Progressive and Regressive Equilibria in a Tax Competition Game

James Green-Armytage*

Department of Economics, Bard College,

Annandale-on-Hudson, NY 12504

armytage@bard.edu

Accepted for publication in *Public Finance Review* on September 22, 2015

Abstract: This paper models inter-jurisdictional competition over non-linear taxes on the incomes of mobile individuals. Each individual has exogenous wealth and a location preference that is drawn from a continuous distribution. We find that more concave utility of consumption functions lead to more progressive tax structures, as richer people place less value on marginal consumption relative to location. In the benchmark model, a relative risk aversion coefficient of one is the boundary between progressivity and regressivity. The exercise helps us to understand which types of jurisdictions are more likely to have progressive taxes as their optimal policies.

Journal of Economic Literature classification codes: H2, H7

Keywords: tax competition, optimal taxation, nonlinear taxes, migration, Nash equilibrium taxes

---

* I am grateful for comments by James Alm, Ted Bergstrom, Ted Frech, Marek Kapička, Charlie Stuart, Nicolaus Tideman, and three anonymous referees.

1. Introduction

The threat of migration serves as a potential limit to redistribution.[1] Therefore, when one is either designing a tax or seeking to understand the reasons for the shape of an existing tax, it is important to keep this constraint in mind, along with other limiting factors such as labor supply elasticity,[2] the possibility of people shifting income into other forms,[3] a political aversion to redistribution,[4] etc. This limit is more consequential when inter-jurisdictional mobility is less costly; therefore, it tends to constrain the tax policies of smaller jurisdictions (for example, state and local governments) more than larger ones (for example, a federal government).[5]

The primary message of this paper is as follows: Ceteris paribus, progressive taxation is more likely to be optimal, and thus more likely to be observed, when a typical individual's marginal utility of consumption decreases more rapidly with consumption relative to the total utility he gains from residing in his preferred jurisdiction. On the other hand, regressive taxation will be more likely when the marginal utility of consumption decreases more slowly with consumption relative to locational utility, and proportional taxes will be more likely when it can be approximated by a simple inverse proportion of consumption.

When is each of these the case? Consider first a jurisdiction that provides something to its residents that impacts their utility directly, in a way that is not easily replaceable by money. For example, perhaps living in the jurisdiction is important to residents' ability to spend time with friends and family members. Perhaps it allows access to scarce and highly valued resources, whether in the form of natural resources (such as a beach) or man-made resources (such as a cityscape). Perhaps living elsewhere requires one to speak a different language, or to become used to different social customs. In this case, as people become more wealthy, their marginal utility of consumption becomes small relative to the utility they gain from living where they prefer to live, so the jurisdiction is able to charge them a relatively high tax without inducing them to emigrate. Thus, a progressive tax is more likely to be optimal for such jurisdictions.

On the other hand, consider a jurisdiction where location preferences are reducible to moving costs in the narrow sense of transporting belongings and establishing a new household elsewhere. If these moving costs are independent of income, the jurisdiction will not be able to extract any more tax revenue from a rich person than a poor person without inducing them to emigrate. Thus, the optimal tax is likely to be regressive. In this case, the marginal utility of consumption does not shrink relative to locational utility as consumption increases, because the utility loss associated with paying the moving cost decreases at the same rate.

---

[1] Cremer and Pestieau (2004) survey the literature on tax competition as it relates to the ability of governments to redistribute wealth. Epple and Romer (1991) present a model in which endogenously formed communities vote on a single, redistributive property tax. Hindriks (1999) presents a model in which jurisdictions vote on a single redistribution variable, which is either the tax rate or the transfer payment. Wilson (1999) and Brueckner (2003) provide surveys of the theoretical and empirical literatures (respectively) on tax competition in general. This literature often explores the idea of 'market failure' resulting from inter-jurisdictional externalities, whereas the literature following Tiebout (1956) tends to highlight the 'market efficiency' resulting from inter-jurisdictional competition.

[2] Meltzer and Richard (1981) provide a seminal exploration of this idea.

[3] Feldstein (1995), among many others, explores this idea; Goolsbee (1999) refers to it as the central hypothesis of the "New Tax Responsiveness Literature".

[4] For a survey of limits to redistribution within a democratic process, see Harms and Zink (2003).

[5] See Stigler (1957), Musgrave (1971), Oates (1972), and Brown and Oates (1987). Similarly, Sinn (1990) argues for tax harmonization agreements between countries.

From the above discussion, it begins to become clear that smaller jurisdictions are less likely to have progressive taxes than larger ones, because (for example) the town one lives in is usually less critical to the maintenance of social relationships than the country one lives in. Further, it becomes clear that jurisdictions with unique, highly-valued features are more likely to have progressive income taxes than similar-sized jurisdictions that lack such features.

The primary message of the paper as stated above includes a sweeping ceteris paribus clause, which contains assumptions that we use in the service of simplicity and clarity. One such assumption is that individuals have exogenous wealth. That is, the model developed here supposes that each individual's pre-tax income is independent of his location[6] and the tax rate he faces.[7] In practice, labor supply incentive compatibility and potential migration both constrain taxation, but the importance of each constraint relative to the other varies from jurisdiction to jurisdiction. This paper explores one extreme of that spectrum, where the labor supply constraint is negligible in comparison to the migration constraint. Thus it more closely resembles the reality of smaller jurisdictions, who are more likely to induce individuals to move by raising their taxes, and less likely to change individuals' income-generating behavior.[8]

This assumption of exogenous wealth sets the paper apart from most of the migration-constrained optimal taxation literature, which usually extends the Mirrlees's (1971) model by adding migration as a second constraint in addition to labor supply elasticity.[9] The value of this predominant approach is beyond doubt, since both constraints are present in reality, but just as there are several useful papers considering the labor supply constraint without the migration constraint, it is also useful to consider the migration constraint without the labor supply constraint.[10] This provides a theoretical benchmark which helps us understand some underlying dynamics in the general case by first improving our understanding the special case with only a migration constraint. An advantage of this approach is that it can reveal features of the optimal tax functions in easily-graspable analytical terms, without reliance on numerical simulations. In particular, our goal is to focus in a maximally simple and clear manner on the question of

[6] This assumption is actually not crucial to the logic of the model, as it could potentially accommodate differing income opportunities in the two jurisdictions via a modification of the location preferences function. But we do not explore this idea in detail here.

[7] Individual taxable incomes are closer to exogenous when there are fewer loopholes that individuals can use to shift income into untaxed forms while remaining in the jurisdiction.

[8] It is worth noting that the smaller jurisdictions where this assumption is more realistic are also the ones where the threat of migration is most relevant to tax policy.

[9] The application of this approach to nonlinear taxation is an active frontier in public finance theory. For example, Simula and Trannoy (2010) extend the Mirrleesian framework to the case in which individuals may exit with some cost to a country with a fixed low-tax policy. Other recent papers allow both countries to have flexible tax policies, and focus on the Nash equilibrium in their strategic interaction, as we do here: Morelli et al (2012) assume country symmetry, and focus primarily on the contrast with the autarkic case. Bierbrauer et al (2011) assume costless migration. Lehmann et al (2013) build a framework that can accommodate elements of both Brewer et al (2010) and Blumkin et al (2012), finding that the a key difference between their approaches and results (particularly concerning whether top rates of zero are optimal) can be framed in terms of differences in the "semi-elasticity of migration". Lehmann et al (2013) is closely related to this paper, but doesn't focus on the question of when the equilibrium tax function is likely to be progressive. An older literature investigates competition over linear tax structures; for example, see Wilson (1980).

[10] In that it focuses on an extensive margin (in this case, emigration) rather than the intensive margin (e.g. changes in hours of work, or effort), our exercise is analogous to Diamond (1980) and the Saez (2002). These papers focus on the extensive margin of exiting the labor force, and thus create room in the Mirrleesian model for negative marginal tax rate schemes like the Earned Income Tax Credit.

progressivity under tax competition – specifically, how it is affected by the role that location plays in the utility of individuals.

In the Mirrleesian literature with labor supply elasticity as the primary constraint on taxes, there is a common result beginning with Sadka (1976) that the optimal marginal tax rate eventually decreases to zero. Others have questioned the relevance of this result,[11] but it still appears in many specifications and remains as a well-known consequence of the model. However, in the model presented here with migration as the primary constraint, marginal tax rates continue to increase in income as long as the marginal utility of consumption continues to decrease relative to locational utility. This gives a new way to think about the question of what the optimal tax looks like in a more realistic world with both labor supply constraints and migration constraints: that is, when these two effects go in opposite directions, we ask which is dominant.

A few other key details of the model are worth mentioning at the outset. The first is our assumption about the governments' objective function. Here we choose the simplest and cleanest alternative, which can be thought of in a few ways: (a) Governments are using a maximin social welfare function.[12]
(b) 'Leviathan'[13] governments are seeking to maximize revenue. (c) There is a democratic process in which the poor and middle-class people form a coalition to redistribute as much wealth as possible from the rich. (d) Governments are using a utilitarian social welfare function, but – as Diamond and Saez (2011) argue – revenue maximization is a reasonable first approximation for this when we are concerned with taxes on the rich.

Some discussion on the conditional equivalence of these assumptions within our model: In the mathematics below, governments' objectives are represented in terms of revenue maximization.
(a) If governments use a minimax social welfare function, they are only concerned with the well-being of the poorest individuals, so they seek to maximize revenue from the non-poorest in order to effect maximal redistribution. Thus our model applies at all income levels above the very poorest in this case.[14]
(b) If governments behave as Leviathans, this analysis applies at all income levels by construction.
(c) If governments are driven by a democratic coalition of the poor and middle classes, the model applies to those with incomes above the levels included in this coalition, i.e. above the median income, at minimum. (d) If we are thinking of the conditions derived from a revenue maximization assumption as a first approximation of the conditions that would result from a utilitarian social welfare maximization assumption, there is no specific income level where the results are precisely equal, but they are closest when we are dealing with taxes on the very rich.

---

[11] For example, Diamond (1998) creates examples in which optimal marginal tax rates are increasing at high incomes.
[12] The objective in this case is to maximize the utility of the poorest individuals. Rawls (1971) argues for a social welfare function of this type. Recent tax competition papers such as Simula and Trannoy (2010) and Lehmann et al (2013) also employ this function in the interest of simplicity.
[13] Brennan and Buchanan (1980) use this term to refer to a revenue-maximizing government.
[14] As Mirrlees (1982) points out, utilitarian governments in an open economy face further questions as to whether they care about the utility of their own native residents, their own residents including immigrants, all people regardless of birthplace or residence, etc. The results of the model are largely independent of the answer to this question when we are using a maximin function, since no positive welfare weight is placed on anyone but the very poorest individuals. By construction, these individuals are outside our immediate analysis in that we are not seeking to determine how much they should be taxed. Instead, we assume only that their welfare increases with the government's tax revenue because they benefit from resulting increases in spending; whether this spending goes to natives, non-natives, etc. may be left open for now.

