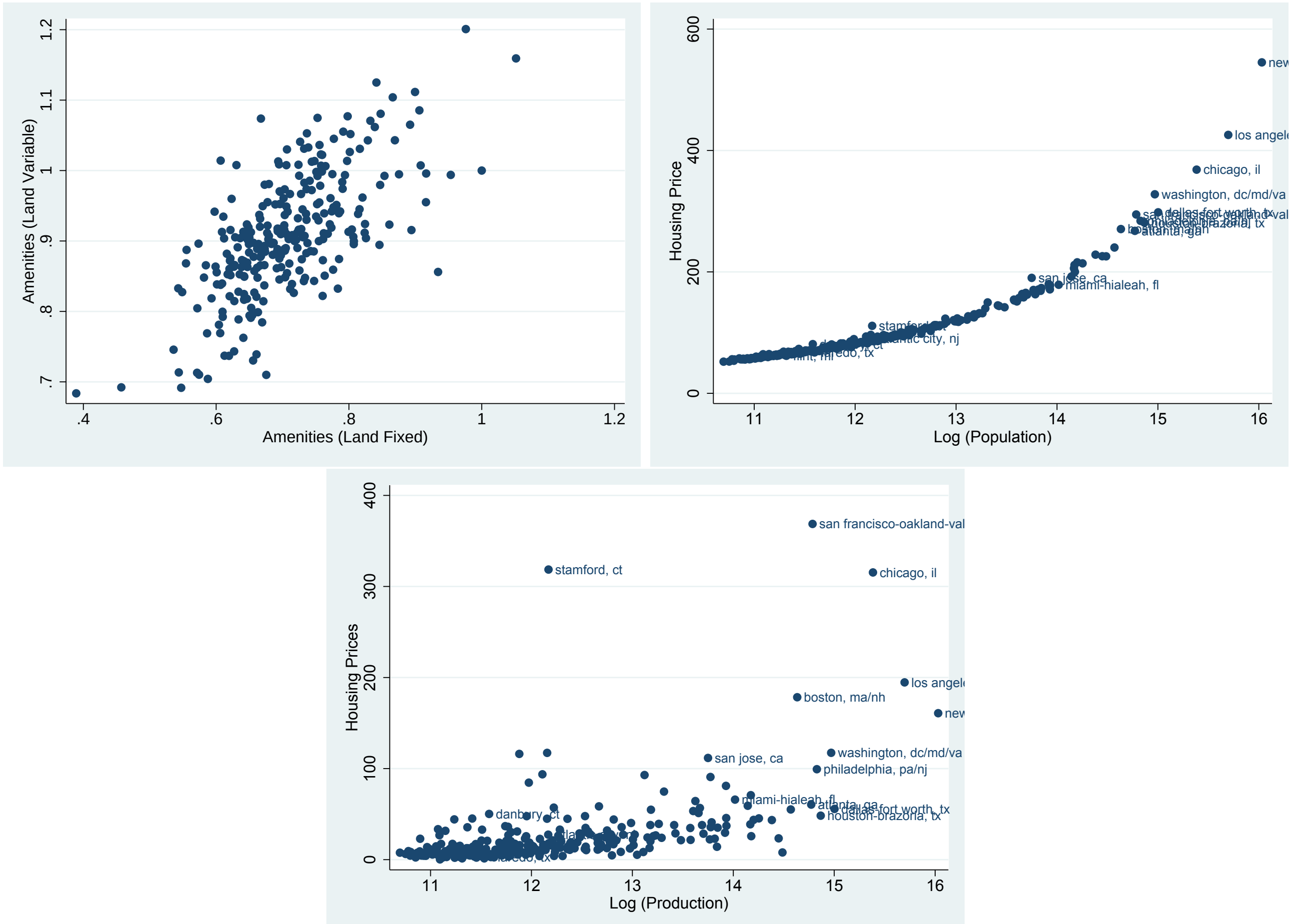

Figure 11: Estimated amenities with fixed land versus the benchmark (variable land): A. Amenities when land is fixed and variable; B. Housing prices when land is fixed; C. Housing prices when there is no housing production.

the benchmark economy. When housing is a stock, reallocation of workers to larger cities has a bigger impact on housing prices. This limits the planner's willingness to lower taxes in more productive cities.

