Appendix: For Online Publication

A Proofs

A.1 Revenue-Maximizing Linear Tax Rate

Federal Government The federal government maximises the tax revenue function, that does not depend on the number of taxpayers because of the absence of tax-driven migration in the federal union. For notation purposes, I present the problem such that there are N_i individuals characterised by preferences $u^i(c,y)$, and Y denotes the sum of earnings' function over individuals with various preferences $Y = \sum_i N_i y_i$.

The first order condition with respect to goverment rax revenue $R = \sum_i N_i y_i (1-\tau)\tau = Y(1-\tau)\tau$ is given by $(d\tau/d\tau)Y - (dY/d(1-\tau))\tau = 0$. Using the definition of the labor supply elasticity,we can show that:

$$\tau^f = \frac{1}{1+e}$$

Where e is the income weighted elasiticities such that $e = \sum \frac{N_i y_i}{Y} e_i$.

Competing Government In the presence of tax competition, the number of taxpayers becomes a function of the net-of-tax rate determined in tax competition. As a result, the government tax revenue can be written $R = \sum_i N_i(1-\tau) y_i(1-\tau)\tau$. The first order condition with respect to the tax rate is $\sum_i [(d\tau/d\tau) y_i N_i - (dy_i/d(1-\tau))\tau N_i - (dN_i/d(1-\tau))\tau y_i] = 0$. Using the definition of ε_i and e_i, we obtain:

$$\tau^f = \frac{1}{1+e+\varepsilon}$$

Where is e and ε are the income weighted elasiticities such that $e = \sum_i \frac{y_i N_i}{Y} e_i$ and $\varepsilon = \sum_i \frac{y_i N_i}{Y} \varepsilon_i$. The optimal tax can be easily retrieved by studying a small deviation in the tax schedule τ. Consider an infra-marginal change in the linear tax schedule $d\tau$. The small tax deviations induces a change in the government tax revenue equal to $d\tau Y$, due to a mechanical increase in tax revenue. As pre-tax earnings are endogeneously determined by a labor-leisure trade-off, the reform causes an aggregated change in earnings $-e\frac{\tau}{1-\tau} Y d\tau$. In the presence of tax competition, individuals have an extensive margin of response to the tax change through migration. Individuals react to $d\tau$ through an additional migration effect $-\varepsilon\frac{\tau}{1-\tau} Y d\tau$, that captures mobility response to the net effect of the reform on their post-tax earnings. The total effect on tax revenue is therefore given by $dR = (1 - e\frac{\tau}{1-\tau} - \varepsilon\frac{\tau}{1-\tau}) Y d\tau$ in the competing union, and $dR = (1 - e\frac{\tau}{1-\tau}) Y d\tau$ in the federal union. Summing behavioral and mechanical effects to zero yields the inverse tax rate formula for the Laffer rate that maximizes tax revenue.

A.2 Transfer Maximizing Rate

Let's now consider the case where the government wants to maximize the amount of transfer to the poorest individuals (that is to say the government is Rawlsian). In the absence of tax competition, the population can always be normalized to one without loss of generality, and the revenue-maximizing rate corresponds to the optimal linear rate chosen by the Rawlsian government. In the presence of tax competition, individuals are able to respond to taxation trough migration, and the absolute number of taxpayers may be changed by these migration responses. Therefore, the amount that can be redistributed to individuals can be indirectly affected by the number of individuals in the country through the tax-driven migration channel. As a result, when the absolute number of taxpayers in changed by a change in the linear tax rate, the Ralwsian linear rate is no longer equivalent to the revenue-maximizing rate. The optimal linear rate of the Rawlsian government maximizes $T_0 = \frac{1}{N}\sum_i \tau y_i N_i$. Denoting $R = \sum_i \tau y_i N_i$, the first order condition with respect to τ is given by $\frac{dR \times N - dN \times R}{N^2}$. Formally, the FOC is:

$$\frac{1}{N}(\sum_i y_i N_i - \sum_i \frac{\tau}{1-\tau} e_i y_i N_i - \sum_i \frac{\tau}{1-\tau} \varepsilon_i y_i N_i) + \frac{1}{N^2}(\sum_i \frac{\tau}{1-\tau} \varepsilon_i N_i \sum_i y_i N_i) = 0 \tag{9}$$

The equation can be rewritten as

$$1 - \sum_i \frac{\tau}{1-\tau} \frac{e_i N_i y_i}{Y} - \sum_i \frac{\tau}{1-\tau} \frac{\varepsilon_i y_i N_i}{Y}) + \frac{1}{N} \sum \frac{\tau}{1-\tau} \varepsilon_i N_i = 0$$

And it follows that the linear tax rate that maximizes the amount of transfer is given by

$$\frac{\tau}{1-\tau} = \frac{1}{e + \bar{\varepsilon}}$$

Where $e = \sum \frac{N_i y_i}{Y} e_i$ and $\bar{\varepsilon} = \sum \frac{N_i y_i}{Y} \varepsilon_i - \sum_i \frac{N_i}{N} \varepsilon_i$.

A.3 Optimal Linear Tax Rate with Welfare Weights

In this section, I consider the case where the government maximizes the total welfare in a country. The welfare function is modeled as a sum of weighted utilities, where the welfare weights capture the social preferences of the government, and are exogeneously determined. In each country, there is a mass N_i of type-i individuals characterized by the same preferences and the same income y_i. In the federal symmetric union, the size of the population does not matter for the welfare maximization and can always be normalized to one without loss of generalities. Without tax-driven migration, the welfare maximized is always the welfare of individuals located in the country, as location choices are exogeneous to the coordinated tax policy. Any change in the tax rate will not affect the government welfare function through the number of individuals entering in this sum. In a free mobility union with competing countries, the maximization of total welfare for the optimal tax policy becomes less evident, as the competing government could aim to maximize the welfare of its nationals, initial residents, or may want to take into account the welfare of residents arriving after a change in the tax rate. The issues related to which individuals should be included in the social welfare functions of competing countries is normative, and beyond the scope of this paper. I discuss below two alternative welfare functions and their implications for the optimal linear tax rate set at the optimuum.

Formally, the government attributes a general welfare weight g_i to the utility of type i individuals in the economy such that the optimal linear tax rate maximizes $\sum_i N_i g_i u_i(c_i, y_i)$. Because of the envelop theorem, the government can ignore the effect of τ through y_i at the optimum. When countries compete, the density N_i is affected by tax policy. Therefore, the total welfare of the government may be affected through two channels by tax competition. The sum of weighted utilities may be changed by *(i)* the intensive change in welfare through the change in taxes paid and transfers received but also through *(ii)* the extensive change in the number of individuals that enter in the welfare maximization of the government due to migration responses to taxation. This can be simply observed by taking the first order condition of the government social welfare function with respect to $1-\tau$:

$$-\sum_i N_i g_i \frac{\partial u_i(c_i)}{\partial(1-\tau)} - \sum_i g_i u_i(c_i) \frac{\partial N_i}{\partial(1-\tau)} \tag{10}$$

The first term captures the effect of tax-driven migration on the level of consumption in the country, because mobility responses to taxation affect (1) taxes collected by the government and (2) the number of individuals in the country who have to split the amount collected by the government. The second term captures the effect of mobility responses to taxation of the number of individuals entering in the welfare function of the government. In the case where the welfare maximization is endogeneous to the population changes, the government may have the incentives to increase the number of taxpayers to increase the total welfare in the economy. Said differently, the government could have the incentive to maximize the amount of individuals entering in the country in order to maximize the total sum of welfare, rather than maximizing the amount of welfare for a given population size.

In the main specification, I ignore the second term of Equation (10) in order to avoid considerations related to the size of the population that maximizes the weighted sum of utilities. Rather, I consider a government that maximizes the welfare of a given population N_i, that is endogeneously determined in equilibrium, but that is taken as given for the welfare aggregation. With this specification, the number of individuals leaving country A only affect the total welfare in country A through the effect of their migration on the level of tax revenue and transfer for individuals residing in country A. The fact that individuals leave country A and thus the total welfare of country A does not matter, the government only cares about individuals residing in country A. Similarly, the total welfare in country A cannot be increased by the entry of new residents would increase the total welfare in country A by increasing the number of individuals entering in the sum. In spirit, this approach would consist in a government that maximizes the welfare of non-movers, taking into account the effect of movers on non movers utility through the revenue and transfer effects. Using the quasi-linearity in consumption because of the absence of income effects, it is possible to write the first order condition with respect to the linear tax rate as:

$$\sum N_i g_i (-y_i + \frac{1}{N}(Y - \sum_i \frac{\tau}{1-\tau} e_i y_i N_i - \sum_i \frac{\tau}{1-\tau} \varepsilon_i y_i N_i) + \frac{Y}{N^2} \sum_i \frac{\tau}{1-\tau} N_i \varepsilon_i) = 0 \tag{11}$$

$$\sum_i N_i g_i y_i . \sum_i N_i = \sum_i N_i g_i (Y - \sum_i \frac{\tau}{1-\tau} e_i y_i N_i - \sum_i \frac{\tau}{1-\tau} \varepsilon_i y_i N_i + Y \sum_i \frac{\tau}{1-\tau} \frac{N_i}{N} \varepsilon_i$$

$$\frac{\sum_i N_i g_i y_i . \sum_i N_i}{(\sum_i N_i Y_i) . \sum_i N_i g_i} = 1 - \frac{\tau}{1-\tau} e - \frac{\tau}{1-\tau} \bar{\varepsilon} \tag{12}$$