In summary, the assumptions used here fit better with a larger number of underlying stories when we apply them to taxes on higher income levels.[15] It is a highly convenient feature of this model – which is developed rigorously in Section 2 below, and which follows from the assumption of exogenous wealth – that the equilibrium taxes on each income level are independent on the equilibrium taxes on every other income level. It is this particular feature that allows us to be flexible with regard to which income levels we consider to be part of the analysis, and thus which allows us to choose liberally among these underlying stories.

Furthermore, although it is not part of the main thrust of the paper, we do explore in an appendix the alternative assumption that governments pursue Benthamite (i.e. additive) utilitarian social welfare maximization. A sharper focus on this is one possible direction for an extension of the model in a future paper, but the contrast with the base model suggested by our limited foray is an intuitive starting point: The equilibrium under utilitarian social welfare maximization is likely to be more progressive than the equilibrium under revenue maximization when we are comparing taxes on two income levels that are both within the analysis, e.g. when we are comparing the very-rich to the slightly-less-rich. This is intuitive because utilitarian tax policy most closely resembles revenue maximization where the highest incomes are concerned.

To complete the introductory description of our model, we specify that it includes two jurisdictions (a logical first step that is common in the literature), and focuses on pure strategy Nash equilibrium outcomes. Modeling Nash equilibria between jurisdictions with respect to nonlinear tax structures is a logical frontier for investigation because nonlinear taxes are common, migration is an important constraint on taxes, and pure strategy Nash equilibrium is the most natural theoretical starting point for the prediction of long-run policy outcomes.

The remainder of the paper is organized as follows. Section 2 defines the model, Section 3 presents the results, and Section 4 concludes. Appendix 1 gives proofs and examples, and Appendix 2 discusses the case where governments pursue a Benthamite utilitarian social welfare function.

2. Model

In this model, two jurisdictions attempt to maximize revenue from those with wealth above a certain threshold, but each jurisdiction is constrained by the fact that more individuals may choose to emigrate as it raises its taxes. Individuals differ from one another by wealth and by location preference. In the interest of simplicity, we examine a static case, and abstract away constraints on taxation other than the threat of migration, such as labor supply incentive compatibility, etc. Thus, the initial distribution of wealth is

[15] Conversely, the model is not meant to be applied to the taxation of low-income individuals. That is, the objective of maximizing revenue from the poor is generally incompatible with conventional ideas of social welfare. For example, the revenue maximization objective here implies that taxes will necessarily be non-negative for each wealth type in the analysis, but when it comes to the poor, policy makers should consider negative taxes (i.e. net transfers), both in the form of negative marginal tax rates and lump-sum transfers that are gradually taxed away with increases in income. The latter type of transfer appears in Appendix 2 of this paper, where we consider the consequences of a utilitarian social welfare function and thus increase the model's relevance to low-income taxation.

treated as exogenous here; that is, each person starts with a given amount of wealth that does not depend on chosen effort.

Taxes and equilibrium

Let j and $k \neq j$ be jurisdiction indices. Let $\theta_j(W)$ represent the tax that jurisdiction j charges those with wealth W. (Note that this is the *total* tax, as opposed to the marginal tax rate or the average tax rate.)[16] Assume that each jurisdiction is trying to maximize the revenue that it receives from those with wealth greater than or equal to a threshold value $\underline{W}$. Let Θ_j represent jurisdiction j's overall tax function on those with $W \geq \underline{W}$. (We make no comments or conjectures about taxes on lower wealth levels – these are simply outside of our analysis.)

Assume also a maximum tax constraint of $\theta_j(W) \leq W$; this means that it is not possible to take more from a person than he has. Define $\mathcal{S}$ as the set of tax functions that satisfy the maximum tax constraint; that is, $\Theta_j \in \mathcal{S} \leftrightarrow \theta_j(W) \leq W, \forall W \geq \underline{W}$.

Let $\Omega_j(\Theta_j, \Theta_k)$ represent jurisdiction j's objective function, i.e. the total revenue it receives from those with wealth greater than $\underline{W}$. So we represent jurisdiction j's optimization problem as

$$\max_{\Theta_j \in \mathcal{S}} \Omega_j(\Theta_j, \Theta_k)$$

Our goal is to study pure strategy Nash equilibria in this tax competition game. Therefore, an equilibrium will be defined as a situation in which

$$\Theta_j = \operatorname*{argmax}_{\Theta_j \in \mathcal{S}} \Omega_j(\Theta_j, \Theta_k), \forall j = 1, 2$$

That is, each jurisdiction's tax function maximizes revenue from the rich, given the other jurisdiction's tax function, and subject to the maximum tax constraint.

Each jurisdiction's total revenue is given by the integral[17] over wealth types above $\underline{W}$ of revenue from each wealth type, which we will represent as $R_j(\theta_j(W), \theta_k(W), W)$. That is,

$$\Omega_j(\Theta_j, \Theta_k) = \int_{\underline{W}}^{\infty} R_j(\theta_j(W), \theta_k(W), W)\, dW$$

It is important to note that jurisdiction j's tax revenue from those with a particular wealth level of W depends only on $\theta_j(W)$, $\theta_k(W)$, and W; that is, it does *not* depend on either j or k's taxes on any other wealth level W'. This is true specifically because of our assumption that wealth is exogenous; as a result of this, a person with wealth W has no possibility of having any more or less wealth, so he does not take into account what people with more or less wealth are being charged when he makes his location decision. Therefore, we may define equilibrium equivalently as a situation in which

$$\theta_j(W) = \operatorname*{argmax}_{\theta_j(W) \leq W} R_j(\theta_j(W), \theta_k(W), W), \forall j = 1, 2; \forall W \geq \underline{W}$$

[16] A point of linguistic clarification: Our default use of the word 'tax' will be this total tax; when we are speaking instead of the marginal tax rate or the average tax rate, we will specify this explicitly.

[17] Alternatively, we may suppose a finite number of wealth levels, and thus a discrete distribution of wealth rather than a continuous one, without altering our fundamental results.

That is, at each wealth level W, each jurisdiction j chooses a tax on that wealth level, $\theta_j(W)$, that maximizes revenue from those with W, given the other jurisdiction's tax on that wealth level, $\theta_k(W)$. This is the definition of equilibrium that we will make direct use of in our results below.

Populations and revenues

Let $\pi_j(\theta_j(W), \theta_k(W), W)$ represent jurisdiction j's population of those with wealth W, i.e. those who choose to live in j once the tax rates have been decided. Let $\Pi(W)$ be a density function giving the total population (in both jurisdictions combined) of those with each wealth level W. Because everyone must live in one jurisdiction or the other, we have

$$\pi_j(\theta_j(W), \theta_k(W), W) + \pi_k(\theta_j(W), \theta_k(W), W) = \Pi(W)$$

Jurisdiction j's tax revenue from those with wealth W is the product of its tax, $\theta_j(W)$, and its per-type population $\pi_j(\theta_j(W), \theta_k(W), W)$. This can be expressed as follows.[18]

$$R_j(\theta_j(W), \theta_k(W), W) = \theta_j(W)\pi_j(\theta_j(W), \theta_k(W), W)$$

Utility

In order to determine the precise way in which j's per-type population $\pi_j(\theta_j(W), \theta_k(W), W)$ depends on the two jurisdictions' taxes, we must first describe the utility functions that underlie each individual's migration decision. We will assume that this utility function is

$$U_{ij} = V(C_j) + \xi_{ij}$$

Here, U_{ij} denotes the utility of a person i if he chooses jurisdiction j, and $C_j(\theta_j(W), W)$ denotes the consumption of an individual with wealth W who chooses jurisdiction j; consumption is defined as wealth minus tax, i.e.

$$C_j(\theta_j(W), W) = W - \theta_j(W)$$

The function $V(C)$ represents utility from consumption. We simplify by assuming that everyone shares the same $V(C)$ function, and that this function is increasing and at least weakly concave, i.e. that $V'(C) > 0$ and $V''(C) \leq 0$.

ξ represents the individuals' location preferences; that is, ξ_{ij} gives the utility that person i derives from living in jurisdiction j, with consumption held aside. In our two-jurisdiction case, it will simplify the analysis to define

$$x_i \equiv \xi_{i1} - \xi_{i2}$$

That is, if x_i is much greater than zero, person i has a strong preference for jurisdiction 1, if x_i is much less than zero, person i has a strong preference for jurisdiction 2, and if x_i is close to zero, person i has only a relatively weak preference for one jurisdiction over the other.

We assume that these preferences take moving costs into account, along with whatever other factors might be relevant. That is, we may suppose that each individual starts out in one jurisdiction or the other,

---

[18] It is worth noting tangentially that this situation is partially isomorphic to a Bertrand duopoly game (with differentiated products) in which two firms (with equal and constant marginal costs) divide up a market with perfectly inelastic demand (because everyone must live in one jurisdiction or the other). In this analogy, the taxes are like prices net of marginal costs, and the populations are like quantities. A major distinction between such a duopoly game and our current tax competition model lies in the particular role and structure of individual utilities, which we develop next.

and that where he starts has some bearing on his location preferences, because moving to the other jurisdiction entails some cost; by assumption, this cost is included in his location preference value x_i. For example, if he considers jurisdictions 1 and 2 to be equally desirable places ceteris paribus, but he starts in jurisdiction 1, and moving is costly, his ξ_{i1} is greater than his ξ_{i2}, and so his x_i is positive. In other words, his moving cost shows up in the model as decreasing his utility for the jurisdiction that he does not start in. It is common in the literature to conceive of location preference as being *solely* a product of moving cost, but in reality it is possible to prefer a place where one does not currently live. So, the approach used here adds generality.

Utility and willingness to pay for residence

Now that we have defined the utility function and its components, we may ask how its shape bears on the relationship between wealth and the additional tax that an individual is willing to pay to satisfy his location preference. So, let this willingness to pay be defined roughly as the maximum additional tax amount $\theta_j - \theta_k$ that an individual who prefers jurisdiction j will accept without moving to k. This concept will not play a direct role in any of the formal results in Section 3, but it is useful as a source of intuition.

First suppose a linear utility of consumption function $V(C) = C$, and consider an individual i who prefers jurisdiction 1 when taxes are equal, so that $\xi_{i1} > \xi_{i2}$ and $x_i > 0$. This person can be convinced to move to jurisdiction 2 if and only if $C_2 + \xi_{i2} > C_1 + \xi_{i1}$, or $C_2 - C_1 > x_i$. That is, jurisdiction 2 has to *add* a certain amount or more to his consumption in jurisdiction 1 in order to acquire him as a resident. Or in other words, i is willing to pay an additional tax of x_i in order to live in his preferred jurisdiction.

