CESifo

Working Papers

www.cesifo.org/wp

When a Price is Enough:
Implementation in Optimal Tax Design

Sander Renes
Floris T. Zoutman

CESifo
Center for Economic Studies & Ifo Institute

When a Price is Enough:
Implementation in Optimal Tax Design

Abstract

This paper studies the design of tax systems that implement a planner's second-best allocation in a market economy. An example shows that the widely used Mirrleesian (1976) tax system cannot implement all incentive-compatible allocations. Hammond's (1979) "principle of taxation" proves that any incentive-compatible allocation can be implemented through at least one tax system. However, this tax system is often undesirable since it severely restricts the choice space of agents in the economy. In this paper we derive necessary and sufficient conditions to verify whether a given tax system can implement a given incentive-compatible allocation. We show that when an incentive-compatible allocation is on the Pareto frontier, and/or surjective onto the choice space, a tax system that equates the marginal tax rates to the optimal wedges can implement the second best, without restricting the choice space of the agents. It follows that the Mirrleesian tax system can successfully implement the second best in the identified classes. Since the second-best allocation of welfarist planners is always on the Pareto frontier, our results (ex post) validate most tax systems proposed in the literature. Outside of the identified classes, the planner may need to restrict the choice space of agents to implement its second best in the market. This sheds new light on rules, quotas and prohibitions used in real-world tax and benefit systems.

JEL-Code: H210, H220, D820, H240.

Keywords: optimal non-linear taxation, redistribution, tax system, market implementation, price mechanism, private information.

Sander Renes*
Erasmus School of Economics
Erasmus University Rotterdam
PO box 1738
The Netherlands – 3000 DR Rotterdam
srenes@ese.eur.nl

Floris T. Zoutman
NHH Norwegian School of Economics
Department of Business and
Management Science
Norway – 5045 Bergen
floris.zoutman@nhh.no

*corresponding author

1 Introduction

Governments in modern welfare states have developed elaborate systems of taxes, benefits, and subsidies (further: tax system). These tax systems are designed to raise revenue for the provision of public goods, to redistribute resources from the fortunate to the less fortunate, to insure constituents against adverse outcomes when insurance markets fail, and to correct for externalities. Tax systems combine non-linear tax rates on labor income, capital income and commodities, with non-linear subsidies on healthcare, housing, education and other observable choices of the constituents.

The formal study of optimal non-linear tax systems was pioneered by Mirrlees (1971, 1976). In his model agents are heterogeneous in their earnings ability. They make several observable choices, such as labor income, portfolio and savings choices, and consumption levels of various goods. The social planner wants to redistribute from agents with high to agents with low earnings ability. However, earnings ability is private information, and hence, the first-best allocation is not attainable. Instead, the main problem in optimal taxation is to find a tax system that implements the social planner's second-best allocation in a market-based economy.

The technique, pioneered by Mirrlees, to identify the optimal tax system is to split the problem into two subproblems. The first subproblem is to characterize the best attainable outcome given the informational restrictions, the second-best allocation, through a direct mechanism. The second subproblem is to identify a tax system under which the second-best allocation is a market equilibrium, and hence, a tax system that implements the second-best allocation in a market-based economy.

There is an extensive literature that studies the first subproblem and describes the properties of the second-best allocation. In addition, Mirrlees (1976) proposes to solve the second subproblem of designing the optimal tax system by equating the marginal tax rates to the optimal wedge between the marginal rates of substitution and transformation (further: equating marginal tax rates to wedges). In his model agents differ only in their earnings ability, such that equating marginal tax rates to wedges yields a separable tax system in which the marginal tax rate on each good only depends on the consumption of that particular good. The Mirrleesian tax system has become the canoncial solution in, and has subsequently been applied throughout, the literature on optimal non-linear taxation (see e.g. Atkinson and Stiglitz, 1976, Diamond, 1998, Saez, 2001, Bovenberg and Jacobs, 2005, and Choné and Laroque, 2010).

By equating the marginal tax rate to the wedge, the Mirrleesian tax system solves the first-order conditions for implementation. However, to the best of our knowledge, general conditions under which this tax system satisfies the second-order conditions for market implementation have never been derived. To understand this issue in more detail note that a tax system can only implement the second-best allocation in the market if no agent can increase his utility by deviating from the bundle that would have been assigned to him in the direct mechanism. However, agents typically have more choices available to them in a market economy than they have in a direct mechanism. In the direct mechanism of Mirrlees (1976, sect. 3), for example, agents can choose which earnings ability they report to the planner. Based on this reported ability the planner assigns each agent a bundle of choice variables, such as labor effort, consumption levels and savings. On the market the agent can make each of these choices separately, only constrained by his budget. The combination of these choices allows agents to create bundles that are not available to them in the direct mechanism. Hence, even if an allocation is incentive compatible in

the direct mechanism, and the tax system satisfies first-order implementability conditions it does not follow that the tax system implements the allocation in a market economy. The bundles that are not part of the second-best allocation are known in the literature as "joint" or "double" deviations (see e.g. Kocherlakota, 2005). The potential for these joint deviations implies that second-order implementability constraints on the market are generally more stringent than the incentive-compatibility constraints in the direct mechanism.

This issue has not gone entirely unnoticed. Specifically, the principle of taxation derived by Hammond (1979) proves that for any incentive-compatible allocation there exists at least one tax system that satisfies first- and second-order implementability constraints.[1] This tax system stops agents from creating joint deviations in the market by simply prohibiting such deviations. The incentives created by this tax system are mathematically equivalent to the incentives in a direct mechanism. Hence, incentive compatibility in the direct mechanism implies implementability through this tax system.

Hammond's result is of importance to the study of optimal tax systems because it shows that there always exists at least one tax system that can implement the second-best allocation in the market. However, restricting the choices of economic agents to the choices available to them in a direct mechanism effectively removes the benefits of free choice and limited administrative costs associated with a market mechanism. Therefore, this tax system has limited appeal to policy makers in market economies.[2] Moreover, the principle of taxation does not tell us whether less restrictive tax systems, like the Mirrleesian tax system, can implement the second-best allocation.

In addition, in recent literature several papers have focused on identifying the second-best allocation when agents differ in multiple dimensions (see e.g. Cremer et al. (2001), Saez (2002a), Kleven et al. (2009), Choné and Laroque (2010), Renes and Zoutman (2014), Rothschild and Scheuer (2014) and Jacquet and Lehmann (2014)). However, to our knowledge the second subproblem of designing the optimal non-linear tax system in these models has not yet been addressed.

In this paper we address both gaps in our understanding of tax design. We derive conditions under which the Mirrleesian tax system, and indeed any tax system that satisfies first-order implementability conditions, successfully implements the second-best allocation in the market. These results hold independent of the dimension of heterogeneity. In models of optimal taxation where these conditions apply the optimal tax system follows directly from the optimal wedges in the second-best allocation. If the planner adjusts relative prices in the market by equating taxes to wedges, this suffices and the planner does not need to restrict the choice space of agents in the market.

The analytic starting point of this paper is a set of agents that differ in multiple continuously distributed characteristics, such as earnings ability and tastes. The agents make multiple choices pertaining, for instance, consumption levels of several goods and income from labor and capital. Preferences are described by a completely general utility function. The economy is guided by a planner who wants to implement a second-best allocation of which we know three things: i.) it satisfies the economy's resource constraint,

---

[1]A similar result has been derived in Maskin (1999) for any countable number of agents.

[2]The name 'principle of taxation' stems from later applications of the Hammond's result to taxation by e.g. Rochet (1985), Guesnerie (1995) and Bierbrauer (2009). However, the principle can also apply to the multi-product monopolist pricing problem and auction design (see e.g. Armstrong, 1996). In such a setting the application leads to perfectly realistic implementations, since the monopolist and the auctioneer can choose what (not) to produce/sell and how to bundle their goods.

ii.) it is incentive compatible in a direct mechanism, iii.) it maximizes a general welfare function under the constraints given by the other two conditions. It is well-known in the literature that the derivation of the second-best allocation under multi-dimensional heterogeneity of agents is technically very complex. In the companion paper, Renes and Zoutman (2014), we take up this issue and set the first steps towards fully characterizing the second-best allocation with multi-dimensional heterogeneity of agents. In this paper, however, we take the second-best allocation as given, and use the common elements of such allocations to study sufficient conditions for the implementation of second-best allocations in a market economy.

We first show the relevance of this analysis trough a simple, but instructive example. The example shows that even when agents only differ in a uni-dimensional characteristic the canonical Mirrleesian tax system may fail to implement the second best. In our example at least one agent prefers a bundle available to him in the market over the bundle assigned to him in the direct mechanism. The planner can only implement its second best by restricting these joint deviations or taxing them at prohibitive rates. Neither the restrictions nor the prohibitive tax rates can be derived from the second-best allocation. Hence, in this particular example designing the optimal tax system is significantly more complex than deriving the second-best allocation.

We proceed by deriving a lemma which describes the general conditions under which a tax system implements the desired allocation using standard micro-economic theory. The first order implementation constraints require that marginal tax rates are equal to optimal wedges. The second-order implementation constraints require that indifference curves are more convex than budget constraints in all linear combinations of the decision variables. Economists can use these implementability constraints to verify whether a proposed tax system implements the desired allocation. That is, after solving the maximization problem of the planner and formulating the entire tax system, it can be verified whether the tax system satisfies these constraints for the proposed allocation. Unfortunately, most optimal allocations in the literature do not have a closed-form solution. Therefore, verification of implementation can only be performed in the special cases that have been simulated. This verification is useful (and necessary) in such simulations, but it does not provide insights in the general properties of optimal tax systems.

Therefore, our main contribution lies in identifying two classes of optimal tax problems in which a tax system can always implement the allocation, provided that tax rates are equated to wedges. In the identified classes joint deviations are never optimal, and hence, the second-order implementation constraints are automatically satisfied. This implies there is no need to impose restrictions, or levy prohibitive tax rates, on the choice space available to agents in the market, and that implementation does not have to be verified ex post.

The first class of models satisfies the following conditions: i.) the allocation is feasible, incentive compatible and on the Pareto frontier, ii.) there are no externalities iii.) tax rates are equated to wedges, and iv.) the resulting tax system does not have an internal maximum. For this class of problems we show that in any allocation that allows profitable joint deviations, at least one joint deviation exists which increases the utility of the agent and weakly increases tax revenue. Such a deviation entails a Pareto improvement over the original allocation. Hence, the initial allocation could not have been on the Pareto frontier. For optimal taxation, the most important implication of this first class of models is that the second-best allocation of a welfarist planner can always be implemented by equating taxes to wedges, since the second best of a welfarist planner always resides on

the Pareto frontier.

A simple corollary to this proposition shows that the Mirrleesian tax system can implement the second-best allocation under uni-dimensional heterogeneity of agents. It follows that a separable tax system can implement the second best in models with uni-dimensional heterogeneity provided the four conditions outlined above are satisfied.

Intuitively, if the planner is welfarist the preferences of the planner and the agent are aligned in the sense that the objective function of the planner is, *ceteris paribus*, increasing in the utility of the agent. Hence, budget-neutral joint deviations that increase the agents' utility cannot exist in a market mechanism. It is this property of alignment between the objective of the planner and the agents that allows the planner to implement its second best through a relatively simple tax system.