6.2.4 Rebates and Transfers

In the benchmark economy, we assume that 82% of the tax revenue is rebated to households, in the form of transfers that are independent of city size. Table 6 shows the results when we do not redistribute any tax revenue back to households, i.e. $TR = 0$. Since the tax rebate is lump sum, it has exactly the same effect as an increase in government expenditure G. We know from the theory that an increase in G leads to lower taxes in large cities, and hence a lower τ. This is confirmed here as we set the rebate to zero, though quantitatively the impact is small.

7 Conclusions

We have studied the role of federal income taxation on the misallocation of labor across geographical areas. More productive cities pay higher wages, and with progressive taxes, those workers also pay higher average taxes. Given perfect mobility, the tax schedule affects the incentives of workers where to locate. Our objective has been to calculate the shape of the optimal tax schedule in general equilibrium. When taxes change, citizens respond by relocating, but that in turn affects equilibrium prices. Those equilibrium effects determine both the optimal spatial tax schedule as well as the quantitative implications.

Our findings are first, that the optimal spatial tax schedule is not flat and is sensitive to the level of government spending, to the concentration of housing wealth, and to presence of agglomeration externalities. From a welfare viewpoint, what matters for the population allocation is not only the amount of government revenue and hence where it is best generated across differentially productive locations, but also the implied value of housing. While lower taxes in big cities can generate higher aggregate output and government revenue, they also make it more expensive to live.

Second, quantitatively, the optimal tax is less progressive than the existing schedule in the data. Implementing the optimal schedule therefore favors the more productive cities. In equilibrium this leads to output growth economy-wide and population growth in the largest cities. The output growth is 1.42%. At the same time, there is first order stochastic dominance in the city size distribution where the fraction of population in 5 largest cities grows around 8%. The welfare effects however are small, 0.07%. Welfare obviously goes up, but in small amounts. This is due to the fact that the cost of living in the productive cities has increased commensurately. Our quantitative exercise also shows that the size of the government, the concentration of housing wealth, as well as the presence of agglomeration externalities play a critical role in determining the optimal tax differences between large and small cities.

Appendix

Characterization of Equilibrium

Consider first the problem of construction firms. The First Order Conditions are given by

$$p_j B \frac{1}{\rho}[(1-\beta)K_j^{\rho} + \beta T_j^{\rho}]^{\frac{1}{\rho}-1}(1-\beta)\rho K_j^{\rho-1} = 1,$$

and,

$$p_j B \frac{1}{\rho}[(1-\beta)K_j^{\rho} + \beta T_j^{\rho}]^{\frac{1}{\rho}-1}\beta\rho T_j^{\rho-1} = r_j.$$

These conditions imply

$$K_j^{\star} = \left(\frac{1-\beta}{\beta} r_j\right)^{\frac{1}{1-\rho}} T_j,$$

and

$$H_j = B\left[(1-\beta)\left(\frac{1-\beta}{\beta} r_j\right)^{\frac{\rho}{1-\rho}} + \beta\right]^{1/\rho} T_j. \tag{7}$$

The zero-profit condition then implies (after factoring out T_j and r_j):

$$p_j = r_j \frac{\left(1 + \left(\frac{1-\beta}{\beta}\right)^{\frac{1}{1-\rho}} r_j^{\frac{\rho}{1-\rho}}\right)}{B\left[(1-\beta)\left(\frac{1-\beta}{\beta} r_j\right)^{\frac{\rho}{1-\rho}} + \beta\right]^{1/\rho}} \tag{8}$$

From the household problem we know that $p_j h_j = \alpha(\tilde{w}_j + R_j + TR)$. Since market clearing in the housing market requires that $h_j l_j = H_j$, this implies $\alpha(\tilde{w}_j + R_j + TR)l_j = p_j H_j$ which can be written as

$$p_j B\left[(1-\beta)\left(\frac{1-\beta}{\beta} r_j\right)^{\frac{\rho}{1-\rho}} + \beta\right]^{1/\rho} T_j = \alpha l_j(\tilde{w}_j + R_j + TR),$$

or, after substituting equation (8), rearranging and canceling terms:

$$r_j\left(1 + \left(\frac{1-\beta}{\beta}\right)^{\frac{1}{1-\rho}} r_j^{\frac{\rho}{1-\rho}}\right) = \frac{\alpha l_j(\tilde{w}_j + R_j + TR)}{T_j}. \tag{9}$$

Observe that this expression consist of one equation in one unknown, r_j. Given the solution for r_j, we can use equation (8) to find p_j.

