Probabilistic Voting over Nonlinear Income Taxes with International Migration

Craig Brett[a]

[a] *Department of Economics, Mount Allison University, 144 Main Street, Sackville NB E4L 1A7, Canada. E-mail:* cbrett@mta.ca

July 2015; this draft May 2016

Please do not cite

Abstract. The choice of income tax policy by office-seeking politicians when workers are internationally mobile is examined. Optimal tax rules are derived, showing the separate effects of voting at the ballot box and with the feet. The resulting formulas are used to highlight when and how political considerations can attenuate (or exacerbate) the downward pressure on income tax rates typically associated with international mobility.

Journal of Economic Literature classification numbers. D72, D82, H21, H24

Keywords. income taxation, multi-dimensional Leibniz Rule, probabilistic voting

---

1. Introduction

At least as far back as Stigler (1957), conventional wisdom in economics suggests that geographical mobility of workers limits the ability of governments to redistribute income via the tax-transfer system. There has been much work in the theory of taxation aimed at testing this conventional wisdom. This literature falls into two broad categories. The first analyzes the problem of a single government whose policies affect neither the opportunities available to workers in other jurisdictions nor the tax treatment of these opportunities. This single government tries to design a tax system that reconciles its redistributive goals with two types of behavioral responses on the part of workers: standard labor supply responses and the potential movement of workers to avoid taxes and/or seek transfers. By adding to the overall elasticity of the tax base, geographical mobility tends to lead to lower optimal tax rates and less redistribution. This finding, with varying degrees of nuance, is found in Wilson (1980, 1992),Simula and Trannoy (2010), and Lehmann et al. (2014).

This paper takes a positive approach to the determination of tax schedules in the presence of worker mobility. Tax policy arises from electoral competition rather than from the optimization problem of a social planner. People vote both at the ballot box and with their feet. Specifically, a two-party probabilistic voting model with migration is developed. The parties wish to maximize the expected plurality of votes cast in their favor. They can commit to tax policies, and these polices are their election platforms. The tax polices must be feasible; that is, they must be budget-balanced and incentive-compatible. Fully nonlinear tax policies are permitted. When setting tax policy, each party is mindful that tax increases may result the loss of votes to the other party and both votes and tax base to migration.

Voter-workers differ in skill, preference for parties, and attachment to home. Differences in skill give rise to the need for distortionary taxes. Without these, political competition would be an all-out bidding war for swing voters via the use of lump-sum subsidies. Differences in political preference gives rise to the probabilistic nature of party competition. Without this feature, every voter of a particular skill level would vote for the party that offers the better tax treatment to its skill level. Differences in attachment to home imply that some members of a skill group are more apt to leave the country when faced with tax increases than are others. Anyone who leaves the country enjoys an exogenous level of utility elsewhere and does not vote. This introduces features of random participation, in the spirit of Charles Rochet and Stole (2002), and voter abstention, in the tradition of Hinich et al. (1972) into the analysis.

I show that the optimal tax strategies for a the political parties can be described by a modified Diamond (1998)–Saez (2001) ABCD rule. This rule incorporates the distortionary and demographic effects of the traditional ABC rule. It also includes a term to account for the revenue effects of tax changes due solely to migration (random participation). This term is analogous to the one found by Lehmann et al. (2014) in their analysis of a revenue-maximizing tax authority. The characterization also includes a political weighting term, which takes the place of the social welfare weighting term in the traditional ABC formula. The political weights themselves have three components. A skill-group has a larger political weight in a party's calculus: the more swing voters it has, the more likely are potential in-migrants of that type to vote for that party (or potential out-migrants to vote for its competitor), and the more tax revenue potential in-migrants of that type would bring with them. This final effect arises because the extra resources brought in have political value in that they can be used to strategically lower the tax liabilities of other types, thereby gaining even more votes.

The remainder of the paper organized as follows. The next section summarizes some of the related literature. This is followed by a description of the model. Section 4 describes how voters respond to changes in tax policy, both at the ballot box and via migration. Politicians' optimal tax strategies are described and interpreted in Section 5. Section 6 contains concluding remarks. Longer derivations and proofs are found in Appendix A and Appendix B.

2. Related Literature

Two papers very closely related to this work are those of Bierbrauer and Boyer (2015) and Chen (2000). Bierbrauer and Boyer model the equilibrium of two-party competition when nonlinear taxation can be supplemented by type-specific side payments. They show that first-best taxation is optimal when voters have no party preference but that equilibrium features distortionary taxes when voters have heterogeneous party preferences. In that case, they derive an optimal tax rule with a political weighting term. Because they have no geographic mobility and no voter abstention, their political weights are determined entirely by swing voters. Chen considers probabilistic voting over nonlinear taxes when the distribution of preferences over the political parties has a logistic form. Under this assumption, the equilibrium features parties acting as if maximizing a logarithmic social welfare function. In other words, specific assumptions about the distribution of political preferences gives rise to structured political weights. Chen's model has neither abstention nor geographic mobility.

There has been other work on the political economy of nonlinear taxes. Hamilton and Pestieau (2005) provide a majority-rule mode of taxation in the presence of worker mobility, but their analysis is confined to economies with just two types of workers. Likewise, much of the existing literature on voting over nonlinear income taxes in one-country models with variable labor supply is restricted to two-type worlds. Roemer (2012) derives the equilibrium positions of two partisan political parties who choose anonymous taxation schedules for two types of workers. Bierbrauer and Boyer (2013) consider a similar problem under the assumption that parties wish to maximize votes. They also allow for one party to be more efficient in running the affairs of government than another. Analyses of median-voter models of nonlinear income taxes with a continuum of skill types are conducted by Bohn and Stuart (2013) and Brett and Weymark (2016). They highlight how the tendency to redistribute toward the middle can give rise to wage subsidies at lower income levels.

A separate category of the literature on income taxes with international mobility develops models of income tax competition, which feature strategic interactions among the tax policies of different countries. Much of this research finds a tendency for countries to engage in a race-to-the-bottom in redistributive policy as they lower tax rates on high earners in order to attract more of them into their country and to cut welfare benefits in order to limit inflows of low earners. Among the works in this tradition are Hamilton et al. (2002), Cremer and Pestieau (2004), Piaser (2007) and Lipatov and Weichenrieder (2010). In the extreme case of perfect mobility of workers, Bierbrauer et al. (2013) find an equilibrium in which tax competition is so fierce that it results in transfers to the rich financed by the taxes of the poor.

