Nonlinear Taxation and International Mobility in General Equilibrium*

Eckhard Janeba† **Karl Schulz**‡

University of Mannheim University of Mannheim

March 13, 2022

Abstract

We study the nonlinear taxation of internationally mobile workers in general equilibrium. First, we show in a theoretical model with N ability types that, contrary to conventional wisdom, in general equilibrium, migration may lower the bottom tax rate but raises the top tax rate, moving marginal tax rates closer to those in an economy with fixed wages. Then, we turn to a standard quantitative framework of optimal tax progressivity, in which we discover a novel wage effect on migration in addition to a migration effect on wages. Both call for a higher rate of tax progressivity to amplify pre-tax wage inequality and blunt trickle-down forces. We show that the presence of general equilibrium effects can offset almost half of the migration-induced reduction in tax progressivity. Altogether, depending on the shape of the migration semi-elasticity, the common concern of a "migration threat" limiting governments' scope for redistribution is exaggerated.

Keywords: Optimal Taxation, General Equilibrium, Trickle-Down Effects, Migration, Tax/Subsidy Competition

JEL Classification: H21, H24, H73, F22, R13

*Eckhard Janeba gratefully acknowledges the support from the Collaborative Research Center (SFB) 884 "Political Economy of Reforms", funded by the German Research Foundation (DFG). Karl Schulz gratefully acknowledges the support by the University of Mannheim's Graduate School of Economic and Social Sciences, funded by the German Research Foundation (DFG). For helpful comments and discussions, we thank Hans Peter Grüner, Duk Gyoo Kim, Mathilde Muñoz, Nicolas Werquin and numerous seminar and conference participants.

†University of Mannheim, Department of Economics, L7 3-5, 68131 Mannheim, Germany, janeba@uni-mannheim.de.

‡University of Mannheim, Center for Doctoral Studies in Economics, L7 3-5, 68131 Mannheim, Germany, karl.schulz@gess.uni-mannheim.de.

1 Introduction

International (and inter-regional) mobility of high-income individuals has been at the center of recent theoretical and empirical research due to its far-reaching implications for the taxation of mobile individuals and the progressivity of the income tax code. Mirrlees (1971) has already recognized the importance of international migration but focuses on a closed economy case in his formal analysis.[1] The literature on optimal income taxation has demonstrated the effects of migration on the level and shape of optimal marginal income tax rates. In particular, migration tends to decrease optimal marginal tax rates over the entire income distribution (see Simula and Trannoy (2010)) and, depending on the shape of migration semi-elasticities, even lead to negative tax rates at the top (Lehmann, Simula, and Trannoy (2014)). The mobility of labor reduces tax payments by top-income workers relative to a closed economy. The driving force has been labeled the "threat of migration" and is even present in a situation in which no net migration occurs in equilibrium. A standard assumption in these models is, however, that workers' wages are exogenous.

The role of wage endogeneity for tax policy has been highlighted in another strand of the optimal income taxation literature. For instance, in the closed economy model of Stiglitz (1982), the government lowers the top tax rate to encourage the labor supply of high-skilled individuals, thereby raising wages at the bottom and reducing them at the top.[2] Contributors to the literature refer to this indirect redistribution of pre-tax wages as "predistribution." However, this literature has so far disregarded the presence of labor mobility.

This paper bridges the gap between these two strands of the literature by studying optimal income taxation and international mobility in general equilibrium, i.e., with endogenous wages. We shed light on the interaction of general equilibrium and migration responses in two steps. As a first step, we introduce labor mobility (as in Lehmann et al. (2014)) into the N-type version of the Stiglitz (1982) model. We show that the optimal general equilibrium marginal tax rate at the top is higher (i.e., a lower labor subsidy) in the open than in the closed economy. Moreover, we derive conditions under which the optimal marginal tax rate at the bottom declines. Thus, migration may lead to a more progressive tax code in terms of marginal tax rates.

[1] See Kleven, Landais, Munoz, and Stantcheva (2020) for a survey of the empirical literature.

[2] While Stiglitz (1982) makes this observation about general-equilibrium wage effects in a two-type framework, Sachs, Tsyvinski, and Werquin (2020) demonstrate in a continuous-type framework that lower tax rates in general equilibrium may also apply to middle-class workers (see Figure 4 in Sachs et al. (2020)).

The intuition for this finding is that the predistribution effect, which the government achieves by taxing incomes less progressively, is self-limiting under labor mobility: Any reduction in pre-tax wage inequality comes along with lower immigration (or higher emigration) of high-skilled workers, which, in turn, raises their wages, partly offsetting the benefits of predistribution.

In addition to the finding concerning the tax rates at the ends of the ability distribution, we also characterize the effect of mobility on optimal marginal tax rates in the "middle" and show that it relates to the sign of the optimal marginal tax rate in the closed economy. While the above insights on marginal tax rates complement well the existing literature, they are only to some extent informative about the shape of the income tax scheme (see Heathcote and Tsujiyama (2021b) for a discussion). The N-type model does not reveal profound results on average tax rates that would be necessary to characterize the overall effect on tax progressivity.

Therefore, we set up a continuous version of the model and calibrate it to the U.S. economy as a second step. We depart from the Mirrleesian approach of solving for the optimal arbitrarily nonlinear tax system by restricting attention to the class of nonlinear tax functions with a constant rate of progressivity (CRP). The CRP tax scheme approximates well the current U.S. code (see Heathcote, Storesletten, and Violante (2017)) and delivers similar policy prescriptions as the fully nonlinear optimum (e.g., Heathcote and Tsujiyama (2021a)). This modified approach allows us to study the interaction of migration and general equilibrium effects throughout the entire income distribution and to speak to the impact on average tax rates that are more critical for migration decisions than marginal rates.

As a first step, we analyze the incidence of changing tax progressivity. We discover two novel effects revealing the underlying mechanisms that limit predistribution and weaken the migration threat. Firstly, there is a *wage effect on migration*: Raising tax progressivity, a government can amplify wage inequality (predistribute less to low-skilled workers). This triggers high-skilled immigration and expands tax revenues.

Secondly, a rise in tax progressivity provokes high-skilled emigration out of a country and, thus, lowers high-skilled workers' labor supply in that country. As a result, in general equilibrium, wage inequality rises, which broadens the tax base. We label this as a *migration effect on wages*. Both the wage effect on migration and the migration effect on wages work against the predistribution rationale established in the literature on the optimal taxation in general equilibrium and against the argument in the taxation and migration literature that migration responses lower a government's scope for redistribution.

We demonstrate that, depending on the shape of the semi-elasticity of migration, the presence of general equilibrium effects offsets between 6% and 47% of the classic migration-induced erosion in the optimal rate of tax progressivity. In our simulations, the wage effect on migration appears quantitatively more important than the migration effect on wages. We conclude that the threat of migration is, depending on the shape of the migration semi-elasticity, exaggerated, and the canonical Mirrlees (1971) model with fixed wages and no migration provides a more realistic benchmark than previously expected.

Related literature. Our work is related to several recent contributions to the literature on optimal nonlinear income taxation. Stiglitz (1982) initiated the debate on labor income taxation with endogenous wages in a two-type setting. Generalizing Stiglitz (1982) to a continuum of types, Sachs et al. (2020) consider reforms of arbitrarily nonlinear tax schedules and the optimal taxation in general equilibrium. They demonstrate that increasing tax rates in an initially progressive tax system increases government revenue more with endogenous than with exogenous wages. Therefore, depending on the initial tax code, it may be beneficial to raise tax progressivity. Our quantitative framework confirms their result for reforms of the rate of tax progressivity instead of elementary tax reforms (as in Sachs et al. (2020)). However, the setting of Sachs et al. (2020) ignores workers' extensive margin migration responses.

Rothschild and Scheuer (2013) examine the optimal nonlinear income tax schedule in a multi-sector Roy model with endogenous wages. Trickle-down effects are central for their finding that the optimal tax system is more progressive than in an environment without occupational choice. At first glance, one might think that their two-sector setting nests our two-country economy. The key difference to their paper is that, in our model, two governments set tax policies for each country separately. In Rothschild and Scheuer (2013), only one government chooses the tax schedule for both sectors. The latter, however, is equivalent to the coordinated tax policy setup in our model, which we consider in the Appendix.

Our discrete N-type framework connects the closed economy setup with endogenous wages by Stiglitz (1982) to the two-country environment of Lehmann et al. (2014) with internationally mobile workers and heterogeneous migration costs. Contrary to the fixed-wage economies in the tax competition literature (e.g., Lehmann et al. (2014)), in our environment, workers are imperfect substitutes in producing a composite output good under constant returns to scale. One exception departing from exogenous wages is Tsugawa (2021), who also studies tax competition under endogenous wages but restricts attention to a two-type setup.

Our continuous-type framework nests those of Mirrlees (1971), Lehmann et al. (2014), and Sachs et al. (2020). We derive and quantify the traditional effects occurring in the former papers. Moreover, we discover novel effects that enlighten the interaction between migration and general equilibrium responses and compare their quantitative importance to the traditional effects.

Finally, our paper adds to the literature on optimal taxation in the spirit of Ramsey (1927). By restricting attention to the class of CRP tax functions (developed in Feldstein (1969), Persson (1983), and Benabou (2000)), we follow Heathcote et al. (2017), who document that the current U.S. tax code is close to CRP and, then, explore this tax function in a dynamic setting with earnings risk. Heathcote and Tsujiyama (2021a) and Heathcote and Tsujiyama (2021b) study the CRP tax scheme in a Mirrlees (1971) model showing that it approximates well the full optimum.

Outline. In Section 2, we solve a discrete N-type Stiglitz (1982) model with general equilibrium and migration responses. We briefly discuss extensions to our framework and relate our model and findings to the empirical literature. In Section 3, we present our quantitative framework and derive the incidence of reforms to the rate of tax progressivity. Then, we calibrate the framework to the U.S. economy and compute the optimal rate of progressivity in four different policy regimes depending on whether migration and general equilibrium effects are taken into account. Section 4 concludes. We relegate all proofs to the Appendix. Moreover, in the Appendix, we provide a parameterized version of our N-type economy, leading to closed-form expressions for the optimal tax system. Finally, we study the effects of tax coordination under symmetric country sizes.

2 Optimal Marginal Tax Rates and International Migration in General Equilibrium

2.1 The N-Type Model

Economic environment. We extend the canonical model of Stiglitz (1982) to a setting where two countries or regions $i = A, B$ compete for internationally mobile workers. We go beyond the two-type setting studied in Stiglitz (1982) by considering an arbitrary set of skill or productivity types $\theta \in \Theta = \{1, ..., N\}$, which is private information of individuals and not observable to the government. Aside from studying the impact of migration and general equilibrium on bottom and top tax rates, this allows us to speak to the effects on the optimal taxation of the middle class. We order types such that their equilibrium wages are increasing in types $w_{i,\theta} > w_{i,\theta-1}$.[3] Therefore, one can interpret a household's type as its position (e.g., percentile) in the equilibrium wage distribution. Moreover, consider a general class of utility functions, $u(c, l)$, with consumption c, labor supply l, and labor income $y \equiv wl$. Let $u(c, l)$ satisfy the Spence-Mirrlees single-crossing property, $\frac{d}{dw} \frac{-u_l(c,y/w)/w}{u_c(c,y/w)} < 0$, and suppose that labor and consumption are separable (e.g., $u(c, l) = h(c) - v(l)$).

Let $n_{i,\theta}$ be the number of natives (born) in country i with skill θ. Denote $N_{i,\theta}$ as country i's equilibrium mass of θ-type workers and $l_{i,\theta}$ as an individual's labor supply. As we explain later, both the labor supply and the equilibrium population will be endogenous to the tax system. In each country i, competitive firms produce a single composite output under a constant elasticity of substitution (CES)

$$F_i\left(\{l_{i,\theta} N_{i,\theta}\}_{\theta\in\Theta}\right) = \left[\sum_{\theta\in\Theta} a_{i,\theta} \left(l_{i,\theta} N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$

for some $\sigma \in [0, \infty)$ and $a_{i,\theta} \in \mathbb{R}_+$.[4] Consequently, a worker θ's marginal product pins down her wage rate in that country

$$w_{i,\theta} = a_{i,\theta} \left[l_{i,\theta} N_{i,\theta} / F_i\left(\{l_{i,\theta} N_{i,\theta}\}_{\theta\in\Theta}\right)\right]^{-\frac{1}{\sigma}} \quad for\ \theta \in \Theta, \tag{1}$$

[3] In Appendix A, we derive mild conditions under which this ordering is without loss of generality ensuring a one-to-one mapping between skills, wages, and incomes.

[4] In a two-type setup, one can extend the results to any constant-return-to-scale production function.

which she takes as given. Let labor and goods markets clear in each country. Define $\gamma_{i,\theta,\theta} \equiv \frac{\partial w_{i,\theta}}{\partial \left(N_{i,\theta} l_{i,\theta}\right)} \frac{N_{i,\theta} l_{i,\theta}}{w_{i,\theta}} < 0$ and $\gamma_{i,\theta,\theta'} \equiv \frac{\partial w_{i,\theta}}{\partial \left(N_{i,\theta'} l_{i,\theta'}\right)} \frac{N_{i,\theta'} l_{i,\theta'}}{w_{i,\theta}} > 0$ as the own- and cross-wage elasticity. Country i's government taxes labor income $y_{i,\theta} \equiv w_{i,\theta} l_{i,\theta}$ according to a nonlinear tax scheme $T_i\left(y_{i,\theta}\right)$.

