Optimal Taxation under Regional Inequality*

Sebastian G. Kessing[a] Vilen Lipatov[b] Jens M. Zoubek[c]

April 7, 2015

Abstract

Combining an intensive labor supply margin with an extensive, productivity-enhancing migration margin, we determine how regional inequality and labor mobility shape optimal redistribution. We propose the use of delayed optimal control techniques to obtain optimal tax formulae with location-dependent productivity and two-dimensional heterogeneity. Our baseline simulations using the productivity differences between large metropolitan and other regions in the US indicate that efficiency-enhancing inter-regional migration reduces optimal marginal tax rates by up to three percentage points relative to a no migration benchmark. Optimal regionally differentiated marginal tax rates should be substantially lower in high productivity regions.

JEL classification: H11, J45, R12

Keywords: Optimal taxation, redistribution, regional inequality, migration, multi-dimensional screening, delayed optimal control

*We would like to thank David Agrawal, David Albouy, Mireille Chiroleu-Assouline, Bill Hoyt and Laurence Jacquet for their valuable comments. Financial support from the Deutsche Forschungsgemeinschaft (DFG), project No. KE 1693/3-1 is gratefully acknowledged.

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

[b]Goethe University Frankfurt and CESifo, lipatov@em.uni-frankfurt.de

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

1 Introduction

Regional productivity differences are large in many countries. Real per capita GDP in the New England Region was 40% higher than in the Southeast Region in the US in 2013 (BEA, 2014), for example. In Italy, the 2011 real per capita GDP of the Northern and Central Regions was even 71% higher than in the Southern and Islands Region (ISTAT 2013). The spatial dispersion of wages and incomes is well documented and the underlying causes are still subject to debate (Barro and Sala-i-Martin (1991), Ciccone and Hall (1996), Kanbur and Venables (2005), Acemoglu and Dell (2010) and Young (2013), among others). Given such productivity differences, the efficiency-enhancing potential of inter-regional mobility is substantial, and increases in personal income are key drivers of this mobility, see Kennan and Walker (2011). Centralized redistribution schemes such as a federal income tax or federal social transfers reduce inter-regional migration incentives, since an individual who migrates from a low to a high productivity region has to share the realized productivity gains with the government through higher taxes or lower transfers. This generates a trade-off for an inequality-averse policy maker between redistribution and efficiency-enhancing inter-regional migration. Contrary to the emigration of high-income earners to low-tax countries or the immigration of welfare recipients from less generous jurisdictions, the role of internal migration for optimal federal tax policy has been mostly neglected.[1] We develop a conceptual framework to analyze the implications of internal migration for an optimal tax-transfer policy and assess its quantitative importance. While our focus is on efficiency-enhancing migration between regions with permanent productivity differences, our approach may also be used to address the related optimal taxation problem that arises with respect to efficiency-enhancing migration in response to idiosyncratic shocks to regional labor markets, as discussed by Blanchard and Katz (1992), or, more recently, Yagan (2014).

We propose a two-dimensional optimal taxation model which combines an extensive, inter-regional migration decision with an intensive labor supply decision. Our key innovation is the productivity-enhancing nature of the migration margin. The actual or realized productivity of individuals of any given innate productivity is location-dependent, such that individuals can increase their productivity by migrating from a low to a high productivity region. Thus, the extensive migration margin also affects the intensive labor supply decision since productivity and, typically, the relevant marginal tax rate change whenever an individual decides to migrate, even though the same tax schedule applies nationwide.

This framework allows us to determine the optimal federal tax schedule as a function of

[1]Studies addressing external migration include Mirrlees (1982), Wildasin (1991), Wilson (1992), Lehmann et al. (2014) and others.

the government's redistributive preferences, the observed regional earnings distributions, the earnings elasticity, and the inter-regional migration elasticity. Our analysis shows that regional disparities and the possibility of efficiency-enhancing inter-regional labor mobility can be important determinants of the optimal tax schedule. Optimal marginal taxes tend to be below the benchmark without regional inequality since the decision to migrate to an area with higher productivity implies a fiscal externality. The size of this fiscal externality depends on the migration elasticity and on the inter-regional tax differential, which is itself a function of regional productivity differences and the tax schedule. If marginal tax rates are positive throughout the tax schedule, the fiscal externality is always positive, such that optimal marginal tax rates are lower compared to a situation with the same nationwide posterior productivity distribution but without migration. Moreover, for some subset of the productivity distribution, negative marginal tax rates are possible. This latter result is similar to other studies that have analyzed the optimal tax-transfer schedule with an intensive labor supply decision and the participation decision (see e.g. Saez 2002 and Jacquet et al. 2013).

Our framework provides a methodological contribution to the theory of optimal taxation. Making the intensive margin dependent on the extensive margin is a useful extension of the class of multi-dimensional screening models, originally discussed by Rochet and Choné (1998) and Armstrong (1996). We argue that this class of models can be fruitfully studied using the delayed optimal control approach as recently formally analyzed by Göllmann et al. (2008) in its entire generality. The approach is suitable to address a range of other multi-dimensional screening problems, where an extensive margin directly affects an intensive margin. Several other optimal taxation problems are characterized by a similar structure. The discrete decision whether to participate in the labor market or not, for example, affects productivity endogenously, given that non-participation tends to result in the depreciation of human capital. Similarly, discrete education decisions also determine productivity and interact with marginal tax rates and the intensive labor supply margin, such that our framework may also be applied to optimal taxation problems with endogenous education decisions.

We additionally study regionally differentiated tax-transfer schemes. 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 may also be challenged on the grounds of the violation of horizontal equity. Despite these caveats, regional differentiation can be an element of real world tax systems. From 1971 to 1994, the German tax system, for example, treated residents in West-Berlin differently from people in the rest of the country. Another example is the current path towards a more fiscally integrated Europe. As the Eurozone is moving towards deeper fiscal integration, it faces the choice between a system of explicit and implicit transfers between Member States combined with a different tax-transfer scheme within each Member State and the alternative of moving to an integrated Eurozone-wide tax-transfer scheme. This decision requires an understanding of the advantages and the challenges of a differentiated system vis-a-vis an integrated system.[2] Finally, nominally non-differentiated federal income taxation amounts to regionally differentiated taxation in real terms due to cost of living differences. Albouy (2009) has analyzed the efficiency costs associated with the resulting spatial distortions. Our analysis demonstrates that, from an optimal redistribution perspective, the optimal differentiated system should actually tax individuals less heavily in the more productive urban areas. Thus, the implied differentiation resulting from nominally undifferentiated taxation is the exact opposite of what we find to be optimal.

Conceptually, regionally differentiated tax-transfer schemes use the region of residence as an additional tag in the design of tax-transfer schemes. We add to the debate on tagging in optimal taxation by considering the region of residence as an endogenous tag. Moreover, the tag is not only correlated with the unobservable productivity but is, at least partly, directly causal for realized individual productivity. This gives rise to an interesting trade-off: On the one hand, the government can increase redistribution since the additional information reduces the efficiency costs of redistribution. On the other hand, differentiated taxes can be used to encourage productivity-enhancing migration.

To assess the quantitative importance of the additional constraint of internal migration for redistribution, we apply our framework to the US. We focus on the productivity difference between metropolitan and non-metropolitan areas, which is a relevant application of our analysis in many countries. Using micro data from the Panel Study of Income Dynamics (PSID) on rural and urban areas along with the actual tax-transfer treatment, we retrieve the underlying regional productivity distributions and establish the productivity gain from relocating from rural to urban areas for each productivity level. Taking into account the empirical evidence on labor supply and migration elasticities, we then simulate the optimal non-differentiated tax schedule that takes potential efficiency-enhancing migration into account. We contrast this schedule with the optimal tax schedule for the same posterior productivity distribution of individuals, but without migration. The results show that efficiency-enhancing mobility reduces optimal marginal tax rates by

[2] To the extent that the Member States are unrestricted by the center to decide on their own tax-transfer schemes, additional considerations of tax competition have to be additionally taken into account. See Lehmann et al. (2014) for the analysis of such considerations in the optimal taxation framework. Bargain et al. (2013) have contrasted a Member-States-based redistribution scheme to an integrated scheme in Europe. However, they address the implications for macroeconomic stabilization, whereas we study the efficiency of redistribution.

up to three percentage points in the baseline scenario. Productivity-increasing internal migration does not eliminate the case for progressive taxation, but it constitutes a quantitatively important constraint on redistribution. Finally, our simulations for the optimal regionally differentiated tax schedule indicate that marginal tax rates, and the overall tax burden for a given gross income, should be substantially lower in the more productive region, with optimal marginal tax rates differing by more than 15 percentage points for some income levels in the baseline scenario.

2 Related literature

The normative implications of efficiency-enhancing internal migration for optimal redistribution have, to the best of our knowledge, not been studied to date. The constraint of inter-jurisdictional or international mobility for the redistribution policy of a single jurisdiction or country, however, has received considerable attention within the optimal taxation literature and beyond, see, in particular, Mirrlees (1982), Wildasin (1991), Wilson (1992), Lipatov and Weichenrieder (2010), Simula and Trannoy (2011), Lehmann et al. (2014). Our approach reveals that labor mobility within a sufficiently large jurisdiction or between regions within a country can also be important for redistribution.

Conceptually, our analysis belongs 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 it is therefore complementary to our analysis of regionally non-differentiated and differentiated taxation by a single government. Moreover, individual productivity is not location-dependent in their analysis, and they only focus on the threat of migration, whereas actual efficiency-enhancing migration is at the heart of our approach. Gordon and Cullen (2012) also use an optimal taxation approach to study inter-regional migration in a model with several states. However, they focus on the assignment problem of whether redistribution should be carried out at the national or the subnational level and do not consider productivity differences. Jacquet et al. (2013) also study a two-dimensional optimal taxation model but focus on participation.

The structure of our approach owes much to Kleven et al. (2006, 2009) who study the optimal taxation of couples with cooperative households. Their analysis combines the intensive labor supply decision with the household's choice to become a single or a double earner household. However, our analysis differs in several important ways from their framework. First, we consider individuals and not households consisting of two persons whose respective incomes may be taxed separately. Secondly, in our approach individuals originally reside in different regions, such that the tax units not only differ among themselves regarding their costs to change their location, but also differ by the group they originally belong to. Finally and most importantly, we introduce an explicit consideration of endogenous individual productivity as a function of the extensive margin.

Rothschild and Scheuer (2014), and Rothschild and Scheuer (2013) and Gomes et al. (2014) also study optimal taxation of rent seeking activities and optimal taxation in the Roy model, respectively, using two-dimensional screening approaches. Wages are endogenously determined in their work, either by total labor supply in a given sector, or by total rent-seeking activities. Similarly, Scheuer (2014) studies entrepreneurial taxation with an endogenous decision, of whether to become an entrepreneur or a worker, where these decisions determine relative compensation in the aggregate. In our study individual productivity and thus market compensation, however, depend directly on the discrete decision of individuals and not on aggregate outcomes. Accordingly, our argument for optimally adjusting marginal tax rates is not based on the attempt to manipulate relative wages but to encourage efficiency-enhancing regional mobility.

Our analysis of the potential benefits of differentiated taxation relates to the increased interest in tagging in the design of tax-transfer-schemes. The idea that the government's information problem can be relaxed by using additional observable characteristics ("tags") that are correlated with the individual productivity goes back to Akerlof (1978) and has recently been discussed intensively in the optimal taxation literature, see Immonen et al. (1998), Weinzierl (2012), Mankiw and Weinzierl (2011), Boadway and Pestieau (2005), Cremer et al. (2010) and Best and Kleven (2013). We add to this literature in several ways. First, we consider the region of residence as a potential tag. Secondly, we explicitly study a tag that is endogenous and can be adjusted by individuals subject to some cost. Moreover, changing the tag directly affects productivity. In this respect, our paper is related to the literature that studies the interplay between human capital formation and optimal taxation, where the former shapes the productivity distribution and the latter influences incentives for human capital formation, see Stantcheva (2015) and the references therein. The endogeneity of productivity also relates our work to Best and Kleven (2013) who consider a dynamic setting where individual productivity intertemporally depends on the previous intensive labor supply decisions.

Albouy (2009) has argued that non-differentiated nominal federal taxation effectively implies de facto regionally differentiated taxation due to cost-of-living differences. He reasons that differential taxation distorts the spatial allocation in the economy and analyzes the associated efficiency costs and the implied interregional redistribution, but his analysis does not consider the question of optimal redistribution between heterogenous individuals. Our normative approach to regionally differentiated taxation can be regarded as complementary to his work, since we ask the question whether and to what extent federal taxes should be regionally differentiated for redistribution purposes, if such differentiation were possible. Finally, Eeckhout and Guner (2015) also study the effects of a progressive federal income tax on the spatial allocation of economic activity with a heterogenous population, and also consider regionally differentiated taxation, but they do not use a Mirrleesian optimal taxation framework and do not consider jointly the interaction of the intensive labor supply decision and the inter-regional migration decision.

