Nonlinear Taxation and International Mobility in General Equilibrium*

Eckhard Janeba† **Karl Schulz**‡

University of Mannheim University of Mannheim

May 26, 2021

Abstract

We study the nonlinear taxation of internationally mobile workers in general equilibrium. At odds with conventional wisdom, in general equilibrium, migration lowers the optimal marginal tax rate at the bottom but raises the optimal top tax rate, making the tax code more progressive and moving marginal tax rates closer to those in an economy with fixed wages. The intuition is that governments attract high-skilled workers by amplifying pre-tax wage inequality and thereby partly offsetting trickle-down forces from production complementarities. This finding raises doubts about the importance of trickle-down externalities for optimal tax policy. Moreover, it offers a novel explanation for why globalization may lead to higher tax progressivity and wage inequality.

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. The importance of international migration has already been recognized by Mirrlees (1971).[1] According to the classical inverse elasticity rule, the mobility of individuals typically calls for lower tax rates. This paper demonstrates that this claim is overturned in a general equilibrium environment in which governments of two countries compete for mobile workers to pursue a redistributive objective by choosing a nonlinear income tax.

In contrast to Lehmann, Simula, and Trannoy (2014), who prove that mobility tends to lower marginal tax rates throughout the entire income distribution, we show that, by weakened trickle-down effects, the endogeneity of wages reverses the conventional result at the top. With migration in general equilibrium, the marginal tax rate is higher at the top (less negative) and lower at the bottom (less positive) than what would be chosen by a social planner in a closed economy, implying a more progressive tax code. As in Stiglitz (1982), the reason for marginal subsidies to (rather than taxes on) high-skilled workers is to encourage their labor supply, which reduces pre-tax wage inequality and thus becomes an effective instrument for redistribution. However, attracting internationally mobile high-skilled individuals requires relaxing their incentive constraint, which in an endogenous-wage environment is obtained by taxing high-skilled workers more at the margin to boost pre-tax wage inequality.

Related literature. Our work is related to several recent contributions to the literature on optimal nonlinear income taxation. Lipatov and Weichenrieder (2015) analyze tax competition for mobile high-skilled workers between the governments of two identical countries and show that tax competition leads to lower tax payments for high-skilled workers. In contrast to our model, wage rates are assumed to be exogenous.

Stiglitz (1982) initiated the debate on the optimal taxation with endogenous wages in a two-type setting. Generalizing Stiglitz (1982) to a continuum of types, Sachs, Tsyvinski, and Werquin (2020) consider reforms of arbitrarily nonlinear tax schedules and the optimal taxation

[1] See Kleven, Landais, Munoz, and Stantcheva (2020) for a survey of the empirical literature. An early theoretical contribution is Simula and Trannoy (2010), who analyze a government's choice of a nonlinear income tax schedule to attract high-skilled workers. Further theoretical works are discussed below.

in general equilibrium. They demonstrate that raising the progressivity of an initially progressive tax system increases government revenue more in a situation with endogenous wages than with exogenous wages. Therefore, depending on the initial tax code, it may be beneficial to raise tax progressivity, similar to our case with two competing governments. However, their setting ignores worker's migration responses at the extensive margin.

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. In Rothschild and Scheuer (2013), the government sets one tax schedule that applies to all workers regardless of their work sector. However, in our context, competing governments set tax schedules for workers who choose their place of residence, making the equilibrium workforce endogenous to the tax code.

Our framework connects the closed economy two-type 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. Skilled and unskilled workers are imperfect substitutes in producing a composite output good under constant returns to scale. As standard, governments maximize the weighted sum of their residents' utilities and observe only incomes, but not skills. In the partial equilibrium version of our economy, migration leads to a reduction in the bottom tax rate and no change in the top tax rate, nesting the finding by Lehmann et al. (2014). There is also a migration-induced decline in the bottom tax rate with endogenous wages, but the top tax rate goes down (instead of remaining unaffected). The governments' incentives to amplify pre-tax wage inequality, which are absent in partial equilibrium, drive this disparate response. However, low-skilled workers are compensated for this rise in inequality by a more pronounced decline in their marginal tax rate than in partial equilibrium.

Outline. In Section 2, we show our main result that migration tends to raise income tax progressivity in general equilibrium. We briefly discuss extensions to our framework and relate our model and findings to the empirical literature before concluding in Section 3. We relegate all proofs to the Appendix. Moreover, in the Appendix, we provide a parameterized version of our economy, leading to closed-form expressions for the optimal tax system. Finally, we study the effects of tax coordination under symmetric country sizes.

2 The Model

Economic environment. We extend the canonical model of Stiglitz (1982) to a setting where two countries (or regions, more generally) $i = A, B$ compete for internationally mobile workers. There are two unobservable types $\theta = L, H$, indicating low- and high-skilled workers. 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 $F_i\left(N_{i,L}l_{i,L}, N_{i,H}l_{i,H}\right)$ under constant returns to scale. Consequently, a worker θ's marginal product pins down her wage rate in that country

$$w_{i,\theta} = \frac{\partial F_i\left(N_{i,L}l_{i,L}, N_{i,H}l_{i,H}\right)}{\partial\left(N_{i,\theta}l_{i,\theta}\right)}, \tag{1}$$