Labor supply. Conditional on living in country i, a worker optimally chooses labor supply $l_{i,\theta}$ to maximize utility $u\left(c_{i,\theta}, l_{i,\theta}\right)$. Consumption is given by the after-tax income $c_{i,\theta} = y_{i,\theta} - T_i\left(y_{i,\theta}\right)$. Then, the worker's first-order condition

$$u_c\left(c_{i,\theta}, l_{i,\theta}\right) w_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right) = -u_l\left(c_{i,\theta}, l_{i,\theta}\right) \tag{2}$$

pins down optimal labor supply.

Migration. As in Lehmann et al. (2014), a worker θ born in country i draws a migration cost m from a conditional density function $G_i\left(m|\theta\right) = \int_0^m g_i\left(x|\theta\right) dx$, accounting for the fact that migration costs may differ between workers (even conditional on skill-type). Then, a native in country i, for instance, migrates to country j if and only if $u\left(c_{j,\theta}, l_{j,\theta}\right) - m > u\left(c_{i,\theta}, l_{i,\theta}\right)$. Defining $\Delta_i \equiv u\left(c_{i,\theta}, l_{i,\theta}\right) - u\left(c_{j,\theta}, l_{j,\theta}\right)$, one can derive a country's equilibrium mass of θ-workers as

$$N_{i,\theta} \equiv \rho_i\left(\Delta_i|\theta\right) \equiv \begin{cases} n_{i,\theta} + G_j\left(\Delta_i|\theta\right) n_{j,\theta} & for\ \Delta_i \geq 0 \\ \left(1 - G_i\left(-\Delta_i|\theta\right)\right) n_{i,\theta} & for\ \Delta_i \leq 0 \end{cases}. \tag{3}$$

Accordingly, denote the semi-elasticity of migration as $\eta_{i,\theta} \equiv \frac{\partial \rho_i\left(\Delta_i|\theta\right)}{\partial \Delta_i} \frac{1}{N_{i,\theta}} > 0$.

Government problem. We consider a Nash game between the governments of the two countries. Each government chooses its nonlinear income tax schedule, taking the other country's tax schedule as given and correctly anticipating the migration and labor supply effects from its tax policy. As in Simula and Trannoy (2010) and Lehmann et al. (2014), we consider a Rawlsian objective function: The government maximizes the utility of the lowest type. The approach has the advantage that - given the government's objective of redistributing from high and medium types to the lowest type - constraints from mobility become most visible. In addition, one avoids the issue that the aggregate welfare level depends on migration decisions.[5] Formally, country i's government wants to redistribute to the lowest type (i.e., $T_i\left(y_{i,\theta}\right) \geq T_i\left(y_{i,\theta-1}\right)$), and, thus,

[5]Defining Pareto weights $\{\psi_{i,\theta}\}_{\theta \in \Theta}$ with $\psi_{i,\theta-1} \geq \psi_{i,\theta}$, this exposition is very similar for a utilitarian objective. A Rawlsian government is a special case where $\psi_{i,1} = 1/N_{i,1}$ and $\psi_{i,\theta} = 0$ for all $\theta \geq 2$. Note that this assumes that $\theta = 1$ is present in both countries at the equilibrium allocations.

solves

$$\max_{\left\{c_{i,\theta},l_{i,\theta},N_{i,\theta}\right\}_{\theta\in\Theta}} \quad u\left(c_{i,1},l_{i,1}\right) \tag{4}$$

$$\text{subject to} \quad u\left(c_{i,\theta},l_{i,\theta}\right)\ \geq u\left(c_{i,\theta-1},\frac{l_{i,\theta-1}w_{i,\theta-1}}{w_{i,\theta}}\right)\ \text{for}\ \theta\in\{2,...,N\}, \tag{5}$$

$$\sum_{\theta\in\Theta} N_{i,\theta}c_{i,\theta}\leq F_i\left(\{N_{i,\theta}l_{i,\theta}\}_{\theta\in\Theta}\right), \tag{6}$$

as well as subject to the endogeneity of wages (Equation (1)) and the equilibrium population (Equation (3)), and taking the other country j's allocation as given. Equations (5) and (6) are the high-skilled worker's incentive constraint and the government budget (no public good provision, purely redistributive tax). Observe that one can omit non-local incentive constraints where $\theta' < \theta - 1$ and $\theta' > \theta + 1$ (see Milgrom and Shannon (1994)). Following Stiglitz (1982) and Ales, Kurnaz, and Sleet (2015), we focus on solutions where only the local downward incentive constraints bind. With migration responses, this assumption is not entirely innocuous. We discuss this observation in Section 2.3. Also notice that we implicitly assume that governments do not discriminate between natives and immigrant workers in their taxation. Observe that this planner problem is identical to the one in Ales et al. (2015) except for the endogeneity of the equilibrium population (Equation (3)) and the specification of Pareto weights to a Rawlsian objective function.

2.2 Optimal Marginal Tax Rates

Optimal top and bottom marginal tax rate. As a benchmark, consider the marginal tax rate chosen by an exogenous technology planner who ignores migration ($\eta_{i,\theta} = 0,\ \forall\theta \in \Theta$). In Proposition 1, we characterize high-skilled workers' optimal (Nash equilibrium) marginal tax rate with migration and compare it to the closed economy's optimal tax rate (without migration). For the comparison, we compute in the closed economy the optimal tax rate of an exogenous technology planner, $T_i^{ex\prime}\left(y_{i,\theta}\right)$, who ignores migration but takes as given the number of the population groups $\{N_{i,\theta}\}$ that materialize in the open economy Nash equilibrium.[6] This notion includes the self-confirming policy equilibrium proposed by Rothschild and Scheuer (2013,

[6]More directly, one may focus on the symmetric Nash equilibrium. Then, we do not have to take a stand on why a government, which does not consider migration responses in its optimization (e.g., in a Mirrlees (1971) benchmark), does not respond to migration flows when observing them.

2016), where the exogenous technology planner sets the tax scheme such that it generates outcomes for which it is optimal.[7]

Proposition 1. *In the Nash equilibrium of our N-type economy, the optimal marginal tax at the top is higher, $T_i'(y_{i,N}) > T_i^{ex\prime}(y_{i,N})$, than in the closed economy.*

Proof. See Appendix B. □

Perhaps surprisingly, the optimal marginal tax rate at the top is *higher* with migration than without. This finding is at odds with those from the tax competition literature, where migration leads to *lower* marginal tax rates (e.g., Lehmann et al. (2014)).[8] The reason is that general equilibrium externalities are absent in the existing partial equilibrium models (fixed wages). With endogenous wages, trickle-down forces justify lower top tax rates relative to an economy with fixed wages because a tax cut at the top raises low-skilled wages and lowers high-skilled wages (predistribution). With labor migration, these general equilibrium forces may still call for a lower marginal tax of high-skilled workers (compared to the partial equilibrium) but less relative to an economy without migration. The intuition is that a reduction in high-skilled wages causes an outflow of these workers such that the predistribution rationale is self-limiting. In that sense, trickle-down forces are partly offset by labor migration.

Assumption 1. *Let the migration semi-elasticity be increasing or constant, that is, $\eta_{i,\theta+1} \geq \eta_{i,\theta}$. Moreover, suppose that the native population is symmetric $n_{i,\theta+1} = n_{i,\theta}$.*

For the result on the bottom tax rate, we make Assumption 1. The first part of the assumption rules out that the migration semi-elasticity decreases for some parts of the income distribution.[9] The second part is not very restrictive. For instance, one can interpret a worker's type θ as her position in the income distribution, in which case $n_{i,\theta} = \frac{1}{N}$.

Proposition 2. *Let Assumption 1 hold. Then, in the symmetric Nash equilibrium of our N-type economy, the optimal marginal tax at the bottom is lower, $T_i'(y_{i,1}) < T_i^{ex\prime}(y_{i,1})$, than in the closed economy.*

[7] In Appendix D, we demonstrate in a two-type setup that a Cobb-Douglas technology, a linear consumption utility, and an isoelastic disutility of labor yield, in the symmetric Nash equilibrium, closed-form expressions for $T_i^{ex\prime}(y_{i,\theta})$. Then, there is no endogeneity of the right-hand side variables in the underlying tax code.

[8] The possibility that migration opportunities of the rich may increase tax rates has been noted in other situations (absent of general equilibrium wage effects), such as endogenous land quality (see Glazer, Kanniainen, and Poutvaara (2008)).

[9] In Section 3, we also consider a decreasing semi-elasticity.

Proof. See Appendix B. □

With endogenous wages, the intuition for the lower bottom tax rate is, besides the standard migration elasticity argument in partial equilibrium, the same as the one that calls for a lower marginal subsidy at the top. In response to a lower marginal tax rate, low-skilled workers' labor supply rises. On the one hand, this leads to a decline in low-skilled workers' wage rates. But, on the other hand, due to the complementarity of labor, the wages of high-skilled workers increase. Altogether, the lower marginal tax at the bottom amplifies pre-tax wage inequality in the respective country trying to attract high-skilled workers. Thus, the bottom tax rate does not only decline due to the "migration threat" as described in Lehmann et al. (2014) but also due to the amplification of pre-tax wage inequality.[10]

Optimal middle tax rate. The effects on top and bottom tax rates we derived so far can be considered local, as these only refer to a particular worker type. In the following, we show that the effect of migration on the tax rate of middle incomes, $\theta \in \{2, ..., N-1\}$, is case-specific. Firstly, considering the binding incentive constraint of θ-workers

$$u\left(c_{i,\theta}, l_{i,\theta}\right) = u\left(c_{i,\theta-1}, \frac{l_{i,\theta-1} w_{i,\theta-1}}{w_{i,\theta}}\right),$$

observe that the government can raise θ-workers' utility by amplifying pre-tax wage inequality between $\theta - 1$ and θ (higher $\frac{w_{i,\theta}}{w_{i,\theta-1}}$). Secondly, by the binding incentive constraint at $\theta + 1$

$$u\left(c_{i,\theta+1}, l_{i,\theta+1}\right) = u\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right),$$

more pre-tax wage inequality between θ and $\theta+1$ (higher $\frac{w_{i,\theta+1}}{w_{i,\theta}}$) increases the utility of workers with type $\theta + 1$.

To raise θ-workers' utility, the government has an incentive to tax type θ more to reduce their labor supply and, thereby, raise $\frac{w_{i,\theta}}{w_{i,\theta-1}}$. We call this channel "local trickle-down" since a tax cut on the relatively richer θ-workers would trickle down to the poorer workers of type $\theta-1$. At the same time, the second incentive constraint calls for a higher $\frac{w_{i,\theta+1}}{w_{i,\theta}}$, which the government

[10] To see this, set wage responses equal to zero ($\sigma = \infty$). This setup with fixed wages nests the standard migration-induced decline in bottom tax rates and the unaltered no-distortion-at-the-top-result for finite productivities (see Lehmann et al. (2014)). With endogenous wages, the migration induced tax cut at the bottom is higher in absolute terms than in partial equilibrium, reflecting the incentives to boost pre-tax wage inequality.

can achieve by taxing θ-workers less. We label this mechanism as "local trickle-up" because the tax cut on the poor (type θ) trickles up to the rich (type $\theta+1$). To simplify the exposition, we make the following assumption about the interaction of general equilibrium and migration responses.

Assumption 2. *Let the interaction between the migration semi-elasticity and general equilibrium production complementarities be small:* $\frac{1}{\sigma}\eta_{i,\theta} \to 0$ *for all* $\theta \in \Theta$.

Under Assumption 2, one can easily relate the effect of migration in general equilibrium to the exogenous technology planner's marginal tax rate. We summarize this result in Proposition 3. Notice that the proof of the proposition does not rely on Assumption 1 or on any symmetry assumption.

Proposition 3. *Let Assumption 2 hold. Then, in the Nash equilibrium of our* N*-type economy, the optimal marginal tax of a worker with type* θ *is higher in the open economy,* $T_i\left(y_{i,\theta}\right) > T_i^{ex\prime}\left(y_{i,\theta}\right)$*, if and only if the marginal tax rate set by the exogenous technology planner is negative,* $T_i^{ex\prime}\left(y_{i,\theta}\right) < 0$*. The optimal marginal tax of a* θ*-type worker is lower in the open economy,* $T_i\left(y_{i,\theta}\right) < T_i^{ex\prime}\left(y_{i,\theta}\right)$*, if and only if the exogenous technology planner's marginal tax rate is positive,* $T_i^{ex\prime}\left(y_{i,\theta}\right) > 0$.

