Munich Personal RePEc Archive

On the Political Economy of Nonlinear Income Taxation

Berliant, Marcus and Gouveia, Miguel

Washington University in St. Louis, Universidade Católica Portuguesa

20 May 2022

Online at https://mpra.ub.uni-muenchen.de/113140/
MPRA Paper No. 113140, posted 25 May 2022 07:40 UTC

On the Political Economy of Nonlinear Income Taxation*

Marcus Berliant[†] and Miguel Gouveia[‡]

May 20, 2022

Abstract

The literatures dealing with voting, optimal income taxation, implementation, and pure public goods are drawn on here to address the problem of voting over income taxes and a public good. In contrast with previous articles, general nonlinear income taxes that affect the labor-leisure decisions of consumers who work and vote are allowed. Uncertainty plays an important role in that the government does not know the true realizations of the abilities of consumers drawn from a known distribution, but must meet the realization-dependent budget; the tax system must be robust. Even though the space of alternatives is infinite dimensional, conditions on primitives are found to assure existence of a majority rule equilibrium when agents vote over both a public good and income taxes to finance it. JEL numbers: D72, D82, H21, H41 Keywords: Voting; Income taxation; Public good; Robustness

*Support provided by National Science Foundation grant SES-9007800 and the hospitality of the Whiteley Center at the University of Washington are gratefully acknowledged. The comments of Faruk Gul, an associate editor, two anonymous referees, Pierre Boyer, Randy Calvert, Yan Chen, Roy Gardner, Hideo Konishi, John Ledyard, Dilip Mukherjee, Wan-hsiang Pan, Paolo Piacquadio, Michael Smart, Jim Snyder, Max Troshkin and John Weymark are appreciated, but the authors bear full responsibility for any errors. This work was originally a chapter of M. Gouveia's dissertation, completed at the University of Rochester.

†Department of Economics, Washington University, MSC 1208-228-308, 1 Brookings Drive, St. Louis, MO 63130-4899 USA. Fax: (1-314) 935-4156, e-mail: berliant@wustl.edu

‡Universidade Católica Portuguesa, Lisbon, Portugal, e-mail: mig@ucp.pt

1 Introduction

1.1 Background

The theory of income taxation has been an important area of study in economics. Interest in a formal theory of income taxation dates back to at least J.S. Mill (1848), who advocated an equal sacrifice approach to the normative treatment of income taxes. In terms of the modern development, Musgrave (1959) argued that two basic approaches to taxation can be distinguished: the benefit approach, which puts taxation in a Pareto efficiency context; and the ability to pay approach, which puts taxation in an equity context. Some of the early literature, such as Lindahl (1919) and Samuelson (1954, 1955), made seminal contributions toward understanding the benefit approach to taxation and tax systems that lead to Pareto optimal allocations. Although the importance of the problems posed by incentives and preference revelation were recognized, scant attention was paid to solving them, perhaps due to their complexity and difficulty.

Since the influential work of Mirrlees (1971), economists have been quite concerned with incentives in the framework of income taxation. The model proposed there postulates a government that tries to collect a given amount of revenue from the economy. For example, the level of public good provision might be fixed. Consumers have identical utility functions defined over consumption and leisure, but differing abilities or wage rates. The government chooses an income tax schedule that maximizes some objective, such as a utilitarian social welfare function, subject to collecting the needed revenue, resource constraints, and incentive constraints based on the knowledge of only the overall distribution of wages or abilities. The incentive constraints derive from the notion that individuals' wage levels or characteristics (such as productivity) are unknown to the government. The optimal income tax schedule must separate individuals as well as maximize welfare and therefore is generally second best.[1] The necessary conditions for welfare optimization when the distribution of ability is bounded generally include a zero marginal tax rate for the highest wage individual; this result only holds very locally at the top of the distribution. Intuitive and algebraic derivations of this result can be found in Seade (1977), where it is also shown that some of these necessary

[1]If the government knew the type of each agent, it could impose a differential head tax. As is common in the incentives literature, one must impose a tax that accomplishes a goal without the knowledge of the identity of each agent *ex ante*.

conditions hold for Pareto optima as well as utilitarian optima. Existence of an optimal tax schedule for a modified model was demonstrated in Kaneko (1981), and then for the classical model in Berliant and Page (2001, 2006). An alternative view of optimal income taxation is as follows. Head taxes or lump sum taxes are first best, since public goods are not explicit in the model and therefore Lindahl taxes cannot be used. Second best are commodity taxes, such as Ramsey taxes. Third best are income taxes, which are equivalent to a uniform marginal tax on all commodities (or expenditure).[2] In our view, it is not unreasonable to examine these third best taxes, since from a pragmatic viewpoint, the first and second best taxes are infeasible.[3]

1.2 A Positive Political Model

How can we explain (or model) the income tax systems we observe in the real political world? We shall attempt to answer this question with a voting model, a positive political model, in combination with the standard income tax model described above. As noted in the introduction of Roberts (1977), one does not need to believe that choices are made through any particular voting mechanism; one need only be interested in whether choices mirror the outcomes of some voting process. Thus, what is described below is an attempt to construct a potentially predictive model with both political and economic content. It contains elements of the optimal income tax literature as well as positive political theory (an excellent survey of which can be found in Calvert (1986)).

Although much of the optimal income tax literature and most of the work cited above deals with the normative prescriptions of an optimal income tax, there is a relatively small literature on voting over income taxes. Most of this

[2] In certain circumstances, nonlinear income taxes can be second best; see Laroque (2005).

[3] Ramsey (optimal commodity) taxes typically follow the inverse elasticity rule. There are several practical issues. First, to compute Ramsey taxes, one must estimate supply and demand elasticities for each commodity. Second, these elasticities may vary over time, so the taxes would vary over time. Third, there is a question regarding how fine the classification of a commodity for the purposes of taxation might be. Demand for vegetables will typically be more inelastic than that for potatoes and carrots, which in turn will be more inelastic than demand for Yukon gold potatoes. As Stiglitz (2015, footnote 13) notes regarding the reasons for grouping commodities, "Partly it is administrative: the cost of having millions of tax rates, one for each precisely identified commodity, would be large. Partly it is based on information: private parties would have an incentive to try to get their products classified as one of the lower taxed products. It is costly for the government to gather the information required to implement and enforce fine differentiations."

literature is either restricted to consideration of only linear taxes, or does not consider problems due to information (adverse selection and moral hazard), or both. Examples that might fit primarily into the linear tax category which also involve no labor disincentives on the part of agents are Foley (1967), Nakayama (1976) and Guesnerie and Oddou (1981). Aumann and Kurz (1977) use personalized lump sum taxes in a one commodity model. Hettich and Winer (1988) present an interesting politico-economic model in which candidates seek to maximize their political support by proposing nonlinear taxes. Work disincentives are not present in the model. Chen (2000) extends their work to the more standard optimal income tax model in the context of probabilistic voting. Bierbrauer et al (2021) examine redistributive nonlinear income taxes with two parties, probabilistic voting, endogenous turnout, and ethical voters who care about the welfare of members of their party. They find that members of the other party might be demobilized by catering policy to them. Romer (1975), Roberts (1977), Peck (1986), and Meltzer and Richard (1981, 1983) use linear taxes in voting models with work disincentives. Roemer (1999) restricts to quadratic tax functions with no work disincentives but with political parties. Perhaps the model closest in spirit to the one we propose below is in Snyder and Kramer (1988), which uses a modification of the standard (nonlinear) income tax model with a linear utility function. The modification accounts for an untaxed sector, which actually is a focus of their paper. This interesting and stimulating paper considers fairness and progressivity issues, as well as the existence of a majority equilibrium when individual preferences are single peaked over the set of individually optimal tax schedules. (Sufficient conditions for single peakedness are found.) Röell (1996) considers the differences between individually optimal (or dictatorial) tax schemes and social welfare maximizing tax schemes when there are finitely many types of consumers. Of particular interest are the tax schedules that are individually optimal for the median voter type. This interesting work uses quasi-linear utility and restricts voting to tax schedules that are optimal for some type. Brett and Weymark (2017) push this further in a continuum of types model by characterizing individually optimal tax schedules. Then they show, under conditions including quasi-linear utility, that if the set of tax schedules is restricted to individually optimal ones, the individually optimal tax for the median voter is a Condorcet winner.

We propose in this paper to allow general nonlinear income taxes with work disincentives in a voting model. The main problem encountered in trying to find a majority equilibrium, as well as the reason that various sets of restrictive assumptions are used to obtain such a solution in the literature, is as follows. The set of tax schedules that are under consideration as feasible for the economy (under any natural voting rule) is large in both number and dimension. Thus, the voting literature such as Plott (1967) or Schofield (1978) tells us that it is highly unlikely that a majority rule winner will exist. Is there a *natural* reduction of the number of feasible alternatives in the context of income taxation?

1.3 The Role of Uncertainty and Feasibility

The answer appears to be yes. The (optimal) income tax model has a natural uncertainty structure that has yet to be exploited in the voting context. As in the classical optimal income tax model, all *worker/consumers* have the same well-behaved utility function, but there is a nonatomic distribution of wages or abilities. In standard models, such as the Mirrlees model or its modern descendants, the distribution of consumers by type is known by all and the aggregate revenue requirement is fixed at a scalar; it is 0 in models of pure redistribution. (This applies whether the number of consumer/workers is finite or a continuum.) Suppose that a *finite* sample is drawn from this nonatomic distribution.[4] The finite sample will be the true economy, and the revenue requirement imposed by the government can depend on the draw. In fact this dependence is just a natural extension of the standard optimal income tax model. In that model, the amount of revenue to be raised (the revenue requirement in our terminology) is a fixed parameter, something that makes perfect sense since the population in the economy and the distribution of the characteristics of that population are both fixed, and thus we can take public expenditures also as fixed. But consider now the optimal tax problem for the cases when the characteristics of the actual population are unknown. That is exactly what happens when we consider that the true population is a finite draw from a given distribution. In such circumstances, it is not reasonable to fix the revenue requirement at some exogenously given target level, but instead the revenue requirement should be a function of the population characteristics. It is possible to derive the aggregate revenue requirement from primitives in different ways. In our analysis below, *the revenue requirements for a particular draw will be derived from the Pareto efficient level of public good provision for*

[4] This assumption is similar to the one used in Bierbrauer (2011), though the purpose of that work is entirely different from ours.

that draw, leading to intrinsic variation in revenue requirements across draws.[5] We shall make assumptions on primitives implying that the Pareto efficient level of public good is unique but generally different for each draw.

Why do we focus on the Pareto efficient level for a given draw? There are two reasons, to be discussed in more detail shortly. First, given utility functions that are separable in the public good, *ex post* recontracting on the level of public goods provision will not be beneficial. Second, related to robustness, the worst case scenario for the government budget will always be a draw of identical individuals. Given the voting mechanism, these individuals will be at the median of the draw. Thus, for budgetary purposes, the government takes the public good level to be the efficient one with identical individuals at the stated median of the draw.

It seems natural for us to require that any proposed tax system must be feasible (in terms of the revenue it raises) for *any* draw, as *no player (including the government) knows the realization of the draw before a tax is imposed*. In other words, the consumers do not transmit information, such as their labor income, to the government prior to the imposition of the income tax. For example, an abstract government planner might not know precisely the top ability of individuals in the economy, and therefore might not be able to follow optimal income tax rules to give the top ability individual a marginal rate of zero, as described above. The key implication of using finite draws as the true economies is that requiring *ex post* feasibility of any proposable tax system for any draw narrows down the set of alternatives, which we call the *feasible* set, to a manageable number (even a singleton in some cases).[6]

To be clear, the assumption is that *the government must commit to a set of feasible tax systems* (where a tax system maps income to tax liability) before knowing the realization of the draw of abilities from the distribution of abilities, this set of feasible tax systems cannot depend on the draw, and it must raise sufficient draw-dependent revenue no matter the draw. Voting then chooses among the feasible tax systems, and the government must implement this choice. Finally, consumers supply labor and their types are revealed through

[5] As an alternative, the variation in revenue requirements can be seen as variation in fiscal pressure on the government; see Heathcote and Tsujiyama (2021) for discussion.

[6] Pierre Boyer has pointed out that the cost of the public good could be unknown in addition to or instead of the abilities of the individuals in the economy. This would represent an aggregate risk to the economy, in contrast with individual abilities, that are idiosyncratic risks. In accordance with the assumptions concerning asymmetric information in the optimal tax literature, we stick to unknown individual abilities on the part of the tax designer.

their choice of labor supply. If we allowed the set of feasible tax systems to depend on the draw of abilities, we would be back in the situation the rest of the literature has found unsolvable, since in general any tax system can be defeated by a majority for a given draw. In other words, if the government doesn't have to commit and can propose a state or draw contingent set of feasible taxes, we have the same situation as if there is no uncertainty and a finite number of worker/consumers with given types, so there generally will be no Condorcet winner in any given state.

Our model fits naturally into the literature in public finance on robustness, examining allocation when agents have ambiguous beliefs about the state of the world. Recent innovations include an application to public goods by Kocherlakota and Song (2019), that analyzes efficiency in the context of a discrete decision about whether or not to produce a public good in a static environment; and Lensman and Troshkin (forthcoming) that examines optimal allocative policy in a dynamic environment. Robustness has been used in the context of voting as well; see Berliant and Konishi (2005).

A significant difference between the work here and the balance of the literature is how robustness enters the model. In the literature just cited, agents have multiple priors over the state of the world and are ambiguity averse. In contrast, we shall assume that the government must generate sufficient revenue independent of the draw of agent types. One way to envision this is to assume that the *government* is ambiguity averse and each element of the set of distributions it considers possible assigns a draw probability 1. Notice that in this case, ambiguity enters into a government constraint rather than a consumer's objective function.

In sum, given government concerns about robustness, our setup guarantees that a majority rule equilibrium exists and that it can be achieved by the budget.

