

Eeckhout, Jan; Guner, Nezh

Working Paper

Optimal Spatial Taxation: Are Big Cities Too Small?

IZA Discussion Papers, No. 8781

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Eeckhout, Jan; Guner, Nezh (2015) : Optimal Spatial Taxation: Are Big Cities Too Small?, IZA Discussion Papers, No. 8781, Institute for the Study of Labor (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/107574>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

IZA DP No. 8781

Optimal Spatial Taxation: Are Big Cities Too Small?

Jan Eeckhout
Nezih Guner

January 2015

Optimal Spatial Taxation: Are Big Cities Too Small?

Jan Eeckhout

University College London and Barcelona GSE-UPF

Nezih Guner

*ICREA-MOVE, Universitat Autònoma de Barcelona,
Barcelona GSE and IZA*

Discussion Paper No. 8781
January 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Optimal Spatial Taxation: Are Big Cities Too Small?*

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 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 depend on the level of government spending and on the concentration of housing wealth. Our estimates for the US imply higher marginal rates in big cities, but lower than what is 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 8.0% with 3.5% of the country-wide population moving to bigger cities. The welfare gains however are smaller. Aggregate consumption goes up by 1.53%. 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 housing consumption goes down by 1.75%.

JEL Classification: H21, J61, R12, R13

Keywords: misallocation, taxation, population mobility, city size, general equilibrium

Corresponding author:

Nezih Guner
MOVE (Markets, Organizations and Votes in Economics)
Facultat d'Economia, Universitat Autònoma de Barcelona
Edifici B – Campus de Bellaterra
08193 Bellaterra
Cerdanyola del Vallès (Spain)
E-mail: nezih.guner@movebarcelona.eu

* We are grateful to seminar audiences and numerous colleagues, and in particular to Morris Davis, John Kennan, Kjetil Storesletten, Michele Tertilt, Aleh Tsyvinski, and Tony Venables for detailed discussion and insightful comments. Eeckhout gratefully acknowledges support by the ERC, Grant 339186. Guner gratefully acknowledges support by the ERC, Grant 263600.

1 Introduction

What is the role of income taxation for the location choice of agents across different cities? We argue that taxation is an institution that affects the allocation of resources across space and can lead to inefficiency. Wages and productivity for identical workers are considerably higher in larger cities. This is known as the Urban Wage Premium. 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. 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 an average worker is almost 5 percentage points higher in NY than it is in Asheville.

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. The contribution of our work is therefore to move beyond the results of the partial equilibrium models that exist in the literature. Those models do not allow us to evaluate optimal tax policy nor can they be used to perform quantitative tax policy experiments and characterize the optimal tax policy.

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. The optimal tax rates depend on the level of government spending and on the concentration of housing wealth. On the one hand, taxes in big cities relative to those in small cities decrease as government spending increases. Higher government spending increases all taxes, but it is more efficient to generate the revenue by attracting more workers to the big, more productive city. This is achieved by setting relatively low taxes in big cities. On the other hand, relative taxes in the big 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.

Quantifying these findings for the US economy using the current taxation regime, we find a rationale for city specific taxation with higher taxes in big cities relative to small cities as we currently observe due to existing progressive federal income tax schedules, but the optimal tax difference between big and small cities should be lower. Implementing the optimal tax schedule implies that after tax wages increase in large cities taking advantage of the higher TFP of workers in large cities. As a result, there is a first order stochastic dominance shift in the city size distribution.

For US data, the impact of the optimal tax policy are far reaching. In the benchmark economy, the population in five largest cities grows by 7.95%. About 3.5% of the workforce move from smaller to bigger cities countrywide. The aggregate output increases by 1.57%. The gains in terms of utility are, however, much smaller. The experiment that results in an 1.57% 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 away by higher housing prices, which go up by 5.3% on average. As a result, while aggregate consumption goes up by 1.53%, aggregate housing consumption declines by 1.75%. 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 reality. 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. While we do not explicitly model expenditure on public goods, this accounts for the fact that tax revenues also generate benefits.²

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.³ 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. Our results on reallocation of labor across cities echoes

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

²We 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 Glaeser (1998)). In our model, we abstract from this important channel altogether and focus on the role of active, full time workers.

³See also Sahin, Song, Topa, and Violante (2014) for the role of unemployment frictions on spatial mismatch.

Klein and Ventura (2009) and Kennan (2013), who find quite larger output gains from free mobility of workers across countries. 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, 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.

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) and Helpman and Pines (1980) are the first to explicitly consider federal taxation and argue that it creates distortions. They propose taxing the immobile commodity, land, to achieve the efficient allocation. In the legal literature, Kaplow (1995) and Knoll and Griffith (2003) argue for the indexation of taxes to local wages. Albouy (2009) and Albouy and Seegert (2010) quantitatively analyze the question. Starting from the Rosen-Roback tradeoff between equalizing differences across locations in a partial equilibrium model, they calibrate the model and conclude that any tax other than a lump sum tax is distortionary.

To the best of our knowledge, this list of related work is exhaustive. What sets our work apart from the existing literature is a comprehensive 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. These are the three main features of this paper.

2 The Model

Population. The basic model builds on Eeckhout, Pinheiro, and Schmidheiny (2014). The economy is populated by a continuum of identical workers. The country-wide measure of workers is \mathcal{L} . There are J locations (cities), $j \in \mathcal{J} = \{1, \dots, J\}$. 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 US households spend roughly the same fraction of their income on housing of their income level. 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. 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 value

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. 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, 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

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

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. For most of the empirical exercise, we will use a production technology that is linear, i.e., $\gamma = 1$. This is in line with Duranton and Puga (2004) and Combes, Duranton, and Gobillon (2013).

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

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

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.

This 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 will be increasing in the population density. Albouy and Ehrlich (2012) provide evidence that the share of land in housing prices is sharply increasing in city size, ranging from 11% to 48%.

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 equal to

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

where the per-worker value of land across all cities is given by $\frac{\sum_j r_j T_j}{\sum_j l_j}$. The main reason for introducing this ownership structure is to capture the fact that housing ownership is not perfectly diversified. As we will see below, the details of the ownership structure are important for the results. 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.

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 can 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.⁴ To that purpose, we assume that

⁴Of 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%.

the progressive tax schedule can be represented by a two-parameter family 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). Heathcote, Storesletten, and Violante (2013) use the same function to study optimal progressivity of income taxation in the U.S. The average tax rate is given by $\lambda w_j^{-\tau}$ and the marginal tax rate is $\lambda(1-\tau)w_j$. 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 $TR = \frac{\phi G}{T}$.

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 also the rental price of housing capital is therefore 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 Allocation

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

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

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 $h_j = \alpha \frac{(\tilde{w}_j + R_j + TR)}{p_j}$. The indirect utility for a worker is

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

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: $w_j = A_j \gamma l_j^{\gamma-1}$.

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

Given housing supply, and taking the tax schedule as given, the optimal consumption decision will determine the demand for housing. Market clearing then pins down the equilibrium housing prices p_j . This is summarized in the following Proposition.

Proposition 1 *Given amenities a_j , TFP levels A_j , and taxes t_j , the equilibrium populations l_j , allocations c_j, h_j, H_j and prices \tilde{w}_j, p_j, r_j are fully determined by:*

$$\begin{aligned} \frac{a_j}{a_{1j}} &= \frac{l_1^\delta (\tilde{w}_1 + R_1 + TR) ((\tilde{w}_1 + R_1 + TR) l_1)^{-\alpha} H_1^\alpha}{l_j^\delta (\tilde{w}_j + R_j + TR) ((\tilde{w}_j + R_j + TR) l_j)^{-\alpha} H_j^\alpha} \\ c_j &= (1 - \alpha)(\tilde{w}_j + R_j + TR) \quad \text{and} \quad h_j = \alpha \frac{(\tilde{w}_j + R_j + TR)}{p_j} \\ H_j &= B \left[(1 - \beta) \left(\frac{1 - \beta}{\beta} r_j \right)^{\frac{\rho}{1-\rho}} + \beta \right]^{1/\rho} T_j \\ \tilde{w}_j &= (1 - t_j) A_j l_j \\ p_j &= r_j \frac{\left(1 + \left(\frac{1-\beta}{\beta} \right)^{\frac{1}{1-\rho}} r_j^{\frac{\rho}{1-\rho}} \right)}{B \left[(1 - \beta) \left(\frac{1-\beta}{\beta} r_j \right)^{\frac{\rho}{1-\rho}} + \beta \right]^{1/\rho}} \\ r_j &= \frac{\alpha l_j (\tilde{w}_j + R_j + TR)}{T_j} \left(1 + \left(\frac{1 - \beta}{\beta} \right)^{\frac{1}{1-\rho}} r_j^{\frac{\rho}{1-\rho}} \right)^{-1} \end{aligned}$$

for all j together with $\sum_{j=1}^J l_j = \mathcal{L}$, $R_j = (1 - \psi) \frac{\sum_j r_j T_j}{\sum_j l_j}$, $TR = \frac{\phi G}{\mathcal{L}}$, and $\sum_j t_j w_j = G$.

