

OPTIMAL SPATIAL TAXATION*

– preliminary and incomplete –

JAN EECKHOUT[†] AND NEZIH GUNER[‡]

February, 2014

Abstract

We analyze the role of optimal income taxation across different locations. Existing federal income tax schedules have a distortionary effect and result in the misallocation of labor across cities of different size. Because of higher productivity in big cities, wages for identically skilled workers are larger than in small cities. Progressive taxation thus implies that citizens in big cities pay higher taxes than in small cities. With mobility, utility is equalized, and the taxes are reflected in equilibrium wages and house prices. We solve for the optimal level of progressiveness. We find that the optimal level is not zero, but that it is less than what is observed in the US economy. Simulating the US economy under the optimal tax schedule, we find large effects on output and population mobility. GDP increases are in the range of 2.6–8.8%, and the fraction of population in 5 largest cities grows between 1.5–4.9%. The welfare effects however are small, 0.008–0.067%. This is due to the fact that the big output gains are lost in increased costs of living.

Keywords. Misallocation. Cities. Sorting. Population Mobility. City Size. General equilibrium. Skill Distribution.

*We are grateful to numerous colleagues for detailed discussion and insightful comments. Eeckhout gratefully acknowledges support by the ERC, Grant 208068. Guner gratefully acknowledges support by the ERC, Grant 263600.

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

[‡]ICREA-MOVE, Universitat Autònoma de Barcelona and Barcelona GSE, nezih.guner@movebarcelona.eu.

1 Introduction

The size of cities is determined by the location choice of workers, which is driven by worker productivity. Whether due to exogenous and endogenous factors, such as agglomeration externalities, there is substantial variation in Total Factor Productivity (TFP) across different locations. This is most notably reflected in the well established Urban Wage Premium. Nonetheless, the high TFP cities do not attract all workers from the economy at large, since housing prices in big, productive cities are higher and provide a countervailing force. In equilibrium, a worker is indifferent between cities with high wages and high housing prices, and cities with low wages and low housing prices.

In this context, we analyze the role of federal income taxation. A progressive taxation policy taxes earnings of equally skilled workers more in large cities. They are more productive and earn higher wages, and as a result, they pay a higher average tax rate. The location decision of workers is driven by after tax income, and since mobility equates utility between equally skilled workers, this progressive tax scheme does not redistribute utility. In the US for example, wages for identically skilled workers living in an urban area like New York (19 million inhabitants) are about 22% higher than wages of those living in smaller urban areas (say Janesville, WI with a population around 160,000). As a result of progressive taxation, the average tax rate of the median worker is about 3 percentage points higher.

We study a simple parametric representation of progressive taxes on labor incomes in the US and study the optimal level of progressiveness within this particular parametric representation of taxes. Somewhat surprisingly, we find that optimal level of progressiveness is not zero. Still, the optimal level should be lower than what we observe in the US economy. 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, identically skilled workers move into big cities, thus increasing their size at the detriment of smaller cities, and 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. The largest cities grow by more than 1.1%-3.5%, and the smallest cities lose 2.2%-9.2% of their population. But most importantly, aggregate output increases substantially. There is a gain of 2.65%-8.86% in GDP. The impact on output is remarkable, especially in the light of any typical gains that are obtained from adjusting distortions. In the light of the misallocation debate in macro economics on aggregate output differences due to the misallocation of inputs, most notably capital, we add a different insight. Due to existing income taxation schemes, also labor is substantially misallocated across cities within countries that have location-independent progressive taxation. We argue that a spatial tax policy is simple to implement and it has obvious economic consequences that are both desirable and sizable.

While the impact on output is enormous, the gains in terms of utility are rather tiny. The experiment that results in an 2.65% (8.86%) increase in GDP only leads to a 0.008% (0.067%) increase in Utilitarian welfare. This is due to the fact that most of the output gain in the more productive cities is eaten

away by higher expenditure on housing. As a result, together with the output increase there is a commensurate increase in housing prices that offsets much of the utility. 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 some progressiveness.

There are a number of caveats. First, we have not modeled the production of housing. The expenditure on owner occupied housing is also an investment. This in turn implies that the actual housing expenditure is less than the estimated amount. To take this into account, we propose an extension of the model that incorporates the idea that some of the housing expenditure is rebated. We find that this affects the utility level, but – at least in the simplest version of the extension – not the allocation of workers across different cities and therefore total output produced. Second, we have not modeled the expenditure on public goods. All the tax experiments are revenue neutral, that is, they generate the same amount of total income tax, but we do not make explicit what happens to the tax revenue. Like taxes, also benefits can distort the attractiveness of different locations. For example, when benefits go disproportionately to those in less productive locations, this will affect the location decision of the benefit recipients.¹ In addition to taxation, this further exacerbates the impact of the misallocation. In our model, we abstract from this important channel and focus on the role of active, full time workers.

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) is the first to explicitly consider federal taxation and argues that it creates distortions. He proposes 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) quantitatively analyzes the question. Starting from the Rosen-Roback tradeoff between equalizing differences across locations in a partial equilibrium model, he calibrates the model and concludes that any tax other than a lump sum tax is distortionary. In the absence of market clearing he cannot possibly to perform any welfare analysis, and needs to log-linearize preferences and technology in order to derive the marginal rate of substitution. Moreover, in the calibration exercise he does not allow for variation in housing characteristics which is a key feature of living in different locations: in large cities people live in smaller houses because there is substitution away from the expensive housing into consumption goods. We will show that all these aspects are key ingredients of the underlying mechanism.

¹Glaeser (1998) argues that even when benefits are indexed to local prices, there is no efficiency either because it affects utility levels and not marginal utility.

2 The Model

Population. The basic model builds on Eeckhout, Pinheiro, and Schmidheiny (2013). An economy has heterogeneously skilled workers, indexed by a skill type $i \in \mathcal{I} = \{1, \dots, I\}$. Associated with this skill order is a level of productivity x_i . The country-wide measure of skills of type i is M_i . There are J locations (cities) $j \in \mathcal{J} = \{1, \dots, J\}$. The amount of land in a city is fixed and denoted by H . The total workforce in city i is $\sum_i l_{ij}$.

Preferences. All citizens have preferences over consumption c , and the amount of land (or housing) h denoted by $c^{1-\alpha}h^\alpha$, where $\alpha \in [0, 1]$. The consumption good is a tradable numeraire good with price normalized to one. The price for one unit of land is p_j . The expenditure on housing is the flow value that compensates for the depreciation, interest on capital, etc. In a competitive market, the flow payment will equal the rental price. We assume that the housing stock is held by a zero measure of absentee landlords. Workers are perfectly mobile and can relocate instantaneously and at no cost. In equilibrium therefore, identical 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 satisfy $u(c_{ij}, h_{ij}) = u(c_{ij'}, h_{ij'})$, for all skill types $\forall i \in \{1, \dots, I\}$.

Amenities. Cities inherently differ in their attractiveness that is not captured in productivity, but rather in the utility of its citizens. This can be due to geographical features such as water (rivers, lakes and seas), mountains and temperature, but also due to man-made features such as cultural attractions (opera house, sports teams,...).² We denote the city-specific amenity by ε_j , which is a zero-mean term known to the citizens but unobserved to the econometrician. The utility in city j from consuming the bundle (c, h) is therefore written as:

$$u(c, h) = (1 + \varepsilon_j)c^{1-\alpha}h^\alpha. \quad (1)$$

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 observed amenities are indeed uncorrelated with city size.

Technology. Cities differ in their total factor productivity (TFP) which is denoted by A_j . 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 . Output is produced by choosing the right mix of differently skilled workers i . For each skill i , a firm in city j chooses a level of employment l_{ij} and produces output: $A_j F(l_{1j}, \dots, l_{Ij})$.

²We assume amenities are exogenous. In this paper, output produced is assumed to be a homogeneous good, we also abstract from the value of diversity of consumption goods.

We will consider a version of the CES technology, written as

$$F(l_{1j}, \dots, l_{Ij}) = A_j \left(\sum_i l_{ij}^\gamma x_i \right). \quad (2)$$

Firms pay wages w_{ij} for workers of type i . 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. Like land, firms are owned by absentee capitalists (or equivalently, all citizens own an equal share in the country-wide real estate bond).

To match the observed heterogeneity across skilled workers and across cities to the data, we need to introduce an error term η_{ij} at the skill-city level. Moreover, there are systematic patterns (skill distributions with thicker tails in large cities, as established by Eeckhout, Pinheiro, and Schmidheiny (2013)) that may warrant more structure on the technology such as differential complementarities to explain those patterns. In what follows for the heterogeneous agent case, we will not impose any such systematic structure and we will allow all the variation to be absorbed by the error structure in the CES technology: $F(l_{1j}, \dots, l_{Ij}) = A_j \left(\sum_i \eta_{ij} l_{ij}^\gamma x_i \right)$. Without any additional structure, it can easily be seen that this is equivalent to $F(l_{1j}, \dots, l_{Ij}) = \sum_i A_{ij} l_{ij}^\gamma$.