There have been attempts to add some elements of political economy to the analysis of tax-transfer systems when workers are mobile. In a world with some costs to worker mobility, Morelli et al. (2011) derive optimal nonlinear tax schedules when redistribution is carried out by two competing regions and the optimal schedule when the tax system is centralized. They then analyze the choice between centralized and decentralized taxation as a constitutional design problem. Gordon and Cullen (2011) analyze the potential for vertical fiscal externalities between a regional government and a national government when both levels of government have redistributive motives. One of their goals is to gain insight into which order of government responsibility for redistribution should rest. The constitutional design approach to the question of optimal taxation offers an interesting mixture of positive and normative analysis. On the positive side, it takes full account of worker mobility (and labor supply) behavior. On the normative side, it begins with a specification of the government's objective function. In a world with mobile workers, this specification must include consideration of whose well-being should count: initial residents or final residents. Following Wilson (1980, 1992), most studies consider a government that is concerned with the well-being of initial residents only. Alternatives are suggested and defended by Simula and Trannoy (2011) and Bierbrauer et al. (2013). The political economy model of this paper provides a simple resolution. Politicians care about whomever votes.

3. The Model

There are two types of agents in the model economy. Politicians, who set tax policy and engage in electoral competition, and citizens, who choose how much to work, where to live and for whom to vote. There are two political parties, each wanting to maximize the expected plurality of votes cast in their favor. The parties are named A and B. These compete by announcing, and credibly committing to follow, income tax policies. The sole purpose of taxation is to raise money for redistributive purposes. This redistribution is motived by political concerns, rather than by a concern for social welfare. Party i announces a tax function $T^i(y)$, that specifies the tax liability of anyone who earns before-tax labor income y. A citizen's after-tax income is given by $x = y - T^i(y)$ whenever Party

i wins the election.

Citizens (also referred to as workers) vary along three dimensions, one for each aspect of their decision-making. As is customary in models of income taxation, each citizen has a skill level, w, that measures the productivity of a unit of that person's labor. Specifically, a worker with skill level w who choses to work l units of time has before-tax income $y = wl$. In addition, each citizen has an attachment to the home country, parameterized by β. A person gains an additional β units of utility by staying in the country. This gain could represent patriotic feelings or perhaps be the savings of the moving costs needed to set up elsewhere. Lastly, each citizen has a parameter γ that measures preference for Party A. The mere victory of Party A provides the worker with an additional utility γ. This can be interpreted as a "look shock" or the voter's gain from aspects of Party A's platform not bearing on income tax policy. A negative value of γ denotes an all-else-equal preference for Party B. The vector (γ, β, w) is a citizen's type. The space of citizen types is continuous, and their distribution has a density $f(\gamma, \beta, w)$ with support $(-\infty, \infty) \times (-\infty, \infty) \times [\underline{w}, \overline{w}]$.[1]

Preferences are represented by

$$U(x, l; \gamma, \beta, w) = \begin{cases} u(x, l; w) + \gamma + \beta, & \text{if A wins and the worker stays;} \\ u(x, l; w) + \beta, & \text{if B wins and the worker stays;} \\ 0, & \text{if the worker emigrates.} \end{cases} \tag{1}$$

In this formulation of preference, the "outside option" of emigration appears to be type-independent. But it is possible to allow the attractiveness of the outside option to vary with skill or with political preference by allowing β to depend on one or both of these parameters. One can imagine that γ is a decreasing function of w to allow for either: (i) effective moving costs decreasing with the skill level, or (ii) the higher skilled have more lucrative opportunities abroad. This suggests that it is useful to think of the case when β is negatively correlated with w.

When preferences are given by (1), the labor supply decision is separated from the voting decision and from the residential decision. All workers with skill type w who choose to remain in the country work the same amount. They may vote differently, due to differences in γ, or some might choose to leave, depending on their respective values of β. Citizens behave as if using a two-stage decision process. In the first stage, they decide on their hours of work under the tax policies proposed by the candidates. The sub-utilities (of consumption and leisure) resulting from this stage are given by

$$V^i(w) = \max_l u(wl - T^i(wl), l), \quad i = A, B. \tag{2}$$

Under mild assumptions about preferences, before-tax income is increasing in skill, so the tax function can be construed as applying to the skill level, and one can write (with a slight abuse of notation) the tax paid by workers with skills w under the tax function proposed by Party i as $T^i(w)$.

[1] The case of an unbounded distribution of skills is permitted, with $\overline{w} = \infty$.

The political parties hold Nash conjectures about the other's behavior. Thus, they hold the tax function announced by the other party as given when deciding on their policy. Consequently, Party A, say, treats $V^B(w)$ as fixed when deciding on its tax policies. Moreover, it knows that voting and migration behavior are determined by $V^A(w)$, $V^B(w)$, β, and γ alone. In light of (1) and (2), a worker will vote for Party A (and stay) if

$$\gamma \geq V^B(w) - V^A(w), \tag{3}$$

and will not emigrate in the event of Party A winning if

$$\gamma + \beta \geq -V^A(w). \tag{4}$$

Only workers who choose to stay after an election win by Party A can be taxpayers. The mass of final taxpayers with skills w following a victory for Party A is given by

$$\theta(V^A(w), w) = \int_{-\infty}^{\infty} \int_{-V^A(w)-\beta}^{\infty} f(\gamma, \beta, w) d\gamma \, d\beta. \tag{5}$$

For ease of exposition later in the paper we introduce the notations

$$\eta(w) = \frac{\partial \theta(V^A(w), w)}{\partial V^A(w)} \quad \text{and} \quad \varphi(w) = \frac{\partial \ln \theta(V^A(w), w)}{\partial \ln x^A(w)} \tag{6}$$

to denote the response of the mass of taxpayers of type w to an increase in utility offered by Party A and the elasticity of population with respect to the after-tax income offered by Party A, respectively. While these quantities also depend on $V^A(w)$, this dependence is suppressed to make the notation somewhat tidier.

The parties care only for people who reside in the country after the election. Any worker who would vote for a party and then leave if the party of their electoral choice won would also leave the country if the other party won. Such people have attachment to the country so low that they would not remain in the long run. Thus, the notion of equilibrium used here incorporates a notion that citizens have some minimal attachment to the country. With this proviso, the mass of votes cast for Party A is then the mass of workers for whom both (3) and (4) hold. This quantity is given by

$$M^A(V^A(w), w) = \int_{V^B(w)-V^A(w)}^{\infty} \int_{-V^A(w)-\gamma}^{\infty} f(\gamma, \beta, w) d\beta \, d\gamma. \tag{7}$$

The region of integration on the right-hand side of (7) is illustrated in Figure 1.