Our arguments apply to finite numbers of agents. The model has a discontinuity when one goes from a finite to an infinite number of agents. In this latter case there is no uncertainty about the composition of the draw, so we do not have a continuum of *ex post* feasibility restrictions, one for each possible draw. Instead we have only that the revenue constraint needs to be satisfied for the known population. Thus, for our purposes, even a little uncertainty is sufficient, and it is possible to view perfect certainty about the draw as a knife-edge case.[7] Moreover, there are further conceptual issues pertaining to

[7] We are indebted to Jim Snyder for some of these thoughts.

models with a pure public good and a continuum of consumers; see Berliant and Rothstein (2000).

In relation to the literature that deals with voting over *linear* taxes, our model of voting over nonlinear taxes will not yield a linear tax as a solution without very extreme assumptions. This will be explained in section 6 below. Moreover, our second order condition for incentive compatibility will generally be much weaker than those used in the literature on linear taxes; compare our assumptions below with the Hierarchical Adherence assumption of Roberts (1977). As noted by L'Ollivier and Rochet (1983), these second order conditions are generally not addressed in the optimal income taxation literature, though they ought to be addressed there. In what follows, we employ the results contained in Berliant and Gouveia (2001) and more generally in Berliant and Page (1996) to be sure that the second order conditions for incentive compatibility hold in our model.

At this point, it is important to remark on the anatomy of our analysis. We shall introduce two models: an endowment economy, where there is no choice of labor supply, where each consumer knows only their own endowment and the prior distribution from which endowments are drawn, and where all taxes are lump sum; and an optimal income tax economy, where each consumer knows only their own productivity and the prior distribution of types, where the government knows only the prior distribution from which agents' types are drawn, where taxes induce distortions in labor supply, and where incentive constraints must be satisfied. Although results on existence of Condorcet winners for the endowment economy may be of independent interest, our primary objective is to apply these results to the income tax model with distortions. The method for accomplishing this is to use a result on implementation of lump sum taxes from the endowment economy, the first model, in terms of a tax on labor income in the optimal income tax economy, the second model. Under assumptions we shall specify, this result implements the lump sum tax system in the sense that each consumer facing a labor-leisure choice ends up paying exactly the tax specified by the lump sum tax system; notice that income taxes are an indirect mechanism. In the literature on optimal income taxation, this is called the "Taxation Principle." Moreover, we show that characteristics of the lump sum tax systems, such as single crossing, are inherited by the taxes implementing them in the framework with incentives.

To prove our main result, we combine the restriction on tax systems, as outlined above, with assumptions on the utility function and the cost function for public goods. In particular, we assume that the utility function is quasi-linear in consumption and additively separable in the consumption commodity, labor supply, and the public good level; moreover, the public good subutility is multiplicative in type. Due to quasi-linearity and separability of the utility function, the Pareto efficient level of public good, and thus aggregate revenue requirements, will be unique for each draw. Furthermore, the aggregate revenue requirements will be concave in the draw, where the worst case for the government budget occurs when the draw consists of identical individuals.[8] The tangent to the aggregate revenue requirement function where all individuals in a draw are identical (a one dimensional domain for the function) represents a minimal feasible individual revenue requirement. These are single crossing, and their implementations in terms of income taxes or net income schedules are also single crossing. Among these, the tax system most preferred by the median voter is a majority rule equilibrium, since induced preferences over net income schedules are single crossing, as is standard in the optimal income tax literature.

The structure of the paper is as follows. First, we introduce our framework and notation in section 2. In section 3 our main result on voting over both public goods and income taxes is stated. Section 4 contains a discussion of the techniques we use in the proofs, whereas section 5 contains two examples of interest. Finally, section 6 contains conclusions and suggestions for further research. The appendix contains proofs of most results.

2 The Model

2.1 Basic Notation and Definitions

We shall develop an initial model of an endowment economy as a tool. Although it might be of independent interest, our primary purpose is to apply this model and the results we obtain to the standard optimal income tax model in the succeeding sections.

There is a single consumption good c and consumers' preferences are identical and given by the utility function $v(c) = c$, with $c \in \mathbb{R}_+$. A consumer's en-

[8] We have succeeded in proving analogous results when the aggregate draw-dependent revenue requirements function is convex rather than concave, so the worst case scenario for the government budget is when the draw consists of a given type and the extreme type most unlike it. However, we were unable to derive this kind of aggregate revenue requirements function from primitives.

dowment, which is also her type, is described by $w \in [\underline{w}, \overline{w}]$, where $[\underline{w}, \overline{w}] \subseteq \mathbb{R}_+$. In this section the endowment can also be seen as pre-tax income or, following classical terminology in Public Finance, the *ability to pay* of each agent. References to measure are to Lebesgue measure on $[\underline{w}, \overline{w}]$.

The distribution of consumers' endowments has a measurable density $f(w)$, where $f(w) > 0$ $a.s.$

Let k be a positive integer and let $\mathcal{A} \equiv [\underline{w}, \overline{w}]^k$, the collection of all possible draws of k individuals from the distribution with density f.[9] Formally, a *draw* is an element $(w_1, w_2, ..., w_k) \in \mathcal{A}$.

In order to be able to determine what any particular draw can consume, it is first necessary to determine what taxes are due from the draw. Hence, we first assume that there is a given net revenue requirement function $R : \mathcal{A} \to \mathbb{R}$. For each $(w_1, w_2, ..., w_k) \in \mathcal{A}$, $R(w_1, w_2, ..., w_k)$ represents the total taxes due from a draw. For example, if the revenues from the income tax are used to finance a good such as schooling, then $R(w_1, w_2, ..., w_k)$ can be seen as: the per capita revenue requirement for providing schooling to the draw $(w_1, w_2, ..., w_k)$ multiplied by k.[10]

Although we shall begin by taking revenue requirements as a primitive, in the end we will justify this postulate by deriving revenue requirements from the technology for producing a public good. But this is simply an important example illustrating where aggregate revenue requirements come from.

It is important to be clear about the interpretation of R. One easy interpretation is that the taxing authority provides a schedule giving the taxes owed by any draw. There are several reasons that revenue requirements might differ among draws, including differences in taste for a public good that is implicitly provided, income or wealth differences, a non-constant marginal cost for production of the public good, differences in the cost of revenue collection,

[9] Note that $f(\cdot)$ plays almost no role in the development to follow, in contrast with its preeminent role in the standard optimal income tax model. It may be interpreted as a subjective distribution describing the planner beliefs about the characteristics of the agents in the economy, but that consideration is immaterial for the model presented here. We have implicitly assumed that the abilities are drawn independently, but since we never use this, correlation would also be permissible provided that the support of the joint distribution is fixed at $[\underline{w}, \overline{w}]^k$. In multistage voting in a representative democracy, the equilibria are likely to be a function of f, as is often the case in signaling games. We expect to study that problem in the future.

[10] Actually, regarding schooling, there is a separate literature on the political economy of public supplements for such goods. The formal structure is slightly different from what we consider in this paper; see Gouveia (1997).

and so forth.

The government and the agents in the economy know the prior distribution f of types of agents in the economy[11] as well as the mapping R. Before moving on to consider the game - theoretic structure of the problem, it is necessary to obtain some facts about the set of tax systems that are feasible for any draw in $\mathcal{A}$. These are the only tax systems that can be proposed, for otherwise the voters and social planner would know more about the draw than that it consists of k people drawn from the distribution with density f. Voters can use their private information (their endowment) when voting, but not in constructing the feasible set. For otherwise either each voter will vote over a different feasible set, or information will be transmitted just in the construction of the feasible set.

An *individual revenue requirement*[12] is a function $g : [\underline{w}, \overline{w}] \to \mathbb{R}$ that takes w to tax liability. It is a lump sum tax function.

Clearly, there will generally be a range of individual revenue requirements consistent with any map R. Our next job is to describe this set formally. Fix k and R. Let

$$G \equiv \Big\{ g : [\underline{w}, \overline{w}] \to \mathbb{R} \mid g \text{ is measurable}, \sum_{i=1}^{k} g(w_i) \geq R(w_1, w_2, ..., w_k) \text{ a.s. } (w_1, w_2, ..., w_k) \in \mathcal{A} \Big\}$$

G is the set of all *individual* revenue requirements that collect enough revenue to satisfy R. $G \neq \emptyset$ if almost surely for $(w_1, w_2, ..., w_k) \in \mathcal{A}$, $\sum_{i=1}^{k} w_i \geq R(w_1, w_2, ..., w_k)$. The constraint that the aggregate revenue requirement be satisfied for each draw restricts the feasible set G significantly.

2.2 From Collective to Individual Revenue Requirements

In order to examine the set of feasible individual revenue requirements described above, more structure needs to be introduced. It is obvious that some feasible g's will raise strictly more taxes than necessary to meet $R(w_1, w_2, ..., w_k)$ for almost any $(w_1, w_2, ..., w_k)$. For example, take any $g \in G$. Then for $\epsilon > 0$, $g + \epsilon$ will satisfy $\sum_{i=1}^{k} [g(w_i) + \epsilon] \geq R(w_1, w_2, ..., w_k)$ a.s. $(w_1, w_2, ..., w_k) \in \mathcal{A}$, but clearly the consumers would all prefer g to $g + \epsilon$. The point is that some

[11] Actually, all they need to know is the support of that distribution.

[12] Even though this is simply a tax function on endowments, we will reserve the terminology "tax function" for an environment with incentives to simplify the exposition.

individual revenue requirements functions can be dominated unanimously, and thus can be eliminated from consideration. We now search for the minimal elements of the set G. We call the set of such elements G^*. In other words, we search for individual revenue requirements $g \in G^* \subseteq G$ with the following property: there is no g' such that almost surely for $(w_1, w_2, ..., w_k) \in \mathcal{A}$, $R(w_1, w_2, ..., w_k) \leq \sum_{i=1}^{k} g'(w_i)$; almost surely for $w \in [\underline{w}, \overline{w}]$, $g'(w) \leq g(w)$; and there exists a set of positive Lebesgue measure in $[\underline{w}, \overline{w}]$ where $g'(w) < g(w)$.

To this end, define a binary relation $\succeq$ over G by $g \succeq g'$ if and only if $g(w) \geq g'(w)$ for almost all $w \in [\underline{w}, \overline{w}]$. Let

$$\mathcal{G} \equiv \{B \subseteq G \mid B \text{ is a maximal totally ordered subset of } G\}.$$

By Hausdorff's Maximality Theorem (see Rudin (1974, p. 430)), $\mathcal{G} \neq \emptyset$. Finally, define

$$G^* \equiv \{g : [\underline{w}, \overline{w}] \to \mathbb{R} \mid \exists B \in \mathcal{G} \text{ such that } g(w) = \inf_{g' \in B} g'(w) \text{ a.s.}\}.$$

G^* is nonempty.

If $g \in G \backslash G^*$ is proposed as an alternative to $g^* \in G^*$, $\exists g' \in G^*$ that is unanimously weakly preferred to g.

2.3 Notation for the Optimal Income Tax Model

Having dispensed with preliminaries, we now turn to the voting model with incentives based on Mirrlees (1971). The three goods in the model are a composite consumption good, whose quantity is denoted by c; labor, whose quantity is denoted by l; and a pure public good, whose quantity is denoted by x. Consumers have an endowment of 1 unit of *labor/leisure*, no consumption good, and no public good.[13] Let $u : \mathbb{R}_+ \times [0, 1] \times \mathbb{R}_+ \times [\underline{w}, \overline{w}] \to \mathbb{R}$ be the utility functions of the agents, writing $u(c, l, x, w)$ as the utility function of type w, where u is twice continuously differentiable. Subscripts represent partial derivatives of u with respect to the appropriate arguments. The parameter w, an agent's type, is now to be interpreted as the wage rate or productivity of an agent. Thus w is the value of an agent of type w's endowment of labor. The gross income earned by an agent of type w is $y = w \cdot l$ and it equals consumption when there are no taxes.

[13] It would be easy to add an endowment of consumption good for consumers, but that would complicate notation.

A *tax system* is a function $\tau : \mathbb{R} \to \mathbb{R}$ that takes y to tax liability. A *net income function* $\gamma : \mathbb{R} \to \mathbb{R}$ corresponds to a given τ by the formula $\gamma(y) \equiv y - \tau(y)$.

First we discuss the typical consumer's problem under the premise that the consumer does not lie about its type, and later turn to incentive problems. A consumer of type $w \in [\underline{w}, \overline{w}]$ is confronted with the following maximization problem in this model:

$$\begin{aligned} \max_{c,l} u(c,l,x,w) \text{ subject to } w \cdot l - \tau(w \cdot l) &\geq c \text{ with } \tau, x \text{ given,} \\ \text{and subject to } c &\geq 0,\, l \geq 0,\, l \leq 1. \end{aligned}$$

For fixed τ, we call arguments that solve this optimization problem $c(w)$ and $l(w)$ (omitting τ and x) as is common in the literature. Define $y(w) \equiv w \cdot l(w)$.

The public good financed by the revenue raised through the income tax is usually excluded from models of optimal income taxation due to the complexity introduced, but here the cost of the public good will be used to derive the revenue requirements function. Let the cost function for the public good in terms of consumption good be $H(x)$, which is assumed to be C^2. The basic set of tax functions for the optimal income tax model is defined as:

$$T \equiv \{\tau : \mathbb{R}_+ \to \mathbb{R} |\ \tau \text{ is measurable}\}$$

As is standard in the literature, for $\tau \in T$ we shall generally write $\tau(y)$ to denote the tax liability of a worker earning income y.

These basic assumptions will be maintained throughout the remainder of this paper.

We will now use ideas inspired by Bergstrom and Cornes (1983) to obtain a unique Pareto optimal level of public good for each draw, so the revenue requirement function is well-defined.