Proof. See Appendix C. □

Thus, if the exogenous technology planner's marginal tax rate of type θ is negative, the tax rate will rise due to the presence of migration. In this situation, the exogenous technology planner's incentive to decrease θ's tax rate to raise pre-tax wage inequality between $\theta-1$ and θ is sufficiently strong. The mechanism that calls for a rise in wage inequality $\frac{w_{i,\theta}}{w_{i,\theta-1}}$ (local trickle-down) dominates the local trickle-up mechanism that calls for a rise in $\frac{w_{i,\theta+1}}{w_{i,\theta}}$. Vice versa, if the exogenous technology planner sets a positive marginal tax rate, then the presence of migration responses has the intuitive negative impact on the optimal marginal tax rate.

Another way to read this proposition is that even under the admittedly strong Assumption 2, the presence of general-equilibrium effects may increase or decrease a worker θ's marginal tax rate depending on the local effect on pre-tax wage inequality. On the one hand, a rise in the worker's wage, $w_{i,\theta}$, increases wage inequality (relative to workers below type θ). On the other hand, the wage rise lowers wage inequality (relative to workers above type θ). The proposition shows that the overall impact depends on whether the worker's wage initially raised local wage inequality to justify negative marginal tax rates in the N-type Stiglitz (1982) setup ($T_i^{ex\prime}(y_{i,\theta}) < 0$).

Proposition 3 further illustrates Propositions 1 and 2. At the top, migration reduces the optimal tax rate since the closed economy's optimal tax rate is negative. However, the bottom tax rate rises in response to migration because the exogenous technology planner's tax rate is positive.

2.3 Discussion and Extensions

We start our discussion of the results with a note of caution. The presence of international migration opportunities makes optimal tax codes more progressive at the ends of the type distribution in terms of marginal tax rates. Nonetheless, the effect of migration on net-tax payments and transfers could be the opposite. To understand this ambiguity, consider the case with only two types, $N = 2$. Starting from the optimal consumption allocation without migration and holding labor supply fixed, a revenue-neutral reduction in lump-sums to low-skilled workers leads to high-skill immigration and low-skill emigration. The government can use the resulting fiscal surplus for transfers to low-skilled workers leading to a welfare improvement. However, this line of reasoning is not complete to prove lower tax burdens on high-skilled workers because labor supplies also change, thereby affecting tax payments (which are the difference between consumption and gross income). Therefore, one cannot infer from the results on the responses of marginal tax rates to the reaction of average tax rates and, thus, tax progressivity. This observation motivates the quantitative framework in Section 3. Moreover, the quantitative framework is free of restrictions on the migration semi-elasticities and their interaction with general equilibrium forces which are necessary to obtain theoretical results on the middle and bottom tax rates.

Moreover, we have focused on solutions where only the local downward incentive constraints bind. This assumption makes our results directly comparable to those in the literature on optimal taxation in general equilibrium (e.g., Stiglitz (1982) and Ales et al. (2015)). In the presence of migration, however, the assumption is debatable because sufficiently increasing migration semi-elasticities may imply decreasing tax payments at the very top (see Lehmann et al. (2014)). In such a situation, local upward incentive constraints would bind. This issue is local as it would only concern workers at the very top – in our model the worker $\theta = N$. Notwithstanding, our solution does not consider the possibility of binding local upward incentive constraints.

Our setup allows us to consider the effects of tax coordination on marginal income tax rates. Tax coordination provides a way to overcome the inefficiencies from the non-cooperative setting of tax policies, although typically in the context of representative household models. In Appendix E, we show that in our framework, under cross-country symmetry, governments can restore the autarky solution by coordinating their income taxation. The intuition is that governments internalize the cross-country externalities from international labor migration when coordinating their tax policies. Thus, in general equilibrium, international coordination of income taxation leads to *less* tax progressivity in terms of marginal tax rates. This finding is in contrast to the conventional view that fiscal competition between governments limits the amount of redistribution, and tax coordination may, therefore, raise the level of tax progressivity (for a survey of the literature on tax competition and coordination, see Keen and Konrad (2013)). Moreover, notice that the two-type version of the coordinated tax policy setup is equivalent to Rothschild and Scheuer (2013). In their model, a policymaker sets a tax scheme under occupational mobility. In our coordination setting, a planner chooses the tax system in both countries subject to the international mobility of labor.

3 Optimal Tax Progressivity and International Migration in General Equilibrium

3.1 Quantitative Framework

In the following, we develop a standard framework of optimal tax progressivity and consider a Mirrleesian economy with migration and general equilibrium wage responses. However, we depart from the Mirrleesian approach of solving for an optimal arbitrarily nonlinear tax scheme and focus on finding the optimum of a parameterized nonlinear tax function. This complements and extends our analysis of the previous N-type model and allows us to go beyond statements about marginal tax rates. We, thereby, contribute to the recent literature on Ramsey vs. Mirrlees taxation (Heathcote and Tsujiyama (2021a)), optimal taxation in general equilibrium (Sachs et al. (2020)), and provide a connection to optimal taxation with migration in partial equilibrium (Lehmann et al. (2014)).

Setup. We retain the two-country framework. In contrast to Section 2, we assume that, in country $i \in \{A, B\}$, there is a continuum of workers $\theta \in \Theta = [0,1] \sim F_i(\theta)$. Without loss of generality, a worker's type θ can be interpreted as her position in the wage distribution. Each worker's utility function is quasilinear with an isoelastic disutility from labor (measured by e)

$$u(c_{i,\theta}, l_{i,\theta}) = c_{i,\theta} - \frac{l_{i,\theta}^{1+\frac{1}{e}}}{1+\frac{1}{e}}.$$

Consumption is given by the after-tax labor income $c_{i,\theta} = y_{i,\theta} - T_i(y_{i,\theta})$, where $y_{i,\theta} \equiv w_{i,\theta} l_{i,\theta}$ denotes a worker's labor income. Instead of deriving a Diamond-Saez formula for the optimal nonlinear tax function (see Diamond (1998) and Saez (2001)), we restrict attention to a standard constant-rate-of-progressivity (CRP) tax function

$$T_i(y_{i,\theta}) = y_{i,\theta} - \frac{1-\tau_i}{1-p_i} y_{i,\theta}^{1-p_i},$$

for $\tau_i \in \mathbb{R}$ and $p_i < 1$. The parameter p_i represents the rate of progressivity and is the key focus of our analysis. The other parameter τ_i is relevant for the level of tax revenues and therefore of less interest for our purposes). It has been shown that this CRP tax scheme not only approximates well the shape of the current U.S. tax code (see Heathcote et al. (2017)) but also delivers policy prescriptions close to the Mirrleesian optimum in a closed economy (Heathcote and Tsujiyama (2021a) and Heathcote and Tsujiyama (2021b)). Given this parametrization, each worker's optimal labor supply reads as $l_{i,\theta} = (1-\tau_i)^{\frac{e}{1+p_i e}} w_{i,\theta}^{\frac{(1-p_i)e}{1+p_i e}}$. The reduced-form elasticities of optimal labor supply along the nonlinear budget with respect to the wage rate and the tax scheme's level parameter and progressivity rate, respectively, set are given by

$$\varepsilon_{i,\theta}^{l,(1-\tau)} \equiv \frac{\partial l_{i,\theta}}{\partial(1-\tau_i)} \frac{1-\tau_i}{l_{i,\theta}} = \frac{e}{1+p_i e} = \varepsilon_i^{l,(1-\tau)},$$

$$\varepsilon_{i,\theta}^{l,w} \equiv \frac{\partial l_{i,\theta}}{\partial w_{i,\theta}} \frac{w_{i,\theta}}{l_{i,\theta}} = (1-p_i)\,\varepsilon_i^{l,(1-\tau)} = \varepsilon_i^{l,w},$$

and

$$\varepsilon_{i,\theta}^{l,(1-p)} \equiv \frac{\partial l_{i,\theta}}{\partial(1-p_i)} \frac{1-p_i}{l_{i,\theta}} = \varepsilon_i^{l,w} \varepsilon_i^{l,(1-\tau)} \left[\left(1+\frac{1}{e}\right) log(w_{i,\theta}) + log(1-\tau_i) \right].$$

As before and in Lehmann et al. (2014), a conditional migration cost distribution $G_i(m|\theta)$

gives rise to an endogenous equilibrium population distribution $N_{i,\theta} \equiv \rho_i(\Delta_{i,\theta}|\theta)$ for each worker θ and country i where $\Delta_{i,\theta}$ denotes the utility difference from living in country i relative to country j. Again, denote $\eta_{i,\theta} \equiv \frac{\partial \rho_i(\Delta_{i,\theta}|\theta)}{\partial \Delta_{i,\theta}} \frac{1}{N_{i,\theta}}$ as the semi-elasticity of migration. In each country i, a mass-one continuum of identical firms produces a single final output good using the labor of each worker type. Moreover, let firms produce under a constant elasticity of substitution (CES)

$$F_i\left(\{l_{i,\theta} N_{i,\theta}\}_{\theta \in \Theta}\right) = \left[\int_{\Theta} a_{i,\theta} \left(l_{i,\theta} N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}} d\theta\right]^{\frac{\sigma}{\sigma-1}}$$

for some $\sigma \in [0, \infty)$ and $a_{i,\theta} \in \mathbb{R}_+$. Firms earn zero profits, and, as in the discrete-type setting, workers' wages are equal to the respective marginal productivity of labor

$$w_{i,\theta} = a_{i,\theta}\left[l_{i,\theta} N_{i,\theta} / F_i\left(\{l_{i,\theta} N_{i,\theta}\}_{\theta \in \Theta}\right)\right]^{-\frac{1}{\sigma}} \;\; for\ \theta \in \Theta. \tag{7}$$

For this CES technology, $\sigma = 0$, $\sigma = 1$, and $\sigma = \infty$ correspond to Leontieff, Cobb-Douglas, and exogenous-wage technologies.

As in the N-type setting above, we focus on the optimal Rawlsian tax system. In each country i, the government chooses the progressivity of the tax schedule p_i[11] to maximize the indirect utility of the lowest worker type subject to the aggregate budget constraint (with exogenous revenue requirement E):

$$\max_{p_i} \;\; \mathcal{U}_{i,0} \;\; subject\ to \;\; \mathcal{R}_i \equiv \int_{\Theta} T_i\left(y_{i,\theta}\right) N_{i,\theta} d\theta \geq E, \tag{8}$$

as well as subject to the equilibrium supply and demand of labor, the endogenous migration responses, and taking as given the other government's tax system.

Policy scenarios. We consider a situation with symmetric countries and characterize the optimal progressivity parameter in the symmetric Nash equilibrium ($N_{i,\theta} = \rho_i(0|\theta)$), in which no migration takes place in equilibrium while the threat of migration (Lehmann et al. (2014)) is affecting tax policies. By setting $\sigma = \infty$ and $\eta_{i,\theta} = 0$, $\forall \theta \in \Theta$, accordingly, we compare the Nash equilibrium optimal policies to three well-studied, counterfactual policy environments: Mirrlees (1971) (partial equilibrium/no migration), Lehmann et al. (2014) (partial

[11] For simplicity, we ignore the optimal choice of the tax level parameter τ_i. In the absence of migration responses, one can show that wages do not respond to this parameter ($\frac{dw_{i,\theta}}{d\tau_i} = 0$, $\forall \theta \in \Theta$) and, thus, a social planner chooses the same tax level parameter irrespective of general equilibrium effects. Also see Section 3.3 for a more detailed discussion. By omitting τ_i as a tax instrument in this analysis, all fiscal effects are reflected in the optimal choice of the tax progressivity p_i.

equilibrium/migration), and Sachs et al. (2020) (general equilibrium/no migration).

Aggregate incidence of tax reforms. Before simulating the optimal tax progressivity, we characterize the incidence of reforming the tax progressivity in each policy scenario. This allows us to form an intuition about the underlying economic forces that drive the optimal choice of tax progressivity. As a byproduct, we obtain the planner's first-order condition that characterizes the solution to the taxation problem: Again, denoting ξ_i as the marginal value of public funds, the government chooses, in each policy scenario, the progressivity of the income tax schedule such that $\frac{d(\mathcal{U}_{i,0}/\xi_i+\mathcal{R}_i)}{dp_i}=0$.

For any small policy reform, $dp_i > 0$, one can decompose the first-order impact on the government objective function into five terms

$$\mathcal{ME}_i + \mathcal{BE}_i + \mathcal{GE}_i + \mathcal{MME}_i + \mathcal{WEM}_i. \tag{9}$$

The first term captures the direct effect of a small rise in p_i (i.e., higher tax progressivity) on aggregate tax revenues and the indirect utility of the lowest type

$$\mathcal{ME}_i \equiv \int_{\theta\in\Theta} \frac{\partial T_i\left(y_{i,\theta}\right)}{\partial\left(1-p_i\right)} N_{i,\theta} d\theta \left(-dp_i\right) - \frac{\partial T_i\left(y_{i,0}\right)}{\partial\left(1-p_i\right)} \frac{1}{\xi_i} \left(-dp_i\right). \tag{10}$$

In the literature (e.g., Saez (2001)), this effect is referred to as the *mechanical effect*. The second and third terms collect labor supply and demand effects. Higher tax progressivity reduces the workers' incentives to supply labor (see later for an explicit characterization of $dl_{i,\theta}$ in terms of dp_i) which is commonly labeled as the *behavioral effect*

$$\mathcal{BE}_i \equiv \int_{\theta\in\Theta} \frac{dl_{i,\theta}}{l_{i,\theta}} y_{i,\theta} T_i'\left(y_{i,\theta}\right) N_{i,\theta} d\theta. \tag{11}$$

The combination of the mechanical and the behavioral effect leads to a Diamond-Saez formula for the optimal tax progressivity in partial equilibrium without migration.