The other people who remain in the country vote for Party B. The mass of these people is given by

$$M^B(V^A(w), w) = \int_{-\infty}^{V^B(w)-V^A(w)} \int_{-V^A(w)-\gamma}^{\infty} f(\gamma, \beta, w) d\beta \, d\gamma, \tag{8}$$

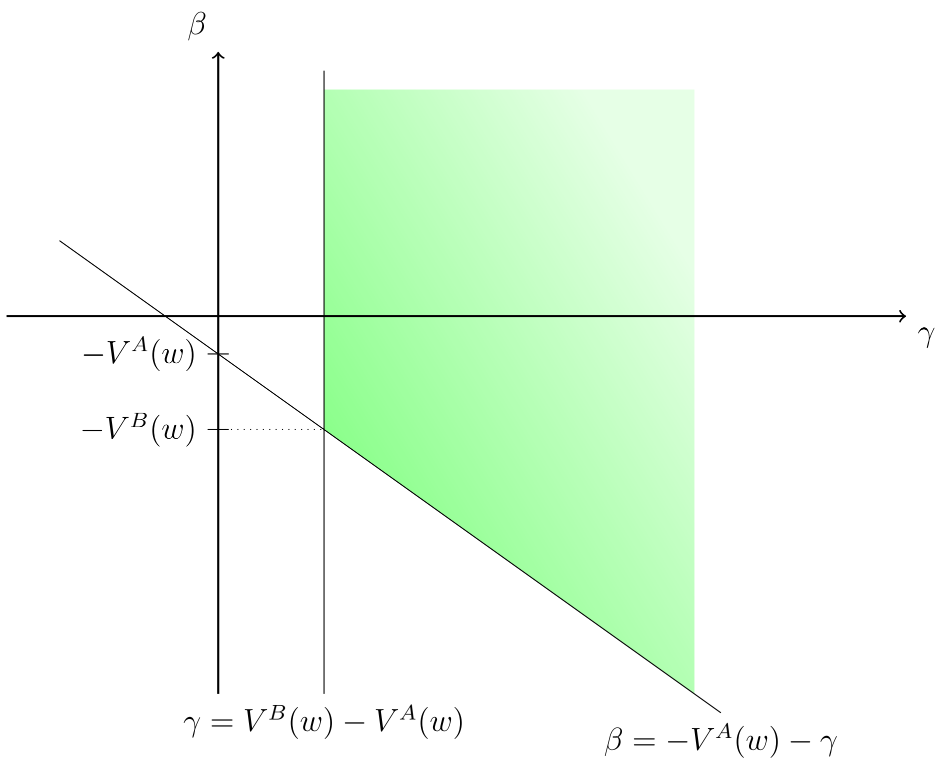

Figure 1: The Set of People who Vote for Party A

which is illustrated in Figure 2. The expected plurality of Party A is given by the difference between the votes cast for Party A and those cast for Party B. Given (7) and (8), the expected plurality of voters of skill level w is given by

$$P(V^A(w), w) = \int_{V^B(w)-V^A(w)}^{\infty} \int_{-V^A(w)-\gamma}^{\infty} f(\gamma, \beta, w) d\beta\, d\gamma - \int_{-\infty}^{V^B(w)-V^A(w)} \int_{-V^A(w)-\gamma}^{\infty} f(\gamma, \beta, w) d\beta\, d\gamma. \quad (9)$$

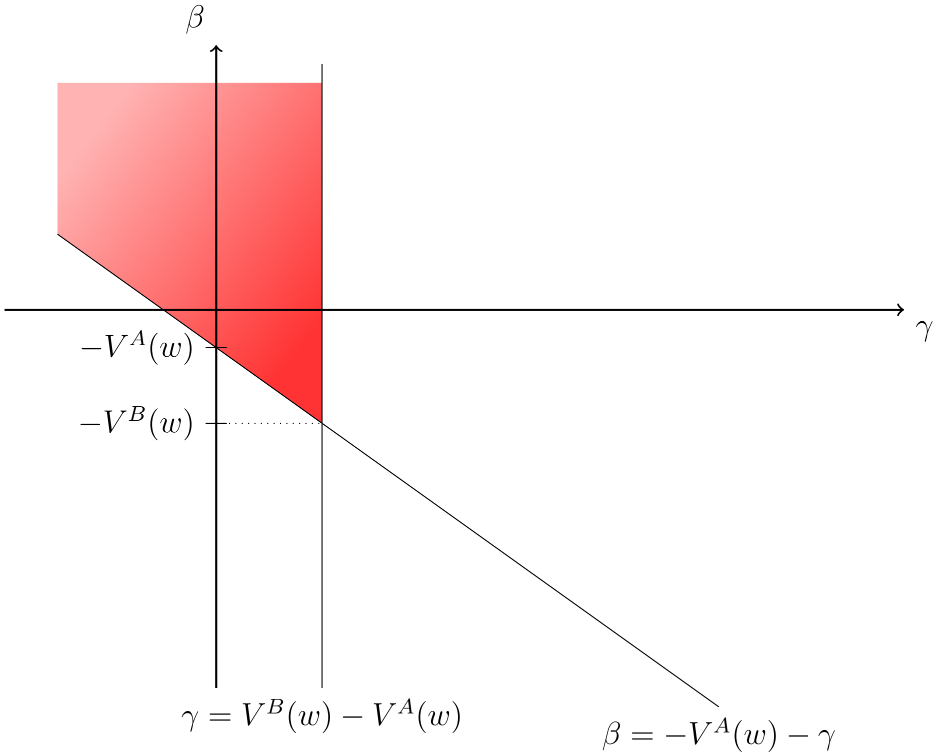

Figure 2: The Set of People who Vote for Party B

Given that taxation is purely redistributive, Party A's reaction function arises out of the maximization of

$$\int_{\underline{w}}^{\overline{w}} P(V^A(w), w) dw \quad (10)$$

subject to the budget constraint

$$\int_{\underline{w}}^{\overline{w}} T^A(w)\theta(V^A(w), w) dw \geq 0. \quad (11)$$

4. Citizen-Retention and Vote-Getting

Changes in tax policy can affect the maximized value of utility enjoyed by each type of worker. Changes in utility, in turn, affect location and voting choices. Parties are concerned with the respective masses of voters and taxpayers of each skill type, because these quantities affect the electoral outcome and the set of feasible policy choices. An application of the standard Leibniz rule implies that

$$\frac{\partial \theta(V^A(w), w)}{\partial V^A(w)} = \int_{-\infty}^{\infty} f(-V^A(w) - \beta, \beta, w) d\beta. \tag{12}$$

The calculation of the the effect of maximized utility on total votes is more involved. Whenever Party A offers more utility to a skill-type e, it induces some voters to switch allegiance from Party B to Party A and it induces migration into the country. Some of these migrants vote for A. With the help of the multidimensional Leibniz rule (Flanders (1973)), it is shown in Appendix A that

$$\frac{\partial M(V^A(w), w)}{\partial V^A(w)} = \int_{-\infty}^{-V^B(w)} f(-V^A(w) - \beta, \beta, w) d\beta + \int_{-V^B(w)}^{\infty} f(V^B(w) - V^A(w), \beta, w) d\beta. \tag{13}$$