This can be interpreted in two potentially interrelated ways: The first possibility (which is the more literal interpretation of the mathematical expressions used here) is that the marginal utility of consumption and the total utility associated with location are both constant in income. The second possibility, which may be more broadly applicable, is that both types of utility decrease with income *at the same rate*. When does this make sense? The most likely explanation is that location preference is reducible to a simple monetary moving cost (because jurisdictions are viewed as being largely similar in other important respects, such as language, culture, social life, etc.), in which case the utility loss associated with this cost declines with income just as the marginal utility of consumption does.

Next, suppose a logarithmic utility of consumption function $V(C) = \ln C$, and consider once again a person with $x_i > 0$. Jurisdiction 2 can acquire this person if and only if $\ln C_2 + \xi_{i2} > \ln C_1 + \xi_{i1}$; this inequality can also be represented as $e^{\ln C_2} > e^{\ln C_1 + x_i}$, or $C_2/C_1 > e^{x_i}$. That is, jurisdiction 2 has to *multiply* his consumption in jurisdiction 1 by a certain amount or more in order to acquire him as a resident. When we compare a wealthy person and a poor person with the same location preference x_i, the wealthy person is willing to pay more in absolute terms to live his preferred jurisdiction, though he is willing to pay approximately the same proportion of his income.

Again, we have at least two possible interpretations. One is that location preference enters into the utility function as a constant, and marginal utility of consumption decreases at a rate consistent with logarithmic utility. An alternative that could yield similar results is that marginal utility of consumption is constant but moving cost increases with wealth in a way that is sufficiently rapid to be mathematically parallel to the first story.

Thus, although our mathematics use a locational utility that is constant in income and a marginal utility of consumption that is variable in income, the intuition of the model can easily be extended to cases with variable locational utility, by adjusting the utility of consumption function accordingly. What matters to the result is the rate at which marginal utility of consumption decreases in income *relative to* locational utility, which in turn can be described in terms of the types of reasons that individuals have for preferring one jurisdiction over another.

Location preference distribution and population shares

A person with wealth W decides to live in jurisdiction 1 or 2 depending on this inequality:

$$V\big(W-\theta_1(W)\big)+x_i \gtrless V\big(W-\theta_2(W)\big)$$

Therefore, the x_i value of a person who is precisely indifferent about his location is

$$x^*\big(\theta_j(W),\theta_k(W),W\big)=V\big(W-\theta_2(W)\big)-V\big(W-\theta_1(W)\big)$$

Note that, although we represent the jurisdictions generally as j and k when possible, such that j can be either jurisdiction 1 or 2, the location preference variables x_i and x^* are defined in such a way that it does matter which jurisdiction is which, so we use the subscripts 1 and 2 when this distinction must be made, as is the case immediately above.

Define $f(x_i|W)$ to be the density function of x_i, conditional on W, with support $x_i \in \mathbb{R}$. Define $F_1(x_i)$ and $F_2(x_i)$ to be the reliability function and cumulative distribution function of x_i, respectively, conditional on W. That is, define

$$F_1(x_i|W)=\int_{x_i}^{\infty} f(x|W)\,dx \qquad F_2(x_i|W)=\int_{-\infty}^{x_i} f(x|W)\,dx \qquad F_1(x_i|W)+F_2(x_i|W)=1$$

Therefore, $F_1\big(x^*\big(\theta_j(W),\theta_k(W),W\big)|W\big)$ is the share of people with wealth W who have x_i greater than $x^*\big(\theta_j(W),\theta_k(W),W\big)$, and thus choose to live in jurisdiction 1, once the taxes have been set. Likewise, $F_2\big(x^*\big(\theta_j(W),\theta_k(W),W\big)|W\big)$ is the share of people with wealth W who have x_i less than $x^*\big(\theta_j(W),\theta_k(W),W\big)$, and thus choose to live in jurisdiction 2, once taxes have been set.

We assume that the density function $f(x_i|W)$ is continuous in x_i,[19] which can be thought of in one of two ways: The first possibility is that the number of people at any given wealth level is infinite, or large enough that it can be treated as infinite for practical purposes. The second possibility is that there is a finite number of people, but the jurisdictions do not know the precise location preference of each person, only the distribution of location preferences. In this latter case, several variables must be expressed in probabilistic terms; for example, jurisdictions are maximizing expected revenue rather than revenue per se. Therefore, we will stick to the first interpretation for terminological purposes, while noting that the second interpretation leads to functionally equivalent results.

[19] A formal discussion is omitted here, but if we were to assume instead a discrete distribution of location preferences, consisting of a finite number of location preference ‘types’, pure strategy Nash equilibrium would not exist, except in a few unlikely cases.

Populations and revenues in terms of taxes

With the definitions of the boundary location preference value and the location preference distribution functions in hand, we may now write jurisdiction j's population of those with wealth W in terms of $x^*\big(\theta_j(W), \theta_k(W), W\big)$, and thus in terms of both jurisdictions' taxes:

$$\pi_j\big(\theta_j(W), \theta_k(W), W\big) = F_j\big(x^*\big(\theta_j(W), \theta_k(W), W\big)|W\big)\Pi(W)$$

$$\pi_j\big(\theta_j(W), \theta_k(W), W\big) = F_j\big(V\big(W - \theta_2(W)\big) - V\big(W - \theta_1(W)\big)|W\big)\Pi(W)$$

Similarly, jurisdiction j's revenue from those with wealth W can be expressed as

$$R_j\big(\theta_j(W), \theta_k(W), W\big) = \theta_j(W)F_j\big(V\big(W - \theta_2(W)\big) - V\big(W - \theta_1(W)\big)|W\big)\Pi(W)$$

Recall from earlier in this section that an equilibrium in this model is characterized by

$$\theta_j(W) = \underset{\theta_j(W)\leq W}{\operatorname{argmax}} R_j\big(\theta_j(W), \theta_k(W), W\big), \forall j = 1, 2; \forall W \geq \underline{W}$$

Now, we can write this condition more explicitly in terms of wealth W, the location preference distribution functions $F_j(x_i|W)$, and the utility of consumption function $V(C)$, as follows:

$$\theta_j(W) = \underset{\theta_j(W)\leq W}{\operatorname{argmax}} \theta_j(W)F_j\big(V\big(W - \theta_2(W)\big) - V\big(W - \theta_1(W)\big)|W\big)\Pi(W), \forall j = 1,2; \forall W \geq \underline{W}$$

Notational simplifications

In the following sections, we will simplify the notation by suppressing the arguments of the tax function $\theta_j(W)$, the revenue function $R_j\big(\theta_j(W), \theta_k(W), W\big)$, consumption $C_j\big(\theta_j(W), W\big)$, and the boundary location preference value $x^*\big(\theta_j(W), \theta_k(W), W\big)$, writing them instead as θ_j, R_j, C_j and x^*. We will partially simplify the expression of the location preference functions $f(x_i|W)$ and $F_j(x_i|W)$, writing them instead as $f(x_i)$ and $F_j(x_i)$. Thus, we use θ_j to indicate j's tax at the particular wealth level under consideration, $f(x_i)$ to indicate the location preference density function at the particular wealth level under consideration, etc.

We will also simplify by setting $\Pi(W) = 1$, which can be done without loss of generality in that it has no effect on either jurisdiction's tax rate – that is, $\Pi(W)$ multiplies each jurisdiction's revenue from those with wealth W by a constant, so it does not affect either jurisdiction's choice of $\theta_j(W)$. In this case, each jurisdiction's population function $\pi_j\big(\theta_j(W), \theta_k(W), W\big)$ can be represented as its population share, which we will write as $F_j(x^*)$.

3. Results

Proposition 1: *Given that the reliability function $F_1(x_i)$ and distribution function $F_2(x_i)$ are log-concave at all wealth levels, a pure strategy Nash equilibrium exists.*

Proposition 2: *Given that the reliability function $F_1(x_i)$ and distribution function $F_2(x_i)$ are log-concave at all wealth levels, the equilibrium is unique.*

Here we demonstrate existence and uniqueness of the pure strategy Nash equilibrium, given the added assumption that location preference distribution functions for both jurisdictions are log-concave. Bagnoli and Bergstrom (2005) show that a wide variety of commonly studied probability distributions (including the uniform distribution, the normal distribution, etc.) have log-concave density functions, and that this implies that they have log-concave distribution functions and reliability functions as well. Thus, our assumption is compatible with a fairly broad range of distributions, though it does rule out some potentially relevant ones, such as bimodal distributions.

The establishment of existence and uniqueness are standard first steps in seeking to understand the nature of an economic situation under investigation. Do we expect policies to converge to a stable point when all else is still, or do we expect them to adjust in response to one another in a way that forms an endless, restless cycle? If the former, is there only one stable point, or are there multiple possibilities, so that the final outcome is path-dependent? Here we demonstrate that there is a single set of stable tax policies, under conditions that are sufficient but by no means necessary.

In brief, the existence proof proceeds by showing that each jurisdiction's revenue function R_j is continuous with respect to both taxes θ_j and θ_k, and strictly quasi-concave with respect to its own tax θ_j. The uniqueness proof shows that the reaction functions of the two jurisdictions will cross only once. The remaining proofs make use of the jurisdictions' first-order conditions for revenue maximization, so we turn to this subject next.

First-order conditions for revenue maximization

Since each country's revenue function is $R_j = \theta_j F_j(x^*)$, the first order condition for maximization is

$$\frac{\partial R_j}{\partial \theta_j} = F_j(x^*) + \theta_j \frac{\partial F_j(x^*)}{\partial x^*}\frac{\partial x^*}{\partial \theta_j} = 0$$

That is, if the tax is revenue-maximizing, the gain from increasing the tax by one unit on the current population $F_j(x^*)$ must be just offset by the resulting decrease in the population, multiplied by the current tax rate. Because $x^* = V(W-\theta_2) - V(W-\theta_1)$, the first order condition can be re-written as

$$F_j(x^*) = \theta_j f(x^*) V'\left(W-\theta_j\right)$$

Intuitively, the revenue loss from population loss (right hand side) is greater when there are more people who are close to indifferent between the jurisdictions, and when the marginal utility of income is high (because this makes people more strongly averse to paying the extra tax).