If the tax function has an internal maximum on the allocation, the proposition fails to apply. Intuitively, from such a maximum a deviation that weakly increases tax revenue does not necessarily exist. However, most tax systems that result from equating taxes to wedges are either monotonic or convex in all choice variables and as such they usually do not exhibit internal maxima.

More importantly, the proposition fails for non-welfarist planners or in the presence of externalities. The second-best allocation of a non-welfarist planner is not necessarily on the Pareto frontier (see Kaplow and Shavell, 2001). Similarly, in the case of externalities, even if both the utility for the agent and tax revenue weakly increase, this does not necessarily imply a Pareto improvement, since other agents might be adversely affected through the externality.

The second class of problems where a tax system that equates taxes to wedges can always implement the second-best allocation, is when the second-best allocation is surjective onto the choice space. Intuitively, when each bundle that exists in the market is assigned to at least one type in the direct mechanism, the agents' problem on the market is identical to the agents' problem in the direct mechanism. Therefore, incentive-compatibility and implementability constraints coincide. Since we assumed the original allocation was incentive compatible, it must also be implementable in the market. The prime example of this class of models is the Mirrlees (1971) model where agents only differ in earnings ability and their only choice variables are consumption and labor income.

Our main results hold independently of the number of choice variables, and independently of the number of characteristics in which agents differ. As such, our results have strong implications for the extensively studied models of optimal tax design with uni-dimensional heterogeneity, as well as for the technically more complex study of optimal taxation under multi-dimensional heterogeneity. Moreover, preferences do not have to satisfy a Spence-Mirrlees or single-crossing condition, since our results remain valid even if there is bunching, such that at least two types receive the same bundle, in the optimal allocation.

Note that the conditions we derive are sufficient conditions, not necessary conditions. There are cases outside of these two classes for which a mechanism that equates taxes to wedges (without additional restrictions) suffices to implement the second-best allocation. Implementation cannot be guaranteed ex ante through our propositions in that case. The two identified classes, however, are of enormous importance since they encompass virtually all non-stochastic models based on Mirrlees (1971, 1976) and the generalization of these models in Renes and Zoutman (2014) and as a result validate almost all tax systems proposed in the literature.[3]

---

[3]A notable exception is Jacobs and De Mooij (2011) which extends the Mirrlees model with external-

5

This paper primarily provides a guide to the relatively understudied second step of optimal tax design. More generally, our results show an imperfect link between direct and indirect mechanisms. A central planner that perfectly observes choices and can price/tax them non-linearly, might still want to rely on quotas or legal prohibitions to reach the second-best allocation. This provides some intuition for the existence of (possibly optimal) complexities in the tax systems in modern welfare states. To prevent abuse of social insurance schemes, a central planner may have to restrict the choices of (potential) beneficiaries, and force them to study, apply for jobs, or enroll in debt counseling for instance.

The rest of the paper is organized as follows. Section 2 discusses related literature. Section 3 introduces the model. Section 4 contains an example that shows the potential problems in implementation. Section 5 derives our main results and section 6 concludes.

2 Related Literature

A large and growing literature has been devoted to deriving second-best allocations under multi-dimensional heterogeneity. Early examples of multi-dimensional screening problems include Mirrlees (1976), Armstrong (1996), Rochet and Choné (1998) and Armstrong and Rochet (1999), and recently there have been several applications in the optimal taxation problem, including Cremer et al. (2001), Saez (2002a), Kleven et al. (2009), Choné and Laroque (2010), Renes and Zoutman (2014), Rothschild and Scheuer (2014) and Jacquet and Lehmann (2014). The tax systems that implement the second-best allocation in the market are usually left implicit. This paper complements this literature by guiding the design of a tax system that implements the second-best allocations in the market.

In recent literature it has become popular to derive the optimal tax system using a perturbation method (see e.g. Roberts, 2000, Saez, 2001 and Jacquet et al., 2013). With the perturbation method one can directly derive the optimal tax system in the market without resorting to the direct mechanism. However, the perturbation method is mainly applied as a heuristic tool, to find first- but not second-order optimality conditions. Moreover, the perturbation method does not formally show that the derived tax system also achieves the second-best allocation. Hence, in practice articles that use the perturbation method usually supplement their analysis with a formal proof applying the direct mechanism. Therefore, the insights derived in this paper apply equally to articles that use the perturbation method to identify the optimal tax system.

Stiglitz (1987) studies implementation in a setting with discrete types. His analysis shows that a discrete distribution of characteristics adds another layer of complexity, since the optimal wedge is determined for each type in the economy, but not for the 'holes' between the types. The insights derived in this paper do not apply directly to the case with discrete types, and we leave this issue for future research.

Our paper also relates to the more general problem of implementation theory, which identifies economic mechanisms that can implement the planner's optimal allocation (see e.g. Hurwicz, 1960, Vickrey, 1961, Hurwicz, 1973, Dasgupta et al., 1979, Myerson, 1979 and Maskin, 1999). In many applications it has been shown that the planner should apply restrictions to the choice space of the agents. In particular, in optimal taxation Guesnerie and Roberts (1984) and Hammond (1979) shows that quantity restrictions on some goods are in the general case welfare improving, even if the planner sets its linear,

---

ities.

respectively, non-linear tax rates optimally. Adams and Yellen (1976) and Armstrong (1996) show that the multi-product monopolist should optimally bundle its products in order to maximize its profits, effectively restricting the combinations of goods available to customers. A similar result has been found in the literature on multi-product auctions, where the auctioneer should optimally bundle its products in order to maximize auction revenue (see e.g. Palfrey, 1983, Chakraborty, 1999, Armstrong, 2000, and Jehiel et al., 2007). Moreover, it has been shown that if several goods are procured or auctioned in a single contract, but the planner is uncertain about the exact quantities required at the time of the procedure, the principal should optimally restrict the range of possible bids. This prevents agents from using skewed bids, bids where the price is too high for one good, and too low for another (see e.g. Athey and Levin, 2001, Ewerhart and Fieseler, 2003, and Renes, 2011). We add to the implementation theory literature by deriving two special cases where the planner can implement its optimum in a market mechanism without imposing any restrictions on the choice space of the agents, provided the planner has access to non-linear taxation/pricing. In particular, we show that when preferences of the planner and the agents are sufficiently aligned, the planner does not need to restrict the choice set of the agents. Hence, the optimal restrictions in the applications above may stem from a misalignment of preferences between the planner and the agents.

The New Dynamic Public Finance has generalized the Mirrlees model to a setting where earnings ability follows a stochastic dynamic process.[4] Kocherlakota (2005) shows that in this setting equating the tax rate to the optimal wedge generally does not implement the second-best allocation, and hence, the planner should restrict the choice set of the agents. Albanesi and Sleet (2006) show that even in the simplest setting, where earnings ability follows an iid process, the planner can implement the optimal allocation only if it supplements income taxes by borrowing constraints. We provide an intuition for this result. As Kocherlakota (2010) a.o. argues, in dynamic stochastic models of taxation, savings by an agent in period t create an "externality" on the labor supply decision of future incarnations of this agent through the wealth effect. Our first proposition does not apply when there are externalities, and hence, the planner may need to restrict the savings decisions of agents in order to implement its second-best allocation. Intuitively, the externality causes a misalignment between the preferences of the agents and the planner, and hence, restrictions are required.

3   The Model

In this section we lay down the formal structure of our model. First, we define the preferences of the agents in the economy. Second, we define the general properties of a second-best allocation. Third, we define the agents' maximization problem on the market. Finally, we discuss the two methods of implementation that are most commonly applied in the literature: the principle of taxation due to Hammond (1979) and the Mirrleesian tax system due to Mirrlees (1976).[5]

---

[4]See Golosov et al. (2007) and Kocherlakota (2010) for an overview of the literature.

[5]The model description below closely follows that of the companion paper Renes and Zoutman (2014).

3.1 Preferences

The economy is populated by a unit mass of agents characterized by a twice-differentiable utility function:

$$u\left(\mathbf{x}, y, \mathbf{n}\right), \tag{1}$$

where $\mathbf{x} \in \mathbf{X} \subseteq \mathcal{R}^k$ is a vector of choice variables, $y \in Y \subseteq \mathcal{R}$ is an untaxed numeraire choice variable, and $\mathbf{n} \in \mathbf{N} \subseteq \mathcal{R}^p$ is the type of an individual. Variables in $\mathbf{x}$ can include e.g. effective labor supply, consumption of (different) commodities, or savings. Choice variables $\mathbf{x}$ and y are observable at the individual level and the social planner can tax all choices in $\mathbf{x}$ non-linearly, but cannot tax y.[6] Throughout the paper we will sometimes refer to the choice variables in $\{\mathbf{x}, y\}$ as goods, even though they can be both inputs and outputs to the production process.

We assume that the untaxed good y is a normal good such that $u_y > 0, u_{yy} \leq 0$ for all values of $\{\mathbf{x}, y, \mathbf{n}\}$. Note that the choice of the numeraire variable has no effect on the optimal allocation, since a tax on y can always be replicated by a uniform tax on all goods in $\mathbf{x}$. Hence, the choice of the untaxed good does not affect the optimal allocation. Therefore, the assumption that y is a normal good is equivalent to the assumption that among all the choice variables available to the agents, there is at least one normal good.

The assumption that y is a normal good directly implies that the utility function is non-satiated everywhere. In addition, it eases interpretation of the results in the remainder of the paper. Since y will act as a numeraire, whether a good is taxed or subsidized can be evaluated by direct comparison to the numeraire good y. Moreover marginal preferences for all goods are neatly summarized by the vector of marginal rates of substitution with respect to the numeraire good:

$$\mathbf{s}\left(\mathbf{x}, y, \mathbf{n}\right) \equiv -\frac{u_{\mathbf{x}}\left(\mathbf{x}, y, \mathbf{n}\right)}{u_y\left(\mathbf{x}, y, \mathbf{n}\right)}.$$

Here element s_i is the negative of the marginal rate of substitution for choice variable x_i with respect to the numeraire y. Therefore, s_i represents the marginal utility loss of receiving an extra unit of x_i, expressed in units of the numeraire variable y. In case of the choice of labor supply, for example, the marginal utility of providing an extra unit of labor is usually assumed negative, it then follows that s_i will be positive in our notation.

Each element in the type vector $\mathbf{n}$ is referred to as a characteristic. Characteristics may include variables such as ability and health status as well as taste parameters. We assume $\mathbf{n}$ follows a multi-dimensional differentiable cumulative distribution function $F\left(\mathbf{n}\right)$, with $F : \mathbf{N} \to [0, 1]$ and a probability density function $f\left(\mathbf{n}\right)$. Both are defined over the closure of the convex type space $\mathbf{N}$.

The type is private information to each individual and unobservable to the planner. Note that we do not restrict ourselves to static models. Different choices can occur in different periods. However, we do assume that both the type and the mechanism used by the planner are revealed to the individuals before their first choice.

---

[6]To be able to apply non-linear taxes the planner has to observe each choice variable at the individual level. Hammond (1987) and Guesnerie (1995) show that if anonymous transactions in a good are possible, the planner can only tax that good at a linear rate. Mirrlees (1976, sect. 5), Christiansen (1984), and Boadway and Jacobs (forthcoming) extend the non-linear tax framework to allow for goods that can be traded anonymously, and as such should be taxed linearly. Although, the main intuition of our approach remains valid in this setting, formally extending our model to allow for goods that can be traded anonymously is beyond the scope of this paper.