Proof. In Appendix. ■

This is a system of non-linear equations that we will solve and estimate computationally. Now we turn to the optimal policy by the planner.

4 The Planner's Problem

We distinguish between an Optimal Ramsey taxation problem where the planner chooses *tax instruments* in order to affect the equilibrium allocation, and one in which the *full allocation* is chosen by the planner. In the former type of exercise, the planner assumes agents operate in a decentralized economy with equilibrium prices and free choice of consumption and location decisions, albeit affected by taxes. In the latter, there are no prices and the planner directly makes allocation and location decisions.

We focus on the problem in which, rather than choosing allocations and telling workers in which city to live, the planner can use tax schedules to affect the equilibrium allocation. Agents freely choose consumption bundles and decide where to locate, but the planner can affect their labor earnings with a city specific tax t_j where $\tilde{w}_j = (1 - t_j)w_j$. Consider 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. Analytically, we can only explicitly analyze a simple economy with two cities, no government spending, a degenerate wealth distribution ($\psi = 0$) and one specific type of preferences. This gives us the following equivalence result:⁵

Proposition 2 *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:

⁵We are grateful to John Kennan for pointing us to this equivalence.

1. as government expenditure G increases, relative taxes in big cities decrease (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 cities: 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 .

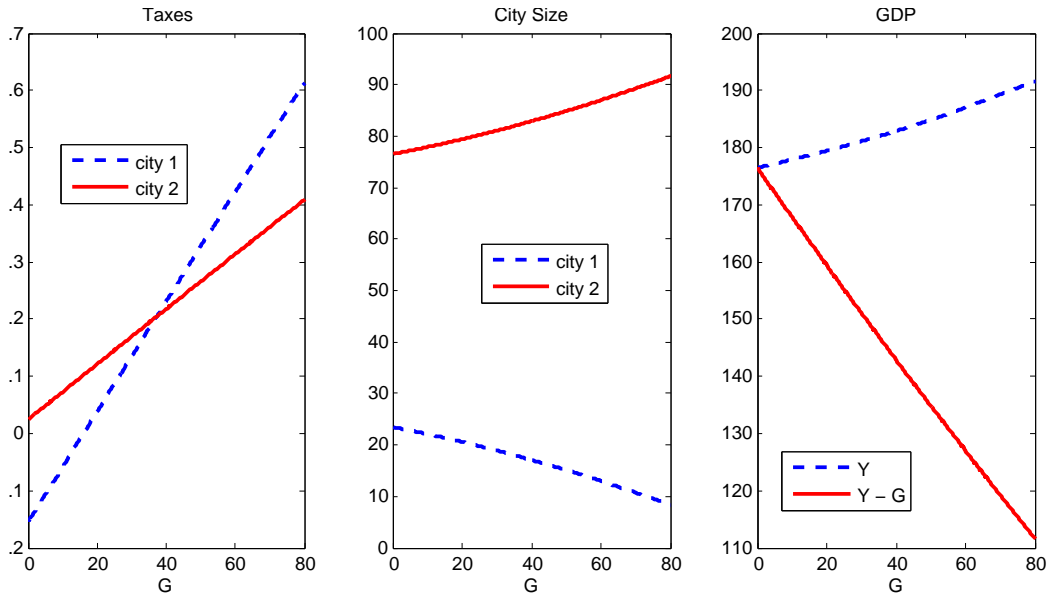


Figure 1: Optimal Ramsey taxes given G in a two city example with a fraction ψ of housing wealth concentration ($\psi = 0.35$): $A_1 = 1, A_2 = 2, \mathcal{L} = 100, \alpha = 0.31$: A. Optimal tax rates t_1, t_2 ; B. populations l_1, l_2 ; C. Output Y and output net of government expenditure $Y - G$.

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. A 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 2.⁶

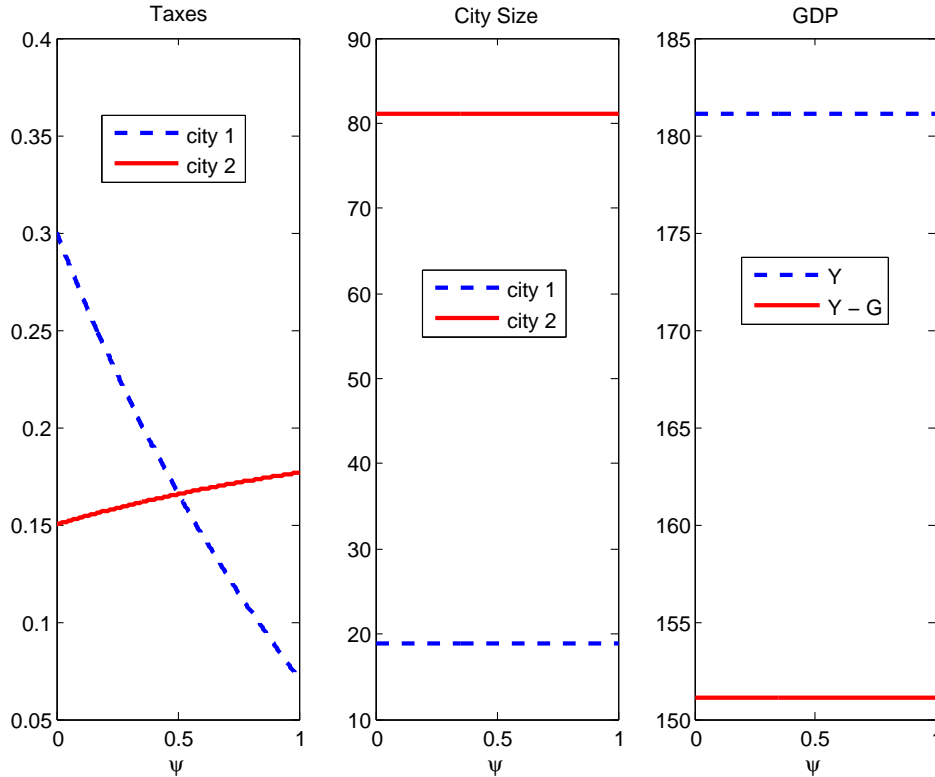


Figure 2: Optimal Ramsey taxes given ψ in a two city example given government expenditure ($G = 30$, 16% of total output): $A_1 = 1, A_2 = 2, \mathcal{L} = 100, \alpha = 0.31$: A. Optimal tax rates t_1, t_2 ; B. populations l_1, l_2 ; C. Output Y and output net of government expenditure $Y - G$.

One could ask what the optimal solution is when the planner is not constrained by mobility of workers. This implies that she can assign workers to cities even if the utility obtained in different cities is not equalized. We analyze this case in detail in the Appendix. What transpires from this is that the unconstrained planner wants to locate a lot of agents in the big cities. There they are very productive, but given housing constraints, they consume little housing and will as a result have a low marginal utility. The planner therefore assigns a lot of consumption to the few workers in the small, unproductive city. There they have a lot of housing and a high marginal utility. This planner’s solution has big ex post inequality in utility.

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

⁶In Figure 1, we set $\psi = 0.35$, 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.

out the productivity parameters A_j . Second, given (l_j, w_j) , a representation of current US taxes on labor income, $(\lambda^{US}, \tau^{US})$, 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.⁷ Figure 3.A and B show the distribution of population and wages across MSAs. The average labor force is 436,613, with a maximum (New York-Northern New Jersey-Long Island) of more than 9.2 million and a minimum (Yuma, AZ) of about 87,707. The population distribution is highly skewed, close to log-normal, where the top 5 MSAs account for 22.3% of total labor force. Average weekly wages is \$605. The highest weekly wage is more than twice as high as the mean level (Stamford, CT) and the lowest is 75% of the mean level (Brownsville-Harlingen-San Benito, TX). Figure 3.C shows the positive relation between population size and wages, well-known urban wage premium in the data. We take both population and wage data as inputs to simulate the benchmark economy. The elasticity of wages with respect to population size is about 0.08.

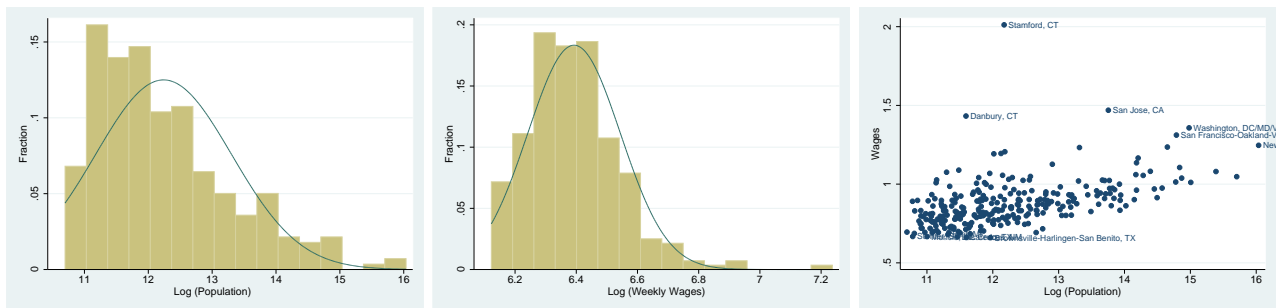


Figure 3: A. Histogram and Kernel density of labor force; B. Histogram and Kernel density of wages; C. Urban Wage premium.