The major assumption that we make to obtain results, beyond requiring sufficient revenue to finance the public good for each draw, is that utility is quasi-linear and separable to a certain degree:[14]

Assumptions:

$$u(c,l,x,w) = c + b(l,w) + r(x,w)$$

[14] In this case we are also using w as a taste parameter. That interpretation is quite common in both the optimal tax literature and the literature on self-selection.

We assume throughout that $\partial b/\partial l \leq 0$ a.s., $\partial^2 b/\partial l^2 < 0$, $\left.\frac{\partial b(l,\overline{w})}{\partial l}\right|_{l=1} \leq -\overline{w}$, $b(1,\underline{w}) - b(0,\underline{w}) \geq -\underline{w}$, $\frac{\partial b(l,w)}{\partial l}$ is weakly increasing in w, $\partial r/\partial x > 0$, $\partial^2 r/\partial x^2 < 0$; $dH(x)/dx > 0$ and $d^2H(x)/dx^2 \geq 0$.

From this, it follows that utility is strictly monotonic in consumption commodity (a good) and labor (a bad). The assumption $\left.\frac{\partial b(l,\overline{w})}{\partial l}\right|_{l=1} \leq -\overline{w}$ is a (weak) boundary condition on utility. There are constraints on labor supply ($0 \leq l \leq 1$) that could bind. The assumption implies that the constraint $l \leq 1$ does not bind. In contrast, $b(1,\underline{w}) - b(0,\underline{w}) \geq -\underline{w}$ is the lower boundary condition in the context of quasi-linear utility. The assumption that $\frac{\partial b(l,w)}{\partial l}$ is weakly increasing in w is the single crossing property. There are several more remarks to be made. First, if we had more than one efficient level of public good possible for given parameters, as is standard in public goods models without the Bergstrom-Cornes type of assumptions, then we would have another dimension to vote over, namely the level of the public good. Generally speaking, this would cause Condorcet cycles and thus no Condorcet winner. Second, if we made utility more general, for example allowing the subutility function $r(x,w)$ to depend on consumption good c or labor l or both, then the public good level and hence the aggregate revenue requirement function would depend on the tax function, and that tax function would depend on the public good level and hence the aggregate revenue requirement function. Thus, the aggregate revenue requirement would not be exogenous and likely not uniquely defined. Probably it is a solution to a fixed point problem, possibly a contraction under some circumstances. Third, when production of the public good is not constant returns to scale, there is a potential issue of profit distribution. However, when utility is quasi-linear, this isn't really an issue.

The bottom line is that something has to be done to shut down the feedback between tax liabilities and the optimal level of the public good. The Bergstrom and Cornes (1983) specification is a natural starting point and actually is more general than some of the separability assumptions used in the optimal nonlinear income tax literature.

3 Voting Over Income Taxes and a Public Good

3.1 Preliminaries

The next step in our analysis is to find the Pareto efficient level of public goods provision for each draw using the Lindahl-Samuelson condition for our specialized economy, a technique pioneered by Bergstrom and Cornes (1983).

Let $(w_1, w_2, ..., w_k) \in \mathcal{A}$, and let c_i and l_i denote the consumption and labor supply of the ith member of the draw respectively. Then production possibilities for this given draw are:

$$\sum_{i=1}^{k} w_i \cdot l_i - \sum_{i=1}^{k} c_i \geq H(x). \tag{1}$$

Fix $(w_1, w_2, ..., w_k) \in \mathcal{A}$. We define an allocation to be *interior* if the associated level of public good x satisfies $x > 0$ and $H(x) < \sum_{i=1}^{k} w_i$. Given our assumptions, a necessary and sufficient condition for an interior Pareto optimum is: $\sum_{i=1}^{k} \partial r(x, w_i)/\partial x \mid_{x=0} > dH(x)/dx \mid_{x=0}$ and there is $\overline{x}$ such that $\sum_{i=1}^{k} \partial r(x, w_i)/\partial x \mid_{x=\overline{x}} < dH(x)/dx \mid_{x=\overline{x}}$ and $H(\overline{x}) < \sum_{i=1}^{k} w_i$. More usefully, we shall assume the following sufficient boundary condition:

For all $w \in [\underline{w}, \overline{w}]$, $\partial r(x, w)/\partial x \mid_{x=0} > dH(x)/dx \mid_{x=0}$, *and* $k \cdot \partial r(x, w)/\partial x \mid_{x=H^{-1}(k\underline{w})} < dH(x)/dx \mid_{x=H^{-1}(k\underline{w})}$.

The first part of the condition says that for each type, at zero public good level, the marginal willingness to pay exceeds the marginal cost. The second part of the condition says that for each type, at public good level that is the maximum possible for a draw of only the lowest type, the sum of marginal willingnesses to pay is less than the marginal cost.

See Bergstrom and Cornes (1983, p. 1757) for a detailed explanation of why we need to restrict the analysis to interior allocations. Briefly, the issue is boundary optima where some consumer has no private good. In this case, there may be multiple efficient levels of public good production. It is ruled out by the second part of our sufficient condition. The entire sufficient condition allows the use of the Lindahl-Samuelson condition with equality. Without the first part, a unique efficient level of public good (possibly 0) could still be obtained, but its characterization is more difficult. Boundary Pareto optima could occur, for example, if the sum of the marginal tax rate at zero income and the marginal disutility of labor at zero labor supply is greater than 1 for some type.

Lemma 1: Under the basic assumptions listed above, for any given draw $(w_1, w_2, ...w_k)$, there exists an interior Pareto optimal allocation; moreover, for all interior Pareto optimal allocations, the public good level x^* is the same.

Proof: The Pareto optimal allocations are solutions to: $\max u(c_1, l_1, x, w_1)$ subject to $u(c_i, l_i, x, w_i) \geq \overline{u}_i$ for $i = 2, 3, ...k$ and subject to (1) where the maximum is taken over c_i, l_i, $(i = 1, ..., k)$ and x.[15] The Lindahl-Samuelson condition for this problem is:

$$\sum_{i=1}^{k} \partial r(x, w_i)/\partial x = dH(x)/dx. \tag{2}$$

Given our assumptions on r and H, there exists a unique interior level of public good x^* that solves (2). Since this equation is independent of c_i and l_i for all i, the interior Pareto optimal level of public good provision is independent of the distribution of income and consumption for the given draw.

For the class of utility functions defined above we can thus solve for x^* as an (implicit) function of $(w_1, w_2, ..., w_k)$, and obtain the revenue requirement function

$$R(w_1, w_2, ..., w_k) \equiv H(x^*(w_1, w_2, ..., w_k)).$$

Let $F \subseteq T$ be the *feasible set* defined by:

$$F \equiv \left\{ \tau \in T \mid \text{a.s. for } (w_1, w_2, ..., w_k) \in \mathcal{A}, \sum_{i=1}^{k} \tau(y(w_i)) \geq R(w_1, w_2, ..., w_k) \right\}.$$

Notice that the feasible set does not depend on the level of public good x, since revenue requirements for Pareto efficient x must be satisfied for all draws, and hence for the draw that is actually realized.

With this in hand, a straightforward definition of majority rule equilibrium follows: a *majority rule equilibrium* is a mapping from profiles of preferences into outcomes such that there is no profile in which $n > k/2$ agents can agree on an outcome that is strictly better for all n of them.

It is important to make a couple of remarks about this definition. First, in the case where aggregate revenue requirements R are taken as primitive

[15] Bergstrom and Cornes (1983, Theorem 3) show that in their context, all interior Pareto optima can be found by solving the utilitarian maximization problem. The same holds in our context, but the direct analysis of interior Pareto optima here is short and self-contained.

and exogenous, we can simply take the public good level x to be fixed as an argument in consumers' utility functions. Then there is no voting over the public good component of the bundle. Second, why do we focus on the first best level of public good provision? There are a couple of answers to this question.

Since our income tax is distortive, it is reasonable to inquire why we should impose the Pareto efficient level of public goods provision for each draw. To be specific, a second-best, lower level of public good provision could be used, where the amount of revenue that must be collected, and thus the distortion imposed by the income tax, could be reduced. For example, a draw of all high types will require a high level of public good provision and thus high per capita taxes. To ensure that high types do not try to mimic low types, the per capita tax on low types will generally have to be higher than it would be if the draw were known to consist of all low types and the first best level of public good provision for that draw is used. Thus, excess revenue will be generated from some draws, specifically of all low types, and it might be more efficient to reduce the income tax for all draws and types and provide a second best, lower level of public good and consequent lower tax for each draw.[16] The problem with this argument is that, *ex post*, after labor is supplied and taxes are collected, the workers would like to voluntarily contribute lump sum to raise the level of public good back to first best, say through a constant per capita tax. In other words, they would like to renegotiate. So we stick to first best public good levels to avoid this problem. Naturally, this involves some cooperative behavior *ex post*.

This example also illustrates how the incentive constraints affect the feasible set of tax instruments. What we have presented here is a benchmark, in that the first best public good level is unique for each draw, thus cutting off feedback between public good level and income tax selection. Such feedback would make our analysis much more difficult.

An alternative, and perhaps more convincing justification, invokes robustness again and in fact does not require the government to provide the public good at the first best level. Only in the last stage of the game do voters make their labor supply decision subject to an income tax. At that point, the tax system and public good level are already chosen, so the incentive constraints on the tax system do not need to account for this. Prior to that, they vote over feasible tax systems, with an understanding that the level of public good

[16] These interesting comments belong to Paolo Piacquadio.

provided will be determined as follows. The level is the one where the Lindahl-Samuelson condition is met contingent on the draw consisting of k identical individuals with marginal willingness to pay for the public good at the stated median. This is the worst case for the government budget constraint, in the sense that it will be met with equality rather than strict inequality. If there is dispersion around the median of the draw, the level of public good provision will not be efficient (subject to the exceptions represented by Bowen's theorem discussed in detail below). This game turns the private optimization problem for the median voter into a social optimization problem for the median voter with clones. Although the outcome is unlikely to be efficient, it is feasible, the median voter has an incentive to be honest, and others in the draw have no incentive to change it when this mechanism is used.

One way to implement this idea is to have voters state their type, where the government chooses the stated median's favorite tax and public good level when the draw consists of median clones. Another is to vote over the tax system, where the public good level is determined through optimization of welfare with clones of the median as specified above. The solution to the optimization problem for the median voter is the same, whether determined through statement of type or by vote.

Of course, in cases where R is a primitive and there is no public good in the model, this entire argument is superfluous.

3.2 The Main Result

To simplify notation, we shall abbreviate derivatives of functions of only one variable using primes, e.g. $H'(x) \equiv dH(x)/dx$. Beyond uniqueness of the efficient public goods level for each draw, further properties of aggregate revenue requirements R, and the implied minimal individual revenue requirements g, are needed for our main result. To be specific, we make assumptions implying that R is a strictly concave function of the sum of the types in a draw. Since R is derived from efficient levels of public good production, assumptions on both the consumer and producer side of the model are required:

Definition: A utility function - production function pair (u, H) is called *manageable* if the following conditions hold:

$u(c, l, x, w) = c + b(l, w) + w \cdot s(\widehat{r}(x))$, $H(x) = m \cdot \widehat{r}(x)$, where $\widehat{r}'(x) > 0$, $\widehat{r}''(x) \geq 0$, $s'(r) > 0$, $s''(r) < 0$, $2s''(r)^2 > s'''(r) \cdot s'(r)$, $m > 0$.

Theorem 1: Let $k \geq 2$ and let (u, H) be manageable. Then there exists a majority rule equilibrium.

Proof: See the Appendix.

Examples covered by this theorem include the following:

A. $u(c, l, x, w) = c + b(l, w) + w \cdot s(x)$, $H(x) = m \cdot x$, where $\widehat{r}(x) = x$, $s'(x) > 0$, $s''(x) < 0$, $s'''(x) \leq 0$, and $m > 0$.

B. $u(c, l, x, w) = c + b(l, w) + \phi \frac{w}{1-\alpha} x^{1-\alpha}$, $H(x) = \frac{m}{\beta} \cdot x^{\beta}$, with $\alpha > 1$, $\beta \geq 1$, $\phi > 0$. In this case, $\widehat{r}(x) = \frac{1}{\beta} x^{\beta}$, $s(r) = \phi \frac{\beta^{\frac{1-\alpha}{\beta}}}{1-\alpha} r^{\frac{1-\alpha}{\beta}}$.

Example 1 below is covered by B, with $\alpha = 3$, $\beta = 2$.

A complicating factor in proving such theorems is that aggregate revenue requirements derived here depend not only on the first derivative of the cost function, but on its level as well.

An important assumption used in the theorem is that utility from the public good is multiplicative in type w. Although this is a strong assumption, it is often used in empirical work; see for example Bishop and Timmins (2019, equation (5)).

3.3 Outline of the Proof of the Main Result

Before proceeding into details of the proof, we outline its structure:

- All feasible $g \in G^*$ take a very specific form, and any two elements of G^* cross at most once (Lemma 2).
- The implementations of any given $g \in G^*$ are Pareto ordered, and the best of these, called $\tau \in T^*$, is well-defined (Lemma 3).
- Any two elements of T^* satisfy a single crossing property inherited from G^* (Lemma 4).
- For any k-tuple $(w_1, ..., w_k)$, define τ^* to be the optimal tax schedule from the point of view of the median type(s).
- Now suppose there is another τ strictly preferred by a majority. By Lemma 4, τ and τ^* satisfy single-crossing.

- The net income functions induced by τ and τ^* also satisfy single crossing, whereas the indifference curves induced over gross and after tax income satisfy single crossing in type, so τ^* is not majority defeated by τ (Lemma 5); see Gans and Smart (1996, Figure 1).

3.4 Single Crossing Individual Revenue Requirements

With a view toward future extensions of Theorem 1, we state some natural assumptions on R that *will be satisfied by the revenue requirements derived in the course of proving Theorem 1.* For instance, we might wish simply to take revenue requirements R as a primitive rather than derived from a model of a pure public good. That generalization is covered in this section.