Our utility function allows a very general description of preferences. Note, for example, that the conventional utility representation (e.g. Mirrlees, 1971, Saez, 2001) $\tilde{u}(y,l)$, where l is labor supply and y is income, is a special case of our utility representation.[7] In addition, we do not require the utility function to be (weakly) separable in any of the goods or characteristics as is usually assumed in the literature. Therefore, agents' preferences for a good may depend on all elements in the type of the agent, $\mathbf{n}$, as well as other consumption choices. Moreover, the utility function does not need to satisfy a single-crossing condition, since our analysis remains valid even if in the optimal allocation types are bunched. Finally, we do not assume any relationship between the number of characteristics p and the number of goods k other than $k, p \geq 1$.

3.2 Incentive Compatibility and Feasibility

The problem of optimal non-linear taxation is to find a tax system that implements the second-best allocation of a social planner in a market economy. Following the methodology of Mirrlees (1971, 1976) this requires us to solve two subproblems. The first subproblem is to find the second-best allocation using a direct mechanism. The second subproblem is to find the optimal tax system that implements the second-best allocation in the market. This paper focuses on the second problem, but before we can find the optimal tax system, we first need to outline the properties of a second-best allocation.

A second-best allocation that has been derived through a direct mechanism must satisfy the following three criteria: i.) the allocation is resource feasible, ii.) the allocation is incentive compatible, and iii.) the allocation maximizes the central planner's objective function under the constraints given by the first two conditions.

Formally, these properties can be defined as follows. Let the second-best allocation of goods be denoted by:

$$\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\} \quad \forall \quad \mathbf{n} \in \mathbf{N}.$$

Here $\mathbf{x}^*$ and y^* is a function mapping from the type space to the good space, $\mathbf{x}^* : \mathbf{N} \to \mathbf{X}$ and $y^* : \mathbf{N} \to Y$. We assume $\mathbf{x}^*(\cdot)$ and $y^*(\cdot)$ are both twice differentiable in all their arguments.[8]

Let $\mathbf{X}^*$ denote the image or range of function $\mathbf{x}^*$, and Y^* the image of y^*. We denote by $\{\mathbf{X}, Y\}^*$ the image of the allocation. This image contains the collection of all bundles assigned to agents in the economy. By definition, the set of assigned bundles are a subset of the goods space, $\mathbf{X}^* \subseteq \mathbf{X}$, $Y^* \subseteq Y$. The sets $\mathbf{X}^*$ and $\mathbf{X}$ (Y^* and Y), are equal if each possible value of choice variables is assigned to at least one agent. Similarly, $\{\mathbf{X}, Y\}^* = \{\mathbf{X}, Y\}$ if all possible combinations of goods are assigned to at least one agent.

Since the economy should be able to produce all goods consumed, we know the optimal allocation must satisfy the economy's resource constraint. We assume the economy's resource constraint takes the form:

$$\int_{\mathbf{N}} y^*(\mathbf{n}) \, dF(\mathbf{n}) + R = \int_{\mathbf{N}} q(\mathbf{x}^*(\mathbf{n})) dF(\mathbf{n}). \tag{2}$$

---

[7]To see this, assume that gross income $x_1 = n_1 l$ where n_1 is earnings ability. It can readily be seen that this utility function can be rewritten into our form: $\tilde{u}(y,l) = \tilde{u}\left(y, \frac{x_1}{n_1}\right) = u(x_1, y, n_1)$.

[8]This assumption makes the rest of the analysis easier. However, by non-satiation of the utility function we know that individual budget constraints are binding, while theorem 1 of Clausen and Strub (2013) guarantees that the FOC binds with equality on interior choices. The combination of these two constraints imply that on any region of the type space that separates our tax system is well defined and our proof structure holds.

In this equation, R is the exogenous revenue requirement of the planner measured in units of y and $q : \mathbf{X} \to Y$ is a reduced form production function that describes the production of y as a function of the total consumption of goods in $\mathbf{x}$. The equation states that total production of y should equal the sum of private consumption and the revenue requirement of the numeraire good. Since the price of the numeraire good is normalized to one, the derivative q_{x_i} represents the individual (negative) marginal rate of transformation between good x_i and the numeraire good y. q_{x_i} Is positive when good x_i is an input and negative when x_i is an output of the production process. We assume weakly decreasing returns to scale such that all $q_{x_i x_i}$ are non-positive. Note that the conventional assumption in the literature that $q(.)$ is linear[9] is a special case of our model. The generalization of the production function allows researchers to investigate the case where individuals face decreasing returns to scale because they, for example, have a lower marginal productivity as they supply more hours of labor. An allocation that satisfies condition (2) is said to be feasible.

Incentive compatibility requires that in a direct revelation game each agent reveals his type. In a revelation game the planner presents the agents with a menu of choices over bundles, $\{\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m})\}$, from which each agent chooses his preferred bundle by sending the corresponding p-dimensional message about his type, $\mathbf{m} \in \mathbf{N}$, to the planner. An allocation $\{\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m})\}$ is incentive compatible if each individual truthfully reveals all his unobserved characteristics through message $\mathbf{m}$ and receives the bundle designed for him. That is, incentive compatibility implies the agent maximizes his utility by sending the message $\mathbf{m} = \mathbf{n}$.

Hence, an incentive-compatible and feasible allocation can be defined as follows:

Definition 1 *An allocation $\{\mathbf{x} = \mathbf{x}^*(\mathbf{n}), y = y^*(\mathbf{n})\} \quad \forall \quad \mathbf{n} \in \mathbf{N}$ is incentive compatible and feasible if each agent truthfully reveals his entire type in a direct mechanism:*

$$\mathbf{n} = \arg \max_{\mathbf{m}} u(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}) \quad \forall \mathbf{n} \in \mathbf{N}. \tag{3}$$

and in addition satisfies constraint (2).

Note that incentive compatibility does not imply that there is no bunching of types. Two different types may well receive exactly the same bundle of goods, provided it is in both types best interest to choose this bundle over all other bundles in $\{\mathbf{X}, Y\}^*$.

A second-best allocation must maximize the objective function of the planner among the set of incentive-compatible and feasible allocations defined in definition 1. However, at this point we deliberately leave the objective function of the social planner entirely generic, such that the second-best allocation of the social planner may reside anywhere within the incentive-compatible and feasible set. This allows us to study implementation in the most generic setting possible. We will later show that whether or not an allocation can be implemented can crucially depend on the assumptions made about the objective function of the planner.

Characterizing the optimal allocation when agents differ in multiple dimensions may be very complex and is beyond the scope of this paper. The common approach is to assume that the second-order incentive constraints are not binding. Mirrlees (1976); McAfee and McMillan (1988); Renes and Zoutman (2014) describe characteristics of optimal incentive-compatible allocations in the context of a first-order approach where

---

[9]See e.g. Mirrlees (1971), Mirrlees (1976), Diamond (1998), Saez (2001) and Jacquet et al. (2013).

second-order conditions of maximization problem have to be checked after the entire allocation is derived, while Rochet and Choné (1998) deals with the problems caused by second-order incentive constraints. However, in this paper we are not interested in deriving the second-best allocation. We start instead from an incentive-compatible and feasible allocation that satisfies definition 1. Following proposition 2 in Renes and Zoutman (2014) we know that such an allocation can be described by the optimal wedges:

$$\mathcal{W}_i(\mathbf{n}) = q_{x_i}(\mathbf{x}^*(\mathbf{n})) - s_i(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \quad \forall \mathbf{n} \in \mathbf{N}. \tag{4}$$

These wedges represent the difference between the marginal rate of substitution, s_i, and the marginal rate of transformation, q_{x_i}, for each agent and each good on the second-best allocation. If the wedge is positive the marginal rate of transformation is larger than the marginal rate of substitution such that the distorted consumption level of the good is below its laissez-faire value. If the wedge is negative the consumption of a good is above its laissez-faire value.

3.3 Market Implementation

The social planner aims to implement the second-best allocation in a market economy through a tax system. In the market, agents maximize their utility function (1) with respect to their choice variables $\mathbf{x}$ and y subject to their budget constraint:

$$y \leq q(\mathbf{x}) - T(\mathbf{x}), \tag{5}$$

where the tax system, T, maps from the good space to the numeraire, $T : \mathbf{X} \to Y$. How much a consumer can spend on y depends on his choice of $\mathbf{x}$, the production function $q(\cdot)$ and the tax system $T(\cdot)$.

A tax system implements an allocation if each agent weakly prefers his bundle over all other combinations of goods available to him in the market. This concept is formally defined in definition 2:

Definition 2 *A tax system $T(\mathbf{x})$ implements an allocation $\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}$ in the market if each agent selects the same bundle on the market that is assigned to him in the allocation:*

$$
\begin{aligned}
\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\} &= \underset{\mathbf{x}, y}{\arg \max} \{u(\mathbf{x}, y, \mathbf{n}) : y = q(\mathbf{x}) - T(\mathbf{x}), \mathbf{x} \in \mathbf{X}\} \\
&\forall \quad \mathbf{n} \in \mathbf{N}
\end{aligned}
\tag{6}
$$

Note that we do not need to check whether the planner's budget constraint is satisfied as long as the tax system is successful in implementing the second-best allocation. By definition 1 the second-best allocation is feasible and the non-satiation of the utility function implies that on the market the agents' budget constraints must hold with equality. By Walras' law the planner's budget constraint must, therefore, also be satisfied.

The difficulty of implementability can be understood by comparing definitions 1 and 2. In the direct mechanism each agent maximizes his utility by sending the optimal p-dimensional message, containing all p characteristics of his type, to the planner. In doing so, he can choose his most preferred bundle in the set $\{\mathbf{X}, Y\}^*$. Agents can deviate by mimicking another type, but they cannot receive a bundle that has not been assigned to any type. However, in the market each agent can choose his optimal bundle out of all

points in the choice space $\{\mathbf{X}, Y\}$ within his budget set. The market allows the agents to create new bundles that were not assigned to any type in the direct mechanism. Such a strategy is called a joint deviation since, in order to create a new bundle that satisfies the budget constraint, an agent has to deviate in at least two goods relative to his bundle in the second-best allocation. If a tax system allows for profitable joint deviations, it cannot satisfy definition 2, even if the allocation satisfies definition 1. Because the market allows agents to create joint deviations, the conditions for implementability are generally stricter than the conditions for incentive compatibility. A graphical example based on the canonical Mirrleesian tax system will illustrate the issue in more detail in the next section. Before discussing the example it is convenient to discuss the most important approaches to implementation in the literature in our notation.