Denoting $\bar{g} = \frac{\sum N_i g_i y_i . N}{Y . \sum N_i g_i}$, we obtain the optimal linear tax rate formula $\tau^c = \frac{1-\bar{g}}{1-\bar{g}+e+\bar{\varepsilon}}$ where $\bar{\varepsilon}$ is a combination of the income weighted migration elasticity $\varepsilon = \sum_i \frac{N_i y_i \varepsilon_i}{Y}$ and the population-weighted migration elasticity $\varepsilon_p = \sum_i \frac{N_i \varepsilon_i}{N}$. In the case where the absolute number of taxpayers is unchanged by tax-driven migration (only the composition of the population changes), the population-weighted elasiticity is zero, and the optimal linear tax rate only depends on the standard income-weighted average mobility parameter ε, similarly than for the revenue-maximizing government. Importantly, the terms $\sum N_i g_i y_i$ and $\sum N_i g_i$ of the average welfare weight $\bar{g}$ depends on the densities that are taken as exogeneous in the government for the welfare aggregation (for instance, densities of stayers).

Endogeneous Size of the Welfare Sum Let's now explore the case where the government maximizes the sum of weighted utilities taking into account the change in the composition of the population and thus the set of individuals entering in the sum of welfare. With the envelop theorem and quasi-linearity in consumption, this is equivalent to $\sum_i N_i g_i \left[(1-\tau) y_i + \frac{\tau \sum_i N_i y_i}{\sum_i N_i}\right]$. The first order condition of the government with respect to τ would be given by $\sum_i N_i g_i \frac{\partial u_i(c_i)}{\partial \tau} + \sum_i g_i u_i(c_i) \frac{N_i}{\partial \tau}$. With these social preferences, the endogeneous change of N_i caused by tax-driven mobility will affect the total welfare through its effect on taxes and transfer as captured by Equation (11), and an additional effect through the change in the number of individuals who compose the welfare sum in equilibrium. The first order condition is more precisely given by:

$$\sum_i N_i g_i \left(-y_i + \frac{1}{N}\left[Y - \sum_i \frac{\tau}{1-\tau} e_i y_i N_i - \sum_i \frac{\tau}{1-\tau} \varepsilon_i y_i N_i\right] + \frac{Y}{N^2} \sum_i \frac{\tau}{1-\tau} N_i \varepsilon_i\right) - \sum_i g_i u_i(c_i, y_i) \frac{\partial N_i}{d(1-\tau)} = 0 \tag{13}$$

As in the previous section, the first term of Equation (13) captures the effect of tax-driven migration on welfare through its effects on residents' consumption: *(i)* a change in taxes paid *(ii)* a change in the amount of transfers received because of change in tax liabilities (*revenue effect*) and change in absolute number of transfer beneficiaries (*transfer channel*). The formula is augmented by an additional term capturing the effect of tax-driven migration of the amount of individuals who enter in the total welfare of the country. The underlying intuition is that any change in the tax rate τ causes a change in the total welfare through the amount of individuals who leave the country and somehow *"take their welfare"* with them. This effect is of magnitude $\frac{\partial N_i}{d(1-\tau)}$ that captures the magnitude of migration responses to taxation, and has a welfare cost $g_i u_i$ as any type-i individual leaving the country decreases the sum of total welfare by its consumption weighted by its corresponding welfare weight. Note that for individuals below the break even point, N_i is an increasing function of the net-of-tax rate, and therefore an increase in τ increases the total welfare by including immigrants in the welfare sum. The FOC of the government can be rewritten:

$$\sum_i N_i g_i \left(-y_i + \frac{1}{N}\left[Y - \sum_i \frac{\tau}{1-\tau} e_i y_i N_i - \sum_i \frac{\tau}{1-\tau}\varepsilon_i y_i N_i\right] + \frac{Y}{N^2}\sum_i \frac{\tau}{1-\tau} N_i \varepsilon_i\right) - \sum_i g_i u_i(c_i, y_i)\frac{N_i}{1-\tau}\varepsilon_i = 0$$

$$\sum_i N_i g_i \left(-y_i + \frac{1}{N}\left[Y - \sum_i \frac{\tau}{1-\tau} e_i y_i N_i - \sum_i \frac{\tau}{1-\tau}\varepsilon_i y_i N_i\right] + \frac{Y}{N^2}\sum_i \frac{\tau}{1-\tau} N_i \varepsilon_i\right) - \sum_i g_i \frac{N_i}{1-\tau}\varepsilon_i \left[(1-\tau)y_i + \tau Y/N\right] = 0$$

$$\sum_i N_i g_i \left(-y_i + \frac{1}{N}\left[Y - \sum_i \frac{\tau}{1-\tau} e_i y_i N_i - \sum_i \frac{\tau}{1-\tau}\varepsilon_i y_i N_i\right] + \frac{Y}{N^2}\sum_i \frac{\tau}{1-\tau} N_i \varepsilon_i\right) - \sum_i g_i N_i \varepsilon_i y_i - \sum_i g_i N_i \frac{\tau}{1-\tau}\varepsilon_i \frac{Y}{N}) = 0$$

$$\sum_i N_i g_i y_i (1+\varepsilon_i) = \sum_i N_i g_i \left(\frac{1}{N}\left[Y - \sum_i \frac{\tau}{1-\tau} e_i y_i N_i - \sum_i \frac{\tau}{1-\tau}\varepsilon_i y_i N_i\right] + \frac{Y}{N^2}\sum_i \frac{\tau}{1-\tau} N_i \varepsilon_i\right) - \sum_i N_i g_i \frac{\tau}{1-\tau}\varepsilon_i \frac{Y}{N}$$

$$\frac{\sum_i N_i g_i y_i (1+\varepsilon_i) \times N}{\sum N_i y_i} = \sum_i N_i g_i \times \left(1 - \frac{\tau}{1-\tau} e - \frac{\tau}{1-\tau}\varepsilon + \sum_i \frac{\tau}{1-\tau}\frac{N_i}{N}\varepsilon_i\right) - \frac{\tau}{1-\tau}\sum_i N_i g_i \varepsilon_i$$

$$\frac{\sum_i N_i g_i y_i (1+\varepsilon_i) \times N}{\sum N_i y_i . \sum_i N_i g_i} = 1 - \frac{\tau}{1-\tau}\left(e - \varepsilon + \varepsilon_p - \sum_i \frac{N_i g_i \varepsilon_i}{\sum_i N_i g_i}\right) \tag{14}$$

How is the optimal linear tax rate changed if the size of the welfare sum, or said differently the population that is taken into account in the total welfare of one country, is changed by taxation? There is a first change through the average welfare weight parameter, that is now affected by the migration elasticity. The welfare weight of individuals is augmented (or lowered) by the strength of their mobility elasticity. The optimal tax rate is also directly affected by an additional term $\varepsilon_w = \sum_i \frac{N_i g_i \varepsilon_i}{\sum_i N_i g_i}$ that is a *welfare-weighted average mobility elasticity*, capturing the effect of tax-driven mobility on total welfare through its effect on the number of individuals included in the welfare definition. For individuals with a negative mobility elasticity, this term is positive, meaning that the government has incentives to *increase* the linear tax rate in order to attract bottom earners and to capture their additional welfare. To summarize, when the government maximizes the total welfare in the country by also maximizing the amount of individuals included in the computation of the welfare, there are three main effects of tax-driven migration on the optimal linear tax rate. First, tax-driven migration changes the revenue collected in equilibrium through the *revenue channel* that is captured by the *income-weighted* parameter ε. Second, tax driven migration changes the amount that can be redistributed to everyone remaining in the country through the *transfer channel* that is captured by the *population weighted* parameter ε_p. These effects are the one affecting welfare through residents' consumption, as in Equation (12). Third, tax-driven migration changes the total welfare through the changed number of individuals included in the welfare aggregation in equilibrium. This *size channel* affects the optimal linear tax rate through the *welfare weighted* parameter ε_w that captures the amount of welfare that can be attracted or loss due to the absolute change in the number of individuals that enter in the government sum of weighted utilities.

B Non Linear Tax Schedule with Discrete Income Brackets

To take into how the tax progressivity of tax systems may be affected by the presence of tax competition, I develop a discrete version of the Mirrlees model, following Piketty (1997) and Saez (2002), in the presence of migration. The discrete non linear case is more tractable than the non linear model with continuous types. It also allows to focus on marginal tax rates in the top and bottom brackets. Compared to the linear analysis developed before, it captures how the progressivity of the tax system can be affected by the magnitude and the distribution of migration elasticities.