⁷We remove wages that are larger than 5 times the 99th percentile threshold and less than half of the 1st percentile threshold.

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 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.⁸ 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 (2013) 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 closer to the ones provided by Guvenen, Burhan, and Ozkan (2013) 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.⁹ 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 .¹⁰

⁸ <http://www.oecd.org/social/soc/benefitsandwage-tax-benefit-calculator.htm>, accessed on March 15, 2013.

⁹ 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

¹⁰ Guvenen, Burhan, and Ozkan (2013) estimate a progressivity wedge of 0.15. Given the particular tax function we are using, the progressivity only depends on τ .

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 Flint, MI.

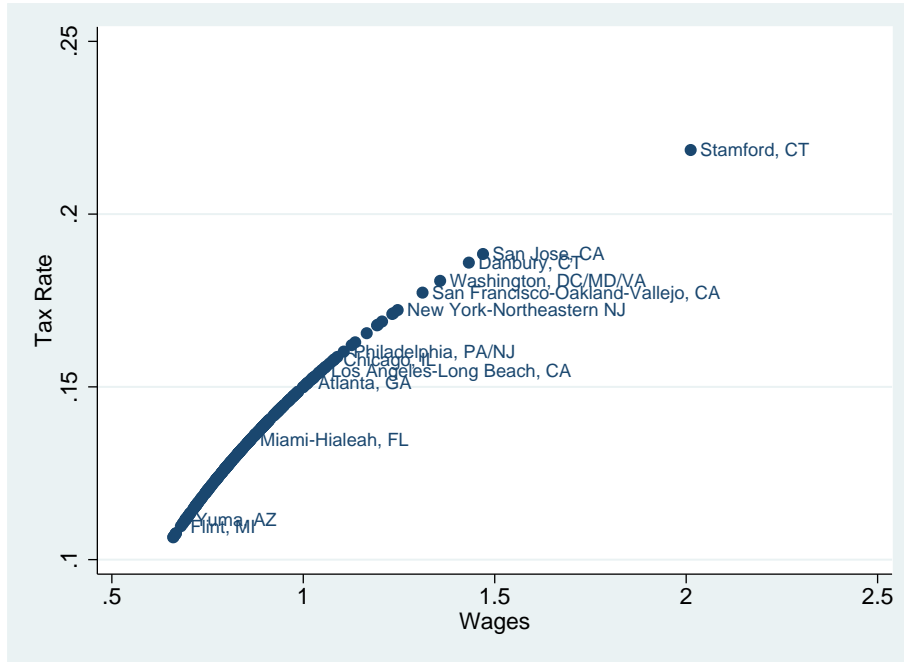


Figure 4: Taxes across cities

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. $TR = 0.82\frac{G}{Y}$.¹¹

5.3 Housing Production

The data on land areas of cities (MSAs), T_j , is taken from the Census Bureau.¹² Average land area of MSAs is about 5254 km² and there is very large variation in land areas across MSA.¹³ The largest MSA in terms of land areas is huge with 70630 km² (Riverside-San Bernardino, CA) while the smallest one has an area of only 312 km² (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 200m².

¹¹National Income and Product Accounts, Bureau of Economic Analysis, Table 3.16. Government Current Expenditures by Function, http://www.bea.gov/iTable/index_nipa.cfm

¹²We use data available at <https://www.census.gov/population/www/censusdata/files/90den.ma.txt> and published in U.S.Census.Bureau (2004).

¹³Figure 12 in the Appendix shows the distribution of land across MSAs.

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 2010.¹⁴ Therefore, we set $1 - \psi$ to be 35% (67% of 52.4%).

5.5 Preferences and Productivity

As we mentioned above, we set $\gamma = 1$ and 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,

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

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 $\alpha/\lambda = \frac{0.24}{0.85} = 0.2824$

¹⁴US Census Bureau Table 5. Homeownership Rates for the United States: 1968 to 2014, available at <http://www.census.gov/housing/hvs/data/q314ind.html>

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 $(0.13)(0.2) = 0.026$.¹⁵

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 0.91. The highest levels of a_j , above 1.1, are calculated for Chicago (IL), Los Angeles-Long Beach (CA) and El Paso (TX). 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 El Paso (TX) also requires a high a to justify its size.

The lowest values are below 0.7, for Stamford (CT), Anchorage (AK) and Danbury (CT). These are MSAs with very high wages but 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 0.11, 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 407 per km² in San Francisco-Oakland-Vallejo (CA), followed by Stamford (CT) and Chicago (IL) where housing prices are 377 and 374, respectively. The lowest housing prices are computed for Flagstaff (AZ-UT), 32, and Yuma (AZ), 46. While housing consumption

¹⁵In 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 5.8, incorporating the exact land area in the model is an important ingredient to fit the data.

is about 200m^2 across MSA, those in Chicago live in houses that are about 80m^2 and about 8 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.

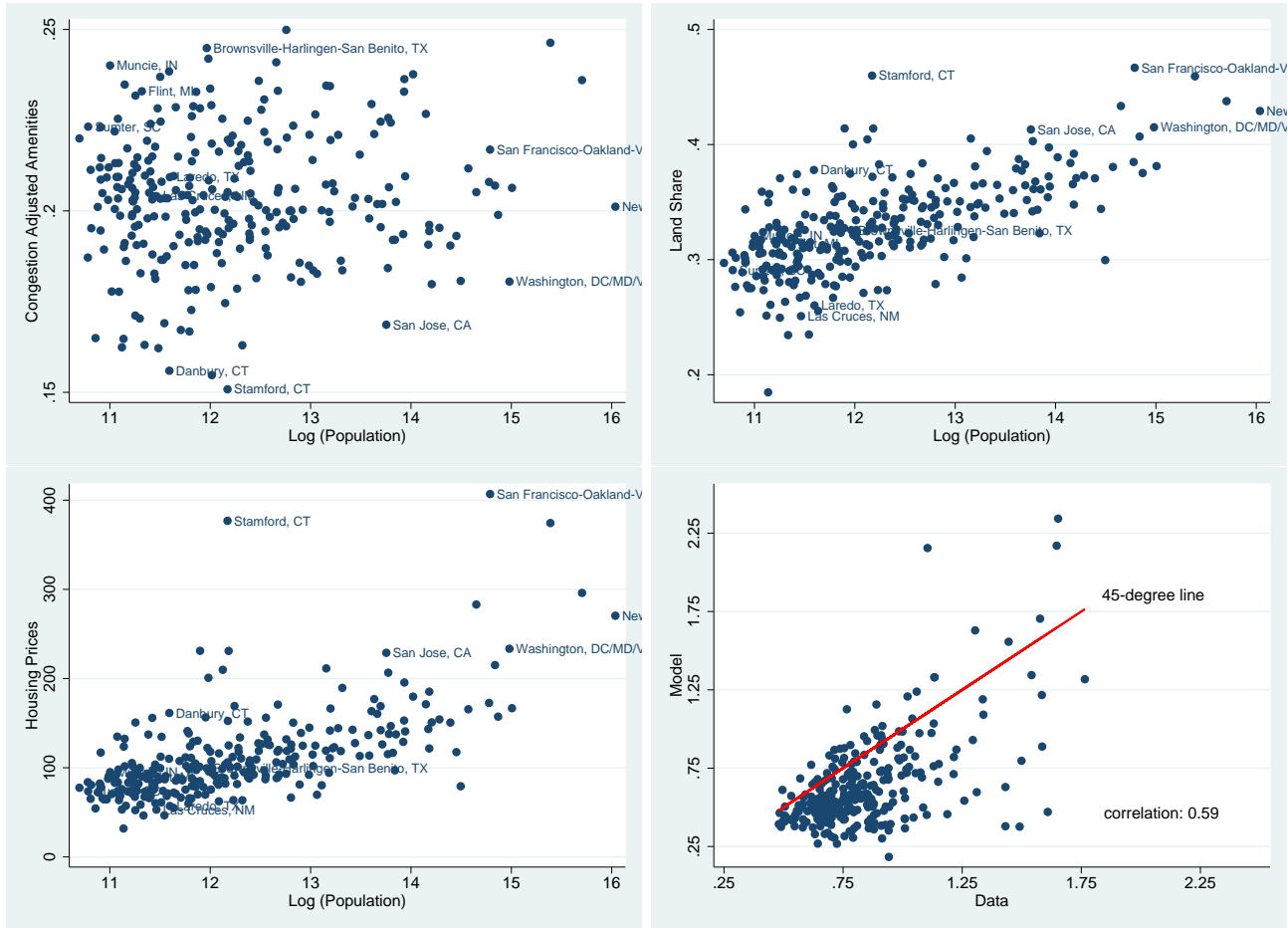


Figure 5: Benchmark Economy. A. Amenities and Population; B. Land Share in the Value of Housing and Population; C. Housing Prices and Population; D. Housing Prices: model versus data.