The first of these assumptions means that position in the draw (first, second, etc.) does not matter. All that matters in determining the revenue to be extracted from a draw is which types are drawn from the distribution.

Definition: A revenue requirement function R is said to be *symmetric* if for each k and for each $(w_1, w_2, ..., w_k) \in \mathcal{A}$, for any permutation σ of $\{1, 2, ..., k\}$, $R(w_1, w_2, ..., w_k) = R(w_{\sigma(1)}, w_{\sigma(2)} ..., w_{\sigma(k)})$.

We will use the assumption that R is C^2. This is not a strong assumption, because the assumption that R is C^2 is generic in the appropriate topology; that is, C^2 R's will uniformly approximate any continuous R (Hirsch (1976, Theorem 2.2)).[17] We will also assume that R is smoothly monotonic:

Definition: A revenue requirement function R is said to be *smoothly monotonic* if for any $(w_1, w_2, ..., w_k) \in \mathcal{A}$, $\partial R(w_1, w_2, ..., w_k)/\partial w_i > 0$ for $i = 1, 2, ..., k$.

This assumption requires that increasing the ability or wage of any individual in a draw increases the total tax liability of the draw. One could successfully use weaker assumptions with this framework, but at a cost of greatly complicating the proofs.[18]

[17] This idea is also used to justify differentiability in the smooth economies literature.

[18] One particular case ruled out is the one of constant per capita revenues. In our model this situation implies constant individual revenue requirements, i.e. a head tax, clearly an uninteresting situation even though it is first-best. It also includes the particular situation where the government wants to raise zero fiscal revenue. Constant per capita revenues can be handled as a limit of the cases considered here.

A major step in our analysis that we have relegated to other papers that are cited in the bibliography, Berliant and Gouveia (2001) and Berliant and Page (1996), is to implement the individual revenue requirement g using an income tax, an indirect mechanism. A sufficient (and virtually necessary) condition is that g be increasing in type, w.[19] If g is anywhere decreasing in type, the net income function can cut the indifference curve of an agent, creating a gap in the assignment of types to tax liability and ruining the implementation of g by an income tax. To use the first order approach to incentive compatibility, for example, we must make further assumptions, namely the second order conditions.[20] These second order conditions are equivalent to the property that g is increasing.

Turning next to aggregate revenue requirements R, we relate the property of increasing R ($\partial R(w_1, w_2, ..., w_k)/\partial w_i > 0$ for $i = 1, 2, ..., k$) to increasing $g \in G^*$. Suppose that there are $w, w' \in [\underline{w}, \overline{w}]$ with $w' > w$. Then, by definition of G^*, there is a draw $(w_1, w_2, ..., w_k)$ with $w = w_i$ for some i and $\sum_{j=1}^{k} g(w_j) = R(w_1, w_2, ..., w_k)$. Now replace w with w', namely set $w_i = w'$, leaving all other elements of the draw the same. Then $R(w_1, w_2, ..., w', ..., w_k) > R(w_1, w_2, ..., w_k)$. Since g is feasible, $g(w_1) + g(w_2) + \cdots + g(w') + \cdots + g(w_k) \geq R(w_1, w_2, ..., w', ..., w_k) > R(w_1, w_2, ..., w_k) = \sum_{j=1}^{k} g(w_j)$, so $g(w') > g(w)$.

The next step is to introduce a set of assumptions where the elements of the set of feasible and minimal individual revenue requirements G^* are single crossing,[21] i.e. each pair of g's will cross only once.[22] They will be implied by the postulates of Theorem 1. Thus, future generalizations of our main results will likely use the lemmas below. The assumptions have collective revenue requirements decreasing as a draw becomes more polarized.

Definition: A revenue requirement function $R(w_1, ...w_k)$ is *argument-additive* if $R(w_1, w_2, ..., w_k) \equiv Q(\sum_{i=1}^{k} w_i)$. Let Q' denote $\frac{dQ}{d\sum_{i=1}^{k} w_i}$.

[19] The case $g'(w) = 0$ for some types w could be handled, but it creates some technical problems because g is not necessarily invertible.

[20] We note that much of the recent literature on optimal taxation verifies the second order conditions *ex post*, not *ex ante*; see Kapička (2013) for example.

[21] In fact, under stronger assumptions, it is possible to show that the set of feasible and minimal individual revenue requirements is a singleton, rendering voting trivial. In that analysis, it's useful to have the size of the draw, k, unknown to the planner as well. We omit this analysis for the sake of brevity.

[22] A G^* with single crossing g's generates a trade-off where raising more taxes from one type of voter allows less revenue to be raised from another type, as in the conventional income tax model.

The goal of the next lemma is to characterize minimal individual revenue requirements, in other words elements of G^*, and show that they are single crossing.

Lemma 2: Let $k \geq 2$ and let the revenue requirement function $R(w_1, w_2, ..., w_k)$ be argument-additive with $Q' > 0$ and $Q'' < 0$.

(a) There exists $\overline{g} : [\underline{w}, \overline{w}]^2 \to R$ such that for all $g \in G^*$, there exists $\widetilde{w} \in [\underline{w}, \overline{w}]$ such that $g(w) = \overline{g}(w, \widetilde{w})$ for all $w \in [\underline{w}, \overline{w}]$ where $\overline{g}$ is defined pointwise as follows:

– For $\widetilde{w} \geq (\underline{w} + \overline{w})/2$:

A) $\overline{g}(w, \widetilde{w}) = Q(k\widetilde{w})/k + Q'(k\widetilde{w}) \cdot (w - \widetilde{w})$ if $w \leq \widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})$.

B) $\overline{g}(w, \widetilde{w}) = Q((k-1)\underline{w} + w) - ((k-1)/k) \cdot Q(k\widetilde{w}) + (k-1) \cdot Q'(k\widetilde{w}) \cdot (\widetilde{w} - \underline{w})$ if $w > \widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})$.

– For $\widetilde{w} < (\underline{w} + \overline{w})/2$:

C) $\overline{g}(w, \widetilde{w}) = Q(k\widetilde{w})/k + Q'(k\widetilde{w}) \cdot (w - \widetilde{w})$ if $w \geq \widetilde{w} - (k-1) \cdot (\overline{w} - \widetilde{w})$

D) $\overline{g}(w, \widetilde{w}) = Q((k-1)\overline{w} + w) - ((k-1)/k) \cdot Q(k\widetilde{w}) + (k-1) \cdot Q'(k\widetilde{w}) \cdot (\widetilde{w} - \overline{w})$ if $w < \widetilde{w} - (k-1) \cdot (\overline{w} - \widetilde{w})$

(b) $\partial \overline{g}(w, \widetilde{w})/\partial w > 0$ for each fixed $\widetilde{w} \in [\underline{w}, \overline{w}]$ except at a finite number of points $w \in [\underline{w}, \overline{w}]$.

(c) $\forall w \in [\underline{w}, \overline{w}]$, $\overline{g}(w, \widetilde{w})$ is single caved[23] in $\widetilde{w}$ and attains a minimum at $\widetilde{w} = w$.

(d) Any pair of functions in G^* will cross once: for any $\widetilde{w}, \widetilde{w}' \in [\underline{w}, \overline{w}]$, there exist $\widehat{w}, \widehat{w}' \in [\underline{w}, \overline{w}]$ such that $\overline{g}(\underline{w}, \widetilde{w}) > \overline{g}(\underline{w}, \widetilde{w}')$ implies $\overline{g}(w, \widetilde{w}) > \overline{g}(w, \widetilde{w}')$ for all $w \in [\underline{w}, \widehat{w})$, $\overline{g}(w, \widetilde{w}) = \overline{g}(w, \widetilde{w}')$ for all $w \in [\widehat{w}, \widehat{w}']$ and $\overline{g}(w, \widetilde{w}) < \overline{g}(w, \widetilde{w}')$ for all $w \in (\widehat{w}', \overline{w}]$.

Proof: See the Appendix.

The implication of our feasibility approach in this case is that feasible tax functions turn out to be parameterized by $\widetilde{w}$. The intuition for this result is quite simple. Consider (for the moment) the case where the distribution of endowments is not bounded above or below. Since the revenue requirement Q

[23] A function g is single-caved if $-g$ is single peaked.

is concave, so is the per capita revenue requirement Q/k. But then, only the tangents to Q/k can be tax functions, since any linear combination of taxes has to be greater than or equal to the per capita requirement. The $\widetilde{w}$'s correspond to the arguments of the per capita revenue functions at the tangency points. The statement of the theorem is slightly more complex because this intuition may not work near the bounds $\underline{w}$ and $\overline{w}$.

Note that the marginal rates in branch B are lower than the rates in branches A and C (the tangent branches), that in turn are lower than those in branch D. In the argument-additivity case, concavity implies that per-capita revenue requirements decrease with the polarization of the draw.

Notice that the shape of the distribution of endowments f does not have in itself any relevant information to predict the shape of the income tax schedules chosen by majority rule, since we have not used it anywhere. Revenue requirements function R is all that is needed.[24]

4 Single Crossing Optimal Tax Functions

Next, some results from the literature on optimal income taxation and implementation theory are used to construct the best income tax function that implements a given individual revenue requirement. The discussion will be informal, but made formal in the lemmas and their proofs.

The problem confronting a *worker/consumer* of type w given net income schedule γ is $\max_l u(\gamma(w \cdot l), l, x, w)$. Using the particular form of utility that we have specified, the first order condition from this problem is $\frac{d\gamma}{dy} \cdot w + \partial b/\partial l = 0$. Rearranging,

$$\frac{d\gamma}{dy} = -\frac{\partial b(l,w)}{\partial l} \cdot \frac{1}{w}. \tag{3}$$

For this tax schedule, we want the consumer of type w to pay exactly the taxes due, which are $g(w)$ for some $g \in G^*$. If g is strictly increasing, g is invertible. If we assume (for the moment) that $g(w)$ is continuously differentiable, then g^{-1}, which maps tax liability to ability (or wage), is well-defined and continuously differentiable. Substituting into the last expression,

[24] With these preliminary results in hand, it would be possible to prove that a majority rule equilibrium exists for the endowment economy where there is no choice of labor supply; voting occurs over feasible taxes, then types are truthfully revealed to the planner for tax purposes. Since this not our main aim, for the sake of brevity it is omitted.

$$\frac{d\gamma}{dy} = -\frac{\partial b(\frac{y}{g^{-1}(y-\gamma)}, g^{-1}(y-\gamma))}{\partial l} \cdot \frac{1}{g^{-1}(y-\gamma)} \equiv \Phi(\gamma, y). \qquad (4)$$

As in Berliant (1992), a standard result from the theory of differential equations yields a family of solutions to this differential equation.[25] Berliant and Gouveia (2001) show that (4) has global solutions if $g' > 0$, $g(\underline{w}) \geq 0$.[26]

Of course, as L'Ollivier and Rochet (1983) point out, the second order conditions must be checked to ensure that solutions to (4) do not involve bunching, which means that consumers do optimize in (4) at the tax liability given by g.[27] This was done in Berliant and Gouveia (2001), where the Revelation Principle[28] was used to construct strictly increasing post tax income

[25]The method used above originates with the signaling model in Spence (1974), further developed by Riley (1979) and Mailath (1987). Equation (4) is best seen as defining an indirect mechanism where gross income is the signal sent by each agent to the planner, much as in Spence's model education is the signal sent to the firm. However, finding the equilibria of this game is only part of the problem. The remaining part of the problem relates to implementation. By this we mean that the social planner's problem is to define reward/penalty functions that induce each type of agent to choose, in equilibrium, the behavior the planner desires of that type of agent. A reference closer to our work is Guesnerie and Laffont (1984). However, there is a difference between our results and the other literature on implementation using the differentiable approach to the revelation principle. The difference is that in the other literature the principal cares only about implementing the action profiles of the agents (labor supply schedules in our model). In contrast, we consider the implementation of explicit maps from types to tax liability. That is, the principal cares about agents' types, which are hidden knowledge. These maps from types to tax liability are not action profiles, and are motivated by the ability to pay approach in classical public finance. They play the same role here as reduced form auctions play in the auction literature.

[26]The relationship between individual and aggregate revenue constraints, if R is the fixed aggregate revenue requirement, is given in the continuum of agents model by:

$$\int_{\underline{w}}^{\overline{w}} g(w)f(w)dw \geq R$$

The point of Berliant and Gouveia (2001) is exactly to study implementation of individual revenue requirements in terms of an income tax, show that they are Pareto ranked, and to examine the properties of individual revenue requirements functions or their implementations that correspond to optimal income taxes when there is an aggregate revenue constraint.

[27]That is, we have a separating equilibrium.

[28]In (4) the planner first chooses a net income function $\gamma(y)$, the agents then take the chosen net income function as given and maximize utility by selecting a gross income level y (or the corresponding level of labor supply). This is the implementation approach described in Laffont (1988). The Revelation Principle allows us to write an equivalent mechanism where agents are simply asked to report their type w. It is easier to check second order conditions of the problem for this direct mechanism. They essentially say that both pre and

functions $\theta(w) = y(w) - g(w)$ that implement $g(w)$, where $g'(w) > 0$. Since we then have that $y(w)$ is invertible, we immediately obtain $\gamma(y) = \theta(w(y))$ and $\tau(y) = g(w(y))$.

It is almost immediate from this development that the set of solutions to (4) for a given g is Pareto ranked. We focus on the best of these for each given g. Define

$$\begin{aligned} T^* \equiv \{ & \tau \in T | \ \gamma \text{ is a solution to (4) for some } g \in G^*, \ \tau(y) = y - \gamma(y), \\ & \text{and } \tau \text{ Pareto dominates all other solutions to (4) for the given } g\}. \end{aligned}$$

Any element of T^* has the property that the marginal tax rate for the top ability $\overline{w}$ consumer type is zero.