3.4 The principle of taxation and the Mirrleesian implementation

We are not the first to study tax systems that meet definition 2. The two most prominent approaches to implementation in the literature are the principle of taxation and the Mirrleesian tax system. The principle of taxation, derived in Hammond (1979), shows that for any incentive-compatible allocation at least one tax system exists that can implement it. This tax system has two properties. First, if the agent chooses the bundle $\mathbf{x}^*(\mathbf{n})$ designed for him, he will receive the corresponding value of $y^*(\mathbf{n})$. That is, the tax function satisfies:

$$T(\mathbf{x}^*(\mathbf{n})) = q(\mathbf{x}^*(\mathbf{n})) - y^*(\mathbf{n}) \qquad \forall \quad \mathbf{n} \in \mathbf{N}.$$

Second, it restricts agents to making a choice within $\{\mathbf{X}, Y\}^*$. The implementation creates individual budget sets that are restricted to lie within the image of the allocation. This restriction effectively prohibits all joint deviations. This immediately implies that the underlying problems of the agents in definitions 1 and 2 are isomorph, and hence, the outcome is identical. It follows that the conditions for incentive compatibility and market implementability coincide. Hence, the principle of taxation shows that by sufficiently restricting the choice space available to agents in the market, a tax system can always implement the second-best allocation.

Mirrlees (1976) derives the optimal second-best allocation with multiple goods under uni-dimensional heterogeneity in earnings ability of the agents. The outcome of the planner's maximization defines the wedges, $\mathcal{W}_i(n)$, for each level of earnings ability, n, and each good x_i. Under certain regularity conditions[10] the consumption level of each good can be used to infer n. That is, there exists an inverse function of the allocation $\mathbf{x}^*(n)$, such that $(\mathbf{x}^*)^{-1}(x_i) = n$ for each good x_i.

Mirrlees (1976) proposes to implement the second-best allocation through a tax system that has the following properties. First, like before, each agent should be able to afford his bundle:

$$T(\mathbf{x}^*(n)) = q(\mathbf{x}^*(n)) - y^*(n) \qquad \forall \quad n \in N,$$

However, unlike in the principle of taxation, Mirrlees does not limit the choice set to $\{\mathbf{X}, Y\}^*$. Instead, he proposes to set the marginal tax on each good x_i equal to the

---

[10]To be precise this requires non-satiation of the utility function, the Spence-Mirrlees condition on preferences and a monotonicity constraint on the allocation, as well as single-dimensional heterogeneity.

wedge:

$$T_i'(x_i) = \mathcal{W}_i(n) = \mathcal{W}_i((\mathbf{x}_i^*)^{-1}(x_i))$$

In the literature this approach is known as equating taxes to wedges (see e.g. Kocherlakota, 2005). The tax system is separable in the sense that the tax rate on good x_i only depends on the consumption of good x_i. Additionally, if all $x_i \in \mathbf{X}$ are awarded to a type in the direct mechanism, this defines the unique separable tax system because a marginal tax rate is assigned to each level of the choice variables. Furthermore, the tax system is entirely defined by the optimal wedges in the direct mechanism. Therefore, the optimal allocation in the direct mechanism contains all the information necessary to design the optimal tax system.

Note that the tax system described does not prohibit choices of the agent in any way, apart from setting a budget constraint. Unlike the tax system described by the principle of taxation, the Mirrleesian tax system does not change the agents' choice set compared to laissez faire, but only affects the relative prices through the marginal tax rates. As a result, the agents can choose joint deviations if they so desire. Therefore, unlike the tax system prescribed by the principle of taxation, there is no easy proof that shows that this tax system can implement all incentive-compatible allocations. Conditions under which this tax system implements the second best are derived in our corollary 2 further on.

Unfortunately Mirrlees (1976) only defined implementation under uni-dimensional heterogeneity. Generally, when agents are heterogeneous in multiple characteristics it is no longer possible to create a separable tax system, since the inverse allocation function $(\mathbf{x}^*)^{-1}(\cdot) = \mathbf{n}$ can no longer be written as a function of a single good (see also the discussion in Renes and Zoutman, 2014). Since the tax system is no longer separable, there may be multiple natural ways to extend the marginal tax rates to consumption bundles that are not awarded to any type in the direct mechanism. Moreover, there might be multiple tax systems that equate the marginal tax rate to the optimal wedge, since the inverse allocation function is generally not unique. Because of this potential multiplicity we do not focus on a particular tax system. Instead, we derive general conditions under which tax systems that affect the relative prices, but do not restrict the choice space can implement the second-best allocation, even if agents are heterogeneous in multiple dimensions. However, given the focus on uni-dimensional heterogeneity of agents in the literature, the most important example tax systems that focus on affecting relative prices without introducing restricting the choice set is the Mirrleesian tax system.

4 Failure of the Mirrleesian implementation: A Simple Example

This section will use a graphical example to show intuitively why the standard Mirrleesian implementation may fail to implement an incentive-compatible allocation. The purpose of the example, depicted in figures 1 and 2, is to illustrate the issue of implementation in the simplest possible setting. For that purpose the example is highly stylized and very much simplified. The idea is specifically not to give a realistic example, but to show that even in very simple settings, with agents that differ only in their earnings ability, the Mirrleesian tax system can sometimes fail.[11]

---

[11]In the appendix we describe one particular set of preferences for planner and the agents for which this allocation is second best.

The agents in the example are couples that maximize a joint utility function. For simplicity we assume that spouses within a couple have matched perfectly assortative, such that within each couple the spouses have exactly the same earnings ability. Moreover, we assume couples differ only in their earnings ability, such that this is a model with uni-dimensional heterogeneity of agents, $p = 1$. The spouses have to decide how much time either partner works and how much each of them tends to the household and children. The optimal allocation specifies how much labor income is generated by each spouse, $x_1^*(n)$ and $x_2^*(n)$, and how much each couple consumes, $y^*(n)$, as a function of the earnings ability of the couple. We assume the standard production function $x_i = nl_i$, where l_i represents the labor effort of spouse i. Clearly, high ability couples have to provide less labor to reach a certain income level than low ability couples. Moreover, we assume that couples' marginal utility of consumption is decreasing in consumption, such that the planner has an incentive to redistribute from high to low ability couples.

Since there is only one hidden characteristic for each couple, the bundles assigned to the types in a second-best allocation form a line in $X_1 \times X_2 \times Y$ space. This line is represented by the black line in figures 1 and 2. In figure 1 the line is sloping upward, indicating that the planner wants couples with higher ability to earn more in equilibrium. In addition, an increase in gross income, $x_1 + x_2$, leads to a less than one-to-one increase in consumption, y. This shows that the planner is redistributing from higher to lower ability couples. Finally, on any point of the line $x_1 = x_2$, indicating that the planner wants the spouses in each couple to supply the same amount of labor effort. In both pictures the dot represents the bundle assigned to one particular couple in the second-best allocation.

The hyperplane shows the budget constraint that results from the Mirrleesian implementation. The budget constraint is uniquely determined by equating taxes to wedges on the allocation, and assuming it is separable in each choice variable. Each point on the surface represents a bundle of labor income of the husband, x_1, labor income of the wife, x_2, and the corresponding amount of consumption after taxes, y, given by the budget constraint.

In figure 2 the vertical axis shows the utility level of the couple whose second-best bundle is located at the dot. The surface represents the reduced-form utility function, with the budget constraint substituted into the utility function, for all combinations $\{x_1, x_2, y\}$ given the tax system imposed. In figure 2 we can see that the assigned bundle (dot) marks the highest utility level on the allocation (line), such that the couple prefers their bundle over any of the other bundles in the allocation. The allocation is therefore incentive compatible for this couple, and indeed for all couples, in the direct mechanism. Mimicking another couple, and receiving the corresponding bundle, would decrease their utility level. Moreover, the depicted allocation is feasible.

However, in the market the couple is also allowed to trade working hours of the husband for working hours of the wife and vice versa. In figure 2 we can see that such joint deviations give the couple located at the dot more utility than their assigned bundle. In the market this couple prefers bundles that are not part of the planner's second-best allocation over the bundles in the allocation. Therefore, the Mirrleesian tax schedule to fail to implement the second-best allocation in this case.

In this particular example, the failure results from divergence in preferences between the planner and agents. Our fictitious couple prefers partial specialization, where one of the partners earns most of the income and the other partner stays at home. This could be rationalized in a model where spouses have increasing returns to scale or increasing utility to specialization. However, in the planner's optimal allocation spouses in couples
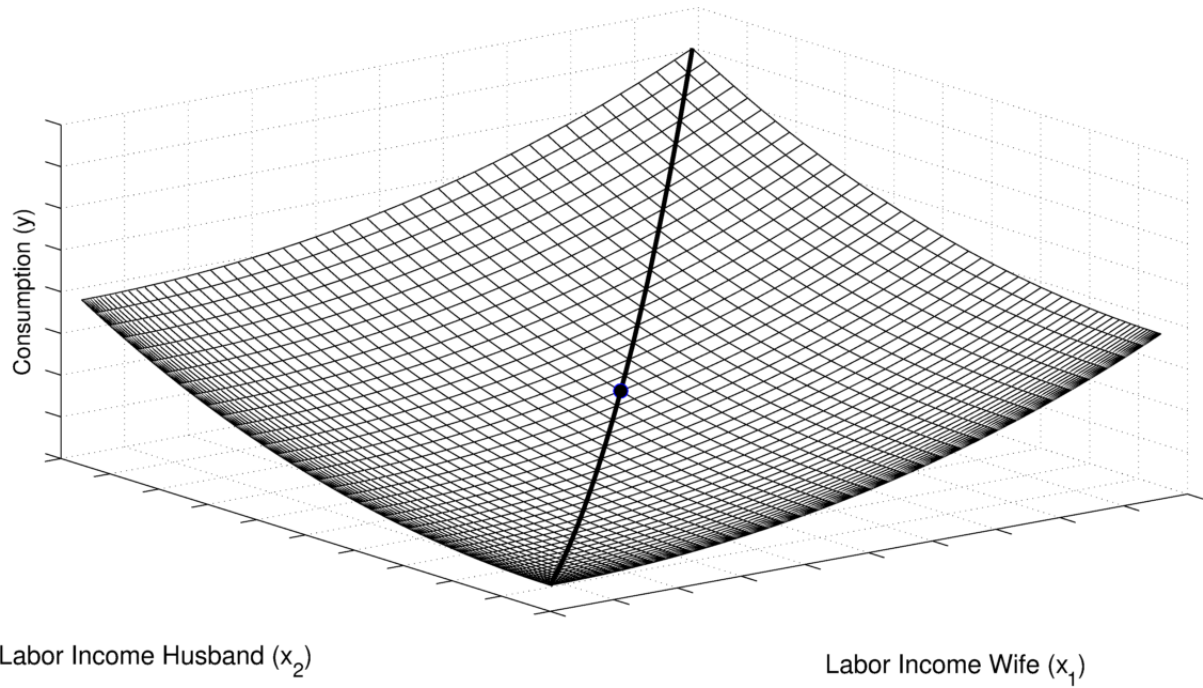
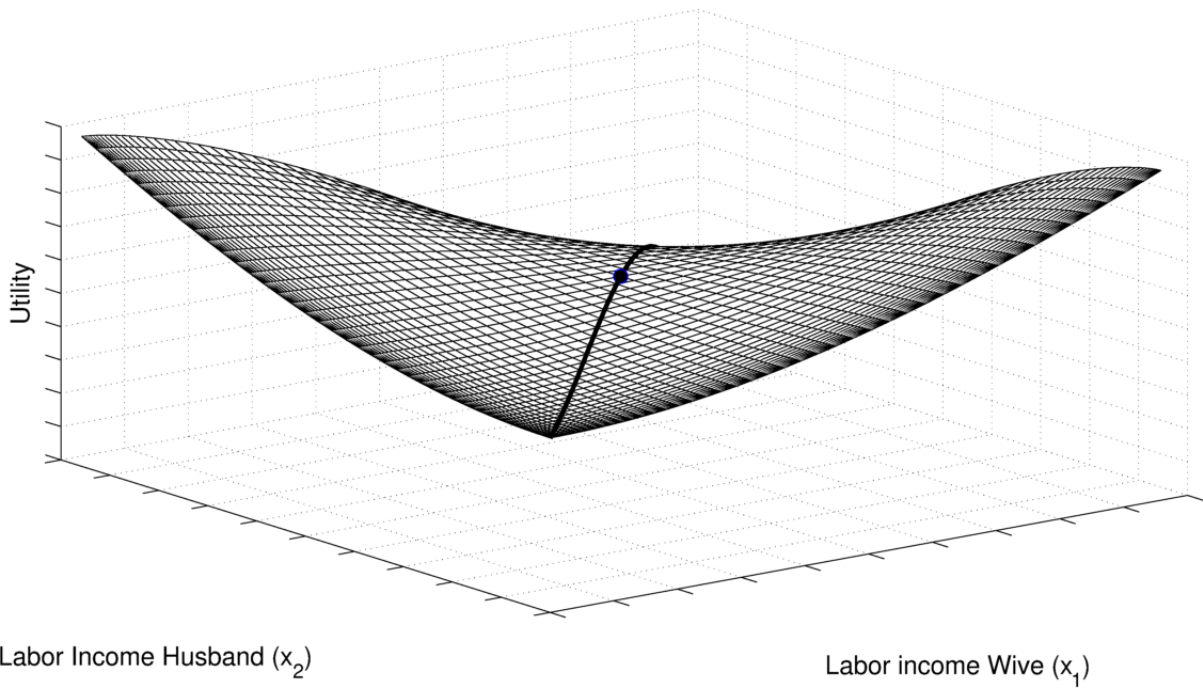Figure 1: An optimal allocation and a budget constraint.