Finally, Figure 5.D 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 as age number of rooms, age of the unit, and the units structure (one family detached unit vs. one family attached unit

etc.).¹⁶ 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

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

which can be used to calculate new allocations for any τ

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

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 6 shows the percentage change in utility from the benchmark economy for different values of τ . The optimal value τ^* , is 0.0145. The optimal τ^* 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 6 shows the implied tax schedule under $(\lambda^{US}, \tau^{US})$ and (λ^*, τ^*) . 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 (λ^*, τ^*) . As a result, for $w = 0.5$, $w = 2$ and $w = 5$, the tax rates are 14.1%, 15.9% and 17.0%, respectively.

Now we can evaluate the implications of a tax change in the tax schedule from τ^{US} to τ^* , both for

¹⁶We use 2010 American Community Survey (ACS) data on housing rentals and housing characteristics.

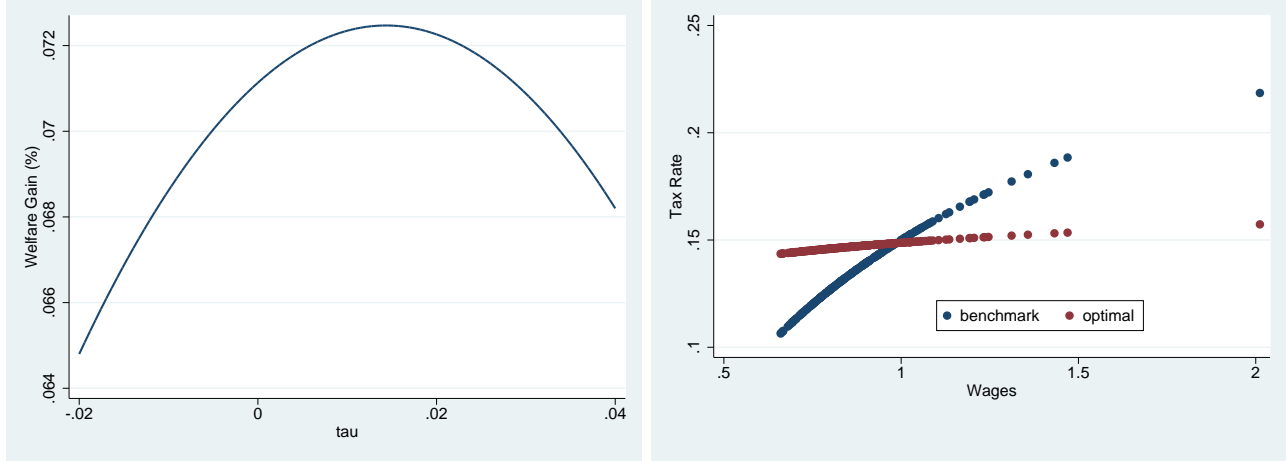


Figure 6: A. Welfare gain for different values of τ ; B. The optimal tax schedule τ^* compared to that in the benchmark economy τ^{US} .

individual cities and in the aggregate. Consider first the impact on individual cities which is summarized in Figure 7 and Table 1. The table gives the numerical values for those cities with extreme values either for TFP A or for amenities a . Cities with 5 highest and lowest values of A are explicitly identified in the scatter plots in Figure 7.¹⁷

Since the optimal degree of tax difference τ^* is below existing τ^{US} , the optimal policy lowers tax payments in high productivity cities. Figure 7.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 Las Cruces (NM) loses 24% of its population. As is apparent in Figure 7.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 7.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 7.D shows the change in housing prices. High productivity cities are up to 23% more expensive while low productivity cities face housing price drops of up to 8%.

Figure 8 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 20%. 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

¹⁷There notable outlier, Stamford CT. All our results are robust if we remove that observation.

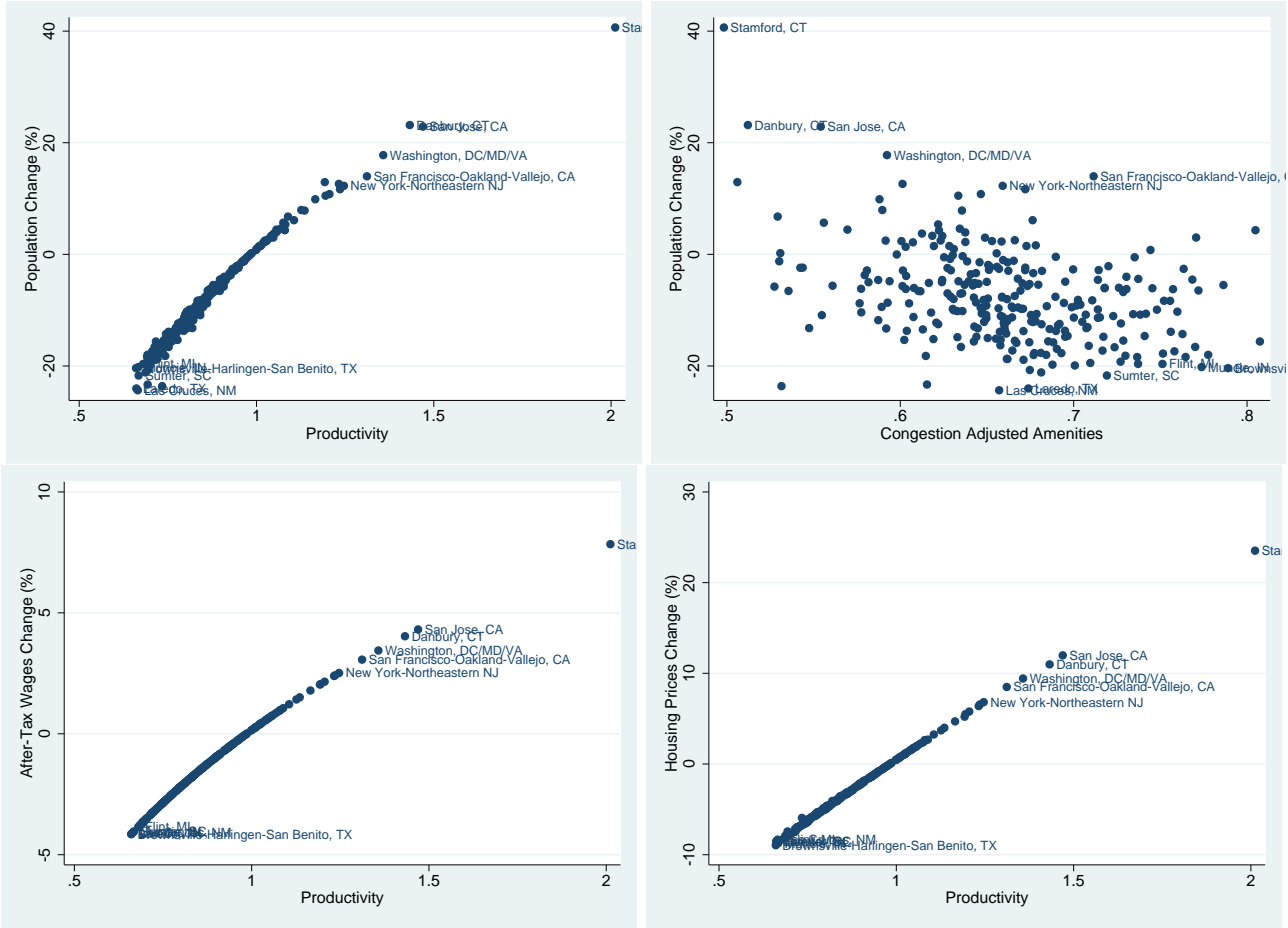


Figure 7: Implied changes of implementing the optimal policy τ^* . A. Change in population by TFP; B. Change in population by a ; C. Change in \tilde{w} by TFP; D. Change in housing prices p by TFP.