B.1 Baseline Framework

Agents are indexed by k and are endowed with continuously distributed skills w_k, but there is a finite numbers of tax brackets, or occupations, $i = 0, ..., I$. Each tax bracket provides a wage y_i for $i = 0, .., I$, with $y_0 = 0$. Earnings y_i are increasing with i, and I assume perfect substitution of labor types in the production function, implying that pre-tax wages are fixed. The government cannot directly observe individuals' skills and has to condition taxation on the observable income levels. The tax function is non-linear, and depends on the level of earnings, such that individuals in bracket-i have an overall tax liability $T(y_i) = T_i$.

Individuals have a utility function $u^k(c_i, k(i))$ that is a function of their tax-bracket choice $k(i)$ and the after-tax income level in their bracket c_i. Given their abilities, preferences and the tax and transfer schedule $(c_0, ..., c_I)$, agents choose their bracket i in order to maximize their utility. There is a population h_i of agents in each bracket $i \in I + 1$ and $\sum_i h_i = N$ is the total population in the country. As in the linear model, agents respond to distortions created by taxation through labor supply changes, that are captured by their choices of income brackets. Labor supply decisions of individuals are thus loaded in the function $h_i(c_0, c_1 ..., c_I)$. I assume that the tastes for work embodied in the individual utilities are smoothly distributed so that the aggregate functions h_i are differentiable. Like before, I consider the case with no income effects. This means that increasing all after-tax consumption levels by a constant amount does not affect the distribution of individuals across brackets.

In this model, $T(y_i)$ embeds all taxes and transfers received by each individual. Compared to the simplistic linear case studied before, it allows to explore at a finer level variations in the profile of transfers depending on the level of revenues. The non-linear tax system is characterized by two key concepts. First, the universal demogrant $-T(0) = T_0$ that is distributed to everyone. Second, the marginal tax rate $T'(y_i)$ that captures the taxation on transitions from one bracket to another. A negative value for $T_i(y_i)$ means that individuals receive a net transfer from the government, and has to be distinguished from a negative value for the taxation rate on occupations transitions defined by $(T_i - T_{i-1})/(c_i - c_{i-1})$. There is an i for which $T_i = 0$, and as in the linear case I call this income level the *break-even point*. An important feature of the optimal tax problem is that it does not produce an explicit formula for the optimal transfer -T(0). The guaranteed income is determined in general equilibrium, and results from the optimal tax and transfer schedule T_i, and the empirical densities h_i determined by the tax schedule. The amount of taxes and transfers in each bracket T_i are set by the government in order to maximize a total welfare function. The government budget constraint is a function of the tax schedule and the endogeneously determined density of individuals in each income bracket:

$$R = \sum_{i=0}^{I} h_i T_i \tag{15}$$

I extend the model to the case where individuals can react to taxation with migration. In the presence of tax competition, taxation affects individuals' choices at the intensive margin regarding their choice of income bracket, and at the extensive margin through their location choices. Conditional of being in the bracket i, individuals choose to migrate from A to country B if their utility is higher in country B. The migration condition considered in the linear case is unchanged, except that the tax system is now non-linear, and T_i directly loads the total tax liability of type-i individual:

$$u_i^A = y_i - T_i^A - v_i(y_i, l_i) + \theta_i^A - m \tag{16}$$

$$u_i(c_i^A, y_i, \theta_i^A, m) \geq u_i(c_i^B, y_i, \theta_i^B, m) \tag{17}$$

As before, the migration condition establishes that location choices are driven by differences in tax liabilities between the two countries, and the density of individuals in one country is therefore a function of its tax liability in this country. The migration responses to taxation can be summarized in terms of elasticity concepts. For simplicity, I define the migration elasticity as the change in the density of type-i individuals locating in country A when their *disposable income* in country A is increased by one percent:

$$\xi_i = \frac{\partial h_i}{\partial c_i} \times \frac{c_i}{h_i} \tag{18}$$

Note that ξ_i is equal to ε_i (defined with respect to the average tax rate $\bar{\tau} = T_i/y_i$) for individuals with earnings above the break-even point, and of opposite sign of ε_i for individuals with income levels below the break-even point. This is because $c_i = y_i(1-\bar{\tau})$ and migration responses are affected by the average not the marginal tax rate as already discussed in the main text. Defining the migration elasticity with respect to consumption in the non-linear case allows to work with a positive parameter at all income levels.

B.2 Intensive Model

In this model, a change in consumption level in any bracket i relative to another bracket $i-1$ induces individuals to switch from bracket i to bracket $i-1$ (Piketty, 1997; Saez, 2002). For simplicity, I assume that agents can only choose between adjacent occupations, and therefore, h_i is only a function of c_i, c_{i+1} and c_{i-1}. I define the elasticity of the number of individuals in bracket i with respect to the differences in consumption $c_i - c_{i-1}$

$$\eta_i = \frac{\partial h_i}{\partial(c_i - c_{i-1})} \times \frac{(c_i - c_{i-1})}{h_i} \tag{19}$$

Where η_i captures the transition of individuals to bracket $i-1$ to bracket i when the difference in consumption between the two brackets is increased. The parameter η_i captures the participation of each individual in bracket i, and can be easily linked to the earnings elasticity e_i. Following

Saez (2002), I use the relationship $\eta_i y_i = e_i(y_{i-1} - y_i)$. Hence, with intensive responses at the labor supply margin, a change in the tax liability in the bracket i will affect the transition rate between the bracket i and the adjacent occupations. The maximization of the government tax revenue leads to the first order condition:

$$h_i = T_{i-1}\frac{\partial h_{i-1}}{\partial(c_i - c_{i-1})} - T_{i+1}\frac{\partial h_{i+1}}{\partial(c_{i+1} - c_i)} + T_i\frac{\partial h_i}{\partial(c_i - c_{i-1})} - T_i\frac{\partial h_i}{\partial(c_{i+1} - c_i)}$$

The optimal tax liability of the revenue-maximizing government is given by:

$$\frac{T_i - T_{i-1}}{c_i - c_{i-1}} = \frac{h_i + h_{i+1} + ... + h_I}{h_i \eta_i} \tag{20}$$

The proof is formally derived in the Appendix B.7.1. Using τ_i the implicit marginal tax rate on bracket i such that $\tau_i = (T_i - T_{i-1})/(Y_i - Y_{i-1})$, where $1 - \tau_i = c_i - c_{i-1}/Y_i - Y_{i-1}$, and $a_i = Y_i/(Y_i - Y_{i-1})$, we obtain the formula for the optimal marginal tax rate on bracket i in the case where individuals can only respond to taxation through labor supply choices. This corresponds to the case of a federal government composed by symetric countries. When countries set the same tax and transfer schedule and are symetric such that before-tax salaries are equal, there is no differences in consumption between home and abroad that is affected by taxation. Therefore, migration decisions are independent from T_i, and do not affect the optimal tax formula.

Proposition 3. *Optimal Marginal Tax Rate of the Revenue-Maximizing Federal Government:*

$$\tau_i^f = \frac{h_i + h_{i+1} + ... + h_I}{h_i + h_{i+1} + ... + h_I + h_i a_i e_i} \tag{21}$$

Proof. The proof is formally derived in the Appendix. □

As outlined by Saez (2002), in the absence of extensive margin responses to taxation, the optimal tax liabilities are always increasing with i, and negative marginal tax rates are therefore never optimal.[6] As a result, the marginal tax rate in the first bracket is very high, and is maximal in the Rawlsian case with high redistributive taste. In complement to the formal maximization of the government problem given in the Appendix, it is possible to provide a simple and intuitive proof of Equation 21 by studying a small deviation in the tax schedule. Consider a small change dT for all brackets $i, i+1, ..., I$. This change in taxation changes $c_i - c_{i-1}$, leaving all other differences in consumption levels unchanged. This change in tax liabilities induces a mechanical increase in collected revenue equal to $(h_i + h_{i+1} + ... + h_I)dT$. The change in taxation also induces a behavioral response through the change in transition from bracket i to $i-1$. Using the definition of the participation elasticity, the mass of taxpayers in bracket i changes by $dh_i = -h_i \eta_i dT/(c_i - c_{i-1})$, inducing a loss in tax revenue of $dhi(T_i - T_{i-1})$. Summing the behavioral and mechanical effects to zero, we retrieve the formula for the optimal tax formula in the pure intensive model.

[6] The fact that negative marginal tax rates are never optimal would plausibly hold even considering participation margin at the bottom of the income distribution in that model. This is because the government is Ralwsian, which implies that the underlying welfare weights are very high for unemployed, and lower for poor workers.