From a practical viewpoint, for instance in solving examples such as those presented here, the use of these techniques and in particular equation (4) makes sense. However, for the general theory, in our application we do not have the conditions required by Berliant and Gouveia (2001); for example, the standard boundary condition is not satisfied due to the quasi-linear form of utility. Thus, we use Berliant and Page (1996), which is more general than Berliant and Gouveia (2001).

Lemma 3: If each minimal, feasible individual revenue requirement function is non-decreasing, then any such function is implementable by an income tax. Without loss of generality, for all $w \in [\underline{w}, \overline{w}]$, $\frac{1}{w} \cdot \frac{\partial b(\frac{y(w)}{w}, w)}{\partial l} \geq -1$. Moreover, the implementations are Pareto ranked, and there is a best one under the Pareto ranking with the property that the marginal tax rate for the top ability consumer type, if it exists, is zero.

Proof: See the Appendix.

Remarks: The theorem says that any non-decreasing and feasible revenue requirement function can be implemented by a continuum of tax schedules. These tax schedules are Pareto ranked and furthermore a maximal tax schedule under the Pareto ranking exists. The result on the top marginal tax rate is extended to non-differentiable functions in Berliant and Page (1996), but is a little complicated and, in fact, irrelevant to our purpose here.

The next step is to characterize a class of individual revenue requirements for which we will be able to obtain results. This class contains the cases discussed in Theorem 1 and may possibly include other sets of assumptions.

---

post tax incomes should be increasing functions of w. In our case they are strictly increasing functions and there is no bunching.

Definition: A collection $\mathbf{E}$ of functions mapping $[\underline{w}, \overline{w}]$ into $\mathbb{R}$ is called *strongly single crossing* if each $g \in \mathbf{E}$ is:

1. Continuous.
2. Twice continuously differentiable except possibly at a finite number of points.
3. $dg/dw > 0$ except possibly at a finite number of points.
4. Individual revenue requirements cross each other only once, i.e. for any pair $g, g' \in \mathbf{E}$, there exist $\widehat{w}, \widehat{w}' \in [\underline{w}, \overline{w}]$, $\widehat{w} < \widehat{w}'$ such that $g(\underline{w}) > g'(\underline{w})$ implies $g(w) > g'(w)$ for all $w \in [\underline{w}, \widehat{w})$, $g(w) = g'(w)$ for all $w \in [\widehat{w}, \widehat{w}']$ and $g(w) < g'(w)$ for all $w \in (\widehat{w}', \overline{w}]$.

Definition: A collection T' of functions mapping non-negative incomes to tax liabilities is called *single crossing* if for all $\tau, \tau' \in T^*$, letting $y(\cdot), y'(\cdot)$ be the gross income functions associated with τ and τ', respectively, for incomes $y_1, y_2, y_3 \in y([\underline{w}, \overline{w}]) \cap y'([\underline{w}, \overline{w}])$, $y_1 < y_2 < y_3$, $\tau(y_3) < \tau'(y_3)$ and $\tau(y_2) > \tau'(y_2)$ implies $\tau(y_1) \geq \tau'(y_1)$.

Lemma 4 proves that when individual revenue requirements are strongly single crossing, the income tax systems in T^* cross at most once.

Lemma 4: Suppose that minimal individual revenue requirements, G^*, are strongly single crossing. Then their best implementations T^* are single crossing.

Proof: See the Appendix.

Remarks: The notion of strongly single crossing is the analog of condition (SC) of Gans and Smart (1996) in this specific context. Outside of $y([\underline{w}, \overline{w}])$, τ can be extended in an arbitrary fashion subject to incentive compatibility, for example in a linear way.

Lemmas 3 and 4 are used to prove Lemma 5:

Lemma 5: Suppose that R implies strongly single crossing minimal individual revenue requirements, G^*. Then a majority rule equilibrium exists.

Proof: See the Appendix.

Strongly single crossing is used intensively to prove this. It has the implication that induced preferences over tax systems have properties shared by single peaked preferences over a one dimensional domain. The winners will be the tax systems most preferred by the median voter (in the draw) out of tax systems in T^*.

The proof consists of two parts. The first part shows that there is a tax schedule that is weakly preferred to all others by the median voter. The second part shows that this tax schedule is a majority rule winner. This second part could be replaced by Gans and Smart (1996, Theorem 1). But it would take as much space to verify the assumptions of that Theorem as it does to prove our more specialized result directly.

A referee has asked, "Can majority rule equilibrium and optimal tax functions be the same?" There are two answers, depending on whether the interpretation of *optimal* is Pareto optimal or utilitarian optimal.

Consider first the interpretation as Pareto optimal Then, in our restricted set of third best tax functions (satisfying our notion of feasibility for all draws), evidently any majority rule equilibrium will be third best efficient, since an objection of the grand coalition is necessarily an objection of a majority. The hard part here is proving existence of a majority rule winner, not showing that the majority rule winner is Pareto optimal.

Consider next the interpretation of efficiency as utilitarian optima, as is common in the optimal tax literature. The answer is related to Bowen's (1943) theorem and the extension by Bergstrom (1979). We focus on the quasi-linear and additively separable (in public goods and labor supply) case. First consider the expenditure or public good side of the problem, which is closer to this literature. Without getting into the technicalities, the main intuition for the primary result, namely Bergstrom (1979, Theorem 3), is that if the marginal willingness to pay for the public good of the median voter is equal to the mean marginal willingness to pay for the draw, then an efficient allocation can be achieved by a majority rule equilibrium. As Bergstrom remarks, this would be rare, and in our view is likely in the complement of a generic set of utility functions. But clearly it is possible.

Inspired by this literature, we can use a similar argument here, in the context of labor supply. We must modify it because we address utility levels rather than first order conditions for optima, and because labor supply is a private good. Assume that

$$b(l, w) = \widehat{b}(l \cdot w) = \widehat{b}(y)$$

Then the (indirect) utility function of every agent (neglecting public good level, that is irrelevant at this stage of the game) is:

$$y - \tau(y) + \widehat{b}(y)$$

The first order condition for incentive compatibility is:

$$1 - \frac{d\tau(y)}{dy} + \frac{d\widehat{b}(y)}{dy} = 0$$

The interpretation is that for each income tax τ, each type of agent earns the same income and enjoys the same level of utility, though they work different hours.[29] That is because the optimized gross income $y = w \cdot l$ is the same for each type w, but since w varies, so does l. Hence, what the median voter selects as a best tax out of a feasible set will also be unanimously best and thus utilitarian best. Again, this seems to be a possible but rare occurrence.

5 Examples

Next we provide a pair of simple examples that can be solved.

Example 1: Take

$$\begin{aligned} u(c, l, x, w) &= c - l^2 - \phi w \cdot \frac{x^{-2}}{2} \\ H(x) &= \frac{x^2}{2} \end{aligned}$$

The marginal cost of the public good is x. The marginal willingness to pay of type w for the public good is $\phi w \cdot x^{-3}$, so the total marginal willingness to pay for the draw $(w_1, w_2, ..., w_k)$ is $\phi x^{-3} \sum_{i=1}^{k} w_i$. Setting this equal to marginal cost to solve for the Pareto efficient level of public good provision (that will be unique), we obtain:

$$x^*(w_1, w_2, ..., w_k) = \left(\phi \sum_{i=1}^{k} w_i \right)^{\frac{1}{4}}$$

A reason why the isoelastic case might be interesting comes from the fact that it is a suitable case for the purpose of carrying out empirical tests of the model, given that the correct way to aggregate abilities (or tastes) in this particular case is simply to sum them.

[29] We neglect second order conditions here for brevity and simplicity.

The aggregate revenue requirement function is:

$$R(w_1, w_2, ..., w_k) = H\left(x^*(w_1, w_2, ..., w_k)\right) = \frac{1}{2}\sqrt{\phi \sum_{i=1}^{k} w_i}$$

Next, take $\underline{w} = 1$, $\overline{w} = 2$, and let $\widetilde{w}$ be the median type of a draw. Then as in Lemma 2, if $k \geq 2$ and $1.5 \leq \widetilde{w} \leq 2$, the minimal individual revenue requirements are indexed by $\widetilde{w}$ and given by[30]

$$\begin{aligned} \overline{g}(w, \widetilde{w}) &= \sqrt{\phi}\left[\frac{1}{2}\left(k\widetilde{w}\right)^{\frac{1}{2}}/k + \frac{1}{4}\left(k\widetilde{w}\right)^{-\frac{1}{2}} \cdot (w - \widetilde{w})\right] \\ &= \frac{\sqrt{\phi}}{4}\left[\sqrt{\frac{\widetilde{w}}{k}} + \sqrt{\frac{1}{k\widetilde{w}}} \cdot w\right] \end{aligned}$$

In an endowment economy, this is the tax on endowments most preferred by type $\widetilde{w}$ among those satisfying the aggregate revenue constraints. In this particular case, it is a linear tax. The next step is to implement it in an optimal income tax economy.

Applying the first order approach to incentive compatibility[31] given in the differential equation (4) above, $\frac{d\tau}{dy} = 1 - \frac{d\gamma}{dy}$, and $g^{-1}(y - \gamma) = w$, the income tax function is given by the solution to:[32]

$$\frac{d\tau}{dy} = 1 - \frac{2y}{w^2}$$

Inverting $\overline{g}(\cdot, \widetilde{w})$ and solving for w in terms of τ,

$$w = 4\sqrt{\frac{k\widetilde{w}}{\phi}}\tau - \widetilde{w}$$

so

$$\frac{d\tau}{dy} = 1 - \frac{2y}{\left[4\sqrt{\frac{k\widetilde{w}}{\phi}}\tau - \widetilde{w}\right]^2}$$

This ordinary differential equation has a solution through every point. To choose the best of these, take the one that has the marginal tax rate zero for the top type $\overline{w} = 2$. For the top type, it is the solution that goes through $(\overline{\tau}, \overline{y}) = \left(\frac{\sqrt{\phi}}{4\sqrt{k\widetilde{w}}}\left[2 + \widetilde{w}\right], 2\right)$. This will be the majority rule equilibrium for any

[30] To keep calculations simple, we focus on draws where the median is at least 1.5.

[31] The second order condition for incentive compatibility will be satsified because $\frac{\partial g(w;\widetilde{w})}{\partial w} > 0$.

[32] Although we know that a solution exists and through any point it is unique, actually solving the ODE explicitly is another matter entirely.

draw with median $\widetilde{w} \geq 1.5$. Several comparative statics can be derived in this example: The total tax paid by the top type is decreasing in the size of the economy (k), but increasing in the marginal utility of public good (ϕ). The income of the top type is independent of both the size of the economy (k) and the marginal utility of the public good (ϕ). This is a consequence of the separability assumption.

Example 2: One point of this example is that although we will restrict to quasi-linear utility functions for the general theory, that might not be necessary. Take

$$\begin{aligned} u(c, l, x, w) &= \min\left(c, w \cdot [1 - l]\right) - w \cdot \frac{x^{-2}}{2} \\ H(x) &= \frac{x^2}{2} \end{aligned}$$

The aggregate revenue requirements function, and thus $\overline{g}(\cdot, \widetilde{w})$, is unchanged from Example 1. Setting $c = w \cdot [1 - l]$,

$$y - \tau = w - y$$

Therefore,

$$\begin{aligned} \tau(y) &= 2y - w \\ &= 2y - 4\sqrt{\frac{k\widetilde{w}}{\phi}} \cdot \tau(y) + \widetilde{w} \end{aligned}$$

and thus

$$\tau(y) = \frac{2y + \widetilde{w}}{1 + 4\sqrt{\frac{k\widetilde{w}}{\phi}}}$$

Comparative statics are apparent.

Remarks: *The single crossing of individual revenue requirements results from the combination of: the assumptions on utility and cost of the public good, and the idea that the aggregate revenue requirements must be satisfied for any draw.* We prove in Lemma 4 that when we implement the individual revenue requirements and impose second best efficiency, the single crossing property is inherited by the income tax implementations. One common feature of our individual revenue requirement functions is that there is a *switch point*, indexed by $\widetilde{w}$ in our examples here, that represents the individual revenue requirement that minimizes that type's tax liability among all individual revenue requirements satisfying the aggregate revenue requirements for all draws. This is not actually necessary for our general results, and is not used in the proofs once we obtain single crossing of individual revenue requirements. However, as seen from Example 1, provided that g is strictly increasing, the optimization point for type w under the (optimal) income tax framework will correspond to tax liability $\overline{g}(w, \widetilde{w})$. Therefore, using the standard diagrams from optimal tax theory,[33] the majority rule equilibrium will correspond to the best implementation (solution to the ordinary differential equation) of the revenue requirement function that minimizes the tax liability of the median type of the draw, $\overline{g}(w, \widetilde{w})$. Thus, the switch point is inherited by the optimal income tax implementation of the individual revenue requirements. The fact that we do not use the switch point once we have single crossing of individual revenue requirements allows room for expansion of our results.

6 Conclusions

Two different but related issues deserve some discussion at the outset. The first is whether information on the likelihood of each draw can be used. The second is how to deal with possible excess revenues. As for the opposite situation of insufficient revenues, the reader should note that imposing a penalty for not meeting the requirement simply results in a new revenue requirement function.

We first discuss the information issue. One obvious possibility would be to define as feasible all individual revenue requirement functions that generate an *expected* revenue equal to or larger than the *expected* collective revenue requirement. In contrast with what we use, this would be a single constraint rather than a constraint for each draw. The problem with this notion is that single crossing conditions for individual revenue requirements would likely fail to be satisfied for most cases, including the ones studied in this paper. But one could consider weakening our feasibility restriction and still have enough "bite" to generate single crossing g's. Here is a suggestion:

One option is to use a class of probability measures over draws and constrain the expectation of revenues for each probability measure. Expected revenue according to f would be one particular member of this class. The class could be chosen to generate a continuum of constraints, binding enough for the single crossing result to survive, and we would be back to our initial setup although with different feasibility conditions. This is similar to a model of government behavior using ambiguity aversion or Knightian uncertainty. Perhaps this could be justified as a way to aggregate risk averse voter preferences over

[33] See Seade (1977).

budget deficits.