Figure 2: Utility of a couple faced with the budget constraint in Figure 1.

need to work the same number of hours. Such misalignments could have several causes. First, the planner could be paternalistic, forcing households to equally divide the task between both partners, simply because he thinks this is fair. Alternatively, the couples in the economy might have children. The children in turn may be better off with attention of both spouses than with attention of only the stay-at-home spouse. If the utility of the child is not sufficiently weighted in the utility function of the couple, the planner may want to correct the externality and make both spouses stay at home part of the time. Note that both of these arguments can lead to exactly the same first-order optimality conditions for the second-best allocation. This example indicates that both non-welfarist motives and correction of externalities may lead to a failure of the Mirrleesian tax system, independently of the preferences of the agents. In section 5 we prove that absent this type of misalignment any tax system that equates taxes to wedges and does not impose any other restrictions on the choice set of agents, like the Mirrleesian tax system, suffices to implement the allocation.

To complete the picture, note that the solution offered by the principle of taxation is to disallow all consumption bundles that are not on the black line. That is, it forces both spouses to work an equal number of hours. In that case the observation that the assigned bundle (the dot) represents the highest utility level on the allocation (line) directly implies implementability.

However, in this example, as in many other cases, implementing the allocation through the principle of taxation may be socially undesirable. First, the planner should have the practical means to disallow joint deviations. The planner could, for instance, tax joint deviations at an infinite rate, or prohibit joint deviations explicitly, giving severe punishment to trespassers. Such measures, if they are feasible to begin with, are sensitive to small mistakes or a trembling hand. In our example, a slight deviation where one partner works more than the other for some unforeseen reason would be outright prohibited or subject to a severe punishment.

More generally, the planner could tax joint deviations by prohibitive, but non-infinite tax rates. However the calculations required to determine the second-best allocation do not provide us with information about when tax rates become prohibitive. Obviously, this crucially depends on the preferences of the agents. Hence, another set of complicated calculations is necessary to determine whether or not specific tax rates indeed lead agents to refrain from joint deviations in equilibrium.

Additionally, it may be costly and complex to administer prohibitions or prohibitive tax rates. In order to determine whether a specific bundle entails a joint deviation the entire vector of $\mathbf{x}$ has to be known and compared to the bundles in $\{\mathbf{X}, Y\}^*$ and this may be a very costly process, in terms of computational complexity, monitoring and administration.

This example clearly shows that a Mirrleesian tax system cannot implement all second-best allocations, and that the principle of taxation provides policy makers with tax systems that are unrealistic in market economies. Hence, in order to guide tax design for policy makers, we have to find conditions under which more realistic tax systems can implement a given allocation.

5   Implementation through Taxation

Now that we have shown that the canonical Mirrleesian tax system may sometimes fail to implement the second-best allocation. We will proceed by deriving the necessary and sufficient conditions for a tax system to implement the second-best allocation. These conditions are a useful test to verify whether a specific tax system implements a specific allocation. However, in order to perform this test, one first needs to derive both the allocation and the tax system. In many cases an explicit closed-form solution for the second-best allocation does not exist. Numerical solutions are available, but these describe only special cases by definition. Implementation can only be verified for the specific tax system and parametrization studied. Therefore, these explicit solutions cannot be used to say anything about tax systems in general.

   To overcome this issue we use to lemma derived in subsection 5.1 to derive two classes of problems for which a tax system that equates taxes to wedges always implements the second-best allocation in subsections 5.2 and 5.3. In the final subsection we discuss the consequence of this result for the existing literature on optimal taxation.

5.1   Conditions for Implementation

The necessary and sufficient conditions for implementation can be found by formally solving the problem of definition 2 in terms of the first - and second-order conditions of optimality. Although this approach is standard, the non-linear nature of prices in our setting requires some attention. The solution to this problem is summarized in lemma 1:

Lemma 1 *An incentive-compatible and feasible allocation can be implemented through a twice differentiable tax system $T(\mathbf{x})$ iff a.e.:*

   i.)
$$y^{*}(\mathbf{n}) = q(\mathbf{x}^{*}(\mathbf{n})) - T(\mathbf{x}^{*}(\mathbf{n})), \tag{7}$$

   ii.)
$$T_{i}'(\mathbf{x}^{*}(\mathbf{n})) = \mathcal{W}_{i}(\mathbf{n}), \tag{8}$$

   iii.)

$$-\frac{\partial \mathbf{s}(\mathbf{x}, y(u, \mathbf{x}, \mathbf{n}), \mathbf{n})}{\partial \mathbf{x}}|_{\{\mathbf{x}=\mathbf{x}^{*}(n), y=y^{*}(n)\}} + q''(\mathbf{x}^{*}(\mathbf{n})) - T''(\mathbf{x}^{*}(\mathbf{n})) \lessdot 0. \tag{9}$$

   where $\lessdot 0$ denotes that the matrix on the left-hand side should be negative semi-definite.

Proof. The proof can be found in the appendix.   ∎

   Equation (7) ensures that the amount of taxes paid for any bundle of $\mathbf{x}^{*}(\mathbf{n})$ within the allocation is uniquely determined. If the total tax level $T(\mathbf{x}^{*}(\mathbf{n}))$ is too high, the tax schedule cannot implement the allocation because people receive too little $y^{*}(\mathbf{n})$, and vice versa. Equation (8) is the first-order condition for a market implementation. It states that marginal taxes are equated to the wedges. There are always as many marginal tax rates in T' as there are goods in $\mathbf{X}$, for all $\mathbf{n} \in \mathbf{N}$. Such that there is always a unique vector of marginal tax rates $T'(\mathbf{x}^{*}(\mathbf{n}))$ that satisfies (8) within any possible incentive-compatible allocation.

   In effect, this means that the first order conditions of this problem can always be met and that the solution is unique on the allocation, but undefined for consumption bundles outside of $\mathbf{X}^{*}$. In our example in figures 1 and 2, this translates to a tax system that

17

is well defined on the line, but undefined everywhere else. Hence, although equation (8) tells us how to design the tax system on the allocation, it does not tell us how to extend it to bundles that are not assigned in the second-best allocation. Therefore, in order to design a complete tax system one has to make additional assumptions. In the case of the Mirrleesian tax system this additional assumption is that marginal tax rates are separable in each choice variable. One can also think of many other solutions, and the optimal tax system is generally not unique. However, any tax system that successfully implements the second-best allocation must also satisfy the second-order implementability constraints, equation (9).

Equation (9) states that the indifference curves of any linear combination of $\mathbf{x}$'s with respect to y should be more convex than the budget constraint for the same linear combination of $\mathbf{x}$'s. This condition is different from the standard second-order condition of utility maximization with two goods (see e.g. Mas-Collel et al., 1995) in two ways. First, in standard micro-economic theory the budget constraint is linear and hence the condition can be simplified to convexity of the indifference curves. Second, since there are multiple choices, sufficiency requires that the indifference curve of all linear combinations of $\mathbf{x}$ with respect to y are more convex than the budget constraint, with two goods only one such combination exists.

5.2 Pareto Efficiency and Welfarist Planner

As we have seen in figure 2 a Mirrleesian tax system may sometimes allow agents to create profitable joint deviations. In the next proposition we show that this issue does not occur if an allocation is i.) incentive compatible and feasible as in definition 1 and, ii.) within this set, on the Pareto frontier, such that no agent can be made better off without reducing the utility of any other agent. The most important implication of this proposition is presented in corollary 1, the second-best allocation of a welfarist planner can always be implemented by equating taxes to wedges.

Proposition 1 *If an allocation satisfies the constraints given by definition 1 and, within this set, resides on the Pareto frontier, then it can be implemented by any tax system, $T(\mathbf{x})$, that; i.) equates taxes to wedges on the allocation, $T_i'(\mathbf{x}^*(\mathbf{n})) = \mathcal{W}_i(\mathbf{n})$, and ii.) does not contain an internal maximum in any vector subspace of $\mathbf{X}^*$.*

Corollary 1 *The second-best allocation of a planner that maximizes a welfare function of the form $\int_{\mathbf{N}} W(u(\mathbf{n}), \mathbf{n}) d\mathbf{n}$ with $W_u(\cdot) > 0$ for all $u, \mathbf{n}$, commonly known as a welfarist planner, can be implemented under the conditions in proposition 1.*

Proof. The proof can be found in the appendix. ∎

Intuitively, consider again our example in figure 2. The increase in the reduced-form utility of the couple is approximately the same whether the couple turns "left" or "right" of the bundle assigned to them. That is, in the example it does not matter which spouse specializes in the official labor market, and which spouse specializes in house work. If tax payments from the couple to the planner are not maximized at their allocated bundle, then one of the two opposite deviation strategies must weakly increase tax revenue as well. Hence, one of the two deviations must be resource feasible. By assumption the utility of other couples is not directly affected by either deviation. Hence, there exists at least one joint deviation that has the following properties i.) it increases the utility of one agent, ii.) it occurs through voluntary transactions on the market and

is hence incentive compatible, iii.) it weakly increases the resources in the economy, and iv.) it leaves the utility of other agents unaffected. This then must be a Pareto improvement. However, an incentive-compatible Pareto improvement does not exist for allocations located on the Pareto frontier, and hence, we arrive at a contradiction. The proof to the proposition shows formally that this logic is not partial to our example. If taxes are equated to wedges, but second-order implementability constraints are not satisfied there always exists at least one incentive-compatible deviation that increases the utility of an agent and weakly increases the economy's resources. In our setting such a deviation is always a Pareto improvement.