B.3 Extensive Model

I now turn to the extension of the canonical model, allowing migration responses to taxation. To emphasize how the tax driven mobility affects the non linear tax schedule, I start by considering the pure extensive model, where individuals can only respond to taxation through migration. This means that given their location decision, individuals' earnings are fixed. The only effect of T_i on h_i is therefore through migration responses to taxation summarized by Equation (16). The revenue-maximizing government chooses the optimal T_i taking into account the endogeneous changes in h_i due to migration responses to taxation, and the optimal tax and transfer schedule for type-i individuals satisfies:

$$\frac{T_i}{y_i - T_i} = \frac{1}{\xi_i} \tag{22}$$

The proof of Equation Equation (22) is derived in the Appendix. I give a simple intuition of the formula studying a small deviation in the tax schedule when individuals respond to taxation through migration only. In the pure extensive model, the change in tax liability T_i has an effect on migration decisions of individuals in bracket i, but does not affect transition to adjacent brackets. The change in T_i produces a mechanical increase in collected tax revenue $h_i dT_i$. The reform also creates a behavioral effect through migration, with a change in taxpayers mass of $dh_i = -h_i dT_i / c_i \xi_i$. Each individual emigrating from the country induces a loss of its overall tax liability T_i, and the overall behavioral effect is therefore $-h_i dT_i / c_i \xi_i T_i$. Summing behavioral and mechanical effects to zero gives the formula for the optimal tax and transfer schedule for each income bracket i. The marginal tax rate from bracket i to bracket $i-1$ is therefore given by

$$\frac{T_i - T_{i-1}}{c_i - c_{i-1}} = \frac{1}{ci - c_{i-1}}\left(\frac{y_i}{1+\xi_i} - \frac{y_{i-1}}{1+\xi_{i-1}}\right)$$

B.4 Optimal Linear Tax Rate in Tax Competition

I finally put together the pure intensive and extensive model to consider the case where individuals respond to taxation through migration and labor supply behavioral reponses. With a sligh abuse of notation, I rewrite the population function of taxpayers in the country as $h_i(c_i - c_{i-1}, c_{i+1} - c_i, c_i)$. The first two terms capture the effect of taxation on transition to adjacent tax brackets, while the last term captures the effect of taxation on utility differentials between home and abroad through consumption at home c_i. The derivation of the government tax revenue with respect to T_i is given by:

$$h_i = (T_i - T_{i-1}) \frac{\partial h_i}{\partial (c_i - c_{i-1})} + (T_{i+1} - T_i) \frac{\partial h_i}{\partial (c_{i+1} - c_i)} + T_i \frac{\partial h_i}{\partial c_i}$$

Making use of the set of first order condition and the fact that $\frac{\partial h_i}{\partial (c_i - c_{i-1})} = \frac{-\partial h_{i-1}}{\partial (c_i - c_{i-1})}$, I obtain an expression for the optimal non linear tax rate chosen by a revenue-maximizing government in the competition union.

Proposition 4. *Optimal Revenue-Maximizing Marginal Tax Rate in Tax Competition:*

$$\tau_i^c = \frac{h_i(1-b_i\xi_i) + h_{i+1}(1-b_{i+1}\xi_{i+1}) + ... + h_I(1-b_I\xi I)}{h_i(1-b_i\xi_i) + h_{i+1}(1-b_{i+1}\xi_{i+1}) + ... + h_I(1-b_I\xi_I) + h_i a_i e_i} \tag{23}$$

Proof. The optimal tax rate formulas are formally derived in the Appendix. □

The formula for the optimal non linear tax rate in tax competition can also be retrieved by using a small deviation in the tax schedule. I consider a small deviation dT for all tax bracket $i, i+1, ..I$. As in the pure intensive model, this change in taxation modifies $c_i - c_{i-1}$, leaving all other differences in consumption levels unchanged. The change in tax liabilities induces a mechanical increase in collected revenue equal to $(h_i + h_{i+1} + ... + h_I)dT$. The change in taxation also induces a behavioral response through the change in transition from bracket i to $i-1$. Using the definition of the participation elasticity, the mass of taxpayers in bracket i changes by $dh_i = -h_i\eta_i dT/(c_i - c_{i-1})$, inducing a loss in tax revenue of $dhi(T_i - T_{i-1})$. In the presence of tax competition, there is an additional effect on tax revenue due to migration responses to taxation. As migration decisions are driven by overall tax liabilities, the change dT creates a migration response in all brackets affected by the change. Using the definition of the migration elasticity, the change in the number of taxpayers due to the tax reform can be written $-dT(\frac{h_i}{c_i}\xi_i + \frac{h_{i+1}}{c_{i+1}}\xi_{i+1} ... + \frac{h_I}{c_I}\xi_I)$. Any individual migrating from the country imposes a loss in tax revenue equal to its overall tax liability, such that the effect of tax-driven migration on tax revenue is equal to $-dT(\frac{h_i}{c_i}T_i\xi_i + \frac{h_{i+1}}{c_{i+1}}T_{i+1}\xi_{i+1} ... + \frac{h_I}{c_I}T_I\xi_I)$. Note that by contrast to the linear case, the *revenue and transfer channels* are simultaneously captured by the overall tax liability T_i, that can either be positive or negative. Summing the behavioral and mechanical effects to zero yields the optimal tax formula.

Letting one of the two elasticities e_i and ξ_i tend to zero in Equation (23), we retrieve the optimal formula of the pure extensive and intensive models. The optimal tax formulas make use of two fundamental parameters that relate to the two distortions implied by the intensive and extensive behavioral responses to taxation. The labor supply responses is weighted by the discrete equivalent of the usual Pareto parameter $a_i = Y_i/(Y_i - Y_{i-1})$ that captures the relative gain from the income bracket transition taxed at the marginal rate τ_i, while the migration response is weighted by the parameter $b_i = T_i/(y_i - T_i)$, emphasizing how location choices are driven by average tax rates. The migration wedge b_i is negative for individuals with income level below the break-even point, leading the overall migration response $b_i\xi_i$ of bottom earners to be negatively weighted in the optimal tax formula.

The optimal marginal rates in the discrete linear case are, as it is well known, U-curved, with high and decreasing marginal rates at the bottom of the distribution, and increasing marginal rate at the top of the distribution. Of particular interests are the optimal top and bottom marginal tax rates. The optimal formulas in Equation (23) relate to the extreme case of a revenue-maximizing government, where the social planner exclusively values redistribution towards zero earners. As emphasized by Piketty and Saez (2013), this specific case of social preferences is likely to generate high values for the optimal bottom marginal tax rate τ_1. This is because increasing the transfers by increasing the phase-out rate produces a moderate behavioral cost, as individuals who decide to leave the labor force would have had low earnings should they work. As a result, in the presence of extensive and intensive responses to taxation, the optimal phase-out rate chosen by the Ralwsian government is positive, and very high.

As emphasized by Proposition 4, tax competition modifies the optimal tax and transfers schedule compared to the federal system summarized by Proposition 3, to an amount that is proportional to the migration tax wedges b_i that are negative for low income levels, and the mobility elasticities ξ_i. In the non-linear case, the effect of migration on the universal demogrant through the amount of taxes paid and the amount of transfer received is directly loaded in the term T_i that captures the net amount of taxes and transfers received or paid by the individual in bracket i. To simply illustrate this fact, and link the migration wedge b_i to the trade-off between the income and population weighted mobility parameter faced by the government in the linear case, I present the derivation of the optimal non-linear tax rate in the presence of migration by making the distinction between the amount of taxes paid $\tilde{T}_i$ and the amount of transfer received by everyone T_0. The resulting optimal linear tax rate is the same than the one presented in Proposition 3 with alternative notations. In this case, the government seeks to maximize $\frac{1}{\sum_i h_i}\sum_i h_i\tilde{T}_i$, where the total number of taxpayers $N=\sum_i h_i$ is endogeneously affected by the tax schedule through migration responses to taxation. With $N=\sum_i h_i$ and $R=\sum_i h_i\tilde{T}_i$, the first order condition of the government with respect to the amount of taxes $\tilde{T}_i$ can be rewritten as $\frac{1}{N^2}(\frac{\partial R}{\partial \tilde{T}_i}\times N-\frac{\partial N}{\partial \tilde{T}_i}\times R)=0$ which is equivalent to $\frac{\partial R}{\partial \tilde{T}_i}=\frac{\partial N}{\partial \tilde{T}_i}\times\frac{R}{N}$. Intuitively, the government first order condition with respect to $\tilde{T}_i$ indicates that at the optimum, the change in tax revenue due to a distortion in type-i individuals tax liability has to be offset by the change in transfer caused by the change in the number of type-i taxpayers implied by the tax reform, such that the net effect of the reform is equal to zero in the optimum. As before, I consider a small change $d\tilde{T}$ on tax brackets $i,i+1...,I$. The change $d\tilde{T}$ causes a change in the density of taxpayers in all brackets affected by the tax reform that is equal to $d\tilde{T}(-h_i\frac{\xi_i}{c_i}-h_{i+1}\frac{\xi_{i+1}}{c_{i+1}}-...-h_I\frac{\xi_I}{c_I})$. Each migration response to taxation induces a fiscal lost equal to individual's overall tax liability. When the absolute number of taxpayers is changed by the small tax deviation, there is an additional effect of migration on the government tax revenue, through the change in the number of transfers' benificiaries, and each individual emigration yields a fiscal gain equal to the universal demogrant $T_0=-T(0)$. It follows that the net effect of migration responses to taxation of $d\tilde{T}$ is equal to $-d\tilde{T}(h_i\frac{\xi_i}{c_i}(\tilde{T}_i-T_0)+h_{i+1}\frac{\xi_{i+1}}{c_{i+1}}(\tilde{T_{i+1}}-T_0)+...+h_I\frac{\xi_I}{c_I}(\tilde{T}_I-T_0))$. The overal effect of $d\tilde{T}$ on the universal demogrant is therefore $d\tilde{T}(h_i+h_{i+1}+...+h_I-\frac{T_i-T_{i-1}}{c_i-c_{i-1}}h_i\eta_i-\xi_i\frac{\tilde{T}_i-T_0}{c_i}h_i-\xi_{i+1}\frac{\tilde{T_{i+1}}-T_0}{c_{i+1}}h_{i+1}-...-\xi_I\frac{\tilde{T}_I-T_0}{c_I}h_I)$. The mobility response of individuals in each bracket-i is weighted by $b_i=(\tilde{T}_i-T_0)/c_i=T_i/c_i$, meaning that the trade off between the revenue and the transfer channel of tax driven migration responses emphasized in the linear case is captured by the migration wedge b_i in the non linear case.