We now address the issue of excess revenue. Consider first the case of utility quasi-linear in consumption good that we have used throughout this paper. It is possible to return the *ex post* excess revenue in a lump-sum fashion, as there are no income effects. Or the government could use them for another purpose, such as production of yet another public good.

When we consider general preferences and technologies the problem becomes more difficult. Clearly, the excess revenue cannot be returned to taxpayers in a lump sum fashion, as it will affect their behavior in optimizing against the income tax. However, once we deviate from quasi-linear utility, other issues would arise before we get to this point, most importantly the presence of multiple Pareto optimal levels of public good provision. From the point of view of applications, analysis of these more general models will be much more difficult.

The bottom line is whether the alternative models have more to offer. Is it better to restrict ourselves to fixed revenue and voting over a parameter of a prespecified functional form for taxes (as in the previous literature), which are also generally Pareto dominated, or is the model proposed here a useful complement? Differences of opinion are clearly possible.

We note here that unlike much of the earlier literature on voting over *linear* taxes, the majority equilibria are not likely to be linear taxes without strong assumptions on utility functions and on the structure of incentives. The reason is simple: in the optimal income tax model, Pareto optimality requires that the top ability individuals face a marginal tax rate of zero.[34] Therefore, poll taxes are the only linear taxes that could possibly be equilibria. In our model, such taxes are not generally majority rule equilibria, since consumers at the lower ability end of the spectrum will object. All majority rule equilibria derived in this paper are second best Pareto optimal (for a given individual revenue requirement), and hence satisfy the property that the top marginal tax rate is zero.

In that sense, the results obtained here are a step forward relative to Romer (1975) and Roberts (1977). In another sense, they also improve on Snyder and

[34] We know of only one case where an optimal tax is linear: Snyder and Kramer (1988). But this and other results derived in that paper are due to the use of a peculiar model that departs significantly from the other models used in the study of income taxation. There are no income nor substitution effects on effort induced by taxation up to the point where workers switch to the underground sector, and from that point on the same holds since, by definition, income realized in the underground sector is not taxed.

Kramer (1988) by using a standard optimal income tax model as the framework to obtain the results.

Although majority rule equilibrium taxes in our model have the property that the top marginal tax rate is zero, a property in common with second best Pareto efficient taxes, in our model the majority rule equilibrium taxes satisfy a revenue requirement for each draw (or equivalently, an individual revenue requirement), stronger than a single aggregate revenue requirement. Thus, our majority rule equilibrium taxes are not necessarily second best Pareto efficient. Notice also that our majority rule equilibrium differs by draw, whereas the second best Pareto efficient tax systems for the aggregate distribution from which the population is drawn does not. However, if we restrict to tax systems and public good levels that are feasible in our sense, majority rule equilibria will be Pareto efficient for the draw among the feasible tax systems in our restricted sense; the latter are third best.

There are a few strategies that may be productive in pursuing research on voting over taxes.[35] One strategy is to use probabilistic voting models such as in Ledyard (1984). Another is to take advantage of the structure built in this paper and, with our results in hand, look at multi-stage games in which players' actions at the earlier stages might transmit information about types. Of course, it might be necessary to look at refinements of the Nash equilibrium concept to narrow down the set of equilibria to those that are reasonable (at least imposing subgame perfection as a criterion).

A two-stage game of interest is one in which k is fixed and each player in a draw proposes a tax system in T^* (simultaneously). The second stage of the game proceeds as in the single stage game above, with voting restricted to only those tax systems in T^* that were proposed in the first stage.

A three stage game of interest is one in which k is again fixed and the players in a draw elect representatives and who then propose tax systems and proceed as in the two stage game (see Baron and Ferejohn (1989)).

Work remains to be done in obtaining comparative statics results, as in the examples. Finally, the predictive power of the model will be the subject of empirical research. That will certainly be the focus of future work.

[35] For a discussion of dynamics in both the general literature and a model closely related to this one, see Berliant and Boyer (2021).

7 Appendix

7.1 Proof of Lemma 2

For part (b), it is straightforward to prove by direct calculation that $\forall \widetilde{w} \in [\underline{w}, \overline{w}]$, $\overline{g}(w, \widetilde{w})$ is continuous everywhere, continuously differentiable in w (except where the transition is made between branches), and has positive derivative in w. Below, we will also use the fact that for $\forall w \in [\underline{w}, \overline{w}]$, $\overline{g}(w, \widetilde{w})$ is continuously differentiable in $\widetilde{w}$ (except where the transition is made between branches).

To provide some intuition for the next part of the proof, it is important to inquire: Where does the relation $w \lesseqgtr \widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})$ (for branches A and B) come from? It simply describes when it is possible or impossible to construct a draw containing w so that the average ability is equal to $\widetilde{w}$. Beyond that, the function $Q(kw)/k$ will be crucial in the proof below.

Now we proceed to prove part (a). First we shall show that $\overline{g}(\cdot, \widetilde{w})$ is feasible and minimal for each $\widetilde{w}$, in other words it is in G^*, and then show that there are no other functions that are feasible and minimal. Fix a draw $(w_1, w_2, ..., w_k)$. To consider branches A and B separately to begin, define:

$$\begin{aligned} I_L &\equiv \{i \mid w_i \leq \widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})\} \\ I_H &\equiv \{i \mid w_i > \widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})\} \end{aligned}$$

Let k_L be the number of elements of I_L. Further, let

$$w_L \equiv \frac{1}{k_L} \sum_{i \in I_L} w_i$$

Since Q is concave, if $w_i \leq \widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})$, branch A yields that $\overline{g}(w, \widetilde{w})$ is the tangent line to $Q(kw)/k$ at $w = \widetilde{w}$:

$$\overline{g}(w_i, \widetilde{w}) = Q(k \cdot \widetilde{w})/k + Q'(k \cdot \widetilde{w})(w_i - \widetilde{w}) \geq Q(kw_i)/k.$$

Therefore, since $\overline{g}(w, \widetilde{w})$ is linear in w on branch A,

$$\sum_{i \in I_L} \overline{g}(w_i, \widetilde{w}) = k_L \cdot \overline{g}(w_L, \widetilde{w})$$

So *if* for all $i = 1, ..., k$, $i \in I_L$, then

$$\sum_{i=1}^{k} \overline{g}(w_i, \widetilde{w}) = k \cdot \overline{g}(w_L, \widetilde{w}) \geq Q(kw_L) = Q(\sum_{i=1}^{k} w_i) = R(w_1, ..., w_k)$$

In particular, if we set the draw to be $w_i' \equiv \min\{w_i, \widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})\}$, then

$$\sum_{i=1}^{k} \overline{g}(w_i', \widetilde{w}) \geq R(w_1', ..., w_k') \tag{5}$$

For branch B, let $\widehat{w}_j \in [\widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w}), \overline{w}]$ for $j \in I_H$ and $\widehat{w}_j = w_j$ for $j \in I_L$. Using concavity of Q and setting $i \in I_H$,

$$\frac{\partial}{\partial w_i} \sum_{j \in I_H} \overline{g}(w_j, \widetilde{w}) \mid_{w_j = \widehat{w}_j} = Q'((k-1)\underline{w} + \widehat{w}_j) \geq Q'(\sum_{j=1}^{k} \widehat{w}_j) = \frac{\partial}{\partial w_i} Q(\sum_{j=1}^{k} w_j) \mid_{w_j = \widehat{w}_j \forall j}$$

Integrating via a line integral and applying Stokes' theorem,

$$\begin{aligned}
& \sum_{j \in I_H} [\overline{g}(w_j, \widetilde{w}) - \overline{g}(\widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w}), \widetilde{w})] \\
= & \sum_{j \in I_H} \int_{\widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})}^{w_j} \frac{\partial}{\partial \widehat{w}_j} \overline{g}(\widehat{w}_j, \widetilde{w}) d\widehat{w}_j \\
\geq & \sum_{j \in I_H} \int_{\widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})}^{w_j} Q'(\sum_{j=1}^{k} \widehat{w}_j) d\widehat{w}_j \\
= & Q(\sum_{j=1}^{k} w_j) - Q(\sum_{j=1}^{k} w_j') \\
= & R(w_1, ..., w_k) - R(w_1', ..., w_k')
\end{aligned}$$

Adding this inequality to (5) yields feasibility of $\overline{g}(\cdot, \widetilde{w})$. Feasibility of branches C and D is proved similarly.

We now prove that the $\overline{g}(\cdot, \widetilde{w})$ are minimal among those feasible. Consider branch A. Clearly, if a draw consists of k individuals of type $\widetilde{w}$, $\overline{g}(\widetilde{w}, \widetilde{w})$ is minimal by definition. To show that $\overline{g}(w, \widetilde{w})$ is minimal, suppose the opposite. Take $h(w)$ with $h(w) \leq \overline{g}(w, \widetilde{w})$ and with strict inequality for some $w_1 \in [\underline{w}, \widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})]$ (the case $w_1 \in (\widetilde{w} + (k-1)(\widetilde{w} - \underline{w}), \overline{w}]$ will be considered in the next paragraph). It is possible to construct a draw $(w_1, w_2, ...w_k)$ with mean $\widetilde{w}$ and $w_i \in [\underline{w}, \widetilde{w} + (k-1) \cdot (\widetilde{w} - \underline{w})]$ for $i = 2, ...k$. Then, $R(w_1, w_2, ...w_k) = Q(k \cdot \widetilde{w}) = \sum_{i=1}^{k} \overline{g}(w_i, \widetilde{w})$. But $\sum_{i=1}^{k} h(w_i) < \sum_{i=1}^{k} \overline{g}(w_i, \widetilde{w})$, so $h(w)$ is not feasible. Similar reasoning holds for branch C.

Now consider branch B. Take $h(w)$ with $h(w) \leq \overline{g}(w, \widetilde{w})$ and with strict inequality for some $w_1 \in (\widetilde{w} + (k-1)(\widetilde{w} - \underline{w}), \overline{w}]$. The logic used for branches A and C does not hold in this case: it is not possible to find $k-1$ ability levels in this interval to construct a draw with mean $\widetilde{w}$. Consider a draw with

$w_j = \underline{w}$ for $j = 2, 3, ...k$. For this draw,

$$\begin{aligned}
&\sum_{i=1}^{k} \overline{g}(w_i, \widetilde{w}) \\
= \;& Q((k-1)\cdot \underline{w} + w_1) - ((k-1)/k)\cdot Q(k\cdot \widetilde{w}) + (k-1)\cdot Q'(k\cdot \widetilde{w})\cdot(\widetilde{w} - \underline{w}) \\
&+ \sum_{i=2}^{k} Q(k\cdot \widetilde{w})/k + Q'(k\cdot \widetilde{w})\cdot(\underline{w} - \widetilde{w}) \\
= \;& Q((k-1)\cdot \underline{w} + w_1) - ((k-1)/k)\cdot Q(k\cdot \widetilde{w}) + (k-1)\cdot Q'(k\cdot \widetilde{w})\cdot(\widetilde{w} - \underline{w}) \\
&+ \frac{k-1}{k} Q(k\cdot \widetilde{w}) - (k-1)Q'(k\cdot \widetilde{w})(\widetilde{w} - \underline{w}) \\
= \;& Q((k-1)\cdot \underline{w} + w_1)
\end{aligned}$$

Hence, for the draw $(w_1, \underline{w}, ..., \underline{w})$, $\sum_{i=1}^{k} h(w_i) < \sum_{i=1}^{k} \overline{g}(w_i, \widetilde{w}) = Q((k-1)\cdot \underline{w} + w_1) = R(w_1, \underline{w}, ..., \underline{w})$, so h is not feasible. Similar reasoning holds for Branch D.

Next, suppose there is $h \in G^*$ that is not of the form given in the statement of the Lemma. Then for each $\widetilde{w} \in [(\underline{w} + \overline{w})/2, \overline{w}]$ there is some $w \in [\underline{w}, \overline{w}]$ with $h(w) < g(w, \widetilde{w})$. Fix any such $\widetilde{w}$. Repeating the arguments just above showing that $g(w, \widetilde{w})$ is minimal in the cases of Branch A ($w \leq \widetilde{w} + (k-1)\cdot(\widetilde{w} - \underline{w})$) and Branch B ($w > \widetilde{w} + (k-1)\cdot(\widetilde{w} - \underline{w})$), it follows that h is not feasible, a contradiction. Similar reasoning hold for Branches C and D.

Next we prove part (c). To prove single cavedness in $\widetilde{w}$, one need only differentiate $\overline{g}(w, \widetilde{w})$ with respect to the parameter $\widetilde{w}$, fixing w. For branches A and C we obtain:

$$\frac{\partial \overline{g}(w, \widetilde{w})}{\partial \widetilde{w}} = Q''(k\cdot \widetilde{w})\cdot k\cdot(w - \widetilde{w}).$$

The derivative above is positive if $w < \widetilde{w}$ and negative for $w > \widetilde{w}$.

For branch B we have:

$$\frac{\partial \overline{g}(w, \widetilde{w})}{\partial \widetilde{w}} = k\cdot(k-1)\cdot Q''(k\cdot \widetilde{w})\cdot(\widetilde{w} - \underline{w}) < 0.$$

which applies only for $w > \widetilde{w} + (k-1)\cdot(\widetilde{w} - \underline{w})$. Otherwise Branch A applies.