Proposition 1 is important for public economists and policy makers. It shows that if the conditions of proposition 1 are fulfilled, the design of the optimal tax system is relatively straightforward. If a second-best allocation is Pareto efficient, the optimal tax system can be designed by equating the marginal tax rates to the optimal wedges. By proposition 1 this tax system must satisfy first and second-order implementability constraints, and can hence implement the planner's optimal allocation.

In addition, by corollary 1 such a tax system can also implement the second-best allocation of a welfarist planner. This follows from the fact that the second-best allocation of a welfarist planner always resides on the Pareto frontier, as was already shown in Werning (2007) and Brendon (2013). Intuitively, if the planner is welfarist, preferences of the agents and the planner are aligned in the sense that welfare strictly increases in the utility of all agents. Therefore, an allocation which is not on the Pareto frontier cannot be optimal for a welfarist planner. Hence, the proposition also applies to the second-best allocation of a welfarist planner.

The conditions of the proposition only applies if the tax system does not have an internal maximum in any vector subspace of $\mathbf{X}^*$, there are no externalities, and the allocation is on the Pareto frontier. In practice, the first restriction is almost always satisfied. The sign of the marginal tax rate is equal to the sign of the wedge. In most models of optimal taxation the optimal wedge does not change sign, such that the resulting tax system is monotonic, and hence, the tax system does not have an internal maximum. Even in models where the optimal wedge does change sign, such as Saez (2002b) and Choné and Laroque (2010), it changes sign from negative to positive. As such, the resulting tax system has an internal minimum, but not an internal maximum. We are not aware of any articles in the literature where this assumption is violated.

More importantly, the logic of our proof does not carry over to a model with externalities or to situations where the planner is non-welfarist. With externalities the deviation of any agent can influence the utility of other agents through the externality. Therefore, if a deviation increases the utility of a single agent, and weakly increases tax revenue, this no longer implies the deviation is a Pareto improvement. For instance, if a deviation increases the amount of gasoline purchased by an agent, this may lead to an increase in the utility of the agent as well as the tax revenue of the planner. However, the externality generated by the extra carbon emissions may harm the utility of all other agents in society. Finally, Kaplow and Shavell (2001) show that for a planner with a non-welfarist objective function there exist at least one Pareto improvement for the agents that decreases the objective function of the planner. Hence, even if a Pareto improvement exists, it is unclear whether this increases the objective function of the non-welfarist planner.

5.2.1 The Mirrleesian Tax System

As a corollary to proposition 1 we show when the Mirrleesian tax system can implement a second-best allocation.

Corollary 2 *Under uni-dimensional heterogeneity, $p = 1$, when an allocation i.) lies on the Pareto frontier, or ii.) is optimal to a welfarist planner, and the Mirrleesian tax system, as defined in section 3.4, does not contain an internal maximum in any vector subspace of $\mathbf{X}^*$, the Mirrleesian tax system can implement the second-best allocation.*

Proof. In the Mirrleesian tax system tax rates are equated to wedges such that all conditions in i.) proposition 1, respectively ii.) corollary 1 are satisfied. ■ Corollary 2 is a simple extension of proposition 1 and corollary 1. The Mirrleesian tax system, which is only defined under uni-dimensional heterogeneity, equates taxes to wedges. Hence, it can implement any incentive-compatible allocation on the Pareto frontier, as well as the second-best allocation of a welfarist planner.

This corollary is important primarily because the Mirrleesian tax system is separable. Hence, our result show that a separable tax system can implement the second-best allocation of a welfarist planner under uni-dimensional heterogeneity of agents. In static models this implies, for example, that the tax rate on capital income is independent of labor income and vice versa. In non-stochastic dynamic models in all periods, it implies that tax rates are independent over time, such that the government does not have to keep records over time to attain its optimal allocation.

5.3 Surjective allocations

There is an important second class of maximization problems in which equating taxes to wedges will always implement the second-best allocation. If the optimal allocation is surjective, such that all possible bundles of goods that are available in the market are assigned to at least one type in the second-best allocation, then only one tax system can satisfy first-order implementability conditions (7) and (8), and this tax system implements the second-best allocation. Proposition 2 formalizes this result.

Proposition 2 *If the mapping $\mathbf{x}^*(\mathbf{n})$ is surjective, then the tax system is fully described by equations (7) and (8), and this is the unique differentiable tax schedule that implements the second-best allocation.*

Proof. The proof can be found in the appendix. ■

Note that surjectiveness of the mapping $\mathbf{x}^*(\mathbf{n})$ is a relatively strict requirement. It requires that every possible bundle of $\mathbf{x} \in \mathbf{X}$ is assigned to at least one type. In this situation every choice in the market corresponds to the choice of a type in the direct mechanism. Since all types prefer their own bundle over the bundles assigned to other types and all bundles are assigned to a type, it follows that all types prefer their bundle above any other bundle in the economy.

Note that the tax system described in proposition 2 is unique. Wedges are defined for every possible value of $\mathbf{x} \in \mathbf{X}$, such that only one implementing tax system exists. Moreover, since surjectiveness implies that $\mathbf{X}^* = \mathbf{X}$ the principle of taxation does not restrict the choice set of agents in this particular case.

The allocation derived in Mirrlees (1971) is an example of a surjective allocation provided ability is continuously distributed in $\mathcal{R}_+$, and the second-best allocation assigns

all positive gross income levels to at least one type.[12] In this case, the function $x^*(n)$, mapping ability to gross income, is surjective. Then by definition incentive-compatibility constraints coincide with the constraints for implementability.

5.4 Implications for implementation in the existing literature

Our paper provides new insights about tax systems for a number of articles in the literature. In most models considered in the existing literature the planner is welfarist, there are no externalities, agents are heterogeneous in only one dimension and the optimal resulting tax system is monotonic or convex (see e.g. Mirrlees, 1971, Mirrlees, 1976, Atkinson and Stiglitz, 1976, Saez, 2001,Bovenberg and Jacobs, 2005, Golosov et al., 2013). Additionally, a number of articles study the implementation of allocations on the Pareto frontier, without specifically assuming a benevolent planner. (see e.g. Werning, 2007 and Brendon, 2013). By proposition 1 and its corollary 2 the second-best allocation in these papers can be implemented by a Mirrleesian tax system. In all aforementioned articles, except Golosov et al. (2013), this tax system is also suggested. Golosov et al. (2013) proposes to implement the allocation through a non-separable tax system that also equates taxes to wedges. Proposition 1 proves that this tax system implements the allocation derived in the article as well.

A growing literature considers optimal taxation under multi-dimensional heterogeneity of agents, where the planner is welfarist, there are no externalities, and the tax system that naturally results from equating taxes to wedges is monotonic or convex (see e.g. Mirrlees, 1976, Choné and Laroque, 2010, Renes and Zoutman, 2014, Rothschild and Scheuer, 2014 and Jacquet and Lehmann, 2014). Proposition 1 also applies to these models, such that any tax system that equates taxes to wedges implements the allocation derived in these articles.[13]

A number of articles study optimal non-linear taxation in the presence of a non-welfarist planner (see e.g. Fleurbaey and Maniquet, 2006, Fleurbaey and Maniquet, 2007, Jacquet and Van de Gaer, 2011 and Gerritsen, 2013) under uni-dimensional heterogeneity in earning ability. In each of these articles the agents only choose how much income to earn and how much to consume. Therefore, provided the distribution of earnings ability is continuous and unbounded, the allocation is likely to be surjective. The surjectiveness of the allocation can easily be verified by checking whether all non-negative income levels are assigned, as is usually the case. In that case proposition 2 shows that there is a unique tax system that implements the second-best allocation by equating the tax to the wedge.

6 Concluding Remarks

The results presented in this paper cover the relatively understudied second step of tax design for a general class of models. Propositions 1 and 2 show that a relatively simple class of tax systems, that is completely described by the optimal wedges in the second-best allocation, can implement the optimal allocation found in most problems studied in the existing literature. In these models implementation can proceed through any tax system that equates taxes to wedges. In addition, lemma 1 provides the conditions that need to

---

[12]We implicitly assume agents can also choose their gross income level in $\mathcal{R}_+$.

[13]Note that Rothschild and Scheuer (2014) also extend their analysis by introducing an externality. For this extension proposition 1 does not apply.

be checked if the problem does not fit one of the two classes outlined in our propositions. Importantly, our results are not restricted to the often-studied case of uni-dimensional heterogeneity of agents, but are also directly applicable to the multi-dimensional problem studied in the companion paper Renes and Zoutman (2014).

Proposition 1 highlights a unique feature of the Mirrleesian optimal tax model. Unlike the design problem of auctioneers and monopolists, the maximization problem of a welfarist central planner is aligned with that of the agents he faces. In fields such as monopoly pricing and auction theory the objectives of the principal and the agents are usually opposed, in the sense that the payoff of the planner increases in the payment made by the agents. An increase in a monopolist's profits (at fixed quantities) automatically comes at the expense of the consumers. As such, implementation will generally be more difficult than in the Mirrleesian taxation literature. As a result, our paper shows that the planner in these settings generally needs to restrict the choice space of its agents by for example bundling goods or disallowing certain bids (see also Armstrong, 1996, Renes, 2011, Rochet and Choné, 1998).

The alignment between agents and planner allows a relatively broad class of tax systems to implement the second best of a welfarist planner. In many cases the planner can let the agents maximize their utility subject to a budget constraint, without further restriction in the choice space, irrespective of the actual utility function of individuals.

This result is important because restricting the choice space of agents in a real-world economy is generally very complex, and in most cases unrealistic and undesirable in a market economy. However, our results may also help explain cases where the price mechanism fails and the government has to restrict the choice space of agents. This may for instance be the case in settings where externalities play an important role.

Future work could focus on the possibility to extend this work to dynamically stochastic settings. Additionally, finding implementing tax systems that allow individual freedom of choice in cases when equating taxes to wedges does not lead to the optimal allocation is of first order importance to the normative public finance literature that deals with issues like externalities and non-welfarist motives.