Tax base elasticities

The economic meaning of this first order condition can be made more clear by considering it in terms of elasticities. So, define the tax base elasticity for jurisdiction j (and some wealth level W) as follows:

$$\varepsilon_j = \frac{\partial F_j(x^*)}{\partial \theta_j}\frac{\theta_j}{F_j(x^*)}$$

Evaluating the derivative, we can write the absolute value of the tax base elasticity more explicitly as

$$\left|\varepsilon_j\right| = f(x^*)V'\left(W-\theta_j\right)\frac{\theta_j}{F_j(x^*)}$$

Thus, a jurisdiction's tax base is more elastic ceteris paribus when there are more people who are close to indifferent between the two jurisdictions, when marginal utilities of income are higher, when the jurisdiction's tax is lower, or when its population is lower. Setting $|\varepsilon_j| = 1$ (a standard corollary of revenue maximization), we reproduce the first order condition.

Further, we will find that most of the results below can be explained intuitively in terms of this elasticity. Importantly, much of this intuition applies not only in our minimalist two-jurisdiction case, but also in the more general case with any number of jurisdictions greater than one.

The maximum tax constraint and Inada utility

For propositions 3-8, assume that we are considering wealth values such that the maximum tax constraint does not bind.

This is a minor assumption, which only needs to be stated as a formality. For example, if the utility of consumption function $V(C)$ satisfies the Inada condition[20] that $\lim_{C\to 0} V'(C) = \infty$, a corner solution where any jurisdiction taxes one hundred percent of any individual's wealth (i.e. sets $\theta_j = W$) could not exist in the equilibrium, because the other jurisdiction could poach all of its residents by offering even an arbitrarily small amount of after-tax consumption.

In these propositions, we focus mostly on constant relative risk aversion utility of consumption functions, which have marginal utility of the form $V'(C) = C^{-\rho}$; of these, all satisfy the $\lim_{C\to 0} V'(C) = \infty$ assumption, with the exception of the linear utility function $V(C) = C$, which corresponds to $\rho = 0$. We consider this utility function in Proposition 6, and find that it results in a head tax, i.e. a total tax that is independent of wealth. In this case, our assumption merely adds the exception that people can't be required to pay more than they possess. Given the other utility functions we consider, the extra assumption is not needed at all.

Proposition 3: *In an equilibrium, the jurisdiction with the larger per-type population of those with wealth W (after taxes and migration have been decided) also has the higher total tax on W.*

Intuitively, if more people want to live in jurisdiction 1 than jurisdiction 2 with taxes aside, this allows jurisdiction 1 to charge a higher tax, and also causes it to have a higher per-type population, ceteris paribus. However, higher taxes also lead to lower per-type populations, so we may wonder which effect is stronger. What this proof demonstrates is that the primary effect (greater popularity causing higher population) dominates the secondary effect (greater popularity causing higher taxes, causing lower population) under fairly general circumstances, i.e. whenever both jurisdictions' revenue-maximizing taxes satisfy the standard first order conditions. This result suggests a natural empirical test: Do jurisdictions with more rich people tend to have higher taxes on the rich?

We can also reproduce this result using tax base elasticities. That is, setting $|\varepsilon_j| = |\varepsilon_k|$ (since both are equal to one when the first order conditions hold), we obtain the condition

$$\frac{\theta_j V'(W-\theta_j)}{F_j(x^*)} = \frac{\theta_k V'(W-\theta_k)}{F_k(x^*)}$$

[20] From Inada (1963).

Thus, if one jurisdiction has a higher tax (and thus also a higher marginal utility of consumption for those paying the tax), it must also have a higher population. Note that this logic applies regardless of the number of jurisdictions.

Proposition 4: *In an equilibrium, a lower density of location preferences at the point of indifference, i.e. a lower $f(x^*)$, is associated ceteris paribus with a higher total tax in both jurisdictions.*

This is a common-sense result. Intuitively, if there are fewer people with weak location preferences, both jurisdictions will be able to charge higher taxes in the equilibrium.

This can also be seen as an immediate consequence of the form taken by the tax base elasticity. That is, since $\left|\varepsilon_j\right| = f(x^*)V'\left(W-\theta_j\right)\frac{\theta_j}{F_j(x^*)}$, a lower value of $f(x^*)$ implies that both jurisdictions face a more elastic tax base ceteris paribus. This in turn means that the revenue maximization condition $\left|\varepsilon_j\right| = 1$ will be reached at a lower value of j's tax θ_j.

Definition of symmetrical jurisdictions

'Symmetrical jurisdictions' will be defined to mean that the location preferences density function $f(x_i)$ is symmetrical across zero at all wealth levels. When this is true, the reaction functions $\theta_j\langle\theta_k\rangle$ and $\theta_k\langle\theta_j\rangle$ must be identical, and so the equilibrium taxes θ_j and θ_k must be identical as well. (This follows from Proposition 2; that is, if there were an equilibrium in which θ_j and θ_k were not equal, exchanging their values would also have to be an equilibrium, but this is ruled out by the uniqueness result.) Therefore, populations are equal as well, i.e. $F_j(x^*) = F_k(x^*)$. This assumption of symmetrical jurisdictions is used to simplify the mathematics in propositions 5 and 8 below.

Independence of wealth and the location preferences distribution

For propositions 5-8, assume that the location preference distribution $f(x_i)$ is the same for all wealth levels $W \geq \underline{W}$ under analysis. In other words, assume that $f(x_i|W) = f(x_i|W'), \forall x_i \in \mathbb{R};\ \forall W, W' \geq \underline{W}$.

This is a major assumption, which is important to the remaining results. That is, we will now begin to consider how the equilibrium tax rates vary with wealth, so we need some prior idea of how the location preference distribution at one wealth level compares to the location preference distribution at another. Here we begin with the simplest and clearest possible model by assuming that they are the same.

What intuitive reality does this assumption correspond to? First, it suggests that if a jurisdiction has any special benefits, they can be accessed by people of all income levels. For example, a town that was attractive because of a public swimming lake would fit this assumption well, but one that was attractive because of a lake owned by an expensive members-only club would not. For another example, the assumption fits well to the extent that the beauty of a city can be observed from public spaces (as is true of the exterior architecture of buildings that make up a skyline), and less well to the extent that it can only be observed from privileged private spaces (such as the interior architecture of expensive residences).

To the extent that location preferences are driven by individuals' desire to be close to friends and family, this assumption suggests that this desire is felt about equally by the rich and the poor. Further, it suggests that if moving cost is a major component of location preference, the prior distribution of income in the two jurisdictions is similar. For example, if jurisdiction 1 has a wealthier population than jurisdiction 2, and moving is costly, this causes the location preference distribution $f(x_i)$ to have greater density in the positive range $x_i > 0$ at higher wealth levels than at lower wealth levels.

In Propositions 5-8, we hold location preference constant in wealth in order to illustrate important ceteris paribus relationships, especially between the shape of the utility of consumption function and the progressivity of the equilibrium tax. In the surrounding discussion, we also consider the effects of relaxing this assumption, i.e. allowing location preference to vary directly with wealth.

Proposition 5: *Given symmetrical jurisdictions, the marginal tax rate is non-negative and less than one in an equilibrium, i.e.* $\frac{d\theta_j}{dW} \in [0, 1)$.

Roughly speaking, the non-negativity of marginal tax rates follows from the assumption that $V''(C) \leq 0$. That is, if marginal utility of income is non-increasing in income, willingness to pay to satisfy location preference will be non-decreasing, and so it is logical that (total) taxes will be non-decreasing in the equilibrium.

In terms of the tax base elasticities, since $|\varepsilon_j| = f(x^*)V'(W - \theta_j)\frac{\theta_j}{F_j(x^*)}$, greater wealth is associated with lower marginal utility of consumption, which implies lower elasticity (except in the case of linear utility, where wealth has no impact on elasticity).

If we were to relax our assumption that the location preference distribution $f(x_i)$ is independent of wealth W, negative marginal tax rates would be possible (e.g. in a case where location preferences were weaker at higher wealth levels, and the utility of consumption function were only minimally concave), as could marginal tax rates in excess of unity (e.g. in a case where location preferences were stronger at higher wealth levels, and the utility of consumption function were highly concave).

Proposition 6: *Given a linear utility of consumption function $V(C) = C$, equilibrium total taxes are independent of wealth (a 'head tax').*

This result is similar to the boundary case of Proposition 5 (where the expression for $\frac{d\theta}{dW}$ simplifies to zero given $V''(C) = 0$), but is proven in a slightly more general setting, where jurisdiction symmetry is not assumed. Intuitively, if marginal utility of income is constant in income, willingness to pay for location preference is constant in income as well.

Given $V(C) = C$, the absolute value of the tax base elasticities simplifies to $|\varepsilon_j| = \frac{f(x^*)}{F_j(x^*)}\theta_j$, which is independent of wealth for any given value of the (total) tax θ_j, so long as location preferences are independent of wealth.

Alternatively, if $f(x^*)/F_j(x^*)$ increases directly with wealth (e.g. if wealthier people are more indifferent about their location), then elasticity $|\varepsilon_j|$ rises with wealth, and the total tax θ_j falls, making something more regressive than a head tax. Or, if $f(x^*)/F_j(x^*)$ decreases directly with wealth, then $|\varepsilon_j|$ falls, and θ_j rises, making something less regressive than a head tax.

Average tax rates

The last two propositions consider the relationship between average tax rates and wealth, thus exploring the conditions under which equilibrium taxes are progressive, regressive, and proportional. We define jurisdiction j's average tax rate on those with wealth W to be its tax θ_j, divided by W:

$$\tau_j \equiv \frac{\theta_j}{W}$$

Proposition 7: *Given a logarithmic utility of consumption function $V(C) = ln\,C$, the equilibrium average tax rates are independent of wealth (a 'flat tax').*

Like Proposition 6, Proposition 7 illustrates the consequence of a particular type of utility function, without an assumption of jurisdiction symmetry. Given $V(C) = \ln C$, the rich do not differ from the poor in terms of their proportional willingness to pay for the jurisdiction they prefer (as we discussed in Section 2), so it is logical that equilibrium taxes do not vary with wealth in proportional terms.

Given $V(C) = \ln C$, the absolute value of the tax base elasticity can be written as

$$|\varepsilon_j| = \frac{f(x^*)}{F_j(x^*)} \frac{\theta_j}{(W - \theta_j)}$$

Since the total tax can be expressed as the average tax rate times wealth, we can re-express this as

$$|\varepsilon_j| = \frac{f(x^*)}{F_j(x^*)} \frac{\tau_j}{(1 - \tau_j)}$$

Thus, for any given value of the average tax rate τ_j, the tax base elasticities are independent of wealth if location preferences are independent of wealth.

Alternatively, if $f(x^*)/F_j(x^*)$ increases directly with wealth, then elasticity $|\varepsilon_j|$ rises with wealth, and the average tax τ_j falls, making a regressive tax. Or, if $f(x^*)/F_j(x^*)$ decreases with wealth, then $|\varepsilon_j|$ falls, and τ_j rises, making a progressive tax.