Market Clearing. In the country-wide market for skilled labor, markets for skills clear market by market, $\sum_{j=1}^J l_{ij} = L_i$, $\forall i$ and for housing, there is market clearing within each city $\sum_{i=1}^I h_{ij} l_{ij} = H$, $\forall j$. 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. Denote the pre-tax income by w and the post-tax income by $\tilde{w}$. We assume that 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:

$$\tilde{w}_{ij} = \lambda w_{ij}^{1-\tau},$$

where λ is the level of taxation and τ indicates the progressiveness ($\tau > 0$). This is the tax schedule proposed by Bénabou (2002). Heathcote, Storesletten, and Violante (2013) use the same function to study optimal progressiveness of income taxation in the U.S. The average tax rate is given by $\lambda w_{ij}^{-\tau}$ and the marginal tax rate is $\lambda(1 - \tau)w_{ij}^{-\tau}$. Taxes are proportional when $\tau = 0$, in which case the average rate is equal to the marginal rate and equal to λ . Under progressive taxes, $\tau > 0$ and the marginal rate exceeds the average rate.

3 The Equilibrium Allocation

Subject to after tax income, the worker solves

$$\begin{aligned} \max_{\{c_{ij}, h_{ij}\}} u(c_{ij}, h_{ij}) &= c_{ij}^{1-\alpha} h_{ij}^\alpha \\ \text{s.t. } c_{ij} + p_j h_{ij} &\leq \tilde{w}_{ij} \end{aligned} \quad (3)$$

for all i, j and the allocations are $c_{ij}^* = (1 - \alpha)\tilde{w}_{ij}$ and $h_{ij}^* = \alpha \frac{\tilde{w}_{ij}}{p_j}$. The indirect utility for a type i is:

$$u_{ij} = (1 + \varepsilon_j) \alpha^\alpha (1 - \alpha)^{1-\alpha} \frac{\tilde{w}_{ij}}{p_j^\alpha}. \quad (4)$$

Optimality in goods production implies that the first order condition of the firm holds: $w_{ij} = A_{ij} \gamma l_{ij}^{\gamma-1}$. Optimality in the production of housing requires that construction companies maximize $p_j h_j - B l_j h_j^\beta$ subject to $l_j h_j = H$. ?????????

Then using market clearing, optimality in production and the tax schedule, we can fully represent the equilibrium by the following system:

Lemma 1 *The equilibrium l_{ij} is completely determined by the system of $I \times J$ conditions:*

$$l_{ij} = \left(\frac{A_{ij}}{A_{i1}} \right)^{\frac{1}{1-\gamma}} \left[(1 + \varepsilon_j) \left(\frac{\sum_i A_{i1}^{1-\tau} l_{i1}^{1-(1-\gamma)(1-\tau)}}{\sum_i A_{ij}^{1-\tau} l_{ij}^{1-(1-\gamma)(1-\tau)}} \right)^\alpha \right]^{\frac{1}{(1-\gamma)(1-\tau)}} l_{i1}, \quad (5)$$

for all i, j together with $\sum_{j=1}^J l_{ij} = L_i$ for all i .

Proof. In Appendix. ■

This is a system of non-linear equations that we will solve computationally in the application. To makes some further progress theoretically, we first analyze the case of identical worker types. This turns out to have an explicit solution.

Identical Workers. For the special case of identical workers, with ε_1 normalized to 0, we can write (5) as (where we drop the subscript i):

$$l_j = l_1 \left[(1 + \varepsilon_j) \left(\frac{A_j}{A_1} \right)^{(1-\alpha)(1-\tau)} \right]^{\frac{1}{(1-\gamma)(1-\alpha)(1-\tau)+\alpha}},$$

for and $\sum_{i=j}^N l_j = \mathcal{L}$ and where $\mathcal{L}$ is the total labor force in the economy. This is a system of J equations

in J unknowns. The system is linear and we can solve it explicitly. Denote by K_i

$$K_j = \left[(1 + \varepsilon_j) \left(\frac{A_j}{A_1} \right)^{(1-\alpha)(1-\tau)} \right]^{\frac{1}{(1-\gamma)(1-\alpha)(1-\tau)+\alpha}}, \quad (6)$$

so we can solve for l_1 from the consistency condition $\sum_{j=1}^N l_j = \mathcal{L}$ which is equal to

$$l_1 = \frac{\mathcal{L}}{\sum_{j=1}^J K_j}, \quad (7)$$

and where $K_1 = 1$. With the solution for l_1 , we can immediately solve for all $l_j = K_j l_1$, and once we have l_j , we can also solve for the equilibrium wages $w_j = \gamma A_j l_j^{\gamma-1}$.

We can use the theoretical model to obtain predictions about the equilibrium values of output and welfare. First, we consider the effect of taxes on output. To that end, we consider the identical agent economy with two cities. Since the output maximizing planner does not care about utility, we set $\varepsilon_j = 0$. We also assume that city 1 is the smallest (with the lowest TFP) and that it is used as the reference city, i.e. with $K_1 = 1$. First, we derive the the population allocation that maximizes output. Total output is given by $Y = A_1 l_1^\gamma + A_2 l_2^\gamma$. Solving for the optimal l_1, l_2 that maximizes Y , using the fact that $l_1 + l_2 = \mathcal{L}$ we get

$$l_1 = \frac{\mathcal{L}}{1 + \left(\frac{A_2}{A_1} \right)^{\frac{1}{1-\gamma}}}. \quad (8)$$

We can state the following result:

Proposition 1 *Let there be 2 cities, identically skilled workers and no amenities. Then no tax τ will achieve maximal output, and output is weakly decreasing in the tax progressiveness τ .*

Proof. In Appendix. ■

The impact of progressive taxes puts a higher burden on high wage, high productivity cities. In equilibrium therefore, more progressiveness results in underpopulation of large cities and overpopulation of small cities relative to the output maximizing allocation. In other words, increasing τ draws more workers away from the productive cities, thus resulting in lower output being produced. Observe that this is not bounded by $\tau = 0$. Even for $\tau < 0$ – that is regressive taxation – this continues to hold. Now there is a subsidy to locate in big, productive cities, and output will continue to grow as τ decreases. Therefore, output keeps increasing even as τ becomes negative.

The optimal allocation has workers in both cities. This is because of decreasing returns and the fact that the marginal return is infinite when the city is nearly empty (the Inada conditions hold). Nonetheless, this particular tax schedule with a power functional form will never be sufficiently strong to move workers beyond the output maximizing allocation.

4 The Planner's Problem

We distinguish between the optimal allocation chosen by the planner, and the Ramsian optimal taxation problem where the planner chooses tax instruments in order to affect the equilibrium allocation. In the latter exercise, the planner assumes agents operate in a decentralized economy with equilibrium prices, albeit affected by taxes. In the former there are no prices.

4.1 The optimal allocation

The planner chooses the bundles l_j, c_j, h_j in all cities j to maximize Utilitarian welfare:

$$\begin{aligned} \max_{l_j, c_j, h_j} \quad & \sum_j (1 + \varepsilon_j) c_j^{1-\alpha} h_j^\alpha l_j^{-\delta} \\ \text{s.t.} \quad & \sum_j c_j l_j + \sum_j B l_j h_j^\beta + G = \sum_j A_j l_j \\ & l_j h_j = H \\ & \sum_j l_j = \mathcal{L}, \end{aligned} \tag{9}$$

where G is the exogenously given government spending and $B h_j^\beta$ is the production technology of housing with $\beta \leq 1$. The cost of housing is increasing in the size of the house, but at a (weakly) decreasing rate: small apartments still need a bathroom and a kitchen, so the unit cost per square meter is higher. Think of this as the flow payment (mortgage or rent) of the investment.

We solve the reduced version of the model by substitution for $h_j = \frac{H}{l_j}$.

$$\begin{aligned} \max_{l_j, c_j, h_j} \quad & \sum_j (1 + \varepsilon_j) H^\alpha c_j^{1-\alpha} l_j^{-(\alpha+\delta)} \\ \text{s.t.} \quad & \sum_j c_j l_j + \sum_j B H^\beta l_j^{1-\beta} + G = \sum_j A_j l_j \\ & \sum_j l_j = \mathcal{L}, \end{aligned} \tag{10}$$

Observe that if $\beta = 1$, the total production cost of housing is independent of the population l_j and equal to BH .

