Joint Discussion Paper
Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 13-2020

Sebastian G. Kessing, Vilen Lipatov and J. Malte Zoubek

Optimal Taxation under Regional Inequality

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Optimal Taxation under Regional Inequality*

Sebastian G. Kessing[a] Vilen Lipatov[b] J. Malte Zoubek[c]

March 23, 2020

Abstract

We study how regional productivity differences and labor mobility shape optimal Mirrleesian tax-transfer schemes. When tax schedules are not allowed to differ across regions, productivity-enhancing inter-regional migration exerts a downward pressure on optimal marginal tax rates. When regionally differentiated taxation is allowed, marginal tax rates in high-(low-)productivity regions should be corrected downwards (upwards) relative to the benchmark without migration. Simulations of the productivity differences between metropolitan and other areas of the US indicate that migration affects the optimal tax-transfer schedule more strongly in the regionally differentiated rather than in the undifferentiated case.

JEL classification: H11, J45, R12

Keywords: Optimal taxation, place-based redistribution, regional inequality, migration, multidimensional screening, delayed optimal control

*We would like to thank David Agrawal, David Albouy, Spencer Bastani, Mireille Chiroleu-Assouline, Bill Hoyt, Christian Gollier, Laurence Jacquet, Luca Micheletto, Casey G. Rothschild, and two anonymous referees for their comments. Kessing and Zoubek gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG), project No. KE 1693/3-1 and KE 1693/3-2.

[a]University of Siegen and CESifo, kessing@vwl.wiwi.uni-siegen.de (corresponding author)

[b]Compass Lexecon Brussels and CESifo, vlipatov@compasslexecon.com

[c]University of Siegen, malte.zoubek@uni-siegen.de

1 Introduction

Regional productivity differences are large in many countries. In the US, real GDP per capita in the Mideast was 45% higher than that in the Southeast in 2018 (BEA 2020). Real GDP per capita in the Italian northwest was 90% higher than in the South of Italy in 2018 (ISTAT 2020). The spatial dispersion of wages and incomes is well documented, and the underlying causes are still subject to debate, see Barro and Sala-i-Martin (1991), Ciccone and Hall (1996), Kanbur and Venables (2005), Acemoglu and Dell (2010), and Young (2013).

Such productivity differences provide a chance of "moving to opportunity." Indeed, income gains play a significant role in the individual migration decision; see Kennan and Walker (2011). Centralized redistribution schemes, such as a federal income tax, reduce migration incentives. This is because an individual who migrates from a low- to a high-productivity region, and earns a higher income as a result, has to give up a share of the income gain to taxes. Under such circumstances, redistributive policies distort labor supply and migration decisions, so that the equity-efficiency trade-off becomes more complex. We develop a conceptual framework to analyze the implications of this additional distortion for an optimal tax-transfer policy and assess its quantitative importance.

We propose a multi-dimensional optimal-taxation model which features heterogeneity in (i) innate productivity; (ii) the cost of moving; (iii) the initial region of residence. Each individual resides in one of two regions and has to make a migration decision and a labor supply decision. Our key innovation is the introduction of the productivity-changing nature of migration into the optimal tax analysis. The actual productivity of individuals of any given innate productivity is location-dependent, so that individuals can increase their productivity by migrating from a low- to a high-productivity region. Thus, the extensive migration margin affects the intensive labor supply margin via the change in productivity and via the change in the marginal tax rate.

This framework allows us to determine the optimal country-level tax schedule as a function of the government's redistributive preferences, the observed regional earnings distributions, the labor supply elasticity, and the distribution of migration costs. Optimal marginal tax rates tend to be below the benchmark without migration, since the decision to migrate to a higher productivity region bears a fiscal externality: for the same tax rate, higher income increases the taxes paid, but also, the tax rate increases if marginal tax rates are positive. This result is preserved if potential inter-regional cost-of-living differences are taken into account, since the size of the fiscal externality does not depend on them. Moreover, for some productivity distributions, the migration opportunity makes negative marginal tax rates optimal. This is reminiscent of optimal taxation studies with an extensive participation margin, e.g., Saez (2002), Diamond (1980), and Jacquet et al. (2013).

We contribute to the theory of optimal taxation in that we endogenize individual productivity through the extensive margin, similar to Gomes et al. (2018). This is a useful extension to the class of multidimensional screening models, originally discussed by Rochet and Choné (1998) and Armstrong (1996), and further developed to study the taxation of couples by Kleven et al. (2009). We argue that variants of these models with systematic endogenous productivity differences can be fruitfully studied using the delayed optimal-control approach as formally analyzed by Göllmann et al. (2008) in its entire generality.

Variations of our framework are suitable to address a range of multidimensional screening problems where an extensive margin directly affects agents' productivity, and thus the intensive margin. The decision to participate in the labor market, for example, affects productivity, because non-participation tends to depreciate human capital. Similarly, the decision to switch to a better job also affects actual individual productivity. The same holds true for discrete education or personal decisions, such as the decision to attend college or the decision to have children. All these examples represent empirically relevant dimensions of the labor supply decision. Our framework and the proposed delayed optimal-control technique provide a readily applicable solution for characterizing optimal tax schedules and applying them to data in such cases.[1]

[1] Our model can be seen as a three-dimensional optimal tax model with migration costs, pro-

We also study regionally differentiated tax-transfer schemes, i.e., place-based redistribution. To the extent that such schemes are explicit, they are often difficult to enforce in practice given the challenge to monitor the actual place of residence of individuals, and they may also be challenged on the grounds of the violation of horizontal equity. Despite these caveats, regional differentiation of labor income taxation can be an element of real-world tax systems. From 1971 to 1994, the German tax-transfer system, for example, treated residents in West Berlin differently from people in the rest of the country. Further, a nominally undifferentiated country-level tax scheme amounts to regionally differentiated taxation in real terms, due to the cost-of-living differences; see Albouy (2009).

The idea that the government's information problem can be relaxed via additional observable characteristics (tags) that are correlated with unobservable productivity goes back to Akerlof (1978) and has recently been discussed intensively in optimal-taxation; see Immonen et al. (1998), Boadway and Pestieau (2006), Blomquist and Micheletto (2008), Cremer et al. (2010), Mankiw and Weinzierl (2010), Weinzierl (2011), and Bastani et al. (2013). We consider regions as endogenous tags, which relax information constraints but can also encourage productivity-enhancing migration.

Our main take from the analysis of differentiated taxation is that the migration opportunity exerts downward (upward) pressure on marginal tax rates in the high- (low-) productivity region. Intuitively, the larger the potential fiscal gains from working in the high-productivity region are, the more the government distorts labor supply in the low-productivity region and the less it distorts labor supply in the high-productivity region.

To illustrate our theoretical analysis, we apply our framework to US data. We focus on the productivity difference between large metropolitan and other areas. The simulations of optimal undifferentiated tax schedules confirm that the government chooses a higher marginal tax rate when migration is ignored

ductivity in the low productivity region, and productivity in the high productivity region being the three dimensions of unobservable individual heterogeneity. Delayed optimal control exploits the systematic regional productivity differences to reduce the problem to a two-dimensional one.

relative to the situation when it is taken into account. Quantitatively, however, the reduction of marginal tax rates is rather small. The deviations from the no-migration benchmark are substantially more pronounced in the simulations of differentiated taxation. These show that marginal tax rates fall in the high-productivity region and rise in the low-productivity region.

The next section discusses the related literature. Section 3 introduces the theoretical framework, and we derive our theoretical results for unified and differentiated taxation in sections 4 and 5, respectively. We present a numerically calibrated illustrative simulation based on US micro data in Section 6, provide a discussion in Section 7, and leave most proofs and the technical derivations to the Appendices and the Supplement.

2 Related literature

Following Krugman (1991), a vast literature on the extent and persistence of regional inequality, as well as on its origins and propagating mechanisms, has emerged (see Breinlich et al. 2014, and Behrens and Nicoud, 2015, for an overview). Within the neoclassical setting, explanations of regional disparities reach from natural differences (mountains, river access, harbors, climate etc.) to differences in local infrastructure and the quality of local governance. Furthermore, the new economic geography has pointed out the role of local agglomeration and location externalities. These include, in particular, labor market pooling effects, the emergence of specialized intermediate inputs, and technological spillovers. These externalities arise from density and industry specialization in particular locations (Fujita et al. 1999; Fujita and Thisse 2002). Moreover, regional productivity is affected by different sorting mechanisms, i.e., heterogeneous firms and workers take heterogeneous location decisions. Finally, selection matters for location-dependent productivity. Local conditions determine decisions to become an entrepreneur, and they affect which businesses survive.

We take regional differences as given and study their consequences for optimal redistribution. These normative implications of productivity-enhancing internal migration have, to the best of our knowledge, not been studied to date.[2] This stands in contrast to the emigration of high-income earners to low-tax countries or the immigration of welfare recipients from less generous jurisdictions. Studies addressing international migration or migration between autonomously tax-setting jurisdictions include Mirrlees (1982), Wildasin (1991), Wilson (1992), Simula and Trannoy (2011), Lehmann et al. (2014), and Lipatov and Weichenrieder (2015), among others.

Our analysis relates to a class of two-dimensional screening models that have been recently used to analyze a range of tax policy questions. Lehmann et al. (2014) combine the intensive labor supply margin with an extensive migration margin. Their focus is on independent governments competing for internationally mobile high-productivity individuals, and where productivity is not location-dependent. Gordon and Cullen (2012) use an optimal-taxation approach to study interregional migration in a model with several states. They focus on the assignment problem of whether redistribution should be carried out at the national or the subnational level. Blomquist and Micheletto (2009) find advantages of regional in-kind transfers in the presence of a federal income tax. Jacquet et al. (2013) also study a two-dimensional optimal-taxation model but focus on participation. Rothschild and Scheuer (2013, 2016) study taxation of rent-seeking activities and optimal taxation in the Roy model. Wages are endogenously determined, either by the total labor supply in a given sector or by total rent-seeking activities. Scheuer (2014) studies entrepreneurial taxation with an endogenous decision of whether to become an entrepreneur. Finally, Best and Kleven (2013) consider a dynamic setting where productivity depends on previous intensive labor supply decisions.[3]

Our framework owes much to Kleven et al. (2006, 2009), who combine households' intensive labor supply with the decision to become a double-earner household. However, our analysis differs in several important ways. First, we consider

[2] Eeckhout and Guner (2017) study how a progressive federal income tax affects the spatial allocation and also consider regionally differentiated taxation, but they do not use an optimal taxation framework and do not consider the interaction of labor supply and migration.

[3] Golosov et al. (2014) provide a general framework to study multidimensional optimal taxation using a variational approach.

individuals and not couples whose incomes may be taxed separately. Secondly, individuals originally reside in different regions, so they differ not only in the switching costs but also in the group they originally belong to. Finally and most importantly, we endogenize productivity through the extensive margin.

While our analysis closely relates to Gomes et al. (2018), who study optimal taxation with sector-specific productivity, there are important differences to their work. Gomes et al. (2018) focus on a sectoral choice that motivates a setting where individuals differ in productivity in each of two sectors. Taxation in either sector directly affects the switching costs along the extensive margin. We focus on a location choice that motivates a setting where individuals differ in productivity in two regions and also in individual moving costs that are not related to income. Taxation does not affect such costs of moving.

In the context of sectoral choice, it seems natural that individuals are not ex ante affiliated to any of the two sectors and can decide without additional cost which sector to work in. In the context of location choice, individuals are ex ante situated in some region – the act of migration may involve substantial disutility per se. Other than that, the model of Gomes et al. (2018) is richer than ours in the sense that it allows for a joint distribution of productivity whereas we collapse this distribution into one dimension by assuming a functional relation between productivities in two regions. That allows them to study skill transferability and skill intensity motives for differential taxation, while these are fixed exogenously in our model.

With idiosyncratic productivity differences that feature Gomes et al.'s (2018) analysis, there is endogenous sorting by productivity with respect to who works in which sector. With undifferentiated taxation, this implies that the sector choice remains undistorted at the margin, since the marginal individuals pay the same taxes in the two sectors. In our approach with common productivity differences (and no idiosyncratic ones), the marginal individuals are indifferent between moving and staying, but they pay different taxes in the two regions despite the tax schedule being the same for both regions. The extensive margin decision is therefore distorted, and we analyze how the optimal tax schedule should address these distortions. We also show how the use of delayed optimal control techniques allows us to solve for the optimal tax schedule in such a setting.