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 can neither observe productivity nor migration costs. There are two regions, $i = A, B$, with total population normalized to two. Originally, half of the population resides 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. Without loss of generality we assume that region B is the more productive region, so that $\omega(n) > n$. The lag function $\kappa(n) \equiv w(n) - n$ describes the increase in ability from migration. Innate productivity is distributed in each region i according to the unconditional probability distribution $f(n)$ on $[n_{\min}, n_{\max}]$.[3] As in most of the optimal taxation literature, we treat wages as exogenous and independent of individual labor supply and aggregate migration decisions. Accordingly, the analysis applies to a situation where the effect of migration flows on wages is negligible. 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) together with D'Amuri et al. (2010), and

[3]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 further aspects.

Frank (2009) for evidence from the German reunification.[4]

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

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

where l is an indicator variable that takes the value of 1 in case of migration. The function $h(\cdot)$ is increasing, convex and twice-differentiable. It is normalized such 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, which itself depend on gross income, $c_i = z_i - T_i(z_i)$. Total migration costs are potentially made up of two components, $q = q^c + q^h$, where q^c is the cost of moving (the need to adapt to new conditions, to learn a new language in case of mobility between regions where different languages are spoken, the transaction costs of selling your old house and buying a new one, etc.), and q^h is the utility derived from being at home and benefitting from the existing social networks. To isolate the impacts of the two types of heterogeneity, it is useful to consider them separately. The pure *cost of moving* model sets $q = q^c$ and $q^h = 0$; the pure *home attachment* model uses $q = q^h$ and $q^c = 0$. Ex post, i.e. after migration has taken place, heterogeneity in q^c reflects the differences between individuals who migrate, whereas heterogeneity in q^h reflects the differences between individuals who stay in their home region. In what follows we focus on the cost of moving case, but, with some minor modifications, the home attachment case is quite analogous. However, our optimal tax schedules and their derivations are sufficiently general to encompass both cases.

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, given 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 as a function of

[4]For similar findings in case of immigration of foreigners see Borjas (1994) and Ottaviano and Peri (2007, 2008).

gross earnings and the region of residence is defined as

$$\varepsilon_i \equiv \frac{1-\tau_i}{z_i} \frac{\partial z_i}{\partial\left(1-\tau_i\right)} = \frac{n_i h'\left(\frac{z_i}{n_i}\right)}{z_i h''\left(\frac{z_i}{n_i}\right)}.$$

To focus on regional productivity differences, we assume that $\varepsilon_i = \varepsilon$ for all individuals and independent of the region. This simple benchmark arises with an iso-elastic formulation, i.e. $h\left(\frac{z_i}{n_i}\right) = \frac{1}{1+\epsilon}\left(\frac{z_i}{n_i}\right)^{1+\epsilon}$, such that $\varepsilon_A = \varepsilon_B = 1/\epsilon$, for example. Furthermore, we require the following property.

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

Consider now the migration decision. We denote by $p\left(q|\, n\right)$ the density of q conditional on n, and by $P\left(q|\, n\right)$ the cumulated distribution of q conditional on n. Conditional on residing in region i, the individuals' choice of gross earnings is determined by (2), which allows to define indirect utility conditional on the place of residence and net of the costs of moving or the benefits of residing in one's home region as

$$V_i\left(n_i\right) = z_i - T_i\left(z_i\right) - n_i h\left(\frac{z_i}{n_i}\right).$$ [5]

Individuals will move from i to $j, j = A, B, i \neq j$, whenever their migration costs are below the net gain from moving, such that $\bar{q}_i \equiv \max\left\{V_j\left(n_j\right) - V_i\left(n_i\right), 0\right\}$ 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

$$\sum_i \int_{n_{\min}}^{n_{\max}} \int_0^{+\infty} \Psi\left(V_i\left(n\right) - q^c l + q^h\left(1-l\right)\right) p\left(q,n\right) f(n) dq dn, \tag{3}$$

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

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

[5] Note that, in a more general model that allowed for region-specific price levels, the indirect utility received in each region would have to be qualified by using the correpsonding local prices, but the migration decision would follow the same general logic.

Moreover, the government's tax schedule needs to be incentive compatible. This implies

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

where the dot above a variable denotes its derivative with respect to n. Moreover, in case of non-differentiated taxation, $T_A(z) = T_B(z)$. We show in the appendix 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 government's redistributive tastes may be represented by region-dependent social marginal welfare weights. In terms of income, our welfare weights will take the form of

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

for the cost of moving model, where $\bar{q}_i(z) \equiv \max\{V_j(z) - V_i(z), 0\}$.

4 Optimal unified taxation

We first investigate the optimal non-differentiated 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 has been analyzed by Göllmann et al. (2008) in its entire generality. In our model, the delay is a non-fixed lag, though, given that we do not restrict the productivity gain from moving to be constant but treat it as a function of the innate productivity. The necessary conditions for optimal control in such a setting are presented in Abdeljawad et al. (2009). While we explicitly solve the problem in the Appendix to derive all our results rigorously, we first follow here the intuitive perturbation approach pioneered by Piketty (1997) and Saez (2001) to derive the optimal tax scheme. This heuristic derivation allows to disentangle the economic forces that determine the shape of marginal tax rates along the optimal tax schedule, including the effects generated by the possibility of efficiency-enhancing migration.

We use the endogenously realized distribution of gross incomes in both regions denoted by $v_i(z_i)$, and we denote by k the endogenously defined, strictly increasing 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. $z_B = k(z_A)$.[6] We consider an increase in taxes for all individuals above

[6] 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 the fact that

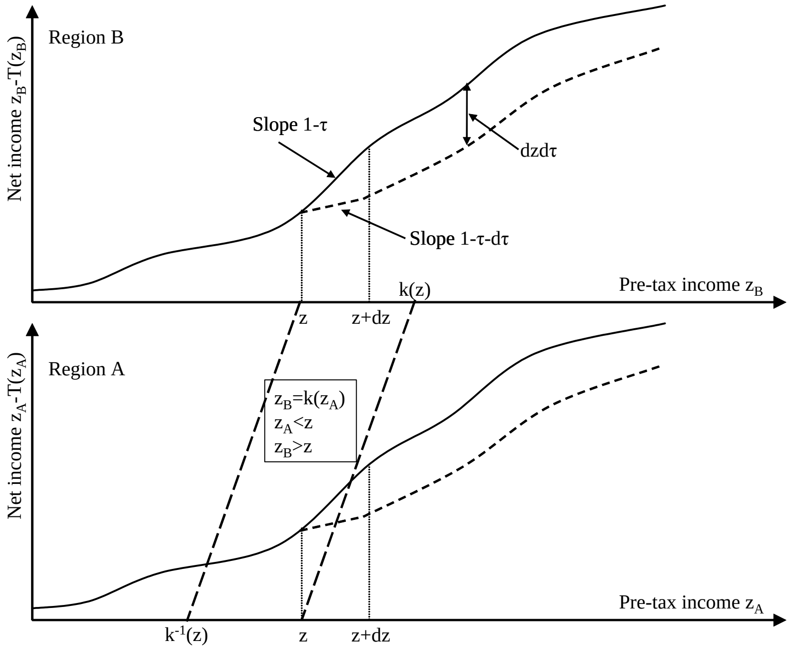

Figure 1: The migration effect comes into play for individuals for which $z_A' < z$ and $z_B' \geq z$.

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 All taxpayers in either region with gross incomes above z pay additional taxes of $dzd\tau$. The net welfare effect of this tax payment for an affected individual in Region i with gross earnings z' is given by $dzd\tau\,(1-g_i\,(z'))$, and the total effect is then

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

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

Behavioral effect Individuals in the band $(z, z+dz)$ will 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 will reduce its income by $-d\tau\varepsilon\frac{z}{1-\tau}$. There are approximately $dz\,[v_A(z)s_A(z) + v_B(z)s_B(z)]$ of these individuals. The total effect on tax revenue is

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

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

the gross income also depends on the tax schedule.

also $z_B' > z$, such that 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 $\bar{q}$ and $\bar{q} - dzd\tau$ will now decide not to migrate. There are $p(\bar{q}|\, z)\, v_A(z) dzd\tau$ affected individuals at any appropriate level of income z with a resulting tax effect of $T_A(z) - T_B(k(z))$ for each of them. The total migration effect is thus

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

where $\tilde{z} \equiv k^{-1}(z)$. Note that there is an endogenous effect on the income distribution in each region. This affect 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*

$$\frac{\tau}{1-\tau} = \mathfrak{A}(z)\mathfrak{B}(z)\left[\mathfrak{C}(z) + \mathfrak{D}(z)\right], \tag{7}$$

where

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

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 the Appendix C. ∎

It is straightforward to compare the result with the alternative 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

$$\frac{\tau}{1-\tau} = \mathfrak{A}(z)\mathfrak{B}(z)\mathfrak{C}(z). \tag{8}$$

With $\mathfrak{D}(z) < 0$, the disincentive effects of higher tax rates on productivity-increasing mobility tend to reduce marginal tax rates, but note that $\mathfrak{B}(z)$ and $\mathfrak{C}(z)$ are endogenously determined by the migration flows, such that (7) and (8) cannot be directly compared, in general. To make the result more formal, we consider the benchmark in which the government faces the same distribution of realized productivity v and of population shares s as in the posterior situation generated by the optimal tax schedule with migration. Given this posterior distribution assume that there is, or the government believes so, no reaction in terms of location choice from the tax system, i.e. that the posterior distribution is fixed and individuals only react to the taxation through their intensive labor supply margin. In this case the optimal tax follows the formula (8) with terms $\mathfrak{A}(z) > 0, \mathfrak{B}(z) > 0, \mathfrak{C}(z) > 0$ identical to the ones in (7). In this case, for $\mathfrak{D}(z) < 0$, we have $\tau_m < \tau_o$, where the subscripts m and o indicate the migration and the no migration case, respectively. This allows us to formulate the following proposition:

Proposition 2 *A government neglecting the effect of taxes on the migration decision, but facing the distribution of realized productivity generated by migration, should set higher marginal tax rates than a government taking the migration decision into account, if marginal tax rates are positive.*

Proof. Going through the derivation of the optimal tax formula in the Appendix A under the assumption that the effect of tax on 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 (that is equivalent to $\mathfrak{D}(z) \leq 0$), the result immediately follows. If marginal tax rates are positive, this condition is always fulfilled. ■

Note that positive marginal tax rates are a sufficient but not a necessary condition for this result. Whenever $\mathfrak{D} < 0$ for any given level of gross income z, marginal tax rates are lower with migration relative to the no-migration benchmark with the posterior distribution. Thus, whenever efficiency-enhancing migration implies a positive fiscal externality at a given innate productivity level, marginal tax rates should be reduced to take the fiscal externality of inter-regional migration appropriately into account. This constrains optimal redistribution beyond the classic adverse labor supply responses.

Another direct implication of the optimal unified taxation formula (7) is stated in the following proposition.

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

Proof. For $\mathfrak{D}(z) < 0$, it is possible that $\mathfrak{C}(z) + \mathfrak{D}(z) < 0$, and thus $\tau < 0$.[7] ■

Similar to the findings of other studies that combine an extensive participation decision with the intensive labor supply decision also endogenous mobility between regions of different productivity can give raise to negative marginal tax rates.

Using the posterior distribution as in Proposition 2 is our preferred benchmark as it allows switching migration on and off while keeping the productivity distribution fixed. This benchmark also corresponds directly to the empirically observed spatial distribution of individuals and productivity at a given point in time. Accordingly, we also focus on it in our simulations in Section 6. However, for completeness, another benchmark to compare our optimal solution to is an economy with the ex ante distribution of productivity and without internal migration. As we show in the Appendix, the comparison of optimal marginal tax rates is less clear cut in this case given the endogeneity of the posterior productivity distribution, when allowing for migration. This also impacts on $\mathfrak{B}(z)$ and $\mathfrak{C}(z)$, and these effects may drive optimal marginal tax rates in the opposite direction. Formally, we provide a sufficient condition for mobility to decrease the marginal tax rates for this alternative benchmark in the Appendix.

Finally, we make the following remark about the welfare comparison in the unified taxation case.

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

Proof. Consider the tax schedule that maximizes welfare if migration is not allowed. Migration brings a Pareto improvement, because individuals move only if they find themselves better-off. Furthermore, with migration to the richer region only, the government budget constraint will not be violated, if the tax is nondecreasing in income. Thus, under the same tax schedule the welfare may not decrease with the introduction of a migration possibility. Finally, the government will change the tax schedule only if it brings a further increase in welfare. Thus, the welfare with migration may not be lower than welfare with no migration, Q.E.D. ■

5 Optimal differentiated taxation

We now consider the possibility that the central government can choose differentiated tax schedules for both regions. If there were regional productivity differences but no migration,

[7] By simulative example (available upon request) it can be illustrated that this is the case for certain parameter values.

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 heterogenous with respect to their migration costs, which are unobservable by the government. Again we employ the perturbation approach and delegate the formal proofs to the Appendix.