The planner's solution to this problem satisfies the FOCs:

$$\begin{aligned} (1 + \varepsilon_j)(1 - \alpha) c_j^{-\alpha} \left(\frac{H}{l_j} \right)^\alpha + \lambda l_j &= 0 \\ -(1 + \varepsilon_j)(\alpha + \delta) c_j^{1-\alpha} H^\alpha l_j^{-\alpha-\delta-1} + \lambda c_j + (1 - \beta) \lambda B H^\beta l_j^{-\beta} - \lambda A_j + \psi &= 0 \end{aligned}$$

where ϕ and ψ are the Lagrangian constraints. So we have a system of $2N + 2$ equations (the FOCs

and the two Lagrangian constraints) in $2N + 2$ unknowns $\{\{l_j, c_j\}, \lambda, \psi\}$

For the case of two cities, with $\beta = 1$, $\delta = 0$ and with $\varepsilon_1 = 0$:

$$\begin{aligned}
(1 + \varepsilon_2)(1 - \alpha)c_2^{-\alpha}H^\alpha l_2^{-(1+\alpha)} + \phi &= 0 \\
(1 - \alpha)c_1^{-\alpha}H^\alpha l_1^{-(1+\alpha)} + \phi &= 0 \\
-(1 + \varepsilon_2)\alpha c_2^{1-\alpha}H^\alpha l_2^{-\alpha-1} + \phi c_2 - \phi A_2 + \psi &= 0 \\
-\alpha c_1^{1-\alpha}H^\alpha l_1^{-\alpha-1} + \phi c_1 - \phi A_1 + \psi &= 0 \\
l_1 + l_2 &= \mathcal{L} \\
c_1 l_1 + c_2 l_2 + HB + G - A_1 l_1 - A_2 l_2 &= 0
\end{aligned}$$

which implies

$$(1 + \varepsilon_2)^{-1}c_2^\alpha l_2^{1+\alpha} = c_1^\alpha l_1^{1+\alpha}$$

and using the expression for ϕ and substituting for ψ

$$\begin{aligned}
\alpha c_2 c_1^{-\alpha} l_1^{-\alpha-1} + (1 - \alpha)c_1^{-\alpha} l_1^{-(1+\alpha)}(c_2 - c_1 - A_2 + A_1) &= \alpha c_1^{1-\alpha} l_1^{-\alpha-1} \\
\alpha c_2 + (1 - \alpha)(c_2 - c_1 - A_2 + A_1) &= \alpha c_1 \\
c_2 - c_1 &= (1 - \alpha)(A_2 - A_1)
\end{aligned}$$

which implies

$$(1 + \varepsilon_2)^{-1}(c_1 + (1 - \alpha)(A_2 - A_1))^\alpha (\mathcal{L} - l_1)^{1+\alpha} = c_1^\alpha l_1^{1+\alpha}.$$

There is no explicit solution for l_1, c_1 from this equation. Hence, we need to solve the model numerically. We can of course do comparative statics exercises on the equilibrium allocation. For example how does population, consumption and utility change as G or B increases. But I think we can do that better numerically to get quantitatively relevant predictions.

4.2 Ramsey optimal tax

But the planner does not care about the output maximizing allocation but rather the welfare maximizing allocation. We consider two cases: 1. a simple progressive tax schedule and 2. a city-specific tax schedule.

4.2.1 A simple progressive tax policy

Consider a Utilitarian planner who chooses the tax schedule (λ, τ) to maximize the sum of utilities subject to the revenue neutrality constraint, i.e. she has to raise the same amount of tax revenue.

In order to derive total welfare, we write utility as:

$$u_j = (1 + \varepsilon_j) \alpha^\alpha (1 - \alpha)^{1-\alpha} \frac{\tilde{w}_j}{p_j^\alpha} = (1 + \varepsilon_j) [(1 - \alpha) \lambda]^{1-\alpha} H^\alpha (\gamma A_j)^{(1-\tau)(1-\alpha)} l_j^{-(1-\gamma)(1-\tau)(1-\alpha)-\alpha}. \quad (11)$$

Then total welfare is given by $\mathcal{W} = \sum_j l_j u_j$. The planner maximizes $\mathcal{W}$ subject to the revenue neutrality constraint, i.e. total tax revenue is constant:³

$$\begin{aligned} \max_{\lambda, \tau, l_j} \quad & [(1 - \alpha) \lambda]^{1-\alpha} \gamma^{(1-\tau)(1-\alpha)} H^\alpha \sum_j (1 + \varepsilon_j) A_j^{(1-\tau)(1-\alpha)} l_j^{1-(1-\gamma)(1-\tau)(1-\alpha)-\alpha} \\ \text{s.t.} \quad & \sum_j w_j l_j - \lambda \sum_j w_j^{1-\tau} l_j = G, \quad u_j = u_{j'}, \quad \forall j \neq j' \quad \text{and} \quad \sum_j l_j = \mathcal{L}, \end{aligned} \quad (12)$$

where G is a constant equal to the tax revenue. The solution to this problem involves solving a system of $N + 4$ equations ($N + 2$ FOCs and 2 Lagrangian constraints⁴) in $n + 4$ variables $\{\{l_j\}, l_j, \tau, \phi, \psi\}$. We cannot derive an analytical solution, so we will characterize the optimal tax schedule from simulating the US economy in the next section.

4.2.2 A city-specific tax policy

Consider a Utilitarian planner who chooses the tax schedule $\{t_j\}$ to maximize the sum of utilities subject to the revenue neutrality constraint, i.e. she has to raise the same amount of tax revenue.

Using the unique economy-wide tax schedule $\tilde{w} = \lambda w^{1-\tau}$ was motivated by realism. It fits existing tax schedules quite closely, and all federal tax schedules we know of apply country-wide. Typically, the progressiveness is justified on the grounds of redistribution between heterogeneous types, without consideration for the spatial implications. Our objective so far was therefore to use the restriction of a *common* federal tax schedule to derive the optimal allocation. Yet, from a spatial viewpoint it is restrictive because we ask a two-parameter schedule to fit the variation in all 264 cities. Increasing the number of parameters will improve the fit. We could go all the way and have a tax schedule that is city specific.

With identical workers, we can fully capture the spatial heterogeneity by a tax T_j where $\tilde{w}_j = (1 - t_j) w_j$.⁵ Utility can be written as:

$$u_j = (1 + \varepsilon_j) (1 - \alpha)^{1-\alpha} H^\alpha ((1 - t_j) \gamma A_j l_j^{\gamma-1})^{1-\alpha} l_j^{-\delta}. \quad (13)$$

Now the optimal tax schedule is a J -tuple $\{t_j\}$ and one for populations $\{l_j\}$ that solves maximizes

³This formulation implicitly assumes workers optimally choose consumption and housing.

⁴Observe that the utility equalization constraint is redundant since it is implicit from optimization of the additively separable objective function.

⁵Since there is no heterogeneity, it does not matter whether it is written additively, proportionally $\tilde{w} = w_j - T_j$ or in any one-parameter functional form for that matter $\tilde{w} = f_j(w)$ as long as it is city specific.

welfare subject to revenue neutrality and labor mobility and

$$\begin{aligned} \max_{\{t_j, l_j\}} \quad & (1 - \alpha)^{1-\alpha} H^\alpha \sum_j (1 + \varepsilon_j) ((1 - t_j) \gamma A_j l_j^{\gamma-1})^{1-\alpha} l_j^{1-\alpha} \\ \text{s.t.} \quad & \sum_j t_j \gamma A_j l_j^\gamma = G, \quad u_j = u_{j'}, \quad \forall j, j', \quad \text{and} \quad \sum_j l_j = \mathcal{L}. \end{aligned} \tag{14}$$

where G is a constant equal to the tax revenue. The solution to this problem involves solving a system of $N + 4$ equations ($N + 2$ FOCs and 2 Lagrangian constraints⁶) in $n + 4$ variables $\{\{l_j\}, l_j, \tau, \phi, \psi\}$. We cannot derive an analytical solution, so we will characterize the optimal tax schedule from simulating the US economy in the next section.