In equilibrium each location has to give the same utility. Given equation (5), and normalizing $a_1 = 1$, we have

$$\frac{a_j}{a_1} = \frac{l_1^{\delta}(\tilde{w}_1 + T_1 + TR)\left((\tilde{w}_1 + R_1 + TR)l_{i1}\right)^{-\alpha} H_1^{\alpha}}{l_j^{\delta}(\tilde{w}_j + T_j + TR)\left((\tilde{w}_j + R_j + TR)l_j\right)^{-\alpha} H_j^{\alpha}}.$$

Using the expression for H_j in (7) we obtain:

$$\frac{a_j}{a_1} = \frac{l_1^{\delta}(\tilde{w}_1 + R_1 + TR)\left((\tilde{w}_1 + R_1 + TR)l_1\right)^{-\alpha}\left[(1-\beta)\left(\frac{1-\beta}{\beta}r_1\right)^{\frac{\rho}{1-\rho}} + \beta\right]^{\alpha/\rho}}{l_j^{\delta}(\tilde{w}_j + R_j + TR)\left((\tilde{w}_j + R_j + TR)l_j\right)^{-\alpha}\left[(1-\beta)\left(\frac{1-\beta}{\beta}r_j\right)^{\frac{\rho}{1-\rho}} + \beta\right]^{\alpha/\rho}} \tag{10}$$

The first order condition of production firms implies $w_j = A_j$, and $\tilde{w}_j$ is given by $(1-t_j)w_j$. In equilibrium, individuals are assumed to own land in proportion to their consumption of housing. Therefore R_j satisfies:

$$R_j = (1-\psi)\frac{\sum_j r_j T_j}{\sum_j l_j}. \tag{11}$$

The population allocation must satisfy feasibility: $\sum_j l_j = \mathcal{L}$. Finally, $TR = \frac{\phi G}{\mathcal{L}}$. Hence, equation (10) can be used to pin down l_j for a given values of a_j.

Finally, in order to arrive at the aggregate resource constraint for this economy, we first aggregate household budget constraints, $c_j + p_j h_j \leq \tilde{w}_j + R_j + TR$, across cities

$$\sum_{j=1}^{J} l_j c_j + \sum_{j=1}^{J} l_j p_j h_j = \sum_{j=1}^{J} l_j \tilde{w}_j + \sum_{j=1}^{J} l_j R_j + \sum_{j=1}^{J} l_j TR.$$

Since $l_j h_j = H_j$, $\sum_{j=1}^{J} l_j R_j = \sum_{j=1}^{J} l_j (1-\psi)\frac{\sum_j r_j T_j}{\sum_j l_j} = (1-\psi)\sum_j r_j T_j$, and $\sum_{j=1}^{J} l_j TR = \sum_{j=1}^{J} l_j \frac{\phi G}{\mathcal{L}} = \phi G$, we have

$$\sum_{j=1}^{J} l_j c_j + \sum_{j=1}^{J} p_j H_j = \sum_{j=1}^{J} l_j \tilde{w}_j + (1-\psi)\sum_j r_j T_j + \phi G.$$

Adding and subtracting $\sum_j K_j$ to the right-hand side of this expression, we get

$$\sum_{j=1}^{J} l_j c_j + \sum_{j=1}^{J} p_j H_j = \sum_{j=1}^{J} l_j \tilde{w}_j + \sum_j r_j T_j - \psi)\sum_j r_j T_j + \sum_j K_j - \sum_j K_j + \phi G.$$

Since the housing production function is constant returns to scale, $\sum_{j=1}^{J} p_j H_j = \sum_j r_j T_j + \sum_j K_j$ (recall that $r^K = 1$). Therefore,

$$\sum_{j=1}^{J} l_j c_j = \sum_{j=1}^{J} l_j \tilde{w}_j - \psi \sum_j r_j T_j - \sum_j K_j + \phi G.$$