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. 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 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 $dz_A d\tau_A \left(1 - g_A\left(z'_A\right)\right)$ and the total effect is

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

Behavioral effect: Individuals in the band $(z_A, z_A + dz_A)$ will 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 will reduce 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, such that the total effect on tax revenue is

$$L_A = -d\tau_A dz_A \varepsilon \frac{\tau_A}{1-\tau_A} z_A v_A(z_A) s_A(z_A).$$

Migration effect: An increase in taxes for all 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 $p\left(\bar{q} | \, z_A\right) v_A(z_A) dz_A d\tau_A$ 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 for people of income z, the argument is analogous, as we show formally in the Appendix. The total effect is thus

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

In the optimum, these effects should cancel out such that optimal marginal tax rates can be characterized by

$$\begin{aligned}\frac{\tau_A}{1-\tau_A} &= \frac{1}{\varepsilon}\frac{1}{z_A v_A(z_A) s_A(z_A)} \\ &\times \int_{z_A}^{\infty} \left\{\left[1-g_A\left(z_A'\right)\right] s_A(z_A') + \left[T_B(k(z_A')) - T_A(z_A')\right] p\left(\bar{q}\right| z_A')\right\} v_A(z_A') dz_A'\end{aligned} \tag{9}$$

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, such that

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

Both optimal tax schedules are derived rigorously in the Appendix. The optimal tax formulae not only differ by the different average welfare weights and the respective productivity distribution above the gross income level for which taxes are increased, but they also take the fiscal externality from the effect on migration into account. Typically, this externality will be negative for the high productivity region and positive for the low productivity area. Accordingly, from the optimal tax schedules under differentiated taxation (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 in the low productivity region τ_A is increasing in the difference in total tax liability between the high and the low productivity regions, and the marginal tax rate in the high productivity region τ_B is decreasing in this difference in total tax liability.*

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

Intuitively, the larger the potential fiscal gains are from working in the high productivity region instead of working in the low productivity region, the more the government distorts labor supply in the low productivity region and the less it distorts labor supply in the high productivity region. This indicates that the marginal tax rates are used to steer migration flows. Differences in the demogrant may be used instead to target redistribution by using the region as a productivity tag. In the Appendix we additionally rearrange the optimal taxation formulae (9) and (10) to show how the regional semi-elasticities of migration act as a correction factor to the region-dependent marginal social welfare weights in the determination of optimal marginal tax rates.

5.1 Asymptotic properties with differentiated taxation

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 $a > 1$ ($f(n) = C/n^{1+a}$). Moreover, we also 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 $n \to \infty$. We assume that for sufficiently large n, $\omega(n) = n + c$, where $c \geq 0$ is a finite constant. In this case, the following proposition arises:

Proposition 5 *Under the assumptions on convergence formulated above, (i) 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*

$$\frac{\tau^\infty}{1-\tau^\infty} = \frac{1}{a\varepsilon^\infty}\left(1 - \frac{\bar{\psi}}{\lambda}\right). \tag{11}$$

Proof. *The proof is left to Appendix AA.* ■

The intuition for zero difference of top taxes is similar to that in Kleven et al. (2009). Namely, starting from a wedge between T_B and T_A, welfare could be increased by marginally reducing this wedge due to the migration effect. If T_B is decreased, some people move to Region B and pay higher taxes; if T_A is increased, some people move to Region B and pay higher taxes. 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.[8]

5.2 Partial differentiation of the tax schedule

Once we allow for differences in the tax schedules of the two regions, it makes sense to ask how different the schedules should be, starting from a situation with undifferentiated marginal tax rates. In particular, in the following we show that (i) starting from identical tax schedules and not allowing different marginal taxes on income, it is optimal to make a transfer to the more productive region; (ii) starting from different tax schedules with the same marginal taxes for the same ability, it is optimal to lower marginal taxes in the high productivity region while raising them in the low productivity region.

[8] 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 (11) under the Pareto distribution and proper rescaling of the Lagrange multiplier.

5.2.1 On desirability of transfers given equal marginal taxes

Consider a tax system that is separable in the sense that the same income in two regions faces the same marginal tax. This is very similar to uniform taxation, but now we are allowed to charge the same incomes different taxes. The maximization problem of the government is the same apart from the feature that instead of the restriction that $\Delta T = 0$ we have the restriction $\Delta T = C$, where C is constant in n and $\Delta T := T_B(n) - T_A(n)$. For this setting, we can formulate the following proposition:

Proposition 6 *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. *The proof is left to Appendix AA.* ■

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. This result is independent of the interpretation of migration costs.

5.2.2 On suboptimality of equal marginal taxes for the same ability

Consider a tax schedule that is separable in the sense that $\tau_A = \tau_B$. Starting from this schedule, the following proposition shows that in the cost of moving model decreasing the marginal tax in Region B and increasing it in Region A would be desirable:

Proposition 7 *If Ψ' is convex, q and n are independently distributed and $\omega'(n) \geq 1$, it is optimal to introduce some wedge in marginal taxes to the system of separable taxation of the two regions. In particular, in the cost of moving model it is optimal to decrease the marginal tax in the high productivity region and increase it in the low productivity region.*
Proof. *The proof is left to Appendix AA.* ■

The proof is quite intuitive: it is based on the fact that in the cost of moving model the difference in marginal welfare weights of residents of regions A and B is decreasing with 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 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 distribution is related to substantially larger productivity gains than migration in the upper part of 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 efficiency-enhancing migration for the design of tax policy and optimal redistribution.[9] 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 implement our framework empirically, we divide the US into a rural (low productivity) and an urban (high productivity) area. We use the empirically 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 standard optimal tax formula 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 these synthetic high and low productivity regions. In what follows, we first specify functional forms and parameters used in the simulations and then describe the calibration procedure.

6.1 Simulation specification

For simulations we use iso-elastic utility $h(\frac{z}{n}) = (\frac{z}{n})^{1+\epsilon}/\left(1+\epsilon\right)$ with a constant earnings elasticity $\varepsilon = \frac{1}{\epsilon}$ as in Saez (2001). Paralleling our theoretical derivations, we concentrate on the cost of moving model, hence $q = q^c$. Moreover, we follow Kleven et al. (2009) by assuming a power law distribution for the costs at the extensive 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 we use the constant rate of risk aversion (CRRA) function $\Psi\left(V\right) = V^{1-\gamma}/(1-\gamma)$, where the parameter γ measures the government's preference for equity. In our baseline scenario we choose $\gamma = 1$, hence $\Psi\left(V\right) = \log(V)$ in line with Chetty (2006). Finally, the simulation is done in a way that, with the optimal tax rates obtained, the ratio of exogenous budget expenditures E to aggregate production is 0.25 as in Saez (2001).

[9] Our simulations use a modified and extended version of the code developed by Kleven et al. (2009). We would like to express our gratitude to them for providing their original code to us.

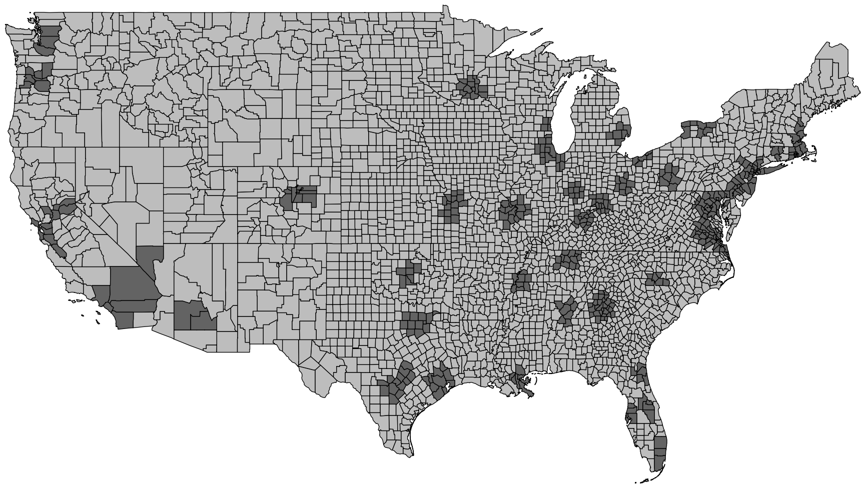

Figure 2: Split of the mainland US districts into two regions: Region A consists of rural counties (light), region B of urban ones (dark). Boundaries taken from US Census Bureau (census.gov: Cartographic Boundary Shapefiles).

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 focussing on the considerable productivity discrepancy between rural and urban regions in the US.[10] 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 PSID data provides the RUCC for each individual's county of residence. We treat all counties belonging to class 1 to be the urban region (Region B), and counties of classes 2 through 9 to be the rural region (Region A) as illustrated by Figure 2.

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 2006 individual gross labor income data from the 2007 PSID for unmarried, working individuals with no children under 18 years with the corresponding marginal tax rate from the NBER TAXSIM model. This procedure is similar to Best and Kleven (2013), who differentiate individuals by age, whereas we use a regional distinction. As suggested by Diamond (1998) and Saez (2001), very high incomes are well approximated by a Pareto distribution. Therefore, the skill distributions are mod-

[10] This procedure specifies regions in an economic rather than administrative or purely geographical sense and is applicable to most countries.

ified by assuming a Paretian shape for each top quintile of the respective distribution that is gross incomes above \$78,000 (\$59,000) in the urban (rural) region. This parallels the assumptions used in Kleven et al. (2009). These Paretian additions exhibit reasonable coefficients for the US. Corresponding to the assumptions by Jacquet et al. (2013) or Best and Kleven (2013), we estimate the specific Pareto parameter for each of our regions by regression of the gross income ratio z_m/z between \$78,000(\$59,000)−\$750,000, where z_m is the average of earnings above z. The recovering procedure then is as follows: each individual's ability is computed from individual maximization using its income data from the PSID, the actual marginal tax rate corresponding to this income level from TAXSIM together with the earnings elasticity ε and the functional assumption for $h(\frac{z}{n})$. For all incomes above \$78,000 (\$59,000) the computed ability is then substituted by the respective Pareto value. This procedure is applied to both regions. Figure 3A depicts the computed skill distributions in regions A and B, where, for visualization, the discrete values are smoothed continuously using kernel estimation. The resulting descriptive statistics for both areas exhibit a median (mean) ability difference, that amounts to, in terms of our theory framework, a difference in potential income of 11.7% (32.8%) between the rural and the urban areas.

Third, we estimate the lag function $\kappa(n)$ from the data by using the differences in mean ability of each frequency percentile of the two regions. This difference is then assumed to be the productivity increase for the mean person (sampling point) of each percentile. The function is estimated by linear interpolation using these sampling points and smoothed afterwards.[11] We obtain an increasing nonlinear function presented in Figure 3B.[12] We find a substantial productivity increase from migration from a rural to an urban district, in particular for the types with potential annual incomes of around \$50,000 − \$200,000 at their origin. For high ability levels (above \$208,000) we assume that $\kappa(n)$ is constant, given that fitting the lag function above that level becomes rather arbitrary given the lack of sampling points in this range.

[11] In detail, Matlab routine 'interp1' is applied.

[12] Our empirical construction of ω may be justified as follows. By the definition of the transformation function ω the cumulative distribution functions of ability in two regions are related as $F_B(n) = F_A(\omega(n)) \forall n \in [n_{\min}, n_{\max}]$. Then, at each α-percentile it is true that $F_B(n_\alpha) = F_A(\omega(n_\alpha)) = \alpha$. From the properties of cdfs assumed (strictly increasing) it follows that function ω can be reconstructed from F_A and F_B at any point $n_\alpha \forall \alpha \in [0, 1]$ or, equivalently, $\forall n \in [n_{\min}, n_{\max}]$. In our simulation, we do not observe the true cdfs, but only their empirical counterparts, $\hat{F}_A$ and $\hat{F}_B$. The proof of statistical properties of our approach is beyond the scope of this paper. Note, however, that under the assumption that we actually observe the true cdfs at a limited number of data points m, a smooth interpolation is the best way to fill in the missing values in the estimates of F_A and F_B , because the cdfs are smooth by continuity of the pdfs. Once we have the estimates of cdfs defined over the whole domain, we can recover the function ω for any point in the domain. The only remaining problem then are the corners. Whereas theoretically we should observe abilities starting from $n_{\min}$ in one region and $\omega(n_{\min})$ in the other region, empirically we observe only the lowest income category and hence $n_A(z_{\min}) = n_B(z_{\min})$.

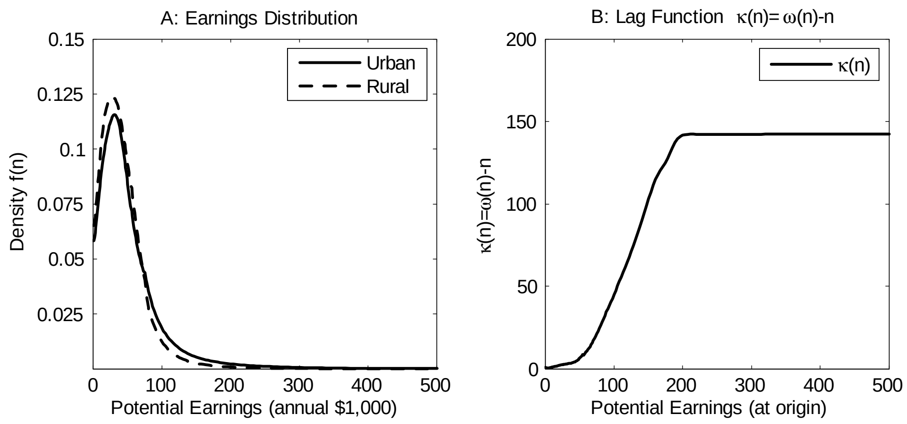

Figure 3: Revealed true abilities (plot A) and revealed lag function $\kappa(n)$ with a fixed productivity gain for highly productive individuals (plot B) for the two chosen regions of the US based on data from PSID 2008/09 and TAXSIM.