5 Quantifying the Spatial Misallocation

We now quantify the magnitude of spatial misallocation for the case of identical agents. We proceed in following steps: First, given the U.S. data on the distribution of labor force across cities (l_j) and wages in each city (w_j), we back out the productivity parameters A_j . Second, given (l_j, w_j) and a representation of current US taxes on labor income, $(\lambda^{US}, \tau^{US})$, we compute ε_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 Data

The data on the distribution of labor force across cities (l_j) and wages in each city (w_j) are calculated from 2009 CPS March Supplement. For each of 264 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 1 shows the distribution of population and wages across MSAs. The average labor force is 475,500, with a maximum (New York-Northern New Jersey-Long Island) of more than 9.5 million and a minimum (Ocean City, NJ) of about 26,000. The population distribution is highly skewed, close to log-normal, where the top 5 MSAs account for 21.8% of total labor force. Average weekly wages is 707\$. The highest weekly wage is 40% above the mean level (Bridgeport-Stamford-Norwalk, CT) and the lowest is about half of the mean level (Bowling Green, KY).

⁶Observe that the utility equalization constraint is redundant since it is implicit from optimization of the additively separable objective function.

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

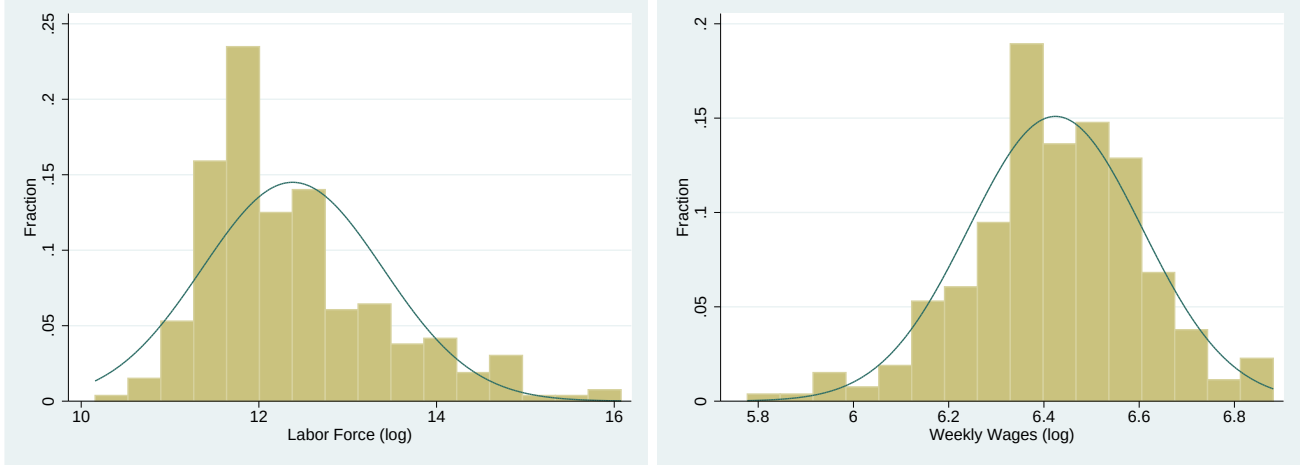


Figure 1: Labor force and wage distribution, histogram and Kernel density estimates: A. Labor force; B. Wages.

5.2 Taxes

As we mentioned above, we assume that the relation between after and before tax wages are given by $\tilde{w} = \lambda w^{1-\tau}$, where λ is the level of taxation and τ indicates the progressiveness ($\tau > 0$). 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 λ and τ from a simple OLS regression

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

The estimated values of λ^{US} is 0.752 and τ^{US} is 0.120. Hence at mean wages ($w = 1$), tax rate is 25%. Tax rates at $w = 0.5$, $w = 2$ and $w = 5$ are 18%, 31% and 38%, respectively. Estimating the same tax function with the U.S. micro data on taxes from the Internal Revenue Services (IRS), Guner,

⁸<http://www.oecd.org/social/soc/benefitsandwagestax-benefitcalculator.htm>, accessed on March 15, 2013.

Kaygusuz, and Ventura (2013) estimate lower values for τ , around 0.03-0.06. 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. With $w_2 = 2.5$ and $w_1 = 0.5$, our estimates imply a progressiveness 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 , while they estimate a progressiveness wedge of 0.15.

5.3 Optimal Progressiveness

Given observations on l_j and w_j , we first calculate productivity level in each city as

$$A_j = \frac{w_j(l_j)^{1-\gamma}}{\gamma}, \forall j.$$

Then, we calculate the error term ε_j from utility equalization condition across cities. Given the indirect utility function in equation (4), for any two locations j and j' , the following equality must hold:

$$(1 + \varepsilon_j)\alpha^\alpha(1 - \alpha)^{1-\alpha} \frac{\tilde{w}_j}{p_j^\alpha} = (1 + \varepsilon_{j'})\alpha^\alpha(1 - \alpha)^{1-\alpha} \frac{\tilde{w}_{j'}}{p_{j'}^\alpha}.$$

Since in equilibrium $p_j = \alpha \frac{\tilde{w}_j}{h_j}$ and $h_j = \frac{H}{l_j}$, if we normalize $\varepsilon_1 = 0$, we obtain ε_j as

$$1 + \varepsilon_j = \frac{l_j^\alpha w_1^{(1-\alpha)(1-\tau^{US})}}{l_1^\alpha w_j^{(1-\alpha)(1-\tau^{US})}}.$$

Calculations for A_j and ε_j obviously depend on the values we assume for α and γ . 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 $\alpha/\lambda = \frac{0.24}{0.752} = 0.319$ from after-tax income at mean income ($w = 1$). Given α , we choose $\gamma = 1.033$, which, consistent with the evidence provided by Albouy (2008), generates zero correlation between ε and A across MSAs in the benchmark economy. Finally we select H , the amount of housing in each city, such that average housing price across cities is 100. Given H and l_j , housing prices in each city are given by $p_j = \alpha \frac{\tilde{w}_j l_j}{H}$.

The computed values of ε_j as well as housing prices differ greatly across metropolitan statistical areas. We set $\varepsilon_1 = 0$ for Acron (OH), the first MSA alphabetically. The mean value of ε_j across MSAs is 0.446. The highest level of ε_j , 1.45, is calculated for New York-Northern New Jersey-Long Island MSA, while the lowest value is -0.62, for Ocean City (NJ). The calibration procedure assigns a high value of ε for New York-Northern New Jersey-Long Island MSA to account for its very large size. Estimated housing prices are about 450 in New York-Northern New Jersey-Long Island, followed by

Los Angeles-Long Beach-Santa Ana (CA) and Chicago-Naperville-Joliet where housing prices are 276 and 202, respectively. The lowest housing prices are computed for Bowling Green (KY), 0.8, and Ocean City (NJ), 1.17.

Given values for A_j and ε_j , the next step is to find counterfactual allocations for any level of $\tau \neq \tau^{US}$. This is done simply by equations (6) and (7). Wages are then calculated as

$$w_j(\tau) = \gamma A_j (l_j(\tau))^{\gamma-1},$$

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 2.A shows the percentage change in utility from the benchmark economy for different values of τ . The optimal value τ^* , is 0.0067 and the associated value for λ^* that guarantees revenue neutrality is 0.7526. The optimal τ^* is less than τ^{US} , i.e. taxes should be less progressive than observed taxes. However, the optimal τ is not zero. While $\tau = 0$ results in larger movements of population to more productive cities and maximizes the output, it does not necessarily maximize consumer's utility as consumers are hurt by higher housing prices in larger cities. Figure 2.B shows the implied tax schedule under $(\lambda^{US}, \tau^{US})$ and (λ^*, τ^*) . While tax rates for $w = 1$ is qvery similar 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 21%, 28% and 32%, respectively.

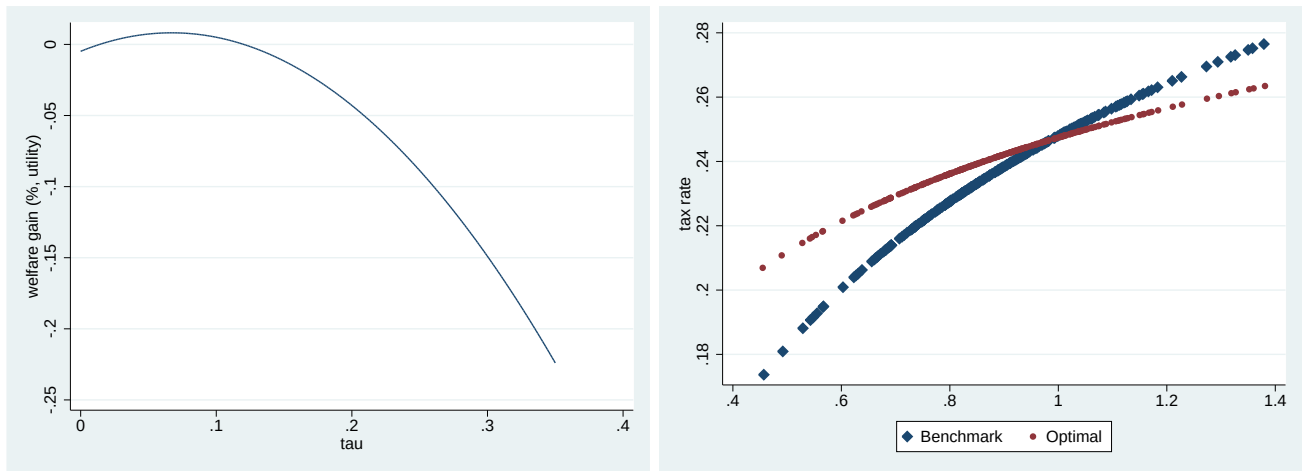


Figure 2: A. Welfare gain for different values of τ ; B. The optimal tax schedule τ^* compared to that in the benchmark economy τ^{US} .