Finally, note that $\sum_{j=1}^{J} l_j \tilde{w}_j = \sum_{j=1}^{J} l_j w_j - G = \sum_{j=1}^{J} l_j w_j - \sum_j t_j w_j$. Hence,

$$\sum_{j=1}^{J} l_j c_j = \sum_{j=1}^{J} l_j w_j - \sum_j t_j w_j - \psi \sum_j r_j T_j - \sum_j K_j + \phi \sum_j t_j w_j,$$

which delivers us the aggregate resource constraint for the economy

$$\sum_{j=1}^{J} l_j c_j + \psi \sum_j r_j T_j + \sum_j K_j + (1-\phi) \sum_j t_j w_j = \sum_{j=1}^{J} l_j w_j.$$

Proof of Proposition 1

Decentralized Equilibrium. Each consumer in city $i \in \{1, 2\}$ maximizes his utility

$$\max_{c_i, h_i} c_i h_i$$

subject to

$$c_i + p_i h_i = w_i + R,$$

where $R = p_1 + p_2$.

Let x be the fraction of population that lives in city 1. Then market clearing conditions are given by

$$\frac{c_1}{x} = \frac{c_2}{1-x}, \; h_1 = \frac{1}{x}, \text{ and } h_2 = \frac{1}{1-x}$$

The entire system then is

$$c_1 = \frac{w_1 + R}{2}, \; c_2 = \frac{w_2 + R}{2}, \; \frac{1}{x} = \frac{w_1 + R}{2p_1}, \text{ and } \frac{1}{1-x} = \frac{w_2 + R}{2p_2}.$$

We can rewrite $R = p_1 + p_2 = x w_1 + (1-x) w_2$ so that

$$\frac{w_1 + R}{x} = \frac{w_2 + R}{1-x},$$

or

$$\frac{w_1 + x w_1 + (1-x) w_2}{x} = \frac{w_2 + x w_1 + (1-x) w_2}{1-x}.$$

The last expression produces a quadratic expression in x

$$2(w_2 - w_1) x^2 - 4 w_2 x + (w_1 + w_2) = 0.$$

We have two explicit solutions for x:

$$x^* = \frac{2w_2 \pm \sqrt{2(w_1^2 + w_1^2)}}{2(w_2 - w_1)}$$

Ramsey Problem. We solve the Ramsey problem where the planner chooses a taxes t_1 and t_2 such that:

$$\begin{aligned}
\max_{t_1,t_2} \quad & c_1 x^{-1} x + c_2 (1-x)^{-1}(1-x) \\
\text{s.t.} \quad & w_1 t_1 x + w_2 t_2 (1-x) = 0 \\
& \frac{c_1}{x} = \frac{c_2}{1-x} \\
& c_1 = \frac{1}{2}\left((1-t_1)w_1 + R\right) \\
& c_2 = \frac{1}{2}\left((1-t_2)w_2 + R\right) \\
& h_1 = \frac{1}{x} = \frac{1}{2}\frac{(1-t_1)w_1 + R}{p_1} \\
& h_2 = \frac{1}{1-x} = \frac{1}{2}\frac{(1-t_2)w_2 + R}{p_2} \\
& R = p_1 + p_2.
\end{aligned}$$

This problem is equivalent to the following problem

$$\begin{aligned}
\max_{t_1,t_2} \quad & c_1 + c_2 = \frac{1}{2}\left[(1-t_1)w_1 + (1-t_2)w_2\right] + R \\
\text{s.t.} \quad & t_2 = \frac{-w_1 t_1 x}{w_2(1-x)} \\
& \frac{(1-t_1)w_1 + R}{x} = \frac{(1-t_2)w_2 + R}{1-x} \\
& p_1 = x\frac{(1-t_1)w_1 + R}{2} \\
& p_2 = (1-x)\frac{(1-t_2)w_2 + R}{2} \\
& R = p_1 + p_2
\end{aligned}$$

where we rewrite $R = (p_1 + p_2) = x(1-t_1)w_1 + (1-x)(1-t_2)w_2$. Then we have,