Finally, for branch D we get:

$$\frac{\partial \overline{g}(w, \widetilde{w})}{\partial \widetilde{w}} = k\cdot(k-1)\cdot Q''(k\cdot \widetilde{w})\cdot(\widetilde{w} - \overline{w}) > 0.$$

which applies only for $w < \widetilde{w} + (k-1)\cdot(\widetilde{w} - \underline{w})$. Otherwise Branch C applies. The results above imply that $\arg\min_{\widetilde{w}} \overline{g}(w, \widetilde{w}) = w$.

Finally, consider part (d). We claim that these $\overline{g}$'s are single crossing. To see this, first note that from the definition of $\overline{g}(w, \widetilde{w})$ in the statement of the Lemma, direct calculation yields that $\frac{\partial \overline{g}(w,\widetilde{w})}{\partial w}$ is weakly decreasing in $\widetilde{w}$ for each w. Therefore, if $\overline{g}(w, \widetilde{w})$ and $\overline{g}(w, \widetilde{w}')$ cross twice, there exist $w, w', w'' \in [\underline{w}, \overline{w}]$, $w < w' < w''$ such that $\overline{g}(w, \widetilde{w}) = \overline{g}(w, \widetilde{w}')$, $\overline{g}(w', \widetilde{w}) \neq \overline{g}(w', \widetilde{w}')$, $\overline{g}(w'', \widetilde{w}) = \overline{g}(w'', \widetilde{w}')$. But this cannot happen in each case: $\widetilde{w}' = \widetilde{w}$, $\widetilde{w}' < \widetilde{w}$, $\widetilde{w}' > \widetilde{w}$. The first case is obvious, and the other cases use the fundamental theorem of calculus applied to integrals of $\frac{\partial \overline{g}(w,\widetilde{w})}{\partial w}$ and $\frac{\partial \overline{g}(w,\widetilde{w}')}{\partial w}$.

7.2 Proof of Lemma 3

We employ Berliant and Page (1996) Theorems 1 and 2 to prove Lemma 3. Notice first that since $H(x) \geq 0$, if we focus on draws of identical individuals, $R \geq 0$ and $g \geq 0$. So we restrict to taxes that are non-negative. The key assumptions to be verified are the boundary conditions and a single crossing property. We will be precise. For the quasi-linear utility we use, we must verify the following boundary conditions on the utility function, numbered (3) and (4) in that paper:

$$\text{For all } w \in [\underline{w}, \overline{w}], \text{ for all } 0 \leq t' \leq t \leq y, \text{ there exists } y',\ t' \leq y' \leq y \text{ with} \quad (6)$$
$$y' - t' + b(\frac{y'}{w}, w) \leq y - t + b(\frac{y}{w}, w)$$

$$\text{For all } w \in [\underline{w}, \overline{w}], \text{ for all } 0 \leq y \leq w \text{ and } 0 \leq t \leq y, \text{ for all } 0 \leq y' < y, \quad (7)$$
$$\text{there exists } 0 \leq t' \leq y' \text{ with } y' - t' + b(\frac{y'}{w}, w) \leq y - t + b(\frac{y}{w}, w)$$

Fix $w \in [\underline{w}, \overline{w}]$. For condition (6), we must show that there is $y' \leq y$ such that $y' - t' + b(\frac{y'}{w}, w) \leq y - t + b(\frac{y}{w}, w)$. We will show, in addition, that without loss of generality, $\frac{1}{w}\frac{\partial b(\frac{y'}{w},w)}{\partial l} \geq -1$. This will imply that for the implementation, $\frac{1}{w}\frac{\partial b(\frac{y(w)}{w},w)}{\partial l} \geq -1$.

Next we show that if $\frac{1}{w}\frac{\partial b(\frac{y}{w},w)}{\partial l} < -1$, there exists $\widehat{y}$, $t \leq \widehat{y} < y$ with $y - t + b(\frac{y}{w}, w) = \widehat{y} - t + b(\frac{\widehat{y}}{w}, w)$ and $\frac{1}{w}\frac{\partial b(\frac{\widehat{y}}{w},w)}{\partial l} \geq -1$.

First, we know from the boundary condition that $b(1, \underline{w}) - b(0, \underline{w}) \geq -\underline{w}$. Using the fundamental theorem of calculus, $\int_0^1 \frac{\partial b(l,w)}{\partial l} dl = b(1, w) - b(0, w)$. Since $\frac{\partial b(l,w)}{\partial l} < 0$ is weakly increasing in w, $b(1, w) - b(0, w)$ is weakly increasing in w, so $b(1, w) - b(0, w) \geq -\underline{w} \geq -w$.

Rearranging, $w + b(1, w) \geq b(0, w)$. Let y^* solve $\frac{\partial b(\frac{y^*}{w},w)}{\partial l} = -w$ if the solution is positive, set $y^* = 0$ otherwise. Note that $\frac{1}{w}\frac{\partial b(\frac{y''}{w},w)}{\partial l} < -1$ for

$y'' > y^*$, so $y'' + b(\frac{y''}{w}, w)$ is *decreasing* in y''. The worst utility level is given by $y'' = w$, or $w + b(1, w) \geq b(0, w)$ using the start of this paragraph. So by the intermediate value theorem, there is a $\widehat{y}$ with $y + b(\frac{y}{w}, w) = \widehat{y} + b(\frac{\widehat{y}}{w}, w)$ and $\widehat{y} \leq y^*$, so $\frac{1}{w}\frac{\partial b(\frac{\widehat{y}}{w}, w)}{\partial l} \geq -1$, and without loss of generality we can take $y = \widehat{y}$.

Next we prove (6). For $t' = t$, this is trivial (take $y' = y$), so take $0 \leq t' < t \leq y$.

First we show that (6) holds for t' in a neighborhood of t, and then expand to any t'.

Using $\frac{\partial^2 b(l,w)}{\partial l^2} < 0$ and $\frac{1}{w}\frac{\partial b(l,w)}{\partial l} \geq \frac{1}{w}\frac{\partial b(l,w)}{\partial l}\Big|_{l=1} \geq -1$, $\frac{1}{w}\frac{\partial b(l,w)}{\partial l} \geq -1$, so for $\widetilde{y}' < \widetilde{y}$,

$$b(\frac{\widetilde{y}}{w}, w) - b(\frac{\widetilde{y}'}{w}, w) = \int_{\widetilde{y}'}^{\widetilde{y}} \frac{1}{w}\frac{\partial b(\frac{\widehat{y}}{w}, w)}{\partial l} d\widehat{y} > \int_{\widetilde{y}'}^{\widetilde{y}} -1 d\widehat{y} = \widetilde{y}' - \widetilde{y} \tag{8}$$

Now take y' with $0 \leq t' \leq y' < y \leq w$. Using (8),

$$y - t + b(\frac{y}{w}, w) > y' - t + b(\frac{y'}{w}, w)$$

So for a size ϵ neighborhood of t, for all t' with $t - \epsilon \leq t' < t$ in that neighborhood, the conclusion of (6) holds. In fact, given the assumption $\frac{\partial^2 b(l,w)}{\partial l^2} < 0$, the size of the neighborhood expands as y decreases, so that we can use the same ϵ size as applies to y and t. Therefore, given any $t' < t$ and y, we can construct a finite chain $y_1 = y$, $y_2 = y' = t$, $t'_1 = t - \epsilon = t_2$ with $t'_k = t'$ so that the conclusion of (6) holds for each element of the chain and by transitivity holds for the original data, namely y, t, and t'.

The proof of (7) proceeds in the same manner. As shown above, for $y' < y$,

$$y - t + b(\frac{y}{w}, w) > y' - t + b(\frac{y'}{w}, w)$$

Thus, for all (t', y') where $t' \leq t$ and $t' \leq y' \leq y$ in a neighborhood of (t, y), the conclusion of (7) holds. Then construct a finite chain as before.

The single crossing condition in Berliant and Page (1996) has two parts:

$$w' > w,\ y > y' \text{ and } y' - t' + b(\frac{y'}{w'}, w') > y - t + b(\frac{y}{w'}, w') \tag{9}$$
$$\text{implies } y' - t' + b(\frac{y'}{w}, w) > y - t + b(\frac{y}{w}, w)$$
$$w' > w,\ y' > y \text{ and } y' - t' + b(\frac{y'}{w}, w) > y - t + b(\frac{y}{w}, w) \tag{10}$$
$$\text{implies } y' - t' + b(\frac{y'}{w'}, w') > y - t + b(\frac{y}{w'}, w')$$

A sufficient condition for (9) to hold is:

$$b(\frac{y'}{w},w)-b(\frac{y'}{w'},w')\geq b(\frac{y}{w},w)-b(\frac{y}{w'},w')$$

or

$$b(\frac{y}{w'},w')-b(\frac{y'}{w'},w')\geq b(\frac{y}{w},w)-b(\frac{y'}{w},w)$$

We now proceed to prove that.

$$b(\frac{y}{w'},w')-b(\frac{y'}{w'},w')=\int_{y'/w'}^{y/w'}\frac{\partial b(l,w')}{\partial l}dl\geq\int_{y'/w'}^{y/w'}\frac{\partial b(l,w)}{\partial l}dl=b(\frac{y}{w'},w)-b(\frac{y'}{w'},w)$$
$$=\int_{y'}^{y}\frac{\partial b(\frac{\widehat{y}}{w'},w)}{\partial l}\cdot\frac{1}{w'}d\widehat{y}\geq\int_{y'}^{y}\frac{\partial b(\frac{\widehat{y}}{w},w)}{\partial l}\cdot\frac{1}{w}d\widehat{y}=b(\frac{y}{w},w)-b(\frac{y'}{w},w)$$

A sufficient condition for (10) to hold is:

$$b(\frac{y'}{w'},w')-b(\frac{y'}{w},w)\geq b(\frac{y}{w'},w')-b(\frac{y}{w},w)$$

or

$$b(\frac{y'}{w'},w')-b(\frac{y}{w'},w')\geq b(\frac{y'}{w},w)-b(\frac{y}{w},w)$$

This follows from the argument just above, replacing y with y' and y' with y.

7.3 Proof of Lemma 4[36]

Let g' and g'' be the elements of G^*, and let (τ',y',γ') and (τ'',y'',γ'') be the tax, gross income, and net income functions associated with g' and g'', respectively. We ignore x and $r(x,w)$ in this proof, due to additive separability and the structure of the game. The proof is by contradiction. Suppose that there exist incomes $y_1<y_2<y_3$ with $\tau'(y_1)<\tau''(y_1)$, $\tau'(y_2)>\tau''(y_2)$ and $\tau'(y_3)<\tau''(y_3)$. Then by the intermediate value theorem applied to utility differences as a function of w, there exists w^a such that $y'(w^a)-\tau'(y'(w^a))+b(y'(w^a)/w^a,w^a)=y''(w^a)-\tau''(y''(w^a))+b(y''(w^a)/w^a,w^a)$, $y''(w^a)>y'(w^a)$, and $\tau'(y'(w^a))<\tau''(y'(w^a))$. Moreover, $g'(w^a)=\tau'(y'(w^a))<\tau''(y'(w^a))$. Now by (3) and Lemma 3, $\frac{1}{w}\frac{\partial b(\frac{y''(w)}{w},w)}{\partial l}\geq-1$ and $\frac{d\tau''}{dy}=1-\frac{d\gamma''}{dy}\geq 0$ and since $y''(w^a)>y'(w^a)$, $g''(w^a)=\tau''(y''(w^a))\geq\tau''(y'(w^a))>\tau'(y'(w^a))=g'(w^a)$. There also exists $w^b>w^a$ with $y'(w^b)-\tau'(y'(w^b))+b(y'(w^b)/w^b,w^b)=y''(w^b)-\tau''(y''(w^b))+b(y''(w^b)/w^b,w^b)$, $y'(w^b)>y''(w^b)$, $\tau'(y''(w^b))>\tau''(y''(w^b))$, and $\tau''(y'(w^b))>\tau'(y'(w^b))$. Hence $\tau'(y''(w^b))>\tau''(y''(w^b))=g''(w^b)$ and since $y'(w^b)>y''(w^b)$, $g'(w^b)=\tau'(y'(w^b))\geq\tau'(y''(w^b))>\tau''(y''(w^b))=g''(w^b)$.

[36] To see how this critical proof works, it is useful to draw the graphs from optimal taxation, net income as a function of gross income, that are standard in the literature; see Seade (1977).

Using strongly single crossing, $g'(\overline{w}) > g''(\overline{w})$.

By construction of T^*, $\tau'(y'(\overline{w})) > \tau''(y''(\overline{w}))$. Note that since the marginal tax rate at $y'(\overline{w})$ and $y''(\overline{w})$ is zero, what we have are essentially lump sum taxes at the top ability level. Hence, $y''(\overline{w})-\tau''(y''(\overline{w}))+b(y''(\overline{w})/\overline{w},\overline{w}) > y'(\overline{w}) - \tau'(y'(\overline{w})) + b(y'(\overline{w})/\overline{w},\overline{w})$. Zero income effects implies $y'(\overline{w}) = y''(\overline{w})$. Moreover, $\tau'(y''(\overline{w})) > \tau''(y'(\overline{w}))$. Since $\tau''(y'(w^b)) > \tau'(y'(w^b))$, there exists y^*, $y''(\overline{w}) > y^* > y'(w^b)$ with $\tau'(y^*) = \tau''(y^*)$, so there exists w^c with $y'(w^c)-\tau'(y'(w^c))+b(y'(w^c)/w^c,w^c) = y''(w^c)-\tau''(y''(w^c))+b(y''(w^c)/w^c,w^c)$, $y''(w^c) > y'(w^c)$, and $\tau'(y'(w^c)) < \tau''(y'(w^c))$. As above, $g'(w^c) = \tau'(y'(w^c)) < \tau''(y'(w^c))$ and since $y''(w^c) > y'(w^c)$, $g''(w^c) > g'(w^c)$.

This contradicts strongly single crossing. So the hypothesis is false, and the lemma is established.