In general equilibrium, not only labor supply but also wages (that is, labor demand) respond to tax reforms which is the third term

$$\mathcal{GE}_i \equiv \int_{\theta\in\Theta} \frac{dw_{i,\theta}}{w_{i,\theta}} y_{i,\theta} T_i'\left(y_{i,\theta}\right) N_{i,\theta} d\theta + \frac{dw_{i,0}}{w_{i,0}} \frac{y_{i,0}\left(1-T_i'\left(y_{i,0}\right)\right)}{\xi_i}. \tag{12}$$

As shown by Stiglitz (1982) and Sachs et al. (2020), these wage responses ($dw_{i,\theta}$) call for lower tax progressivity when the government optimally chooses an arbitrarily nonlinear tax schedule. The intuition is that a government can lower the inequality in pre-tax wages by taxing the poor more and the rich less. This tax-induced reduction in wage inequality is called predistribution. For this specification of welfare, wage changes affect tax revenues and the welfare of the lowest ability type. When describing the wage responses below, we show that the term $\mathcal{GE}_i$ also depends on the presence of migration responses.

The fourth term captures the mechanical effect of changing the tax code on the equilibrium population

$$\mathcal{MME}_i \equiv -\int_{\theta\in\Theta} T_i\left(y_{i,\theta}\right)\frac{\partial N_{i,\theta}}{\partial\left(1-p_i\right)}d\theta\left(-dp_i\right) = -\int_{\theta\in\Theta} T_i\left(y_{i,\theta}\right)\frac{\partial T_i\left(y_{i,\theta}\right)}{\partial\left(1-p_i\right)}\eta_{i,\theta}N_{i,\theta}d\theta\left(-dp_i\right). \tag{13}$$

For instance, a rise in tax progressivity leads to labor emigration and, thus, lower tax revenues. This negative impact of labor mobility limiting a government's ability to levy high taxes is typically referred to as the threat of migration (e.g., Lehmann et al. (2014)).

Finally, the fifth term captures a novel *wage effect on migration* that captures the first-order effects of wage changes on the equilibrium population

$$\mathcal{WEM}_i \equiv \int_{\theta\in\Theta} T_i\left(y_{i,\theta}\right)\frac{\partial N_{i,\theta}}{\partial w_{i,\theta}}\left(dw_{i,\theta}\right)d\theta = \int_{\theta\in\Theta} T_i\left(y_{i,\theta}\right)\frac{dw_{i,\theta}}{w_{i,\theta}}y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta}N_{i,\theta}d\theta. \tag{14}$$

As we show below, $\mathcal{WEM}_i$ is positive. The intuition is as follows: By making the tax code more progressive (in terms of p_i) and, therefore, amplifying pre-tax wage inequality, a government can trigger high-skilled immigration and raise more tax revenues. This effect weakens predistribution and works against the threat of migration that calls for lower tax progressivity in response to a rise in labor mobility. The decomposition allows us to consider the various combinations of wage settings (endogenous vs. exogenous) and migration settings (with and without migration) by shutting down one or more of the five effects shown in (9).

Individual incidence of tax reforms. To shed more light on the nature of aggregate responses, we now characterize individual responses to a change in the policy parameter p_i. Absent of income effects, a worker's labor supply responds in general equilibrium in two respects: directly through the behavioral effect and indirectly due to the adjustment in her wage

$$\frac{dl_{i,\theta}}{l_{i,\theta}} = \varepsilon_{i,\theta}^{l,(1-p)}\frac{-dp_i}{1-p_i} + \varepsilon_{i,\theta}^{l,w}\frac{dw_{i,\theta}}{w_{i,\theta}}. \tag{15}$$

Similarly, we perturb the equilibrium population to show that the response of the equilibrium population consists of a direct (mechanical) effect and a wage effect

$$\frac{dN_{i,\theta}}{N_{i,\theta}} = -\eta_{i,\theta}\frac{\partial T_i\left(y_{i,\theta}\right)}{\partial\left(1-p_i\right)}\left(1-p_i\right)\frac{-dp_i}{1-p_i} + \eta_{i,\theta}y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\frac{dw_{i,\theta}}{w_{i,\theta}}. \tag{16}$$

By the envelope theorem, behavioral effects play no first-order role for the equilibrium population.

To determine the impact on the labor supply and the equilibrium population, we derive the incidence on wages by perturbing the wage equation (7)

$$\frac{dw_{i,\theta}}{w_{i,\theta}} = -\frac{1}{\sigma}\left(\frac{dl_{i,\theta}}{l_{i,\theta}} + \frac{dN_{i,\theta}}{N_{i,\theta}}\right) + \frac{1}{\sigma}\frac{\left[\int_{\theta\in\Theta} a_{i,\theta}\left(\frac{dl_{i,\theta}}{l_{i,\theta}} + \frac{dN_{i,\theta}}{N_{i,\theta}}\right)\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}d\theta\right]}{\left[\int_{\theta\in\Theta} a_{i,\theta}\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}d\theta\right]}. \tag{17}$$

By solving this system of three equations, the wage incidence can be written in closed form[12]

$$\frac{dw_{i,\theta}}{w_{i,\theta}} = -\frac{1}{\sigma}\frac{1}{1+\frac{1}{\sigma}\left(\varepsilon_{i,\theta}^{l,w} + \eta_{i,\theta}y_{i,\theta}\left(1-T'\left(y_{i,\theta}\right)\right)\right)}\left\{\frac{dl_{i,\theta}^{PE}}{l_{i,\theta}} + \frac{dN_{i,\theta}^{PE}}{N_{i,\theta}}\right.$$
$$\left. - \frac{\int_{\Theta}\left(\frac{dl_{i,\theta}^{PE}}{l_{i,\theta}} + \frac{dN_{i,\theta}^{PE}}{N_{i,\theta}}\right)y_{i,\theta}N_{i,\theta}/\left[1+\frac{1}{\sigma}\left(\varepsilon_{i,\theta}^{l,w} + \eta_{i,\theta}y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\right)\right]d\theta}{\int_{\Theta} y_{i,\theta}N_{i,\theta}/\left[1+\frac{1}{\sigma}\left(\varepsilon_{i,\theta}^{l,w} + \eta_{i,\theta}y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\right)\right]d\theta}\right\},$$

where we denote $\frac{dl_{i,\theta}^{PE}}{l_{i,\theta}} + \frac{dN_{i,\theta}^{PE}}{N_{i,\theta}} \equiv \varepsilon_{i,\theta}^{l,(1-p)}\frac{-dp_i}{1-p_i} - \eta_{i,\theta}\frac{\partial T_i(y_{i,\theta})}{\partial(1-p_i)}\left(1-p_i\right)\frac{-dp_i}{1-p_i}$ as the partial equilibrium labor supply and migration response.

Thus, the wage responses depend on the presence of migration responses. To make this more transparent, we set $\eta_{i,\theta} = 0$ for any θ in the expression for the wage incidence and define $\widehat{\frac{dw_{i,\theta}}{w_{i,\theta}}}$ as the wage change absent of migration responses. Then, the general equilibrium effect, $\mathcal{GE}_i = \mathcal{WE}_i + \mathcal{MEW}_i$, can be decomposed into a standard wage effect that captures the aggregate (welfare and revenue) effect of wage changes absent of migration (see Sachs et al. (2020))

$$\mathcal{WE}_i \equiv \widehat{\frac{dw_{i,0}}{w_{i,0}}}\frac{y_{i,0}\left(1-T_i'\left(y_{i,0}\right)\right)}{\xi_i} + \int_{\theta\in\Theta}\widehat{\frac{dw_{i,\theta}}{w_{i,\theta}}}y_{i,\theta}T_i'\left(y_{i,\theta}\right)N_{i,\theta}d\theta \tag{18}$$

[12] Plugging Equation (17) into the sum of (15) and (16) gives an inhomogeneous Fredholm integral equation of the second kind that one can solve using standard techniques.

and a novel *migration effect on wages*

$$\mathcal{MEW}_i \equiv \left(\frac{dw_{i,0}}{w_{i,0}} - \frac{\widehat{dw_{i,0}}}{w_{i,0}}\right)\frac{y_{i,0}\left(1-T_i'\left(y_{i,0}\right)\right)}{\xi_i} + \int_{\theta\in\Theta}\left(\frac{dw_{i,\theta}}{w_{i,\theta}} - \frac{\widehat{dw_{i,\theta}}}{w_{i,\theta}}\right) y_{i,\theta}T_i'\left(y_{i,\theta}\right)N_{i,\theta}d\theta. \quad (19)$$

As the *wage effect on migration*, the *migration effect on wages* also works against predistribution: By raising tax progressivity, a government triggers high-skilled emigration and, thereby, raises pre-tax wage inequality. This boost in wage inequality allows the government to collect more taxes at the top such that tax revenues rise. Altogether, both the migration effect on wages and the wage effect on migration work against the conventional migration threat and weaken predistribution.

3.2 Calibration

We assume that the current U.S. tax schedule can be approximated by a CRP tax schedule with parameters $p_i = 0.151$ and $\tau_i = -3$ (Sachs et al. (2020), Heathcote et al. (2017)). Moreover, to match the current U.S. labor income distribution, we proceed as in Sachs et al. (2020): Let earnings below $\$150\,000$ be log-normally distributed with mean 10 and variance 0.95. Above $\$150\,000$, we append a Pareto distribution with a tail parameter that decreases from 2.5 to 1.5 (above $\$250\,000$). We set the Frisch elasticity of labor supply to $\varepsilon = 0.33$ (Chetty (2012)).

In the calibration, we follow Sachs et al. (2020) who extend the approach of Saez (2001): We use the empirical U.S. income distribution and the individual first-order conditions to infer the wage distribution (Saez (2001)) and, for a given elasticity of substitution, to back out the underlying productivity distribution (Sachs et al. (2020)), $\{a_{i,\theta}\}_{\theta\in\Theta}$. We set the elasticity of substitution to $\sigma = 1.5$ (Katz and Murphy (1992), Card and Lemieux (2001), Card (2009)).[13]

We assume an exponential migration cost distribution $G_i\left(m|\theta\right) = 1 - e^{-\delta_{i,\theta}m}$. Observe that $\eta_{i,\theta} = \frac{g_i(m|\theta)}{1-G_i(m|\theta)} = \delta_{i,\theta}$. Therefore, $\delta_{i,\theta}$ is equal to the migration semi-elasticity and, hence, a model primitive. As in Lehmann et al. (2014), we consider three cases depending on whether the semi-elasticity of migration is decreasing, constant, or increasing. Thus, we depart from the assumption in the N-type model that the semi-elasticity is weakly increasing (used for

[13]This value is considered to be at the lower bound of the likely values (Card (2009)). Setting the elasticity of substitution to a value at the upper bound, $\sigma = 2.5$ (Card (2009)), only slightly changes the quantitative results. In the literature there also exist values outside of this range, e.g., $\sigma = 0.6$ (Dustmann, Frattini, and Preston (2013)) and $\sigma = 3.1$ (Heathcote et al. (2017)).

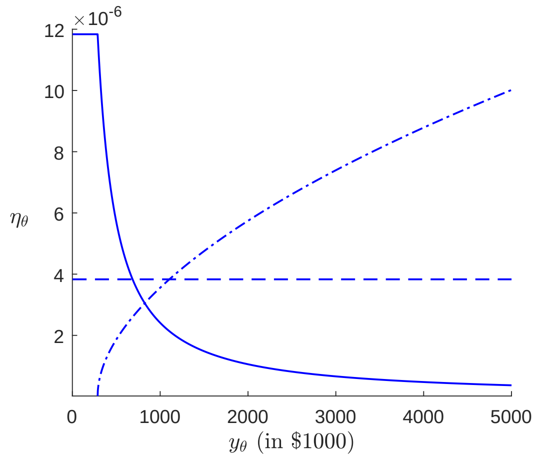

Figure 1: Calibrated Migration (Semi-)Elasticities

Proposition 2), and allow for other possibilities.

In all three cases, we choose the semi-elasticities such that the average migration elasticity, $\nu_{i,\theta} \equiv \eta_{i,\theta} c_{i,\theta}$, of the top 1% of the income distribution is constant $\nu_{i,\theta} = \nu_{top}$, $\forall \theta \geq 0.99$ (Lehmann et al. (2014)). Following Kleven et al. (2020), we choose a medium value for the top migration elasticity: $\nu_{top} = 1.0$. In the first case of Lehmann et al. (2014), the semi-elasticity of migration is a positive constant up to the top 1% and then decreasing. In the second one, it is constant over the whole population and, in the third case, the migration semi-elasticity is zero up to the top 1% and then increasing. In Figure 1, we depict the calibrated migration elasticities and semi-elasticities. In this economy, the calibrated population-wide migration elasticity is the largest in the first case with 0.059 and smaller in the second and third case (0.026 and 0.010, respectively).