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 $u\left(c_{i,\theta}, l_{i,\theta}\right) \equiv c_{i,\theta} - v\left(l_{i,\theta}\right)$, where $v\left(l_{i,\theta}\right)$ is an increasing and convex disutility from labor. Consumption is given by the after-tax income $c_{i,\theta} = y_{i,\theta} - T_i\left(y_{i,\theta}\right)$.[2]

Migration. As in Lehmann et al. (2014), a worker θ born in country i draws a migration cost m from a conditional density function $G_i(m|\theta) = \int_0^m g_i(x|\theta)\,dx$, 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

$$N_{i,\theta} \equiv \rho_i\left(\Delta_i|\theta\right) \equiv \begin{cases} n_{i,\theta} + G_j\left(\Delta_i|\theta\right)n_{j,\theta} & for\ \Delta_i \geq 0 \\ \left(1 - G_i\left(-\Delta_i|\theta\right)\right)n_{i,\theta} & for\ \Delta_i \leq 0 \end{cases}. \tag{2}$$

Accordingly, denote the semi-elasticity of migration as $\eta_{i,\theta} \equiv \frac{\partial\rho_i\left(\Delta_i|\theta\right)}{\partial\Delta_i}\frac{1}{N_{i,\theta}} > 0$. Suppose that high-skilled workers are more mobile than low-skilled ones $\eta_{i,H} \geq \eta_{i,L}$.[3]

---

[2] The usual monotonicity condition applies that $\frac{-u_l(c,y/w)/w}{u_c(c,y/w)}$ is decreasing in w.

[3] We obtain our main result that migration raises tax progressivity for a constant or increasing semi-elasticity of migration. For a negative slope of the migration semi-elasticity, tax progressivity may rise or decline.

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. To solve for the equilibrium of this game, define the Pareto weights $\{\psi_{i,\theta}\}_{\theta=L,H}$, where $\psi_{i,L} > 1 > \psi_{i,H}$. We assume that country i's utilitarian government wants to redistribute income from H to L (i.e., $T_i(y_{i,H}) \geq 0 \geq T_i(y_{i,L})$), and, thus, solves

$$\max_{\{c_{i,\theta},l_{i,\theta},N_{i,\theta}\}_{\theta=L,H}} \sum_{\theta=L,H} \psi_{i,\theta} u(c_{i,\theta}, l_{i,\theta}) N_{i,\theta} \tag{3}$$

$$\text{subject to} \quad u(c_{i,H}, l_{i,H}) \geq u\left(c_{i,L}, \frac{w_{i,L} l_{i,L}}{w_{i,H}}\right), \tag{4}$$

$$\sum_{\theta=L,H} N_{i,\theta} c_{i,\theta} \leq F_i(N_{i,L} l_{i,L}, N_{i,H} l_{i,H}), \tag{5}$$

as well as subject to the endogeneity of wages (Equation (1)) and the equilibrium population (Equation (2)), and taking the other country j's tax scheme as given. Equations (4) and (5) are the high-skilled worker's incentive constraint and the government budget (no public good provision, purely redistributive tax). Notice that we implicitly assume that governments do not discriminate between natives and immigrant workers in their taxation. Moreover, welfare is defined as the Pareto-weighted sum of citizens' utility. As an alternative, one may easily consider natives' welfare.

Optimal marginal tax rate at the top. Observe that in this economy, there are trickle-down forces. As a benchmark, consider the marginal tax rate chosen by an exogenous technology planner who ignores migration ($\eta_{i,\theta} = 0$). With arbitrary values of the two population groups $\{N_{i,\theta}\}_{\theta=L,H}$, the optimal marginal tax on skilled workers set by such a planner who ignores migration is negative

$$\overline{T}_i'(y_{i,H}) = -\frac{(\psi_{i,L} - \psi_{i,H}) N_{i,L}}{\psi_{i,L} N_{i,L} + \psi_{i,H} N_{i,H}} v'\left(\frac{w_{i,L} l_{i,L}}{w_{i,H}}\right) \frac{w_{i,L} l_{i,L}}{l_{i,H} w_{i,H}^2} (\gamma_{i,L,H} - \gamma_{i,H,H}) < 0.$$

This result mirrors the closed-economy finding in Stiglitz (1982). The intuition is that high-skilled workers are subsidized because their labor supply raises low-skilled workers' wages.

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 who ignores migration but takes as given the number of the two population groups $\{N_{i,\theta}\}_{\theta=L,H}$ that materialize in the open economy Nash equilibrium. This notion includes the self-confirming policy equilibrium proposed by Rothschild and Scheuer (2013, 2016), where the exogenous technology planner sets the tax scheme such that it generates outcomes for which it is optimal (i.e., $\overline{T}_i'(y_{i,\theta})|_{\overline{T}_i'}$). In Appendix C, we demonstrate that a Cobb-Douglas technology and an isoelastic disutility of labor yield, in the symmetric Nash equilibrium, closed-form expressions for $\overline{T}_i'(y_{i,\theta})$. Then, there is no endogeneity of the right-hand side variables in the underlying tax code.

Proposition 1. *In the Nash equilibrium, the optimal marginal tax of high-skilled workers is*

$$T_i'(y_{i,H}) = \Psi_i(\eta_{i,L}, \eta_{i,H}) \cdot \overline{T}_i'(y_{i,H})|_{T_i'}, \tag{6}$$

where $\Psi_i(\eta_{i,L}, \eta_{i,H}) < 1$ *for* $\eta_{i,\theta} > 0$*. Therefore, in the Nash equilibrium, the optimal marginal subsidy for high-skilled workers is lower in the open economy:* $T_i'(y_{i,H}) > \overline{T}_i'(y_{i,H})|_{T_i'}$.