Fourth, the migration cost distribution is calibrated by choosing the migration elasticity η and the parameter $q_{\max}$. We use $\eta = 1.5$ for the migration elasticity, and additionally consider lower and higher values in a range between $\eta = 0.2$ and $\eta = 2$ to assess the sensitivity of the results with respect to this parameter.[13] By choosing $q_{\max}$ we then calibrate the migration costs such that the typical move costs around $\$34,000$ which is the value estimated by Bayer and Juessen (2012) and, likewise, is in the range obtained by Kennan and Walker (2011).

6.3 Results for non-differentiated taxation

Figure 4 illustrates the simulation outcomes for incomes up to $\$500,000$ under our parameter settings. As is standard in the optimal tax literature following Diamond (1998) a U-shaped pattern appears. In the no-migration case (dashed lines of 4A,C,D), which uses the posterior productivity distribution outcome with migration, the government chooses a higher marginal tax rate compared to the migration case (solid lines) as stated in Proposition 2. Apparently, from Figure 4A (baseline case), the migration effect reduces marginal tax rates by up to three percentage points in the most relevant level of mid-abilities. Moreover, productivity-increasing migration seems to smooth the U-shaped pattern (Figure 4A). Interestingly, the feature of smoothing the U-shaped pattern is likewise obtained by Best and Kleven (2013) in their setting. In Figure 4B we plot the marginal tax rate differences between the migration and the no-migration case for alternative values of the

[13]Reliable econometric estimates of the parameter η for our synthetic urban and rural regions are unavailable, but the range of values we consider covers available estimates on inter-state migration in the US, see, inter alia, the results of Greenwood (1969) or Kennan and Walker (2011), to the extent that these can be turned into the elasticity concept we use.

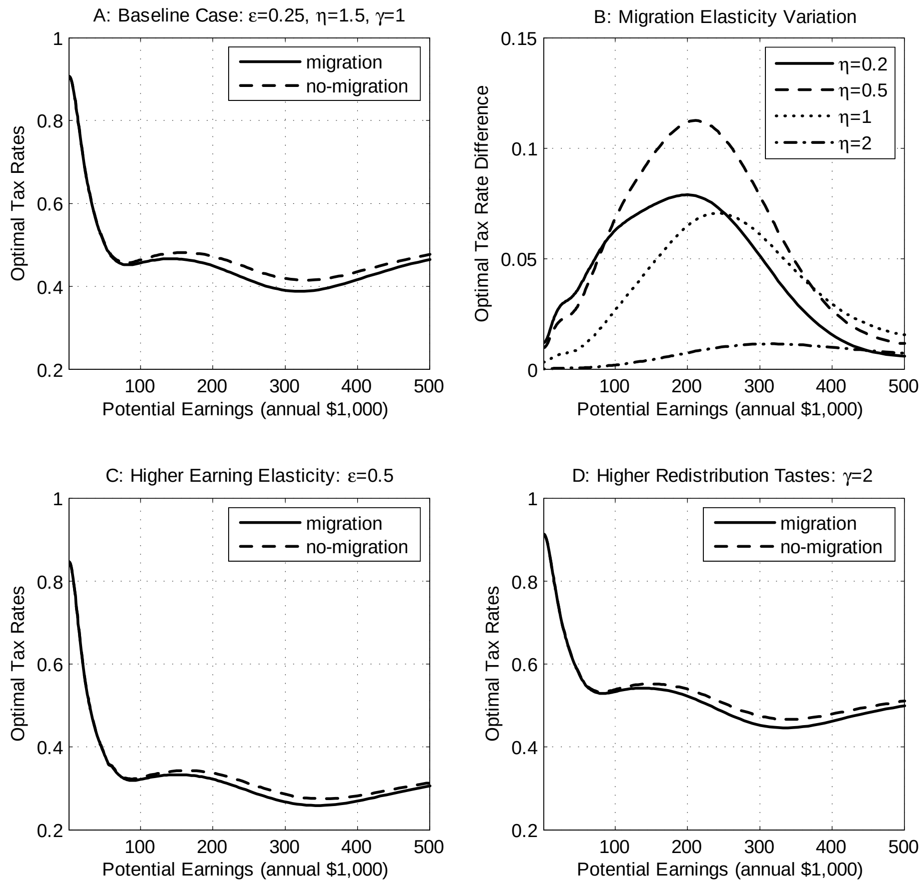

Figure 4: Optimal uniform tax simulations for the US based on data from PSID 2008/09 and TAXSIM.

migration elasticity. A higher migration elasticity reduces the tax rate differences, a lower migration elasticity than in our baseline scenario increases it. Overall, however, this relationship is not monotone. As is also immediate from theory, the differences in marginal tax rates are again reduced for relatively low values of η, compare the tax rate differences for $\eta = 0.2$ and $\eta = 0.5$ in 4B.[14] Note also from 4B that the marginal tax rate differences may be quantitatively substantially more pronounced than the three percentage points in our baseline case, exceeding ten percentage points for $\eta = 0.5$ for some ability levels.

As shown by Figure 4C, a higher labor supply elasticity lowers marginal tax rates throughout the ability distribution and also reduces the gap between the migration and the no-migration tax schedule. Intuitively, lower progressivity tends to reduce the difference in tax payments between individuals of the same innate productivity, such that the fiscal migration externality is reduced. However, it also affects gross earnings in ei-

[14] Our robustness checks indicate that the differences in marginal tax rates are further reduced for relatively low values of the migration elasticity, and completely disappear for $\eta \to 0$.

ther region, which typically affects the magnitude of the fiscal externality in the opposite direction. Apparently, in the case of the higher labor supply elasticity the former effect dominates. Finally, Figure 4D depicts a simulation with higher redistributive tastes which lead to higher marginal tax rates throughout the ability distribution. The fiscal migration externality and the corresponding difference between the optimal tax schedules with and without migration are again subject to opposing effects caused by the change in progressivity. However, in this case, the outcome is more balanced.

6.4 Results for differentiated taxation

The results for differentiated taxation are depicted in Figure 5. We continue to use the same parameter settings for $q_{\max}$ and E. As in our baseline case for non-differentiated taxation, Figure 5A uses $\eta = 1.5$. Figure 5A displays the corresponding optimal marginal tax rates for both the rural and the urban region. Furthermore, the optimal absolute tax liabilities are illustrated for this setting in Figure 5C, too. Additionally, we display the optimal marginal tax rates for a lower and a higher migration elasticity, $\eta = 0.5$ and $\eta = 2$, respectively, in Figure 5B and 5D. For both regions U-shaped marginal tax rate patterns reappear, but there are important differences between the two regions. For all three different values of the migration elasticity we show, the marginal tax rates of both regions are closely together for low abilities but begin to diverge sharply for potential earnings above $\$60,000$. The marginal tax rates in the less productive region (rural) are equal or higher than those in the more productive region (urban) as suggested by Proposition 7. With a lower migration elasticity of $\eta = 0.5$ the difference in marginal tax rates is even considerable for low ability levels (Figure 5B). The U-shaped pattern is more distinctive for the rural region (solid line), but becomes also more obvious for the urban region (dashed line) in this case (Figure 5B). However, for the urban region (dashed line) the U-shaped pattern seems to be smoothed out by migration to cover a wider range, in general. With the lower migration elasticity of $\eta = 0.5$ as shown in Figure 5B the marginal tax rate difference increases. The marginal tax rate in the urban region goes down to approximately 0.3 in this case. The marginal tax rate difference is up to 30 percentage points, see Figure 5B. Conversely, for a higher migration elasticity the opposite is true as drawn in Figure 5D, which uses $\eta = 2$.

In comparison to the optimal non-differentiated tax schedule, marginal tax rates are distinctly lower in both regions for low ability levels. However, for mid-abilities and individuals with high potential earnings in the urban region the marginal tax rates are lower relative to the non-differentiated case. The opposite holds true for these skill groups in the rural region (Figure 4A and 5A).

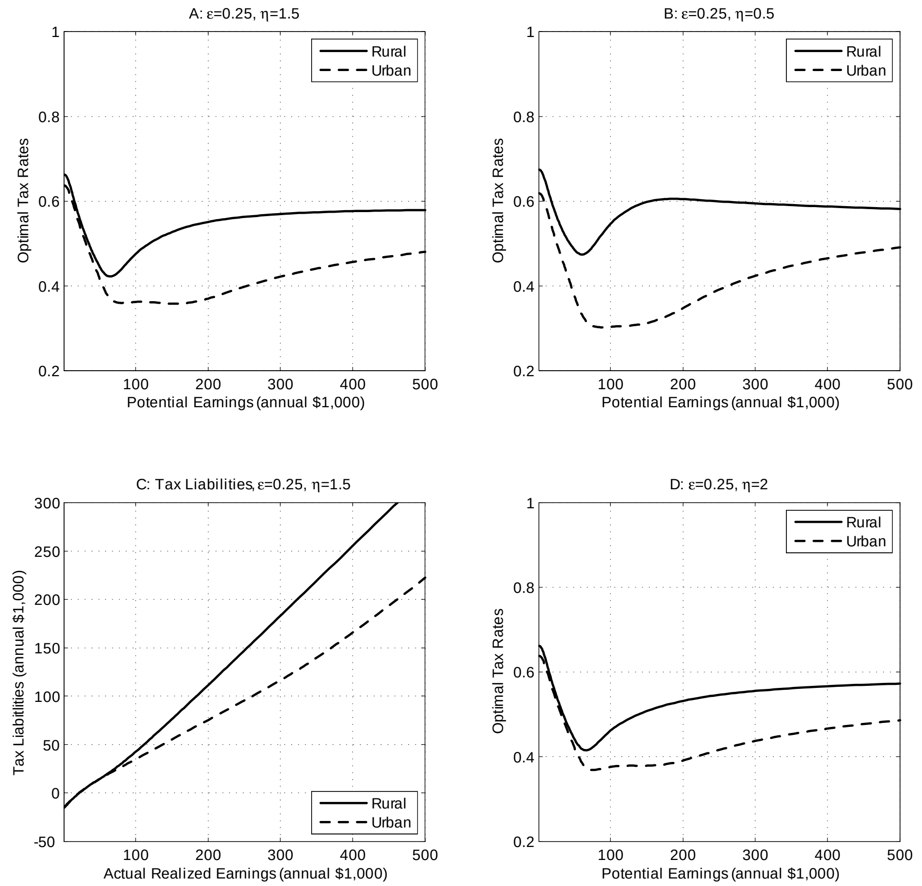

Figure 5: Optimal differentiated tax simulations for the US based on data from PSID 2008/09 and TAXSIM.

In Figure 5C we additionally depict absolute tax liabilities as a function of actual realized earnings, the z_A and z_B in terms of our theory. The outcome displays a basic income support of roughly the same amount ($\$15,000$) in both regions, which is taxed away quickly in both regions as income grows. It is evident from Figure 5C, and corresponding to the marginal tax rate schedule (Figure 5A), that the tax liability schedules diverge from around $\$60,000$ in actual earnings. Individuals with the same realized income should face a lower tax burden in the more productive region for all incomes above that level. Finally, given the posterior regional distribution of the population, the average per capita tax liability amounts to $\$8,000$ in the rural and to $\$12,500$ in the urban region.

It is interesting to compare our optimally differentiated tax schedule for redistribution with the actual regional differentiation of income taxation due to cost of living differences as discussed by Albouy (2009). He argues that the inhabitants of urban agglomerations are effectively taxed more heavily than individuals residing in rural areas. In our specification, optimal differentiated taxation bears more heavily on the individuals residing in the rural areas, with the exception of the very low income earners. These findings imply that the current favorable tax treatment of low productivity regions which arises from the interaction of cost-of-living differences with nominally non-differentiated federal income taxation may not only be criticized from pure efficiency grounds as argued by Albouy (2009), but also appears to be opposite to an optimal differentiated tax system derived from our redistribution perspective with location-specific productivity.

7 Concluding remarks

Regional inequality and the corresponding possibility of efficiency-enhancing migration can be an important determinant of the optimal redistributive tax-transfer scheme. A government that is constrained to use a unified redistribution policy faces an additional equity-efficiency trade-off beyond the intensive labor supply margin. Optimal taxation needs to take the fiscal migration externality into account, and our simulations indicate that this additional constraint to redistribution is quantitatively important. In our analysis we have abstracted from some aspects that are relevant in practice.