3.3 Numerical Simulations

We now use the calibrated model to simulate the optimal rate of tax progressivity. That is, we find the tax progressivity parameter, p_i, that solves the planner's respective first-order condition for each of the four policy scenarios (partial or general equilibrium; with or without migration). In the Mirrlees (1971) scenario, $\mathcal{ME}_i + \mathcal{BE}_i = 0$, whereas, in the setting of Lehmann et al. (2014), $\mathcal{ME}_i + \mathcal{BE}_i + \mathcal{MME}_i = 0$. In both partial equilibrium environments, we assume that the policymaker correctly infers the current wage distribution from the income distribution (as

	$p^{Mirrlees}$	$p^{Lehmann\ et\ al.}$	$p^{Sachs\ et\ al.}$	$p^{GE\ \&\ Migration}$
$\eta_{i,\theta}$ decreasing	0.268	0.214 (-20.1%)	0.268	0.217 (-19.0%)
$\eta_{i,\theta}$ constant	0.268	0.233 (-13.1%)	0.268	0.239 (-10.8%)
$\eta_{i,\theta}$ increasing	0.268	0.238 (-11.2%)	0.268	0.252 (-6.0%)
GE Effects	X	X	✓	✓
Migration	X	✓	X	✓

Table 1: Optimal Tax Progressivity in Each Policy Scenario; Tax Level Parameter Set to $\tau_i = -3$; Changes in Brackets Relative to Optimal Mirrleesian Progressivity

in Saez (2001)) but incorrectly assumes wages to be fixed. In the Sachs et al. (2020) policy scenario, we choose p_i such that $\mathcal{ME}_i + \mathcal{BE}_i + \mathcal{WE}_i = 0$. Finally, in our setting with migration and general equilibrium effects, p_i solves $\mathcal{ME}_i + \mathcal{BE}_i + \mathcal{GE}_i + \mathcal{MME}_i + \mathcal{WEM}_i = 0$. Since we consider the symmetric Nash equilibrium, in which net migration flows are zero at each point in the income distribution, we do not have to adjust the population mass for the differently chosen rates of tax progressivity in the policy scenarios.

In Table 1, we display the optimal rates of tax progressivity in each of the described policy scenarios. Overall, the effects are moderate. We note that in the absence of migration responses, general equilibrium effects are negligible, leading to virtually the same optimal tax progressivity ($p^{Mirrlees} \approx p^{Sachs\ et\ al.} \approx 0.268$). This optimal rate of progressivity is in line with the values found in the literature (for example, Heathcote and Tsujiyama (2021a)).

As the decomposition below reveals, wage effects, $\mathcal{WE}_i$, are very small in this specification. There are two reasons for that: the production function and the specified preferences. Firstly, starting from the Mirrleesian optimum, the tax scheme is close to linear $T_i'(y_{i,\theta}) \approx \tau_i$ such that the general equilibrium revenue effect can be approximated by

$$\int_{\theta\in\Theta} \widehat{\frac{dw_{i,\theta}}{w_{i,\theta}}} y_{i,\theta} T_i'(y_{i,\theta}) N_{i,\theta} d\theta \approx \tau_i \int_{\theta\in\Theta} \widehat{\frac{dw_{i,\theta}}{w_{i,\theta}}} y_{i,\theta} N_{i,\theta} d\theta.$$

However, for a constant returns to scale production function $\int_{\theta\in\Theta}\widehat{\frac{dw_{i,\theta}}{w_{i,\theta}}}y_{i,\theta}N_{i,\theta}d\theta = 0$.[14] Secondly, without migration and given our assumption of an isoelastic labor disutility, the labor supply elasticity with respect to the progressivity $\varepsilon_{i,\theta}^{l,(1-p)}$ is quantitatively close to constant across skill types θ. This makes wage responses $\widehat{\frac{dw_{i,\theta}}{w_{i,\theta}}}$ small.[15]

This model feature that general equilibrium wage effects are negligible has two advantages for studying the impact of migration responses. For one, it provides us with a benchmark level of optimal progressivity that does not depend on the presence of general equilibrium effects ($p^{Mirrlees} \approx p^{Sachs\ et\ al.}$). For another, the quantitative magnitude of the interaction between general equilibrium and migration effects appears conservative, as the general equilibrium effects are initially small (absent of migration).

We now turn to the impact of migration responses. Depending on the shape of the migration semi-elasticity, in partial equilibrium, migration lowers the optimal tax progressivity by 11% to 20% (see the second column in Table 1). The migration-induced reduction is most pronounced when the semi-elasticity is decreasing because the population-wide migration elasticity is the largest in this case. This confirms the insight by Lehmann et al. (2014) that the adjustment in the tax scheme to migration responses crucially depends on the slope of the migration semi-elasticity. Similarly, the optimal tax progressivity in general equilibrium declines due to the migration responses (third and fourth columns). This observation echoes our prediction from the N-type model (Proposition 3) that migration reduces marginal tax rates (lower p) starting from a tax system with positive marginal tax rates.

However, in all three specifications of the migration-semi elasticity, the migration-induced reduction in the optimal tax progressivity is smaller in general equilibrium. Thus, irrespective of the shape of the migration semi-elasticity (increasing, constant, or decreasing), the presence of general equilibrium effects makes the migration responses less important for tax policy. In this economy, general equilibrium effects offset between 6% and 47% (e.g., from -11.2% to

[14]This observation echoes Corollary 3 in Sachs et al. (2020).

[15]To further illustrate this, consider an alternative specification with income effects where consumption utility is logarithmic. Then, workers' labor supplies and their elasticities are constant, $l_{i,\theta} = (1-p_i)^{\frac{\varepsilon}{1+\varepsilon}}$ and $\varepsilon_{i,\theta}^{l,(1-p)} = \frac{\varepsilon}{1+\varepsilon}$, and wages do not depend on tax progressivity (and wage changes are exactly zero). By a similar argument, the level of tax progressivity, τ_i, plays no role in the current specification with a linear consumption utility. Absent migration, the wage response to a change in τ_i, $\widehat{\frac{dw_{i,\theta}}{w_{i,\theta}}}$, is equal to zero since $\varepsilon_{i,\theta}^{l,(1-\tau)} = \frac{e}{1+p_i e}$ is constant across workers.

	$\mathcal{ME}_i$	$\mathcal{BE}_i$	$\mathcal{WE}_i$	$\mathcal{MME}_i$	$\mathcal{MEW}_i$	$\mathcal{WEM}_i$
η_θ decreasing	49.5%	−23.9%	0.2%	−23.7%	1.1%	1.6%
η_θ constant	48.1%	−31.4%	0.2%	−16.4%	1.1%	2.8%
η_θ increasing	45.4%	−34.7%	0.2%	−13.3%	1.0%	5.4%

Table 2: Decomposition of Total Effect into Mechanical Effect ($\mathcal{ME}_i$), Behavioral Effect ($\mathcal{BE}_i$), Mechanical Migration Effect ($\mathcal{MME}_i$), General Equilibrium Wage Effect ($\mathcal{WE}_i$), Migration Effect on Wages ($\mathcal{MEW}_i$), and Wage Effect on Migration ($\mathcal{WEM}_i$) (all effects evaluated at the optimum and relative to total absolute effects)

-6.0%) of the migration-induced reduction in the optimal tax progressivity.[16] The offsetting effect is largest for an increasing semi-elasticity. Interestingly, the interaction of migration and general equilibrium responses is thus most important from a policy perspective in the case in which the overall migration elasticity is the smallest.

To shed further light on the underlying mechanisms, in Table 2, we decompose the effects at the optimal progressivity rate with migration and general equilibrium effects. As theoretically expected, the effect of endogenous wages absent of migration ($\mathcal{WE}_i$) is small.[17] However, their interaction with migration responses ($\mathcal{MEW}_i$) makes general equilibrium effects ($\mathcal{GE}_i = \mathcal{WE}_i + \mathcal{MEW}_i$) non-negligible. As argued, the sign of the migration effect on wages is positive. The magnitude appears unrelated to the shape of the migration semi-elasticity. In all three specifications, the wage effect on migration ($\mathcal{WEM}_i$) is quantitatively more important than the other general equilibrium effects. Thus, the wage effect on migration is the main driver offsetting the migration-induced reduction in tax progressivity. It is weaker for a decreasing than for a constant or increasing migration semi-elasticity, as assumed in the N-type model of Section 3.

Finally, we quantify the implications for income inequality. In this exercise, we consider an increasing semi-elasticity of migration. In Figure 2, we compare the income distributions resulting from the optimal tax progressivities chosen in each policy regime. Both the level and the variance in incomes with migration and general equilibrium effects lie, as the optimal

[16]For a purely revenue-maximizing government, the offset ranges between 10% and 50%.
[17]As in Sachs et al. (2020), the sign is positive, starting from an initially progressive tax code.

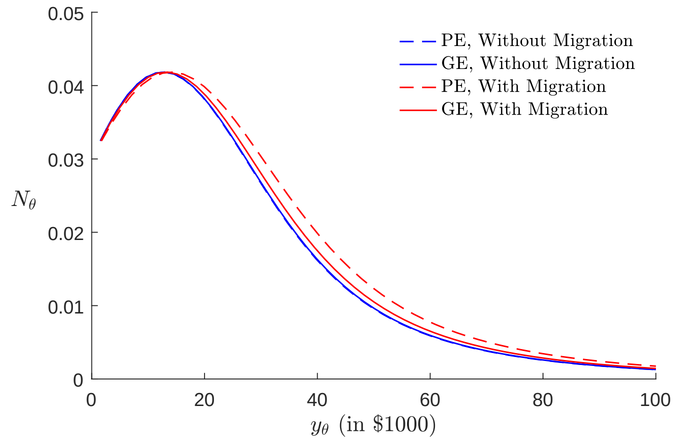

Figure 2: Income Distribution under the Four Regimes (η_θ increasing)

tax progressivity, in between the Lehmann et al. (2014) specification and the Mirrlees (1971) benchmark. One can make a similar statement about marginal and average taxes. For example, in partial equilibrium, migration reduces the average tax rate in the population by ten percentage points. However, the respective reduction in general equilibrium is only five percentage points. Altogether, depending on the shape of the migration semi-elasticity, the common concern of a migration threat limiting a government's scope for redistribution appears overblown.

4 Conclusion

In this paper, we introduce migration into the nonlinear taxation in general equilibrium. By adding an extensive margin in our N-type model and our quantitative framework, we make the canonical Stiglitz (1982) and Mirrlees (1971) models with endogenous labor supply and wages more realistic. As we have shown, contrary to conventional wisdom, migration leads to a more progressive tax code in terms of marginal tax rates. This finding is at odds with Lehmann et al. (2014), who conclude that introducing migration in partial equilibrium reduces marginal tax rates. By weakening general equilibrium trickle-down forces, migration responses move optimal marginal tax rates closer to the partial equilibrium optimum. We explore the underlying mechanisms in our quantitative framework, discovering a novel wage effect on migration as well as a migration effect on wages. Depending on the shape of the migration semi-elasticity, general equilibrium wage effects offset almost half of the migration-reduced decline in optimal progressivity. Thus, the threat of migration appears exaggerated, and the canonical Mirrlees (1971) model with fixed wages provides a more realistic benchmark than previously expected.

Furthermore, our theoretical results suggests an alternative explanation of the recent empirical literature on the effects of globalization on redistribution and inequality. Most prominently, Egger, Nigai, and Strecker (2019) demonstrate that increases in both international trade and migration in OECD countries in the 1980s and early 1990s led to higher average tax burdens (but less in the following years). A well-known explanation for this finding is that redistribution, as well as government size (e.g., Rodrik (1998)), compensates for the adverse effects of globalization on workers from lower parts of the income distribution (see Autor, Dorn, and Hanson (2013)). Our finding does not call into question this widespread view. Instead, it offers an alternative explanation for why globalization may lead to higher tax rates along with a rise in wage inequality. The main difference is that, in the former view, globalization directly amplifies pre-existing inequities, whereas, in our framework, the policy response to international mobility induces the rise in inequality.

References

Ales, L., Kurnaz, M., and Sleet, C. (2015). "Technical change, wage inequality, and taxes." *American Economic Review*, *105*(10), 3061–3101.

Autor, D., Dorn, D., and Hanson, G. H. (2013). "The china syndrome: Local labor market effects of import competition in the united states." *American Economic Review*, *103*(6), 2121–68.

Benabou, R. (2000). "Unequal societies: Income distribution and the social contract." *American Economic Review*, *90*(1), 96–129.

Card, D. (2009). "Immigration and inequality." *American Economic Review*, *99*(2), 1–21.

Card, D., and Lemieux, T. (2001). "Can falling supply explain the rising return to college for younger men? a cohort-based analysis." *The Quarterly Journal of Economics*, *116*(2), 705–746.

Chetty, R. (2012). "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply." *Econometrica*, *80*(3), 969–1018.

Diamond, P. A. (1998). "Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates." *American Economic Review*, 83–95.

Dustmann, C., Frattini, T., and Preston, I. P. (2013). "The effect of immigration along the distribution of wages." *Review of Economic Studies*, *80*(1), 145–173.