B.5 Welfare Weights

I finally turn to the derivation of the optimal non linear tax tax and transfers schedule relying on the more general concept of generealized social marginal welfare weights. These formulas will be used in the numerical simulations, as they allow to capture the effects of government's tastes for redistribution on the optimal tax rates, and ultimately on the welfare effects of tax competition. As I will show later, the formulas of the optimal non linear tax rate with discrete earnings and welfare weights allow to emphasize simply the effects of tax-driven mobility on redistribution. I use the concept of generalized marginal social welfare weights, to be fully consistent with the approach presented for the linear framework. As before, the government attributes a weight g_i to type-i individuals' consumption, and the optimal tax schedule is such that any small tax deviation is welfare neutral.

Proposition 5. *Optimal Non-Linear Marginal Tax Rates in Federal Union:*

$$\tau_i^f = \frac{h_i(1-\bar{g}_i)+h_{i+1}(1-\bar{g}_{i+1})+...+h_I(1-\bar{g}_I)}{h_i(1-\bar{g}_i)+h_{i+1}(1-\bar{g}_{i+1})+...+h_I(1-\bar{g}_I)+h_ia_ie_i} \tag{24}$$

Proof. With $\bar{g}_i = g_i/(\sum_{m=0}^{I} h_m g_m \times N)$. The proof is derived below. □

Proposition 6. *Optimal Non-Linear Marginal Tax Rates in Tax Competition:*

$$\tau_i^c = \frac{h_i(1-b_i\xi_i-\bar{g}_i)+h_{i+1}(1-b_{i+1}\xi_{i+1}-\bar{g}_{i+1})+...+h_I(1-b_I\xi_I-\bar{g}_I)}{h_i(1-b_i\xi_i-\bar{g}_i)+h_{i+1}(1-b_{i+1}\xi_{i+1}-\bar{g}_{i+1})+...+h_I(1-b_I\xi_I-\bar{g}_I)+h_ia_ie_i} \tag{25}$$

Proof. With $\bar{g}_i = g_i/(\sum_{m=0}^{I} h_m g_m \times N)$. The proof is derived below. □

As in the linear case, I derive the optimal non linear tax rate using the small perturbation method with generalized social marginal weights, where the optimal tax schedule is such that no welfare gain can be acchieved through a small reform.[7] Consider again a small tax deviation dT on tax brackets $i, i+1..., I$. The deviation causes a mechanical increase in tax revenue $dT(h_i+h_{i+1}+...+h_I)$. In addition to the mechanical change in revenue collected due to the tax reform, the small tax deviation creates behavioral responses to taxation. In the federal union, there is no migration responses to taxation, and the only behavioral response to taxation is through labor supply responses to the tax reform. The tax reform dT modifies transition to bracket $i-1$ to bracket i, and the density of taxpayers in the bracket i is changed by the amount $dh_i = -h_i\eta_i dT/(c_i - c_{i-1})$ at a net fiscal cost $(T_i - T_{i-1})$. The total effect of the reform on the government tax revenue is therefore $dT(h_i+h_{i+1}+...+h_I - \frac{T_i - T_{i-1}}{c_i - c_{i-1}}h_i\eta_i)$. What is the effect on individuals welfare of the reform in the federal union? By definition of the general welfare weights, any increase in type-i individuals' consumption has a value g_i for the government. From bracket i to I, individuals have to pay additional taxes and the effect on their welfare is $-dT(h_ig_i+h_{i+1}g_{i+1}+...+h_Ig_I)$.The tax reform also increases the amount of transfer received by everyone such that the effect of the tax reform on total welfare through the change in tax revenue

[7]See Saez and Stantcheva (2016) for a discussion on the local derivation of the optimuum with generalized social marginal weights.

collected is given by $-\frac{dT}{N}(\sum_{m=0}^{I} h_m g_m(h_i + h_{i+1} + ... + h_I - \frac{T_i - T_{i-1}}{c_i - c_{i-1}} h_i \eta_i))$. At the optimum, the welfare effects sum to zero and the optimal marginal tax rate in the federal union is given by $\frac{\tau_i^f}{1-\tau_i^f} = \frac{h_i(1-\bar{g_i}) + h_{i+1}(1-\bar{g_{i+1}} h_{i+1})... + h_I(1-\bar{g_I})}{h_i a_i e_i}$ with $\bar{g_i} = g_i/(\sum_{m=0}^{I} h_m g_m \times N$ the normalized welfare weight, and where the population N can be normalized to one without loss in generality. Note that this formula is a discretization of the optimal non linear tax formula provided in Saez and Stantcheva (2016).[8]

I now turn to the evaluation of the optimal non linear tax rate in the case where individuals respond to taxation through migration. As before, the tax reform dT modifies transition to bracket $i-1$ to bracket i, and the density of taxpayers in the bracket i is changed by the amount $dh_i = -h_i \eta_i dT/(c_i - c_{i-1})$ at a net fiscal cost $(T_i - T_{i-1})$. Second, the change dT causes a change in the number of taxpayers in all brackets affected by the tax reform, as migration decisions are driven by overall tax liabilities. This change in the mass of taxpayers is equal to $dT(-h_i \frac{\xi_i}{c_i} - h_{i+1}\frac{\xi_{i+1}}{c_{i+1}} - ... - h_I \frac{\xi_I}{c_I})$. Each migration response induces a net fiscal cost of T_i for the government and it follows that the net effect of migration responses to the reform on the government revenue is $-dT(h_i \frac{\xi_i}{c_i} T_i + h_{i+1}\frac{\xi_{i+1}}{c_{i+1}} T_{i+1} + ... + h_I \frac{\xi_I}{c_I} T_I)$. The overal effect of dT on the government tax revenue is $dT(h_i + h_{i+1} + ... + h_I - \frac{T_i - T_{i-1}}{c_i - c_{i-1}} h_i \eta_i - \xi_i \frac{T_i}{c_i} h_i - \xi_{i+1}\frac{T_{i+1}}{c_{i+1}} h_{i+1} - ... - \xi_I \frac{T_I}{c_I} h_I)$. This increase in tax revenue is rebated lump-sum such that the small reform is budget neutral. What are the effects on welfare of the tax reform dT? The increase in taxes for individuals in brackets $i, i+1, ..., I$ generates a change in the universal demogrant for all individuals in the economy, and this welfare effect can be written $\frac{-dT}{N}(\sum_{m=0}^{I} h_m g_m(h_i + h_{i+1} + ... + h_I - \frac{T_i - T_{i-1}}{c_i - c_{i-1}} h_i \eta_i - \xi_i \frac{T_i}{c_i} h_i - \xi_{i+1}\frac{T_{i+1}}{c_{i+1}} h_{i+1} - ... - \xi_I \frac{T_I}{c_I} h_I)$. The welfare effect for individuals who have to pay the increase in taxes that is equal to $-dT(h_i g_i + h_{i+1} g_{i+1} + ... + h_I g_I)$. Summing the welfare effects to zero, we obtain the formula for the optimal non linear tax rate in tax competition presented in Equation (25).

Comparing Equation (25) with Equation (24) allows to see how mobility responses to taxation affects the implicit weights given by the government to mobile individuals. Because of b_i, the implicit welfare weight given to mobile individuals is *lowered* for individuals below the break-point, and *increased* for individuals above. The optimal tax formulas in the non-linear discrete framework therefore allow to emphasize how tax competition modifies implicitly the redistributive preferences on the government, and how these *mobility-adjusted* welfare weights may have different values, but also different distribution, because of the transfer channel of tax-driven mobility discussed in the previous section.