First, we restrict the central government to use a unified or regionally differentiated tax scheme, but it is not allowed to use targeted subsidies to 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 fixed migration subsidy typically cannot eliminate the problem completely, since the fiscal migration externality differs by earnings level. Even if the migration subsidy could be adjusted by earnings level, these need to be incentive compatible, implying an additional constraint for tax policy.

Secondly, regional productivity differences are partly reflected in local prices of non-tradable goods, rents and house prices, which also reduces migration incentives. While this is an important additional aspect, it does not challenge the economic intuition for the modification of the optimal tax schedule. The fiscal migration externality still exists in this case and should be taken into account accordingly.

Thirdly, we have also abstracted from redistributive taxation and welfare programs at the state level, which are an important component in some countries, including the US. Such state-level policies additionally affect migration incentives. Depending on whether these policies increase or decrease the fiscal migration externality, they may strengthen or weaken the constraint of productivity-enhancing inter-regional migration for federal redistribution.

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

Our results have immediate implications for redistributive tax policy. Policy makers should not only worry about external migration, but also need to consider the potential role of internal migration for a progressive tax-transfer-schedule. This constraint is less important for countries that are characterized by low regional inequality, which is typically the case for small countries. However, in countries where regional inequality is substantial, policy makers should carefully assess how tax progressivity may hurt productivity-enhancing inter-regional migration and design their tax-transfer scheme accordingly.

8 Appendix A: Formal derivation of the optimal tax formulae

We now show that the optimal tax formulae (7) and (9) and (10) can also be rigorously derived by standard optimal control techniques. The equivalence of the expression in terms of n and z is shown in Appendix D. We start with the differentiated case, since the unified case can be interpreted as the same problem with the additional constraint of the tax schedules to be identical in both regions.

8.1 Regionally differentiated taxation

The government maximizes

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

where $\bar{q}_A = \max\left\{V_B\left(\omega(n)\right) - V_A\left(n\right), 0\right\}$, $\bar{q}_B = \max\left\{V_A\left(n\right) - V_B\left(\omega(n)\right), 0\right\}$, $q = q^c + q^h$, and either $q^c = 0$ or $q^h = 0$. The first term in this expression stands for the social welfare from the population of region B who did not move, the second term stands for that of the population 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 term is equal to zero, because migration in both direction at the same ability level is not possible.

The maximization is subject to

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

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 $H(z_A, z_B, V_A, V_B, \lambda, \mu_A, \mu_B, n)$. The 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 the similar way 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. Indeed, the FOCs can be rewritten as

$$\begin{aligned} \frac{\mu_A(n)}{n}\frac{z_A}{n}h''\left(\frac{z_A}{n}\right) + \lambda\left(1 - h'\left(\frac{z_A}{n}\right)\right)\left(1 + P(\bar{q}_B|n) - P(\bar{q}_A|n)\right) f(n) = 0, \\ \varphi\left(\frac{z_A}{n}\right) = -\frac{\mu_A(n)}{\lambda n f(n)\left(1 + P(\bar{q}_B|n) - P(\bar{q}_A|n)\right)} \end{aligned}$$

for region A and

$$\frac{\mu_B(n)}{\omega(n)}\frac{z_B}{\omega(n)}h''\left(\frac{z_B}{\omega(n)}\right)+\lambda\left(1-h'\left(\frac{z_B}{\omega(n)}\right)\right)\left(1+P(\bar{q}_A|n)-P(\bar{q}_B|n)\right)f(n)=0$$
$$\varphi\left(\frac{z_B}{\omega(n)}\right)=-\frac{\mu_B(n)}{\lambda\omega(n)f(n)\left(1+P(\bar{q}_A|n)-P(\bar{q}_B|n)\right)}$$

for region B. In both cases, LHS is decreasing in z_A/n ($z_B/\omega(n)$) whereas RHS is constant, which implies that $z_i(n)$ is a unique solution and a global maximum indeed. Continuity can be then established in a way similar to Kleven et al (2009).

The conditions for $\dot{\mu}_i(n)$ imply

$$\begin{aligned}
-\dot{\mu}_A(n) &= f(n)[\int_{\bar{q}_A}^{+\infty}\Psi'\left(V_A(n)+q^h\right)p(q|n)dq+\int_0^{\bar{q}_B}\Psi'\left(V_A(n)-q^c\right)p(q|n)dq \\
&\quad -\lambda\left(1+P(\bar{q}_B|n)-P(\bar{q}_A|n)\right) \\
&\quad +\lambda\left(T_A-T_B\right)\left(p(\bar{q}_B|n)+p(\bar{q}_A|n)\right)],
\end{aligned}$$

and

$$\begin{aligned}
-\dot{\mu}_B(n) &= f(n)[\int_{\bar{q}_B}^{+\infty}\Psi'\left(V_B(\omega(n))+q^h\right)p(q|n)dq+\int_0^{\bar{q}_A}\Psi'\left(V_B(\omega(n))-q^c\right)p(q|n)dq \\
&\quad -\lambda\left(1+P(\bar{q}_A|n)-P(\bar{q}_B|n)\right) \\
&\quad +\lambda\left(T_B-T_A\right)\left(p(\bar{q}_B|n)+p(\bar{q}_A|n)\right)],
\end{aligned}$$

Integrating this, we have

$$\begin{aligned}
-\frac{\mu_A(n)}{\lambda} &= \int_n^{n_{\max}}[-\frac{1}{\lambda}\left(\int_{\bar{q}_A}^{+\infty}\Psi'\left(V_A(n)+q^h\right)p(q|n')dq+\int_0^{\bar{q}_B}\Psi'\left(V_B\left(\omega(n')\right)-q^c\right)p(q|n')dq\right) \\
&\quad +1+P(\bar{q}_B|n')-P(\bar{q}_A|n') \\
&\quad -\left(T_A-T_B\right)\left(p(\bar{q}_B|n')+p(\bar{q}_A|n')\right)]f(n')dn',
\end{aligned}$$

Analogously, for region B we get

$$\begin{aligned}
-\frac{\mu_B(n)}{\lambda} &= \int_n^{n_{\max}}[-\frac{1}{\lambda}\left(\int_{\bar{q}_B}^{+\infty}\Psi'\left(V_B\left(\omega(n')\right)+q^h\right)p(q|n')dq+\int_0^{\bar{q}_A}\Psi'\left(V_B\left(\omega(n')\right)-q^c\right)p(q|n')dq\right) \\
&\quad +\left(1+P(\bar{q}_A|n')-P(\bar{q}_B|n')\right) \\
&\quad -\left(T_B-T_A\right)\left(p(\bar{q}_B|n')+p(\bar{q}_A|n')\right)]dn'.
\end{aligned}$$

Defining by $g_A(n)$ the average marginal social welfare weight of the region A residents with inborn ability n, by $g_B(n)$ the average marginal social welfare weight of the region B initial residents with inborn ability n, we have

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

Using these, we can rewrite the optimality conditions as

$$\begin{aligned} -\frac{\mu_A(n)}{\lambda} &= \int_n^{n_{\max}} [\left(1-g_A(n')\right)\left(1+P(\bar{q}_B|n')-P(\bar{q}_A|n')\right) \\ &\quad -\left(T_A-T_B\right)\left(p(\bar{q}_B|n')+p(\bar{q}_A|n')\right)]f(n')dn', \end{aligned}$$

$$\begin{aligned} -\frac{\mu_B(n)}{\lambda} &= \int_n^{n_{\max}} [\left(1-g_B(n')\right)\left(1+P(\bar{q}_A|n')-P(\bar{q}_B|n')\right) \\ &\quad -\left(T_B-T_A\right)\left(p(\bar{q}_B|n')+p(\bar{q}_A|n')\right)]f(n')dn'. \end{aligned}$$

Inserting into the FOCs, we get

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

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

for the marginal rates in region A and in region B, respectively. The formulae are similar to Kleven et al (2009) except that two terms (rather than one) reflect the possibility of either immigration to or emigration from the given region. Note that for each n, there are two mutually exclusive scenarios: either there is migration from A to B (and $V_B\left(\omega(n)\right) > V_A\left(n\right)$) so that the formulae take the form

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

and

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

or there is migration from B to A (and $V_B\left(\omega(n)\right) < V_A\left(n\right)$) so that the formulae turn to

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

and

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

The average marginal tax rate is then

$$f(n)[n\frac{\tau_A}{1-\tau_A}\varepsilon_A\left(1+P(\bar{q}_B|n)-P(\bar{q}_A|n)\right)$$
$$+\omega(n)\frac{\tau_B}{1-\tau_B}\varepsilon_B\left(1+P(\bar{q}_A|n)-P(\bar{q}_B|n)\right)]=\int_n^{n_{\max}}\left(2-\bar{g}\left(n'\right)\right)f(n')dn',$$

where

$$\bar{g}\left(n\right):=g_A(n)\left(1+P(\bar{q}_B|n)-P(\bar{q}_A|n)\right)+g_B(n)\left(1+P(\bar{q}_A|n)-P(\bar{q}_B|n)\right)$$

is the average social marginal welfare weight of the individuals with innate ability n and we have 2 instead of 1 simply because our total population in two regions is of measure 2. Clearly, $\tau_A\left(n_{\max}\right)=0=\tau_B\left(\omega\left(n_{\max}\right)\right)$ and $\tau_A\left(n_{\min}\right)=0=\tau_B\left(\omega\left(n_{\min}\right)\right)$ from the transversality conditions.

Define migration semi-elasticities $\mu_i^+\left(n\right):=\frac{1}{1+P(\bar{q}_i|n)}\frac{\partial P(\bar{q}_i|n)}{\partial\bar{q}}=\frac{p(\bar{q}_i|n)}{1+P(\bar{q}_i|n)}$ for the region with inflow of population and $\mu_i^-\left(n\right):=\frac{1}{1-P(\bar{q}_i|n)}\frac{\partial P(\bar{q}_i|n)}{\partial\bar{q}}=\frac{p(\bar{q}_i|n)}{1-P(\bar{q}_i|n)}$ for the region with outflow of population. Define migration elasticity as $\nu_i:=\mu_i\left(T_A-T_B\right)$, whereby normalizing in terms of tax differential rather than utility differential V_B-V_A is for notational conviniency. We have

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

and

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

The effect of the migration elasticity as a top-up to the marginal social welfare weight is evident from the resulting formulae. Indeed, the marginal tax rate in region A (source region) is reduced by the migration elasticity in the same way it is reduced by the welfare weight of region A citizens. Conversely, the marginal tax rate in region B (recepient region) is increased by migration elasticity in the same way it is reduced by the welfare weight of region B citizens. Intuitively, marginal increase of tax for all skill levels above n in region A will lead to outflow of people resulting in the loss of revenue differential T_A-T_B between two regions, properly accounted for by the term $\nu_A^-\left(n'\right)$ at each skill level n'. In region B, the same mechanism is in action, only the loss of revenue differential is properly accounted for by the term $-\nu_A^+\left(n'\right)$. From the formulae above we can also see that more elastic migration response leads to higher marginal tax rates in region A and lower marginal tax rates in region B (migration elasticity is negative whenever $T_A<T_B$).

8.2 Non-differentiated tax schedule

In this case the tax schedules in two regions must be identical, and hence the indirect utilities also are (there are no differences in preferences). The government problem is to maximize

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

where $\bar{q}=V\left(\omega(n)\right)-V\left(n\right)$, and either q^h or q^c is equal to zero. We have also dropped the subscript B from the omega function for more parsimonious notation. The maximization is subject to

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

where the superscript w stands for the individuals with productivity $\omega(n)$. 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 established in the way similar 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 be then established in a way similar to Kleven et al (2009). Further,

$$\begin{aligned}-\dot{\mu}(n) &= f(\omega^{-1}(n))\int_{0}^{+\infty}\Psi'\left(V\left(n\right)+q^{h}\right)p(q|\omega^{-1}(n))dq\\&+f\left(\omega^{-1}(n)\right)\int_{0}^{\bar{q}_{1}}\Psi'\left(V\left(n\right)-q^{c}\right)p(q|\omega^{-1}(n))dq\\&+f(n)\int_{\bar{q}}^{+\infty}\Psi'\left(V\left(n\right)+q^{h}\right)p(q|n)dq\\&+\lambda[-\left(1-P(\bar{q}|n)\right)f(n)-\left(1+P(\bar{q}_{1}|\omega^{-1}(n))\right)f(\omega^{-1}(n))\\&-\left(z^{w}-n^{w}h\left(\frac{z^{w}}{n^{w}}\right)-V^{w}-\left(z-nh\left(\frac{z}{n}\right)-V\right)\right)p(\bar{q}|n)f(n)\\&+\left(z-nh\left(\frac{z}{n}\right)-V-\left(z^{-w}-n^{-w}h\left(\frac{z^{-w}}{n^{-w}}\right)-V^{-w}\right)\right)p(\bar{q}_{1}|\omega^{-1}(n))f(\omega^{-1}(n))],\end{aligned}$$

where superscript $-w$ stands for the argument $\omega^{-1}(n)$ of corresponding functions. Also, $\bar{q}_{1}=$