Constant relative risk aversion utility of consumption functions

. Both the linear utility of wealth function $V(C) = C$ and the logarithmic utility of wealth function $V(C) = \ln C$ are special cases of constant relative risk aversion[21] (CRRA) functions, for which the marginal utility of consumption $V'(C)$ can be expressed as $C^{-\rho}$. $V(C) = \ln C$ corresponds to $\rho = 1$, and $V(C) = \frac{1}{1-\rho} C^{1-\rho}$ corresponds to any $\rho \neq 1$, including $\rho = 0$, which gives $V(C) = C$.

It is interesting to explore these utility functions in particular because we can characterize their concavity using a single parameter, ρ, and then determine the effect of this parameter on the progressivity of the tax code. ρ reflects the rate at which the marginal utility of consumption changes; a higher value of ρ means that marginal utility of consumption (relative to the benefits of living in a particular location) decreases more rapidly as consumption increases. Given greater concavity of the $V(C)$ function, wealthier people gain less utility from additional consumption, and thus are willing to pay more to live in the jurisdiction they prefer. This is illustrated in Proposition 8 below.

---

[21] They are called this because relative risk aversion, or $-C \cdot V''(C)/V'(C) = \rho$, is a constant.

Proposition 8: *In an equilibrium with symmetrical jurisdictions and a CRRA utility of consumption function such that* $V'(C)$ *has the form* $C^{-\rho}$, $\rho \in [0,1)$ *leads to a regressive tax structure (*$\frac{d\tau}{dW} < 0$*),* $\rho = 1$ *leads to a proportional tax structure (*$\frac{d\tau}{dW} = 0$*), and* $\rho > 1$ *leads to a progressive tax structure (*$\frac{d\tau}{dW} > 0$*).*

The strong assumption of symmetrical jurisdictions is used to simplify the mathematics and to produce tidy closed-form expressions for both $\frac{d\theta}{dW}$ and $\frac{d\tau}{dW}$, but we are confident that it is far from a necessary condition for the result. That is, we have already shown in propositions 6 and 7 that risk aversion coefficients of $\rho = 0$ and $\rho = 1$ lead respectively to head taxes (which are regressive) and flat taxes (which are proportional), without any reliance on a symmetry assumption.

Further, we can use tax base elasticities to explore the equilibria given other values of ρ, again without assuming symmetry. That is, with CRRA utility, the absolute value of these elasticities can be expressed as

$$|\varepsilon_j| = \frac{f(x^*)}{F_j(x^*)} \frac{\theta_j}{(W - \theta_j)^\rho}$$

Once again using the identity $\theta_j = \tau_j W$, we can rearrange to isolate the role of wealth in relation to the average tax rate:

$$|\varepsilon_j| = \frac{f(x^*)}{F_j(x^*)} \frac{\tau_j}{(1 - \tau_j)^\rho} W^{1-\rho}$$

In this equation, we see that for any given average tax rate τ_j, and holding location preferences constant, the tax base elasticity depends positively on wealth if and only if the risk aversion coefficient ρ is less than 1.

As we discussed above for the cases of risk aversion coefficients of 0 and 1, allowing the location preference distribution to vary directly with wealth affects the results in a straightforward manner: If preferences for the jurisdiction grow stronger with wealth, progressivity increases, and if they grow weaker with wealth, progressivity decreases. Here we will briefly formalize this for the CRRA case, assuming jurisdictional symmetry once again to avoid complications.

Defining $\Gamma \equiv F(x^*)/f(x^*)$ as the ratio of population shares to density at indifference, the first order condition is

$$\tau(1-\tau)^{-\rho} = \Gamma W^{\rho-1}$$

Taking the total derivative as in the proof of Proposition 8, but now allowing Γ to vary with W, we have

$$\{(1-\tau)^{-\rho} + \rho\tau(1-\tau)^{-\rho-1}\}d\tau = \left\{\frac{\partial\Gamma}{\partial W} W^{\rho-1} + (\rho-1)\Gamma W^{\rho-2}\right\} dW$$

Defining $\eta \equiv (\partial\Gamma/\partial W)(W/\Gamma)$ as the elasticity of Γ with respect to wealth, we can solve for $d\tau/dW$ and simplify to

$$\frac{d\tau}{dW} = \frac{[(\rho-1)+\eta]\Gamma W^{\rho-2}}{(1-\tau)^{-\rho} + \rho\tau(1-\tau)^{-\rho-1}}$$

Since the rest of the right hand side is positive, the sign of $d\tau/dW$ must be the same as the sign of $(\rho - 1) + \eta$.

That is, a higher value of Γ corresponds to less location preference density near indifference, and higher equilibrium taxes. Thus, because η measures the rate at which Γ increases directly with wealth, higher values of η lead to greater progressivity.

4. Conclusion

Several factors can either increase or decrease the progressivity of the equilibrium tax structures. For example, as we discuss in the Appendix 2, a social welfare function that places decreasing weight on the marginal consumption of increasingly wealthy people tends to increase progressivity. Further, since the location preferences of individuals depend in part on where they reside initially, location preferences may depend on wealth via the prior distribution of wealth in different jurisdictions. Thus, a jurisdiction with a higher current concentration of rich people may be able to sustain a more progressive tax. If taxes are used to finance public goods, the relative enjoyment of these goods by different income groups becomes a factor. And of course progressivity depends also on how the elasticity of labor supply (and other similar behavioral elasticities) varies with income.

Our approach emphasizes the idea that progressive tax codes are more likely to be optimal when marginal utilities of consumption decrease more rapidly relative to total locational utilities as wealth increases. Representing locational utility as constant for each individual, we have shown that utility of consumption functions with relative risk aversion coefficients greater than one lead toward equilibria with progressive taxes, and that those with relative risk aversion coefficients greater than one lead toward equilibria with regressive taxes, ceteris paribus.

The model can also be used to consider cases where locational utility varies with consumption. For example, to the extent that location preferences are driven by a simple monetary moving cost that is independent of wealth, total locational utility decreases with consumption along with the marginal utility of consumption. In our analysis, this can be represented by maintaining the assumption of constant locational utility but reducing the concavity of the utility of consumption function, so that the ratio of the total locational utility to the marginal utility of consumption is still represented accurately.

Thus, it is in cases where the impact of location on utility is roughly constant in income (e.g. where it is driven more by quality of life considerations than by moving costs that are reducible to a monetary value) where it is most valid to estimate the shape of our utility of consumption function using risk aversion estimates from the literature. For example, Friend and Blume (1975) find evidence for a relative risk aversion coefficient greater than one, as do Layard et al (2008), who provide an estimate of $\rho = 1.26$; under the specified conditions, both of these suggest an equilibrium with progressive taxation.

Appendix 1: Proofs and examples

Proof of Proposition 1 (existence of equilibrium): We consider both jurisdictions' taxes, θ_1 and θ_2, on a particular wealth value W. We want to demonstrate that there must be a θ_1, θ_2 pair such that

$$\theta_j = \underset{\theta_j \leq W}{\operatorname{argmax}} R_j \, , \forall j = 1, 2$$

As we showed in Section 2, if these equations hold for all wealth levels in the relevant range (i.e. $\forall W \geq \underline{W}$), the model is in equilibrium. Therefore, we may treat the competition for those with wealth W as if it were an independent game, in which each jurisdiction j's strategy is its tax θ_j, its payoff is its revenue R_j, and the relationship between taxes and revenues is mediated by the utility of consumption function $V(C)$ and the location preference distribution for that wealth level, $f(x_i)$. We want to show that such a game has a pure strategy Nash equilibrium.

In addition to the upper bound restriction on each jurisdiction's tax given by the maximum tax constraint $\theta_j \leq W$, we can also impose a lower bound of $\theta_j \geq 0$ without loss of generality, because a negative tax cannot be revenue-maximizing. Thus, we have a compact strategy space $\theta_j \in [0, W]$, so we can demonstrate the existence of a pure strategy Nash equilibrium by showing that each jurisdiction j's payoff (revenue, R_j) is quasi-concave in its own strategy (tax, θ_j) and continuous in both strategies (taxes, θ_j and θ_k), as held e.g. in Fudenberg and Tirole (1991).

As developed in Section 2, the revenue function for jurisdiction j is

$$R_j = \theta_j F_j(x^*)$$

It is straightforward to show that each revenue function is continuous in each tax, by taking the derivatives below, and verifying that they are finite:

$$\frac{\partial R_j}{\partial \theta_j} = F_j(x^*) - \theta_j f(x^*) V'\left(W - \theta_j\right)$$

$$\frac{\partial R_j}{\partial \theta_k} = \theta_j f(x^*) V'(W - \theta_k)$$

The second equation here also shows us that if jurisdiction j increases its tax, this has a direct effect of increasing jurisdiction k's revenue, which may be called a positive fiscal externality.

Next, making use of the assumption that $F_1(x_i)$ and $F_2(x_i)$ are log-concave, we will show that $\ln R_1$ and $\ln R_2$ are strictly concave in the taxes θ_1 and θ_2, respectively, which in turn implies that the revenue functions R_1 and R_2 are strictly quasi-concave in the same variables.

$$\ln R_j = \ln \theta_j + \ln F_j(x^*)$$

Note that the sum of two concave functions is another concave function. Thus, because $\ln \theta_j$ is concave in θ_j, it only remains to show that the second right hand side term is concave as well. That is, defining j's 'log-distribution function' at wealth W as

$$\Phi_j(x_i) \equiv \ln F_j(x_i)$$

it will be sufficient to show that it is concave in the same jurisdiction's tax, i.e. $\frac{\partial^2 \Phi_j}{\partial^2 \theta_j} < 0$. By the definitions of the distribution functions $F_1(x_i)$ and $F_2(x_i)$, and by the log-concavity assumption,

$$\Phi_1'(x_i) < 0 \qquad \Phi_2'(x_i) > 0$$

$$\Phi_1''(x_i) < 0 \qquad \Phi_2''(x_i) < 0$$

Taking the first and second derivatives of either Φ_j with respect to θ_j, we find

$$\frac{\partial \Phi_j}{\partial \theta_j} = \Phi_j'(x^*) \frac{\partial x^*}{\partial \theta_j}$$

$$\frac{\partial^2 \Phi_j}{\partial^2 \theta_j} = \Phi_j''(x^*)\left(\frac{\partial x^*}{\partial \theta_j}\right)^2 + \Phi_j'(x^*)\frac{\partial^2 x^*}{\partial^2 \theta_j}$$

In Section 2, we found that $x^* = V(W - \theta_2) - V(W - \theta_1)$. Evaluating the first and second derivatives of x^* with respect to each tax rate, we find

$$\frac{\partial x^*}{\partial \theta_1} = V'(W - \theta_1) > 0 \qquad \frac{\partial x^*}{\partial \theta_2} = -V'(W - \theta_2) < 0$$

$$\frac{\partial^2 x^*}{\partial^2 \theta_1} = -V''(W - \theta_1) \geq 0 \qquad \frac{\partial^2 x^*}{\partial^2 \theta_2} = V''(W - \theta_2) \leq 0$$

$$\frac{\partial^2 \Phi_j}{\partial^2 \theta_j} = \Phi_j''(x^*)\big(V'(W - \theta_1)\big)^2 + \left|\Phi_j'(x^*)\right| V''(W - \theta_2) < 0$$

Therefore, the second derivative of the log-distribution function $\frac{\partial^2 \Phi_j}{\partial^2 \theta_j}$ is negative for both jurisdictions, which implies that the log of each jurisdiction's revenue function $\ln R_j$ is concave in its own tax rate, which implies in turn that each jurisdiction's revenue function R_j is quasi-concave in its own tax rate. Therefore, a pure strategy Nash equilibrium exists. ■

Proof of Proposition 2 (uniqueness of equilibrium): Define $\theta_j\langle\theta_k\rangle$ to be jurisdiction j's reaction function, which represents the best response value of jurisdiction j's tax θ_j on wealth W given the other jurisdiction k's tax θ_k, and subject to the maximum tax constraint of $\theta_j \leq W$.