References

Adams, William J., and Janet L. Yellen (1976) 'Commodity bundling and the burden of monopoly.' *Quarterly Journal of Economics* 90(3), 475–498

Albanesi, Stefania, and Christopher Sleet (2006) 'Dynamic optimal taxation with private information.' *Review of Economic Studies* 73(1), 1–30

Armstrong, Mark (1996) 'Multiproduct nonlinear pricing.' *Econometrica* 64(1), 51–75

_ (2000) 'Optimal multi-object auctions.' *Review of Economic Studies* 67(3), 455–81

Armstrong, Mark, and Jean-Charles Rochet (1999) 'Multi-dimensional screening: A user's guide.' *European Economic Review* 43(4), 959–979

Athey, Susan, and Jonathan Levin (2001) 'Information and competition in US forest service timber auctions.' *Journal of Political Economy* 109(2), 375–417

Atkinson, Anthony B., and Joseph E. Stiglitz (1976) 'The design of tax structure: Direct versus indirect taxation.' *Journal of Public Economics* 6(1-2), 55–75

Bierbrauer, Felix (2009) 'A note on optimal income taxation, public goods provision and robust mechanism design.' *Journal of Public Economics* 93(5), 667–670

Boadway, Robin, and Bas Jacobs (forthcoming) 'Optimal linear commodity taxation under optimal non-linear income taxation.' *Journal of Public Economics*

Bovenberg, A. Lans, and Bas Jacobs (2005) 'Redistribution and education subsidies are siamese twins.' *Journal of Public Economics* 89(11-12), 2005–2035

Brendon, Charles (2013) 'Efficiency, equity, and optimal income taxation.' Max Weber Program Working Paper MWP 2013/22 Florence

Chakraborty, Indranil (1999) 'Bundling decisions for selling multiple objects.' *Economic Theory* 13(3), 723–733

Choné, Philippe, and Guy Laroque (2010) 'Negative marginal tax rates and heterogeneity.' *American Economic Review* 100(5), 2532–2547

Christiansen, Vidar (1984) 'Which commodity taxes should supplement the income tax?' *Journal of Public Economics* 24(2), 195–220

Clausen, Andrew, and Carlo Strub (2013) 'A general and intuitive evelope theorem.' mimeo Zurich

Cremer, Helmuth, Pierre Pestieau, and Jean-Charles Rochet (2001) 'Direct versus indirect taxation: The design of the tax structure revisited.' *International Economic Review* 42(3), 781–800

Dasgupta, Partha S., Peter J. Hammond, and Eric S. Maskin (1979) 'The implementation of social choice rules: Some general results on incentive compatibility.' *Review of Economic Studies* 46(2), 185–216

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

Ewerhart, Christian, and Karsten Fieseler (2003) 'Procurement auctions and unit-price contracts.' *RAND Journal of Economics* 34(3), 569–581

Fleurbaey, Marc, and Francois Maniquet (2006) 'Fair income tax.' *Review of Economic Studies* 73(1), 55–83

Fleurbaey, Marc, and François Maniquet (2007) 'Help the low skilled or let the hardworking thrive? A study of fairness in optimal income taxation.' *Journal of Public Economic Theory* 9(3), 467–500

Gerritsen, Aart (2013) 'Optimal taxation when people do not maximize well-being.' mimeo Erasmus University Rotterdam

Golosov, Mikhail, Aleh Tsyvinski, and Iván Werning (2007) 'New dynamic public finance: A user's guide.' In *NBER Macroeconomics Annual 2006, Volume 21,* ed. Daron Acemoglu, Kenneth Rogoff, and Micheal Woodford (Cambridge-MA: MIT Press) pp. 317–388

Golosov, Mikhail, Aleh Tsyvinski, and Matthew Weinzierl (2013) 'Preference heterogeneity and optimal commodity taxation.' *Journal of Public Economics* 97(1), 160–175

Guesnerie, Roger (1995) *A contribution to the pure theory of taxation,* vol. 25 (Cambridge-MA: Cambridge University Press)

Guesnerie, Roger, and Kevin Roberts (1984) 'Effective policy tools and quantity controls.' *Econometrica* 52(1), 59–86

Hammond, Peter J. (1979) 'Straightforward individual incentive compatibility in large economies.' *Review of Economic Studies* 46(2), 263–282

_ (1987) 'Markets as constraints: Multilateral incentive compatibility in continuum economies.' *Review of Economic Studies* 54(3), 399–412

Hurwicz, Leonid (1960) 'Optimality and informational efficiency in resource allocation processes.' In *Mathematical Methods in the Social Sciences,* ed. Kenneth Arrow, S. Karlin, and P. Suppes (Redwood City, CA: Stanford University Press)

_ (1973) 'The design of mechanisms for resource allocation.' *American Economic Review* 63(2), 1–30

Jacobs, Bas, and Ruud A. De Mooij (2011) 'Pigou meets Mirrlees: On the irrelevance of tax distortions for the second-best pigouvian tax.' CESifo Working Paper No. 3250 Munich

Jacquet, Laurence, and Dirk Van de Gaer (2011) 'A comparison of optimal tax policies when compensation or responsibility matter.' *Journal of Public Economics* 95(11), 1248–1262

Jacquet, Laurence, and Etienne Lehmann (2014) 'Optimal nonlinear income taxation with multidimensional types: The case with heterogeneous behavioral responses.' THEMA (THéorie Economique, Modélisation et Applications), Université de Cergy-Pontoise Paris

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden (2013) 'Optimal redistributive taxation with both extensive and intensive responses.' *Journal of Economic Theory* 148(5), 1770–1805

Jehiel, Philippe, Moritz Meyer-Ter-Vehn, and Benny Moldovanu (2007) 'Mixed bundling auctions.' *Journal of Economic Theory* 134(1), 494–512

Kaplow, Louis, and Steven Shavell (2001) 'Any non-welfarist method of policy assessment violates the Pareto principle.' *Journal of Political Economy* 109(2), 281–286

Kleven, Henrik J., Claus T. Kreiner, and Emmanuel Saez (2009) 'The optimal income taxation of couples.' *Econometrica* 77(2), 537–560

Kocherlakota, Narayana R. (2005) 'Zero expected wealth taxes: A Mirrlees approach to dynamic optimal taxation.' *Econometrica* 73(5), 1587–1621

_ (2010) *The new dynamic public finance* (Princeton-NJ: Princeton University Press)

Mas-Collel, Andreau, Micheal D. Whinston, and Jerry R. Green (1995) *Microeconomics* (Oxford: Oxford University Press)

Maskin, Eric (1999) 'Nash equilibrium and welfare optimality.' *Review of Economic Studies* 66(1), 23–38

McAfee, R. Preston, and John McMillan (1988) 'Multidimensional incentive compatibility and mechanism design.' *Journal of Economic Theory* 46(2), 335–354

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

— (1976) 'Optimal tax theory: A synthesis.' *Journal of Public Economics* 6(4), 327–358

Myerson, Roger B. (1979) 'Incentive compatibility and the bargaining problem.' *Econometrica* 47(1), 61–73

Palfrey, Thomas R. (1983) 'Bundling decisions by a multiproduct monopolist with incomplete information.' *Econometrica* 51(2), 463–83

Renes, Sander (2011) 'Balancing the bids, solutions for unit price auctions.' Tinbergen Institute Discussion Paper 11-047/1 Rotterdam

Renes, Sander, and Floris T. Zoutman (2014) 'As easy as ABC? Multi-dimensional screening in public finance.' Rotterdam. mimeo

Roberts, Kevin (2000) 'A reconsideration of the optimal income tax.' In 'Incentives and Organization, Papers in Honour of Sir James Mirrlees' (Oxford: Department of Economics (University of Oxford))

Rochet, Jean-Charles (1985) 'The taxation principle and multi-time Hamilton-Jacobi equations.' *Journal of Mathematical Economics* 14(2), 113–128

Rochet, Jean-Charles, and Philippe Choné (1998) 'Ironing, sweeping, and multidimensional screening.' *Econometrica* 66(4), 783–826

Rothschild, Casey, and Florian Scheuer (2014) 'A theory of income taxation under multidimensional skill heterogeneity.' NBER Working Paper No. 19822 Cambridge, MA

Saez, Emmanuel (2001) 'Using elasticities to derive optimal income tax rates.' *Review of Economic Studies* 68(1), 205–229

— (2002a) 'The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes.' *Journal of Public Economics* 83(2), 217–230

— (2002b) 'Optimal income transfer programs: Intensive versus extensive labor supply responses.' *Quarterly Journal of Economics* 117(3), 1039–1073

Stiglitz, Joseph E. (1987) 'Pareto efficient and optimal taxation and the new new welfare economics.' *Handbook of Public Economics* 2, 991–1042

Vickrey, William (1961) 'Counterspeculation, auctions, and competitive sealed tenders.' *Journal of Finance* 16(1), 8–37

Werning, Ivan (2007) 'Pareto efficient income taxation.' Mimeo MIT Cambridge, MA

A   Appendix

For bookkeeping, the Jacobian of first derivatives $\phi'(\cdot)$ of any function $\phi(\cdot) : \mathcal{R}^a \to \mathcal{R}^b$, is of dimension $b \times a$, while the Hessian of second derivatives $\phi''(\cdot)$ is of dimension $ab \times a$. For any multi-vector functions $\psi(\mathbf{z}^1, \mathbf{z}^2, \ldots) : \mathcal{R}^{a^1} \times \mathcal{R}^{a^2} \ldots \to \mathcal{R}$ the vector of first derivatives ψ_{z^i} is of dimension $a^i \times 1$ and the matrix of second derivatives $\psi_{z^i z^j}$ is of dimension $a^i \times a^j$ where the dimension of the matrix follows the order of the subscripts. In addition, let superscript T be the transpose operator. Vectors and multi-dimensional constructs are denoted in bold, scalars are in normal font.

A.1   Proof of Lemma 1

Proof. Due to non-satiation of the utility function we know that the budget constraint will hold with equality such that we know that:

$$y^*(\mathbf{n}) = q(\mathbf{x}^*(\mathbf{n})) - T(\mathbf{x}^*(\mathbf{n}))$$

Direct substitution of the budget constraint into the utility function allows us to write the first-order conditions to problem (2) as:

$$\mathbf{0} = u_{\mathbf{x}} + (q' - T')^T u_y \tag{10}$$

which directly implies equations (7) and (8).

Now take the second-order derivative of the utility function with respect to $\mathbf{x}$ to get the second-order conditions:

$$u_{\mathbf{xx}} + \left(2u_{\mathbf{x}y} + u_{yy}(q'(\mathbf{x}^*) - T'(\mathbf{x}^*))^T\right)(q'(\mathbf{x}^*) - T'(\mathbf{x}^*)) + u_y(q''(\mathbf{x}^*) - T''(\mathbf{x}^*)) < 0 \tag{11}$$

Differentiate the marginal rate of substitution, $\mathbf{s}$, toward $\mathbf{x}$, using the definition of $\mathbf{s}$ and the implicit function theorem to define $y(u, \mathbf{x}, \mathbf{n})$, to get:

$$\frac{\partial \mathbf{s}(\mathbf{x}, y(u, \mathbf{x}, \mathbf{n}), \mathbf{n})}{\partial \mathbf{x}} = -\frac{\left(u_{\mathbf{xx}} + 2u_{\mathbf{x}y}\mathbf{s}^T\right) - u_{yy}\mathbf{s}\mathbf{s}^T}{u_y} \tag{12}$$