This quantity is illustrated in Figure 3. The first term on the right-hand side of (13) represents the lower right portion of the change in the set of voters. Along that boundary, the marginal voters are potential migrants, so the effect of an increase in $V^A(w)$ on votes is the same as its effect on the mass of taxpayers. In accordance with intuition, the migration boundary is salient for voters with a smaller value of β, the attachment to home. The second term is the mass of voters situated along the vertical line segment in Figure 3. Those voters with an attachment to home larger than $-V^B(w)$ will remain in the country regardless of the winner of the election. Party A can entice some of these voters away from Party B by increasing $V^A(w)$. The swing voters are those for which $\gamma = V^B(w) - V^A(w)$, as indicated by the argument in the first component of f in the second term on the right-hand side of (13).

The actions of Party A also have an influence on the votes cast for Party B. Every voter who switches from Party B to Party A also counts as a subtraction to B's support.[2] Moreover, some of the people who enter the country when Party A offers more utility vote for B. The total effect on support for Party B is given by

$$\frac{\partial M^B(V^A(w), w)}{\partial V^A(w)} = \int_{-V^B(w)}^{\infty} f(\beta - V^A(w), \beta, w) d\beta - \int_{-V^B(w)}^{\infty} f(V^B(w) - V^A(w), \beta, w) d\beta, \tag{14}$$

which is illustrated in Figure 4.

[2]This double-counting is necessary when using an expected plurality objective.

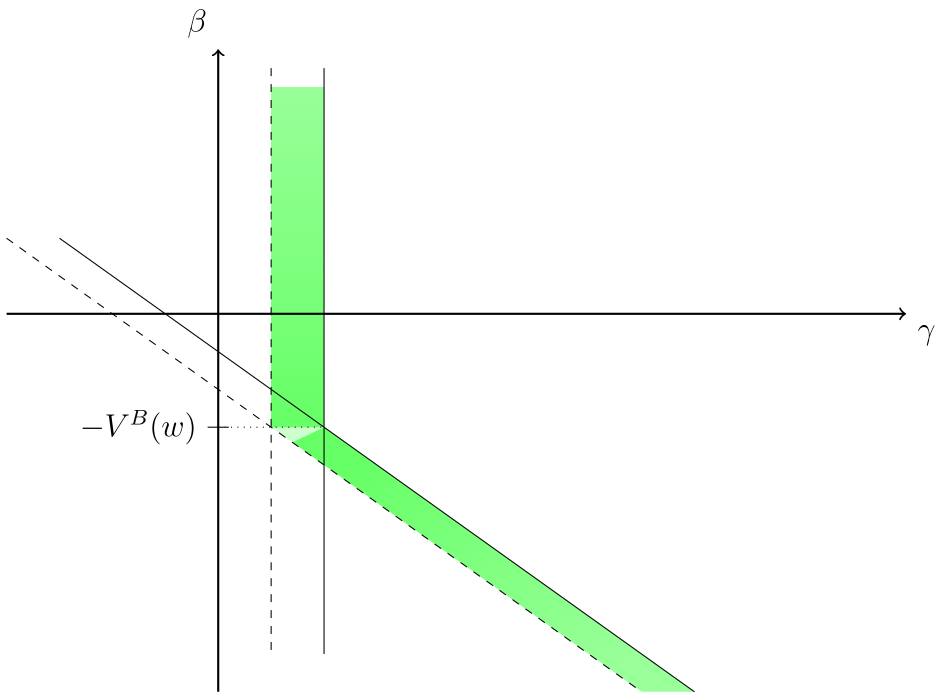

Figure 3: The Leibniz Rule First-Order Approximation to the Change in the Mass of Voters

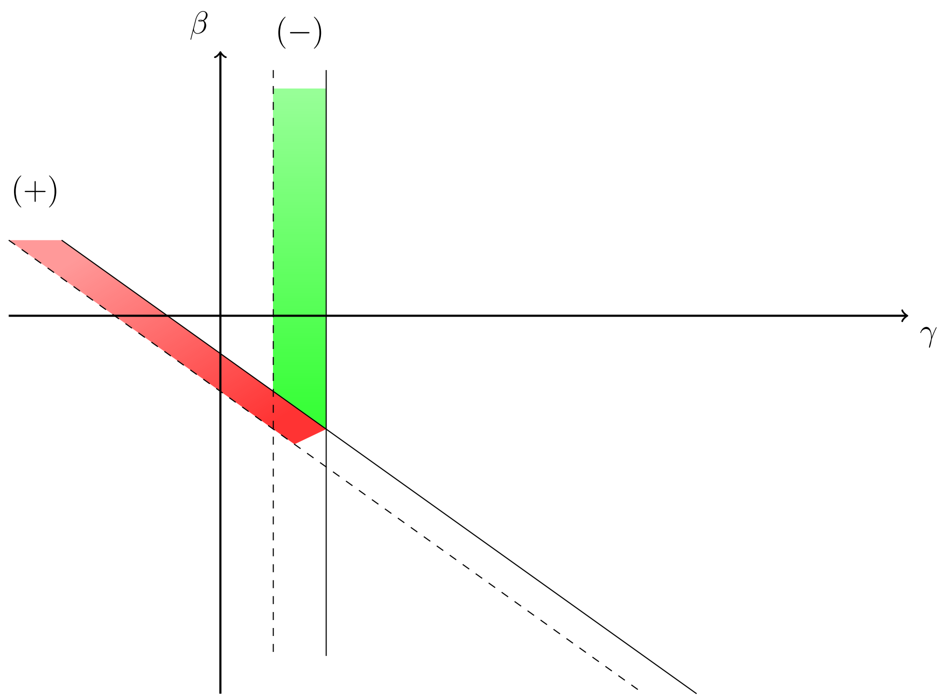

Figure 4: The Effect of an Increase in V^A on the Mass of Votes for Party B

Combing the effects of $V^A(w)$ on the votes cast in favor of the two parties yields the following expression for the contribute of skill-type w toward Party A's expected plurality is affected by an increase in $V^A(w)$:

$$\begin{aligned}\nu(w) &= \frac{\partial P(V^A(w), w)}{\partial V^A(w)} \\ &= 2\int_{-V^B(w)}^{\infty} f(V^B(w) - V^A(w), \beta, w)d\beta + \int_{-\infty}^{-V^B(w)} f(\beta - V^A(w), \beta, w)d\beta \quad (15) \\ &- \int_{-V^B(w)}^{\infty} f(\beta - V^A(w), \beta, w)d\beta.\end{aligned}$$

5. Best Response Nonlinear Income Taxes

Political discussions often include debate over the targeting of tax increases or reductions on specific segments of the income distribution. Thus, it is reasonable to allow the political parties to use a fully nonlinear tax scheme. For ease of exposition, we consider the often-studied case of quasi-linear in consumption preferences, so that

$$u(x, l : w) = x - h(l) = x - h\left(\frac{y}{w}\right). \quad (16)$$

The tax functions $T^i(y)$ must be incentive-compatible. Equivalently, one can image the political parties offering realized utility levels $V^i(w)$ to the workers, along with an associated before-tax income schedule $y^i(w)$. Lemma 1 collects the conditions satisfied by incentive-compatible schemes.