Egger, P. H., Nigai, S., and Strecker, N. M. (2019). "The taxing deed of globalization." *American Economic Review*, *109*(2), 353–90.

Feldstein, M. S. (1969). "The effects of taxation on risk taking." *Journal of Political Economy*, *77*(5), 755–764.

Glazer, A., Kanniainen, V., and Poutvaara, P. (2008). "Income taxes, property values, and migration." *Journal of Public Economics*, *92*(3-4), 915–923.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). "Optimal tax progressivity: An analytical framework." *The Quarterly Journal of Economics*, *132*(4), 1693–1754.

Heathcote, J., and Tsujiyama, H. (2021a). "Optimal income taxation: Mirrlees meets ramsey." *Journal of Political Economy*, *129*(11), 000–000.

Heathcote, J., and Tsujiyama, H. (2021b). "Practical optimal income taxation." *Working Paper*.

Katz, L. F., and Murphy, K. M. (1992). "Changes in relative wages, 1963–1987: supply and demand factors." *The Quarterly Journal of Economics*, *107*(1), 35–78.

Keen, M., and Konrad, K. A. (2013). "The theory of international tax competition and coordination." *Handbook of Public Economics*, *5*, 257–328.

Kleven, H., Landais, C., Munoz, M., and Stantcheva, S. (2020). "Taxation and migration: Evidence and policy implications." *Journal of Economic Perspectives*, *34*(2), 119–42.

Lehmann, E., Simula, L., and Trannoy, A. (2014). "Tax me if you can! optimal nonlinear income tax between competing governments." *The Quarterly Journal of Economics*, *129*(4), 1995–2030.

Milgrom, P., and Shannon, C. (1994). "Monotone comparative statics." *Econometrica*, 157–180.

Mirrlees, J. A. (1971). "An exploration in the theory of optimum income taxation." *The Review of Economic Studies*, *38*(2), 175–208.

Persson, M. (1983). "The distribution of abilities and the progressive income tax." *Journal of Public Economics*, *22*(1), 73–88.

Ramsey, F. P. (1927). "A contribution to the theory of taxation." *The Economic Journal*, *37*(145), 47–61.

Rodrik, D. (1998). "Why do more open economies have bigger governments?" *Journal of Political Economy*, *106*(5), 997–1032.

Rothschild, C., and Scheuer, F. (2013). "Redistributive taxation in the roy model." *The Quarterly Journal of Economics*, *128*(2), 623–668.

Rothschild, C., and Scheuer, F. (2016). "Optimal taxation with rent-seeking." *The Review of Economic Studies*, *83*(3), 1225–1262.

Sachs, D., Tsyvinski, A., and Werquin, N. (2020). "Nonlinear tax incidence and optimal taxation in general equilibrium." *Econometrica*, *88*(2), 469–493.

Saez, E. (2001). "Using elasticities to derive optimal income tax rates." *The Review of Economic Studies*, *68*(1), 205–229.

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

Stiglitz, J. E. (1982). "Self-selection and pareto efficient taxation." *Journal of Public Economics*, *17*(2), 213–240.

Tsugawa, S. (2021). "Income tax competition with endogenous wages." *Working Paper*.

A Monotonicity

We show that the CES production function implies a monotone mapping between types and wages in an open economy with migration, so that the ordering of types we impose is without loss of generality. This observation is, for instance, also necessary to ensure that, in the calibration of Section 3.2, the ordering of types is the same for the calibrated wage distribution and the one evaluated at the optimal tax system.

For a CES production function, a worker θ's wage is given by Equation (1) in the N-type setup (Equation (7) in the continuous-type framework). As the number of types becomes very large and thus the distance between types infinitesimal (or if we consider the continuous-type setup of Section 3), we can write

$$\frac{dw_{i,\theta}}{w_{i,\theta}} = \frac{da_{i,\theta}}{a_{i,\theta}} - \frac{1}{\sigma}\left[\frac{dl_{i,\theta}}{l_{i,\theta}} + \frac{dN_{i,\theta}}{N_{i,\theta}}\right].$$

Then, note that $\frac{dl_{i,\theta}}{l_{i,\theta}} = \varepsilon_{i,\theta}^{l,w}\frac{dw_{i,\theta}}{w_{i,\theta}}$ and $\frac{dN_{i,\theta}}{N_{i,\theta}} = \eta_{i,\theta}y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right)\frac{dw_{i,\theta}}{w_{i,\theta}}$ such that

$$\frac{dw_{i,\theta}}{w_{i,\theta}}\left[1 + \frac{1}{\sigma}\left(\varepsilon_{i,\theta}^{l,w} + \eta_{i,\theta}y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right)\right)\right] = \frac{da_{i,\theta}}{a_{i,\theta}}.$$

Therefore, the type ordering is preserved, i.e., $\frac{dw_{i,\theta}}{w_{i,\theta}}$ and $\frac{da_{i,\theta}}{a_{i,\theta}}$ have the same sign whenever $1+\frac{1}{\sigma}\left(\varepsilon_{i,\theta}^{l,w} + \eta_{i,\theta}y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right)\right) > 0$. This condition holds, for instance, when $T_i'\left(y_{i,\theta}\right) \leq 1$ and $\sigma \geq 1$ noting that, by the Spence-Mirrlees condition, $\varepsilon_{i,\theta}^{l,w} > -1$. In this case, there is also a monotone mapping between wages and income, as $\frac{dy_{i,\theta}}{y_{i,\theta}} = \left(1 + \varepsilon_{i,\theta}^{l,w}\right)\frac{dw_{i,\theta}}{w_{i,\theta}}$.

B Proof of Propositions 1 and 2

The optimal tax code can be implicitly described by the households' first-order condition (2). In the following, we, firstly, characterize the solution to the "inner" problem. That is, we solve for the optimal allocation $\{c_{i,\theta}, l_{i,\theta}\}_{\theta\in\Theta}$ for a given population $\{N_{i,\theta}\}_{\theta\in\Theta}$. Secondly, we maximize welfare by choosing $\{N_{i,\theta}\}_{\theta\in\Theta}$, which is the "outer" problem.

Inner problem. The Lagrangian function of the benevolent social planner in country i is defined by

$$\mathcal{L}_i\left(\{N_{i,\theta}\}_{\theta\in\Theta}\right) \equiv u\left(c_{i,1},l_{i,1}\right) + \sum_{\theta\in\{2,\dots,N\}} \mu_{i,\theta}\left[u\left(c_{i,\theta},l_{i,\theta}\right) - u\left(c_{i,\theta-1},\frac{l_{i,\theta-1}w_{i,\theta-1}}{w_{i,\theta}}\right)\right]$$
$$+\xi_i\left[F_i\left(\{N_{i,\theta}l_{i,\theta}\}_{\theta\in\Theta}\right) - \sum_{\theta\in\Theta} N_{i,\theta}c_{i,\theta}\right] + \sum_{\theta\in\Theta}\lambda_{i,\theta}\left[N_{i,\theta} - \rho_i\left(\Delta_i;\theta\right)\right]$$

for a given population. Assuming that the optimization problem is convex and using the definitions of wages, wage elasticities, and migration semi-elasticities, the following first-order conditions describe the unique optimum

$$[c_{i,\theta}] : 0 = \mathbb{1}\left[\theta=1\right]u_c\left(c_{i,\theta},l_{i,\theta}\right) - N_{i,\theta}\xi_i + \mathbb{1}\left[\theta>1\right]\mu_{i,\theta}u_c\left(c_{i,\theta},l_{i,\theta}\right) - \mathbb{1}\left[\theta<N\right]\mu_{i,\theta+1}u_c\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)$$
$$-\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}u_c\left(c_{i,\theta},l_{i,\theta}\right) \tag{20}$$
$$[l_{i,\theta}] : 0 = \mathbb{1}\left[\theta=1\right]u_l\left(c_{i,\theta},l_{i,\theta}\right) + \mathbb{1}\left[\theta>1\right]\mu_{i,\theta}u_l\left(c_{i,\theta},l_{i,\theta}\right) - \mathbb{1}\left[\theta<N\right]\mu_{i,\theta+1}u_l\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)\frac{w_{i,\theta}}{w_{i,\theta+1}}$$
$$+\xi_i w_{i,\theta}N_{i,\theta} - \lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}u_l\left(c_{i,\theta},l_{i,\theta}\right)$$
$$-\sum_{k\in\{2,\dots,N\}}\mu_{i,k}u_l\left(c_{i,k-1},\frac{y_{i,k-1}}{w_{i,k}}\right)\frac{y_{i,k-1}}{l_{i,\theta}w_{i,k}}\left(\gamma_{i,k-1,\theta}-\gamma_{i,k,\theta}\right), \tag{21}$$

for $\theta\in\Theta$, where $\mathbb{1}[\cdot]$ is the indicator function. Inserting Equation (20) into (21) and making use of the high-skilled worker's first-order condition, the marginal tax rate of a worker θ can be written as

$$T_i'\left(y_{i,\theta}\right) = \frac{\mathbb{1}\left[\theta<N\right]\frac{\mu_{i,\theta+1}}{\xi_i}\frac{u_c\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}\left[1+\frac{u_l\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{w_{i,\theta+1}u_c\left(c_{i,\theta},l_{i,\theta}\right)}\right]}{1+\mathbb{1}\left[\theta<N\right]\frac{\mu_{i,\theta+1}}{\xi_i}\frac{u_c\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}}$$
$$+\frac{\frac{1}{y_{i,\theta}N_{i,\theta}}\sum_{k\in\{2,\dots,N\}}\frac{\mu_{i,k}}{\xi_i}u_l\left(c_{i,k-1},\frac{l_{i,k-1}w_{i,k-1}}{w_{i,k}}\right)\frac{y_{i,k-1}}{w_{i,k}}\left(\gamma_{i,k-1,\theta}-\gamma_{i,k,\theta}\right)}{1+\mathbb{1}\left[\theta<N\right]\frac{\mu_{i,\theta+1}}{\xi_i}\frac{u_c\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}}.$$

In their Proposition (1), Ales et al. (2015) show that one can decompose the formula for the optimal tax rate in the N-type Stiglitz (1982) setting without migration ($\eta_{i,\theta}=0,\ \forall\theta\in\Theta$) into a Mirrleesian and a wage compression term. A similar decomposition applies here with the difference that one needs to account for migration responses ($\eta_{i,\theta}>0$).

On the one hand, the Mirrleesian term is augmented by the direct partial equilibrium impact of migration on the objective function, which Lehmann et al. (2014) label as the "migration threat." On the other hand, migration interacts with the wage compression term. Thus, this setting nests the well-known partial equilibrium effect of migration on the optimal taxation and adds general equilibrium moderation effects. In the following, we derive conditions under which the classical partial equilibrium downward force of migration on taxes is offset by our novel general equilibrium moderation effects.

Under a CES production function, this expression for the optimal marginal tax rate simplifies to

$$T_i'\left(y_{i,\theta}\right)=\frac{\mathbb{1}\left[\theta<N\right]\frac{\mu_{i,\theta+1}}{\xi_i}\frac{u_c\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}\left[1+\frac{u_l\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)}{w_{i,\theta+1}u_c\left(c_{i,\theta},l_{i,\theta}\right)}\frac{\sigma-1}{\sigma}\right]+\mathbb{1}\left[\theta>1\right]\frac{\mu_{i,\theta}}{\xi_i}\frac{y_{i,\theta-1}}{y_{i,\theta}N_{i,\theta}}\frac{1}{\sigma}\frac{u_l\left(c_{i,\theta-1},\frac{y_{i,\theta-1}}{w_{i,\theta}}\right)}{w_{i,\theta}}}{1+\mathbb{1}\left[\theta<N\right]\frac{\mu_{i,\theta+1}}{\xi_i}\frac{u_c\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}}.$$

Therefore, setting $\theta=1$ and $\theta=N$, the bottom and top tax rates are characterized by

$$T_i'\left(y_{i,1}\right)=\frac{\frac{\mu_{i,2}}{\xi_i}\frac{u_c\left(c_{i,1},\frac{y_{i,1}}{w_{i,2}}\right)}{N_{i,1}}}{1+\frac{\mu_{i,2}}{\xi_i}\frac{u_c\left(c_{i,1},\frac{y_{i,1}}{w_{i,2}}\right)}{N_{i,1}}}\left[1+\frac{u_l\left(c_{i,1},\frac{y_{i,1}}{w_{i,2}}\right)}{w_{i,2}u_c\left(c_{i,1},l_{i,1}\right)}\frac{\sigma-1}{\sigma}\right]$$

and

$$T_i'\left(y_{i,N}\right)=\frac{\mu_{i,N}}{\xi_i}\frac{y_{i,N-1}}{y_{i,N}N_{i,N}}\frac{1}{\sigma}\frac{u_l\left(c_{i,N-1},\frac{y_{i,N-1}}{w_{i,N}}\right)}{w_{i,N}}$$

respectively. These expressions depend on the (relative) Lagrange multipliers $\frac{\mu_{i,\theta}}{\xi_i}$.