[8]The formula can be rewritten as $\frac{\tau_i^f}{1-\tau_i^f} = \frac{\sum_{m>i} h_m - \bar{G_i}}{h_i a_i e_i}$ where $\sum_{m>i} h_m$ denotes the mass of taxpayers with income above the income level where the small tax reform applies and $\bar{G_i} = \sum_{m>i} h_m g_m / (\sum_{m=0}^{I} h_m g_m \times N)$ is the discrete equivalent of the average welfare weight parameter in t$\bar{G}(y) = \int_{\{i:y_i>y\}} g_i d_i / (\int_i g_i d_i)$ with population normalized to one.

B.6 Effect of Tax Competition on Individuals' Welfare

I then turn to the numerical simulations of the optimal non linear marginal tax rates schedules. The numerical simulations for the non linear tax schedule are more complex, as they require to find the non linear tax schedule and the distribution of earnings that simultaneously satisfy the government first order condition. In order to compute the welfare effects of tax competition in the non linear tax schedule, I first conduct numerical simulations in order to find the optimal tax and transfer schedules in the federal and competing unions. For this purpose, I use a fixed-point algorithm such that the optimal tax formulas in Equation (25) and Equation (24) and the optimal conditions of individuals summarized by the behavioral elasticities in Equation (19) and Equation (18) are simultaneously satisfied. Then, I compute individuals' welfare taking into account the change in the taxes and transfer schedules and the changes in labor supply implied by the changes in taxes. I describe these two steps in details below.

I use a discrete grid of earnings with eight tax brackets based on the same empirical French earnings distribution used in the linear framework. I define h_i as the number of individuals whose earnings fall in the range $[y_i - (y_i - y_{i-1})/2; y_i + (y_{i+1} - y_i)/2]$. The resulting discretized distribution of earnings used for the numerical simulations is presented in Table B.II. In the discrete non linear model, the earnings grid and income levels are fixed, while intensive labor supply responses are loaded in the endogeneously determined population functions h_i. The functional forms chosen for the population functions h_i need to be consistent with the structure of behavioral elasticities defined in Equation (19) and Equation (18) and should coincide with empirical populations h_i^0 when the tax schedule is equal to the actual tax schedule. As in the linear framework, with symmetric countries, there is no tax-driven migration in equilibrum and the effect of migration on taxpayers' population and densities can be ignored. From Equation (19) it is possible to write:

$$h_i = h_i^0 \left(\frac{c_i - c_{i-1}}{c_i^0 - c_{i-1}^0} \right)^{a_i e_i} \tag{26}$$

Where $(c_{i-1}^0, c_i^0 ..., c_I^0)$ are the actual after tax schedules. The after-tax schedule used for the simulation is a very simple approximation of the real current after tax schedule, with a linear tax rate of 50 percent and a constant transfer of 5,000 euros. However, the results of the numerical simulations show very little sensitivity to the initial tax schedule used to solve the model. Using the functional form for h_i and the exogeneously chosen g_i, I find the tax and transfer schedules such that the optimal conditions of the government summarized in Proposition 5 and Proposition 6 and the behavioral responses summarized by Equation (26) are simultaneously satisfied.

With these optimal non-linear tax schedules at hand, I turn to the welfare analysis. As in the linear case, individuals' welfare is computed using the functional form described by Equation (8). The change in welfare from the federal union to the competition union is caused by the change in the optimal tax and transfer schedule and the change in labor supply that is loaded in the change of the endogeneous mass of tax payers h_i. The results of the numerical simulations are presented in Table B.IV and Figure A.II, while the shape of the optimal tax schedule is displayed in Figure A.I.

B.7 Formal Derivation of the Non Linear Optimal Tax Rates

B.7.1 Intensive Model

The Ralwsian government maximizes the tax revenue $R = \sum_{i=0}^{J} T_i h_i$, given that h_i is a function of $(c_i - c_{i-1}, c_{i+1} - c_i)$. The first order condition is given by the system of equation:

$$h_i = \sum_{j=0}^{I} \frac{-dh_j}{dT_i} T_j = \sum_{j=0}^{I} \frac{dh_j}{dc_i} T_j$$

As individuals can only choose between adjacent occupation, it is easy to rewrite the first order condition such that:

$$h_i = T_{i-1} \frac{\partial h_{i-1}}{\partial (c_i - c_{i-1})} - T_{i+1} \frac{\partial h_{i+1}}{\partial (c_{i+1} - c_i)} + T_i \frac{\partial h_i}{\partial (c_i - c_{i-1})} - T_i \frac{\partial h_i}{\partial (c_{i+1} - c_i)} \tag{27}$$

Using $\partial h_{i+1} / \partial (c_{i+1} - c_i) = -\partial h_i / \partial (c_{i+1} - c_i)$, we obtain:

$$h_i = (T_i - T_{i-1}) \frac{\partial h_i}{\partial (c_i - c_{i-1})} + (T_{i+1} - T_i) \frac{\partial h_i}{\partial (c_i - c_{i+1})}$$

Using Equation 27 for $i = i+1 ... I$ and the participation elasticity $\eta_i = \partial h_i / \partial (c_i - c_{i-1}) \times (c_i - c_{i-1}) \times h_i$ we obtain:

$$\frac{T_i - T_{i-1}}{c_i - c_{i-1}} = \frac{h_i + h_{i+1} + ... + h_I}{h_i \eta_i}$$

To express the optimal tax schedule as a function of the standard labor supply elasticity e_i, I follow Saez (2002) and use $(y_i - y_{i-1})\eta_i = e_i y_i$, that yields to $\eta_i = e_i y_i / (y_i - y_{i-1})$. Using τ_i the marginal tax rate on bracket i such that $\tau_i = (T_i - T_{i-1})/(Y_i - Y_{i-1})$, where $1 - \tau_i = c_i - c_{i-1}/Y_i - Y_{i-1}$, we obtain the formula for the optimal marginal tax rate on bracket i. As outlined by Saez (2002), in the absence of extensive margin responses to taxation, the optimal tax liabilities are always increasing with i, and negative marginal tax rates are therefore never optimal. As a result, the marginal tax rate in the first bracket is very high, and is maximal in the Rawlsian case with high redistributive taste.

B.7.2 Extensive Model

Let's consider now the case where individuals respond to taxation through migration. Conditional of being in the bracket i, individuals can choose to migrate from A to B if $u_i(c_i^B, y_i) \geq u_i(c_i^A, y_i)$. In that case, the fraction of individuals in a given tax bracket h_i is a function of the overall tax schedule in the bracket i T_i. Consider first the case where there are only extensive margin responses to taxation. In that case, the number of individuals in the tax bracket i is only a function of the overall tax liability T_i that determines migration decisions. The system of first order conditions follows:

$$h_i = \frac{\partial h_i}{\partial c_i} T_i \tag{28}$$

Making use of the migration elasticity formula:

$$\frac{T_i}{y_i - T_i} = \frac{1}{\xi_i} \tag{29}$$

B.7.3 Mixed Intensive and Extensive Model

I finally put together intensive and extensive responses to taxation. In this case, individuals in bracket i can choose adjacent occupations, but can also choose to locate abroad. The first order condition becomes:

$$h_i = (T_i - T_{i-1})\frac{\partial h_i}{\partial(c_i - c_{i-1})} + (T_{i+1} - T_i)\frac{\partial h_i}{\partial(c_{i+1} - c_i)} + T_i\frac{\partial h_i(c_i)}{\partial(c_i)} \tag{30}$$

This relationship illuminates the effects of a tax reform at the extensive and intensive margins. The two first terms of Equation 30 capture the effect of a change in the marginal rate on transitions between brackets. The last term captures the extensive margin on individuals that decide to migrate after that their overall tax liabilitie T_i has been changed. It is easy to derive the optimal top marginal tax rate from Equation 30 :

$$\frac{T_I - T_{I-1}}{c_I - c_{I-1}} = \frac{1}{a_I e_I}\left(1 - \frac{T_I}{c_I}\xi_I\right)$$

That simplifies to the following formula, with $b_I = T(y_I)/(y_I - T(y_I))$ that captures the overall wedge of labor taxation at income level y_I and $a_I = y_I/(y_I - y_{I-1})$ the local discrete pareto parameter.

$$\frac{\tau_I}{1 - \tau_I} = \frac{1 - b_I\xi_I}{a_I e_I} \tag{31}$$

For any other income level, making use of the set of first order conditions, the optimal tax formula is given by:

$$\frac{\tau_i}{1 - \tau_i} = \frac{h_i(1 - b_i\xi_i) + h_{i+1}(1 - b_{i+1}\xi_{i+1}) + ... + h_I(1 - b_I\xi_I)}{a_i h_i e_i} \tag{32}$$

If we let one of the two elasticities e_i and ξ_i tend to zero in Equation 32, we retrieve the optimal formula of pure extensive and intensive models. The model mixing labor supply and migration responses is similar to the pure intensive model, with weights on income groups replaced by $b_i\xi_i$, implicitly attributing more welfare weights to individuals who are more likely to respond to taxation through migration.