In cases such that the best response tax $\theta_j\langle\theta_k\rangle$ is not bound by this maximum tax constraint, it must satisfy the first order condition for maximizing revenue. To make use of the log-concavity assumption, we find a version of this first order condition by setting $\frac{\partial(\ln R_j)}{\partial \theta_j} = 0$:

$$\frac{1}{\theta_j} + \Phi_j'(x^*)\frac{\partial x^*}{\partial \theta_j} = 0$$

Taking the total derivative of this equation with respect to the taxes θ_j and θ_k, we find:

$$\left\{-\frac{1}{\theta_j^2} + \Phi_j''(x^*)\left(\frac{\partial x^*}{\partial \theta_j}\right)^2 + \Phi_j'(x^*)\frac{\partial^2 x^*}{\partial^2 \theta_j}\right\} d\theta_j + \left\{\Phi_j''(x^*)\frac{\partial x^*}{\partial \theta_j}\frac{\partial x^*}{\partial \theta_k}\right\} d\theta_k = 0$$

(There is an additional term equal to $\Phi_j'(x^*)\frac{\partial^2 x^*}{\partial \theta_j \partial \theta_k} d\theta_k$, but from the equation $x^* = V(W - \theta_2) - V(W - \theta_1)$, developed in Section 2, we can see that $\frac{\partial^2 x^*}{\partial \theta_j \partial \theta_k} = 0$.) Solving for the reaction function slope $\frac{d\theta_j\langle\theta_k\rangle}{d\theta_k}$, and evaluating the derivatives of x^*, we find

$$\frac{d\theta_j\langle\theta_k\rangle}{d\theta_k} = \frac{-\Phi_j''(x^*)V'(C_j)V'(C_k)}{-\Phi_j''(x^*)\left[V'(C_j)\right]^2 + \frac{1}{\theta_j^2} - \left|\Phi_j'(x^*)\right| V''(C_j)}$$

The sign of this expression is positive, which means that j's tax θ_j is increasing in k's tax θ_k. That is, whenever jurisdiction k charges a higher tax, jurisdiction j's optimal tax increases.

Define the 'graphical slope' of either reaction function as its slope when graphed in a θ_j, θ_k Cartesian space. Thus, the graphical slope of jurisdiction k's reaction function is $\frac{d\theta_k\langle\theta_j\rangle}{d\theta_j}$, and the graphical slope of jurisdiction j's reaction function is the multiplicative inverse of $\frac{d\theta_j\langle\theta_k\rangle}{d\theta_k}$, because its independent variable is represented by the vertical axis, but we are measuring the slope as the vertical change divided by the horizontal change. In order to demonstrate equilibrium uniqueness, we will show that, at any equilibrium point, the graphical slope of j's reaction function must be greater than the graphical slope of k's reaction function. Given this, and given that both reaction functions are continuous, they cannot cross each other more than once, which means that the Nash equilibrium must be unique.

We begin by rearranging $\frac{d\theta_j\langle\theta_k\rangle}{d\theta_k}$ into the following form:

$$\frac{d\theta_j\langle\theta_k\rangle}{d\theta_k} = \frac{V'(C_k)}{\left\{V'(C_j) + \frac{\frac{1}{\theta_j^2} - \left|\Phi_j'(x^*)\right|V''(C_j)}{-\Phi_j''(x^*)V'(C_j)}\right\}}$$

Next, to make this more visually manageable, we define Ψ_j and Ψ_k as follows:

$$\Psi_j \equiv \frac{\frac{1}{\theta_j^2} - \left|\Phi_j'(x^*)\right|V''(C_j)}{-\Phi_j''(x^*)V'(C_k)} \qquad \Psi_k \equiv \frac{\frac{1}{\theta_k^2} - \left|\Phi_k'(x^*)\right|V''(C_k)}{-\Phi_k''(x^*)V'(C_k)}$$

Note that both Ψ_j and Ψ_k are unambiguously positive. Using the definition of Ψ_j, we can express the graphical slope of jurisdiction j's reaction function as follows:

$$\frac{1}{\left[\frac{d\theta_j\langle\theta_k\rangle}{d\theta_k}\right]} = \frac{V'(C_j) + \Psi_j}{V'(C_k)}$$

Likewise, using the definition of Ψ_k, the graphical slope of k's reaction function is

$$\frac{d\theta_k\langle\theta_j\rangle}{d\theta_j} = \frac{V'(C_j)}{V'(C_k) + \Psi_k}$$

So, when neither jurisdiction's tax is bound by the maximum tax constraint, we can use the calculations above to see that the graphical slope of j's reaction function is greater than the graphical slope of k's reaction function at any equilibrium:

$$\frac{V'(C_j) + \Psi_j}{V'(C_k)} > \frac{V'(C_j)}{V'(C_k) + \Psi_k} \qquad \frac{1}{\left[\frac{d\theta_j\langle\theta_k\rangle}{d\theta_k}\right]} > \frac{d\theta_k\langle\theta_j\rangle}{d\theta_j}$$

When jurisdiction j's tax θ_j is bound by the maximum tax constraint, it must be true that the maximum tax of $\theta_j = W$ gives the highest revenue of available taxes, because j's revenue is quasi-concave with respect to its own tax (from Proposition 1). In this case, a marginal increase in θ_k has no impact on θ_j; that is, $\frac{d\theta_j\langle\theta_k\rangle}{d\theta_k} = 0$. Therefore, in our θ_j, θ_k space, j's reaction function when constrained is a vertical line, so its graphical slope is greater than any finite number. Likewise, when k is bound by the constraint, its reaction function is a horizontal line, with slope zero. Thus, whether one, both, or neither jurisdiction is constrained at the equilibrium, the graphical slope of j's reaction function is greater than the graphical slope of k's reaction function. Therefore, there can only be one equilibrium. ■

Figures 1 and 2 below give two examples of the intersecting reaction functions described in Proposition 2. Figure 1 illustrates an equilibrium in which the maximum tax constraint binds neither jurisdiction, and figure 2 illustrates one in which the constraint binds both jurisdictions. In both cases, the graphical slope of j's reaction function is greater than the graphical slope of k's reaction function at their intersection (equilibrium) point; in figure 2, the graphical slopes of j's and k's reaction function at the equilibrium point are infinity and zero, respectively.

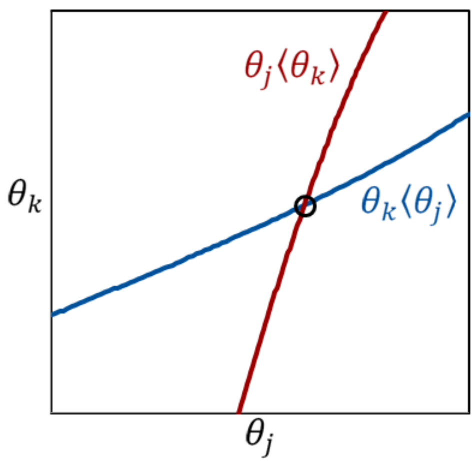

Figure 1: Reaction functions with maximum tax constraint binding neither jurisdiction

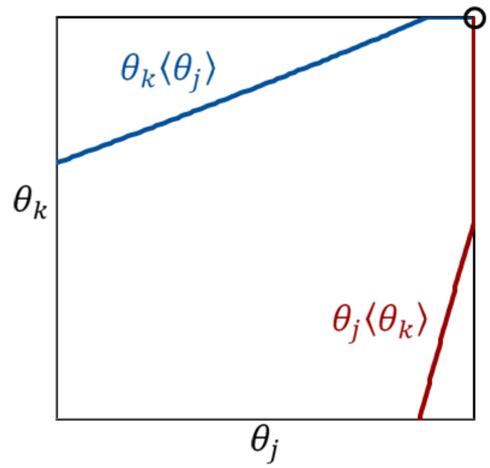

Figure 2: Reaction functions with maximum tax constraint binding both jurisdictions

Proof of Proposition 3 (populations and taxes): From Section 2, we know that

$$x^* = V(W-\theta_2) - V(W-\theta_1) \qquad\qquad R_j = \theta_j F_j(x^*)$$

Setting the derivatives of a revenue function $\frac{\partial R_j}{\partial \theta_j}$ equal to zero, we get the condition

$$F_j(x^*) = \theta_j f(x^*) V'(W-\theta_j)$$

We can divide jurisdiction j's first order condition by jurisdiction k's first order condition to get this equation, which must be true in an equilibrium not bound by the maximum tax constraint:

$$\frac{F_j(x^*)}{F_k(x^*)} = \frac{\theta_j V'(W-\theta_j)}{\theta_k V'(W-\theta_k)}$$

Thus, given any weakly concave utility of consumption function $V(C)$, $\theta_j > \theta_k$ implies $F_j(x^*) > F_k(x^*)$. That is, in the equilibrium, the jurisdiction with the higher tax on a given wealth type is also the jurisdiction that has more residents who belong to that type. ■

Proof of Proposition 4 (location preference density and taxes): This follows from the first order conditions developed in Proposition 3 above. That is, they can be rearranged in the following form:

$$\frac{F_j(x^*)}{f(x^*)} = \theta_j V'(W-\theta_j)$$

From this we can see that, holding constant the population shares $F_j(x^*)$, a higher tax in each jurisdiction is clearly associated with a lower density of location preferences close to the point of indifference. (This is intuitive: when more people have strong preferences about where they live, jurisdictions can perform more revenue extraction.) ■