Now we can evaluate the implications of a tax change in the tax schedule from τ^{US} to τ^* , both for individual cities and in the aggregate. Consider first the impact on individual cities which is summarized in Figure 3 and Table 1. The table gives the numerical values for those cities with extreme values either for TFP A or for amenities ε . Those same cities are explicitly identified in the scatter plots in Figure 3.

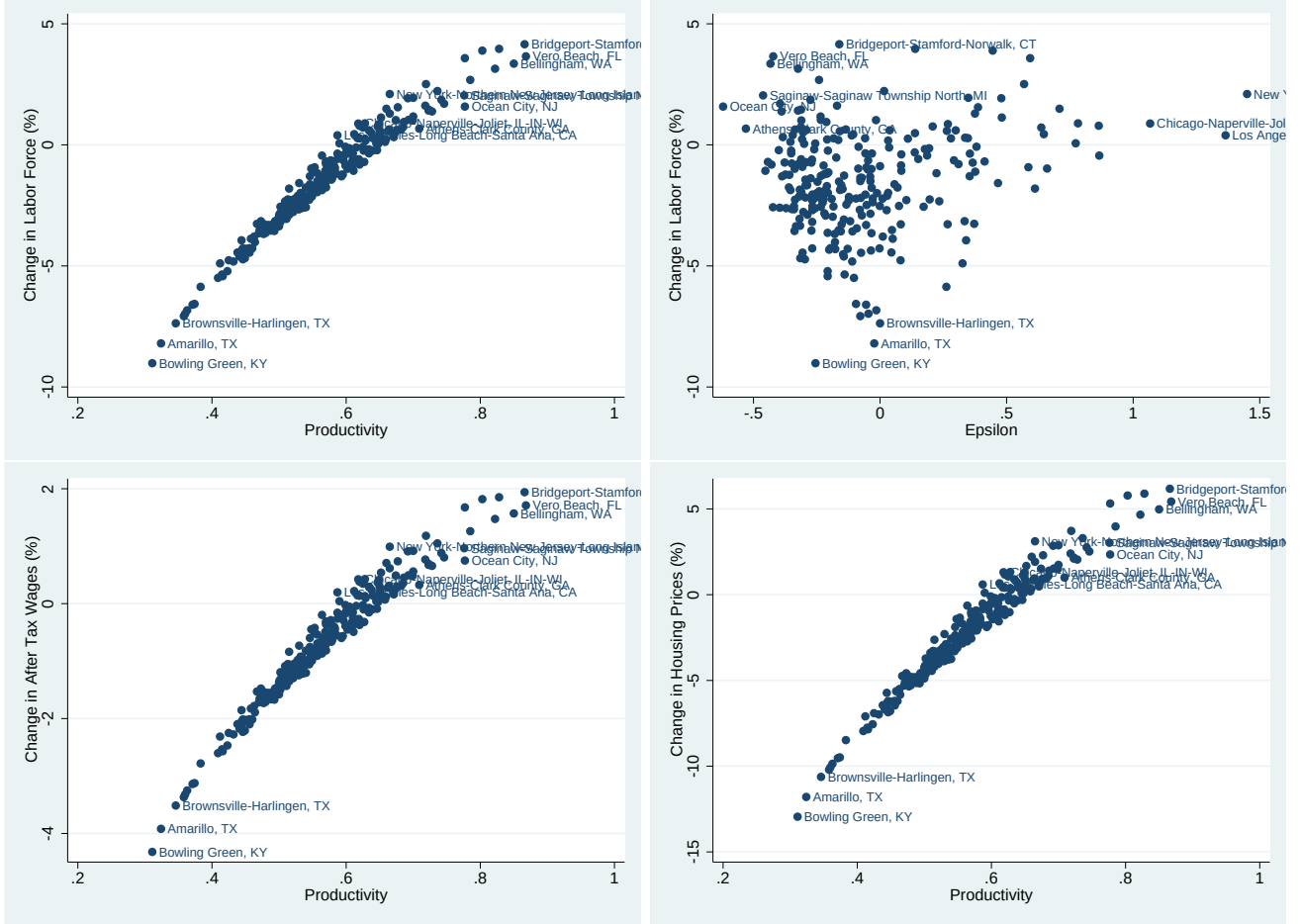


Figure 3: Implied changes of implementing the optimal policy τ^* . A. Change in population by TFP; B. Change in population by ε ; C. Change in $\tilde{w}$ by TFP; D. Change in housing prices p by TFP.

Since the optimal degree of progressiveness τ^* is below existing τ^{US} , the optimal policy lowers tax payments in high productivity cities. Figure 3.A shows that the high A cities grow in size while the low productivity A cities lose population. The largest population growth rate is 3.6% whereas Bowling Green loses 9% of its population. The role of amenities is important and sizable, as is apparent in Figure 3.B. Because the γ is chosen to obtain zero correlation between A and ε , there is no systematic pattern.

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 3.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 not decreasing returns to scale. In fact, the economy is calibrated to mildly increasing returns ($\gamma =$), so that should provide even further mobility. Rather, what eventually deters further population growth is housing prices. Figure 3.D shows the change in housing prices. High productivity cities are up to 6% more expensive while low productivity cities face housing price drops of up to 12.9%.

Of course with higher housing prices goes substitution of housing for consumption. In the high productivity cities, workers live in even smaller housing while increasing goods consumption. Housing consumption decreases by up to 4% in the high productivity cities in substitution for nearly 2% higher goods consumption. In the less productive cities housing consumption increase by up to 9.9% at the cost of decreased goods consumption by 4.3%. Given homothetic preferences, the marginal rate of substitution is constant 4.A.

Table 1: Benchmark Economy, move from τ to τ^* : Outcomes for selected cities.

MSA	A	ε	Δl	$\% \Delta p$	$\% \Delta c$	$\% \Delta h$
Highest A						
Vero Beach, FL	0.87	-0.42	3.66	5.43	1.71	-3.53
Bellingham, WA	0.87	-0.16	3.66	6.18	1.94	-4.00
San Jose-Sunnyvale-Santa Clara, CA	0.83	-0.43	3.35	4.98	1.57	-3.25
Lowest A						
Brownsville-Harlingen, TX	0.35	0.00	-7.37	-10.69	-3.51	7.96
Amarillo, TX	0.32	-0.02	-8.20	-11.80	-3.92	8.93
Bowling Green, KY	0.31	-0.26	-9.02	-12.95	-4.32	9.91
Highest ε						
New York-Northern New Jersey-Long Island	0.67	1.45	2.10	3.11	0.99	-2.06
Los Angeles-Long Beach-Santa Ana, CA	0.59	1.37	0.40	0.59	0.20	-0.39
Chicago-Naperville-Joliet, IL-IN-WI	0.62	1.07	0.88	1.31	0.42	-0.87
Lowest ε						
Saginaw-Saginaw Township North, MI	0.78	-0.46	2.04	3.07	0.97	-2.00
Athens-Clark County, GA	0.71	-0.53	0.67	1.00	0.33	-0.67
Ocean City, NJ	0.78	-0.63	1.58	2.34	0.75	-1.56

Table 2 shows the aggregate outcomes from moving the benchmark allocation to the optimal. On average output goes up by about 2.65%. This increase is substantial. This is driven by the population moving to the more productive cities. The population in the 5 largest cities grows by 1.5%, despite the fact that the top three are large in part because they also offer high amenities ε . 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 4.B. Not surprisingly, aggregate housing prices go up by 3.2%.