7.4 Proof of Lemma 5

Definition: Let C^1 be the space of continuously differentiable functions (with domain $[\underline{w},\overline{w}]$ and range $\mathbb{R}$) endowed with the uniform topology. We consider T^* to be a subset of this space by extending any $\tau \in T^*$ to the whole domain, if necessary, in a C^1 and linear fashion.

Fix $\tau \in T^*$. First we claim that $0 \leq \frac{d\tau}{dy} \leq 1$. From $\partial b/\partial l \leq 0$ a.s., (3), and Lemma 3, we obtain that $0 \leq \frac{d\gamma}{dy} = -\frac{1}{w}\frac{\partial b}{\partial l} \leq 1$. Since $\frac{d\tau}{dy} = 1 - \frac{d\gamma}{dy}$, the claim is proved. So every $\tau \in T^*$ is Lipschitz in income with constant 1, and thus T^* is equicontinuous. Since $k \cdot g(w) \geq R(w,w,...,w) \geq 0$, T^* is also norm bounded by $\overline{w}$. Using Ascoli's theorem (see Munkres (1975, p. 290)), $\overline{T}^*$ (the closure of T^* in C^1) is compact.[37]

Fix k and let $(w_1, w_2, ..., w_k) \in \mathcal{A}$. For any $\tau \in T$, let $v(\tau, w) = \max_y u(y - \tau(y), y/w)$, the utility induced by the tax system τ for type w. It is easy to verify that for each w, $v(\tau, w)$ is continuous in its first argument.

Let τ^* be a maximal element of $\overline{T}^*$ using $v(\cdot, w^M)$ as the objective, where w^M is the median ability level in $(w_1, w_2, ..., w_k)$ if k is odd, and $w^M \in [w_{k/2}, w_{k/2+1}]$ (where the wage rates are ordered in an increasing fashion) if k is even. Using Lemmas 2 and 3, $\tau^* \in T^*$.

Now suppose there exists $\tau \in T$ such that there is a subset D of $\{w_1, w_2, ..., w_k\}$ with $v(\tau, w) > v(\tau^*, w)$ for all $w \in D$ and where the cardinality of D is greater than $k/2$. Then using Lemma 3, we can take τ to be in T^* without loss of

[37] An alternative proof, pointed out by a referee, would show that the best implementations of $\overline{g}(w, \widetilde{w})$ are continuous in $\widetilde{w}$.

generality. Using Lemma 4, $\tau^*$and τ are single crossing, or alternatively, their after tax income functions $y - \tau^*(y)$ and $y - \tau(y)$ are single crossing. Notice also that, due to non-inferiority of consumption good, indifference curves (as a function of gross and after tax income) are single crossing in w; see Seade (1977, footnote 8). Thus, there exist intervals $W, W' \subseteq [\underline{w}, \overline{w}]$ such that W and W' partition $[\underline{w}, \overline{w}]$ and $D \subseteq W$. Let $\widehat{W}$ be the smallest interval (in the sense of set inclusion) such that $\widehat{W}$ and its complement $\widehat{W}'$ are both intervals, $\widehat{W}$ and $\widehat{W}'$ partition $[\underline{w}, \overline{w}]$, and $D \subseteq \widehat{W}$.

Then by definition of τ^*, $w^M \notin \widehat{W}$. Hence D cannot contain a majority of the draw, a contradiction. Hence the hypothesis is false and τ^* cannot be defeated by any other feasible tax system.

7.5 Proof of Theorem 1

For a draw $(w_1, w_2, ..., w_k) \in \mathcal{A}$, the Lindahl-Samuelson condition for this model is:

$$\sum_{i=1}^{k} w_i \cdot s'(\widehat{r}(x)) \cdot \widehat{r}'(x) = m \cdot \widehat{r}'(x)$$

Hence,

$$x = \widehat{r}^{-1}\left[s'^{-1}\left(\frac{m}{\sum_{i=1}^{k} w_i}\right)\right]$$

and thus

$$R(w_1, w_2, ..., w_k) = m \cdot s'^{-1}\left(\frac{m}{\sum_{i=1}^{k} w_i}\right)$$

Hence, R is argument additive. Computing the first derivative,

$$\frac{dR}{d\sum_{i=1}^{k} w_i} = -m^2 \cdot \frac{1}{s''\left(s'^{-1}\left(\frac{m}{\sum_{i=1}^{k} w_i}\right)\right) \cdot \left(\sum_{i=1}^{k} w_i\right)^2} > 0$$

Computing the second derivative,

$$\frac{d^2R}{d\left(\sum_{i=1}^{k} w_i\right)^2}$$

$$= \quad m^2 \cdot \frac{2s''\left(s'^{-1}\left(\frac{m}{\sum_{i=1}^{k} w_i}\right)\right) \cdot \left(\sum_{i=1}^{k} w_i\right) + s'''\left(s'^{-1}\left(\frac{m}{\sum_{i=1}^{k} w_i}\right)\right) \cdot \frac{-m}{s''\left(s'^{-1}\left(\frac{m}{\sum_{i=1}^{k} w_i}\right)\right)}}{\left[s''\left(s'^{-1}\left(\frac{m}{\sum_{i=1}^{k} w_i}\right)\right) \cdot \left(\sum_{i=1}^{k} w_i\right)^2\right]^2}$$

$$= \quad m^2 \cdot \frac{2s''(r) \cdot \left(\sum_{i=1}^{k} w_i\right) + s'''(r) \cdot \frac{-m}{s''(r)}}{\left[s''(r) \cdot \left(\sum_{i=1}^{k} w_i\right)^2\right]^2}$$

where

$$r = s'^{-1}\left(\frac{m}{\sum_{i=1}^{k} w_i}\right)$$

Thus,

$$\begin{aligned}
\frac{d^2R}{d\left(\sum_{i=1}^{k} w_i\right)^2} &< 0 \text{ if and only if} \\
2s''(r) \cdot \left(\sum_{i=1}^{k} w_i\right) &< s'''(r) \cdot \frac{m}{s''(r)} \text{ or} \\
2s''(r)^2 &> s'''(r) \cdot s'(r)
\end{aligned}$$

The last expression holds by assumption. Therefore, R is argument additive with negative second derivative. The result then follows from Lemmas 2 and 5.

References

[1] Aumann, R.J. and M. Kurz, 1977, Power and taxes, Econometrica 45, 1137-1161.

[2] Baron, D. and J. Ferejohn, 1989, Bargaining in legislatures, American Political Science Review 83, 1181-1206.

[3] Bergstrom, T.C., 1979, When does majority rule supply public goods efficiently?, The Scandinavian Journal of Economics 81, 216-226.

[4] Bergstrom,T.C. and R. Cornes, 1983, Independence of allocative efficiency from distribution in the theory of public goods, Econometrica 51, 1753-1765.

[5] Berliant, M., 1992, On income taxation and the core, Journal of Economic Theory 56, 121-141.

[6] Berliant, M. and P.C. Boyer, 2021, Politics and income taxes: Progress and progressivity. Working paper.

[7] Berliant, M. and M. Gouveia, 2001, Incentive compatible income taxation, individual revenue requirements and welfare. https://mpra.ub.uni-muenchen.de/84374/1/MPRA_paper_84374.pdf

[8] Berliant, M. and H. Konishi, 2005, Salience: Agenda choices by competing candidates, Public Choice 125, 129-149.

[9] Berliant, M. and F.H. Page, Jr., 1996, Incentives and income taxation: The implementation of individual revenue requirement functions, Ricerche Economiche 50, 389-400.

[10] Berliant, M. and F.H. Page, Jr., 2001, Income taxes and the provision of public goods: Existence of an optimum, Econometrica 69, 771-784.

[11] Berliant, M. and F.H. Page, Jr., 2006, Budget balancedness and optimal income taxation, Journal of Public Economic Theory 8, 409-451.

[12] Berliant, M. and P. Rothstein, 2000, On models with an uncongestible public good and a continuum of consumers, Journal of Urban Economics 48, 388-396.

[13] Bierbrauer, F.J., 2011, On the optimality of optimal income taxation, Journal of Economic Theory 146, 2105-2116.

[14] Bierbrauer, F.J., Tsyvinski, A. and N. Werquin, 2021, Taxes and turnout: When the decisive voter stays home, CESifo Working Paper No. 8954.

[15] Bishop, K.C. and C. Timmins, 2019, Estimating the marginal willingness to pay function without instrumental variables, Journal of Urban Economics 109, 66-83.

[16] Bowen, H.R., 1943, The interpretation of voting in the allocation of economic resources, The Quarterly Journal of Economics 58, 27-48.

[17] Brett, C. and J.A. Weymark, 2017, Voting over selfishly optimal nonlinear income tax schedules, Games and Economic Behavior 101, 172-188.

[18] Calvert, R., 1986, Models of Imperfect Information in Politics (Harwood Academic Publishers, Chur).

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

[20] Foley, D., 1967, Resource Allocation and the Public Sector, Yale Economic Essays 7, 45-98.

[21] Gans, J.S. and M. Smart, 1996, Majority voting with single-crossing preferences, Journal of Public Economics 59, 219-237.

[22] Gouveia, M., 1997. Majority rule and the public provision of a private good, Public Choice 93, 221-244.

[23] Guesnerie, R. and J.-J. Laffont, 1984, A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm, Journal of Public Economics 25, 329-369.

[24] Guesnerie, R. and C. Oddou, 1981, Second best taxation as a game, Journal of Economic Theory 25, 67-91.

[25] Heathcote, J. and H. Tsujiyama, 2021, Optimal income taxation: Mirrlees meets Ramsey, Journal of Political Economy 129, 3141-3184.

[26] Hettich, W. and S.L. Winer, 1988, Economic and political foundations of tax structure, American Economic Review 78, 701-713.

[27] Hirsch, M., 1976, Differential Topology (Spinger-Verlag, New York, NY).

[28] Kaneko, M., 1981, On the existence of an optimal income tax schedule, Review of Economic Studies 48, 633-642.

[29] Kapička, M., 2013, Efficient allocations in dynamic private information economies with persistent shocks: A first-order approach, Review of Economic Studies 80, 1027-1054.

[30] Kocherlakota, N.R. and Y. Song, 2019, Public goods with ambiguity in large economies, Journal of Economic Theory 182, 218-246.

[31] Laffont, J.-J., 1988, Fundamentals of Public Economics (MIT Press, Cambridge, MA).

[32] Laroque, G.R., 2005, Indirect taxation is superfluous under separability and taste homogeneity: A simple proof, Economics Letters 87, 141-144.

[33] Ledyard, J., 1984, The pure theory of large two-candidate elections, Public Choice 44, 1-43.

[34] Lensman, T. and M. Troshkin, 2020, Implications of uncertainty for optimal policies, Journal of Economic Theory, forthcoming.

[35] Lindahl, E., 1919, Positive lösung, die gerechtigkeit der besteuerung, translated as Just taxation — a positive solution, in: R.A. Musgrave and A.T. Peacock, eds., Classics in the Theory of Public Finance (Macmillan, London).

[36] L'Ollivier, S. and J.-C. Rochet, 1983, Bunching and second order conditions: A note on optimal tax theory, Journal of Economic Theory 31, 392-400.

[37] Mailath, G., 1987, Incentive compatibility in signaling games with a continuum of types, Econometrica 55, 1349-1365.

[38] Meltzer, A. and S. Richard, 1981, A rational theory of the size of government, Journal of Political Economy 89, 914-927.

[39] Meltzer, A. and S. Richard, 1983, Tests of a rational theory of the size of government, Public Choice 41, 403-418.

[40] Mill, J.S., 1848, Principles of Political Economy (Longmans Green, London).

[41] Mirrlees, J.A., 1971, An exploration in the theory of optimum income taxation, Review of Economic Studies 38, 175-208.

[42] Munkres, J.R., 1975, Topology: A First Course (Prentice-Hall, Englewood Cliffs, NJ).

[43] Musgrave, R.A., 1959, The Theory of Public Finance (McGraw Hill, New York).

[44] Nakayama, M., 1976, A note on the core of an income redistribution game, Journal of Economic Theory 12, 347-349.

[45] Peck, R.M., 1986, Power and linear income taxes: An example, Econometrica 54, 87-94.

[46] Plott, C.R., 1967, A notion of equilibrium and its possibility under majority rule, American Economic Review 67, 787-806.

[47] Riley, J., 1979, Informational equilibrium, Econometrica 47, 331-359.

[48] Roberts, K.W.S., 1977, Voting over income tax schedules, Journal of Public Economics 8, 329-340.

[49] Röell, A., 1996, Voting over nonlinear income tax schedules, mimeo.

[50] Roemer, J.E., 1999, The democratic political economy of progressive income taxation, Econometrica 67, 1-19.

[51] Romer, T., 1975, Individual welfare, majority voting, and the properties of a linear income tax, Journal of Public Economics 4, 163-186.

[52] Rudin, W., 1974, Real and Complex Analysis (McGraw-Hill, New York, NY).

[53] Samuelson, P.A., 1954, The pure theory of public expenditure, Review of Economics and Statistics 36, 387-389.

[54] Samuelson, P.A., 1955, Diagrammatic exposition of a theory of public expenditure, Review of Economics and Statistics 37, 350-356.

[55] Schofield, N., 1978, Instability of simple dynamic games, Review of Economic Studies 45, 575-594.

[56] Seade, J.K., 1977, On the shape of optimal tax schedules, Journal of Public Economics 7, 203-235.

[57] Snyder, J.M. and G.H. Kramer, 1988, Fairness, self-interest, and the politics of the progressive income tax, Journal of Public Economics 36, 197-230.

[58] Stiglitz, J.E., 2015, In praise of Frank Ramsey's contribution to the theory of taxation, The Economic Journal 125, 235-268.

[59] Spence, M., 1974, Market Signaling (Harvard, Cambridge, MA).