3 The framework

We consider two sources of heterogeneity across workers: *innate* productivity n and migration costs q. These original individual characteristics are distributed over $[n_{\min}, n_{\max}] \times [0, +\infty)$, and the government cannot observe either of them. There are two regions, $i = A, B$, with the total population normalized to two. Originally, half of the population reside in each region, but the endogenous migration decisions of individuals change these population shares. Our key assumption is that the regions differ in their productivity. An individual's *actual* or *realized* productivity n_i is a function of her innate productivity and her region of residence: $n_i = \omega(n, i) = \omega_i(n)$, where ω_i is strictly increasing in n. We normalize $n_A = \omega_A(n) = n$. Accordingly, the function $n_B = \omega_B(n) = \omega(n)$ not only assigns the actual productivity to all original residents of region B, but also indicates the transformation of productivity for individuals who migrate from A to B. We further assume that region B is the more productive region, so that $\forall n, \omega(n) > n$. Innate productivity is distributed in each region i according to the unconditional probability distribution $f(n)$ on $[n_{\min}, n_{\max}]$.[4] As in most of the optimal-taxation literature, we treat wages as exogenous and independent of individual labor supply and aggregate migration decisions. The empirical evidence supports the view that, for sufficiently large regions, the effects of internal migration on wages are rather small; see, for the US, Boustan et al. (2010), and for evidence from the German reunification, D'Amuri et al. (2010) and Frank (2009).[5]

Following Diamond (1998), we use preferences that are separable in consump-

[4]It is straightforward to extend the analysis to the case in which regions also differ in their distribution of innate productivity. Similarly, we could allow for negative migration costs for some subset of individuals at each innate productivity level without affecting the results qualitatively. The latter can generate migration in both directions. For clarity, we abstract from these aspects.

[5]For similar findings in the case of the immigration of foreigners see Borjas (1994) and Ottaviano and Peri (2007, 2008). We discuss endogenous wages in Section 7.

tion and labor. The utility function of a worker of type (n, q) is similar to the formulation in Kleven et al. (2009) but depends on the region of residence:

$$u(c, z, l) = c_i - n_i h\left(\frac{z_i}{n_i}\right) - ql, \tag{1}$$

where l is an indicator variable that takes the value 1 in the case of migration.[6] The function $h(\cdot)$ is increasing, convex, and twice differentiable. It is normalized so that $h'(1) = 1$ and $h(0) = 0$. The other variables have standard interpretations. Consumption c_i equals gross income z_i minus taxes T_i, the latter depending on gross income: $c_i = z_i - T_i(z_i)$.

Each individual chooses l and z_i to maximize (1) for a given tax schedule, i.e., she decides whether to move or not and determines her gross earnings, given that she resides in region i. The first-order condition for gross earnings is

$$h'\left(\frac{z_i}{n_i}\right) = 1 - \tau_i(z_i), \tag{2}$$

where τ_i is the marginal tax rate. Accordingly, n_i can be interpreted as potential income, in that individuals facing a marginal tax rate of zero would realize this level of gross earnings. The elasticity of gross earnings with respect to net-of tax-rate is $\varepsilon_i \equiv \frac{1-\tau_i}{z_i}\frac{\partial z_i}{\partial(1-\tau_i)} = \frac{n_i h'\left(\frac{z_i}{n_i}\right)}{z_i h''\left(\frac{z_i}{n_i}\right)}$, and is a function of gross earnings and the region of residence. Finally, we require the following property:

Assumption The function $x \to \frac{1-h'(x)}{xh''(x)}$ is decreasing.

Consider now the migration decision. We denote by $p(q|\, n)$ the density of q conditional on n, and by $P(q|\, n)$ the cumulated distribution of q conditional on n. Conditional on residing in region i, the individuals' choice of gross earnings is

[6] Note that in this and following expressions, we usually drop the argument of $z_i(n_i)$ for parsimony, but z_i is understood as the income obtained by the individual of innate productivity n in territory i. Moreover, migration costs can alternatively be modeled as home attachment, i.e., in a way that the decision to move implies the loss of an endowment which is kept by the remaining individuals (i.e., a social network), and the unobservable heterogeneity corresponds to the individual valuation of this endowment. However, the optimal tax formulae we derive below are sufficiently general to encompass this case, see Kessing et al. (2015), and we use the simple formulation (1) for clarity of exposition.

determined by (2), which allows us to define the indirect utility conditional on the place of residence and net of the migration costs as

$$V_i\left(n_i\right) = z_i - T_i\left(z_i\right) - n_i h\left(\frac{z_i}{n_i}\right).$$

Individuals will move from i to j, $j = A, B$, $i \neq j$, whenever their migration costs are below the net gain from moving, so that $\bar{q}_i(n) \equiv \max\left\{V_j\left(n_j\right) - V_i\left(n_i\right), 0\right\}$ is the critical level of migration costs that determines the actual number of migrants for any innate productivity level.

3.1 The government's optimal-tax problem

The government wants to maximize the social welfare function

$$\sum_{i=A,B} \int_{n_{\min}}^{n_{\max}} \left[\int_{\bar{q}_i}^{+\infty} \Psi\left(V_i\left(n_i\right)\right) p(q|n) dq + \int_0^{\bar{q}_{-i}} \Psi\left(V_i\left(n_i\right) - q\right) p(q|n) dq\right] f(n) dn, \quad (3)$$

where $\Psi(.)$ is a concave and increasing transformation of individual utilities.[7] Denoting by E the exogenous expenditure requirement, it needs to respect the budget constraint

$$\sum_i \int_{n_{\min}}^{n_{\max}} \int_0^{+\infty} T_i(z_i) p\left(q|\, n\right) f(n) dq dn \geq E. \quad (4)$$

Moreover, the government's tax schedule needs to be incentive-compatible. Denoting by a dot above a variable its derivative with respect to n, this implies

$$\dot{V}_i\left(n\right) = \left[-h\left(\frac{z_i}{n_i}\right) + \frac{z_i}{n_i} h'\left(\frac{z_i}{n_i}\right)\right] \omega_i'\left(n\right) \geq 0. \quad (5)$$

Moreover, in the case of nondifferentiated taxation, $T_A\left(z\right) = T_B\left(z\right)$. We show in the Supplement that a path for z_A and z_B can be truthfully implemented by the government using a non-linear tax schedule.

Let $\lambda > 0$ be the multiplier associated with the budget constraint (4). The

[7] This means that we restrict our attention to a concave utilitarian social planner.

government's redistributive tastes may be represented by region-dependent social marginal welfare weights. In terms of income, our welfare weights take the form

$$g_i(z) = \frac{\Psi'\left(V_i\left(z\right)\right)\left(1 - P(\bar{q}_i|z)\right) + \int_0^{\bar{q}_j} \Psi'\left(V_i\left(z\right) - q\right) p(q|z) dq}{\lambda\left(1 + P(\bar{q}_j|z) - P(\bar{q}_i|z)\right)},$$

where $\bar{q}_i\left(z\right) \equiv \max\left\{V_j\left(z\right) - V_i\left(z\right), 0\right\}$.

4 Optimal unified taxation

4.1 Analysis in the general setting

We first investigate the optimal nondifferentiated tax-transfer system. The government maximizes (3) subject to (4) and (5) through its choice of $T(z)$. This problem formally amounts to a delayed optimal-control problem as analyzed by Göllmann et al. (2008) in its entire generality. In our model, though, the delay is a nonfixed lag, given that we do not require the productivity gain from moving to be constant. The necessary conditions for optimal control in such a setting are presented in Abdeljawad et al. (2009). We describe in Appendix A how the delayed optimal-control approach can be applied to solve the optimal-taxation problem, and we derive all our results rigorously there with further technical details provided in the Supplement. Below, however, we follow the intuitive perturbation approach pioneered by Piketty (1997) and Saez (2001) to derive the optimal tax scheme. This heuristic derivation disentangles the economic forces that determine the shape of the optimal marginal tax rate schedule, including the effects generated by productivity-enhancing migration.

We denote the endogenous distribution of gross incomes in both regions by $v_i\left(z_i\right)$, and we denote by k the endogenously defined function that maps gross income in the low-productivity region to the gross income this individual would earn in the high-productivity region, given his innate productivity and the respective tax treatment, i.e., $z_B = k(z_A)$.[8] We consider an increase in taxes for all individ-

[8]In terms of our previous formulation, an individual of ability n receives gross income $z_A = z(n)$ in region A and gross income $z_B = z(\omega(n))$ in region B, where this notation abstracts from

uals above gross income z. The increase is engineered through an increase in the marginal tax rate $d\tau$ in the small band $(z, z+dz)$, such that for all individuals with gross earnings above z the tax payments increase by $dzd\tau$. This tax increase gives rise to three different effects.

Revenue effect Taxpayers in both regions with gross incomes above z pay additional taxes of $dzd\tau$. The net welfare effect for an affected individual in region i with gross earnings z' is $dzd\tau\left(1-g_i\left(z'\right)\right)$, and the total effect is then

$$R = dzd\tau \int_z^\infty \left\{\left[1-g_A\left(z'\right)\right] v_A\left(z'\right) s_A\left(z'\right) + \left[1-g_B\left(z'\right)\right] v_B\left(z'\right) s_B\left(z'\right)\right\} dz',$$

where $s_A(z) \equiv 1 - P\left(\bar{q}_A \middle| z\right)$ and $s_B(z) \equiv 1 + P\left(\bar{q}_A \middle| k^{-1}(z)\right)$.

Behavioral effect Individuals in the band $(z, z+dz)$ change their labor supply in response to the increase in the marginal tax rate. Given that $\varepsilon \equiv \frac{1-\tau}{z}\frac{dz}{d(1-\tau)}$, each individual in the band reduces its income by $-d\tau\varepsilon\frac{z}{1-\tau}$. There are approximately $dz\left[v_A(z)s_A(z) + v_B(z)s_B(z)\right]$ of these individuals. To take the non-linearity of the tax schedule into account, we correct this with the "feedback" effect $1+\frac{\tau'}{1-\tau}z\varepsilon$ following Golosov et al. (2014). The total effect on tax revenue is

$$L = -dzd\tau\frac{\tau z\varepsilon}{1-\tau}\left[v_A(z)s_A(z) + v_B(z)s_B(z)\right]\left(1+\frac{\tau'}{1-\tau}z\varepsilon\right).$$

Migration effect An increase in taxes for all individuals above gross income z does not affect the migration decision of individuals with gross income $z'_A \geq z$, and accordingly also $z'_B > z$, since the tax increase affects them in both regions alike. The same holds true for all individuals for which $z'_B = k(z'_A) < z$ and accordingly $z'_A = k^{-1}(z'_B) < z$. However, as illustrated in Figure 1, for all individuals for which $z'_A < z$ and $z'_B \geq z$ the migration decision is negatively affected. In this range, all individuals whose cost of moving is between $\bar{q} - dzd\tau$ and $\bar{q}$ will now decide not to migrate. There are $p\left(\bar{q} \middle| z\right) v_A(z)dzd\tau$ affected individuals at any concerned level of income, with a resulting tax effect of $T_A(z) - T_B(k(z))$ for each of them. The

the fact that the gross income also depends on the tax schedule.

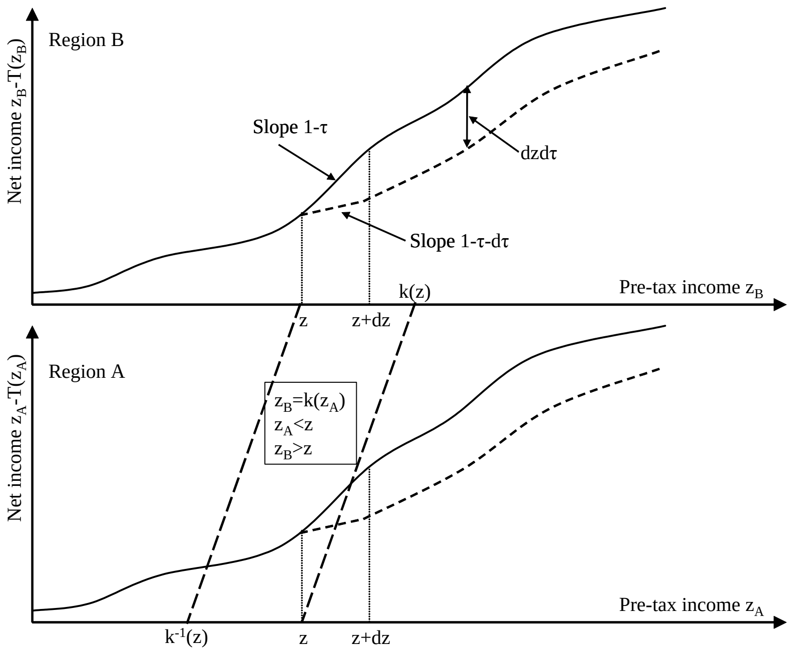

Figure 1: The migration effect comes into play for individuals for which $z'_A < z$ and $z'_B \geq z$.

total migration effect is thus

$$M = dzd\tau \int_{\tilde{z}}^{z} \left[T(z') - T(k(z'))\right] p\left(\bar{q}|\, z'\right) v_A(z')dz',$$

where $\tilde{z} \equiv k^{-1}(z)$. Note that there is an endogenous effect on the income distribution in each region. This effect does not come into play explicitly here, since we express the effects in terms of the posterior distribution. The three effects must balance out in the optimum: $R + L + M = 0$. From this we have our first result.