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

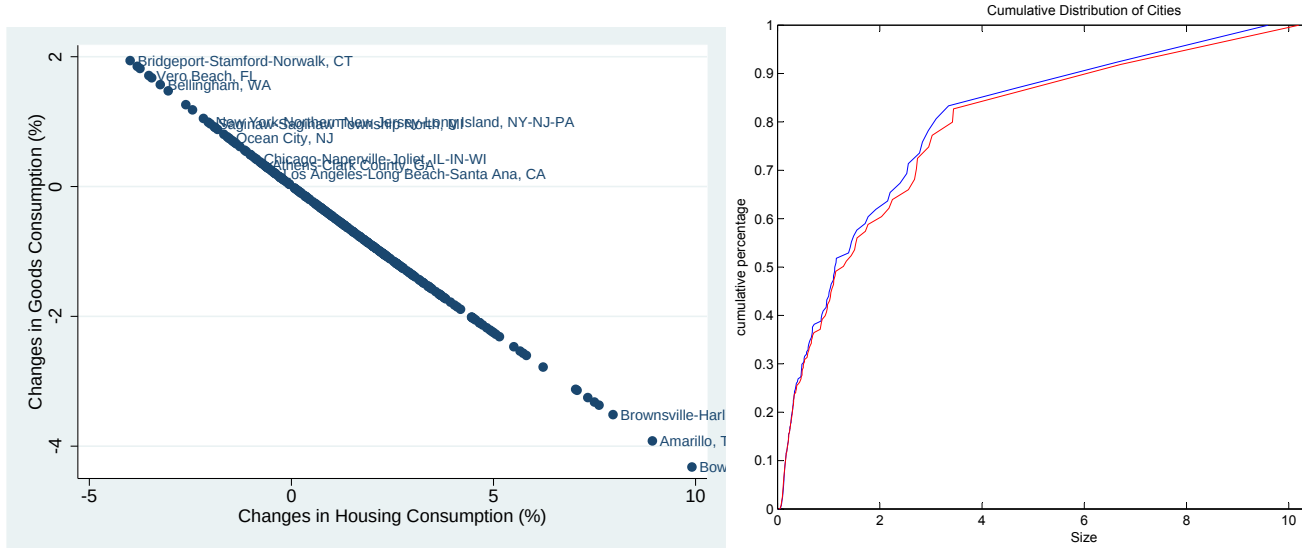


Figure 4: Implied changes of implementing the optimal policy τ^* : A. Substitution between c and h ; B. Cumulative Distribution of city sizes.

increases by 0.0082%, almost nothing. The reason for such tiny welfare gains is quite simple. Under the optimal taxes, after tax wages in cities that have initially high productivity increases. These cities, however, also get more crowded and housing prices goes up. With higher prices, housing consumption in these cities declines. True, from substitution of housing for goods, 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^* : Aggregate Outcomes.

Outcomes	% Δ
Output gain	2.65
Population in 5 largest cities	1.5
Average housing prices	3.2
Welfare gain (utility)	0.0082

5.4 Sensitivity

There are two key forces in the model economy that determine τ^* . On the one hand, a higher value of γ pushed the optimal value of τ^* towards zero, as there are large agglomerations effects and it is optimal to move as many people as possible to more productive cities. On the other hand, a high value for α pushes the optimal value of τ^* up, as consumers spend a larger fraction of their income on housing. As a result, they are hurt more by housing price increases in more productive cities that result from more workers living there. Table 3 shows results for two different parametrization of γ and α . First,

Table 3: Sensitivity Analysis.

	$\alpha = 0.24, \gamma = 1, \tau^* = -0.0082$	$\alpha = 0.3191, \gamma = 1.2, \tau^* = -0.0834$
Outcomes	% Δ	% Δ
Output gain	8.86	20.30
Population in 5 largest cities	4.91	9.63
Average housing prices	10.36	23.39
Welfare gain (utility)	0.067	0.167

we set $\alpha = 0.24$, the value estimated by Davis and Ortalo-Magné (2011) and $\gamma = 1$. Second, we set $\alpha = 0.319$ as in the benchmark economy, but we set $\gamma = 1.2$. In both of these cases (a low value for α or a high value for γ), the optimal τ turns out to be negative, i.e. the taxes are regressive. As Table 3 shows with $\tau^* < 0$, there are much larger gains in output, up to 20.3%. Not surprisingly, this is driven by bigger population movements into large cities, which in turn results in much bigger average housing price increases. Also welfare increases substantially more, but the magnitudes of the welfare improvements remain small.

5.5 Heterogeneous Skills

For the model with heterogeneous skills, we use the same logic as above. Now we divide the population in three skill categories. We obtain the values for A_{ij} and ε_j from the firm's first order condition and the consumer's utility equalization using wages in city j for skill i as well as employment by skill i in city j . Once we have pinned down the baseline values for these parameters (A_{ij}, ε_j), we can calculate counterfactual allocations using the system (5). From a welfare point of view, we have utility equalization across cities *within* a given skill level, but not between skill levels. Hence we cannot make welfare comparisons across skill types without a social welfare criterion. However we can calculate the welfare changes for each group and see the impact of different tax schemes by skill level. The actual calibration exercise is current being completed.

6 Extensions

6.1 Rebating Housing Expenditure

So far, the housing stock was held by zero measure absentee landlords. In reality, housing is held by citizens. To properly account for housing expenditure, we would need to model the technology for housing construction. Because this is beyond the scope of the paper, we propose the following reduced form way of modeling housing supply. We assume that a fraction ω of the housing expenditure is a foregone cost of construction, maintenance and capital, and a fraction $1 - \omega$ flows back to the households. We distinguish between two cases: 1. the housing stock of a city is exclusively held by the residents of

that city; 2. the housing stock is held economy wide, and all households hold an equal share. In what follows we continue to focus on the representative agent economy.

Within city housing ownership. Now each household also has a return on housing equal to $1 - \omega$ times the total expenditure on housing. Therefore the budget constraint of the household is $c_j + p_j h_j = \tilde{w}_j + (1 - \omega)p_j h_j$ or:

$$c_j + \omega p_j h_j = \tilde{w}_j$$

Now this is simply rescaling the cost of housing by ω . It is easily verified that in the condition for utility equalization across cities $p_j h_j = \alpha \tilde{w}_j$, ω cancels out since $p_j = \frac{\alpha l_j \tilde{w}_j}{\omega H}$ and the allocation of workers across different cities is not affected by ω :

$$(1 + \varepsilon_j) \frac{\tilde{w}_j}{p_j^\alpha} = \frac{\tilde{w}_1}{p_1^\alpha}.$$

The allocation across cities is essentially unaltered, and therefore total output produced is unaltered. Note though that taxes still have an impact since workers face progressive income taxes. What changes is the utility level obtained. Because the expenditure share on housing is lower, the equilibrium utility will, all other parameters equal, be different and increasing in ω . The equilibrium quantities consumed are $c = (1 - \alpha)\tilde{w}$ and $h = \frac{\alpha \tilde{w}}{\omega p}$, and therefore:

$$u_j = (1 + \varepsilon_j) \alpha^\alpha (1 - \alpha)^{1-\alpha} \frac{\tilde{w}_j}{(\omega p_j)^\alpha}.$$

If we had an idea of the magnitude of ω , we can immediately redo the entire analysis. The important implication though from this exercise is that the equilibrium allocation is unchanged. For all ω , output is the same and so is the city size distribution. As a result, there is no effect of rebating housing expenditure on any observed outcomes and prices. And while utility level changes, the utility gain or loss is unaltered since utility is proportional to ω .

Of course, if income from housing is taxed, then we need to take that into account as well. Suppose taxes on capital (housing) is proportional. Let τ_k denote the tax on capital income, then writing $\tilde{\omega} = \omega + \tau_k(1 - \omega)$ we can write the budge constraint as $c_j + \tilde{\omega} p_j h_j = \tilde{w}_j$. Again all is the same as long as we use $\tilde{\omega}$ instead of ω .

Country-wide housing ownership. When housing is owned by any worker from any part of the country, the logic is similar, but unfortunately the analytics are very different. This is due to the fact that we can no longer conveniently represent expenditures as a proportion of the population. The budget constraint of the household is $c_j + p_j h_j = \tilde{w}_j + (1 - \omega) \sum_j p_j h_j$ and therefore the solution is now a fixed point problem that takes into account housing expenditures in all other cities.

6.2 Congestion Externalities and Endogenous City Size

We focus on the benefits from agglomeration. But there are also costs to agglomeration. Kahn (2010) estimates the elasticity of time with respect to city size as 0.13 (double for males, 0.26!): as city size doubles, commuting time increases by 13%. Combes, Duranton, and Gobillon (2013) find an estimate for the cost of agglomeration (congestion, commuting,...) with respect to size of 0.041. As the city size doubles, the agglomeration cost increases by 4.1%. Observe that the cost of agglomeration is nearly identical to the benefits, hence the net effect is close to constant returns.