$$\begin{aligned}
\max_{t_1,t_2} \quad & \frac{1}{2}\left[(1-t_1)w_1 + (1-t_2)w_2\right] + x(1-t_1)w_1 + (1-x)(1-t_2)w_2 \\
\text{s.t.} \quad & t_2 = \frac{-w_1 t_1 x}{w_2(1-x)} \\
& \left[(1+x)(1-t_1)w_1 + (1-x)(1-t_2)w_2\right](1-x) = \left[x(1-t_2)w_1 + (2-x)(1-t_2)w_2\right]x
\end{aligned}$$

or

$$
\begin{aligned}
\max_{t_1,t_2} \quad & \left(\frac{1}{2}+x\right)(1-t_1)w_1 + \left(\frac{3}{2}-x\right)(1-t_2)w_2 \\
\text{s.t.} \quad & t_2 = \frac{-w_1 t_1 x}{w_2(1-x)} \\
& 2\left[(1-t_2)w_2 - (1-t_1)w_1\right]x^2 - \left[4(1-t_2)w_2\right]x + \left[(1-t_1)w_1 + (1-t_2)w_2\right] = 0
\end{aligned}
$$

Note that if $t_1 = t_2 = 0$, the x chosen is equal to the x^* in the competitive equilibrium.

$$2(w_2 - w_1)x^2 - 4w_2 x + (w_1 + w_2) = 0.$$

Constrained Planner Choosing Populations. Finally, we also show that the Ramsey planner's solution can be attained by a planner who chooses the population across cities and who takes the constraint that workers are mobile and equate utility as given.

The planner maximizes weighted aggregate utility by choosing where the population lives:

$$
\begin{aligned}
\max_{x} \quad & c_1 h_1 x + c_2 h_2 (1-x) \\
\text{s.t.} \quad & x c_1 + (1-x)c_2 = x w_1 + (1-x) w_2 \\
& \frac{c_1}{x} = \frac{c_2}{1-x}. \\
& h_1 = \frac{1}{x} \\
& h_2 = \frac{1}{1-x}
\end{aligned}
$$

Which implies

$$
\begin{aligned}
\max_{x} \quad & c_1 \frac{1}{x} \\
\text{s.t.} \quad & x^2 c_1 + (1-x)^2 c_1 = x^2 w_1 + (1-x) x w_2,
\end{aligned}
$$

or

$$\max_{x} \frac{x w_1 + (1-x) w_2}{x^2 + (1-x)^2}.$$

The FOC is:

$$\frac{(w_1 - w_2)(x^2 + (1-x)^2) - (x w_1 + (1-x) w_2)(2x - 2(1-x))}{(x^2 + (1-x)^2)^2} = 0.$$

The last expression can be simplified as

$$\begin{aligned}(w_1 - w_2)(x^2 + (1-x)^2) - 2(xw_1 + (1-x)w_2)(2x-1) &= 0\\(w_1 - w_2)(2x^2 + 1 - 2x) + 2(xw_1 + (1-x)w_2) - (4x^2 w_1 + (4x - 4x^2)w_2) &= 0\\2(w_2 - w_1)x^2 - 4w_2 x + (w_1 + w_2) &= 0\end{aligned}$$

We thus find the same outcome x^* as before.