Proposition 1 *The optimal unified tax schedule is characterized by*

$$\begin{aligned} \frac{\tau}{1-\tau} &= \bar{A}(z)\bar{B}(z)\left[\bar{C}(z) + \bar{D}(z)\right], \ \textit{where} \qquad (6) \\ \bar{A}(z) &\equiv \frac{1}{\varepsilon}\frac{1-\tau}{1-\tau+\tau' z\varepsilon}, \qquad \bar{B}(z) \equiv \frac{1}{z\left(v_A(z)s_A(z) + v_B(z)s_B(z)\right)}, \\ \bar{C}(z) &\equiv \int_z^\infty \left\{\left[1 - g_A\left(z'\right)\right] v_A\left(z'\right) s_A + \left[1 - g_B\left(z'\right)\right] v_B\left(z'\right) s_B\right\} dz', \\ \bar{D}(z) &\equiv \int_{\tilde{z}}^{z} \left[T(z') - T(k(z'))\right] p\left(\bar{q}|\, z'\right) v_A(z')dz'. \end{aligned}$$

Proof. This follows from the exposition above. The equivalence to the optimal tax formula formally derived by using the delayed optimal-control technique is presented in Appendix A and the Supplement. ■

It is straightforward to compare the result to the benchmark without migration. The optimal tax schedule then follows the usual Diamond (1998) and Saez (2001) results for the earnings distribution in the entire country without a migration effect. In this case, optimal marginal tax rates are determined by

$$\frac{\tau}{1-\tau} = \bar{A}(z)\bar{B}(z)\bar{C}(z). \tag{7}$$

With $\bar{D}(z) < 0$, the disincentive effects of higher tax rates on productivity-increasing mobility tend to reduce marginal tax rates, but note that $\bar{B}(z)$ and $\bar{C}(z)$ are endogenously determined by the migration flows, so that (6) and (7) cannot be directly compared in general. To make the result more formal, consider the following proposition:

Proposition 2 *If marginal tax rates are positive, a government taking into account the effect of taxes on the migration decision sets lower optimal marginal tax rates than a government that faces the same posterior income distribution generated by migration but disregards the migration decision.*

Proof. See Appendix B. ■

With positive marginal tax rates, any individual moving to the high productivity region increases tax revenues. Migration entails a positive fiscal externality. Individuals do not take this externality into account in their migration decision. The government optimally responds to this by reducing marginal tax rates relative to the optimal level without migration. This rate reduction amounts to a migration subsidy to correct the fiscal externality. The size of the optimal corrective subsidy depends on the size of the migration effect at any income level. Whenever $\bar{D} < 0$ for any given level of gross income z, marginal tax rates are lower when, under the same posterior distribution, the government takes migration into account rather than disregarding it.

Next, we make a remark about the welfare comparison in the unified taxation case, highlighting the desirability of productivity-enhancing migration.

Remark The welfare achieved with unified taxation in the no-migration case is not higher than the welfare achieved with migration.

Proof. Start from the welfare-maximizing tax schedule without migration. Migration brings a Pareto improvement, because individuals move only if they will find themselves better off. With migration to the richer region only, the government budget constraint will not be violated, since the original optimal Mirrleesian tax schedule is nondecreasing in income without migration. Thus, under the same tax schedule welfare cannot decrease after the introduction of migration. Moreover, the government will change the tax schedule only if it brings a further increase in welfare. Thus, welfare with migration cannot be lower than welfare without migration. ■

Another direct implication of the optimal unified taxation formula (6) is stated in the following proposition.

Proposition 3 *Optimal marginal tax rates can be negative.*

Proof. For $\bar{D}(z) < 0$, it is possible that $\bar{C}(z) + \bar{D}(z) < 0$, and thus $\tau < 0$. ■

Negative marginal tax rates arise whenever $\bar{D} > \bar{C}$. This result parallels studies of optimal taxation with an extensive participation margin.

4.2 Exogenously fixed number of working hours

For the intuition of Proposition 3, consider a pure extensive margin version of our model. We retain our assumptions about the regional distribution of productivity and migration costs, but assume that everyone works a fixed amount of hours. Preferences are given by

$$u(c, z, l) = c_i - k_i - ql, \tag{8}$$

where k_i is the disutility of working, assumed to be sufficiently small, so that all individuals work. Without an intensive margin, we do not have to distinguish between gross income z and potential income n. An individual of innate ability n earns gross income n in region A. This individual has realized productivity $\omega(n)$, and earns gross income $\omega(n)$, in region B. The only decision individuals take is whether to reside in region A or in region B. An individual moves from A to B, if $\omega(n) - T(\omega(n)) - [n - T(n)] \geq q$, which again defines a critical level of migration cost $\bar{q}_i(n) \equiv \max\{V_j(n_j) - V_i(n_i), 0\}$ which determines the number of migrants for any innate productivity level.

Assume that there is only migration from A to B. A perturbation which increases taxation by dT in the income range between n and $\omega(n)$ generates a revenue effect

$$\begin{aligned}\hat{R} &= dT \int_n^{\omega(n)} \{[1 - g_A(n')] f(n') (1 - P(\bar{q}_A | n')) \\ &\quad + [1 - g_B(n')] f(n') \left(1 + P\left(\bar{q}_A | \omega^{-1}(n')\right)\right)\} dn'.\end{aligned}$$

Moreover, it generates two migration effects. It reduces the number of migrants with productivity $n_A \in [\omega^{-1}(n), n)$, which reduces tax revenues, but increases the number of migrants with productivity $n_A \in [n, \omega(n)]$, which increases tax revenues. The total migration effect is

$$\begin{aligned}\hat{M} &= dT[\int_{\omega^{-1}(n)}^{n} p(\bar{q}_A | n') f(n') [T(n') - T(\omega(n'))] dn' \\ &\quad + \int_n^{\omega(n)} p(\bar{q}_A | n') f(n') [T(\omega(n')) - T(n')] dn'].\end{aligned}$$

Marginal tax rates can be negative for sufficiently high redistributive tastes ($\hat{R} < 0$), if the migration reducing effect (on those with $n_A \in [\omega^{-1}(n), n)]$) sufficiently outweighs the migration increasing effect (on those with $n_A \in [n, \omega(n)]$).

The intuition parallels Diamond (1980) and Saez (2002). For contradiction, assume that, for some low income individuals, $T(n) - T(\omega^{-1}(n)) > 0$ and $\tau(n) \geq 0$ for $n \in [\omega^{-1}(n), \omega(n)]$. Consider a policy of reducing $T(.)$ marginally by dT on the interval $[n, \omega(n)]$. This makes individuals with an income between n and

$\omega(n)$ better off and represents a welfare improvement if their average marginal welfare weight is above one. Moreover, the policy change generates additional tax revenues from the additional migrants just below the affected band, given that $T(n) - T(\omega^{-1}(n)) > 0$ and $\tau(n) \geq 0$ for $n \in [\omega^{-1}(n), \omega(n)]$, but lower tax revenues from the inhibited migrants within the band. If the latter tax revenue effect is sufficiently small, the original policy of a positive marginal tax cannot have been optimal.

The parallel to Diamond (1980) and Saez (2002), who focus on the participation decision, can also be seen by considering a degenerated ability distribution in region A. If the realized productivity distribution in region A is such that $n_A = 0$ for all individuals, irrespective of their innate productivity, but the realized productivity in region B is continuously distributed and equal to $\omega(n) > 0$, our framework corresponds to that of the participation margin in Diamond (1980).[9]

5 Optimal differentiated taxation

We now return to the model of location-dependent productivity that allows choice along both intensive and extensive margins, but consider the possibility that the central government can choose differentiated tax schedules, i.e., place-based taxation. If there were regional productivity differences but no migration, this setting would correspond to the analysis of an optimal tax scheme with tagging on the region of residence. However, we continue to assume that migration between the regions is possible, that productivity is location-dependent, and that individuals are heterogeneous with respect to their migration costs, which are unobservable by the government. Again, we employ the perturbation approach and delegate the formal proofs to Appendix B.

We first study the optimal tax schedule in the low-productivity region. Consider an increase of taxes in region A for all individuals above gross income z_A.

[9] Increasing the tax liability by dT over some band $[n, \omega(n)]$ will only generate a single migration effect of reduced migration. There is no migration-increasing effect, since the income range $[n, \omega(n)]$ is empty in region A. Thus, in this case, whenever the revenue effect $\hat{R}$ is negative because of sufficiently large welfare weights, negative marginal tax rates will be optimal.

The increase is engineered through an increase in the marginal tax rate $d\tau_A$ in the small band $(z_A, z_A + dz_A)$, such that all individuals with gross earnings above z_A increase their tax payments by $dz_A d\tau_A$. This generates three effects.

Revenue effect: All taxpayers in A with incomes above z_A pay additional taxes of $dz_A d\tau_A$. The net welfare effect of this tax payment for an individual with gross earnings z'_A is given by $dz_A d\tau_A (1 - g_A(z'_A))$, and the total effect is

$$R_A = dz_A d\tau_A \int_{z_A}^{\infty} [1 - g_A(z'_A)]\, v_A(z'_A)\, s_A(z'_A)\, dz'_A.$$

Behavioral effect: Individuals in the band $(z_A, z_A + dz_A)$ change their labor supply in response to the increase in the marginal tax rate. Given that $\varepsilon \equiv \frac{1-\tau_i}{z_i}\frac{dz_i}{d(1-\tau_i)}$, each individual in the band reduces its income by $-d\tau_A \varepsilon \frac{z_A}{1-\tau_A}$. There are approximately $dz_A v_A(z_A) s_A(z_A)$ of these individuals, so that the total effect on tax revenue, taking into account the nonlinearity correction term $1 + \frac{\tau'_A z_A \varepsilon_A}{1-\tau_A}$ is

$$L_A = -dz_A d\tau_A \varepsilon \frac{\tau_A}{1-\tau_A} z_A v_A(z_A) s_A(z_A) \left(1 + \frac{\tau'_A z_A \varepsilon_A}{1-\tau_A}\right).$$

Migration effect: An increase in taxes for individuals above gross income z_A affects the migration decision of individuals with gross income in region A above this level. At any income level $z \geq z_A$ individuals whose cost of moving is between $\bar{q}$ and $\bar{q} + dz_A d\tau_A$ will now decide to migrate. There are $p(\bar{q}|\, z_A)\, v_A(z_A) dz_A d\tau_A$ affected individuals, with a resulting tax effect of $T_B(k(z_A)) - T_A(z_A)$ for each of them. If the schedule results in migration from region B, the argument is analogous, as we show formally in the Supplement. The total effect is

$$M_A = dz_A d\tau_A \int_{z_A}^{\infty} [T_B(k(z'_A)) - T_A(z'_A)]\, p(\bar{q}|\, z'_A)\, v_A(z'_A) dz'_A.$$

In the optimum, these effects cancel out, $R_A + L_A + M_A = 0$, so that optimal marginal tax rates can be characterized by

$$\begin{aligned}\frac{\tau_A}{1-\tau_A} &= \frac{1}{\varepsilon}\frac{1}{z_A v_A(z_A) s_A(z_A)}\frac{1-\tau_A}{1-\tau_A+\tau'_A z_A \varepsilon_A} \\ &\quad \times \int_{z_A}^{\infty} \left\{\left[1-g_A\left(z'_A\right)\right] s_A(z'_A) + \left[T_B(k(z'_A)) - T_A(z'_A)\right] p\left(\bar{q}\middle|\, z'_A\right)\right\} v_A(z'_A) dz'_A.\end{aligned} \tag{9}$$

We turn now to the optimal tax schedule in the high-productivity region. We consider a small increase in taxes by $dz_B d\tau_B$ for all individuals above z_B in region B. This again generates three effects, which must balance out along the optimal tax schedule, so that

$$\begin{aligned}\frac{\tau_B}{1-\tau_B} &= \frac{1}{\varepsilon}\frac{1}{z_B v_B(z_B) s_B(z_B)}\frac{1-\tau_B}{1-\tau_B+\tau'_B z_B \varepsilon_B} \\ &\quad \times \int_{z_B}^{+\infty} \left\{\left[1-g_B\left(z'_B\right)\right] s_B(z'_B) - \left[T_B(z'_B) - T_A(k^{-1}(z'_B)\right] p\left(\bar{q}\middle|\, z'_B\right)\right\} v_B\left(z'_B\right) dz'_B.\end{aligned} \tag{10}$$

Both optimal tax schedules are rigorously derived in Appendix A. The optimal tax formulae not only differ in the average welfare weights and in the corresponding productivity distributions above the relevant gross income level, but they also take the fiscal migration externality into account. Typically, this externality will be negative for the high-productivity region and positive for the low-productivity region. Accordingly, from (9) and (10) we have the following result.