Proof. See Appendix A. □

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)). The reason is that general equilibrium externalities are absent in the existing partial equilibrium models (fixed wages). These trickle-down forces provide a rationale for a lower marginal tax rate at the top relative to an economy with fixed wages. 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. In that sense, trickle-down forces are partly offset by labor migration.

To gain some intuition for this result, consider the high-skilled workers' incentive constraint (4) which binds in the optimum. Therefore, to attract high-skilled workers, the government needs to raise the utility of a high-skilled worker who mimics a low-skilled worker: $u\left(c_{i,L}, \frac{w_{i,L} l_{i,L}}{w_{i,H}}\right)$. For a given consumption-labor-bundle of the low-skilled, this expression is increasing in pre-tax wage inequality measured by the high-skilled workers' relative wage $\frac{w_{i,H}}{w_{i,L}}$. One way to boost this pre-tax wage inequality is to tax high-skilled workers more. In response, they reduce their labor supply such that low-skilled (high-skilled) workers' wages decline (go up).

Optimal marginal tax rates at the bottom. In Proposition 2, we compare the optimal marginal tax rate at the bottom without migration

$$\overline{T}_i'(y_{i,L}) = \frac{(\psi_{i,L} - \psi_{i,H})N_{i,H}}{\psi_{i,L}(N_{i,H} + N_{i,L})}\left[1 + v'\left(\frac{w_{i,L}l_{i,L}}{w_{i,H}}\right)\frac{1}{w_{i,H}}(\gamma_{i,H,L} - \gamma_{i,L,L} - 1)\right] > 0$$

to the Nash equilibrium marginal tax rate with migration. Again, we utilize for the comparison the optimal tax rate chosen by a social planner in a closed economy who takes the population distribution under the open economy Nash equilibrium as given, but mistakenly assumes no migration.

Proposition 2. *In the Nash Equilibrium, the optimal marginal tax of low-skilled workers is*

$$T_i'(y_{i,L}) = f\left[\Psi_i(\eta_{i,L}, \eta_{i,H})\right] \cdot \overline{T}_i'(y_{i,L})\,|_{T_i'}, \tag{7}$$

where $f\left[\Psi_i(\eta_{i,L}, \eta_{i,H})\right] < 1$ *for* $\eta_{i,\theta} > 0$. *Therefore, in the Nash equilibrium, the optimal marginal tax for low-skilled workers is lower in the open economy:* $T_i'(y_{i,L}) < \overline{T}_i'(y_{i,L})\,|_{T_i'}$.

Proof. See Appendix B. □

With endogenous wages, the intuition for the lower bottom tax rate in Proposition 2 is, besides the standard migration elasticity argument in partial equilibrium (see Lehmann et al. (2014)), 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. 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 that wants 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.[4]

In summary, relative to the closed economy, international migration decreases the optimal marginal tax rate of low-skilled workers but raises high-skilled workers' optimal tax rate making

[4]To see this, set wage responses equal to zero ($\gamma_{i,\theta} = 0$). 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.

the tax code *more progressive*. This adjustment in marginal tax rates augments pre-tax wage inequality. As a consequence, migration mitigates the general equilibrium trickle-down effects. Pre-tax wages and optimal marginal tax rates move closer to the partial equilibrium setting. With this more realistic economic structure in which workers are internationally mobile, general equilibrium forces become less important. Depending on the degree of international mobility ($\eta_{i,L}$ and $\eta_{i,H}$), the canonical Mirrlees (1971) model with fixed wages may, therefore, serve as a reasonable proxy, even in the presence of general equilibrium wage externalities.

Discussion. We start our discussion of the results with a note of caution. The presence of international migration opportunities makes optimal tax codes more progressive in terms of marginal tax rates. Nonetheless, the effect of migration on net-tax payments and transfers from high-skilled to low-skilled workers could be negative. 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 as labor supplies also change, thereby affecting tax payments, which are the difference between consumption and gross income.

Our setup allows us to consider the effects of tax coordination on income tax progression. 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 D, 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)).

The empirical literature estimates migration responses, for example, by top earners to the respective marginal tax rate (e.g., Kleven et al. (2020)). Therefore, it is important to note that our setup contains the well-known incentives to migrate in response to changes in the marginal tax rate despite the general equilibrium wage responses. In the example of Appendix C, we show that, for an isoelastic disutility of labor and a Cobb-Douglas technology, the standard reduced-form relationship holds: The high-skilled workers' income, which can be written as $log(y_{i,H}) = (1+\varepsilon)\, log(w_{i,H}) + \varepsilon log[1 - T_i'(y_{i,H})]$, and, thus, their consumption declines ceteris paribus in the marginal top tax rate. On the one hand, a rise in the tax rate directly reduces high-skilled workers' labor supply. On the other hand, it raises their wages in general equilibrium. The difficulty is that one needs to account for wage changes from simultaneous labor supply and migration responses of all worker types. We demonstrate that, for the given parametrization, the negative direct labor supply response outweighs the positive wage response to a tax rise. Therefore, consistent with the empirical literature, there is a negative reduced-form relationship between high-skilled workers' marginal tax rate and their income, leading to top-earner immigration in response to top tax cuts. However, in alternative settings with other production functions, this may not hold anymore. Altogether, it is essential to also account for wage responses in empirical settings when estimating migration elasticities.