Proof of Proposition 5 (marginal tax rates non-negative and below unity): As in Proposition 2, we begin by taking the total derivative of the first order condition $\frac{1}{\theta_j} + \Phi_j'(x^*)\frac{\partial x^*}{\partial \theta_j} = 0$. We now allow wealth to vary as well, but we assume that the location preference distribution $f(x_i)$, and thus the related distribution $\Phi_j(x_i)$, do not depend directly on wealth. We use the symmetry assumption on the first line, in the form of $C \equiv C_j = C_k$ and $\Phi \equiv \Phi_j = \Phi_k$.

$$\left\{\Phi''(x^*)[V'(C)]^2 + |\Phi'(x^*)|V''(C) - \frac{1}{\theta^2}\right\} d\theta_j + \{-\Phi''(x^*)V'(C)^2\}d\theta_k + \{-|\Phi'(x^*)|V''(C)\}dW = 0$$

Rewriting this in the form $\mathcal{A}d\theta_j + \mathcal{B}d\theta_k + \mathcal{C}dW$, we rearrange to find $\frac{d\theta_j}{dW} = -\frac{\mathcal{B}}{\mathcal{A}}\frac{d\theta_k}{dW} - \frac{\mathcal{C}}{\mathcal{A}}$. Then, we can use the symmetry assumption in the form of $d\theta_j = d\theta_k$ to solve for $\frac{d\theta}{dW}$ as follows:

$$\frac{d\theta}{dW} = \frac{-\mathcal{C}}{\mathcal{A}+\mathcal{B}} = \frac{-|\Phi'(x^*)|V''(C)}{-|\Phi'(x^*)|V''(C) + \frac{1}{\theta^2}} \in [0, 1) \qquad ■$$

Proof of Proposition 6 (linear utility and head taxes): Given $V(C) = C$, the boundary location preference value is

$$x^* = \theta_1 - \theta_2$$

Thus, the revenue functions are

$$R_j = \theta_j F_j(\theta_1 - \theta_2)$$

Setting the derivatives of the revenue functions $\frac{\partial R_j}{\partial \theta_j}$ equal to zero, we get

$$F_j(\theta_1 - \theta_2) = \theta_j f(\theta_1 - \theta_2)$$

From this, we can see that if the location preference distribution doesn't depend directly on wealth – i.e. if $f(x_i)$ and $F_j(x_i)$ are independent of wealth – the linear utility function $V(C) = C$ leads to equilibrium conditions such that the taxes θ_j are independent of wealth. ■

Example: Suppose that there is a linear utility of consumption function $V(C) = C$, and a uniform distribution of location preferences $x_i \sim \mathcal{U}[-B, A]$. The location preference distribution can be expressed by

$$F_1(x^*) = \frac{A - x^*}{A + B} \qquad F_2(x^*) = \frac{B + x^*}{A + B} \qquad f(x^*) = \frac{1}{A + B}$$

Plugging these in to the first order conditions, we can find the reaction functions; from these we can find the taxes and the per-type populations:

$$\theta_1\langle\theta_2\rangle = \frac{1}{2}\theta_2 + \frac{1}{2}A \qquad \theta_2\langle\theta_1\rangle = \frac{1}{2}\theta_1 + \frac{1}{2}B$$

$$\theta_1 = \frac{2}{3}A + \frac{1}{3}B \qquad \theta_2 = \frac{1}{3}A + \frac{2}{3}B$$

$$x^* = \frac{1}{3}A - \frac{1}{3}B \qquad F_1(x^*) = \frac{1}{A+B}\left(\frac{2}{3}A + \frac{1}{3}B\right) \qquad F_2(x^*) = \frac{1}{A+B}\left(\frac{1}{3}A + \frac{2}{3}B\right)$$

Thus, if $A > B$, jurisdiction 1 has both a higher tax and more people with wealth W, whereas if $B > A$, the opposite is true. (This is an example of Proposition 3's result.) In the symmetrical case, with $A = B$, the equilibrium taxes are $\theta_1 = \theta_2 = A$. Intuitively, if more people prefer jurisdiction j ceteris paribus, j can charge higher taxes. Furthermore, both jurisdiction's tax rates depend positively on the overall spread of the location distribution function. (This is an example of Proposition 4's result.)

Proof of Proposition 7 (logarithmic utility and flat taxes): Here we first find the boundary value of x_i, and then observe that, as the taxes θ_j can be re-written as $\tau_j W$, the W terms drop out of the equation:

$$x^* = \ln(W - \theta_2) - \ln(W - \theta_1)$$

$$x^* = \ln(1 - \tau_2) - \ln(1 - \tau_1)$$

Expressing the revenue functions in terms of the average tax rates τ_j, we have

$$R_j = \tau_j W F_j(x^*)$$

Setting the revenue function derivatives $\frac{\partial R_j}{\partial \tau_j}$ to zero, we obtain

$$F_j(x^*) = \frac{\tau_j}{1 - \tau_j} f(x^*)$$

So long as $F_j(x_i)$ and $f(x_i)$ are independent of W, these first order conditions are independent of W also, so we have average tax rates that are independent of wealth when they hold. ■

Example: Suppose a logarithmic utility function $V(C) = \ln C$, and symmetrical jurisdictions. Given symmetrical jurisdictions, $F_1(x^*) = F_2(x^*) = \frac{1}{2}$, and $x^* = 0$, so we have

$$\frac{1}{2} = \frac{\tau_j}{1 - \tau_j} f(0)$$

$$\tau_j = \frac{1}{2f(0) + 1}$$

Example: Suppose a logarithmic utility function $V(C) = C$, and a uniform location preference distribution $x_i \sim \mathcal{U}[-B, A]$. Applying the location preference distribution to the first order conditions, we find

$$A - x^* = \frac{\tau_1}{1 - \tau_1} = \frac{\theta_1}{W - \theta_1} \qquad B + x^* = \frac{\tau_2}{1 - \tau_2} = \frac{\theta_2}{W - \theta_2}$$

From these, it should be clear that a higher population in jurisdiction 1 must go with a higher average tax rate, and a higher tax. In the symmetrical case with $A = B$, $f(0) = \frac{1}{2A}$, so the average tax rates are $\tau_1 = \tau_2 = \frac{A}{A+1}$, and the taxes themselves are $\theta_1 = \theta_2 = \frac{A}{A+1}W$.

Proof of Proposition 8 (relative risk aversion and progressivity): From the first order conditions for revenue maximization, we know that in this symmetrical case,

$$\theta f(0) V'(W - \theta) = \frac{1}{2}$$

Defining $\Gamma \equiv F(x^*)/f(x^*) = \frac{1}{2f(0)}$, and assuming $V'(C) = C^{-\rho}$, the same condition can be represented as

$$\theta(W-\theta)^{-\rho} = \Gamma$$

In order to determine the impact of wealth (W) on the tax (θ), we can take the total derivative of this equation and solve for $\frac{d\theta}{dW}$ as follows. (Here we assume as specified that $f(x_i)$, and thus Γ, are independent of W.)

$$\{(W-\theta)^{-\rho} + \rho\theta(W-\theta)^{-\rho-1}\}d\theta - \{\rho\theta(W-\theta)^{-\rho-1}\}dW = 0$$

$$\frac{d\theta}{dW} = \frac{\rho\theta(W-\theta)^{-1}}{\rho\theta(W-\theta)^{-1} + 1}$$

Evaluating this expression, we find that it is zero when $\rho = 0$ (as we learned in Proposition 6), that it is positive for all $\rho > 0$, and that it is negative when $\rho < 0$ and $\rho > -\frac{W-\theta}{\theta}$. That is, given $\rho > 0$, equilibrium taxes increase with wealth, and given some values such that $\rho < 0$ (i.e. a 'risk-loving' case such that $V''(C) > 0$, which we assume away in this paper), equilibrium taxes actually decrease as wealth increases. Note that this can be used as an alternate proof of Proposition 5 for the special case of CRRA utility of consumption functions.

To determine the impact of wealth (W) on tax as a fraction of wealth (τ), we can perform a similar analysis, after re-expressing the equilibrium condition in terms of τ instead of θ:

$$\tau(1-\tau)^{-\rho} = \Gamma W^{\rho-1}$$

$$\{(1-\tau)^{-\rho} + \rho\tau(1-\tau)^{-\rho-1}\}d\tau = \{(\rho-1)\Gamma W^{\rho-2}\}dW$$

$$\frac{d\tau}{dW} = \frac{(\rho-1)\Gamma W^{\rho-2}}{(1-\tau)^{-\rho} + \rho\tau(1-\tau)^{-\rho-1}}$$

Evaluating this expression, we find that it is zero when $\rho = 1$ (as we learned in Proposition 7), that it is positive for all $\rho > 1$, and that it is negative when $\rho < 1$ and $\rho > -\frac{W-\theta}{\theta}$. That is, $V(C) = \ln C$ is the boundary case between valuation functions that cause equilibrium average tax rates to be decreasing in wealth versus increasing in wealth, or in other words, it is the boundary between regressive and progressive taxation, within the set of wealth types that are subject to revenue-maximizing taxes. ■

Appendix 2: Utilitarian social welfare maximizing governments

In the main body of the paper, we assume that the objective of each government is to maximize revenue from each wealth type above a particular threshold (which leads to similar results as a maximin social welfare function). Here, let us briefly consider a utilitarian social welfare function as an alternative objective. This analysis is more complicated, so we cannot use the methods above to show equilibrium existence, uniqueness, etc., but we can give preliminary consideration to the relationship between individuals' relative risk aversion and the progressivity of equilibrium taxes.

I define utilitarian social welfare maximization as follows. When dealing with a particular wealth type, each jurisdiction j has two competing objectives: the first is to take in revenue, and the second is to allow the people of this wealth type (regardless of their jurisdiction of origin)[22] to enjoy high utility. We will be general about what the revenue is used for, supposing only that each unit has a shadow value represented by λ_j. Thus, each jurisdiction j's problem can be written as

$$\max_{\theta_j} \Upsilon_j = \lambda_j R_j + U$$

Here, Υ_j is the jurisdiction's objective function, and U is the sum of all utility for those of this particular wealth type, whether they prefer to live in jurisdiction j or jurisdiction k. Further, U includes both the utility that people receive from consumption, and the utility that they receive directly from living in their jurisdiction of residence.