Proposition 4 *For all levels of innate productivity and the corresponding gross incomes, the marginal tax rate τ_A in the low-productivity region is increasing in the difference in total tax liability between the high- and the low-productivity regions, and the marginal tax rate τ_B in the high-productivity region is decreasing in this difference in total tax.*

Proof. The result follows directly from (9) and (10). ■

The optimal differentiated tax schedules differ from the standard tagging case because of the migration effect M_A (and M_B). These migration effects measure the marginal fiscal externality of migration, which is increasing in the tax liability difference. Thus, an increasing tax liability difference puts more downward pressure on marginal tax rates in region B and more upward pressure on the marginal tax rate in region A. In other words, the larger the marginal fiscal gains from migration are, the more the government distorts labor supply in the low-productivity region and the less it distorts labor supply in the high-productivity region.

Finally, we consider how taxation should be regionally differentiated if a full differentiation of the entire tax schedule is not feasible. First, we study a tax system where the same income faces the same marginal tax rate in both regions but allow for region-dependent total tax liabilities. The maximization problem of the government is as in Section 4.1, except that instead of the restriction that $\Delta T \equiv T_B(z) - T_A(z) = 0$ we have $\Delta T = C$, where C is constant. For this setting, we find:

Proposition 5 *Starting from a unified taxation schedule in the two regions, if the government is allowed to make a lump-sum transfer between regions, it will choose to make a transfer from the less productive to the more productive region.*
Proof. *See Appendix B.* ■

While this result may appear surprising, there is a clear economic intuition behind it. The tax on the poor region has to be higher in order to induce extra migration, which is productivity-enhancing. The extensive margin is used to increase efficiency via increased labor mobility, whereas redistribution is engineered through the intensive margin.

Next, consider a tax schedule that is separable in the sense that $\tau_A = \tau_B$. Starting from this schedule, the following proposition shows that decreasing the marginal tax in region B and increasing it in region A would be desirable:

Proposition 6 *If Ψ' is convex, q and n are independently distributed, and $\omega'(n) \geq 1$, then it is optimal to introduce some wedge in marginal taxes into the system of separable taxation ($\tau_A = \tau_B$) of the two regions. In particular, it is optimal to decrease the marginal tax in the high-productivity region and increase it in the low-productivity region.*

Proof. See Appendix B. ■

The proof is based on the fact that the difference in marginal welfare weights of residents of regions A and B is decreasing in productivity if the social welfare exhibits prudence (marginal social welfare is convex). Thus, it makes sense to make the lower part of the productivity distribution in region A marginally happier than in region B, while making the upper part of the productivity distribution in region B marginally happier than in region A. Hence, lower marginal tax rates in region B are optimal. The productivity transformation function $\omega(n)$ may, however, reverse this finding if migration in the lower part of the distribution is subject to substantially larger productivity gains than migration in the upper part of the distribution, i.e., $\omega'(n) < 1$; hence the condition on this function.

6 Simulation and Calibration

In this section we provide numerical simulations for the US, to gain insights into the quantitative importance of productivity-enhancing migration for the design of tax policy and optimal redistribution.[10] We first focus on the difference between an optimal unified tax schedule with and without migration for a given posterior productivity distribution as in propositions 1 and 2. To apply our framework empirically, we divide the US into a high-productivity region (large metropolitan) and a low-productivity region (other). We use the observable income distribution to recover the underlying productivity distributions in both regions, as well as the implied migration gains for workers of different innate productivity. We then simulate the unified optimal tax formulae with and without productivity-enhancing migration for the posterior productivity distribution to gauge the difference between them. Finally, we also simulate the optimal differentiated tax schedules for the high- and low-productivity regions. In what follows, we first specify the functional forms and parameters used in the simulations, and then describe the calibration.

[10] Our code further extends and modifies the code developed by Henrik Kleven, Claus Thustrup Kreiner, and Emmanuel Saez, as applied in Kleven et al. (2009), for our setting. We would like to express our gratitude to them for kindly providing us with their original code.

6.1 Simulation specification

For simulations, we use isoelastic utility $h(\frac{z}{n}) = (\frac{z}{n})^{1+\epsilon}/(1+\epsilon)$ with a constant earnings elasticity $\varepsilon = \frac{1}{\epsilon}$ as in Saez (2001). Moreover, we follow Kleven et al. (2009) by assuming a power law distribution for the costs at the extensive migration margin on the interval $[0, q_{\max}]$ with $P(q) = (q/q_{\max})^{\eta}$ and $p(q) = (\eta/q_{\max}) \cdot (q/q_{\max})^{\eta-1}$. This distribution of q is the same in each region and independent of n, that is, $\partial q_{\max}/\partial n = 0$. The parameter η may be interpreted as a migration elasticity of the form $\eta = \frac{\bar{q}}{P(\bar{q}|n)} \frac{\partial P(\bar{q}|n)}{\partial \bar{q}} = \frac{\bar{q}p(\bar{q}|n)}{P(\bar{q}|n)}$. As the social objective function, we use the constant rate of risk aversion (CRRA) function $\Psi(V) = V^{1-\gamma}/(1-\gamma)$, where the parameter γ measures the government's equity preference. We choose $\gamma = 1$; hence $\Psi(V) = \log(V)$ in line with Chetty (2006).

6.2 Calibration to the US

We proceed with the calibration of our economy to the US in four steps. First, we choose regions by focusing on the productivity discrepancy between regions with different levels of urbanization. To do this, we draw on the Rural Urban Continuum Code (RUCC, also known as the Beale code) that is provided by the US Department of Agriculture. The RUCC assigns each county to one of nine classes. Starting with highly urban counties central in a metropolitan area and with a population of more than 1 million (class 1), the code goes up to 9 for completely rural counties that are not adjacent to a metropolitan area and/or exhibit a population of less than 2,500. The Panel Study of Income Dynamics (2013, PSID) provides the RUCC for each individual's county of residence. We define all counties belonging to class 1 as large metropolitan areas (region B), and counties of classes 2 through 9 as other areas (region A), as illustrated by Figure 2. Regional population shares are calculated using the population information provided with the RUCC-classification, and they amount to 53% and 47% for regions B and A, respectively. The optimal tax formulae are weighted accordingly.[11]

[11] In the Supplement we additionally provide simulation results for an alternative split, where we group all class 1 and class 2 counties into the high productivity region, and classes 3 to 9 counties into the low productivity region.

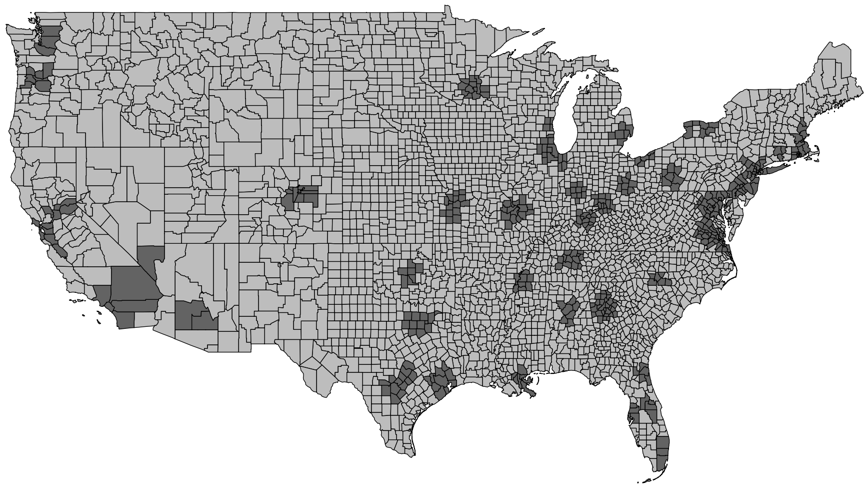

Figure 2: Split of the mainland US districts into two regions: region B consists of the more metropolitan counties (dark), region A of all others (light). Boundaries are taken from the US Census Bureau (census.gov: Cartographic Boundary Shapefiles).

Second, we recover the ability distributions for these regions using individuals' maximization as given by Equation (2) with earnings elasticity $\varepsilon = 0.25$ as suggested by Saez (2001). Specifically, we combine the 2008 individual gross labor income data from the 2009 PSID for households' working heads with the corresponding marginal tax rate from the NBER TAXSIM model.[12] The procedure to recover the productivity distribution is as follows: Each individual's innate ability is computed from the individual utility maximization, using her income data from the PSID, and the actual marginal tax rate corresponding to this income level from TAXSIM together with the earnings elasticity ε and the functional assumption for $h(\frac{z}{n})$. As suggested by Diamond (1998) and Saez (2001), very high incomes are well approximated by a Pareto distribution. Therefore, the skill distributions are modified by assuming a Paretian shape for each top 5% of the respective distribution corresponding to abilities above \$184,717 (\$130,271) in the large metropolitan (other) areas. The Pareto parameter for the US is approximately equal to 2, see Saez (2001). Since income distributions of low income countries and regions exhibit a higher Pareto coefficient than income distributions of high income countries and regions, we slightly adjust the regional Pareto coefficients to 1.95 (2.05) for

[12] The NBER TAXSIM model v9 is applied, see http://www.nber.org/taxsim/.

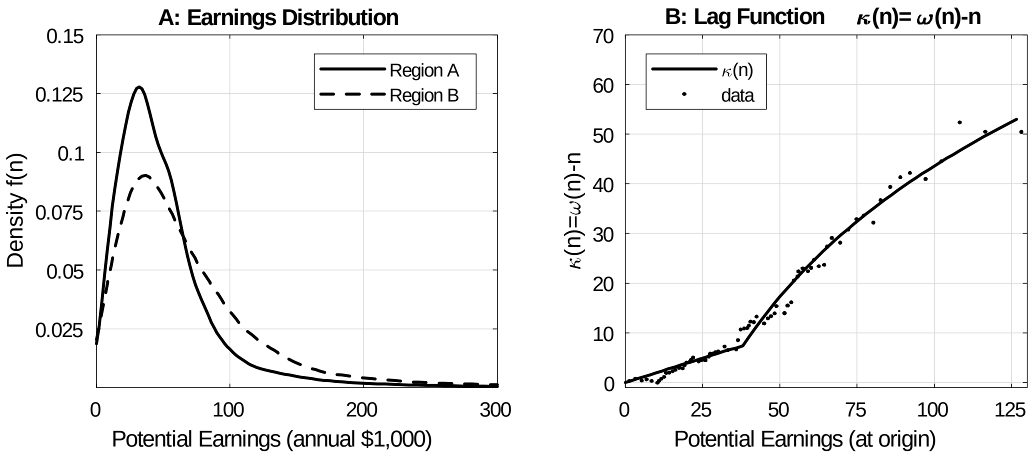

Figure 3: Revealed true abilities (plot A) and revealed lag function $\kappa(n)$ (plot B) for the metropolitan (Region B) and non-metroplitan (Region A) regions of the US, based on data from PSID 2008/09 and TAXSIM. Revealed abilities are continuously smoothed.

the large metropolitan (other) areas. In addition, as is obvious from Figure 3A for abilities above roughly \$80,000 and below the respective Pareto threshold the distribution in region A is always below that of region B which also implies a Pareto coefficient in region A that is higher than in region B. Figure 3A depicts the computed skill distributions in regions A and B.

Third, we estimate the transformation function $\omega(n)$ from the regional ability distributions. We compute the difference in the mean ability in each percentile of the distribution. This difference is assumed to be the productivity increase for the mean person (sampling point) of each percentile. We fit the lag function $\kappa(n) = \omega(n) - n$ for the sampling points that are below the Pareto tail (\$130,271) by minimizing the average squared distance from a linear (for low income) and logarithmic (for high income) function of potential earnings. The fitted function is presented together with the sampling points in Figure 3B.[13] We find a substantial productivity difference between large metropolitan districts and elsewhere.