Furthermore, our theoretical results shed new light on the recent empirical literature on the effects of globalization on tax progressivity 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 more tax progressivity (but less in the following years). A well-known explanation for higher tax progressivity 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 progressivity along with a rise in wage inequality. The main difference is that, in the former view, globalization directly amplifies preexisting inequities, whereas, in our framework, the policy response to international mobility induces the rise in inequality.

Extensions. Our setup allows us to consider the effects of *tax coordination* on income tax progression. 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 D, we show that, under cross-country symmetry, governments can restore the autarky solution by coordinating their income taxation. With coordination, governments internalize the cross-country externalities from international labor migration. 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)).

As in Lehmann et al. (2014), we abstract, for simplicity, from *income effects*. However, our results carry over to more general utility functions with separable labor and consumption, as in Stiglitz (1982) (see Online Appendix). Several other extensions to the framework of Stiglitz (1982) have been proposed in the literature. For example, Sachs et al. (2020) derive the optimal nonlinear taxation in general equilibrium with a continuum of types. For the sake of comparability, we stick to the discrete-type setting but, in the spirit of Ales, Kurnaz, and Sleet (2015), extend our analysis to *N worker types* (see Online Appendix for details). This allows us to speak to the migration-induced effects on middle-class workers' income tax rates in general equilibrium.

We demonstrate that Propositions 1 and 2 continue to hold: the tax system becomes in the open economy more progressive at the top and bottom of the wage distribution than in the closed economy. Compared to the closed economy, the optimal marginal tax of a middle type can be higher or lower. Specifically, we show under the assumption of a small interaction between migration and general equilibrium forces that the tax on a middle-class type is higher (lower) if and only if the tax set by the exogenous technology planner, who ignores migration responses, is locally regressive (progressive). This result suggests that the insights from Propositions 1 and 2 are robust and extend beyond the extreme types in the wage distribution.

An alternative way of considering further heterogeneity is to use our baseline model with two types but allow for *within-type heterogeneity*, as in Acemoglu and Autor (2011). Then, workers differ in terms of their migration costs, skills, and within-skill productivity, leading to overlapping income distributions of skill types. In the Online Appendix, we characterize in the spirit of Sachs et al. (2020) the effects of arbitrary tax reforms in an open economy with endogenous wages. In particular, we characterize novel effects in terms of empirically observable sufficient statistics by delineating the aggregate wage effect on migration on the one hand and the aggregate migration effect on wages on the other hand.

One may reinterpret the workers' extensive migration margin in our model, more generally, as a participation margin, as in Saez (2002). Therefore, our central insight should also hold with *endogenous participation* in the labor force instead of migration. There is, however, one difference. Whereas migrating workers leave a country's tax base, unemployed workers may be eligible for lump-sum transfers by the country's government.

The findings carry over to *more general production functions* (beyond constant returns to scale) as long as the own-wage elasticities are negative, $\gamma_{i,\theta,\theta} < 0$, and cross-wage elasticities are positive, $\gamma_{i,\theta,\theta'} > 0$, generating positive wage externalities. For example, the former assumption ignores superstar effects, as in Scheuer and Werning (2017). The latter rules out negative externalities, for instance, from rent-seeking, as in Rothschild and Scheuer (2016).

Finally, one may also deal with *generalized social marginal welfare weights* (see Saez and Stantcheva (2016)). In principle, marginal welfare weights are endogenous to the optimal allocations and fairness concerns. More subtly, migration may shape a society's redistributive preferences (e.g., Alesina, Miano, and Stantcheva (2018)). For simplicity, we ignore such effects.

3 Conclusion

In this paper, we introduce migration into the nonlinear taxation in general equilibrium. By adding an extensive margin, we make the canonical Stiglitz (1982) model with endogenous labor supply and wages more realistic. As we have shown, contrary to conventional wisdom, migration leads to a more progressive tax code. 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 tax rates closer to those in partial equilibrium. Thus, the canonical Mirrlees (1971) model with fixed wages provides a more realistic benchmark than previously expected.

References

Acemoglu, D., and Autor, D. (2011). “Skills, tasks and technologies: Implications for employment and earnings.” *Handbook of Labor Economics*, *4*, 1043–1171.

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

Alesina, A., Miano, A., and Stantcheva, S. (2018). “Immigration and redistribution.” *National Bureau of Economic Research, Working Paper No. 24733*.

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.

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

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.

Lipatov, V., and Weichenrieder, A. (2015). “Welfare and labor supply implications of tax competition for mobile labor.” *Social Choice and Welfare*, *45*(2), 457–477.

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

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. (2002). “Optimal income transfer programs: intensive versus extensive labor supply responses.” *The Quarterly Journal of Economics*, *117*(3), 1039–1073.

Saez, E., and Stantcheva, S. (2016). “Generalized social marginal welfare weights for optimal tax theory.” *American Economic Review*, *106*(1), 24–45.

Scheuer, F., and Werning, I. (2017). “The taxation of superstars.” *The Quarterly Journal of Economics*, *132*(1), 211–270.

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.