$V(n) - V(\omega^{-1}(n))$ and $\bar{q} = V(\omega(n)) - V(n)$. Rewriting in terms of taxes, we have

$$\begin{aligned}
-\dot{\mu}(n) &= f(\omega^{-1}(n)) \left(\int_0^{+\infty} \Psi' \left(V(n) + q^h \right) p(q|\omega^{-1}(n)) dq + \int_0^{\bar{q}_1} \Psi' \left(V(n) - q^c \right) p(q|\omega^{-1}(n)) dq \right) \\
&+ f(n) \int_{\bar{q}}^{+\infty} \Psi' \left(V(n) + q^h \right) p(q|n) dq \\
&+ \lambda [- (1 - P(\bar{q}|n)) f(n) - \left(1 + P(\bar{q}_1|\omega^{-1}(n)) \right) f(\omega^{-1}(n)) \\
&- (T(\omega(n)) - T(n)) p(\bar{q}|n) f(n) \\
&+ \left(T(n) - T(\omega^{-1}(n)) \right) p(\bar{q}_1|\omega^{-1}(n)) f(\omega^{-1}(n))]
\end{aligned}$$

Defining by $g_i(n)$ the average marginal social welfare weight of the region i residents with inborn ability n, we have

$$\begin{aligned}
g_A(n) &= \frac{1}{\lambda} \frac{\int_{\bar{q}}^{+\infty} \Psi' \left(V(n) + q^h \right) p(q|n) dq}{1 - P(\bar{q}|n)}, \\
g_B\left(\omega^{-1}(n) \right) &= \frac{\frac{1}{\lambda} \left(\int_0^{\bar{q}_1} \Psi' \left(V(n) - q^c \right) p(q|\omega^{-1}(n)) dq + \int_0^{+\infty} \Psi' \left(V(n) + q^h \right) p(q|\omega^{-1}(n)) dq \right)}{1 + P(\bar{q}_1|\omega^{-1}(n))}.
\end{aligned}$$

Thus, we can write

$$\begin{aligned}
-\frac{\dot{\mu}(n)}{\lambda} &= \left(g_B\left(\omega^{-1}(n) \right) - 1 \right) \left(1 + P(\bar{q}_1|\omega^{-1}(n)) \right) f\left(\omega^{-1}(n) \right) \\
&+ (g_A(n) - 1)(1 - P(\bar{q}|n)) f(n) \\
&- (T(\omega(n)) - T(n)) p(\bar{q}|n) f(n) \\
&+ \left(T(n) - T(\omega^{-1}(n)) \right) p(\bar{q}_1|\omega^{-1}(n)) f(\omega^{-1}(n))
\end{aligned}$$

and integrating

$$\begin{aligned}
-\frac{\mu(n)}{\lambda} &= \int_n^{n_{\max}} [\left(1 - g_B\left(\omega^{-1}(n') \right) \right) \left(1 + P(\bar{q}_1|\omega^{-1}(n')) \right) f\left(\omega^{-1}(n') \right) \\
&+ \left(1 - g_A(n') \right) \left(1 - P(\bar{q}|n') \right) f(n') \\
&+ \left(T\left(\omega(n') \right) - T(n') \right) p(\bar{q}|n') f(n') \\
&- \left(T\left(n' \right) - T(\omega^{-1}(n')) \right) p(\bar{q}_1|\omega^{-1}(n')) f(\omega^{-1}(n'))] dn'
\end{aligned}$$

and substituting into the FOC (using the definition of elasticity $\varepsilon = nh'/zh''$),

$$\begin{gathered}
\frac{1}{n\varepsilon} \frac{1}{(1 - P(\bar{q}|n)) f(n) + (1 + P(\bar{q}_1|\omega^{-1}(n))) f\left(\omega^{-1}(n) \right)} \times \\
\int_n^{n_{\max}} [\left(1 - g_B\left(\omega^{-1}(n') \right) \right) \left(1 + P(\bar{q}_1|\omega^{-1}(n')) \right) f\left(\omega^{-1}(n') \right) + \left(1 - g_A(n') \right) \left(1 - P(\bar{q}|n') \right) f(n') \\
+ \left(T\left(\omega(n') \right) - T(n') \right) p(\bar{q}|n') f(n') - \left(T\left(n' \right) - T(\omega^{-1}(n')) \right) p(\bar{q}_1|\omega^{-1}(n')) f(\omega^{-1}(n'))] dn' = \frac{\tau}{1 - \tau}.
\end{gathered}$$

Simplifying the integral expression, we get[15]

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

Defining by $\bar{g}\,(n)$ the average marginal social welfare weight of the people with observed productivity n as

$$\bar{g}\,(n) := g_A(n)\left(1-P(\bar{q}|n)\right) + g_B(n)\left(1+P(\bar{q}|n)\right),$$

we can rewrite the optimal tax formula as

$$\int_{n}^{n_{\max}} \left(2-\bar{g}(n')\right) f(n')dn' + \int_{\omega^{-1}(n)}^{n} \left(\left(1-g_B(n')\right)\left(1+P(\bar{q}|n')\right) - \left(T\left(\omega(n')\right)-T(n')\right) p(\bar{q}|n')\right) f\left(n'\right) dn' \tag{12}$$
$$= \frac{\tau}{1-\tau} n\varepsilon\left(\left(1-P(\bar{q}|n)\right) f(n) + \left(1+P(\bar{q}_1|\omega^{-1}(n))\right) f\left(\omega^{-1}(n)\right)\right)$$

which is analogous to the celebrated Mirrlees formula apart from the integral from $\omega^{-1}(n)$ to n that takes care of the revenue effect ($(1-g_B(n))\,(1+P(\bar{q}|n))$ term) and migration effect ($(T(\omega(n'))-T\,(n'))\,p(\bar{q}|n')$). Clearly, when $n \in [n_{\min}, \omega(n_{\min})]$, only non-migrated region A inhabitants have this productivity, so the formula becomes

$$\int_{n}^{n_{\max}} [1-g_A\left(n'\right)] f(n')dn' = \varepsilon\frac{\tau}{1-\tau} n f(n),$$

which is exactly the Mirrleesean formula. Note that the additional terms admit straightforward interpretation: $\int_{\omega^{-1}(n)}^{n} \left(\left(T\left(\omega(n')\right) - T(n')\right) p(\bar{q}|n')\right) f\left(n'\right) dn'$ is the tax paid by all migrants with skill from $\omega^{-1}(n)$ to n over and above the tax they would have paid if remaining in their home region. This characterizes a distortion that the government creates on extensive margin, stimulating $(T\left(\omega(n)\right) < T(n))$ or discouraging $(T\left(\omega(n)\right) < T(n))$ migration. The need for distortion comes from differences in social marginal welfare weights; its magnitude is determined, among other things, by the shape of the transformation function $\omega\,(n)$.

The other additional term, $\int_{\omega^{-1}(n)}^{n} \left(1-g_B(n')\right)\left(1+P(\bar{q}|n')\right) f\left(n'\right) dn'$, stands for the welfare effect of marginally increasing the tax for all productivity levels between n and $\omega\,(n)$ who migrate from region B to region A becasue of this increase (and thus realize productivity from $\omega^{-1}(n)$ to n). Using the elasticity defined as $\nu(n) := \left(T\left(\omega(n)\right) - T(n)\right) \frac{p(\bar{q}|n)}{1+P(\bar{q}|n)}$, we can rewrite the optimal tax formula as

$$\int_{\omega^{-1}(n)}^{n_{\max}} \left(1-g_B\left(n'\right)\right)\left(1+P(\bar{q}|n)\right) f\left(n'\right) dn' + \int_{n}^{n_{\max}} \left(1-g_A(n')\right)\left(1-P(\bar{q}|n')\right) f(n')dn'$$
$$-\int_{\omega^{-1}(n)}^{n} \nu(n')\left(1+P(\bar{q}|n')\right) f\left(n'\right) dn' = \frac{\tau}{1-\tau}\times$$
$$\times n\varepsilon\left(\left(1-P(\bar{q}|n)\right) f(n) + \left(1+P(\bar{q}_1|\omega^{-1}(n))\right) f\left(\omega^{-1}(n)\right)\right).$$

[15] Note that we have assumed $\omega(n_{\max}) = n_{\max}$, that is why $\omega^{-1}(n_{\max}) = n_{\max}$ and we get the expressions below. In a more general setting, we would have upper limit of the first integral term equal to $\omega^{-1}(n_{\max})$, and we would also have an additional integral term $\int_{\omega^{-1}(n_{\max})}^{n_{\max}} \left(T\left(\omega(n')\right) - T(n')\right) p(\bar{q}|n') f\left(n'\right) dn'$.

We see that more elastic migration puts downward pressure on marginal tax rates whenever the migration elasticity is positive on $[\omega^{-1}(n), n]$.

9 Appendix AA: Further propositions and proofs

9.1 Two propositions using the alternative benchmark of the ex ante productivity distribution

Proposition 8 *In a country with regional productivity differences and internal migration, the optimal non-differentiated marginal tax rates may be higher or lower relative to a benchmark without internal migration and the ex ante productivity distribution. Assuming exogenous marginal welfare weights, they are lower if*

$$\begin{aligned}\int_{\omega^{-1}(n)}^{n} (1-g_B(n'))\, P(\bar{q}|n')f(n')\,dn' - \int_{\omega^{-1}(n)}^{n} (T(\omega(n')) - T(n'))\, p(\bar{q}|n')f(n')\,dn' \quad (13)\\ < \frac{P(\bar{q}_1|\omega^{-1}(n))f(\omega^{-1}(n)) - P(\bar{q}|n)f(n)}{f(n)+f(\omega^{-1}(n))}\int_{n}^{n_{\max}} [1-g_A(n')]f(n')dn' \\ + \int_{n}^{n_{\max}} [g_B(n') - g_A(n')]\, P(\bar{q}|n')f(n')dn'\end{aligned}$$

Proof. *The optimal tax formula in case of unified taxation is presented by (12). For a government that does not take into account the possibility of migration, optimal marginal tax rates are implicitly defined by*

$$\begin{aligned}&\int_{n}^{n_{\max}} [2-g_A(n')-g_B(n')]f(n')dn' + \int_{\omega^{-1}(n)}^{n} (1-g_B(n'))\, f(n')\,dn' \\ = \ & \varepsilon\frac{\tau}{1-\tau}\left(f(n)+f\left(\omega^{-1}(n)\right)\right).\end{aligned}$$

Comparing the two expressions, we arrive at the condition (13). ■

Intuitively, there are three channels through which migration affects the magnitude of the marginal tax. The right hand side of (13) reflects how migration affects the revenue effect of a marginal change in the tax schedule. The first term on the right hand side is the difference of the revenue effects for migrants of productivity n and the migrants of productivity $\omega^{-1}(n)$, normalized by the total mass of people with such productivity. The second term takes care of the difference in social marginal welfare weights that all migrants of productivity n and above get upon migration from A to B. Loosely speaking, the possibility of migration enhances the revenue effect for region B and weakens it for region A simply because the migration flow is from A to B. The total change in the revenue depends on the difference in the migration flows at the initial and the new productivity levels, $P(\bar{q}_1|\omega^{-1}(n))f(\omega^{-1}(n))$ and $P(\bar{q}|n)f(n)$, as well as on the difference in the social weights g_B and g_A. Each difference contributes to lowering the marginal tax in case of migration.

Another channel also works through altering the revenue effect, but only for the migrants between productivity levels $\omega^{-1}(n)$ and n. This effect is positive, it is represented by the first term on the left hand side of (13). Finally, the third channel is through the migration effect as the difference between the new and the old tax on migrants between productivity $\omega^{-1}(n)$ and n. This effect is negative as long as the tax schedule on the appropriate productivity interval is increasing.

Since, for this alternative benchmark using the ex ante distribution there is no unambiguous answer as to whether migration decreases or increases optimal marginal tax rates, we may ask whether it does so marginally. To answer this question we study what happens if, starting from two identical regions, we introduce a marginal productivity differencem. Formally, assume $\omega(n) = n + \Delta$, where Δ is infinitely small. It turns out that even in this case the effect on the optimal marginal tax is theoretically ambiguous and, as the following proposition tells us, in general depends on the shape of the ability distribution and the migration costs distribution.

Proposition 9 *Starting from two identical regions, introducing a marginal difference in productivity distribution lowers optimal marginal tax if and only if*

$$\frac{1-g_B(n)}{2} + \frac{\int_n^{n_{\max}} [2 - g_A(n') - g_B(n')] f(n')dn'}{4(f(n))^2} \left(\frac{P(\bar{q}|n)}{\partial n} f(n) + f'(n) \right) < 0. \tag{14}$$

Proof. *We express the marginal tax rate $\frac{\tau}{1-\tau}$ from the optimal tax formula (12) under the assumption that $\omega(n) = n + \Delta$, take a derivative of it with respect to Δ, and evaluate it at $\Delta = 0$, keeping in mind that there is no migration at this point. The resulting expression is proportional to the left hand side of (14). Correspondingly, the marginal tax rate decreases with the introduction of marginal productivity differences if this expression is negative and it increases in case it is positive.* ■

We can see that the terms in (14) related to the revenue effect are always positive. Thus, a sufficient condition for increase in marginal tax is that $\frac{P(\bar{q}|n)}{\partial n} f(n) + f'(n) \geq 0$. This is satisfied for independent distribution of costs ($\frac{P(\bar{q}|n)}{\partial n} = 0$) and a uniform distribution of ability. On the other hand, if the distribution of ability is sufficiently "decreasing", like the Pareto distribution, for example, then introducing marginal productivity differences puts downward pressure on marginal taxes.