Lemma 1. *Let Party i offer tax schedule $T^i(y)$, $y^i(w)$ be the before-tax income chosen by workers with skills w, and $V^i(w)$ be the associated utility level. If $T^i(y)$ is incentive compatible then*

1. $T^i(y^i(w)) = y^i(w) - V^i(w) - h\left(\frac{y^i(w)}{w}\right)$;
2. $\frac{dV^i(w)}{dw} = h'\left(\frac{y^i(w)}{w}\right)\frac{y^i(w)}{w^2}$;
3. *$y^i(w)$ is non-decreasing.*

Parts 1 and 2 of Lemma 1 can be used to formulate the problem facing Party A as

$$\max_{V^A(w), y^A(w)} \int_{\underline{w}}^{\overline{w}} P(V^A(w), w)dw \quad \text{subject to:} \quad (17)$$

$$\int_{\underline{w}}^{\overline{w}} \left[y^A(w) - V^A(w) - h\left(\frac{y^A(w)}{w}\right)\right]\theta(V^A(w), w)dw = 0, \quad (18)$$

$$\frac{dV^A(w)}{dw} = h'\left(\frac{y^A(w)}{w}\right)\frac{y^A(w)}{w^2}, \quad (19)$$

and the requirement that $y^A(w)$ be non-decreasing.

In the absence of bunching,[3]the parties' best reply tax schedules are solutions to a modified optimal nonlinear tax problem. The solution to this problem can be characterized in an appropriate extension of the Diamond (1998) formula. This description is given Proposition 1 below. In order to help with the exposition of this proposition, Lemma 2 characterizes the shadow value of money for a candidate and Lemma 3 characterizes the political gain to lump-sum transfers to individuals at or above a specific skill level.

Lemma 2. *When Party A uses a best response tax schedule, the associated shadow value of money is given by*

$$\lambda = \frac{\int_{\underline{w}}^{\overline{w}} \frac{\partial P}{\partial V^A(s)} ds}{\left(\int_{\underline{w}}^{\overline{w}} \theta(V^A(s), s) ds\right) \left[1 - E_\theta\left(\varphi(s)\frac{ATR^A(s)}{1-ATR^A(s)}\right)\right]}, \tag{20}$$

where $ATR^A(w) = \frac{T^A(w)}{y^A(w)}$, $E_\theta(a(w)) = \int_{\underline{w}}^{\overline{w}} a(w)\theta(V^A(w), w)dw.$

With quasilinear utility, an exogenous a one-unit increase in money at the disposal of Party A can be doled out equally to everyone in the economy without violating the incentive-compatibility conditions. If a party engages in this activity, it can buy more votes. The right-hand side of (20) measures the effect of those vote purchases on the expected plurality received by Party A. To see this, suppose that there was no geographic mobility. Then term in square brackets in the denominator of (20) would be unity and λ would equal the ratio of the total change in the plurality of Party A to the total population, which is exactly the added plurality associated with dividing an extra dollar equally among an entire immobile population. The term in square brackets in the denominator of (20) adjusts for the change in tax revenue owing to population changes induced by a lump-sum increase in income.

Lemma 3. *When Party A uses a best response tax schedule, the associated shadow value of dividing an exogenous increase in income equally among individuals of skill type w and above is given by*

$$\lambda(w) = \frac{\int_{w}^{\overline{w}} \frac{\partial P}{\partial V^A(s)} ds}{\left(\int_{w}^{\overline{w}} \theta(V^A(s), s) ds\right) \left[1 - E_\theta\left(\varphi(s)\frac{ATR^A(s)}{1-ATR^A(s)} \middle| s > w\right)\right]} \tag{21}$$

The interpretation of (21) is identical to that of (20), except for the restriction of the political and migration responses to skill levels above w. Indeed, $\lambda = \lambda(\underline{w})$. Expression (21) provides some insight into the sources of political power in this model. A high-income group is more valuable to the party (which is the same as electorally powerful) if: (i) their votes are more responsive to changes in utility; (ii) they are of a smaller mass,

[3] I will abstract from bunching in this paper. The usual procedures for handling bunching can be applied.

which lowers the opportunity cost of providing them a transfer; (iii) the larger is their average tax rate; (iv) whenever they pay a positive (negative) amount of tax on average, the more (less) geographically responsive they are to changes in utility. The final two effects are driven by amount of tax revenue in-migrants from the group brings with them. The more money they bring, the lower is the opportunity cost of nudging them into the country.

It is now possible to write the formula characterizing the marginal tax rates in a best response tax function.

Proposition 1. *When both parties offer incentive-compatible tax systems, Party A's best response tax function satisfies the following:*

$$\frac{T^{A\prime}(y^A)}{1-T^{A\prime}(y^A)} = \left[1+\frac{1}{\epsilon}\right]\left[\frac{\int_w^{\overline{w}}\theta(V^A(s),s)ds}{w\theta(V^A(w),w)}\right] \left[1-E_\theta\left(\varphi(s)\frac{ATR^A(s)}{1-ATR^A(s)}\middle| s>w\right)\right]\left(\frac{\lambda-\lambda(w)}{\lambda}\right) \tag{22}$$

where ϵ is the elasticity of labor supply.

All of the expressions in Proposition 1 are evaluated at their equilibrium values. The terms appearing in the first three sets of square brackets on the right-hand side of (22) all appear elsewhere in the optimal tax literature. The first and second are the efficiency and demographic terms found in the Diamond formula. The third arises from the effects of migration on tax revenue, and is discussed in detail by Lehmann et al. (2014). The fourth term arises from probabilistic voting. It replaces the social welfare weights in the original Diamond formula. A similar term arises in the work of Bierbrauer and Boyer (2015), but their expression contains no terms to account for migration and differs in details relating to the nature of political competition in their model.[4]

The following Corollary helps to provide some interpretation for this final term.