A Proof of Proposition 1

The optimal tax code can be implicitly characterized by the households' first-order condition $1 - T_i'\left(w_{i,\theta} l_{i,\theta}\right) = \frac{v'\left(l_{i,\theta}\right)}{w_{i,\theta}}$. In the following, we, firstly, characterize the solution to the "inner" problem. That is, we solve for the optimal allocation $\left\{c_{i,\theta}, l_{i,\theta}\right\}_{\theta=L,H}$ for a given population $\left\{N_{i,\theta}\right\}_{\theta=L,H}$. Secondly, we maximize welfare by choosing $\left\{N_{i,\theta}\right\}_{\theta=L,H}$, which is the "outer" problem.

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

$$\begin{aligned}\mathscr{L}_i\left(N_{i,L}, N_{i,H}\right) \equiv & \sum_{\theta=L,H} \psi_{i,\theta} u\left(c_{i,\theta}, l_{i,\theta}\right) N_{i,\theta} + \mu_i \left[u\left(c_{i,H}, l_{i,H}\right) - u\left(c_{i,L}, \frac{w_{i,L} l_{i,L}}{w_{i,H}}\right)\right] \\ & + \xi_i \left[F_i\left(N_{i,L} l_{i,L}, N_{i,H} l_{i,H}\right) - \sum_{\theta=L,H} N_{i,\theta} c_{i,\theta}\right] + \sum_{\theta=L,H} \lambda_{i,\theta}\left[N_{i,\theta} - \rho_i\left(\Delta_i; \theta\right)\right]\end{aligned}$$

for a given population. Assuming that the optimization problem is convex, the first-order conditions describe the unique optimum

$$\begin{aligned}
\left[c_{i,H}\right] : 0 = & \left(\psi_{i,H} - \xi_i\right) N_{i,H} + \mu_i - \lambda_{i,H} \eta_{i,H} N_{i,H} && (8) \\
\left[l_{i,H}\right] : 0 = & -v'\left(l_{i,H}\right)\left(\psi_{i,H} N_{i,H} + \mu_i - \lambda_{i,H} \eta_{i,H} N_{i,H}\right) + \xi_i w_{i,H} N_{i,H} \\
& + \mu_i v'\left(\frac{w_{i,L} l_{i,L}}{w_{i,H}}\right) \frac{w_{i,L} l_{i,L}}{w_{i,H} l_{i,H}}\left(\gamma_{i,L,H} - \gamma_{i,H,H}\right) && (9) \\
\left[c_{i,L}\right] : 0 = & \left(\psi_{i,L} - \xi_i\right) N_{i,L} - \mu_i - \lambda_{i,L} \eta_{i,L} N_{i,L} && (10) \\
\left[l_{i,L}\right] : 0 = & -v'\left(l_{i,L}\right)\left[\psi_{i,L} N_{i,L} - \mu_i \frac{v'\left(\frac{w_{i,L} l_{i,L}}{w_{i,H}}\right) w_{i,L}}{v'\left(l_{i,L}\right) w_{i,H}} - \lambda_{i,L} \eta_{i,L} N_{i,L}\right] + \xi_i w_{i,L} N_{i,L} \\
& + \mu_i v'\left(\frac{w_{i,L} l_{i,L}}{w_{i,H}}\right) \frac{w_{i,L}}{w_{i,H}}\left(\gamma_{i,L,L} - \gamma_{i,H,L}\right), && (11)
\end{aligned}$$

where we use the definitions of wages, wage elasticities, and migration semi-elasticities. Inserting Equation (8) into (9) and making use of the high-skilled worker's first-order condition, the marginal tax rate of high-skilled workers can be written as

$$T_i'\left(y_{i,H}\right) = -\frac{\mu_i}{\xi_i} v'\left(\frac{w_{i,L} l_{i,L}}{w_{i,H}}\right) \frac{w_{i,L} l_{i,L}}{N_{i,H} l_{i,H} w_{i,H}^2}\left(\gamma_{i,L,H} - \gamma_{i,H,H}\right).$$

This expression depends on the Lagrange multipliers $\frac{\mu_i}{\xi_i}$. Add Equations (8) and (10) to obtain

$\xi_i = \frac{\psi_{i,L}N_{i,L}+\psi_{i,H}N_{i,H}-\lambda_{i,L}\eta_{i,L}N_{i,L}-\lambda_{i,H}\eta_{i,H}N_{i,H}}{N_{i,L}+N_{i,H}}$ and plug this expression back into Equation (8). Divide the resulting expression for μ_i by the expression for ξ_i to get $\frac{\mu_i}{\xi_i} = \frac{(\psi_{i,L}-\psi_{i,H}+\lambda_{i,H}\eta_{i,H}-\lambda_{i,L}\eta_{i,L})N_{i,L}N_{i,H}}{\psi_{i,L}N_{i,L}+\psi_{i,H}N_{i,H}-\lambda_{i,L}\eta_{i,L}N_{i,L}-\lambda_{i,H}\eta_{i,H}N_{i,H}}$. Observe that, without migration (exogenous technology planner), $\overline{\frac{\mu_i}{\xi_i}} = \frac{(\psi_{i,L}-\psi_{i,H})N_{i,L}N_{i,H}}{\psi_{i,L}N_{i,L}+\psi_{i,H}N_{i,H}}$ such that

$$\overline{T}'_i(y_{i,H}) = -\frac{(\psi_{i,L}-\psi_{i,H})N_{i,L}N_{i,H}}{\psi_{i,L}N_{i,L}+\psi_{i,H}N_{i,H}} v'\left(\frac{w_{i,L}l_{i,L}}{w_{i,H}}\right)\frac{w_{i,L}l_{i,L}}{N_{i,H}l_{i,H}w_{i,H}^2}(\gamma_{i,L,H}-\gamma_{i,H,H}).$$