Fourth, we calibrate the migration cost distribution assuming particular values for the migration elasticity η and the parameter $q_{\max}$. We use $\eta = 1$ as a baseline case, and additionally consider lower and higher values in a range between $\eta = 0.1$

[13] The level of potential earnings at which the functional form switches (\$38,000) is determined endogenously by mimimizing the total squared error.

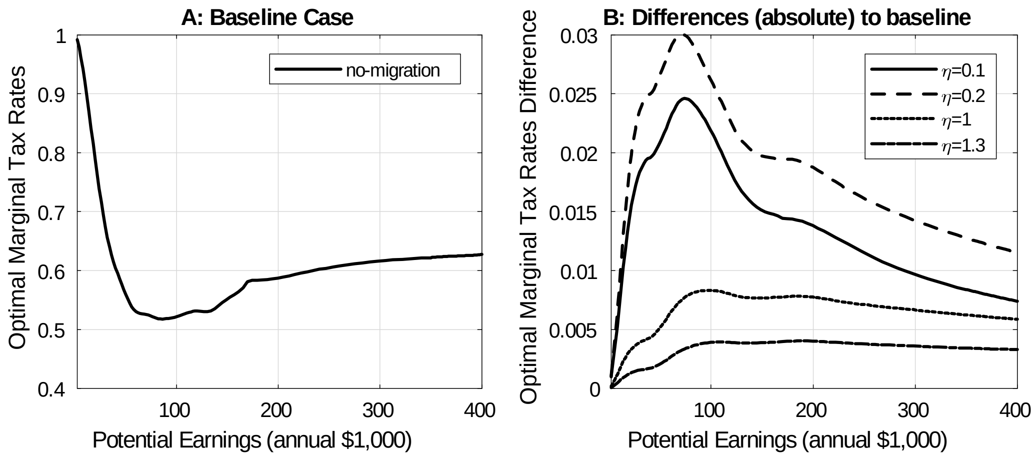

Figure 4: Optimal uniform marginal tax rate simulations for the US, based on data from PSID 2008/09 and TAXSIM (panel A), and absolute tax rate differences between the optimal tax schedules with and without the migration effect for different values of the migration elasticity. Actual differences are negative. Parameter values for panel A: $\varepsilon = 0.25$, $\gamma = 1$; for panel B: $\varepsilon = 0.25$, η varies, $\gamma = 1$.

and $\eta = 1.3$ to assess the sensitivity with respect to this parameter, since there are no reliable estimates that would directly map into the migration elasticity as defined here. We choose $q_{\max}$ such that the average moving costs across all productivity levels amount to approximately \$5,500.[14] Finally, we use the ratio of exogenous expenditures E to aggregate production of 0.25 as in Saez (2001).

6.3 Results for undifferentiated taxation

When the government ignores migration, it chooses a higher marginal tax rate than when it takes it into account, in line with Proposition 2. In Figure 4, we display the marginal tax rate in the oblivious government benchmark in panel A, and the size of the reduction in marginal tax rates due to taking migration into account in panel B. For very low income levels, there is hardly any difference in the marginal tax rate schedules. Migration-induced productivity increases are small in this income range, and tax liabilities are rather low as well, which is why

[14] Our analysis considers migration gains on a yearly basis. Interpreting migration costs as a perpetually accruing disutility and applying a reasonable discount rate, this cost-level corresponds to the range of migration-cost estimates obtained by Kennan and Walker (2011) or Bayer and Juessen (2012).

the migration effect is negligible. However, for higher potential income levels, the differences become more pronounced. Taking the migration effect into account for the optimal tax schedule reduces marginal tax rates by roughly 1 percentage point (p.p.) for mid-income levels in the baseline case. For potential incomes above \$80,000 the deviations from the benchmark are somewhat smaller again. While the individual fiscal externality is large for high-income migrants, there are fewer of them. To sum up, the benchmark results indicate that location-dependent productivity and the possibility of efficiency-enhancing migration constitute only a minor constraint to redistribution via a common tax-transfer scheme.

Figure 4B also illustrates that the relationship between the tax rate reduction due to productivity-enhancing migration relative to the no-migration case is quite sensitive to the migration elasticity, and that, overall, this relationship is not monotone. Consistent with our theory, the differences in marginal tax rates are again reduced for sufficiently low values of η, as can be seen from the comparison of the tax rate differences for $\eta = 0.2$ and $\eta = 0.1$ in Figure 4.[15] Note also from Figure 4B that the marginal tax rate differences may be substantially more pronounced than the 1 p.p. in the baseline case, reaching 3 p.p. for $\eta = 0.2$ for some ability levels.

6.4 Results for differentiated taxation

The results for differentiated taxation are depicted in Figure 5. We use the same baseline parameter values for migration costs and the exogenous government expenditure requirement E as before.

Figure 5A illustrates the case of differentiated taxation without migration (fixed-residence case). This benchmark corresponds to the standard tagging case. The region-specific subsidies at the very bottom (not visible from Figure 5A) amount to $T_A(n_{\min}) = -\$17,500$ and $T_B(\omega(n_{\min})) = -\$16,900$ in this case, showing a substantial across-regions difference. The relatively high marginal tax rates at the bottom in both regions are related to the phasing out of the subsidy at the

[15] Obviously, the differences in marginal tax rates completely disappear for $\eta \to 0$.

very bottom, which is commonly observed with the classic U-shaped marginal tax pattern. At the bottom, marginal tax rates are falling faster in region A than in region B. This is consistent with a higher productivity density in region A in this very low income range.

The marginal tax rates in region B tend to be above the marginal rates in region A at nearly all income levels. Since the productivity distribution in region B is shifted to the right of that in region A, increasing marginal tax rates over a small interval of the distribution generates, in region B, more revenue from the right tail of the distribution than in region A. Moreover, the average marginal social welfare weights of the individuals affected by the higher tax payments tend to be lower. These effects apparently dominate potential efficiency effects stemming from productivity density differences on the interval for which an increase in marginal tax rates is considered.

Over the interval from \$40,000 to \$75,000 of potential income, the regional rate differences are particularly pronounced. Distorting the intensive margin in this range in region A is costly because the productivity density is particularly high here. At the same time, region B is characterized by high-productivity density in the interval from \$75,000 to \$160,000 of potential income, so the disortion from tax is particularly costly here as well. This is reflected in a dent in the marginal tax schedule in region B over this range, with marginal tax rates in region B slightly below those in region A at around \$120,000. At high income levels (>\$180,000 of potential income), marginal rates in region B are somewhat above those in region A.

We compare the fixed residence benchmark to the case of productivity-increasing migration displayed in Figure 5B, which correspond to our optimal tax schedules given by (9) and (10). Moving from the fixed residence case to the migration case substantially changes the differentiated taxation.

First, with migration, the subsidies at the bottom in different regions are much closer to each other, $T_A(n_{\min}) = -\$17.100$ and $T_B(\omega(n_{\min})) = -\$17.200$ (not visible from Figure 5B). To encourage migration, the subsidy is reduced in region

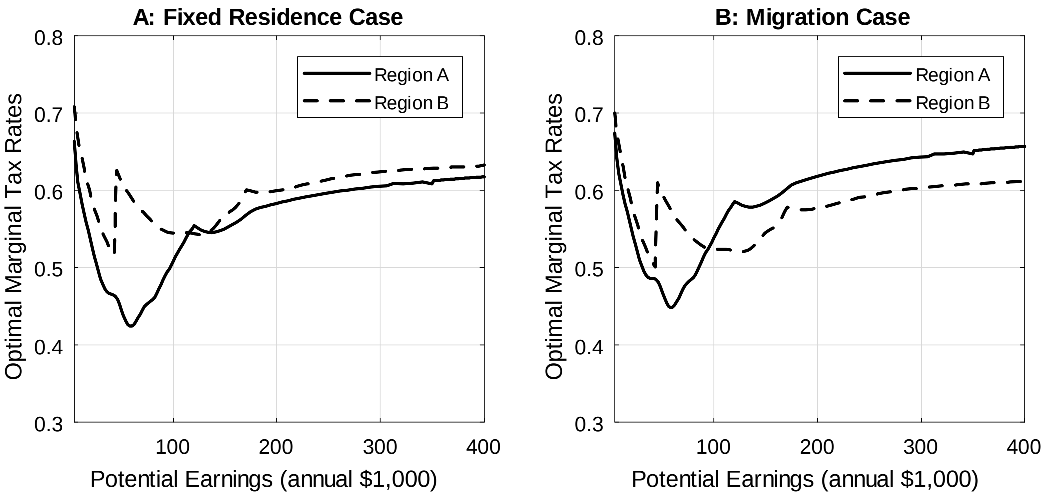

Figure 5: Optimal differentiated marginal tax rate simulations for the US based on data from PSID 2008/09 and TAXSIM. Parameter values for panel A: $\varepsilon = 0.25$, $\eta = 0$, $\gamma = 1$; for panel B: $\varepsilon = 0.25$, $\eta = 1$, $\gamma = 1$.

A and increased in region B relative to the fixed residence case.

Second, the pattern of marginal tax rates is rather similar in both cases. This is because the underlying regional productivity distributions determine the efficiency cost and revenue benefits of distortion along the intensive margin, which largely shape the pattern of the marginal tax schedule.

Third, the marginal tax rates at each income level are shifted due to migration: they fall in region B and rise in region A. These level shifts are less pronounced for low levels of income, where the fiscal migration externality matters less. However, at the potential income level of $200,000, for example, the marginal tax rate in region A rises by over 4 p.p., and the marginal tax rate in region B falls by over 2 p.p. These shifts illustrate the downward pressure of the migration effect on marginal tax rates in the high-productivity region and the reverse pressure in the low-productivity region as suggested by Proposition 4.

It is interesting to compare our optimally differentiated tax schedule with the actual regional differentiation of income taxation due to cost-of-living differences. Albouy (2009) argues that undifferentiated federal income taxation taxes inhabitants of urban agglomerations effectively more heavily than those of rural areas, resulting in a spatial distortion of economic activity. Our analysis suggests that optimally differentiated taxation should result in higher marginal taxes for middle to high income earners in the less-productive region. Thus, the de facto differentiation in favor of low-productivity regions due to cost of living differences is, at least for the high income earners, the opposite of an optimally differentiated tax system. Implicit regional tax differentiation may therefore additionally be criticized from a redistribution perspective with location-specific productivity.

Finally, we compare the welfare implications of the different tax schedules. In particular, we calculate the welfare increases from taking the fiscal migration externalities into account relative to the benchmark of the oblivious government setting a unified tax schedule according to (7). We measure the welfare change in terms of equivalent change in income for all individuals and report it as a percentage of the average income in the economy. For undifferentiated taxation, taking the migration into account leads to a welfare increase of 0.07%. This rather small effect corresponds to the small difference between the baseline tax schedule and the optimal tax schedule which takes the fiscal migration externality into account. For differentiated taxation we find a higher, but still moderate, welfare increase of 0.2%.

7 Discussion and Conclusion

Our analysis has shown how inequality-averse and information-constrained governments should take regional productivity differences and the corresponding fiscal migration externalities into account in the design of their tax-transfer schemes. However, our analysis abstracts from a number of potentially important aspects. First, we restrict the central government to the use of a unified or regionally differentiated tax scheme; it is not allowed to use subsidies targeted at migrants only. If such targeted transfers were available to the government, they could potentially loosen the trade-off between redistribution and internal migration. However, a migration subsidy is typically not straightforward to implement, and it presents a monitoring challenge to prevent fraudulent use. Moreover, it also cannot eliminate the problem completely, since the fiscal migration externality differs by earnings level.

Secondly, regional productivity differences are partly reflected in the local prices of non-tradable goods, rents, and house prices, which also reduces migration incentives. While this is important, it does not challenge our intuition. The value of an additional dollar of tax revenue for the government is independent of the region where this dollar is raised. Thus, the fiscal migration externality also exists with regional cost-of-living differences and should be taken into account.

Thirdly, real regional wages can also be affected by crowding effects that can arise from immigration into high-productivity regions. These effects may drive housing rents up, which could hurt in particular the resident low-productivity individuals. On the other hand, emigration may reduce housing costs in the low-productivity regions, which should benefit the remaining population.[16]

Fourthly, inter-regional migration may directly affect the real wages of non-migrants in both regions. If the marginal productivity of each type in a region depends on the number and distribution of taxpayers in that region, migration endogenously affects wages. If labor is perfectly substitutable across skill levels, and its marginal product is decreasing in the total regional labor supply, migration from low- to high-productivity regions is likely to have a damping effect on inequality. In this case, the optimal tax schedule(s) might have to encourage migration even more than in the exogenous wage case.