Corollary 1. *Provided that $1-E_P\left(\varphi(s)\frac{ATR^A(s)}{1-ATR^A(s)}\middle| s>w\right)>0$, a worker with skill level w faces a positive marginal tax rate in equilibrium if $\lambda(w)<\lambda$.*

The provision in the statement of Corollary 1 ensures that the revenue effects of migration alone are not enough to render a tax cut self-financing. The Corollary states that type w individuals face a positive marginal tax rate if the migration-adjusted political value of individuals above w is less than the analogous political value of all types. In standard (no migration, no politics) optimal nonlinear income tax models, a declining social marginal value of income is sufficient to ensure that the average social marginal utility of persons above w is a decreasing function of w, so that (almost) everyone faces a

[4]Their $\beta^2(w)$ contains terms that reflect the responsiveness of voting to increases in income. These are analogous to the terms in $\lambda(w)$ in (22).

positive marginal tax rate. With political competition and migration, the sign of marginal tax rates is influenced by the relative masses of swing voters at different points of the skill distribution and by the profile of migration elasticities.

6. Conclusion

This paper has provided some insight into the nonlinear tax schemes that would be proposed by political parties when voters are geographically mobile. The resulting tax rule highlights a complex interaction between voting and migration. The rule expressed here is somewhat robust to slight modifications to the model. Indeed, the political influence terms $\lambda(w)$ provide a reduced-form way of describing how politics affects nonlinear taxes. The exact form of these terms depends on the precise details of the political model. I conjecture that other models of politics would generate similar optimal tax formulas, albeit with a different form for the political weights. It is also possible to calibrate the model so as to back out the implicit political weights embodied in existing tax systems.

But the analysis has done more than organize terms into a neat tax formula. It has cataloged some of the sources of political influence. As one would expect in any probabilistic voting model, swing votes attract political attention. But so, too, do segments of the population where flows of people across borders can change the composition of political preferences. Money also plays a role. In the model presented here, politicians are keen to attract and keep segments of the population that pay high (total) taxes even if they do not directly affect electoral outcomes. The tax money they bring to the table can be used by politicians to lower taxes and thereby curry favor with some or all of the other voters. Perhaps this might be called a political economy limit to redistribution.

Appendix A. Derivation of Equation (13)

The region of integration in (7) is sketched in Figure 1. The variable $V^A(w)$ appears only in the limits of integration on the right-hand side of (7), and not in the quantity being integrated. According to the multi-dimensional Leibniz rule of Flanders (1973),

$$\frac{\partial M^A(V^A(w), w)}{\partial V^A(w)} = \oint_C (fud\beta - fvd\gamma), \tag{A.1}$$

where C is the boundary of integration and (u, v) is the velocity vector of the boundary. The region of integration is unbounded on the right and from above, so portions of the boundary that move (that is, that have a velocity vector) are the vertical segment and the diagonal line. Call these two portions C_1 and C_2, respectively. Thus,

$$\frac{\partial M^A(V^A(w), w)}{\partial V^A(w)} = \oint_{C_1} (fud\beta - fvd\gamma) + \oint_{C_2} (fud\beta - fvd\gamma). \tag{A.2}$$

In order to compute the line integrals on the right-hand side of (A.2), it is first necessary to parameterize C_1 and C_2 and to characterize the velocity vectors along those curves.

The curves C_1 and C_2 intersect where

$$\gamma = V^B(w) - V^A(w) \quad \text{and} \quad \gamma + \beta = -V^A(w), \tag{A.3}$$

which implies

$$(\gamma, \beta) = (V^B(w) - V^A(w), -V^B(w)) \tag{A.4}$$

C_1 is naturally described as the vertical half-line extending upward from $(\gamma, \beta) = (V^B(w) - V^A(w), -V^B(w))$. As $V^A(w)$ increases by ϵ, this half-line moves to the left by ϵ. Thus, its velocity vector is $(u, v) = (-1, 0)$. Hence,

$$\oint_{C_1} (fud\beta - fvd\gamma) = \oint_{C_1} -f\, d\beta. \tag{A.5}$$

C_1 can be parameterized by

$$\tilde{C}_1 : \quad (\gamma, \beta) = (V^B(w) - V^A(w), s), \quad s \in (-V^B(w), \infty). \tag{A.6}$$

This parameterization is convenient, but it is in the clockwise direction (opposite to the desired orientation). Nevertheless, (A.6) is used below. With this parameterization, the following holds on C_1:

$$-f\, d\beta = -f(V^B(w) - V^A(w), s, w)\, ds. \tag{A.7}$$

Combining (A.5) and (A.7) yields

$$\begin{aligned} \oint_{C_1} (fud\beta - fvd\gamma) &= -\int_{-V^B(w)}^{\infty} -f(V^B(w) - V^A(w), s, w)\, ds \\ &= \int_{-V^B}^{\infty} f(V^B(w) - V^A(w), \beta, w)\, d\beta. \end{aligned} \tag{A.8}$$

The last expression in (A.8) follows from re-labeling the variable of integration.

C_2 is naturally described as a lower half-line derived from the line in (γ, β)-space with vertical intercept $-V^A(w)$ and slope -1. As $V^A(w)$ increases by ϵ, this half-line moves downward by ϵ. Thus, its velocity vector is $(u, v) = (0, -1)$. Hence,

$$\oint_{C_2} (fud\beta - fvd\gamma) = \oint_{C_2} f\, d\gamma. \tag{A.9}$$

C_2 can be parameterized by

$$C_2 : \quad (\beta, \gamma) = (V^B(w) - V^A(w) + s, -V^B(w) - s), \quad s \in (0, \infty). \tag{A.10}$$

This parameterization is in the counterclockwise direction, because it starts at the point $(\gamma, \beta) = (V^B(w) - V^A(w), -V^B(w))$ and moves outward along C_2. With this parameterization, the following holds on C_2:

$$f\, d\gamma = f(V^B(w) - V^A(w) + s, -V^B(w) - s, w)\, ds. \tag{A.11}$$

Combining (A.9) and (A.11) yields

$$\oint_{C_2} (fud\beta - fvd\gamma) = \int_0^\infty f(V^B(w) - V^A(w) + s, -V^B(w) - s, w)\, ds. \tag{A.12}$$

With the substitution $q = -V^B(w) - s$, (A.12) becomes

$$\begin{aligned}\oint_{C_2} (fud\beta - fvd\gamma) &= -\int_{-V^B(w)}^{-\infty} f(-V^A(w) - q, q, w)\, dq \\ &= \int_{-\infty}^{-V^B(w)} f(-V^A(w) - q, q, w)\, dq = \int_{-\infty}^{-V^B(w)} f(-V^A(w) - \beta, \beta, w)\, d\beta.\end{aligned} \tag{A.13}$$

Adding the two components of the line integral (A.2), as computed in (A.8) and (A.13), yields (13) in the main text.