Wage and Population Distributions

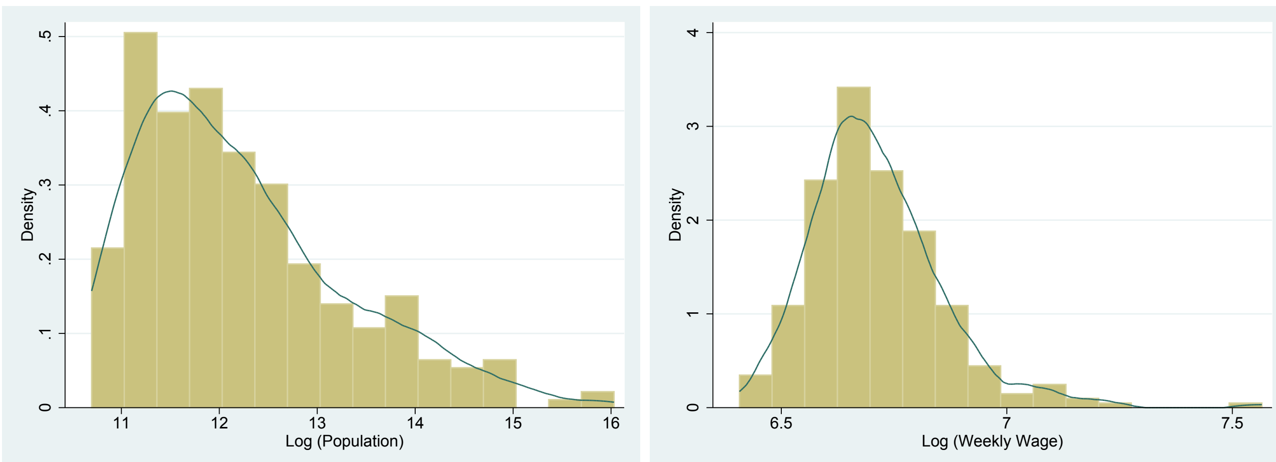

Figure 12: A. Histogram and Kernel density of labor force; B. Histogram and Kernel density of wages.

Estimating the Tax Functions

The OECD tax-benefit calculator provides the gross and net (after taxes and benefits) labor income at every percentage of average labor income on a range between 50% and 200% of average labor income, by year and family type. We simulate values for after and before taxes for increments of 25% of average labor income. As the OECD tax-benefit calculator only allows us to calculate wages up to 200% of average labor income, we use the procedure proposed by Guvenen, Burhan, and Ozkan (2014). In particular, let w denote average wage income before taxes as a multiple of mean wage income before taxes, and $t(w)$ and $\bar{t}(w)$ the marginal and average tax rates on wage income w. Also let t_{top} and w_{top} be the top marginal tax rate and top marginal income tax bracket.[29] Suppose $w > 2$ and $w_{top} < 2$, i.e. top income bracket is less than 2. Then,

$$t(w) = \frac{(\bar{t}(2) \times 2 + t_{top} \times (w-2))}{w}.$$

[29] Top marginal tax rate is taken from http://www.oecd.org/tax/tax-policy/oecdtaxdatabase.htm, Table I.7.

If $w_{top} > 2$ (which is the case for the US), we do not know the marginal tax rate between $w = 2$ and w_{top}. First set

$$t(2) = \frac{(\bar{t}(2) \times 2 - \bar{t}(1.75) \times 1.75)}{0.25}$$

and use linear interpolation between $t(2)$ and t_{top}

$$t(w) = \begin{cases} (t(2) + \frac{t_{top} - t(2)}{w_{top} - 2}(w - 2) & \text{if } 2 < w < w_{top} \\ t_{top} & \text{if } w > w_{top} \\ \end{cases}$$

Then average tax rate function for $w > 2$ is

$$\bar{t}(w) = \begin{cases} (\bar{t}(2) \times 2 + t(w) \times (w - 2))/w & \text{if } 2 < w < w_{top} \\ (\bar{t}(2) \times 2 + \frac{t_{top} + t(2)}{2}(w_{top} - 2) + t_{top} \times (w - w_{top}))/w & \text{if } w > w_{top} \end{cases}$$

Land Distribution across MSA

The following figure shows the distribution of land across MSA.