9.2 Proof of Proposition 5

Under the assumptions formulated in the text, $V_A(n)$ 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, we have

$$g_A(n) = \frac{\int_{\bar{q}_A}^{+\infty} \Psi'\left(V_A(n) + q^h\right) p(q|n')dq + \int_0^{\bar{q}_B} \Psi'\left(V_A(n) - q^c\right) p(q|n)dq}{\lambda\left(1 + P(\bar{q}_B|n) - P(\bar{q}_A|n)\right)}, \tag{15a}$$

$$g_B(n) = \frac{\int_{\bar{q}_B}^{+\infty} \Psi'\left(V_B(\omega(n')) + q^h\right) p(q|n')dq + \int_0^{\bar{q}_A} \Psi'\left(V_B(\omega(n)) - q^c\right) p(q|n)dq}{\lambda\left(1 + P(\bar{q}_A|n) - P(\bar{q}_B|n)\right)}, \tag{15b}$$

which converge to

$$g_A^\infty = g_B^\infty = \frac{\bar{\psi}}{\lambda}. \tag{16a}$$

If $T_B - T_A$ converges, it must be that $\tau_A^\infty = \tau_B^\infty = \tau^\infty$. But since

$$h'\left(\frac{z_i}{n_i}\right) = 1 - \tau_i(z_i), \tag{17}$$

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

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

for the marginal rates in region A and

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

for the region B. The right hand sides are equal, so we need $\Delta T^\infty=0$ for the left hand sides to be equal as well.

9.3 Proof of Proposition 6

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

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

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

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

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,

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

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

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

9.4 Proof of Proposition 7

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

$$\bar{q}_A = V_B - V_A = (\omega(n) - n)\left(\frac{z}{n} - h\left(\frac{z}{n}\right)\right) - (T_B - T_A),$$

so we can write

$$\dot{q}_A = \left(\omega'(n) - 1\right)\left(\frac{z}{n} - h\left(\frac{z}{n}\right)\right) - \left(T'_B - T'_A\right).$$

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

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

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 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:

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

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 separable taxation, $\tau_A = \tau_B$, 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 initial difference $T_A - T_B$ is constant, the gain in revenue from migrants above n will be compensated by the loss in revenue 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, we need that our reasoning holds for $\omega'(n) > 1$, i.e. $\dot{g}_A - \dot{g}_B < 0$ also for this case. Differentiating $g_A - g_B$ in this case, we have

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

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

10 Appendix B: Implementability

We follow the supplementary material to Kleven et al. (2009). The same reasoning applies.

In particular, an action profile $(z_A(n), z_B(\omega(n)))_{n\in(n_{\min},n_{\max})}$ is implementable if and only if there exist transfer functions $(c_A(n), c_B(\omega(n)))_{n\in(n_{\min},n_{\max})}$ such that $(z_i(n_i), c_i(n_i))_{i\in\{A,B\},n\in(n_{\min},n_{\max})}$ is a truthful mechanism. A mechanism is called truthful if there is a $\{\bar{q}_A(n), \bar{q}_B(n)\}$ such that (i) for $q < \max_{i\in\{A,B\}} \bar{q}_i(n)$, the set $(i' = \arg\min_{i\in\{A,B\}} \bar{q}_i(n), n' = n)$ maximizes $u(z_{i'}(n'), i', c_{i'}(n'), (n,q))$ over all (i', n'); (ii) for $q \geq \max_{i\in\{A,B\}} \bar{q}_i(n)$, the set $(i' = \arg\max_{i\in\{A,B\}} \bar{q}_i(n), n' = n)$ maximizes $u(z_{i'}(n'), i', c_{i'}(n'), (n,q))$ over all (i', n'). The usual implementability theorem for the one-dimensional case also applies in our setting.

Lemma 10 (Kleven-Kreiner-Saez) *An action profile $(z_A(n), z_B(\omega(n)))_{n\in(n_{\min},n_{\max})}$ is implementable if and only if $z_A(n)$ and $z_B(\omega(n))$ are both nondecreasing in n.*

Proof. *The utility function $c - nh(z/n)$ (and also $c - \omega(n)h(z/\omega(n))$) satisfies the usual single crossing condition. Indeed, it is equivalent to $xh''(x) > 0$ for all $x > 0$, which is ensured by convexity of h. That is why from the one-dimensional case we know that $z(n)$ (and $z(\omega(n))$) is implementable, which means that there is some $c(n)$ such that $c(n) - nh(z(n)/n) \geq c(n') - nh(z(n')/n)$ for all n, n' if and only if $z(n)$ is nondecreasing.*[16]

Suppose $(z_A(n), z_B(\omega(n)))$ is implementable, so that there exists $(c_A(n), c_B(\omega(n)))$ such that $(z_i(n_i), c_i(n_i))_{i\in\{A,B\},n\in(n_{\min},n_{\max})}$ is a truthful mechanism. This implies that $c_A(n) - nh(z_A(n)/n) \geq c_A(n') - nh(z_A(n')/n)$ and $c_B(\omega(n)) - \omega(n)h(z_B(\omega(n))/\omega(n)) \geq c_B(\omega(n')) - \omega(n)h(z_B(\omega(n'))/\omega(n))$ for all n, n', and so by one-dimensional result $z_A(n)$ and $z_B(\omega(n))$ are nondecreasing.

Conversely, suppose $z_A(n)$ and $z_B(\omega(n))$ are nondecreasing. One-dimentional result implies then that there exist such $c_A(n)$ and $c_B(\omega(n))$ that $c_A(n) - nh(z_A(n)/n) \geq c_A(n') - nh(z_A(n')/n)$ and $c_B(\omega(n)) - \omega(n)h(z_B(\omega(n))/\omega(n)) \geq c_B(\omega(n')) - \omega(n)h(z_B(\omega(n'))/\omega(n))$ correspondingly.

We have to show that the mechanism $(z_i(n_i), c_i(n_i))_{i\in\{A,B\},n\in(n_{\min},n_{\max})}$ is actually truthful . We have

$$\begin{aligned}\bar{q}_A &= \max\{V_B(\omega(n)) - V_A(n), 0\},\\ \bar{q}_B &= \max\{V_A(n) - V_B(\omega(n)), 0\},\end{aligned}$$

where we define $V_B(\omega(n)) := c_B(\omega(n)) - \omega(n)h(z_B(\omega(n))/\omega(n))$ and $V_A(n) := c_A(n) - nh(z_A(n)/n)$.

In case $\bar{q}_A > 0$, for all $n, n', q \geq \bar{q}_A(n)$ we have

$$u_A(z_A(n), c_A(n), 0, (n,q)) = V_A(n) \geq V_B(\omega(n)) - q \geq u_B(z_B(n'), c_B(n'), 1, (n,q));$$

for all $n, n', q \leq \bar{q}_A(n)$ we have

$$u_B(z_B(\omega(n)), c_B(\omega(n)), 1, (n,q)) = V_B(\omega(n)) - q \geq V_A(n) \geq u_A(z_A(n'), c_A(n'), 0, (n,q)).$$

In case $\bar{q}_B > 0$, for all $n, n', q \geq \bar{q}_B(n)$ we have

$$u_B(z_B(\omega(n)), c_B(\omega(n)), 0, (n,q)) = V_B(\omega(n)) \geq V_A(n) - q \geq u_A(z_A(n'), c_A(n'), 1, (n,q));$$

[16] Clearly, the same is true if $\omega(n)$ rather than n is the appropriate argument.

for all $n, n', q \leq \bar{q}_B(n)$ *we have*

$$u_A(z_A(n), c_A(n), 1, (n,q)) = V_A(n) - q \geq V_B(\omega(n)) \geq u_B\left(z_B\left(\omega\left(n'\right)\right), c_B\left(\omega\left(n'\right)\right), 0, (n,q)\right).$$

As in Kleven et al (2009), it is the separability of q *in the utility specification that allows us to get these simple results.* ■

The proof for the uniform tax is analogous, with a restriction that $z_A(n) - c_A(n) \equiv z_B(n) - c_B(n)$.

11 Appendix C: Existence

We follow Kleven et al (2009) to establish similar conditions for our problem with differential taxation.

Formally, our maximization problem is the optimal control problem $\dot{V} = b(n, V, z)$ with maximization objective $B^0 = \int_{n_{\min}}^{n_{\max}} b^0(n, V(n))\,dn$ and constraint $\int_{n_{\min}}^{n_{\max}} b^1(n, z(n), V(n)) \geq 0$, where

$$\begin{aligned}
b(n,V,z) &= \left(-h\left(\frac{z_A}{n}\right) + \frac{z_A}{n}h'\left(\frac{z_A}{n}\right), -h\left(\frac{z_B}{\omega(n)}\right) + \frac{z_B}{\omega(n)}h'\left(\frac{z_B}{\omega(n)}\right)\right), \\
b^0(n,V) &= [\int_{\max\{V_A-V_B,0\}}^{+\infty} \Psi\left(V_B + q^h\right) p(q|n)dq + \int_0^{\max\{V_B-V_A,0\}} \Psi\left(V_B - q^c\right) p(q|n)dq \\
&\quad + \int_{\max\{V_B-V_A,0\}}^{+\infty} \Psi\left(V_A + q^h\right) p(q|n)dq \\
&\quad + \int_0^{\max\{V_A-V_B,0\}} \Psi\left(V_A - q^c\right) p(q|n)dq]f(n), \\
b^1(n, z(n), V(n)) &= [\left(z_B - \omega(n)h\left(\frac{z_B}{\omega(n)}\right) - V_B\right) \\
&\quad \times (1 + P(\max\{V_B - V_A, 0\}|n) - P(\max\{V_A - V_B, 0\}|n)) \\
&\quad + \left(z_A - nh\left(\frac{z_A}{n}\right) - V_A\right) \\
&\quad \times (1 + P(\max\{V_A - V_B, 0\}|n) - P(\max\{V_B - V_A, 0\}|n))]f(n) - E
\end{aligned}$$

The functions b, b^0, b^1 are continuous and continuously differentiable in (z, V) by construction. Analogously to Kleven et al (2009), if we assume that there is an a priori bound on the path of admissible z, we need to show that the sets $B(n, V, \lambda) = \{(y, b(n, V, z)|z_0 \geq 0, z_1 \geq 0, y \geq -b^0 - \lambda b^1\}$ are convex for all n, V and $\lambda \geq 0$, then there exists an optimal control z measurable on $(n_{\min}, n_{\max})$.

We have

$$\begin{aligned}
B(n,V,\lambda) &= \{(y, -h\left(\frac{z_A}{n}\right) + \frac{z_A}{n}h'\left(\frac{z_A}{n}\right), \\
-h\left(\frac{z_B}{\omega(n)}\right) + \frac{z_B}{\omega(n)}h'\left(\frac{z_B}{\omega(n)}\right))|z_A &\geq 0, z_B \geq 0, \\
y &\geq -b^0(n,V) \\
& -\lambda f(n)[\left(z_B - \omega(n)h\left(\frac{z_B}{\omega(n)}\right) - V_B\right)(1 + P_B - P_A) \\
& + \left(z_A - nh\left(\frac{z_A}{n}\right) - V_A\right)(1 + P_A - P_B)]
\end{aligned}$$

We denote by $K(.)$.the inverse of the strictly increasing function $x \to -h(x) + xh'(x)$ with $K(0) = 0$, so we can write

$$\begin{aligned}
B(n,V,\lambda) &= \{(y, x_A, x_B | x_A \geq 0, x_B \geq 0, \\
y + b^0(n,V) &\geq \lambda f(n)[\omega(n)\left(h\left(K(x_B)\right) - K(x_B) + V_B\right)(1 + P_B - P_A) \\
& +n\left(h\left(K(x_A)\right) - K(x_A) + V_A\right)(1 + P_A - P_B)].
\end{aligned}$$

Therefore, $B(n,V,\lambda)$ is convex when $x \to h\left(K(x)\right) - K(x) \equiv \phi(x)$ is convex, which it is by the same reasoning as in Kleven et al (2009). In particular, by definition of $K(.)$, $-h(K(x)) + K(x)h'(K(x)) \equiv x$, so that differentiation gives us $K(x)h''(K(x))K'(x) \equiv 1$. Consider $\phi'(x) = \left(h'\left(K(x)\right) - 1\right)K'(x)$. By our previous result, $K'(x) = 1/\left[K(x)h''(K(x))\right]$. Therefore, we can write $\phi'(x) = -\left(1 - h'\left(K(x)\right)\right)/\left[K(x)h''(K(x))\right]$. By assumption 1, $\phi'(x)$ is an increasing function of $K(x)$. Since $x \to K(x)$ is strictly increasing, $\phi'(x)$ is increasing and thus $\phi(x)$ is convex.