One can write $\frac{\mu_i}{\xi_i} = \Psi_i(\eta_{i,L},\eta_{i,H})\cdot\overline{\frac{\mu_i}{\xi_i}}$ and, therefore, $T'_i(y_{i,H}) = \Psi_i(\eta_{i,L},\eta_{i,H})\cdot\overline{T}'_i(y_{i,H})|_{T'_i}$, where

$$\Psi_i(\eta_{i,L},\eta_{i,H}) \equiv \frac{\psi_{i,L}N_{i,L}+\psi_{i,H}N_{i,H}}{\psi_{i,L}-\psi_{i,H}}\frac{\psi_{i,L}-\psi_{i,H}+\lambda_{i,H}\eta_{i,H}-\lambda_{i,L}\eta_{i,L}}{\psi_{i,L}N_{i,L}+\psi_{i,H}N_{i,H}-\lambda_{i,L}\eta_{i,L}N_{i,L}-\lambda_{i,H}\eta_{i,H}N_{i,H}}$$

still depends on the multipliers $\lambda_{i,L}$ and $\lambda_{i,H}$. Finally, it is to show that $\Psi_i(\eta_{i,L},\eta_{i,H}) < 1$ which holds for $\psi_{i,L}\lambda_{i,H}\eta_{i,H} < \psi_{i,H}\lambda_{i,L}\eta_{i,L}$.

Outer problem. To demonstrate that $\psi_{i,H}\lambda_{i,L}\eta_{i,L} - \psi_{i,L}\lambda_{i,H}\eta_{i,H} > 0$, we derive the first-order conditions with respect to population masses

$$[N_{i,H}] : 0 = \psi_{i,H}u(c_{i,H},l_{i,H}) + \mu_i\frac{v'\left(\frac{w_{i,L}l_{i,L}}{w_{i,H}}\right)w_{i,L}l_{i,L}}{N_{i,H}w_{i,H}}(\gamma_{i,L,H}-\gamma_{i,H,H}) + \xi_i(w_{i,H}l_{i,H}-c_{i,H}) + \lambda_{i,H} \tag{12}$$

$$[N_{i,L}] : 0 = \psi_{i,L}u(c_{i,L},l_{i,L}) + \mu_i\frac{v'\left(\frac{w_{i,L}l_{i,L}}{w_{i,H}}\right)w_{i,L}l_{i,L}}{N_{i,L}w_{i,H}}(\gamma_{i,L,L}-\gamma_{i,H,L}) + \xi_i(w_{i,L}l_{i,L}-c_{i,L}) + \lambda_{i,L}, \tag{13}$$

again using the definitions of wages and wage elasticities. Multiply Equations (12) and (13) by $\psi_{i,L}\eta_{i,H}$ and $\psi_{i,H}\eta_{i,L}$, respectively, and rearrange to get

$$\psi_{i,H}\lambda_{i,L}\eta_{i,L} - \psi_{i,L}\lambda_{i,H}\eta_{i,H} = \mathscr{A}+\mathscr{B}+\mathscr{C},$$

where

$$\mathscr{A} \equiv \psi_{i,L}\psi_{i,H}[\eta_{i,H}u(c_{i,H},l_{i,H}) - \eta_{i,L}u(c_{i,L},l_{i,L})],$$

$$\mathscr{B} \equiv \mu_i v'\left(\frac{w_{i,L}l_{i,L}}{w_{i,H}}\right)\frac{w_{i,L}l_{i,L}}{w_{i,H}}\left[\frac{\psi_{i,L}\eta_{i,H}}{N_{i,H}}(\gamma_{i,L,H}-\gamma_{i,H,H}) + \frac{\psi_{i,H}\eta_{i,L}}{N_{i,L}}(\gamma_{i,H,L}-\gamma_{i,L,L})\right],$$

and

$$\mathscr{C} \equiv \xi_i[\psi_{i,L}\eta_{i,H}T_i(y_{i,H}) - \psi_{i,H}\eta_{i,L}T_i(y_{i,L})].$$

Noting that $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$ are positive since $\eta_{i,H} \geq \eta_{i,L} \geq 0$, $u(c_{i,H},l_{i,H}) > u(c_{i,L},l_{i,L})$, $\gamma_{i,\theta,\theta} < 0$, $\gamma_{i,\theta,\theta'} > 0$, and $T_i(y_{i,H}) \geq 0 \geq T_i(y_{i,L})$ concludes the proof.