Without loss of generality, we can monotonically transform each person's utility function so that his location utility from jurisdiction 2 (ξ_{i2}) is set to zero, and his location utility from jurisdiction 1 (ξ_{i1}) is adjusted accordingly, so that the difference $x_i = \xi_{i1} - \xi_{i2}$ is unchanged. In this case, we can write jurisdiction j's objective function Υ_j as follows:

$$\Upsilon_j = \lambda_j\left[\pi_j \theta_j\right] + V(C_1)\pi_1 + V(C_2)\pi_2 + \Pi \int_{x^*}^{\infty} x_i f(x_i)\, dx_i$$

Taking the derivative of this objective function with respect to the jurisdiction's own tax θ_j, we have

$$\frac{\partial \Upsilon_j}{\partial \theta_j} = \lambda_j\left[\pi_j + \theta_j \frac{\partial \pi_j}{\partial \theta_j}\right] - V'(C_j)\pi_j + V(C_1)\frac{\partial \pi_1}{\partial \theta_j} + V(C_2)\frac{\partial \pi_2}{\partial \theta_j} - p x^* f(x^*)\frac{\partial x^*}{\partial \theta_j}$$

The last three of the five terms on the right hand side sum to zero. This is intuitive: these terms quantify the effect on utility of people moving from one jurisdiction to the other as a result of the tax, but on the margin the people moving are just indifferent between the jurisdictions. Setting the derivative of the objective function equal to zero, we obtain the first order condition:

$$\frac{\partial \Gamma_j}{\partial \theta_j} = \lambda_j \frac{\partial R_j}{\partial \theta_j} - V'(C_j)\pi_j \overset{\text{set}}{=} 0$$

$$\lambda_j\left[\pi_j + \theta_j \frac{\partial \pi_j}{\partial \theta_j}\right] - V'(C_j)\pi_j = 0$$

$$\lambda_j\left[F_j(x^*) - \theta_j f(x^*) V'(W - \theta_j)\right] - V'(W - \theta_j) F_j(x^*) = 0$$

This sets the shadow value of revenue, multiplied by the increase in revenue, equal to the loss in consumption utility experienced by those living in the jurisdiction. Of course, in the limit as λ_j goes to infinity, this becomes equivalent to the first order condition for revenue maximization. Given that the value of consumption function $V(C)$ is strictly concave, a shift from revenue-maximizing governments to utility-maximizing governments will tend to add progressivity to the tax code ceteris paribus, because the marginal value of consumption term $V'(C_j)\pi_j$ has greater weight when consumption is lower.

Unlike the revenue maximizing tax, the social welfare maximizing tax on those with any given wealth value W may be either positive or negative. To see this, note that although the first order condition for revenue maximization, $\pi_j + \theta_j \frac{\partial \pi_j}{\partial \theta_j} = 0$ requires a non-negative value of θ_j, the first order condition for

[22] Again, there are various possible ways in which a country's concern (or lack of concern) for migrants' utility can be modeled, but this assumption seems as straightforward (and ethically defensible) as any, so we adhere to it for the purpose of this preliminary discussion.

social welfare maximization, $\lambda_j \frac{\partial R_j}{\partial \theta_j} = V'(C_j)\pi_j$, may require a value of C_j that is either greater than or less than W. Thus, we are likely to end up with something similar to the 'negative income tax' proposed by Friedman (1962), combining a lump-sum transfer with positive marginal tax rates.

Example: Suppose a CRRA utility of consumption function, and symmetrical jurisdictions. It is slightly easier to compare the social welfare maximization case to the revenue maximization case if we define $\mu_j \equiv 1/\lambda_j$ as the multiplicative inverse of the shadow value of revenue, and multiply the first order condition by this value to obtain the equivalent expression

$$\frac{\partial R_j}{\partial \theta_j} - \mu_j V'(C_j) = 0$$

If we assume symmetrical jurisdictions, and CRRA utility of consumption functions, we have

$$\frac{1}{2} - \theta f(0)(W-\theta)^{-\rho} - \frac{1}{2}\mu(W-\theta)^{-\rho} = 0$$

As in Proposition 8, we define $\Gamma \equiv \frac{1}{2f(0)}$ for visual ease. (For example, given a uniform distribution of location preferences $x_i \sim \mathcal{U}[-A, A]$, we have $\Gamma = A$.) Using the equivalence $\frac{1}{2} = \Gamma f(0)$, we can rewrite the first order condition as

$$(\theta + \mu\Gamma)(W-\theta)^{-\rho} = \Gamma$$

Defining $\tau = \theta/W$ as the average tax rate, we can also express this condition as

$$(\tau W^{-\rho+1} + \mu\phi W^{-\rho})(1-\tau)^{-\rho} = \Gamma$$

Taking the total derivative with respect to τ and W, and solving for $d\tau/dW$, we find

$$\frac{d\tau}{dW} = -\frac{\tau(-\rho+1) - \mu\Gamma\rho W^{-1}}{1 + \tau W\rho(1-\tau)^{-1} + \mu\Gamma\rho(1-\tau)^{-1}}$$

For wealth values such that the average tax rate τ is positive, the denominator is positive, so we can determine the sign of this expression by examining the numerator. This shows us that the tax is proportional ($d\tau/dW = 0$) when

$$\rho = \frac{\tau W}{\tau W + \mu\Gamma}$$

Likewise, it is progressive when the risk aversion coefficient ρ is greater than the right hand side of the equation, and regressive when ρ is less than the right hand side of the equation. If $\mu = 0$, we return to our earlier (Proposition 8) result that the boundary value of ρ is 1. However, given positive τ and μ (i.e. positive taxes, and some concern for the welfare of those being taxed), the value of ρ leading to proportional taxation is less than 1.

To make the example more specific and concrete, if we suppose $\rho = 1$, the total tax can be expressed as

$$\theta = \frac{\Gamma}{1+\Gamma}(W - \mu)$$

This indicates a lump sum transfer of $\Gamma\mu/(1+\Gamma)$ for all individuals, and then a constant marginal tax rate of $\Gamma/(1+\Gamma)$. The constant marginal rate following from the utility function $V(C) = \ln C$ is the same as in the paper's main model, but here the lump sum transfer makes the tax system progressive overall.

References

Bagnoli, Mark and Ted Bergstrom. 2005. Log-concave probability and its applications. *Economic Theory* 26: 445-469.

Bierbrauer, Felix, Craig Brett, and John Weymark. 2011. Strategic nonlinear income tax competition with perfect labor mobility. Forthcoming in *Games and Economic Behavior*.

Blumkin, Tomer, Efraim Sadka, and Yotam Shem-Tov. 2012. International tax competition: Zero tax rate at the top re-established. CESifo working paper No 3820.

Brennan, Geoffrey and James Buchanan. 1980. *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. Cambridge University Press.

Brewer, Mike, Emmanuel Saez, and Andrew Shephard. 2010. Means-testing and tax rates on earnings. In *Dimensions of Tax Design: The Mirrlees Review*, James Mirrlees, ed., 90-173. Oxford University Press.

Brown, Charles and Wallace Oates. 1987. Assistance to the poor in a federal system. *Journal of Public Economics* 32: 307-330.

Brueckner, Jan. 2003. Strategic interaction among governments: An overview of empirical studies. *International Regional Science Review* 26 (2): 175-188.

Cremer, Helmuth and Pierre Pestieau. 2004. Factor mobility and redistribution. *Handbook of Regional and Urban Economics* 4: 2529-2560.

Diamond, Peter. 1980. Income taxation with fixed hours of work. *Journal of Public Economics* 13: 101-110.

Diamond, Peter. 1998. Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates. *American Economic Review* 88 (1): 83-95.

Diamond, Peter and Emmanuel Saez. 2011. The case for a progressive tax: From basic research to policy recommendations. *Journal of Economic Perspectives* 25 (4): 165-190.

Epple, Dennis and Thomas Romer. 1991. Mobility and redistribution. *Journal of Political Economy* 99 (4): 828-858.

Feldstein, Martin. 1995. The effect of marginal tax rates on taxable income: A panel study of the 1986 Tax Reform Act. *Journal of Political Economy* 103: 551-72.

Friedman, Milton. 1962. *Capitalism and Freedom*. University of Chicago Press.

Friend, Irwin and Marshall Blume. 1975. The demand for risky assets. *American Economic Review* 65 (5): 900-922.

Fudenberg, Drew and Jean Tirole. 1991. *Game Theory*. MIT Press.

Goolsbee, Austan. 1999. Evidence on the high income Laffer curve from six decades of tax reform. *Brookings Papers on Economic Activity*, 1999 (2): 1-64.

Harms, Philipp and Stefan Zink. 2003. Limits to redistribution in a democracy: A survey. *European Journal of Political Economy* 19: 651-668.

Hindriks, Jean. 1999. The consequences of labour mobility for redistribution: Tax vs. transfer Competition. *Journal of Public Economics* 74: 215-234.

Inada, Ken-Ichi. 1963. On a two-sector model of economic growth: Comments and a generalization. *Review of Economic Studies* 30 (2): 119-127.

Layard, Richard, Guy Mayraz, and Stephen Nickell. 2008. The marginal utility of income. *Journal of Public Economics* 92: 1846-1857.

Lehmann, Etienne, Laurent Simula, and Alain Trannoy. 2013. Tax me if you can! Optimal nonlinear income tax between competing governments. Forthcoming in the *Quarterly Journal of Economics*.

Meltzer, Allan and Scott Richard. 1981. A rational theory of the size of government. *Journal of Political Economy* 89 (5): 914-927.

Morelli, Massimo, Huanxing Yang, and Lixin Ye. 2012. Competitive nonlinear taxation and constitutional choice. *American Economic Journal: Microeconomics* 4 (1): 142-175.

Mirrlees, James. 1971. An exploration in the theory of optimum income taxation. *Review of Economic Studies* 38 (2): 175-208.

Mirrlees, James. 1982. Migration and optimal income taxes. *Journal of Public Economics* 18 (3): 319-341.

Musgrave, Richard. 1971. Economics of fiscal federalism. *Nebraska Journal of Economics and Business* 10 (4): 3–13.

Oates, Wallace. 1972. *Fiscal Federalism*. Harcourt Brace Jovanovich.

Rawls, John. 1971. *A Theory of Justice*. Belknap Press.

Saez, Emmanuel. 2002. Optimal income transfer programs: Intensive versus extensive labor supply responses. *Quarterly Journal of Economics* 117 (3): 1039-1073.

Simula, Laurent and Alain Trannoy. 2010. Optimal income tax under the threat of migration by top-income earners. *Journal of Public Economics* 94: 163-173.

Sinn, Hans-Werner. 1990. Tax harmonization and tax competition in Europe. *European Economic Review* 34: 489–504.

Stigler, George. 1957. The tenable range of functions of local government. In *Federal Expenditure Policy for Economic Growth and Stability*, U.S. Congress, Joint Economic Committee, ed., 213–219.

Tiebout, Charles. 1956. A pure theory of local expenditures. *Journal of Political Economy* 64 (5): 416-424.

Von Neumann, John and Oskar Morgenstern. 1953. *Theory of Games and Economic Behavior*. Princeton University Press.

Wilson, John. 1980. The effect of potential emigration on the optimal linear income tax. *Journal of Public Economics* 14 (3): 339-353.

Wilson, John. 1999. Theories of tax competition. *National Tax Journal* 52 (2): 269-304.