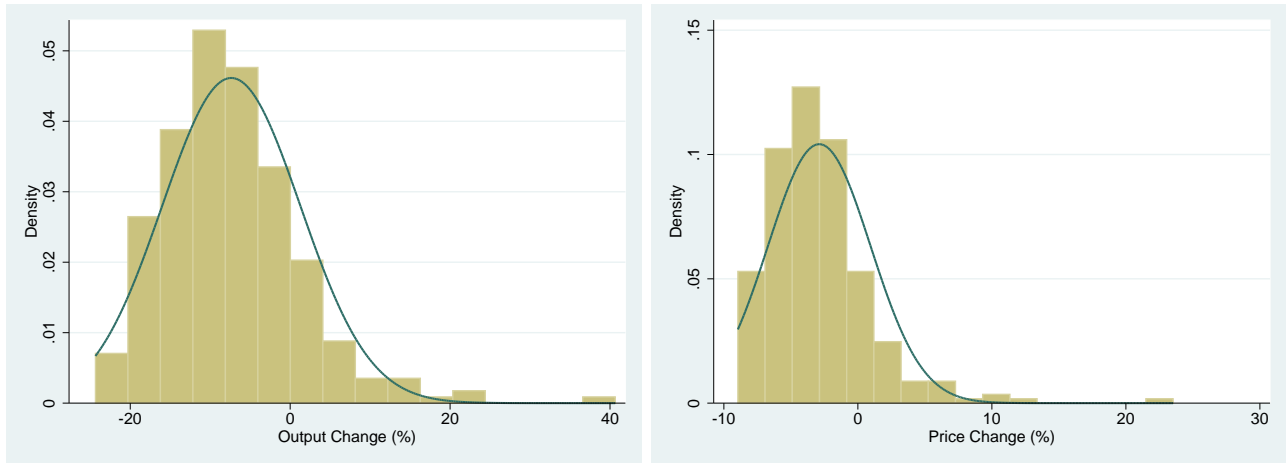


Figure 8: Distribution of changes: A. Output; B. Housing Prices.

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 2%

higher goods consumption. In the less productive cities housing consumption increases by up to 5% at the cost of decreased goods consumption by 2%. Given homothetic preferences, the marginal rate of substitution is constant (see Figure 9.A).

Table 1: Benchmark Economy, move from τ^{USA} to τ^* . Outcomes for Selected Cities.

MSA	A	a	$\% \Delta l$	$\% \Delta p$	$\% \Delta c$	$\% \Delta h$
Highest A						
Stamford, CT	2.01	0.68	40.65	23.52	7.17	-13.34
San Jose, CA	1.47	0.79	22.09	11.98	3.88	-7.24
Danbury, CT	1.43	0.69	23.18	10.99	3.62	-6.64
Lowest A						
Las Cruces, NM	0.67	0.89	-24.35	-8.34	-3.06	5.76
Laredo, TX	0.66	0.91	-24.03	-8.52	-3.11	5.92
Brownsville, TX	0.66	1.08	-20.39	-8.95	-3.12	6.41
Highest a						
Chicago, IL	1.08	1.20	4.33	2.64	0.92	-1.67
Los Angeles-Long Beach, CA	1.05	1.16	3.02	1.76	0.65	-1.10
El Paso, TX	0.72	1.12	-15.62	-7.42	-2.51	5.30
Lowest a						
Danbury, CT	1.43	0.69	23.18	10.95	3.62	-6.64
Anchorage, AK	1.19	0.69	12.94	5.22	1.84	-3.21
Stamford, CT	2.01	0.68	40.65	23.22	7.17	-13.24

Table 2 shows the aggregate outcomes from moving the benchmark allocation to the optimal. On average output and consumption go up by about 1.57% and 1.53%, respectively. This is driven by the population moving to the more productive cities. The population in the 5 largest cities grows by 7.95%, 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 9.B. Not surprisingly, aggregate housing prices go up by 5.32%. Due to higher prices, aggregate housing consumption declines by 1.75%.

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 0.073%, 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 con-

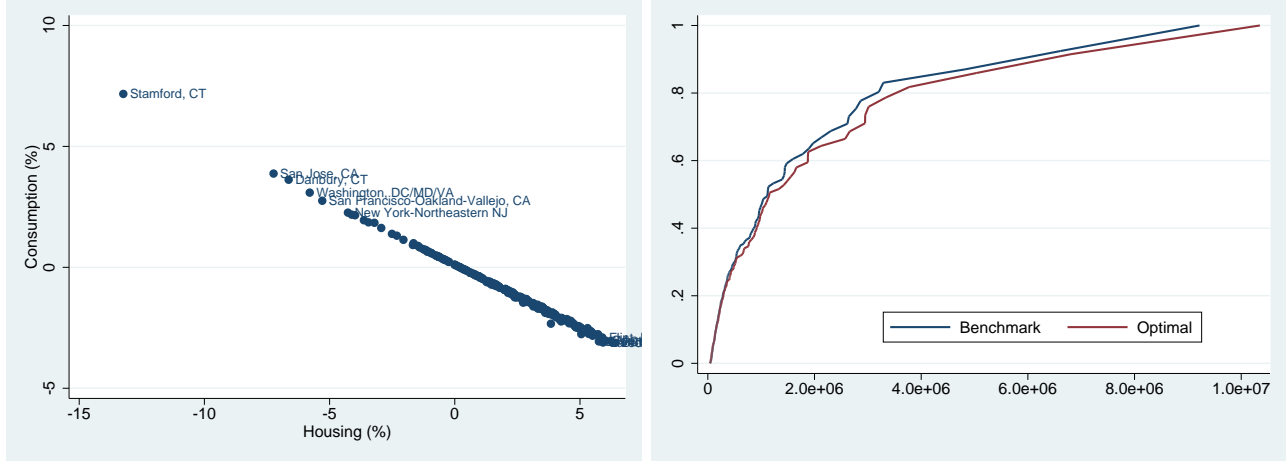


Figure 9: Implied changes of implementing the optimal policy τ^* : A. Substitution between c and h ; B. Cumulative Distribution of city sizes.

sumption 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 τ^*

Outcomes	
Welfare gain (%)	0.0725
Output gain (%)	1.57
Consumption (%)	1.53
Housing Consumption (%)	-1.75
Population change top 5 cities (%)	7.95
Fraction of Population that Moves (%)	3.47
Change in average prices (%)	5.32

5.8 Sensitivity Analysis

In this section, we discuss how sensitive our results are to different aspects of our modeling and calibration choices. First, we focus on λ and τ and show how different values of λ and τ for the benchmark economy affect the optimal level of tax difference. Next, we consider a benchmark economy with different levels of ψ , the concentration of land ownership. We also show the results for an economy in which all cities have an equal amount of land. Finally, we discuss what happens if we change the amount of tax revenue that is transferred back to individuals, TR .

5.8.1 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 are 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%.¹⁸

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 progressivity τ^* , 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.

Table 3: Robustness.

	τ	$\lambda = 0.9$			$\lambda = 0.85$			$\lambda = 0.815$		
		0.053	0.12	0.2	0.053	0.12	0.2	0.053	0.12	0.2
Optimal τ^*		0.0606	0.0646	0.0697	0.0095	0.0145	0.0189	-0.0212	-0.0167	-0.0116
Output gain (%)		-0.13	0.93	2.21	0.64	1.57	2.72	1.02	1.89	2.96
Consumption (%)		-0.12	0.90	2.13	0.62	1.53	2.65	0.99	1.85	2.88
Housing Consumption (%)		0.14	-1.02	-2.48	-0.70	-1.75	-3.10	-1.12	-2.12	-3.39
Pop top 5 (%)		-0.64	4.76	11.21	3.25	7.95	13.69	5.13	9.52	6.41
Pop moves (%)		0.28	2.08	4.88	1.42	3.47	5.95	2.24	4.14	10.48
Avg. prices (%)		-0.40	3.07	7.58	2.10	5.32	9.56	3.38	6.49	10.48
Welfare gain (%)		0.00005	0.0237	0.1315	0.0119	0.0725	0.2105	0.0319	0.1086	0.2585

The most important thing to observe is that as government spending increases (λ decreases), then τ^* declines and taxes becomes relatively lower in bigger cities. This is consistent with the findings in the two city model above. Observe also the 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

¹⁸Source: National Income and Product Accounts (NIPA) Table 3.2. - Federal Government Current Receipts and Expenditures.

to 14% 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.

5.8.2 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 $\tau = -0.1394$ 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 3.7%, the population in the top 5 cities grows by 18% and average housing prices increase by 13%. While the welfare effects are substantially higher, they are still relative small.

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

Outcomes	Benchmark $\psi = 0.35$	All Bond $\psi = 0$	All Landlords $\psi = 1$
Optimal τ	0.0145	-0.1394	0.0783
Output gain (%)	1.57	3.60	0.65
Consumption (%)	1.53	3.69	0.61
Housing Consumption (%)	-1.75	-4.04	-0.71
Pop. change top 5 cities (%)	7.95	18.24	3.31
Frac. of Pop. that Moves (%)	3.47	7.65	1.45
Change in average prices (%)	5.32	13.09	2.11
Welfare gain (%)	0.0725	0.3445	0.0130

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 positive. For our economy, it is equal to $\tau = 0.0783$, which is still lower than than τ^{US} . The effects on output and migration is now much more muted.

5.8.3 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 . We first recalibrate such an economy using exactly the same targets that we used above and then look for the level of τ that maximize welfare.¹⁹ The economy with identical land areas looks very similar to the benchmark economy with different land areas, with one important exception. When we force all cities to have the same land area, amenities play a more important role. Figure 10 shows a scatter plot of amenities in two economies. The range of values that amenities take are larger in an economy with equal land areas.

The optimal level of τ is 0.0303 in 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 5 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.

Table 5: Fixed Land Size at 5000km^2 .