Observe that the agglomeration benefits are already captured in TFP, A_j is increasing in city size. The agglomeration cost is a negative externality that is size dependent. Denote the externality by $C(l_j)$ and using the estimate of Combes, Duranton, and Gobillon (2013) we can write it as $C(l_j) = l_j^\delta$ where $\delta = 0.959$. The utility can thus be written as $u = (1 + \varepsilon_j)l_j^\delta c^{1-\alpha} h^\alpha$. Then the indirect utility is:

$$u_j = (1 + \varepsilon_j)l_j^\delta [(1 - \alpha)\lambda]^{1-\alpha} H^\alpha (\gamma A_j)^{(1-\tau)(1-\alpha)} l_j^{-(1-\gamma)(1-\tau)(1-\alpha)-\alpha} \quad (15)$$

and for the case of $\gamma = 1$:

$$u_j = (1 + \varepsilon_j)l_j^\delta [(1 - \alpha)\lambda]^{1-\alpha} H^\alpha A_j^{(1-\tau)(1-\alpha)} l_j^{-\alpha}. \quad (16)$$

Now we can measure the economic cost of agglomeration due to the mobility. Even if we completely rebate the housing cost ($\omega = 1$), there will be a loss in utility due to agglomeration. We can now take this into account in the utility calculation.

We could also endogenize city size in a model with a central business district and a cost of commuting, the size of the city will be endogenous H_j . We can then introduce the opportunity cost of distance from the center as in Combes, Duranton, and Gobillon (2013). This gives an endogenous H_j dependent on the commuting technology. Since we cannot distinguish amenities from size ($(1 + \varepsilon_j)H_j^\alpha$ is observed jointly) further modeling it without additional data on either size or amenities will not allow us to identify these separately. We therefore continue to treat them jointly.

6.3 The Expenditure Share of Housing

The expenditure share of housing is assumed fixed and constant across levels of income. Davis and Ortalo-Magné (2011) establish that the average expenditure share on housing is remarkably constant and equal to 0.24. There is some variation however within income. Using CEX data, Eeckhout, Pinheiro, and Schmidheiny (2013) find that the expenditure share is decreasing in income. The lowest income households spend around 0.35 of their income on housing while the highest income households spend 0.22. Here we ask whether such variation by income *within* a given city can be affected by taxation.

When gross income is used to measure the expenditure share, then the magnitude of the share will

be underestimated since what matters for the agent's optimization decision is net income. Suppose the measured income share of a worker with income w_i is denoted by s_i . Then using gross income, the expenditure share on housing is equal to $s_i = \frac{h_{ij}p_j}{w_{ij}}$. Now suppose the expenditure share of net income is constant and equal to α , then $hp = \alpha\tilde{w}$. But the measured share, based on gross income is

$$s = \frac{hp}{\left(\frac{\tilde{w}}{\lambda}\right)^{\frac{1}{1-\tau}}} = \alpha\lambda^{\frac{1}{1-\tau}}\tilde{w}^{\frac{-\tau}{1-\tau}}.$$

In the absence of progressive taxation ($\tau = 0$), the estimated share with gross taxes is still constant, but it is too low relative to the share of net income: $s = \alpha\lambda < \alpha$. An estimate of $s = 0.24$ based on gross income would imply a share of net income equal to $\alpha = s/\lambda \approx 0.32$ (taking $\lambda = 0.75$). This variation may also explain why different estimates for the expenditure share are obtained when consumption data are used rather than income data. Albouy (2008), for example, finds an expenditure share of 0.33.

When taxes are progressive ($\tau > 0$), the estimated share of expenditure based on gross income now depends on the income level. In particular, it is decreasing in income:

$$\frac{\partial s}{\partial \tilde{w}} = -\frac{\tau}{1-\tau}\alpha\lambda^{\frac{1}{1-\tau}}\tilde{w}^{\frac{1}{1-\tau}} < 0.$$

The larger the income, the smaller the measured expenditure share based on gross income, even if the expenditure share based on net income is constant. For example, for a wage level normalized to one with $\lambda = 0.75, \tau = 0.115, \alpha = 0.33$, the expenditure share on housing of before tax income is 25%, while those earning twice the normalized income spend 23% and those earning 0.25 times the normalized income spend 31%. This seems to indicate that progressive taxation is an important component of the varying expenditure share of gross income. The bottom line is that maintaining the assumption of a constant expenditure share appears a close approximation of reality when it is calculated based on after tax income.

7 Conclusions

We have studied the role of federal income taxation on the misallocation of labor across geographical areas. More productive cities pay higher wages, and with progressive taxes, those workers also pay higher average taxes. Given perfect mobility, the tax schedule affects the incentives where to locate. Our objective has been to calculate the shape of the optimal tax schedule within a broad class.

Our findings are first, that the optimal tax schedule is progressive and not flat. The reason is that from a welfare viewpoint, what matters for the population allocation is not only productivity but also the cost of living. At flat taxes, the most productive cities become too expensive to live.

Second, 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.

Third, quantitatively, the output growth is substantial, between 2.6–8.8%. 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 between 1.5–4.9%. The welfare effects however are small, 0.008–0.067%. 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 Lemma 1

Proof. Equilibrium expenditure on housing is $p_j h_{ij} = \alpha \tilde{w}_{ij}$. After multiplying by l_{ij} and summing over i , we can write

$$p_j = \frac{\alpha \sum_i (l_{ij} \tilde{w}_{ij})}{H}, \quad (17)$$

since $\sum_i l_{ij} h_{ij} = H$.

Utility equalization for each skill type i in city j and the benchmark city 1 implies:

$$(1 + \varepsilon_j) \frac{\tilde{w}_{ij}}{p_j^\alpha} = \frac{\tilde{w}_{i1}}{p_1^\alpha}, \quad (18)$$

or

$$(1 + \varepsilon_j) \frac{\tilde{w}_{ij}}{(\sum_i l_{ij} \tilde{w}_{ij})^\alpha} = \frac{\tilde{w}_{i1}}{(\sum_i l_{i1} \tilde{w}_{i1})^\alpha}. \quad (19)$$

Further substituting we can write

$$\frac{\tilde{w}_{ij}}{(\sum_i l_{ij} \tilde{w}_{ij})^\alpha} = \lambda^{1-\alpha} \frac{w_{ij}^{1-\tau}}{(\sum_i l_{ij} w_{ij}^{1-\tau})^\alpha} \quad (20)$$

$$= (\lambda \gamma^{1-\tau})^{1-\alpha} \frac{A_{ij}^{1-\tau} l_{ij}^{-(1-\gamma)(1-\tau)}}{\left(\sum_i A_{ij}^{1-\tau} l_{ij}^{1-(1-\gamma)(1-\tau)} \right)^\alpha} \quad (21)$$

where we have used the tax schedule $\tilde{w} = \lambda w^{1-\tau}$ in the first equality, the first order condition of production $w_{ij} = A_{ij} \gamma l_{ij}^{\gamma-1}$ in the second. Utility equalization in (19) then implies

$$(1 + \varepsilon_j) \frac{A_{ij}^{1-\tau} l_{ij}^{-(1-\gamma)(1-\tau)}}{\left(\sum_i A_{ij}^{1-\tau} l_{ij}^{1-(1-\gamma)(1-\tau)} \right)^\alpha} = \frac{A_{i1}^{1-\tau} l_{i1}^{-(1-\gamma)(1-\tau)}}{\left(\sum_i A_{i1}^{1-\tau} l_{i1}^{1-(1-\gamma)(1-\tau)} \right)^\alpha}, \quad (22)$$

and solving for l_{ij} gives:

$$l_{ij} = \left(\frac{A_{ij}}{A_{i1}} \right)^{\frac{1}{1-\gamma}} \left[(1 + \varepsilon_j) \left(\frac{\sum_i A_{i1}^{1-\tau} l_{i1}^{1-(1-\gamma)(1-\tau)}}{\sum_i A_{ij}^{1-\tau} l_{ij}^{1-(1-\gamma)(1-\tau)}} \right)^\alpha \right]^{\frac{1}{(1-\gamma)(1-\tau)}} l_{i1}, \quad (23)$$

for all $i = 1, \dots, I$, and for all $j = 2, \dots, J$, and together with $\sum_{j=1}^J l_{ij} = L_i$, for all $j = 1, \dots, J$. This is a non-linear system of $I \times J$ equations in $I \times J$ unknowns l_{ij} . ■

The case where $\gamma = 1$

Then we can write the solution to l_{ij} explicitly as:

$$l_{ij} = \frac{\left[(1 + \varepsilon_j) \left(\frac{A_{ij}}{A_{i1}} \right)^{1-\tau} \right]^{\frac{1}{\alpha}} (\sum_i A_{i1}^{1-\tau} l_{i1}) - \sum_{k \neq i} A_{kj}^{1-\tau} l_{kj}}{A_{ij}^{1-\tau}}$$