Appendix B. Proofs

Proof of Lemmas 2 and 3 and Proposition 1. Abstracting from bunching, Party A's decision problem is an optimal control problem with state variable $V^A(w)$, control variable $y(w)$, and Hamilton-Langrange function

$$\mathcal{H} = P(V^A(w), w) + \lambda \left[y(w) - V(w) - h\left(\frac{y(w)}{w}\right)\right] \theta(V^A(w), w) + \kappa(w) h'\left(\frac{y(w)}{w}\right)\frac{y(w)}{w^2}, \tag{A.14}$$

where λ is a multiplier on the isoperimetrical constraint (18) and $\kappa(w)$ is a co-state variable on the flow constraint (19). The associated necessary conditions for an optimum are

$$-\kappa'(w) = \mathcal{H}_v = \frac{\partial P}{\partial V^A(w)} + \lambda \left[-\theta(V^A(w), w) + T^A(w) \frac{\partial \theta(V^A(w), w))}{\partial V^A(w)}\right]; \tag{A.15}$$

$$\begin{aligned} 0 &= \mathcal{H}_y \\ &= \lambda \left[1 - h'\left(\frac{y(w)}{w}\right)\frac{1}{w}\right] \theta(V^A(w), w)) + \kappa(w)\left[h'\left(\frac{y(w)}{w}\right)\frac{1}{w^2} + h''\left(\frac{y(w)}{w}\right)\frac{y(w)}{w^3}\right]; \end{aligned} \tag{A.16}$$

$$\kappa(\underline{w}) = \kappa(\overline{w}) = 0. \tag{A.17}$$

Now, from (A.15) and (A.17),

$$0 = \int_{\underline{w}}^{\overline{w}} -\kappa'(w) = \int_{\underline{w}}^{\overline{w}} \frac{\partial P}{\partial V^A(s)} + \lambda \int_{\underline{w}}^{\overline{w}} \left[-\theta(V^A(s), s) + T^A(s) \frac{\partial \theta(V^A(s), s)}{\partial V^A(s)}\right] ds, \tag{A.18}$$

so that

$$\lambda = \frac{\int_{\underline{w}}^{\overline{w}} \frac{\partial P}{\partial V^A(s)} ds}{\int_{\underline{w}}^{\overline{w}} \left[1 - T^A(s) \frac{\partial \ln \theta(V^A(s), s)}{\partial V^A(s)}\right] \theta(V^A(s), s) ds}. \tag{A.19}$$

Now, following the example of Lehmann et al. (2014), we can show that

$$\begin{aligned} &\int_{w}^{\overline{w}} \left[1 - T^A(s)\frac{\partial \ln \theta(V^A(s), s)}{\partial V^A(s)}\right] \theta(V^A(s), s) ds \\ &= \left(\int_{w}^{\overline{w}} \theta(V^A(s), s) ds\right) \left[1 - \frac{\int_{w}^{\overline{w}} T^A(s)\frac{\partial \ln \theta(V^A(s),s)}{\partial V^A(s)}\theta(V^A(s), s) ds}{\int_{w}^{\overline{w}} \theta(V^A(s), s) ds}\right]. \end{aligned} \tag{A.20}$$

Because preferences are quasilinear in x,

$$\begin{aligned} T^A(s)\frac{\partial \ln \theta(V^A(s), s)}{\partial V^A(s)} &= \frac{T^A(s)}{y(s)}\frac{y(s)}{x(s)}x(s)\frac{\partial \ln \theta(V^A(s), s)}{\partial x(s)} \\ &= \frac{ATR^A(s)}{1 - ATR^A(s)}\varphi(s) \end{aligned} \tag{A.21}$$

Substituting (A.21) into (A.20) and using the definition of the expectation E_θ, one obtains

$$\begin{aligned} &\int_{w}^{\overline{w}} \left[1 - T^A(s)\frac{\partial \ln P(V^A(s), s))}{\partial V^A(s)}\right] \theta(V^A(s), s) ds \\ &= \left(\int_{w}^{\overline{w}} \theta(V^A(s), s) ds\right) \left[1 - E_\theta\left(\varphi(s)\frac{ATR^A(s)}{1 - ATR^A(s)}\middle| s > w\right)\right] \end{aligned} \tag{A.22}$$

Substituting the version of (A.22) for $w = \underline{w}$ into (A.19) proves Lemma 2. Lemma 3 follows from (A.22) and the observations in the main text immediately preceding its statement.

Turning now to remainder of the demonstration of Proposition 1, consider the optimization problem for a worker with skills w when faced with a tax schedule $T^A(wl)$. The consumer solves the problem

$$\max_{l} x - h(l) \quad \text{subject to:} \quad x = wl - T^A(wl). \tag{A.23}$$

The first order condition associated with (A.23) is

$$w[1 - T^{A\prime}(w)] = h'(l). \tag{A.24}$$

Let $\tilde{w} = w[1 - T^{A\prime}(w)]$, the after-tax wage rate. Then, differentiating (A.24) yields

$$\frac{dl}{d\tilde{w}} = \frac{1}{h''(l)}, \tag{A.25}$$

and, using (A.24) the labor supply elasticity is given by

$$\epsilon = \frac{dl}{d\tilde{w}}\frac{\tilde{w}}{l} = \frac{1}{h''(l)}\frac{h'(l)}{l}. \tag{A.26}$$

Now, from (A.24),

$$T^{A\prime}(y^A) = 1 - h'\left(\frac{y(w)}{w}\right)\frac{1}{w} \quad \text{and} \quad 1 - T^{A\prime}(y^A) = h'\left(\frac{y(w)}{w}\right)\frac{1}{w}. \tag{A.27}$$

It follows from (A.16) and (A.27) that

$$\frac{T^{A\prime}(y^A)}{1 - T^{A\prime}(y^A)} = \frac{-\kappa(w)}{\lambda P(V^A(w), w)}\left[\frac{1}{w} + \frac{h''\left(\frac{y(w)}{w}\right)}{h'\left(\frac{y(w)}{w}\right)}\frac{y(w)}{w^2}\right]. \tag{A.28}$$

Using the equation $y(w) = wl$ in (A.26), and substituting into (A.28) yields

$$\frac{T^{A\prime}(y^A)}{1 - T^{A\prime}(y^A)} = \frac{-\kappa(w)}{\lambda w \theta(V^A(w), w)}\left[1 + \frac{1}{\epsilon}\right]. \tag{A.29}$$