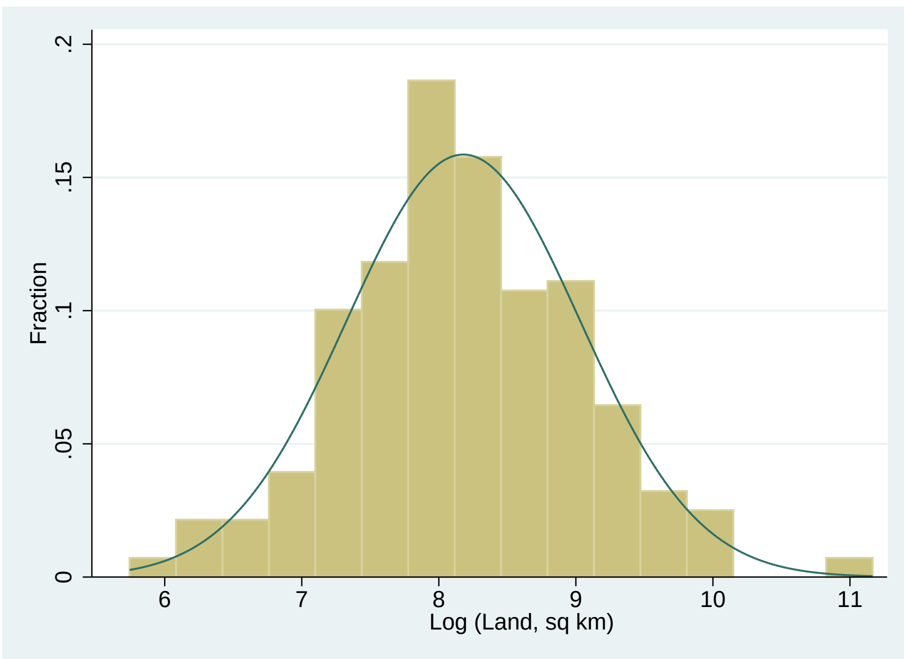

Figure 13: Land Distribution

References

ALBOUY, D. (2008): "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas," NBER Working Papers 14472, National Bureau of Economic Research, Inc.

——— (2009): "The Unequal Geographic Burden of Federal Taxation," *Journal of Political Economy*, 117(4), 635–667.

ALBOUY, D., K. BEHRENS, F. ROBERT-NICOUD, AND N. SEEGERT (2016): "The optimal distribution of population across cities," Discussion paper, National Bureau of Economic Research.

ALBOUY, D., AND G. EHRLICH (2012): "Metropolitan Land Values and Housing Productivity," NBER Working Paper No. 18110.

BÉNABOU, R. (2002): "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?," *Econometrica*, 70(2), 481–517.

COEN-PIRANI, D. (2010): "Understanding gross worker flows across US states," *Journal of Monetary Economics*, 57(7), 769–784.

COMBES, P.-P., G. DURANTON, AND L. GOBILLON (2013): "The Costs of Agglomeration: Land Prices in French Cities," PSE Working Papers halshs-00849078, HAL.

COMBES, P.-P., AND L. GOBILLON (2015): "The Empirics of Agglomeration Economies," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. F. thisse, vol. 5, pp. 247–348. Elsevier B.V.

DAVIS, M., AND J. HEATHCOTE (2007): "The Price and Quantity of Residential Land in the United States," *Journal of Monetary Economics*, 54(8), 2595–2620.

DAVIS, M., AND M. G. PALUMBO (2008): "The Price of Residential Land in Large U.S. Cities," *Journal of Urban Economics*, 63(1), 352–384.

DAVIS, M. A., J. D. FISHER, AND M. VERACIERTO (2013): "Gross migration, housing and urban population dynamics," Chicago Fed mimeo.

DAVIS, M. A., AND F. ORTALO-MAGNÉ (2011): "Household Expenditures, Wages, Rents," *Review of Economic Dynamics*, 14(2), 248–261.

DESMET, K., AND E. ROSSI-HANSBERG (2013): "Urban Accounting and Welfare," *American Eco*, 103(6), 2296–2327.

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

Duranton, G., and D. Puga (2004): "Micro-foundations of urban agglomeration economies," *Handbook of regional and urban economics*, 4, 2063–2117.

Eeckhout, J. (2004): "Gibrat's Law for (All) Cities," *American Economic Review*, 5(94), 1429–1451.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014): "Spatial Sorting," *Journal of Political Economy*, 3(122), 554–620.

Fajgelbaum, P. D., E. Morales, J. C. Suárez-Serrato, and O. Zidar (2016): "State Taxes and Spatial Misallocation," NBER working paper.

Glaeser, E. L. (1998): "Should transfer payments be indexed to local price levels?," *Regional Science and Urban Economics*, 28(1), 1–20.

Glaeser, E. L., and J. M. Shapiro (2003): "The Benefits of the Home Mortgage Interest Deduction," *Tax Policy and the Economy*, 17, 37–82.

Guner, N., R. Kaygusuz, and G. Ventura (2014): "Income Taxation of U.S. Households: Facts and Parametric Estimates," *Review of Economic Dynamics*, 17(4), 559–581.