Outcomes	Benchmark	Fixed Land Area
Optimal τ	0.0145	0.0303
Output gain (%)	1.57	1.30
Consumption (%)	1.53	1.25
Housing Consumption (%)	-1.75	-1.92
Population change top 5 cities (%)	7.95	6.00
Fraction Population that Moves (%)	3.47	2.77
Change in average prices (%)	5.32	5.34
Welfare gain (%)	0.0725	0.0485

5.8.4 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.

¹⁹Recalibrating an economy with identical-sized cities requires rather minor changes in the parameters. In particular, we only change two parameters: $B = 0.033$ and $\beta = 0.25$.

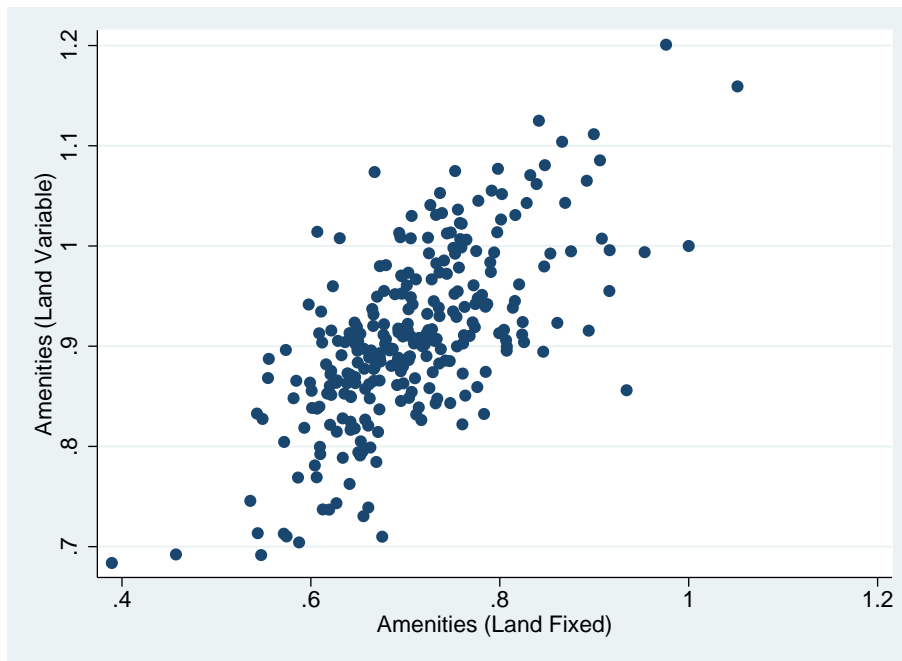


Figure 10: Estimated amenities with fixed land versus the benchmark (variable land).

Table 6: No Rebate of Tax Revenue, $\phi = 0$

Outcomes	Benchmark	No Tax Rebate
Optimal τ	0.0145	0.0073
Output gain (%)	1.57	1.95
Output gain (%)	1.53	2.16
Output gain (%)	-1.75	-2.08
Population change top 5 cities (%)	7.95	9.83
Fraction population that moves (%)	3.47	4.30
Change in average prices (%)	5.32	6.69
Welfare gain (%)	0.0725	0.109

6 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. In contrast, the existing literature has partial equilibrium models and can not evaluate optimal tax policy.

Our findings are first, that the optimal spatial tax schedule is not flat and depends on the level of government spending and on the concentration of housing wealth. 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 current existing schedule. 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 1.57%. 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.

Appendix

Proof of Proposition 1

Proof. The First Order Conditions for the housing production are,

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

and,

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

which implies

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

and

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

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

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

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:

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

Observe that this expression consist of one equation in one unknown, r_j , though there is no explicit solution. Given the (numerical) 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

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

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

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

We can use the condition for optimal production $w_j = A_j l_j$ and the fact that $\tilde{w}_j = (1 - t_j)w_j$. In equilibrium, individuals are assumed to own land in proportion to their consumption of housing. Therefore R_j satisfies:

$$R_j = (1 - \psi) \frac{\sum_j r_j T_j}{\sum_j l_j}. \quad (10)$$

Finally, the population allocation must satisfy feasibility: $\sum_j l_j = \mathcal{L}$. ■

Proof of Proposition 2

Proof. I. Decentralized Equilibrium. Each consumer in city $i \in \{1, 2\}$ optimizes his utility subject to a budget constraint:

$$\begin{aligned} \max_{c_i, h_i} \quad & c_i h_i \\ \text{s.t.} \quad & c_i + p_i h_i = w_i + R, \end{aligned}$$

where $R = p_1 + p_2$. The market clearing conditions are

$$\begin{aligned} \frac{c_1}{x} &= \frac{c_2}{1-x} \\ h_1 &= \frac{1}{x} \\ h_2 &= \frac{1}{1-x} \\ R &= p_1 + p_2 \end{aligned}$$

The entire system then is

$$\begin{aligned}\frac{c_1}{x} &= \frac{c_2}{1-x} \\ c_1 &= \frac{w_1 + R}{2} \\ c_2 &= \frac{w_2 + R}{2} \\ \frac{1}{x} &= \frac{w_1 + R}{2p_1} \\ \frac{1}{1-x} &= \frac{w_2 + R}{2p_2} \\ R &= p_1 + p_2\end{aligned}$$

We can rewrite $R = p_1 + p_2 = xw_1 + (1-x)w_2$ so that

$$\begin{aligned}\frac{w_1 + R}{x} &= \frac{w_2 + R}{1-x} \\ R &= xw_1 + (1-x)w_2\end{aligned}$$

or

$$\frac{w_1 + xw_1 + (1-x)w_2}{x} = \frac{w_2 + xw_1 + (1-x)w_2}{1-x}$$

or

$$2(w_2 - w_1)x^2 - 4w_2x + (w_1 + w_2) = 0$$

We have two explicit solutions for x :

$$\begin{aligned}x^* &= \frac{2w_2 + \sqrt{2(w_1^2 + w_2^2)}}{2(w_2 - w_1)} \\ x^* &= \frac{2w_2 - \sqrt{2(w_1^2 + w_2^2)}}{2(w_2 - w_1)}\end{aligned}$$

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

that:

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

or

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

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

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

or

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

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

$$2(w_2 - w_1)x^2 - 4w_2x + (w_1 + w_2) = 0$$

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:

$$\begin{aligned} \frac{(w_1 - w_2)(x^2 + (1 - x)^2) - (x w_1 + (1 - x) w_2)(2x - 2(1 - x))}{(x^2 + (1 - x)^2)^2} &= 0 \\ (w_1 - w_2)(x^2 + (1 - x)^2) - 2(x w_1 + (1 - x) w_2)(2x - 1) &= 0 \\ (w_1 - w_2)(2x^2 + 1 - 2x) + 2(x w_1 + (1 - x) w_2) - (4x^2 w_1 + (4x - 4x^2) w_2) &= 0 \\ 2(w_2 - w_1)x^2 - 4w_2x + (w_1 + w_2) &= 0 \end{aligned}$$

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

The Unconstrained Optimal Allocation

The Ramsey planner chooses policies that are subject to market forces, equilibrium prices and mobility of workers. As a result, her program is a constrained optimization problem. Now we consider an unconstrained planner who chooses populations across cities and hence production, and also consumption.

She cannot of course unbundle housing consumption from production, but she can choose consumption independent of utility equalization.

Formally, the planner chooses the bundles l_j, c_j, h_j in all cities j as well as housing capital and land K_j, T_j to maximize Utilitarian welfare:

$$\begin{aligned}
& \max_{l_j, c_j, h_j, K_j, T_j} \sum_{i,j} a_j l_j^{-\delta} c_j^{1-\alpha} h_j^\alpha l_j & (11) \\
& \text{s.t.} \quad \sum_j c_j l_j + \sum_j K_j + (1 - \phi)G = \sum_j A_j l_j \\
& \quad H_j = B \left[(1 - \beta)K_j^\rho + \beta T_j^\rho \right]^{1/\rho}, \forall j \\
& \quad \sum_i h_j l_j = H_j, \forall j \\
& \quad T_j = T, \forall j \\
& \quad \sum_j l_j = \mathcal{L}.
\end{aligned}$$

This is a system of $3 \times J + 2J + 2 + 3 \times J$ equations in the same number of variables. Again, there is no explicit solution, so we solve it numerically to get quantitatively relevant predictions.

Proposition 3 *Consider a simple representative agent economy with $\beta = 1$, $\delta = 0$, $\phi = 0$, $a_j = 1$, and $T_j = T$. If $A_1 < A_2$, then the unconstrained optimal allocation satisfies $l_1 < l_2$, $c_1 > c_2$, and $u_1 > u_2$. There is no utility equalization in equilibrium.*

Proof. See below. ■

The result for the unconstrained planner is illustrated in Figure 11. The planner chooses production optimality, and therefore equates marginal product across cities. This inevitably entails locating a lot of the workers in city 2, the high productivity city as illustrated in panel A. Doing that, the city size distribution is not affected by government spending G . Panel C plots output which is constant but of course, net of government spending it is decreasing in G . The consumption that the planner assigns to the citizens does vary differentially across cities as shown in panel B. The higher G , the faster the decline in consumption in city 1 relative to city 2. As less output is available with higher G and production efficiency is not affected, the consumption allocation is purely based on marginal utility. For lower disposable income levels, marginal consumption in the large city is relatively larger. This also helps explain why in the optimal Ramsey problem the tax difference between large and small cities decreases with higher G .