Also from (A.15) and (A.17),

$$\kappa(w) = \int_w^{\overline{w}} \frac{\partial P}{\partial V^A(s)} ds - \lambda \int_w^{\overline{w}} \left[1 - T^A(s)\frac{\partial \ln \theta(V^A(s), s))}{\partial V^A(s)}\right]\theta(V^A(s), s) ds. \tag{A.30}$$

Combining (A.19) and (A.30) yields,

$$\begin{aligned}\frac{-\kappa(w)}{\lambda} = &\int_w^{\overline{w}} \left[1 - T^A(s)\frac{\partial \ln \theta(V^A(s), s))}{\partial V^A(s)}\right]\theta(V^A(s), s) ds \\ &- \left(\int_{\underline{w}}^{\overline{w}} \left[1 - T^A(s)\frac{\partial \ln \theta(V^A(s), s))}{\partial V^A(s)}\right]\theta(V^A(s), s) ds\right)\frac{\int_w^{\overline{w}} \frac{\partial P}{\partial V^A(s)} ds}{\int_{\underline{w}}^{\overline{w}} \frac{\partial P}{\partial V^A(s)} ds}.\end{aligned} \tag{A.31}$$

Factoring out the first term on the right-hand side of (A.31) implies

$$\begin{aligned}\frac{-\kappa(w)}{\lambda} = &\int_w^{\overline{w}} \left[1 - T^A(s)\frac{\partial \ln \theta(V^A(s), s))}{\partial V^A(s)}\right]\theta(V^A(s), s) ds \\ &\left[1 - \frac{\int_{\underline{w}}^{\overline{w}} \left[1 - T^A(s)\frac{\partial \ln \theta(V^A(s),s))}{\partial V^A(s)}\right]\theta(V^A(s), s) ds \int_w^{\overline{w}} \frac{\partial P}{\partial V^A(s)} ds}{\int_w^{\overline{w}} \left[1 - T^A(s)\frac{\partial \ln \theta(V^A(s),s))}{\partial V^A(s)}\right]\theta(V^A(s), s) ds \int_{\underline{w}}^{\overline{w}} \frac{\partial P}{\partial V^A(s)} ds}\right].\end{aligned} \tag{A.32}$$

Repeated use of (A.22) and Lemmas 2 and 3 in (A.32) yields

$$\frac{-\kappa(w)}{\lambda} = \left(\int_w^{\overline{w}} \theta(V^A(s), s) ds\right)\left[1 - E_\theta\left(\varphi(s)\frac{ATR^A(s)}{1 - ATR^A(s)}\middle| s > w\right)\right]\left(1 - \frac{\lambda(w)}{\lambda}\right) \tag{A.33}$$

Substituting (A.33) into (A.29) yields (22). □

References

Bierbrauer, F., Boyer, P., 2013. Political competition and Mirrleesian income taxation: A first pass. Journal of Public Economics 103, 1–14.

Bierbrauer, F., Boyer, P., 2015. Efficiency, welfare, and political competition. Unpublished manuscript, Centre for Macroeconomic Research, University of Cologne.

Bierbrauer, F., Brett, C., Weymark, J. A., 2013. Strategic nonlinear income tax competition with perfect labor mobility. Games and Economic Behavior 82 (No. 3329), 292–311.

Bohn, H., Stuart, C., 2013. Revenue extraction by median voters, unpublished manuscript, Department of Economics, University of California, Santa Barbara.

Brett, C., Weymark, J. A., 2016. Voting over selfishly optimal nonlinear income tax schedules. Games and Economic Behavior, forthcomingIn preparation.

Charles Rochet, J., Stole, L. A., 2002. Nonlinear pricing with random participation. Review of Economic Studies 69, 277–311.

Chen, Y., 2000. Electoral systems, legislative process, and income taxation. Journal of Public Economic Theory 2, 71–100.

Cremer, H., Pestieau, P., 2004. Factor mobility and redistribution. In: Henderson, J. V., Thisse, J. F. (Eds.), Handbook of Urban and Regional Economics. Vol. 4. North-Holland, Amsterdam, pp. 2529–2560.

Diamond, P. A., 1998. Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates. American Economic Review 88, 83–95.

Flanders, H., 1973. Differentiation under the integral sign. American Mathematical Monthly 80, 615–627.

Gordon, R. H., Cullen, J. B., 2011. Income distribution in a federal system of governments. Journal of Public Economics, forthcoming.

Hamilton, J., Lozachmeur, J.-M., Pestieau, P., 2002. Rawlsian governments and the race to the bottom. Economics Bulletin 8 (2), 1–6.

Hamilton, J., Pestieau, P., 2005. Optimal income taxation and the ability distribution: Implications for migration equilibria. International Tax and Public Finance 12, 29–45.

Hinich, M. J., Ledyard, J. O., Ordeshook, P. C., 1972. Nonvoting and the existence of equilibrium under majority rule. Journal of Economic Theory 14, 144–153.

Lehmann, E., Simula, L., Trannoy, A., 2014. Tax me if you can! Optimal nonlinear income tax between competing governments. Quarterly Journal of Economics 129, 1995–2030.

Lipatov, V., Weichenrieder, A., 2010. Optimal income taxation with tax competition. Working Paper No. 3108, CESifo.

Morelli, M., Yang, H., Ye, L., 2011. Competitive nonlinear taxation and constitutional choice. American Economic Journal: Microeconomics, forthcoming.

Piaser, G., 2007. Labor mobility and income tax competition. In: Gregoriou, G., Read, C. (Eds.), International Taxation Handbook: Policy, Practice, Standards and Regulation. CIMA Publishing, Oxford, pp. 73–94.

Roemer, J. E., 2012. The political economy of income taxation under asymmetric information: The two-type case. SERIEs: Journal of the Spanish Economic Association 3, 181–199.

Saez, E., 2001. Using elasticities to derive optimal income tax rates. Review of Economic Studies 68, 205–229.

Simula, L., Trannoy, A., 2010. Optimal income tax under the threat of migration by top-income earners. Journal of Public Economics 94, 163–173.

Simula, L., Trannoy, A., 2011. Shall we keep the highly skilled at home? The optimal income tax perspective. Social Choice and Welfare, forthcoming.

Stigler, G. J., 1957. The tenable range of functions of local government. In: Federal Expenditure Policy for Economic Growth and Stability. Joint Economic Committee, Subcommittee on Fiscal Policy, U. S. Congress, Washington, D. C., pp. 213–219, reprinted in E. S. Phelps (Ed.), 1965. Private Wants and Public Needs. New York: Norton, 167–176.

Wilson, J. D., 1980. The effect of potential emigration on the optimal linear income tax. Journal of Public Economics 14, 339–353.

Wilson, J. D., 1992. Optimal income taxation and international personal mobility. American Economic Review, Papers and Proceedings 82, 191–196.