B Proof of Proposition 2

Plug Equation (10) into (11), use the low-skilled worker's first-order condition, and rearrange to get the marginal tax rate of the low-skilled workers

$$T_i'(y_{i,L}) = \frac{1+v'\left(\frac{w_{i,L}l_{i,L}}{w_{i,H}}\right)\frac{1}{w_{i,H}}\left(\gamma_{i,H,L}-\gamma_{i,L,L}-1\right)}{1+\frac{\xi_i}{\mu_i}N_{i,L}}.$$

Since $\frac{\mu_i}{\xi_i} = \Psi_i(\eta_{i,L},\eta_{i,H}) \cdot \overline{\frac{\mu_i}{\xi_i}}$, one can write the marginal tax rate of low-skilled workers as $T_i'(y_{i,L}) = f\left[\Psi_i(\eta_{i,L},\eta_{i,H})\right] \cdot \overline{T}_i'(y_{i,L})\,|_{T_i'}$, where

$$\overline{T}_i'(y_{i,L})\,|_{T_i'} = \frac{(\psi_{i,L}-\psi_{i,H})N_{i,H}}{\psi_{i,L}(N_{i,H}+N_{i,L})}\left[1+v'\left(\frac{w_{i,L}l_{i,L}}{w_{i,H}}\right)\frac{1}{w_{i,H}}\left(\gamma_{i,H,L}-\gamma_{i,L,L}-1\right)\right]$$

and

$$f\left[\Psi_i(\eta_{i,L},\eta_{i,H})\right] \equiv \frac{1+\overline{\frac{\xi_i}{\mu_i}}N_{i,L}}{1+\overline{\frac{\xi_i}{\mu_i}}N_{i,L}/\Psi_i(\eta_{i,L},\eta_{i,H})} < 1$$

for $\Psi_i(\eta_{i,L},\eta_{i,H}) < 1$ as shown in Appendix A.

C 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. 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 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 disutility from labor be isoelastic $v(l) = \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}$ 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 $\overline{T}_i'(y_{i,H}) = -\frac{(\psi_{i,L}-\psi_{i,H})n_{i,L}}{\psi_{i,L}n_{i,L}+\psi_{i,H}n_{i,H}}\left(\frac{\alpha}{1-\alpha}\frac{n_{i,H}}{n_{i,L}}\right)^{1+1/\varepsilon}\frac{l_{i,H}^{1/\varepsilon}}{w_{i,H}}$. Using the workers' first-order condition and the usual normalization that $\psi_{i,L}n_{i,L}+\psi_{i,H}n_{i,H} = 1$, the marginal tax rate at the top simplifies to

$$\frac{\overline{T}_i'(y_{i,H})}{1-\overline{T}_i'(y_{i,H})} = -\left(\frac{\alpha}{1-\alpha}\right)^{1+1/\varepsilon}\left(\frac{n_{i,H}}{n_{i,L}}\right)^{1/\varepsilon}(\psi_{i,L}-\psi_{i,H})\,n_{i,H} < 0.$$

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

$$\frac{\overline{T}_i'(y_{i,L})}{1-\overline{T}_i'(y_{i,L})} = (\psi_{i,L}-\psi_{i,H})\,n_{i,H} > 0.$$

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 $(l_{i,H})^{1/\varepsilon} = w_{i,H}[1-T'(y_{i,H})]$, their income response to a cut in the top tax rate depends on a direct labor supply and an indirect wage response $\frac{dlog(y_{i,H})}{dlog[1-T'(y_{i,H})]} = (1+\varepsilon)\frac{dlog(w_{i,H})}{dlog[1-T'(y_{i,H})]} + \varepsilon$, whose overall sign is not clear a priori. To calculate the indirect wage response, first derive low-skilled workers' consumption, income, and wage response

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

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

and

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

Accordingly, high-skilled workers' wages change as follows

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

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, the relationship between high-skilled workers' income and the retention rate of the top tax rate is positive (recall $\alpha<1$)

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

D 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 social planner chooses $\left\{c_{i,\theta},l_{i,\theta},N_{i,\theta}\right\}_{\theta=L,H,i=A,B}$ to maximize

$$\sum_{i=A,B}\sum_{\theta=L,H}\psi_{i,\theta}u\left(c_{i,\theta},l_{i,\theta}\right)N_{i,\theta} \tag{14}$$

subject to the high-skilled workers' incentive constraints (Equation (4)), each country's resource constraint (Equation (5)), the endogeneity of wages (Equation (1)), and the equilibrium population (Equation (2)).

Observe that, as before, the set of constraints needs to hold at a country level.[5] Then, the Lagrangian of the outer problem reads as

$$\begin{aligned}\mathscr{L}\left(\left\{N_{i,L},N_{i,H}\right\}_{i=A,B}\right)\equiv&\sum_{i=A,B}\sum_{\theta=L,H}\psi_{i,\theta}u\left(c_{i,\theta},l_{i,\theta}\right)N_{i,\theta}+\sum_{i=A,B}\mu_i\left[u\left(c_{i,H},l_{i,H}\right)-u\left(c_{i,L},\frac{w_{i,L}l_{i,L}}{w_{i,H}}\right)\right]\\&+\sum_{i=A,B}\xi_i\left[\sum_{\theta=L,H}F_i\left(N_{i,L}l_{i,L},N_{i,H}l_{i,H}\right)-\sum_{\theta=L,H}N_{i,\theta}c_{i,\theta}\right]+\sum_{i=A,B}\sum_{\theta=L,H}\lambda_{i,\theta}\left[N_{i,\theta}-\rho_i\left(\Delta_i;\theta\right)\right],\end{aligned}$$

[5] Alternatively, one could consider a planner problem where the aggregate resource constraint (5) only has to hold worldwide, allowing governments to trade consumption levels to achieve cross-country redistribution. Although straightforward to consider, we disregard such incentives for the sake of comparability and due to their limited feasibility.