B.8 Non Linear Discrete Tax Rates with Social Marginal Welfare Weights

I consider alternatively the formal derivation of the optimal tax formulas in the case where the the government maximizes a welfare function. Individual k in the bracket i derives an utility $u^k(c_i, i)$ that is a function if his after-tax income in the bracket i c_i and of its tax bracket choice $i = 0, ..., I$. For a given tax schedule, there is a population $h_i(c_0, ..., c_I)$ of individuals who choose to be in the bracket i. I denote $k(i)$ the tax bracket choice of individual k. The government chooses the tax schedule $(T_0, ..., T_I)$ such that it maximizes the total welfare:

$$SWF = \sum_k w_k u_k$$

$$\sum h_i T_i \geq R$$

Denoting p the multiplier of the government budget constraint, the first order condition yields

$$(1-G_i)h_i = \sum_{m=0}^{I} \frac{\partial h_m}{\partial c_i} T_i \tag{33}$$

With $G_i = \dfrac{1}{ph_i} \sum_{k \in bracket} w_k G'(u^k) u_c^k(c_i, k(i))$. Similarly than before, I can derive the optimal tax schedules in the federal and competition union with these alternative social welfare weights G_i. With the assumption that intensive responses only occur between adjacent occupations, the first order condition can be rewritten in the federal union as

$$(1-G_i)h_i = (T_i - T_{i-1}) \frac{\partial h_i}{\partial (c_i - c_{i-1})} + (T_{i+1} - T_i) \frac{\partial h_i}{\partial (c_{i+1} - c_i)} \tag{34}$$

In the competition union, there is an additional behavioral response to taxation, and the number of individuals in each bracket h_i is affected through migration responses to taxation such that:

$$(1-G_i)h_i = (T_i - T_{i-1}) \frac{\partial h_i}{\partial (c_i - c_{i-1})} + (T_{i+1} - T_i) \frac{\partial h_i}{\partial (c_{i+1} - c_i)} + T_i \frac{\partial h_i(c_i)}{\partial (c_i)} \tag{35}$$

C Additional Tables and Figures

Table B.I: **Migration Elasticities of Top Ten Percent Employees**

Countries	Lower bound	Upper Bound
Belgium	.19	.27
Germany	.16	.24
France	.32	.45
Italy	.05	.07
Luxembourg	.26	.37
Poland	.12	.18
Portugal	.10	.15
Spain	.25	.34
Switzerland	.28	.41
United Kingdom	.52	.83

Notes: This Table summarizes the elasticity of the number of top ten percent employees with respect to the top net-of-tax rate estimated by Muñoz (2019) for the period 2009-2015 using individual-level data from the European labor Force Survey. The empirical strategy exploits within-country variations in top marginal tax rates coming from differences in propensities to be treated by top marginal tax rates between individuals of different earnings levels. Lower bounds are computed using the 8th decile as a control group not affected by top marginal tax rates changes. Upper bounds are compited using the 5th decile as a control group not affected by top marginal tax rates changes.

Table B.II: **Empirical Earnings Distribution Calibration**

Income level (euros)	Density Weight	Cumulative Density
0	.2	.2
5,000	.05	.25
15,000	.11	.36
30,000	.14	.5
45,000	.23	.73
65,000	.17	.9
100,000	.08	.98
200,000	.02	1

Notes: This Table shows the discretized empirical earnings distribution used for the numerical simulations of the non-linear tax and transfer schedule. The data is based on the distribution of labor factor income for France provided by the World Inequality Database. I define the density at each average income level as the density of individuals whose earnings fall in the range $[y_i - (y_i - y_{i-1})/2; y_i + (y_{i+1} - y_i)/2]$.

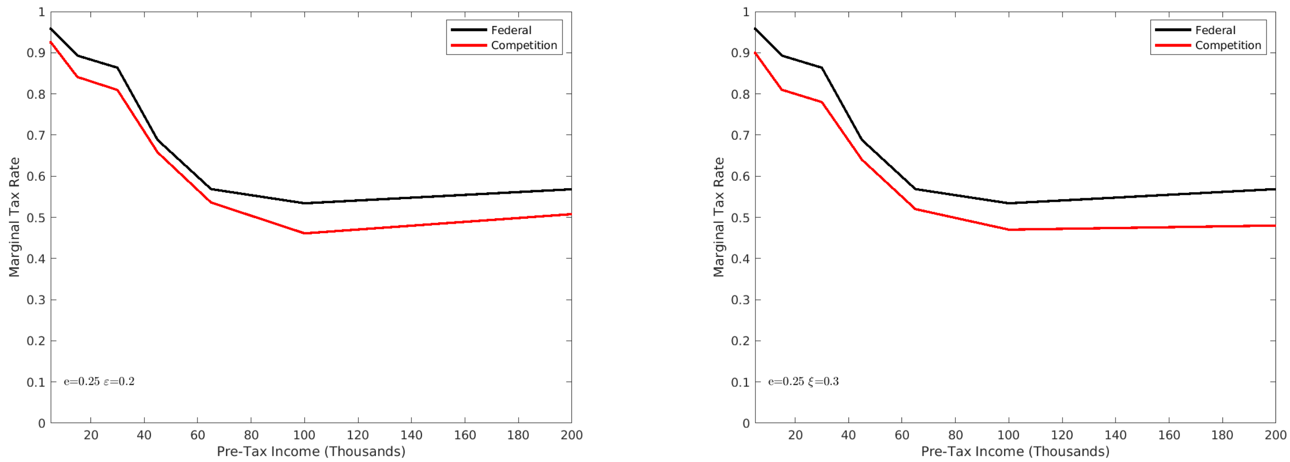
Table B.III: **Tax-Driven Migration Only at the Top With Linear Tax Schedule**

	Scenario 1 Elasticities e=0.25 $\bar{\varepsilon}$=0.02 $\varepsilon_{D_1-D_9}$=0, $\varepsilon_{D_{10}}$=0.1		Scenario 2 Elasticities e=0.25 $\bar{\varepsilon}$=0.04 $\varepsilon_{D_1-D_9}$=0, $\varepsilon_{D_{10}}$=0.2		Scenario 3 Elasticities e=0.25 $\bar{\varepsilon}$=0.06 $\varepsilon_{D_1-D_9}$=0, $\varepsilon_{D_{10}}$=0.3		Scenario 4 Elasticities e=0.25 $\bar{\varepsilon}$=0.09 $\varepsilon_{D_1-D_9}$=0, $\varepsilon_{D_{10}}$=0.4	
I- Optimal Linear Tax Rates	Federal	Competition	Federal	Competition	Federal	Competition	Federal	Competition
Rawlsian	0.73	0.71	0.73	0.70	0.73	0.68	0.73	0.67
Highly Redistributive	0.65	0.63	0.65	0.61	0.65	0.59	0.65	0.57
Mod. Redistributive	0.48	0.45	0.44	0.37	0.42	0.33	0.48	0.40
II- Welfare effect of Tax Competition (%)	Bottom 10	Bottom 50	Bottom 10	Bottom 50	Bottom 10	Bottom 50	Bottom 10	Bottom 50
Rawlsian	-.7	-0.2	-1.7	-.3	-2.6	-.6	-3.6	-.9
Highly Redistributive	-1.7	-.7	-3.5	-1.6	-5.2	-2.4	-6.8	-3.24
Mod. Redistributive	-3.2	-1.9	-6.7	-3.7	-9.7	-5.4	-12.5	-7.0

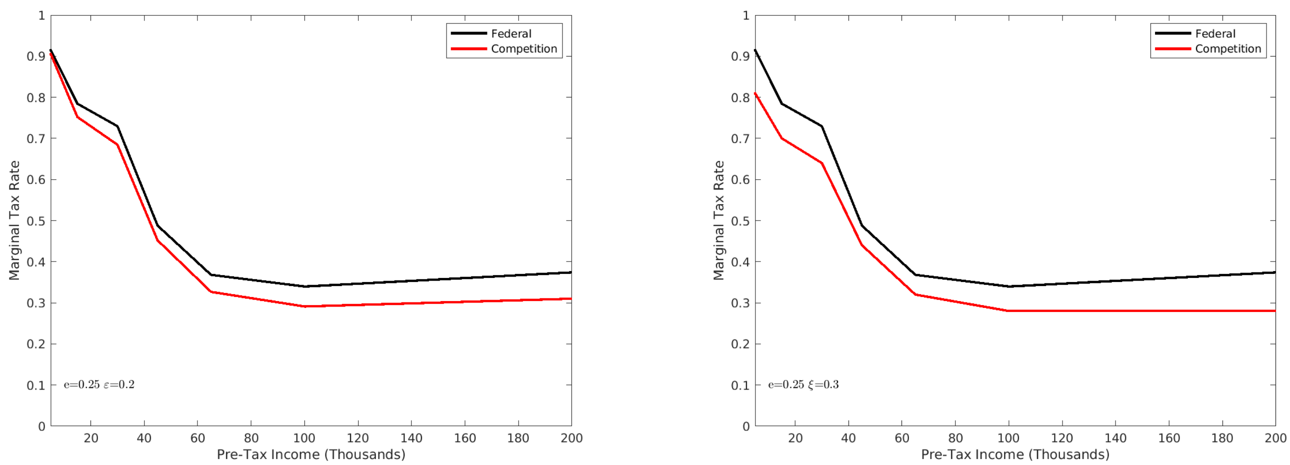
Notes: This Table shows the welfare effects of tax competition when migration responses to taxation are concentrated at the top of the income distribution (top decile) but are zero everywhere else. This implies that $\bar{\varepsilon} = \varepsilon_{D10} - \varepsilon_{p,D10}$, as $\varepsilon_i = 0$ for every individuals who are not in the top decile. The methodology for welfare computation is described extensively in the note below Table 1.