If labor is not perfectly substitutable across skill levels, but rather complementary, productivity of low-productivity individuals depends positively on the regional labor supply of highly productivity individuals, similar to the case of endogenous wages studied by Stiglitz (1982). In this case, the out-migration of high-productivity individuals has a negative effect on low-productivity types in the low productivity region. However, there will be a positive productivity effect on the low-productivity types in the high-productivity region.

The fiscal externalities of migration become more complex with endogenous

[16]In the urban economics literature it is frequently argued that such effects may be rather asymmetric due to the durability of housing. In regions where housing demand is falling, rents may be falling substantially more than the rent increases observed in growing regions, given the inelastically supplied stock of existing housing, see Duranton and Puga (2015).

wages, since the wage effects on the resident population in both regions are also reflected in the changes in their tax liabilities. Overall, the implications of endogenous real wages, either through (de-)congestion or through complementarities in production, for our results are not clear-cut a priori.[17]

Finally, we have used a static framework. One may argue that migration also contains an inherently dynamic aspect. At the same time, to the extent that the migration costs are recurring costs, say, because the disutility of being in a less preferred region accrues every period, our approach easily maps into a dynamic framework. However, further research may explore how the evolution of regional productivity differences and the option of repeated migration may qualify our results.

8 Appendix A: Optimal tax formulae

We show that the optimal tax formulae (6), (9), and (10) can also be rigorously derived by optimal control techniques. We provide a more detailed derivation in the Supplement, where we also show the equivalence of the expressions in terms of n and z. We start with the differentiated case, since the unified case can be interpreted as the same problem with the additional constraint of identical tax schedules in both regions.

8.1 Regionally differentiated taxation

The government maximizes

$$\begin{aligned} W &= \int_{n_{\min}}^{n_{\max}} [\int_{\bar{q}_B}^{+\infty} \Psi\left(V_B\left(\omega(n)\right)\right) p(q|n)dq + \int_0^{\bar{q}_A} \Psi\left(V_B\left(\omega(n)\right) - q\right) p(q|n)dq \\ &\quad + \int_{\bar{q}_A}^{+\infty} \Psi\left(V_A\left(n\right)\right) p(q|n)dq + \int_0^{\bar{q}_B} \Psi\left(V_A\left(n\right) - q\right) p(q|n)dq] f(n)dn, \end{aligned}$$

where $\bar{q}_A(n) = \max\left\{V_B\left(\omega(n)\right) - V_A\left(n\right), 0\right\}, \ \bar{q}_B(n) = \max\left\{V_A\left(n\right) - V_B\left(\omega(n)\right), 0\right\}.$[18] The first term in the expression in brackets stands for the social welfare from the population of region B that did not move, the second term stands for those of the population that moved from A to B, the third term is for those who stayed in A, and the fourth term is for those who moved from B to A. Note that either the second or the fourth

[17] Optimal taxation with endogenous wages has been studied by Stiglitz (1982) and, more recently, by Rothschild and Scheuer (2013), among others. These contributions provide useful starting points to think about endogenous wages in the regional context.

[18] We drop the arguments of $\bar{q}_A$ and $\bar{q}_B$ for ease of notation.

term is equal to zero, because migration in both direction at the same ability level is not possible. The maximization is subject to

$$\int_{n_{\min}}^{n_{\max}} [\left(z_B - \omega(n)h\left(\frac{z_B}{\omega(n)}\right) - V_B(\omega(n))\right)(1 + P(\bar{q}_A|n) - P(\bar{q}_B|n))$$
$$+\left(z_A - nh\left(\frac{z_A}{n}\right) - V_A\right)(1 + P(\bar{q}_B|n) - P(\bar{q}_A|n))]f(n)dn \geq E$$

and the corresponding incentive-compatibility constraints. Note that either $P(\bar{q}_A|n)$ or $P(\bar{q}_B|n)$ is zero for the same reason as discussed above.

Let the Hamiltonian be $H(z_A, z_B, V_A, V_B, \lambda, \mu_A, \mu_B, n)$. Necessary conditions are

1. There exist absolutely continuous multipliers $\mu_A(n), \mu_B(n)$ such that on $(n_{\min}, n_{\max})$ $\dot{\mu}_B(n) = -\frac{\partial H(n)}{\partial V_B(n)}$, $\dot{\mu}_A(n) = -\frac{\partial H(n)}{\partial V_A(n)}$ almost everywhere with $\mu_i(n_{\min}) = \mu_i(n_{\max}) = 0$.

2. We have $H(z_i(n), V_i, \lambda, \mu_i, n) > H(z_i, V_i, \lambda, \mu_i, n)$ almost everywhere in n for all z. The first-order conditions are $\frac{\partial H}{\partial z_A} = 0$, $\frac{\partial H}{\partial z_B} = 0$.

Uniqueness of z_A and z_B that solve the equations above can be established in a similar way to Kleven et al. (2009), using the assumption that $\varphi(x) = (1 - h'(x))/(xh''(x))$ is decreasing in x. The FOCs for region A and B are presented in the Supplement, where we also lead the reader through the derivation steps to arrive at the optimal tax formulae:

$$\frac{\tau_A}{1-\tau_A} = \frac{1}{nf(n)\varepsilon_A(1 + P(\bar{q}_B|n) - P(\bar{q}_A|n))} \int_n^{n_{\max}} [\left(1 - g_A(n')\right)\left(1 + P(\bar{q}_B|n') - P(\bar{q}_A|n')\right)$$
$$- (T_A - T_B)\left(p(\bar{q}_B|n') + p(\bar{q}_A|n')\right)]f(n')dn',$$

$$\frac{\tau_B}{1-\tau_B} = \frac{\omega'(n)}{\omega(n)f(n)\varepsilon_B(1 + P(\bar{q}_A|n) - P(\bar{q}_B|n))} \int_n^{n_{\max}} [\left(1 - g_B(n')\right)\left(1 + P(\bar{q}_A|n') - P(\bar{q}_B|n')\right)$$
$$- (T_B - T_A)\left(p(\bar{q}_B|n') + p(\bar{q}_A|n')\right)]f(n')dn'$$

for the marginal rates in A and B, respectively. Note that for each n, there are two mutually exclusive scenarios: either there is migration from A to B (and $V_B(\omega(n)) > V_A(n)$) or migration from B to A. We discuss this further in the Supplement.

8.2 Delayed optimal control

For the next section, we will need some results in delayed optimal control theory. In particular, we adapt the Göllmann et al. (2008) setting that addresses delayed arguments of constant size. We consider variable "delays" of the size $\omega(n) - n$ in both the state variable $x(n)$ (in our case this is $V(n)$) and in the control variable $u(n)$ (in our case this is $z(n)$). Consider a welfare-maximization problem. Because social welfare depends on the indirect utilities only, our objective function will not depend on control variables. We have the following retarded optimal control problem (ROCP):

$$J(u, x) = \int_a^b L(n, x(n), x(\omega(n)))\, dn,$$

subject to the differential equation (ICC in our context, can be formulated without delayed variables and states on the rhs) and inequality constraint (government budget)

$$\dot{x}(n) = f\left(n, u(n)\right), n \in [a,b] \quad \text{(11a)}$$

$$\int_a^b C\left(n, u(n), u(\omega(n)), x(n), x(\omega(n))\right) dn \geq 0. \quad \text{(11b)}$$

For convenience, the functions

$$\begin{aligned} L &: [a,b] \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R} \\ f &: [a,b] \times \mathbb{R}^n \to \mathbb{R} \\ C &: [a,b] \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R} \end{aligned}$$

are assumed to be twice continuously differentiable wrt all arguments. A pair of functions (u,x) is called an admissible pair for the problem (ROCP) if the state x and control u satisfy (11a) and (11b). An admissible pair $(\hat{u},\hat{x})$ is called locally optimal pair or weak maximum for (ROCP), if

$$J(u,x) \leq J\left(\hat{u},\hat{x}\right)$$

hold for all (u,x) admissible in a neighborhood of $(\hat{u},\hat{x})$. Define the Hamiltonian for (ROCP) as

$$H\left(n, x, x^w, u, u^w, \mu, \lambda\right) := L\left(n, x, x^w\right) + \lambda C\left(n, u, u^w, x, x^w\right) + \mu f\left(n, u\right).$$

Göllmann et al. (2008) show that a necessary condition for (u,x) to be locally optimal is the existence of a costate absolutely continuous function $\mu : [a,b] \to R$, a multiplier function $\lambda : [a,b] \to R$ such that the following conditions hold for all $t \in [a,b]$:

(i) adjoint differential equation

$$\begin{aligned} \dot{\mu}(n) &= -\frac{\partial H\left(n\right)}{\partial x} - \mathbb{I}_{[\omega(a),\omega(b)]}\frac{\partial H\left(\omega^{-1}(n)\right)}{\partial x^w} \\ &= -\frac{\partial H\left(n, x(n), x\left(\omega(n)\right), u(n), u\left(\omega(n)\right), \mu(n), \lambda(n)\right)}{\partial x} \\ &-\mathbb{I}_{[\omega(a),\omega(b)]}\frac{\partial H\left(n, x(\omega^{-1}(n)), x\left(n\right), u(\omega^{-1}(n)), u(n), \mu(\omega^{-1}(n)), \lambda(\omega^{-1}(n))\right)}{\partial x^w}, \end{aligned}$$

(ii) transversality condition

$$\mu(a) = \mu(b) = 0,$$

(iii) local maximum condition

$$\frac{\partial H(n)}{\partial u} + I_{[\omega(a),\omega(b)]}\frac{\partial H\left(\omega^{-1}(n)\right)}{\partial u^w} = 0,$$

(iv) nonnegativity of multiplier and complementarity slackness.

8.3 Nondifferentiated tax schedule

The government now maximizes

$$\begin{aligned} W &= \int_{n_{\min}}^{n_{\max}} [\int_0^{+\infty} \Psi\left(V\left(\omega(n)\right)\right) p(q|n)dq + \int_0^{\bar{q}} \Psi\left(V\left(\omega(n)\right) - q\right) p(q|n)dq \\ &\quad + \int_{\bar{q}}^{+\infty} \Psi\left(V\left(n\right)\right) p(q|n)dq] f(n)dn, \end{aligned}$$

where $\bar{q} = V\left(\omega(n)\right) - V\left(n\right)$. We have also dropped the subscript B from the omega function for more parsimonious notation. The maximization is subject to

$$\begin{aligned} &\int_{n_{\min}}^{n_{\max}} [\left(z\left(\omega(n)\right) - \omega(n)h\left(\frac{z\left(\omega(n)\right)}{\omega(n)}\right) - V\left(\omega(n)\right)\right)(1 + P(\bar{q}|n)) \\ &+ \left(z - nh\left(\frac{z}{n}\right) - V\right)(1 - P(\bar{q}|n))] f(n)dn \geq E. \end{aligned}$$

Note that in the uniform case there cannot be migration from B to A, as this would imply $V\left(\omega(n)\right) < V\left(n\right)$ that contradicts incentive-compatibility (the productivity type $\omega(n)$ can pretend to have productivity n without any costs).

Let the Hamiltonian be $H(z, z^w, V, V^w, \lambda, \mu, n)$. This is a delayed optimal control problem analogous to the one formally analyzed by Göllmann et al. (2008) in its entire generality. The difference is that whereas Göllmann et al. have a lag of fixed size over the whole domain of their functions, our lag is a smooth increasing function of n, namely $\omega(n) - n$. The necessary conditions for optimal control in such a setting is presented in Abdeljawad et al. (2009). Namely, in our context the necessary conditions for the maximum are:

1. There exist absolutely continuous multipliers $\mu(n)$ such that on $(n_{\min}, n_{\max})$ $\dot{\mu}(n) = -\frac{\partial H(n)}{\partial V(n)} - \mathbb{I}_{[\omega(n_{\min}),\omega(n_{\max})]} \frac{\partial H\left(\omega^{-1}(n)\right)}{\partial V^w(n)}$ almost everywhere with $\mu(n_{\min}) = \mu(n_{\max}) = 0$.