Further intuition behind this result can best be obtained by considering a limit case. When productivity is constant and independent of the city population, then from a production efficiency viewpoint, production in the high productivity city is always superior to production in the low productivity city

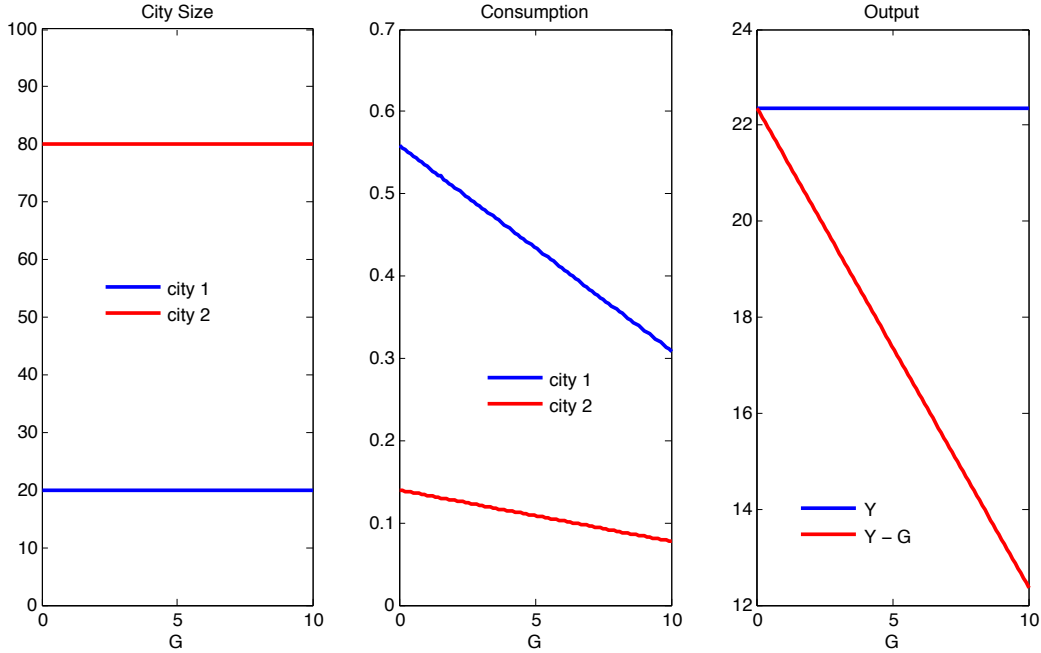


Figure 11: Optimal Taxes for the Unconstrained Planner given G in a two city example: $A_1 = 1, A_2 = 2, \mathcal{L} = 100, \alpha = 0.31, \gamma = 0.5$: A. Populations l_1, l_2 ; B. consumption c_1, c_2 ; C. Output Y and output net of government spending $Y - G$.

and marginal products are never equalized. The following corollary characterizes the planner's solution in that case.

Corollary 1 *Let $A_1 < A_2$. If γ converges to 1, then all production is concentrated in city 2 by nearly all the population, and a minimal fraction of workers gets to consume all the output in city 1.*

The corollary illustrates that the equity implications of the planner's solution are extreme. A minority vanishing in size consumes very large per capita consumption in the unproductive city. The output is generated by the majority in the productive city. All output is generated in the productive city in line with the Diamond and Mirrlees (1971a) and Diamond and Mirrlees (1971b) results. It is optimal not to distort productive efficiency, and as a result, the marginal product of output across cities should be equated. Since the marginal product converges to a constant as γ converges to one, it is optimal to produce all output in the high TFP city. At the same time, those workers are given zero consumption because due to housing supply, their marginal utility of one unit of consumption is lower than that of the those living in small cities. As a result and because the utility is homogeneous of degree one, the planner optimally assigns all consumption and utility to the few in the unproductive city. Observe that when $\gamma \neq 1$, consumption is not independent of the production side, thus violating the premise of Diamond and Mirrlees (1971a). As a result, optimal taxation involves equating marginal productivity across cities.

Proof of Proposition 3

Proof. The planner's problem is:

$$\begin{aligned}
 \max_{c_j, h_j, l_j} \quad & \sum_j c_j^{1-\alpha} h_j^\alpha l_j \\
 \text{s.t.} \quad & \sum_j c_j l_j + G = \sum_j A_j l_j^\gamma \\
 & h_j l_j = BL, \forall j \\
 & \sum_j l_j = \mathcal{T}.
 \end{aligned}$$

For the two city case, the FOCs are (with Langrangian multipliers $\phi, \lambda_1, \lambda_2, \psi$):

$$\begin{aligned}
 (1 - \alpha)c_1^{-\alpha} h_1^\alpha l_1 &= \phi l_1 \\
 (1 - \alpha)c_2^{-\alpha} h_2^\alpha l_2 &= \phi l_2 \\
 \alpha c_1^{1-\alpha} h_1^{\alpha-1} l_1 &= \lambda_1 l_1 \\
 \alpha c_2^{1-\alpha} h_2^{\alpha-1} l_2 &= \lambda_2 l_2 \\
 c_1^{1-\alpha} h_1^\alpha &= \phi(c_1 - A_1 \gamma l_1^{\gamma-1}) + \lambda_1 h_1 + \psi \\
 c_2^{1-\alpha} h_2^\alpha &= \phi(c_2 - A_2 \gamma l_2^{\gamma-1}) + \lambda_2 h_2 + \psi
 \end{aligned}$$

Or

$$\begin{aligned}
 (1 - \alpha)c_1^{-\alpha} h_1^\alpha &= \phi \\
 (1 - \alpha)c_2^{-\alpha} h_2^\alpha &= \phi \\
 \alpha c_1^{1-\alpha} h_1^{\alpha-1} &= \lambda_1 \\
 \alpha c_2^{1-\alpha} h_2^{\alpha-1} &= \lambda_2 \\
 c_1^{1-\alpha} h_1^\alpha &= \phi(c_1 - A_1 \gamma l_1^{\gamma-1}) + \lambda_1 h_1 + \psi \\
 c_2^{1-\alpha} h_2^\alpha &= \phi(c_2 - A_2 \gamma l_2^{\gamma-1}) + \lambda_2 h_2 + \psi
 \end{aligned}$$

or

$$\begin{aligned}
\frac{h_1}{c_1} &= \left(\frac{\phi}{1-\alpha} \right)^{\frac{1}{\alpha}} \\
\frac{h_2}{c_2} &= \left(\frac{\phi}{1-\alpha} \right)^{\frac{1}{\alpha}} \\
\frac{h_1}{c_1} &= \left(\frac{\lambda_1}{\alpha} \right)^{\frac{1}{\alpha-1}} \\
\frac{h_2}{c_2} &= \left(\frac{\lambda_2}{\alpha} \right)^{\frac{1}{\alpha-1}} \\
c_1 \frac{\phi}{1-\alpha} &= \phi(c_1 - A_1 \gamma l_1^{\gamma-1}) + \lambda_1 h_1 + \psi \\
c_2 \frac{\phi}{1-\alpha} &= \phi(c_2 - A_2 \gamma l_2^{\gamma-1}) + \lambda_2 h_2 + \psi
\end{aligned}$$

or

$$\begin{aligned}
\frac{h_1}{c_1} &= \left(\frac{\phi}{1-\alpha} \right)^{\frac{1}{\alpha}} \\
\frac{h_2}{c_2} &= \left(\frac{\phi}{1-\alpha} \right)^{\frac{1}{\alpha}} \\
\frac{h_1}{c_1} &= \left(\frac{\lambda_1}{\alpha} \right)^{\frac{1}{\alpha-1}} \\
\frac{h_2}{c_2} &= \left(\frac{\lambda_2}{\alpha} \right)^{\frac{1}{\alpha-1}} \\
\frac{\alpha}{1-\alpha} \phi c_1 &= -\phi A_1 \gamma l_1^{\gamma-1} + \lambda_1 h_1 + \psi \\
\frac{\alpha}{1-\alpha} \phi c_2 &= -\phi A_2 \gamma l_2^{\gamma-1} + \lambda_2 h_2 + \psi
\end{aligned}$$