One can obtain the shadow value of public funds by summing up Equation (20) over all types and using the separability of consumption and leisure, which yields $\xi_i=\frac{1-\sum_{\theta\in\Theta}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}}{\sum_{\theta\in\Theta}N_{i,\theta}/u_c\left(c_{i,\theta},l_{i,\theta}\right)}$. Plugging this value back into (20), the normalized multiplier on the incentive constraint for $k\in\{2,...,N\}$ reads as

$$\frac{\mu_{i,k}}{\xi_i}=\frac{\sum_{l=k}^{N}N_{i,l}/u_c\left(c_{i,l},l_{i,l}\right)}{1-\sum_{\theta\in\Theta}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}}\left(1+\frac{\sum_{l=1}^{k-1}N_{i,l}/u_c\left(c_{i,l},l_{i,l}\right)}{\sum_{l=k}^{N}N_{i,l}/u_c\left(c_{i,l},l_{i,l}\right)}\sum_{l=k}^{N}\lambda_{i,l}\eta_{i,l}N_{i,l}-\sum_{l=1}^{k-1}\lambda_{i,l}\eta_{i,l}N_{i,l}\right)\equiv\frac{\mu_{i,k}^{ex}}{\xi_i^{ex}}\Delta_{i,k},$$

where we define the normalized multiplier without migration responses ($\eta_{i,\theta} = 0, \ \forall \theta \in \Theta$) as

$$\frac{\mu_{i,k}^{ex}}{\xi_i^{ex}} \equiv \sum_{l=k}^{N} N_{i,l}/u_c\left(c_{i,l}, l_{i,l}\right)$$

and a scaling factor as

$$\Delta_{i,k} \equiv \frac{1 + \frac{\sum_{l=1}^{k-1} N_{i,l}/u_c\left(c_{i,l}, l_{i,l}\right)}{\sum_{l=k}^{N} N_{i,l}/u_c\left(c_{i,l}, l_{i,l}\right)} \sum_{l=k}^{N} \lambda_{i,l}\eta_{i,l}N_{i,l} - \sum_{l=1}^{k-1} \lambda_{i,l}\eta_{i,l}N_{i,l}}{1 - \sum_{\theta \in \Theta} \lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}}.$$

This procedure gives us the exogenous technology planner's optimal bottom and top tax rates

$$T_i^{ex\prime}\left(y_{i,1}\right) \equiv \frac{\frac{\mu_{i,2}^{ex}}{\xi_i^{ex}} \frac{u_c\left(c_{i,1}, \frac{y_{i,1}}{w_{i,2}}\right)}{N_{i,1}}}{1 + \frac{\mu_{i,2}^{ex}}{\xi_i^{ex}} \frac{u_c\left(c_{i,1}, \frac{y_{i,1}}{w_{i,2}}\right)}{N_{i,1}}} \left[1 + \frac{u_l\left(c_{i,1}, \frac{y_{i,1}}{w_{i,2}}\right)}{w_{i,2}u_c\left(c_{i,1}, l_{i,1}\right)} \frac{\sigma - 1}{\sigma}\right] > 0$$

and

$$T_i^{ex\prime}\left(y_{i,N}\right) \equiv \frac{\mu_{i,N}^{ex}}{\xi_i^{ex}} \frac{y_{i,N-1}}{y_{i,N}N_{i,N}} \frac{1}{\sigma} \frac{u_l\left(c_{i,N-1}, \frac{y_{i,N-1}}{w_{i,N}}\right)}{w_{i,N}} < 0,$$

to which we can now compare the optimal Nash equilibrium tax rates. The comparison between $T_i^{\prime}\left(y_{i,\theta}\right)$ and $T_i^{ex\prime}\left(y_{i,\theta}\right)$ depends on the adjustment in the Lagrange multipliers measured by the scaling factor $\Delta_{i,k}$ (i.e., $\frac{\mu_{i,\theta}}{\xi_i}$ vs. $\frac{\mu_{i,\theta}^{ex}}{\xi_i^{ex}}$).[18] More precisely, for Propositions 1 and 2, we need to show that $\Delta_{i,N} < 1$ and $\Delta_{i,2} < 1$, which holds for $\lambda_{i,N}\eta_{i,N}N_{i,N} < 0$ and $\sum_{l=2}^{N} \lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta} < 0$, respectively.

Outer problem. To prove these statements, we now derive the first-order conditions with respect to the population masses

$$\begin{aligned}\left[N_{i,\theta}\right] : 0 = &- \sum_{k \in \{2,\ldots,N\}} \mu_{i,k} u_l\left(c_{i,\theta-1}, \frac{l_{i,k-1}w_{i,k-1}}{w_{i,k}}\right) \frac{l_{i,k-1}w_{i,k-1}}{N_{i,\theta}w_{i,k}} \left(\gamma_{i,k-1,\theta} - \gamma_{i,k,\theta}\right) \\ &+ \xi_i\left(l_{i,\theta}w_{i,\theta} - c_{i,\theta}\right) + \lambda_{i,\theta}. \qquad (22)\end{aligned}$$

[18]Notice that the right-hand side of $T_i^{ex\prime}\left(y_{i,\theta}\right)$ depends on the equilibrium allocation that may be endogenous to tax policy. For simplicity, we evaluate, in the following comparison, the right-hand side at the allocation chosen in the Nash equilibrium. In Appendix D, we provide conditions under which the right-hand side is given in closed form and, thus, independent from the equilibrium allocation.

For a constant elasticity of substitution production function, Equation (22) simplifies to

$$-\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}=\xi_i T_i\left(y_{i,\theta}\right)N_{i,\theta}\eta_{i,\theta}+\frac{1}{\sigma}\eta_{i,\theta}\left[\mu_{i,\theta+1}u_l\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)\frac{y_{i,\theta}}{w_{i,\theta+1}}-\mu_{i,\theta}u_l\left(c_{i,\theta-1},\frac{y_{i,\theta-1}}{w_{i,\theta}}\right)\frac{y_{i,\theta-1}}{w_{i,\theta}}\right].$$

Noting that $T_i\left(y_{i,N}\right)\geq 0$, since $T_i\left(y_{i,\theta}\right)\geq T_i\left(y_{i,\theta-1}\right)$ and $\sum_{\theta\in\Theta}T_i\left(y_{i,\theta}\right)N_{i,\theta}\geq 0$, we conclude that

$$-\lambda_{i,N}\eta_{i,N}N_{i,N}=\xi_i T_i\left(y_{i,N}\right)N_{i,N}\eta_{i,N}-\frac{1}{\sigma}\eta_{i,N}\mu_{i,N}u_l\left(c_{i,N-1},\frac{y_{i,N-1}}{w_{i,N}}\right)\frac{y_{i,N-1}}{w_{i,N}}>0.$$

Thus, do not rely on any assumption about the migration semi-elasticity for the result about top tax rate (Proposition 1).

As mentioned above, for the decline in bottom tax rate postulated in Proposition 2, we need to show that

$$-\sum_{\theta=2}^{N}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}=-\sum_{\theta=1}^{N}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}+\lambda_{i,1}\eta_{i,1}N_{i,1}>0.$$

Let Assumption 1 hold. Then, in the symmetric Nash equilibrium

$$\begin{aligned}-\sum_{\theta=1}^{N}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}&=\xi_i\sum_{\theta=1}^{N}T_i\left(y_{i,\theta}\right)n_{i,\theta}\eta_{i,\theta}\\&\quad+\frac{1}{\sigma}\sum_{\theta=1}^{N}\eta_{i,\theta}\left[\mu_{i,\theta+1}u_l\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)\frac{y_{i,\theta}}{w_{i,\theta+1}}-\mu_{i,\theta}u_l\left(c_{i,\theta-1},\frac{y_{i,\theta-1}}{w_{i,\theta}}\right)\frac{y_{i,\theta-1}}{w_{i,\theta}}\right]\end{aligned}$$

The second summand is positive for $\eta_{i,\theta}\geq\eta_{i,\theta-1}$. For a constant population size ($n_{i,\theta}=n_{i,\theta-1}$), the first summand can be written as

$$\xi_i n_{i,\theta}\sum_{\theta=1}^{N}T_i\left(y_{i,\theta}\right)\eta_{i,\theta}\geq\left(\frac{1}{N}\sum_{\theta=1}^{N}\eta_{i,\theta}\right)\xi_i\sum_{\theta=1}^{N}T_i\left(y_{i,\theta}\right)n_{i,\theta}=0$$

where the first inequality is Chebyshev's sum inequality for $\eta_{i,\theta}\geq\eta_{i,\theta-1}$ and $T_i\left(y_{i,\theta}\right)\geq T_i\left(y_{i,\theta-1}\right)$ and the second inequality follows from the government's budget constraint. Therefore, $-\sum_{\theta=1}^{N}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}\geq 0$. To conclude that $-\sum_{\theta=2}^{N}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}>0$, notice that

$$\lambda_{i,1}\eta_{i,1}N_{i,1}=-\xi_i T_i\left(y_{i,1}\right)n_{i,1}\eta_{i,1}-\frac{1}{\sigma}\eta_{i,1}\mu_{i,2}u_l\left(c_{i,1},\frac{y_{i,1}}{w_{i,2}}\right)\frac{y_{i,1}}{w_{i,2}}>0$$

since $T_i\left(y_{i,1}\right)\leq 0$, for $\sum_{\theta=1}^{N}T_i\left(y_{i,\theta}\right)N_{i,\theta}\geq 0$ and $T_i\left(y_{i,\theta}\right)\geq T_i\left(y_{i,\theta-1}\right)$, and $u_l\left(c_{i,1},\frac{y_{i,1}}{w_{i,2}}\right)<$

0.

C Proof of Proposition 3

Recall that the optimal marginal tax rate of a middle type $\theta \in \{2, ..., N-1\}$ is given by

$$T_i'(y_{i,\theta}) = \frac{\Delta_{i,\theta+1}\frac{\mu_{i,\theta+1}^{ex}}{\xi_i^{ex}}\frac{u_c\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}\left[1+\frac{u_l\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{w_{i,\theta+1}u_c\left(c_{i,\theta},l_{i,\theta}\right)}\frac{\sigma-1}{\sigma}\right]+\Delta_{i,\theta}\frac{\mu_{i,\theta}^{ex}}{\xi_i^{ex}}\frac{y_{i,\theta-1}}{y_{i,\theta}N_{i,\theta}}\frac{1}{\sigma}\frac{u_l\left(c_{i,\theta-1},\frac{l_{i,\theta-1}w_{i,\theta-1}}{w_{i,\theta}}\right)}{w_{i,\theta}}}{1+\Delta_{i,\theta+1}\frac{\mu_{i,\theta+1}^{ex}}{\xi_i^{ex}}\frac{u_c\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}},$$

whereas the exogenous technology planner tax rate reads as

$$T_i^{ex\prime}(y_{i,\theta}) \equiv \frac{\frac{\mu_{i,\theta+1}^{ex}}{\xi_i^{ex}}\frac{u_c\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}\left[1+\frac{u_l\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{w_{i,\theta+1}u_c\left(c_{i,\theta},l_{i,\theta}\right)}\frac{\sigma-1}{\sigma}\right]+\frac{\mu_{i,\theta}^{ex}}{\xi_i^{ex}}\frac{y_{i,\theta-1}}{y_{i,\theta}N_{i,\theta}}\frac{1}{\sigma}\frac{u_l\left(c_{i,\theta-1},\frac{l_{i,\theta-1}w_{i,\theta-1}}{w_{i,\theta}}\right)}{w_{i,\theta}}}{1+\frac{\mu_{i,\theta+1}^{ex}}{\xi_i^{ex}}\frac{u_c\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}}.$$

Then, the presence of migration responses raises θ-workers' optimal tax rate, i.e., $T_i(y_{i,\theta}) > T_i^{ex\prime}(y_{i,\theta})$, if and only if

$$T_i^{ex\prime}(y_{i,\theta}) < \frac{\Delta_{i,\theta+1}-\Delta_{i,\theta}}{1-\Delta_{i,\theta+1}}\frac{\mu_{i,\theta}^{ex}}{\xi_i^{ex}}\frac{y_{i,\theta-1}}{y_{i,\theta}N_{i,\theta}}\frac{1}{\sigma}\frac{-u_l\left(c_{i,\theta-1},\frac{l_{i,\theta-1}w_{i,\theta-1}}{w_{i,\theta}}\right)}{w_{i,\theta}}.$$

However, under Assumption 2, the right-hand side of the expression converges to zero, and Proposition 3 follows. Observe that we do not need to impose any assumption on the shape of the migration semi-elasticity or symmetry assumptions.