2. We have $H(z(n), z^w(n), V, V^w, \lambda, \mu, n) > H(z, z^w, V, V^w, \lambda, \mu, n)$ almost everywhere in n for all z. The first-order condition is

$$\frac{\partial H}{\partial z} + I_{[\omega(n_{\min}),\omega(n_{\max})]} \frac{\partial H\left(\omega^{-1}(n)\right)}{\partial z_B^w} = 0$$

The fact that this condition describes a global maximum can be similarly established to Kleven et al. (2009), using the assumption that $\varphi\left(x\right) = \left(1 - h'\left(x\right)\right) / \left(xh''\left(x\right)\right)$ is decreasing in x.

$$\frac{\mu}{n}\frac{z}{n}h''\left(\frac{z}{n}\right) + \lambda\left(1 - h'\left(\frac{z}{n}\right)\right)\left(\left(1 - P(\bar{q}|n)\right)f(n) + \left(1 + P(\bar{q}_1|\omega^{-1}(n))\right)f\left(\omega^{-1}(n)\right)\right) = 0$$

$$\varphi\left(\frac{z}{n}\right) = -\frac{\mu(n)}{\lambda n\left(\left(1 - P(\bar{q}|n)\right)f(n) + \left(1 + P(\bar{q}_1|\omega^{-1}(n))\right)f\left(\omega^{-1}(n)\right)\right)}$$

LHS is decreasing in z_i/n whereas RHS is constant, which implies that $z_i(n)$ is a unique solution and a global maximum indeed. Continuity can then be established in a way similar to Kleven et al. (2009). From this we arrive at

$$\int_{\omega^{-1}(n)}^{n_{\max}} \left(1-g_B\left(n'\right)\right)\left(1+P(\bar{q}|n')\right) f\left(n'\right) dn' + \int_{n}^{n_{\max}} \left(1-g_A(n')\right)\left(1-P(\bar{q}|n')\right) f(n')dn'$$
$$-\int_{\omega^{-1}(n)}^{n} \left(T\left(\omega(n')\right)-T(n')\right) p(\bar{q}|n') f\left(n'\right) dn' = \frac{\tau}{1-\tau}\times$$
$$\times n\varepsilon\left(\left(1-P(\bar{q}|n)\right) f(n) + \left(1+P(\bar{q}_1|\omega^{-1}(n))\right) f\left(\omega^{-1}(n)\right)\right). \qquad (12)$$

9 Appendix B: Proofs and asymptotic properties

9.1 Proof of Proposition 2

Proof. Going through the derivation of the optimal tax formula in Appendix A under the assumption that the effect of tax on the migration decision is neglected, i.e., $\frac{\partial \bar{q}}{\partial z} = \frac{\partial \bar{q}}{\partial V} = 0$, we arrive at the optimal tax formula (12) short of the term

$$-\int_{\omega^{-1}(n)}^{n} \left(T\left(\omega(n')\right)-T(n')\right) p(\bar{q}|n') f\left(n'\right) dn'.$$

If this is non-positive ($D(z) \leq 0$), the result immediately follows. If marginal tax rates are positive, this condition is always fulfilled. ■

9.2 Proof of Proposition 5

Proof. The maximization problem of the government with the restriction that $\Delta T = C$ is constant in n is

$$W = \int_{n_{\min}}^{n_{\max}} [\int_{0}^{+\infty} \Psi\left(V\left(\omega(n)\right)+q^h\right) p(q|n)dq + \int_{0}^{\bar{q}} \Psi\left(V\left(\omega(n)\right)-q^c\right) p(q|n)dq$$
$$+\int_{\bar{q}}^{+\infty} \Psi\left(V\left(n\right)+q^h+C\right) p(q|n)dq] f(n)dn,$$

where $\bar{q} = V\left(\omega(n)\right) - V\left(n\right) - C$,and we assume C is small enough not to induce "reverse" migration (to the low-productivity region). The maximization is subject to

$$\int_{n_{\min}}^{n_{\max}} [\left(z\left(\omega(n)\right) - \omega(n)h\left(\frac{z\left(\omega(n)\right)}{\omega(n)}\right) - V\left(\omega(n)\right)\right)\left(1+P(\bar{q}|n)\right)$$
$$+\left(z - nh\left(\frac{z}{n}\right) - V - C\right)\left(1-P(\bar{q}|n)\right)] f(n)dn \geq E.$$

Note that we express everything here in terms of region B taxes – that is why C appears in the expressions for region A as a correction term to increase indirect utility (in the objective function) or to reduce the tax revenue (in the government budget constraint).

By the envelope theorem,

$$\begin{aligned}\frac{\partial W^*}{\partial C} &= \int_{n_{\min}}^{n_{\max}} [\left[\Psi\left(V(n)+\bar{q}^h+C\right)-\Psi\left(V(\omega(n))-\bar{q}^c\right)\right]p(\bar{q}|n) \\ &+\int_{\bar{q}}^{+\infty}\Psi'\left(V(n)+q^h\right)p(q|n)dq-\lambda\left(1-P(\bar{q}|n)\right) \\ &-\lambda\left(T(\omega(n))-T(n)+C\right)p(\bar{q}|n)]f(n)dn.\end{aligned}$$

$$\frac{\partial W^*}{\partial C}|_{C=0}=\int_{n_{\min}}^{n_{\max}}[\int_{\bar{q}}^{+\infty}\left(\Psi'\left(V(n)+q^h\right)-\lambda\right)p(q|n)dq-\lambda\left(T(\omega(n))-T(n)\right)p(\bar{q}|n)]f(n)dn,$$

which is negative, if $\Psi'\left(V(n)+q^h\right)/\lambda=g_A\leq 1$ and $T(\omega(n))>T(n)$. (A sufficient condition is that the marginal tax rate is positive everywhere.) ■

9.3 Proof of Proposition 6

Proof. A separable tax schedule implies that T_B-T_A is constant. Since $z_A/n=z_B/\omega(n)$, we have

$$\begin{aligned}\bar{q}_A &= V_B-V_A=(\omega(n)-n)\left(\frac{z}{n}-h\left(\frac{z}{n}\right)\right)-(T_B-T_A)\text{, so that} \\ \dot{q}_A &= \left(\omega'(n)-1\right)\left(\frac{z}{n}-h\left(\frac{z}{n}\right)\right)-\left(T_B'-T_A'\right).\end{aligned}$$

In particular, under separable taxation and $\omega'(n)=1$, we have $\dot{q}_A=0$. At that point

$$\frac{d\left(g_A-g_B\right)}{dn}=\left[\frac{\Psi''\left(V_A(n)\right)}{\lambda}-\frac{\Psi''\left(V_A+\bar{q}_A\right)+\int_0^{\bar{q}_A}\Psi''\left(V_A+\bar{q}_A-q^c\right)p(q)dq}{\lambda\left(1+P(\bar{q}_A)\right)}\right]\dot{V}_A<0$$

iff Ψ' is convex. Similar to Kleven et al. (2006, 2009) we can consider a tax reform introducing a little bit of "negative jointness" (a lower marginal tax for the higher productivity region). This reform has two components. Above ability level n, we increase the tax in region A and decrease the tax in region B. Below ability level n, we decrease the tax in region A and increase the tax in region B. These tax burden changes are associated with changes in the marginal tax rates on earners around n. The direct welfare effect created by redistribution across regions at each income level:

$$\begin{aligned}dW &= \frac{dT}{F(n)}\int_{n_{\min}}^{n}\left(g_A(n')-g_B(n')\right)f(n')dn' \\ &-\frac{dT}{1-F(n)}\int_{n}^{n_{\max}}\left(g_A(n')-g_B(n')\right)f(n')dn'.\end{aligned}$$

Because g_A-g_B is decreasing, $dW>0$. Second, there are fiscal effects associated with earnings responses induced by the changes in τ_A and τ_B around n. Since the reform increases the marginal tax rate in region A around n and reduces it in region B, the earnings responses are opposite. As we start from $\tau_A=\tau_B$, and hence identical elasticities, $\varepsilon_A=\varepsilon_B$, the fiscal effects of earning responses cancel out exactly. Finally, the reform creates migration responses. Above n, migration to B will be induced. Below n, migration to B will be inhibited. The fiscal implications of these responses cancel out exactly only if $\omega'(n)=1$. The elasticity η is constant in this case and since initially

$T_A - T_B$ is constant, the revenue gain from migrants above n will be compensated by the revenue loss from migrants below n. By the same logic, for $\omega'(n) > 1$ the gain from migration will be stronger then the loss from it, so we will have another positive effect. With $\omega'(n) < 1$ the revenue gain from migration is smaller than the loss, so a bit of negative jointness is not necessarily optimal. To complete the proof, our reasoning needs to hold for $\omega'(n) > 1$, i.e. $\dot{g}_A - \dot{g}_B < 0$ also for this case. Differentiating $g_A - g_B$, we have

$$\begin{aligned}\dot{g}_A - \dot{g}_B &= \dot{V}_A \frac{\Psi''(V_A)}{\lambda} - \omega'(n)\dot{V}_A \frac{\Psi''(V_A + \bar{q}) + \int_0^{\bar{q}} \Psi''(V_A + \bar{q} - q^c)\, p(q)dq}{\lambda(1 + P(\bar{q}))} \\ &\quad - \frac{g_A - g_B}{1 + P(\bar{q})} p(\bar{q}) \left(\omega'(n) - 1\right) \dot{V}_A < 0,\end{aligned}$$

The first two terms are negative, because Ψ' is convex by assumption. The third term is negative, because $g_A > g_B$ (which follows from the concavity of Ψ). ■

9.4 Asymptotic properties

It is also interesting to consider the asymptotic properties of the differentiated tax schedule. Suppose the distribution of innate ability, $f(n)$, has an infinite tail ($n_{\max} = \infty$). As is standard in the literature, we assume that $f(n)$ has a Pareto tail with parameter $a > 1$ ($f(n) = \tilde{C}/n^{1+a}$ with $\tilde{C}$ a constant). Moreover, we define the tax differential $\Delta T := T_B - T_A$, and we assume that $P(q|n)$, $T_B - T_A, \tau_A, \tau_B$, $\bar{q}_A, \bar{q}_B$ converge to $P^\infty(q)$, $\Delta T^\infty, \tau_A^\infty < 1, \tau_B^\infty < 1$, $\bar{q}_A^\infty, \bar{q}_B^\infty$ as $n \to \infty$. Finally, we concentrate on the case where, for sufficiently large n, we have $\omega(n) = n + c$, where $c \geq 0$ is a finite constant. In this case, the following proposition arises:

Proposition 7 *Under the stated assumptions on convergence, (i) the average marginal social welfare weights in the two regions converge to the same value $\bar{\psi}/\lambda \geq 0$; (ii) the difference between the taxes in the two regions converges to zero, $\Delta T^\infty = 0$; and (iii) the marginal tax rate in both regions converges to τ^∞ with*

$$\frac{\tau^\infty}{1 - \tau^\infty} = \frac{1}{a\varepsilon^\infty} \left(1 - \frac{\bar{\psi}}{\lambda}\right). \tag{13}$$

Proof. Under our assumptions, $V_A(n)$ and $V_B(n)$ are increasing in n without bound, because $\tau_A^\infty < 1, \tau_B^\infty < 1$. As $\Psi' > 0$ is decreasing, it converges to some $\bar{\psi} \geq 0$. Then,

$$\begin{aligned} g_A(n) &= \frac{\int_{\bar{q}_A}^{+\infty} \Psi'(V_A(n))\, p(q|n')dq + \int_0^{\bar{q}_B} \Psi'(V_A(n) - q)\, p(q|n)dq}{\lambda(1 + P(\bar{q}_B|n) - P(\bar{q}_A|n))}, \\ g_B(n) &= \frac{\int_{\bar{q}_B}^{+\infty} \Psi'(V_B(\omega(n')))\, p(q|n')dq + \int_0^{\bar{q}_A} \Psi'(V_B(\omega(n)) - q)\, p(q|n)dq}{\lambda(1 + P(\bar{q}_A|n) - P(\bar{q}_B|n))}, \end{aligned}$$

which converge to $g_A^\infty = g_B^\infty = \frac{\bar{\psi}}{\lambda}$. If $T_B - T_A$ converges, it must be that $\tau_A^\infty = \tau_B^\infty = \tau^\infty$. But since $h'\left(\frac{z_i}{n_i}\right) = 1 - \tau_i(z_i)$, z_i/n_i converges and hence elasticities converge to the same limit ε^∞. Moreover, $\lim_{n\to\infty} \frac{z_A}{n} = \lim_{n\to\infty} \frac{z_B}{\omega(n)}$. Because $P(q|n)$ and