General CES production function

Let output be given by

$$\left(\sum_i A_{ij} l_{ij}^\gamma \right)^\beta$$

then wages are equal to

$$w_{ij} = \beta \left(\sum_i A_{ij} l_{ij}^\gamma \right)^{\beta-1} A_{ij} \gamma l_{ij}^{\gamma-1}$$

Now we can write Utility equalization in (19) as

$$\begin{aligned} (1 + \varepsilon_j) \left(\sum_i A_{ij} l_{ij}^\gamma \right)^{(\beta-1)(1-\tau)(1-\alpha)} & \frac{A_{ij}^{1-\tau} l_{ij}^{-(1-\gamma)(1-\tau)}}{\left(\sum_i A_{ij}^{1-\tau} l_{ij}^{1-(1-\gamma)(1-\tau)} \right)^\alpha} \\ & = \left(\sum_i A_{i1} l_{i1}^\gamma \right)^{(\beta-1)(1-\tau)(1-\alpha)} \frac{A_{i1}^{1-\tau} l_{i1}^{-(1-\gamma)(1-\tau)}}{\left(\sum_i A_{i1}^{1-\tau} l_{i1}^{1-(1-\gamma)(1-\tau)} \right)^\alpha}, \end{aligned}$$

and solving for l_{ij} gives:

$$\begin{aligned} l_{ij} &= \left(\frac{A_{ij}}{A_{i1}} \right)^{\frac{1}{1-\gamma}} \left[(1 + \varepsilon_j) \left(\frac{\sum_i A_{ij} l_{ij}^\gamma}{\sum_i A_{i1} l_{i1}^\gamma} \right)^{(\beta-1)(1-\tau)(1-\alpha)} \left(\frac{\sum_i A_{i1}^{1-\tau} l_{i1}^{1-(1-\gamma)(1-\tau)}}{\sum_i A_{ij}^{1-\tau} l_{ij}^{1-(1-\gamma)(1-\tau)}} \right)^\alpha \right]^{\frac{1}{(1-\gamma)(1-\tau)}} l_{i1} \\ &= \left(\frac{A_{ij}}{A_{i1}} \right)^{\frac{1}{1-\gamma}} \left[(1 + \varepsilon_j) \left(\frac{\sum_i A_{i1} l_{i1}^\gamma}{\sum_i A_{ij} l_{ij}^\gamma} \right)^{(1-\beta)(1-\tau)(1-\alpha)} \left(\frac{\sum_i A_{i1}^{1-\tau} l_{i1}^{1-(1-\gamma)(1-\tau)}}{\sum_i A_{ij}^{1-\tau} l_{ij}^{1-(1-\gamma)(1-\tau)}} \right)^\alpha \right]^{\frac{1}{(1-\gamma)(1-\tau)}} l_{i1} \end{aligned} \quad (24)$$

Proof of Proposition 1

Proof. Aggregate output is:

$$Y = \frac{\mathcal{L}^\gamma}{(1 + K_2)^\gamma} (A_1 + A_2 K_2^\gamma),$$

and we can write

$$\begin{aligned}
\frac{\partial Y}{\partial \tau} &= \mathcal{S}^\gamma \left\{ -\gamma(1+K_2)^{-\gamma-1} (A_1 + A_2 K_2^\gamma) + (1+K_2)^{-\gamma} A_2 \gamma K_2^{\gamma-1} \right\} \frac{\partial K_2}{\partial \tau} \\
&= \frac{\mathcal{S}^\gamma \gamma}{(1+K_2)^{\gamma+1}} \left\{ -(A_1 + A_2 K_2^\gamma) + (1+K_2) A_2 K_2^{\gamma-1} \right\} \frac{\partial K_2}{\partial \tau} \\
&= \frac{\mathcal{S}^\gamma \gamma}{(1+K_2)^{\gamma+1}} \left\{ -A_1 + A_2 K_2^{\gamma-1} \right\} \frac{\partial K_2}{\partial \tau} \\
&= \frac{\mathcal{S}^\gamma \gamma}{(1+K_2)^{\gamma+1}} \left\{ -A_1 + A_2 \left(\frac{A_2}{A_1} \right)^{\frac{-(1-\gamma)(1-\alpha)(1-\tau)}{(1-\gamma)(1-\alpha)(1-\tau)+\alpha}} \right\} \frac{\partial K_2}{\partial \tau},
\end{aligned}$$

where

$$K_2 = \left(\frac{A_2}{A_1} \right)^X$$

where

$$\begin{aligned}
X &= \frac{(1-\alpha)(1-\tau)}{(1-\gamma)(1-\alpha)(1-\tau)+\alpha} \\
&= \frac{1}{1-\gamma} \left(1 - \frac{\alpha}{(1-\gamma)(1-\alpha)(1-\tau)+\alpha} \right)
\end{aligned}$$

and

$$\frac{\partial X}{\partial \tau} = \frac{-\alpha(1-\alpha)}{[(1-\gamma)(1-\alpha)(1-\tau)+\alpha]^2} < 0$$

and therefore

$$\frac{\partial K_2}{\partial \tau} = K_2 \log \left(\frac{A_2}{A_1} \right) \frac{\partial X}{\partial \tau}$$

Therefore, $\frac{\partial Y}{\partial \tau}$ is negative provided

$$\begin{aligned}
-A_1 + A_2 \left(\frac{A_2}{A_1} \right)^{-(1-\gamma)X} &> 0 \\
-A_1^{1-(1-\gamma)X} + A_2^{1-(1-\gamma)X} &> 0.
\end{aligned}$$

Since $A_1 < A_2$, this inequality is satisfied provided the exponent is positive, or

$$\begin{aligned}\frac{-(1-\gamma)(1-\alpha)(1-\tau)}{(1-\gamma)(1-\alpha)(1-\tau)+\alpha} + 1 &> 0 \\ \frac{\alpha}{(1-\gamma)(1-\alpha)(1-\tau)+\alpha} &> 0 \\ (1-\gamma)(1-\alpha)(1-\tau)+\alpha &> 0.\end{aligned}$$

Since $\tau \leq 1$, this is always satisfied as long as there are weakly decreasing returns to scale ($\gamma \leq 1$). When there are increasing returns ($\gamma > 1$), then this is satisfied provided

$$\tau > 1 - \frac{\alpha}{(\gamma-1)(1-\alpha)}.$$

When increasing returns are too strong, eventually all workers live in the most productive city. Hence there is no interior solution to the location decision and output is not affected by taxes.

Observe that in the neighborhood of $\gamma = 1$, the RHS is $-\infty$, so this is always satisfied. In fact, the RHS is increasing in γ and is equal to zero when $\gamma = \frac{1}{1-\alpha}$, which for our calibrated parameter of $\alpha = 0.24$ implies $\gamma = 1.32$. Even with moderate increasing returns therefore, output is decreasing in τ .

Finally, there is no optimal tax τ that can achieve output maximization. To see this, observe that the equilibrium allocation and the output maximizing allocation coincide when

$$l_1 = \frac{\mathcal{L}}{1+K_2} = \frac{\mathcal{L}}{1+\left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\gamma}}}$$

which implies $\frac{\alpha}{1-\gamma} = 0$. This is satisfied only when $\alpha = 0$. Generically therefore, no tax achieves first best. ■

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.⁹ Suppose $w > 2$ and $w_{top} < 2$, i.e.

⁹Top marginal tax rate is taken from <http://www.oecd.org/tax/tax-policy/oecd-tax-database.htm>, Table I.7.

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

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

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

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

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.
- 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. A., AND F. ORTALO-MAGNÉ (2011): “Household Expenditures, Wages, Rents,” *Review of Economic Dynamics*, 14(2), 248–261.
- ECKHOUT, J., R. PINHEIRO, AND K. SCHMIDHEINY (2013): “Spatial Sorting,” University College London mimeo.
- 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 (2013): “Income Taxation of U.S. Households: Facts and Parametric Estimates,” Barcelona GSE Working Paper No. 705.
- 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.
- KAHN, M. E. (2010): “New Evidence on Trends in the Cost of Urban Agglomeration,” in *Agglomeration Economics*, NBER Chapters, pp. 339–354. National Bureau of Economic Research, Inc.
- KAPLOW, L. (1995): “Regional Cost-of-Living Adjustments in Tax/Transfer Schemes,” *Tax Law Review*, 51, 175–198.
- 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.
- TIEBOUT, C. (1956): “A Pure Theory of Local Expenditures,” *Journal of Political Economy*, 64(5), 416–424.

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