D A Closed-Form Example

The purpose of this section is to provide an example in which one obtains closed-form expressions for the optimal tax rates chosen by the exogenous technology planner. Let $N = 2$ and label L and H as the low- and high-skill worker type. Suppose that $F_i(N_{i,L}l_{i,L}, N_{i,H}l_{i,H}) = A_i(N_{i,L}l_{i,L})^{\alpha}(N_{i,H}l_{i,H})^{1-\alpha}$ for $\alpha \in (0,1)$. Then, the own- and cross-wage elasticities are given by $\gamma_{i,L,L} = -(1-\alpha)$, $\gamma_{i,H,H} = -\alpha$, $\gamma_{i,L,H} = 1-\alpha$, and $\gamma_{i,H,L} = \alpha$. The income share of the unskilled relative to the skilled workers reads as $\frac{N_{i,L}l_{i,L}w_{i,L}}{N_{i,H}l_{i,H}w_{i,H}} = \frac{\alpha}{1-\alpha}$. Moreover, let the consumption utility be linear $u(c) = c$, and assume an isoelastic disutility from labor $v(l) = \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}$ with ε denoting the Frisch elasticity of labor supply. Finally, consider a setup with symmetric countries – thus, the symmetric Nash equilibrium in which no mobility occurs ($N_{i,\theta} = n_{i,\theta}$ for $i = A, B$ and $\theta = L, H$).

Then, one can write the exogenous technology planner's marginal tax rate for the high-skilled workers as $T_i^{ex\prime}(y_{i,H}) = -\left(\frac{\alpha}{1-\alpha}\frac{n_{i,H}}{n_{i,L}}\right)^{1+1/\varepsilon}\frac{l_{i,H}^{1/\varepsilon}}{w_{i,H}}$. Using the workers' first-order condition, the marginal tax rate at the top simplifies to

$$\frac{T_i^{ex\prime}(y_{i,H})}{1-T_i^{ex\prime}(y_{i,H})} = -\left(\frac{\alpha}{1-\alpha}\frac{n_{i,H}}{n_{i,L}}\right)^{1+1/\varepsilon} < 0.$$

Applying similar steps, the marginal tax rate for low-skilled workers reads as

$$\frac{T_i^{ex\prime}(y_{i,L})}{1-T_i^{ex\prime}(y_{i,L})} = \frac{n_{i,H}}{n_{i,L}} > 0.$$

Now, we show that, in this parametrization, there is a negative reduced-form relationship between high-skilled workers' gross income and their marginal tax rate. Notice that this exercise is non-trivial in our setup since one needs to consider general equilibrium wage effects from labor supply and migration. By the high-skilled workers' first-order condition $(l_{i,H})^{1/\varepsilon} = w_{i,H}\left[1-T'(y_{i,H})\right]$, their income response to a cut in the top tax rate depends on a direct labor supply and an indirect wage response $\frac{dlog\left(y_{i,H}\right)}{dlog\left[1-T'\left(y_{i,H}\right)\right]} = (1+\varepsilon)\frac{dlog\left(w_{i,H}\right)}{dlog\left[1-T'\left(y_{i,H}\right)\right]} + \varepsilon$, whose overall sign is not clear a priori. To calculate the indirect wage response, first derive low-skilled workers' consumption, income, and wage responses

$$\frac{dc_{i,L}}{d\left[1-T'(y_{i,H})\right]} = \frac{1-T'(y_{i,L})}{1-T'(y_{i,H})}y_{i,L}(1+\varepsilon)\frac{dlog(w_{i,L})}{dlog\left[1-T'(y_{i,H})\right]},$$

$$\frac{dl_{i,L}}{d\left[1-T'(y_{i,H})\right]} = \frac{1-T'(y_{i,L})}{1-T'(y_{i,H})}l_{i,L}\varepsilon\frac{dlog(w_{i,L})}{dlog\left[1-T'(y_{i,H})\right]},$$

and

$$\frac{dlog(w_{i,L})}{dlog\left[1-T'(y_{i,H})\right]} = \frac{dlog\left[\alpha A_i(N_{i,L}l_{i,L})^{\alpha-1}(N_{i,H}l_{i,H})^{1-\alpha}\right]}{dlog\left[1-T'(y_{i,H})\right]} = -\frac{1-\alpha}{\alpha}\frac{dlog(w_{i,H})}{dlog\left[1-T'(y_{i,H})\right]}.$$

Accordingly, high-skilled workers' wages change as follows

$$\frac{dlog\left(w_{i,H}\right)}{dlog\left[1-T'\left(y_{i,H}\right)\right]}=\frac{\alpha}{1+\alpha\varepsilon}\eta_{i,L}\left[1-T'\left(y_{i,H}\right)\right]y_{i,L}\left(1+\varepsilon\right)\frac{dlog\left(w_{i,L}\right)}{dlog\left[1-T'\left(y_{i,H}\right)\right]}+\frac{\alpha\varepsilon}{1+\alpha\varepsilon}\frac{dlog\left(w_{i,L}\right)}{dlog\left[1-T'\left(y_{i,H}\right)\right]}$$
$$-\frac{\alpha}{1+\alpha\varepsilon}\eta_{i,H}\left[1-T'\left(y_{i,H}\right)\right]y_{i,H}\left[\left(1+\varepsilon\right)\frac{dlog\left(w_{i,H}\right)}{dlog\left[1-T'\left(y_{i,H}\right)\right]}+\varepsilon\right]-\frac{\alpha\varepsilon}{1+\alpha\varepsilon}.$$

Using the expression for low-skilled workers' wage response, one can rewrite high-skilled workers' wage change as

$$\frac{dlog\left(w_{i,H}\right)}{dlog\left[1-T'\left(y_{i,H}\right)\right]}=-\frac{\alpha\varepsilon}{1+\varepsilon}\frac{1+\eta_{i,H}\left[1-T'\left(y_{i,H}\right)\right]y_{i,H}}{1+\left(1-\alpha\right)\eta_{i,L}\left[1-T'\left(y_{i,H}\right)\right]y_{i,L}+\alpha\eta_{i,H}\left[1-T'\left(y_{i,H}\right)\right]y_{i,H}}.$$

Therefore, recalling that $\alpha<1$, the relationship between high-skilled workers' income and the retention rate of the top tax rate is positive

$$\frac{dlog\left(y_{i,H}\right)}{dlog\left[1-T'\left(y_{i,H}\right)\right]}=-\alpha\varepsilon\frac{1+\eta_{i,H}\left[1-T'\left(y_{i,H}\right)\right]y_{i,H}}{1+\left(1-\alpha\right)\eta_{i,L}\left[1-T'\left(y_{i,H}\right)\right]y_{i,L}+\alpha\eta_{i,H}\left[1-T'\left(y_{i,H}\right)\right]y_{i,H}}+\varepsilon>0.$$

E Coordinated Tax Policy

We consider a situation in which the two governments can jointly set their country-specific tax schedules to maximize world welfare. Then, the world social planner chooses $\{c_{i,\theta},l_{i,\theta},N_{i,\theta}\}_{\theta\in\Theta,i=A,B}$ to maximize

$$\sum_{i=A,B}u\left(c_{i,1},l_{i,1}\right)\qquad(23)$$

subject to the high-skilled workers' incentive constraints (Equation (5)), each country's resource constraint (Equation (6)), the endogeneity of wages (Equation (1)), and the equilibrium population (Equation (3)).

Observe that, as before, the set of constraints needs to hold at a country level.[19] Then, the

[19] Alternatively, one could consider a planner problem where the aggregate resource constraint (6) only has to hold worldwide, allowing governments to achieve cross-country redistribution by trading consumption levels. Although straightforward to consider, we disregard such incentives for the sake of comparability and due to their limited feasibility.

Lagrangian of the outer problem reads as

$$\mathcal{L}\left(\{N_{i,\theta}\}_{\theta\in\Theta,i=A,B}\right) \equiv \sum_{i=A,B} u\left(c_{i,1},l_{i,1}\right) + \sum_{i=A,B}\sum_{\theta\in\{2,\ldots,N\}} \mu_{i,\theta}\left[u\left(c_{i,\theta},l_{i,\theta}\right) - u\left(c_{i,\theta-1},\frac{l_{i,\theta-1}w_{i,\theta-1}}{w_{i,\theta}}\right)\right]$$
$$+\sum_{i=A,B}\xi_i\left[F_i\left(\{N_{i,\theta}l_{i,\theta}\}_{\theta\in\Theta}\right) - \sum_{\theta\in\Theta}N_{i,\theta}c_{i,\theta}\right] + \sum_{i=A,B}\sum_{\theta\in\Theta}\lambda_{i,\theta}\left[N_{i,\theta}-\rho_i\left(\Delta_i;\theta\right)\right],$$

which yields the following first-order conditions

$$[c_{i,\theta}]: 0 = \mathbb{1}\left[\theta=1\right]u_c\left(c_{i,\theta},l_{i,\theta}\right) - N_{i,\theta}\xi_i + \mathbb{1}\left[\theta>1\right]\mu_{i,\theta}u_c\left(c_{i,\theta},l_{i,\theta}\right) - \mathbb{1}\left[\theta<N\right]\mu_{i,\theta+1}u_c\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)$$
$$-\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}u_c\left(c_{i,\theta},l_{i,\theta}\right) + \lambda_{j,\theta}\eta_{j,\theta}N_{j,\theta}u_c\left(c_{j,\theta},l_{j,\theta}\right) \tag{24}$$
$$[l_{i,\theta}]: 0 = \mathbb{1}\left[\theta=1\right]u_l\left(c_{i,\theta},l_{i,\theta}\right) + \mathbb{1}\left[\theta>1\right]\mu_{i,\theta}u_l\left(c_{i,\theta},l_{i,\theta}\right) - \mathbb{1}\left[\theta<N\right]\mu_{i,\theta+1}u_l\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)\frac{w_{i,\theta}}{w_{i,\theta+1}}$$
$$+\xi_i w_{i,\theta}N_{i,\theta} - \lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}u_l\left(c_{i,\theta},l_{i,\theta}\right) + \lambda_{j,\theta}\eta_{j,\theta}N_{j,\theta}u_l\left(c_{j,\theta},l_{j,\theta}\right)$$
$$-\sum_{k\in\{2,\ldots,N\}}\mu_{i,k}u_l\left(c_{i,k-1},\frac{y_{i,k-1}}{w_{i,k}}\right)\frac{y_{i,k-1}}{y_{i,\theta}}\left(\gamma_{i,k-1,\theta}-\gamma_{i,k,\theta}\right), \tag{25}$$

for $i=A,B$ and $\theta\in\Theta$. Observe that the world social planner now takes into account cross-country externalities from international migration.

As before, plug (24) into (25) and use the workers' first-order condition as well as the CES production function to get the worker's marginal tax rate

$$T_i^{co\prime}\left(y_{i,\theta}\right) = \frac{\mathbb{1}\left[\theta<N\right]\frac{\mu^{co}_{i,\theta+1}}{\xi^{co}_i}\frac{u_c\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}\left[1+\frac{u_l\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{w_{i,\theta+1}u_c\left(c_{i,\theta},l_{i,\theta}\right)}\frac{\sigma-1}{\sigma}\right]}{1+\mathbb{1}\left[\theta<N\right]\frac{\mu^{co}_{i,\theta+1}}{\xi^{co}_i}\frac{u_c\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}}$$
$$+\frac{\mathbb{1}\left[\theta>1\right]\frac{\mu^{co}_{i,\theta}}{\xi^{co}_i}\frac{y_{i,\theta-1}}{y_{i,\theta}N_{i,\theta}}\frac{1}{\sigma}\frac{u_l\left(c_{i,\theta-1},\frac{l_{i,\theta-1}w_{i,\theta-1}}{w_{i,\theta}}\right)}{w_{i,\theta}}}{1+\mathbb{1}\left[\theta<N\right]\frac{\mu^{co}_{i,\theta+1}}{\xi^{co}_i}\frac{u_c\left(c_{i,\theta},\frac{l_{i,\theta}w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}},$$

where $\frac{\mu^{co}_{i,\theta}}{\xi^{co}_i}$ denote the relative Lagrangian multipliers under tax coordination. Then, sum up Equations (24) over all types and plug the resulting expression for

$$\xi_i^{co} = \frac{1-\sum_{\theta\in\Theta}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}+\sum_{\theta\in\Theta}\lambda_{j,\theta}\eta_{j,\theta}N_{j,\theta}\frac{u_c\left(c_{j,\theta},l_{j,\theta}\right)}{u_c\left(c_{i,\theta},l_{i,\theta}\right)}}{\sum_{\theta\in\Theta}N_{i,\theta}/u_c\left(c_{i,\theta},l_{i,\theta}\right)}$$

back into (24) to solve for $\mu_{i,\theta}^{co}$.

However, under cross-country symmetry, the marginal value of public funds is the same as the one of the exogenous technology planner, $\xi_i^{co} = \xi_i^{ex}$. Moreover, Equation (24) simplifies to

$$0 = \mathbb{1}\left[\theta = 1\right] u_c\left(c_{i,\theta}, l_{i,\theta}\right) - N_{i,\theta}\xi_i^{ex} + \mathbb{1}\left[\theta > 1\right] \mu_{i,\theta}^{co} u_c\left(c_{i,\theta}, l_{i,\theta}\right) - \mathbb{1}\left[\theta < N\right] \mu_{i,\theta+1}^{co} u_c\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right).$$

This implies that the other Lagrange multipliers coincide as well, $\mu_{i,\theta}^{co} = \mu_{i,\theta}^{ex}$, and the coordination solution is equivalent to the autarky allocation: $T_i^{co\prime}\left(y_{i,\theta}\right) = T_i^{ex\prime}\left(y_{i,\theta}\right)$.