$\bar{q}_A, \bar{q}_B$ converge, $P(\bar{q}_i|n)$ and $p(\bar{q}_i|n)$ converge to $P^{\infty}(\bar{q}_i^{\infty})$ and $p^{\infty}(\bar{q}_i^{\infty})$. The Pareto distribution implies that $(1-F(n))/(nf(n)) = 1/a$ in the tail. Take the limit of our optimal tax formulae to get

$$\begin{aligned}
\frac{1}{a\varepsilon^{\infty}}[1-\frac{\bar{\psi}}{\lambda}+\frac{\Delta T^{\infty}\,(p^{\infty}(\bar{q}_B^{\infty})+p^{\infty}(\bar{q}_A^{\infty}))}{1+P^{\infty}(\bar{q}_B^{\infty})-P^{\infty}(\bar{q}_A^{\infty})}] &= \frac{\tau^{\infty}}{1-\tau^{\infty}}, \\
\frac{1}{a\varepsilon^{\infty}}[1-\frac{\bar{\psi}}{\lambda}-\frac{\Delta T^{\infty}\,(p^{\infty}(\bar{q}_B^{\infty})+p^{\infty}(\bar{q}_A^{\infty}))}{1+P^{\infty}(\bar{q}_A^{\infty})-P^{\infty}(\bar{q}_B^{\infty})}] &= \frac{\tau^{\infty}}{1-\tau^{\infty}}
\end{aligned}$$

for the marginal rates in region A and B, respectively. The right-hand sides are equal, so we need $\Delta T^{\infty} = 0$ for the left-hand sides to be equal as well.[19] ■

The intuition for the zero difference of top taxes is similar to that in Kleven et al. (2009). Starting from a wedge between T_B and T_A, say, with $T_B > T_A$, welfare could be increased by marginally reducing this wedge. In particular, consider a reform that increases T_A and decreases T_B above some n, while keeping the tax revenue constant in the absence of migration. The direct welfare effects of such a reform cancel out, because g_A and g_B have converged to g^{∞}. The fiscal effects due to the earnings responses cancel out as well, because ΔT^{∞} is constant and thus the labor supply elasticities are identical. The fiscal migration effect is, however, positive, because some people would move to region B and pay higher taxes. If the starting wedge is characterized by $T_B < T_A$, the welfare can be improved by the opposite reform. Thus, though there are substantial differences between a differentiated and a unified tax schedule, they disappear in the limit at the top of the distribution. Finally, note that our simulations of the differentiated tax schedule indicate convergence at very high income levels (around \$4,000,000).[20]

REFERENCES

ABDELJAWAD, T., F. JARAD, and D. BALEANU (2009): "Variational Optimal-Control Problems with Delayed Arguments on Time Scales," *Advances in Difference Equations*, SpringerOpen, vol. 2009.

ACEMOGLU, D., and M. DELL (2010): "Productivity Differences between and within Countries," *American Economic Journal: Macroeconomics*, vol. 2(1), 169-188.

AKERLOF, G. A. (1978): "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning," *American Economic Review*, vol. 68 (March), 8–19.

ALBOUY, D. (2009): "The Unequal Geographic Burden of Federal Taxation," *Journal of Political Economy*, vol. 117(4), 635-667.

ARMSTRONG, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, vol. 64(1), 51-75.

[19] Note that the optimal tax formula for the uniform case simplifies to

$$2\varepsilon^{\infty}\frac{\tau^{\infty}}{1-\tau^{\infty}} = \frac{1-F(n)}{nf(n)}\left(2-\frac{\bar{\psi}}{\lambda}\right),$$

which is identical to (13) under the Pareto distribution and proper rescaling of the Lagrange multiplier.

[20] Note that, in line with Proposition 7, see Appendix B, the marginal tax rates converge for very high levels of potential income (around \$4,000,000 of potential income).

BARRO, R. J., and X. SALA-I-MARTIN (1991): "Convergences across States and Regions," *Brooking Papers on Economic Activity*, issue 1, pp. 107-182.

BASTANI, S., S. BLOMQUIST, and L. MICHELETTO (2013): "The Welfare Gains of Age-related Optimal Income Taxation," *International Economic Review*, vol. 54(4), 1219-1249.

BAYER, C., and F. JUESSEN (2012): "On the Dynamics of Interstate Migration: Migration Costs and Self-Selection, " *Review of Economic Dynamics*, vol. 15, 337-401.

BEA, (2020): "Interactive Data: GDP & Personal Income: Regional Data: Annual GDP by State: GDP per capita," (SAGDP10N Per capita real GDP by state), US Department of Commerce, Bureau of Economic Analysis, accessed online March 23, 2020, at https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1#reqid=70&step=1&isuri=1.

BEST, M., and H. J. KLEVEN (2013): "Optimal Income Taxation with Career Effects of Work Effort," Working Paper, LSE.

BLOMQUIST, S., and L. MICHELETTO (2009): "Nonlinear Income Taxation And Matching Grants In A Federation With Decentralized In-Kind Transfers," *International Economic Review*, vol. 50(2), 543-575.

BLOMQUIST, S., and L. MICHELETTO (2008): "Age-related Optimal Income Taxation," *The Scandinavian Journal of Economics*, vol. 110(1), 45-71.

BOADWAY, R., and P. PESTIEAU (2006): "Tagging and Redistributive Taxation," *Annales d'Economie et de Statistique*, ENSAE, issue 83-84, 123-147.

BORJAS, G. (1994): "The Economics of Immigration," *Journal of Economic Literature*, vol. 32(4), 1667-1717.

BEHRENS, K., and F. ROBERT-NICOUD (2015): Agglomeration theory with heterogeneous agents. In: *Handbook of regional and urban economics*, Vol. 5, 171-245. Elsevier.

BOUSTAN, L. P., P. V. FISHBACK, and S. KANTOR (2010): "The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression," *Journal of Labor Economics*, vol. 28(4), 719-746.

BREINLICH, H., OTTAVIANO, G. I., and J.R. TEMPLE (2014): Regional growth and regional decline. *Handbook of economic growth*, Vol. 2, 683-779. Elsevier.

CHETTY, R. (2006): "A New Method of Estimating Risk Aversion," *The American Economic Review*, vol. 96(5), 1821-1834.

CICCONE, A., and E. HALL (1996): "Productivity and the Density of Economic Activity," *American Economic Review*, vol. 86(1), 54-70.

CREMER, H., F. GAHVARI, and J.-M. LOZACHMEUR (2010): "Tagging and Income Taxation: Theory and an Application," *American Economic Journal: Economic Policy*, vol. 2(1), 31-50.

D'AMURI, F., G. I. P. OTTAVIANO, and G. PETRI (2010): "The Labor Market Impact of Immigration in Western Germany in the 1990s," *European Economic Review*, vol. 54, 550-570.

DIAMOND, P. (1980): "Income taxation with fixed hours of work," *Journal of Public Economics*, 1980, vol. 13(1), 101-110.

DIAMOND, P. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, vol. 88(1), 83-95.

DURANTON, G. and D. PUGA (2015): "Urban land use", In: *Handbook of regional and urban economics*, Vol. 5, 467-560. Elsevier.

EECKHOUT, J., and N. GUNER (2017): "Optimal Spatial Taxation: Are Big Cities to Small?," CEMFI Working Paper No. 1705, Centro de Estudios Monetarios y Financieros.

FRANK, D. (2009): "The Effect of Migration on Natives' Employment Outcomes: Evidence from the Fall of the Berlin Wall," INSEAD Working Paper 2009/54/ST, Fontainebleau, France.

FUJITA M., P. KRUGMAN and A.J. VENABLES (1999): "The Spatial Economy. Cities, Regions and International Trade," Cambridge (Mass.), MIT Press.

FUJITA, M. and J.-F. THISSE (2002): "Economics of Agglomeration," Cambridge University Press.

GOLOSOV, M., A. TSYVINSKI, and N. WERQUIN (2014): "A Variational Approach to the Analysis of Tax Systems," NBER Working Paper No. 20780, National Bureau of Economic Research.

GOMES, R., J.-M. LOZACHMEUR, and A. PAVAN (2018): "Differential Taxation and Occupational Choice," *The Review of Economic Studies*, vol. 85(1), 511-557.

GORDON, R. H., and J. B. CULLEN (2012): "Income Redistribution in a Federal System of Governments," *Journal of Public Economics*, vol. 96(11-12), 1100-1109.

GöLLMANN, L., D. KERN, and H. MAURER (2008): "Optimal Control Problems with Delays in State and Control Variables Subject to Mixed Control–State Constraints," Optimal Control Applications and Methods, online version doi:10.1002/oca.843.

IMMONEN, R., R. KANBUR, M. KEEN, and M. TUOMALA (1998): "Tagging and Taxing: The Optimal Use of Categorical and Income Information in Designing Tax/Transfer Schemes," *Economica*, vol. 65(258), 179-92.

ISTAT, (2020): "Conti Economici Regionali" ("Regional Economic Accounts"), Communicazione stampa (Press release) January 28, 2020, Italian National Institute of Statistics, accessed online March 5, 2020, at https://www.istat.it/it/archivio/237813.

JACQUET, L., E. LEHMANN, and B. VAN DER LINDEN (2013): "Optimal Redistributive Taxation with both Extensive and Intensive Responses," *Journal of Economic Theory*, vol. 148(5), 1770-1805.

KANBUR, R., and A. VENABLES (2005): "Spatial Inequality and Development," Oxford University Press.

KENNAN, J., and J. R. WALKER (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, vol. 79(1), 211-251.

KESSING, S.G., V. LIPATOV, and J.M. ZOUBEK (2015): "Optimal Taxation under Regional Inequality." CAGE Working Paper 253, University of Warwick.

KLEVEN, H. J., C. T. KREINER, and E. SAEZ (2009): "The Optimal Income Taxation of Couples," *Econometrica*, vol. 77(2), 537-560.

KLEVEN, H. J., C. T. KREINER, and E. SAEZ (2006): "The Optimal Income Taxation of Couples," NBER Working Papers No. 12685, National Bureau of Economic Research.

KRUGMAN, P. (1991): "Geography and Trade," MIT Press.

LEHMANN, E., L. SIMULA, and A. TRANNOY (2014): "Tax Me If You Can! Optimal Nonlinear Income Tax between Competing Governments," *The Quarterly Journal of Economics*, vol. 129(4), 1995-2030.

LIPATOV, V., and A. WEICHENRIEDER (2015): "Welfare and Labor Supply Implications of Tax Competition for Mobile Labor," *Social Choice and Welfare*, vol. 45(2), 457-477.

MANKIW, G., and M. WEINZIERL (2010): "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," *American Economic Journal: Economic Policy*, vol. 2(1), 155-176.

MIRRLEES, J. A. (1982): "Migration and Optimal Income Taxation," *Journal of Public Economics*, vol. 18, 319-341.

OTTAVIANO, G. I. P., and G. PERI (2008): "Immigration and National Wages: Clarifying the Theory and the Empirics," NBER Working Paper No. 14188, National Bureau of Economic Research.

OTTAVIANO, G. I. P., and G. PERI (2007): "The Effects of Immigration on U.S. Wages and Rents: A General Equilibrium Approach," CReAM Discussion Paper Series No 13/07, Center for Research and Analysis of Migration.

PIKETTY, T. (1997): "La redistribution fiscale face au chômage," *Revue française d'économie*, vol. 12(1), 157-201.

PANEL STUDY OF INCOME DYNAMICS (2013): Public use dataset. Produced and distributed by the Institute for Social Research, University of Michigan, Ann Arbor, MI (2013)

ROCHET, J.-C., and P. CHONé (1998): "Ironing, Sweeping, and Multi-Dimensional Screening," *Econometrica*, vol. 66(4), 783-826.

ROTHSCHILD, C., and F. SCHEUER (2013): "Redistributive Taxation in the Roy Model," *The Quarterly Journal of Economics*, vol. 128(2), 623-668.

ROTHSCHILD, C., and F. SCHEUER (2016): "Optimal Taxation with Rent-Seeking," *Review of Economic Studies*, vol. 83(3), 1225-1262.

SAEZ, E. (2002): "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses," *The Quarterly Journal of Economics*, vol. 117(3), 1039-1073.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, vol. 68(1), 205-229.

SCHEUER, F. (2014): "Entrepreneurial Taxation with Endogenous Entry," *American Economic Journal: Economic Policy*, vol. 6, 126-163.

SIMULA, L., and A. TRANNOY (2010): "Optimal income tax under the threat of migration by top-income earners," *Journal of Public Economics*, vol. 94(1-2), 163-173.

STIGLITZ, J. E. (1982): "Self-selection and Pareto efficient taxation," *Journal of Public Economics*, 17(2), 213-240.

WEINZIERL, M. C. (2011): "The Surprising Power of Age-Dependent Taxes," *Review of Economic Studies*, vol. 78(4), 1490–1518.

WILDASIN, D. (1991): "Income Redistribution in a Common Labor Market," *American Economic Review*, vol. 81(4), 757-774.

WILSON, J. D. (1992): "Optimal Income Taxation and International Personal Mobility," *American Economic Review*, Papers and Proceedings, vol. 82(2), 191-196.

YOUNG, A. (2013): "Inequality, the Urban-Rural Gap and Migration," *The Quarterly Journal of Economics*, vol. 128(4), 1727-1785.