Figure A.I: **Optimal Non-Linear Tax Schedules**

Panel A. Highly Redistributive Goverment



Panel B. Moderately Redistributive Government



Notes: This Figure shows the optimal marginal tax rates schedule after the numerical simulations of Proposition 5 and Proposition 6.

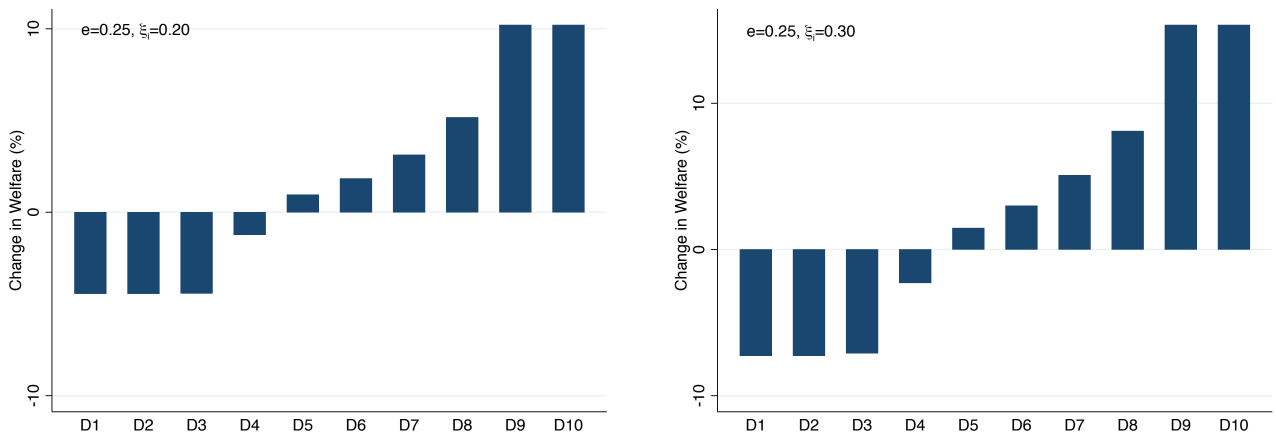
Table B.IV: **Effects of Tax Competition on Optimal Taxes and Welfare With a Non Linear Tax Schedule**

	Scenario 1 Elasticities e=0.25 ξ=0.1		Scenario 2 Elasticities e=0.25 ξ=0.2		Scenario 3 Elasticities e=0.25 ξ=0.3		Scenario 4 Elasticities e=0.25 ξ=0.4	
I- Average Marginal Tax Rates *(Marginal Tax Rate in the Top Bracket in Parentheses)*	Federal	Competition	Federal	Competition	Federal	Competition	Federal	Competition
Rawlsian	.69	.68	.69	.67	.69	.66	.69	.64
	(.64)	*(.63)*	*(.64)*	*(.61)*	*(.64)*	*(.59)*	*(.64)*	*(.55)*
Highly Redistributive	.62	.60	.62	.58	.62	.56	.62	.53
	(.57)	*(.54)*	*(.57)*	*(.51)*	*(.57)*	*(.48)*	*(.57)*	*(.45)*
Mod. Redistributive	.43	.41	.43	.39	.43	.37	.43	.35
	(.37)	*(.35)*	*(.37)*	*(.31)*	*(.37)*	*(.29)*	*(.37)*	*(.26)*
II- Welfare effect of Tax Competition (%)	Bottom 10	Bottom 50	Bottom 10	Bottom 50	Bottom 10	Bottom 50	Bottom 10	Bottom 50
Rawlsian	-.7	-.3	-.8	-.4	-2.9	-2.4	-4.8	-4.3
Highly Redistributive	-2.7	-2.1	-4.5	-3.4	-7.3	-5.9	-11.7	-9.5
Mod. Redistributive	-4.5	-3.4	-8.3	-5.0	-12.8	-9.3	-14.5	-11.5

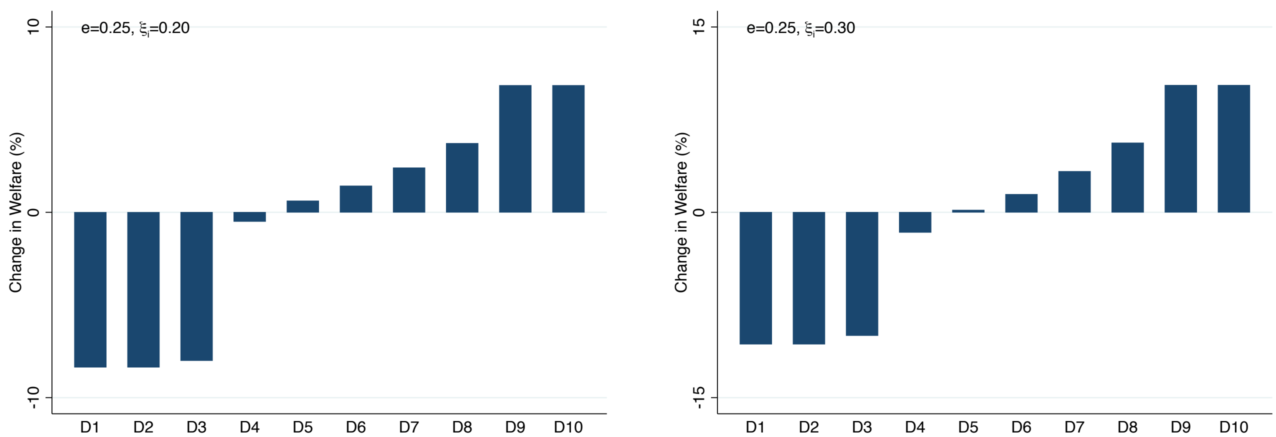
Notes: This Table summarizes the effects of tax competition on optimal tax rates and welfare. The optimal non linear tax rates are computed following the formulas presented with more details in the text and presented in Proposition 6 and Proposition 7. The numerical simulations use a discrete grid of earnings with eight income tax brackets taken from the empirical distribution of labor earnings in France and displayed in Table B.II. The elasticity of migration with respect to taxation ξ_i and the labor supply elasticity e_i are taken as constant across individuals and denoted ξ and e. The average marginal tax rates reports the average marginal tax rates across income tax brackets weighted by income in each bracket. The optimal marginal tax rates in the top income tax bracket are also reported in parentheses. The welfare is computed following Equation (8) and the endogeneous densities are determined following the functional form detailed in Equation (26). The moderately redistributive government corresponds to a government that values the welfare of individuals in the bottom fifty percent two times more than the welfare of individuals in higher earnings' deciles. The highly redistributive government values the welfare of individuals in the bottom fifty percent five times more than the welfare of individuals in higher income deciles. The welfare effect of tax competition is the variation in percentage of individuals' welfare from a federal union to a competition union. A negative welfare variation means that individuals would be better off in a federal union.

Figure A.II: **Distribution of Welfare Gains and Losses from Tax Competition with a Non-Linear Tax Schedule**

Panel A. Highly Redistributive Goverment



Panel B. Moderately Redistributive Government



Notes: This graph shows the distribution of the welfare effects of tax competition across labor earnings' deciles. The welfare effect of tax competition is the variation in percentage of individuals' welfare from a federal union to a competition union. A negative welfare variation means that individuals would be better off in a federal union. The moderatively redistributive government values the welfare of individuals in the bottom fifty percent two times more than individuals in higher income deciles. The highly redistributive government values the welfare of individuals in the bottom fifty percent five times more than individuals in the higher deciles. The tax system is non linear. The parameter ξ_i is the elasticity of migration with respect to the disposable income, and is taken as constant across earnings' deciles. See the note below Table B.IV for more details on the computation of the optimal tax rates and individuals' welfare.