12 Appendix D: on the equivalence of representations via income and via ability

Here we show that the optimal tax formulae obtained in the text are equivalent to those in the appendix. Consider the formula for non-differentiated taxation in the text:

$$\begin{aligned}
\frac{\tau}{1-\tau} &= \frac{1}{z\varepsilon\left(v_A(z)s_A(z) + v_B(z)s_B(z)\right)} \times \\
& [\int_z^\infty \left\{\left[1 - g_A\left(z'\right)\right]v_A\left(z'\right)s_A + \left[1 - g_B\left(z'\right)\right]v_B\left(z'\right)s_B\right\}dz' \\
& + \int_{\tilde{z}}^z \left[T(z') - T(k(z'))\right]p\left(\bar{q}|\, z'\right)v_A(z')dz'],
\end{aligned}$$

where $s_A(z) \equiv 1 - P\left(\bar{q}|\, z(n)\right) = 1 - P\left(\bar{q}|\, n\right)$ and $s_B \equiv 1 + P\left(\bar{q}|\, z'\right) = 1 + P\left(\bar{q}_1|\, \omega^{-1}(n)\right)$ and $v_A(z\,(n)) = f(n)/\dot{z}(n)$, $v_B(z\,(n)) = f(\omega^{-1}(n))/\dot{z}(n)$. Further, $T(z\,(n)) = T(n)$, $T(k(z(n))) = T(\omega(n))$ and $\tilde{z} = k^{-1}(z(n)) = z\left(\omega^{-1}(n)\right)$; $g_i\left(z(n)\right) = g_i(n)$. Plugging into the expression above, we get

$$
\begin{aligned}
\frac{\tau}{1-\tau} \quad = \quad & \frac{1}{\frac{z(n)}{\dot{z}(n)}\varepsilon\left((1-P(\bar{q}|n))\, f(n) + (1+P(\bar{q}_1|\omega^{-1}(n)))\, f\left(\omega^{-1}(n)\right)\right)} \\
& [\int_n^{n_{\max}} \{\left[1-g_A\left(n'\right)\right] \frac{f(n')}{\dot{z}(n')}\left(1-P\left(\bar{q}|\, n'\right)\right) \\
& + \left[1-g_B\left(n'\right)\right] \frac{f(\omega^{-1}(n))}{\dot{z}(n')}\left(1+P\left(\bar{q}_1|\, \omega^{-1}(n)\right)\right)\}\dot{z}(n')dn' \\
& + \int_{\omega^{-1}(n)}^{n} \left[T(n') - T(\omega(n'))\right] p\left(\bar{q}|\, n'\right) \frac{f(n')}{\dot{z}(n')}\dot{z}(n')dn'],
\end{aligned}
$$

compared to

$$
\begin{gathered}
\frac{1}{n\varepsilon}\frac{1}{(1-P(\bar{q}|n))\, f(n) + (1+P(\bar{q}_1|\omega^{-1}(n)))\, f\left(\omega^{-1}(n)\right)}\times \\
\int_n^{n_{\max}} [\left(1-g_B\left(\omega^{-1}(n')\right)\right)\left(1+P(\bar{q}_1|\omega^{-1}(n'))\right) f\left(\omega^{-1}(n')\right) + \left(1-g_A(n')\right)\left(1-P(\bar{q}|n')\right) f(n')]dn' \\
- \int_{\omega^{-1}(n)}^{n} \left(T\left(\omega(n')\right) - T(n')\right) p(\bar{q}|n') f\left(n'\right) dn' = \frac{\tau}{1-\tau}.
\end{gathered}
$$

The expressions are identical, if $\frac{z(n)}{\dot{z}(n)} = n$. To prove that this is indeed the case for the tax schedule linearized around the optimum in our model, simply totally differentiate the first order condition (2):

$$h''\left(\frac{z}{n}\right)\frac{n\dot{z}\left(n\right) - z}{n^2} = -T''\left(z\right)\dot{z}\left(n\right).$$

For a linear approximation, $T''\left(z\right) = 0$, so we get $n\dot{z}\left(n\right) = z(n)$ indeed. This completes the proof of equivalence for the case of non-differentiated taxation. The derivation for differentiated taxation is analogous.

References

[1] Abdeljawad, Thabat; Jarad, Fahd and Dumitru Baleanu (2009): "Variational Optimal-Control Problems with Delayed Arguments on Time Scales," Advances in Difference Equations, SpringerOpen, vol. 2009.

[2] Acemoglu, Daron and Melissa Dell (2010): "Productivity Differences between and within Countries," American Economic Journal: Macroeconomics, vol. 2(1), pp. 169-188.

[3] Akerlof, George A. (1978): "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning," American Economic Review, vol. 68 (March), pp. 8–19.

[4] Albouy, David (2009): "The Unequal Geographic Burden of Federal Taxation," Journal of Political Economy, vol. 117(4), pp. 635-667.

[5] Armstrong, Mark (1996): "Multiproduct Nonlinear Pricing," Econometrica, vol. 64(1), pp. 51-75.

[6] Bargain, Olivier; Dolls, Mathias; Fuest, Clemens; Neumann, Dirk; Peichl, Andreas; Pestel, Nico and Sebastian Siegloch (2013): "Fiscal Union in Europe? Redistributive and Sta-

bilising Effects of a European Tax-Benefit System and Fiscal Equalisation Mechanism," Economic Policy, vol. 28(75), pp. 375-422.

[7] Barro, Robert J. and Xavier Sala-i-Martin (1991): "Convergences across States and Regions," Brooking Papers on Economic Activity, issue 1, pp. 107-182.

[8] Bayer, Christian and Falko Juessen (2012): "On the dynamics of interstate migration: Migration costs and self-selection, " Review of Economic Dynamics, vol. 15, pp. 337-401.

[9] BEA, (2014): "Regional Economic Accounts: Download," US Department of Commersce, Bureau of Economic Analysis, accessed online July 27, 2014 at http://www.bea.gov/regional/downloadzip.cfmU.

[10] Best, Michael and Henrik Jacobsen Kleven (2013): "Optimal Income Taxation with Career Effects of Work Effort," Working Paper, LSE, February 2013.

[11] Blanchard, Olivier and Lawrence Katz (1992): "Regional Evolutions," Brookings Papers on Economic Activity, pp. 1-75.

[12] Boadway, Robin and Pierre Pestieau (2006): "Tagging and redistributive taxation," Annales d'Economie et de Statistique, ENSAE, issue 83-84, pp. 123-147.

[13] Borjas, George (1994): "The economics of immigration," Journal of Economic Literature, vol. 32(4), pp. 1667-1717.

[14] Boustan, Leah Platt, Fishback, Price V. and Shawn Kantor: "The Effect of Internal Migration on Local Labor Markets: American Cities during the Great Depression," Journal of Labor Economics, vol. 28(4), pp. 719-746.

[15] Chetty, Raj (2006): "A New Method of Estimating Risk Aversion," The American Economic Review, vol. 96(5), pp. 1821-1834.

[16] Ciccone, Antonio and Robert E. Hall (1996): "Productivity and the Density of Economic Activity," American Economic Review, vol. 86(1), pp. 54-70.

[17] Cremer, Helmuth; Gahvari, Firouz and Jean-Marie Lozachmeur (2010): "Tagging and income taxation: theory and an application," American Economic Journal: Economic Policy, vol. 2(1), pp. 31-50.

[18] D'Amuri, Francesco; Ottaviano, Gianmarco I. P. and Giovanni Petri (2010): "The labor market impact of immigration in Western Germany in the 1990s," European Economic Review, vol. 54, pp. 550-570.

[19] Diamond, Peter A. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," American Economic Review, vol. 88(1), pp. 83-95.

[20] Eeckhout, Jan and Nezih Guner (2015): "Optimal Spatial Taxation," IZA Discussion Paper No. 8781, Institute for the Study of Labor.

[21] Frank, Douglas (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.

[22] Gomes, Renato; Jean-Marie Lozachmeur and Alessandro Pavan (2014): "Differential Taxation and Occupational Choice, CESifo Working Paper No. 5054, CESifo Group Munich.

[23] Gordon, Roger H. and Julie Berry Cullen (2012): "Income redistribution in a federal system of governments," Journal of Public Economics, vol. 96(11-12), pp. 1100-1109.

[24] Göllmann, L.; Kern, D. 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.

[25] Greenwood, Michael J. (1969): "An Analysis of the Determinants of Geographic Labor Mobility in the United States," Review of Economic and Statistics, vol. 51, pp. 189-194.

[26] Immonen, Ritva; Kanbur, Ravi; Keen, Michael and Matti Tuomala (1998): "Tagging and Taxing: The Optimal Use of Categorical and Income Information in Designing Tax/Transfer Schemes," Economica, vol. 65(258), pp. 179-92.

[27] ISTAT, (2014): "Noi Italia - 100 statistics to understand the country we live in: Macroeconomics: GDP per capita by region," Italian National Institute of Statistics, accessed online July 27, 2014 at http://noi-italia.istat.it/fileadmin/user_upload/allegati/S15I01M02p0_2013_en.xls-

[28] Jacquet, Laurence; Lehmann, Etienne and Bruno van der Linden (2013): "Optimal redistributive taxation with both extensive and intensive responses," Journal of Economic Theory, vol. 148(5), pp. 1770-1805.

[29] Kanbur, Ravi and Anthony Venables (2005): "Spatial Inequality and Development," Oxford University Press, USA.

[30] Kennan, John and James R. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," Econometrica, vol. 79(1), pp. 211-251.

[31] Kleven, Henrik Jacobsen; Kreiner, Claus Thustrup and Emmanuel Saez (2009): "The Optimal Income Taxation of Couples," Econometrica, vol. 77(2), pp. 537-560.

[32] Kleven, Henrik Jacobsen; Kreiner, Claus Thustrup and Emmanuel Saez (2006): "The Optimal Income Taxation of Couples," NBER Working Papers 12685, National Bureau of Economic Research.

[33] Lehmann, Etienne; Simula, Laurent and Alain Trannoy (2014): "Tax Me If You Can! Optimal Nonlinear Income Tax between Competing Governments," The Quarterly Journal of Economics, vol. 129(4), pp. 1995-2030.

[34] Lipatov, Vilen and Alfons Weichenrieder (2010): "Optimal income taxation with tax competition," CESifo Working Paper No. 3108, CESifo Group Munich.

[35] Mankiw, Gregory and Matthew Weinzierl (2010): "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," American Economic Journal: Economic Policy, vol. 2(1), pp. 155-176.

[36] Mirrlees, James A. (1982): "Migration and Optimal Income Taxation," Journal of Public Economics, vol. 18, pp. 319-341.

[37] Ottaviano, Gianmarco I. P. and Giovanni Peri (2008): "Immigration and National Wages: Clarifying the Theory and the Empirics," NBER Working Paper No. 14188, National Bureau of Economic Research.

[38] Ottaviano, Gianmarco I. P. and Giovanni 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.

[39] Picketty, Thomas (1997): "La redistribution fiscale face au chômage," Revue française d'économie, vol. 12(1), pp. 157-201.

[40] Rochet, Jean-Charles and Philippe Choné (1998): "Ironing, Sweeping, and Multi-Dimensional Screening," Econometrica, vol. 66(4), pp. 783-826.

[41] Saez, Emmanuel (2002): "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses," The Quarterly Journal of Economics, vol. 117(3), pp. 1039-1073.

[42] Saez, Emmanuel (2001): "Using Elasticities to Derive Optimal Income Tax Rates," Review of Economic Studies, vol. 68(1), pp. 205-229.

[43] Rothschild, Casey and Florian Scheuer (2014): "Optimal Taxation with Rent-Seeking," CESifo Working Paper No. 5071, CESifo Group Munich.

[44] Rothschild, Casey and Florian Scheuer (2013): "Redistributive Taxation in the Roy Model," The Quarterly Journal of Economics, vol. 128(2), pp. 623-668.

[45] Scheuer, Florian (2014): "Entrepreneurial Taxation with Endogenous Entry," American Economic Journal: Economic Policy, vol. 6, pp. 126-163.

[46] Simula, Laurent and Alain Trannoy (2010): "Optimal income tax under the threat of migration by top-income earners," Journal of Public Economics, vol. 94(1-2), pp. 163-173.

[47] Stantcheva, Stefanie (2015), "Optimal Taxation and Human Capital Policies over the Life Cycle," mimeo, MIT.

[48] Weinzierl, Matthew C. (2011): "The Surprising Power of Age-Dependent Taxes," Review of Economic Studies, vol. 78(4), pp. 1490–1518.

[49] Wildasin, David (1991): "Income Redistribution in a Common Labor Market," American Economic Review, vol. 81(4), pp. 757-774.

[50] Wilson, John D. (1992): "Optimal Income Taxation and International Personal Mobility," American Economic Review, Papers and Proceedings, vol. 82(2), pp. 191-196.

[51] Yagan, Danny (2014): "Moving to Opportunity? Migratory Insurance over the Great Recession," mimeo.

[52] Young, Alwyn (2013): "Inequality, the Urban-Rural Gap and Migration," The Quarterly Journal of Economics, vol. 128(4), pp. 1727-1785.