Guner, N., M. Lopez-Daneri, and G. Ventura (2016): "Heterogeneity and Government revenues: Higher taxes at the top?," *Journal of Monetary Economics*, 80, 69–85.

Guner, N., G. Ventura, and X. Yi (2008): "Macroeconomic Implications of Size-Dependent Policies," *Review of Economic Dynamics*, 11(4), 721–744.

Guvenen, F., K. Burhan, and S. Ozkan (2014): "Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis," *Review of Economic Studies*, 81, 818–850.

Heathcote, J., K. Storesletten, and G. L. Violante (2017): "Optimal tax progressivity: An analytical framework," *Quarterly Journal of Economics*, forthcoming.

Helpman, E., and D. Pines (1980): "Optimal Public Investment and Dispersion Policy in a System of Open Cities," *The American Economic Review*, 70(3), 507–514.

Herkenhoff, K. F., L. E. Ohanian, and E. C. Prescott (2017): "Tarnishing the Golden State: Regulations and the US Slowdown," University of Minnesota mimeo.

Hochman, O., and D. Pines (1993): "Federal Income Tax and Its Effects On Inter- and Intracity Resource Allocation," *Public Finance Quarterly*, 21(3), 276–304.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124(4), 1403–1448.

Hsieh, C.-T., and E. Moretti (2015): "Why Do Cities Matter? Local Growth and Aggregate Growth," NBER Working Paper No. 21154.

Kaplan, G., and S. Schulhofer-Wohl (2017): "Understanding the Long-Run Decline in Interstate Migration," *International Economic Review*, 58(1), 57–94.

Kaplow, L. (1995): "Regional Cost-of-Living Adjustments in Tax/Transfer Schemes," *Tax Law Review*, 51, 175–198.

Karahan, F., and S. Rhee (2014): "Population aging, migration spillovers, and the decline in interstate migration," NY Fed mimeo.

Kennan, J. (2013): "Open Borders," *Review of Economic Dynamics*, 16(2), L1–L13.

Kindermann, F., and D. Krueger (2014): "High marginal tax rates on the top 1%? Lessons from a life cycle model with idiosyncratic income risk," Discussion paper, National Bureau of Economic Research.

Klein, P., and G. Ventura (2009): "Productivity Differences and The Dynamic Effects of Labor Movements," *Journal of Monetary Economics*, 56(8), 1059–1073.

Knoll, M. S., and T. D. Griffith (2003): "Taxing Sunny Days: Adjusting Taxes for Regional Living Costs and Amenities," *Harvard Law Review*, 116, 987–1023.

Lipman, B. J., et al. (2006): "A heavy load: The combined housing and transportation burdens of working families," John D. and Catherine T. MacArthur Foundation.

McKenzie, B., and M. Rapino (2011): "Commuting in the United States: 2009," American Community Survey Reports, ACS-15. U.S. Census Bureau.

Mishel, L., J. Bivens, E. Gould, and H. Shierholz (2012): *The state of working America*. Cornell University Press.

Notowidigdo, M. J. (2011): "The incidence of local labor demand shocks," Discussion paper, National Bureau of Economic Research.

Parkhomenko, A. (2016): "The Rise of Housing Supply Regulation in the U.S.: Local Causes and Aggregate Implications," mimeo.

Restuccia, D., and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Review of Economic Dynamics*, 11(4), 707–720.

Roberto, E. (2008): "Commuting to Opportunity: The Working Poor and Commuting in the United States," The Brookings Institution Series on Transportation Reform.

Rosenthal, S. S., and W. C. Strange (2004): "Evidence on the Nature and Sources of Agglomeration Economies," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. F. thisse, vol. 4, pp. 2119–2171. Elsevier B.V.

Tiebout, C. (1956): "A Pure Theory of Local Expenditures," *Journal of Political Economy*, 64(5), 416–424.

Topa, G., A. Sahin, and G. Violante (2014): "Mismatch Unemployment," *American Economic Review*, 104(11), 3529–3564.

U.S.Census.Bureau (2004): "Population and Housing Unit Counts PHC-3-1," in *2000 Census of Population and Housing*. United States Summary, Washington, DC.

Wildasin, D. E. (1980): "Locational efficiency in a federal system," *Regional Science and Urban Economics*, 10(4), 453–471.