which yields the following first-order conditions

$$[c_{i,H}]: 0 = (\psi_{i,H} - \xi_i) N_{i,H} + \mu_i - \lambda_{i,H} \eta_{i,H} N_{i,H} + \lambda_{j,H} \eta_{j,H} N_{j,H} \tag{15}$$

$$\begin{aligned} [l_{i,H}]: 0 = & - v'(l_{i,H}) (\psi_{i,H} N_{i,H} + \mu_i - \lambda_{i,H} \eta_{i,H} N_{i,H} + \lambda_{j,H} \eta_{j,H} N_{j,H}) + \xi_i w_{i,H} N_{i,H} \\ & + \mu_i v' \left(\frac{w_{i,L} l_{i,L}}{w_{i,H}} \right) \frac{w_{i,L} l_{i,L}}{w_{i,H} l_{i,H}} (\gamma_{i,L,H} - \gamma_{i,H,H}) \end{aligned} \tag{16}$$

$$[c_{i,L}]: 0 = (\psi_{i,L} - \xi_i) N_{i,L} - \mu_i - \lambda_{i,L} \eta_{i,L} N_{i,L} + \lambda_{j,L} \eta_{j,L} N_{j,L} \tag{17}$$

$$\begin{aligned} [l_{i,L}]: 0 = & - v'(l_{i,L}) \left[\psi_{i,L} N_{i,L} - \mu_i \frac{v' \left(\frac{w_{i,L} l_{i,L}}{w_{i,H}} \right) w_{i,L}}{v'(l_{i,L}) w_{i,H}} - \lambda_{i,L} \eta_{i,L} N_{i,L} + \lambda_{j,L} \eta_{j,L} N_{j,L} \right] + \xi_i w_{i,L} N_{i,L} \\ & + \mu_i v' \left(\frac{w_{i,L} l_{i,L}}{w_{i,H}} \right) \frac{w_{i,L}}{w_{i,H}} (\gamma_{i,L,L} - \gamma_{i,H,L}), \end{aligned} \tag{18}$$

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

As before, plug (15) into (16) and use the workers' first-order condition to get

$$\widetilde{T}_i'(y_{i,H}) = -\frac{\widetilde{\mu_i}}{\xi_i} v' \left(\frac{w_{i,L} l_{i,L}}{w_{i,H}} \right) \frac{w_{i,L} l_{i,L}}{N_{i,H} l_{i,H} w_{i,H}^2} (\gamma_{i,L,H} - \gamma_{i,H,H}),$$

and proceed similarly using Equations (17) and (18) to obtain

$$\widetilde{T}_i'(y_{i,L}) = \frac{1 + v' \left(\frac{w_{i,L} l_{i,L}}{w_{i,H}} \right) \frac{1}{w_{i,H}} (\gamma_{i,H,L} - \gamma_{i,L,L} - 1)}{1 + \frac{\widetilde{\xi_i}}{\mu_i} N_{i,L}},$$

where $\frac{\widetilde{\xi_i}}{\mu_i}$ denotes the fraction of the Lagrangian multipliers under tax coordination. Then, add Equations (15) and (17) and plug the resulting expression for $\widetilde{\xi_i}$ back into (15) to solve for $\widetilde{\mu_i}$. Divide the two equations and observe that $\frac{\widetilde{\mu_i}}{\xi_i} = \widetilde{\Psi}_i \left(\{\eta_{i,L}, \eta_{i,H}\}_{i=A,B} \right) \frac{\overline{\mu_i}}{\xi_i}$, where

$$\begin{aligned} \widetilde{\Psi}_i \left(\{\eta_{i,L}, \eta_{i,H}\}_{i=A,B} \right) = & \frac{\psi_{i,L} N_{i,L} + \psi_{i,H} N_{i,H}}{\psi_{i,L} - \psi_{i,H}} \\ & \cdot \frac{\psi_{i,L} - \psi_{i,H} + \lambda_{i,H} \eta_{i,H} \left(1 - \frac{\lambda_{j,H} \eta_{j,H}}{\lambda_{i,H} \eta_{i,H}} \frac{N_{j,H}}{N_{i,H}} \right) - \lambda_{i,L} \eta_{i,L} \left(1 - \frac{\lambda_{j,L} \eta_{j,L}}{\lambda_{i,L} \eta_{i,L}} \frac{N_{j,L}}{N_{i,L}} \right)}{\psi_{i,L} N_{i,L} + \psi_{i,H} N_{i,H} - \lambda_{i,L} \eta_{i,L} N_{i,L} \left(1 - \frac{\lambda_{j,L} \eta_{j,L}}{\lambda_{i,L} \eta_{i,L}} \frac{N_{j,L}}{N_{i,L}} \right) - \lambda_{i,H} \eta_{i,H} N_{i,H} \left(1 - \frac{\lambda_{j,H} \eta_{j,H}}{\lambda_{i,H} \eta_{i,H}} \frac{N_{j,H}}{N_{i,H}} \right)}. \end{aligned}$$

Therefore, under cross-country symmetry, $\widetilde{\Psi}_i \left(\{\eta_{i,L}, \eta_{i,H}\}_{i=A,B} \right) = 1$ such that the coordination solution is equivalent to the autarky allocation ($\frac{\widetilde{\mu_i}}{\xi_i} = \frac{\overline{\mu_i}}{\xi_i}$).