From the first and second equations we obtain $c_1 h_2 = c_2 h_1$. From the first and third equations we obtain $h_1 = \frac{\alpha}{1-\alpha} c_1 \frac{\phi}{\lambda_1}$ and the second and the fourth, $h_2 = \frac{\alpha}{1-\alpha} c_2 \frac{\phi}{\lambda_2}$ or $\frac{\lambda_j}{\phi} = \frac{\alpha}{1-\alpha} \frac{c_j}{h_j}$. The last two equations can be written as:

$$\begin{aligned}
\frac{\alpha}{1-\alpha} c_1 + A_1 \gamma l_1^{\gamma-1} - \frac{\alpha}{1-\alpha} c_1 - \frac{\psi}{\phi} &= 0 \\
\frac{\alpha}{1-\alpha} c_2 + A_2 \gamma l_2^{\gamma-1} - \frac{\alpha}{1-\alpha} c_2 - \frac{\psi}{\phi} &= 0
\end{aligned}$$

or

$$A_1 l_1^{\gamma-1} = A_2 (\mathcal{T} - l_1)^{\gamma-1}$$

or

$$l_1 = \left(\frac{A_2}{A_1} \right)^{\frac{1}{\gamma-1}} (\mathcal{T} - l_1)$$

or

$$l_1 = \left(\frac{A_1}{A_2} \right)^{\frac{1}{1-\gamma}} (\mathcal{T} - l_1)$$

or

$$l_1 = \frac{A_1^{\frac{1}{1-\gamma}}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \mathcal{T}$$

Now we can finalize the whole equilibrium allocation. From feasibility in the housing market, we know that:

$$\begin{aligned} h_1 &= \frac{BL}{\mathcal{T}} \frac{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}}{A_1^{\frac{1}{1-\gamma}}} \\ h_2 &= \frac{BL}{\mathcal{T}} \frac{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}}{A_2^{\frac{1}{1-\gamma}}}. \end{aligned}$$

From the fact that $c_1 h_2 = c_2 h_1$, we get

$$c_2 = c_1 \left(\frac{A_1}{A_2} \right)^{\frac{1}{1-\gamma}},$$

and using the aggregate budget constraint

$$\begin{aligned} c_1 &= \frac{A_1 l_1^\gamma + A_2 l_2^\gamma - G}{l_1 + \left(\frac{A_1}{A_2} \right)^{\frac{1}{1-\gamma}} l_2} \\ &= \frac{A_1^{\frac{1}{1-\gamma}} \left(\frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \right)^\gamma + A_2^{\frac{1}{1-\gamma}} \left(\frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \right)^\gamma - G \left(A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}} \right)^\gamma}{A_1^{\frac{1}{1-\gamma}} \frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} + \left(\frac{A_1}{A_2} \right)^{\frac{1}{1-\gamma}} A_2^{\frac{1}{1-\gamma}} \frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}}} \\ &= \frac{A_1^{\frac{1}{1-\gamma}} \left(\frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \right)^\gamma + A_2^{\frac{1}{1-\gamma}} \left(\frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \right)^\gamma - G}{\frac{A_1^{\frac{1}{1-\gamma}} \mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}}}, \end{aligned}$$

and

$$c_2 = \frac{A_1^{\frac{1}{1-\gamma}} \left(\frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \right)^\gamma + A_2^{\frac{1}{1-\gamma}} \left(\frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \right)^\gamma - G}{\frac{A_2^{\frac{1}{1-\gamma}} \mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}}}.$$

The equilibrium utility levels satisfy:

$$\begin{aligned} u_1 &= \left(A_1^{\frac{1}{1-\gamma}} \left(\frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \right)^\gamma + A_2^{\frac{1}{1-\gamma}} \left(\frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \right)^\gamma - G \right)^{1-\alpha} \left(\frac{BL}{\mathcal{T}} \right)^\alpha \left(A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}} \right) \frac{1}{A_1^{\frac{1}{1-\gamma}}} \\ u_2 &= \left(A_1^{\frac{1}{1-\gamma}} \left(\frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \right)^\gamma + A_2^{\frac{1}{1-\gamma}} \left(\frac{\mathcal{T}}{A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}}} \right)^\gamma - G \right)^{1-\alpha} \left(\frac{BL}{\mathcal{T}} \right)^\alpha \left(A_1^{\frac{1}{1-\gamma}} + A_2^{\frac{1}{1-\gamma}} \right) \frac{1}{A_2^{\frac{1}{1-\gamma}}}. \end{aligned}$$

Clearly, given $A_1 \neq A_2$, this implies that $u_1 \neq u_2$. If $A_1 < A_2$ then $l_1 < l_2, c_1 > c_2, u_1 > u_2$. Moreover, the more productive city is larger under the unconstrained planner's problem than under the Optimal Ramsey Taxation problem. ■

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 (2013). 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.²⁰ Suppose $w > 2$ and $w_{top} < 2$, i.e. top income bracket is less than 2. Then,

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

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

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

²⁰ Top marginal tax rate is taken from <http://www.oecd.org/tax/tax-policy/oecdtaxdatabase.htm>, Table I.7.

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.

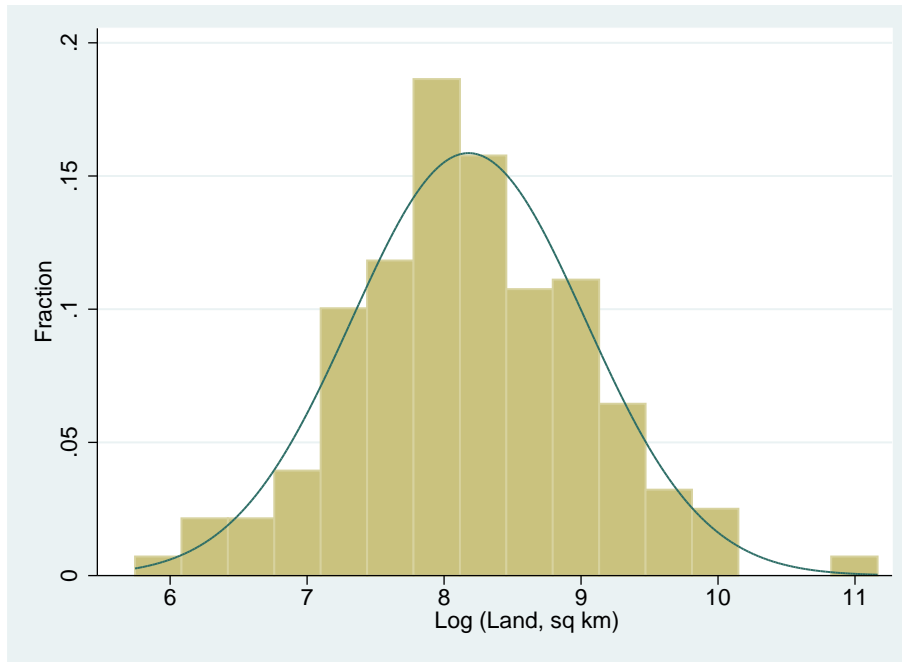


Figure 12: Land Distribution

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., AND G. EHRLICH (2012): “Metropolitan Land Values and Housing Productivity,” NBER Working Paper No. 18110.
- ALBOUY, D., AND N. SEEGER (2010): “The Optimal Population Distribution Across Cities and the Private-Social Wedge,” mimeo.
- 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.
- COMBES, P.-P., G. DURANTON, AND L. GOBILLON (2013): “The Costs of Agglomeration: Land Prices in French Cities,” PSE Working Papers halshs-00849078, HAL.
- 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., 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, P. A., AND J. A. MIRRLEES (1971a): “Optimal Taxation and Public Production: I—Production Efficiency,” *American Economic Review*, 61(1), 8–27.
- (1971b): “Optimal Taxation and Public Production II: Tax Rules,” *American Economic Review*, 61(3), 261–78.
- DURANTON, G., AND D. PUGA (2004): “Micro-foundations of urban agglomeration economies,” *Handbook of regional and urban economics*, 4, 2063–2117.
- ECKHOUT, J. (2004): “Gibrat’s Law for (All) Cities,” *American Economic Review*, 5(94), 1429–1451.
- ECKHOUT, J., R. PINHEIRO, AND K. SCHMIDHEINY (2014): “Spatial Sorting,” *Journal of Political Economy*, forthcoming.

- GLAESER, E. L. (1998): “Should transfer payments be indexed to local price levels?,” *Regional Science and Urban Economics*, 28(1), 1–20.
- 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., 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 (2013): “Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis,” *Review of Economic Studies*, forthcoming.
- HEATHCOTE, J., K. STORESLETTEN, AND G. VIOLANTE (2013): “Consumption and Labor Supply with Partial Insurance: An Analytical Framework,” NYU mimeo.
- 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.
- 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.
- KAPLOW, L. (1995): “Regional Cost-of-Living Adjustments in Tax/Transfer Schemes,” *Tax Law Review*, 51, 175–198.
- KENNAN, J. (2013): “Open Borders,” *Review of Economic Dynamics*, 16(2), L1–L13.
- 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.
- 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.
- SAHIN, A., J. SONG, G. TOPA, AND G. L. VIOLANTE (2014): “Mismatch Unemployment,” *American Economic Review*, 104(11), 3529–3564.
- TIEBOUT, C. (1956): “A Pure Theory of Local Expenditures,” *Journal of Political Economy*, 64(5), 416–424.
- 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.