Now combining (8) with (12) allows us to simplify (11) and obtain the final condition:

$$-\left(\frac{\partial \mathbf{s}(\mathbf{x}, y(u, \mathbf{x}, \mathbf{n}), \mathbf{n})}{\partial \mathbf{x}} + q''(\mathbf{x}^*) - T''(\mathbf{x}^*)\right)u_y \; < \; 0 \Leftrightarrow$$

$$-\frac{\partial \mathbf{s}(\mathbf{x}, y(u, \mathbf{x}, \mathbf{n}), \mathbf{n})}{\partial \mathbf{x}}|_{\{\mathbf{x}=\mathbf{x}^*(n), y=y^*(n)\}} + q''(\mathbf{x}^*) - T''(\mathbf{x}^*) \; < \; 0$$

where the final step follows from the assumption that $u_y > 0$. $\blacksquare$

A.2   Proof to Proposition 1 and corollary 1

Proof. Suppose taxes are equated to wedge for an, presumed optimal, incentive-compatible allocation. Consider a joint deviation of an agent from the bundle assigned to him when the original allocation is incentive compatible, resource feasible, and within this set on the Pareto frontier. Denote the joint deviation by $\alpha\Delta\mathbf{x}$ where $\alpha > 0$ and $\Delta\mathbf{x}$ is a $k \times 1$

vector which we normalize to length one, on the market by a particular agent of type $\mathbf{m}$. The budget constraint (5) defines the amount of y the deviating agent receives as:

$$y = q\left(\mathbf{x}^*\left(\mathbf{m}\right) + \alpha\Delta\mathbf{x}\right) - T\left(\mathbf{x}^*\left(\mathbf{m}\right) + \alpha\Delta\mathbf{x}\right).$$

Note that the resulting allocation after the joint deviation is incentive compatible since the original allocation was incentive compatible and the exchange occurs at market prices. Through the utility function, the utility gain, $\mathcal{U}\left(\alpha\Delta\mathbf{x}\right)$, of such a deviation can therefore be written as:

$$\mathcal{U}\left(\alpha\Delta\mathbf{x}\right) = u\left(\mathbf{x}^*\left(\mathbf{m}\right) + \alpha\Delta\mathbf{x}, q\left(\mathbf{x}^*\left(\mathbf{m}\right) + \alpha\Delta\mathbf{x}\right) - T\left(\mathbf{x}^*\left(\mathbf{m}\right) + \alpha\Delta\mathbf{x}\right), \mathbf{n}\right) - u^*\left(\mathbf{m}\right),$$

where u^* denotes the utility the agent obtains without deviation. By a second-order Taylor approximation around $u^*\left(\mathbf{m}\right)$ the utility gain can be written as:

$$
\begin{aligned}
\mathcal{U}\left(\alpha\Delta\mathbf{x}\right) &\approx \left(u_{\mathbf{x}}^T + u_y\left(q' - T'\right)\right)\alpha\Delta\mathbf{x} + \\
&\qquad \frac{1}{2}\alpha^2\Delta\mathbf{x}^T\left(u_{\mathbf{xx}} + \left(2u_{\mathbf{x}y} + u_{yy}\left(q' - T'\right)^T\right)\left(q' - T'\right) + u_y\left(q'' - T''\right)\right)\Delta\mathbf{x}, \\
&\approx \frac{1}{2}\alpha^2\Delta\mathbf{x}^T\left(u_{\mathbf{xx}} + \left(2u_{\mathbf{x}y} + u_{yy}\left(q' - T'\right)^T\right)\left(q' - T'\right) + u_y\left(q'' - T''\right)\right)\Delta\mathbf{x}, \\
&\approx \frac{1}{2}\alpha^2 u_y\Delta\mathbf{x}^T\left(-\frac{\partial\mathbf{s}\left(\mathbf{x}, y\left(u, \mathbf{x}, \mathbf{m}\right), \mathbf{m}\right)}{\partial\mathbf{x}}\Big|_{\{\mathbf{x}=\mathbf{x}^*, y=y^*, u=u^*\}} + q''\left(\mathbf{x}^*\right) - T''\left(\mathbf{x}^*\right)\right)\Delta\mathbf{x},
\end{aligned}
$$

where the first order terms equals zero by the assumption that taxes are equated to wedges, $T_i'\left(\mathbf{x}^*\left(\mathbf{m}\right)\right) = \mathcal{W}_i(\mathbf{m})$, and we have used equation (12) for simplification.

Now suppose second-order implementation constraints (9) are not satisfied such that:

$$\mathcal{H} = -\frac{\partial\mathbf{s}\left(\mathbf{x}, y\left(u, \mathbf{x}, \mathbf{m}\right), \mathbf{m}\right)}{\partial\mathbf{x}}\Big|_{\{\mathbf{x}=\mathbf{x}^*, y=y^*, u=u^*\}} + q''\left(\mathbf{x}^*\right) - T''\left(\mathbf{x}^*\right),$$

is not negative semi-definite. In that case there exists at least one deviation strategy $\alpha\Delta\hat{\mathbf{x}}$ which yields a positive utility gain: $\mathcal{U}\left(\alpha\Delta\hat{\mathbf{x}}\right) > 0$. To see this note that if this were not the case, by definition $\mathcal{H}$ would be negative semi-definite. In addition, since $\mathcal{H}$ is a matrix of second-order derivatives and therefore symmetric, the opposite deviation $-\alpha\Delta\hat{\mathbf{x}}$ yields approximately the same utility gain. A second-order Taylor approximation then yields: $\mathcal{U}\left(\alpha\Delta\hat{\mathbf{x}}\right) \approx \mathcal{U}\left(-\alpha\Delta\hat{\mathbf{x}}\right)$. Since, the approximation error of the Taylor expansion limits to zero as $\alpha \to 0$, there must exist at least two deviation strategies $\alpha\Delta\hat{\mathbf{x}}$, and $-\alpha\Delta\hat{\mathbf{x}}$ that yield positive utility for sufficiently small α.

The change in tax revenue due to such a deviation can also be approximated by means of a second-order Taylor expansion:

$$
\begin{aligned}
\mathcal{T}\left(\alpha\Delta\hat{\mathbf{x}}\right) &= T\left(\mathbf{x}^*\left(\mathbf{n}\right) + \alpha\Delta\hat{\mathbf{x}}\right) - T\left(\mathbf{x}^*\left(\mathbf{n}\right)\right) \\
&\approx \alpha T'\Delta\hat{\mathbf{x}} + \frac{1}{2}\alpha^2\Delta\hat{\mathbf{x}}^T T''\Delta\hat{\mathbf{x}}.
\end{aligned}
$$

The first-order term will always be non-negative for deviation strategy $\alpha\Delta\hat{\mathbf{x}}$ or for $-\alpha\Delta\hat{\mathbf{x}}$. If for one of the choices, the first-order term is positive, the first-order term dominates the second-order term for sufficiently small α. Hence, there is a deviation that results in higher tax revenue. If on the other hand the first-order term is zero, we need to consider the second-order term. If it is negative the tax schedule contains an internal maximum in at least one vector subspace of $\mathbf{X}$ which violates the assumption stated in the proposition.

Therefore, if the first-order term is zero the second term must be non-negative. Hence, tax revenue always weakly increases in either $-\alpha\Delta\hat{\mathbf{x}}$ or $\alpha\Delta\hat{\mathbf{x}}$ for sufficiently small α. It follows that at least one of the two deviations is resource feasible.

Therefore, under the conditions stated in our proposition, if the tax system is not successful in implementing the allocation on the Pareto frontier, there exist at least one deviation which increase utility of at least one agent, leaves the allocation resource feasible and incentive compatible. The utility of all other agents remains unaffected, since there are no externalities by assumption, and is therefore a Pareto improvement. Hence, we run into a contradiction, since a Pareto improvement cannot exist over an allocation that is already on the Pareto frontier proving the main part of the proposition.

The stated corollary follows from the fact that the second-best allocation of a welfarist planner always lies on the Pareto frontier. To see this note that if the allocation were not on the Pareto frontier, by definition the planner could increase the utility of at least one agent without decreasing the utility of any other agent. Such a deviation increases the value of the planner's objective function, and hence the original allocation could not have been the second best (this is also shown in Werning, 2007,Brendon, 2013). ∎

A.3 Proof to Proposition 2

Proof. Note that in the direct mechanism, reporting your type to the planner is equivalent to choosing a bundle $\{\mathbf{x}, y\}$ among the set of bundles that are assigned to a type, $\{\mathbf{X}, Y\}^*$. Hence, we can rewrite the optimization problem for the agents in the direct mechanism (see definition 1) as follows:

$$\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\} = \arg\max_{\mathbf{x}, y} \{u(\mathbf{x}, y, \mathbf{n}) : \{\mathbf{x}, y\} \in \{\mathbf{X}, Y\}^*\}.$$

In addition, by equation (7), for each bundle of $\mathbf{x}^*(\mathbf{n})$ the tax system provides the corresponding level $y^*(\mathbf{n})$. Hence, if the tax system, $T(\mathbf{x})$, satisfies the equation (7) we can further simplify the optimization problem to:

$$\begin{aligned} \{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\} &= \arg\max_{\mathbf{x}, y} \{u(\mathbf{x}, y, \mathbf{n}) : y = q(\mathbf{x}) - T(\mathbf{x}), \mathbf{x} \in \mathbf{X}^*\}, \\ &= \arg\max_{\mathbf{x}, y} \{u(\mathbf{x}, y, \mathbf{n}) : y = q(\mathbf{x}) - T(\mathbf{x}), \mathbf{x} \in \mathbf{X}\}, \end{aligned}$$

where the final step follows from the fact that the allocation is surjective such that the set of assigned bundles is equal to the set of all available bundles on the market: $\mathbf{X}^* = \mathbf{X}$. This is exactly equal to optimization problem 2, and hence, the tax system that satisfies (7) can implement any surjective allocation. Note that we do not need to check whether the tax system also satisfies equations (8) and (9). Since we have assumed that the original allocation is incentive compatible, it follows immediately that these constraints are also satisfied provided the allocation is surjective.

Finally, note that the tax system that implements a surjective allocation is unique. Equation (7) implicitly defines the value of the tax function for all $\mathbf{x} \in \mathbf{X}^*$:

$$T(\mathbf{x}^*(\mathbf{n})) = q(\mathbf{x}^*(\mathbf{n})) - y^*(\mathbf{n}).$$

Therefore, equation (7) uniquely determines the value of $T(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{X}^*$. Since by surjectiveness $\mathbf{X}^* = \mathbf{X}$ it defines the value of $T(\mathbf{x})$ over its entire domain. Hence, there is only one function $T(\mathbf{x})$ that satisfies (7). Since (7) is a necessary condition for implementation, it follows that there is only one tax system that can implement a surjective allocation. ∎

A.4 Example

A.4.1 The Second-Best Allocation

Our graphical example in section 4 is based on the following model. Suppose the economy is inhabited by couples and their children. Each couple j derives utility from their joint consumption, which is assumed to be the untaxed numeraire commodity Y^j. In addition, the couples derive disutility from hours worked by each spouse i in couple j, l_i^j, as well as disutility from the interaction between hours worked, $l_1^j l_2^j$. This gives couples an incentive for partial specialization where one partner works more hours, while the other spends more time at home. For simplicity we assume preferences for labor are symmetric, such that the couple does not have a preference for which of the two spouses supplies the highest amount of labor. Each partner i in couple j is assumed to have identical ability, n^j. This could be the result of perfect assortative matching. As is standard in the literature, we assume gross income is the product of hours worked and ability: $X_i^j = n^j l_i^j$. We assume the following functional form for the utility function of the parents:

$$
\begin{aligned}
U_p\left(Y, l_1, l_2, l_1 l_2\right) &= \gamma_Y \log\left(Y\right) - \frac{\gamma_l}{\alpha} \sum_{i=1}^{2} l_i^\alpha - \frac{1}{\beta}\left(l_1 l_2\right)^\beta, \\
&= \gamma_Y \log\left(Y\right) - \frac{\gamma_X}{\alpha} \sum_{i=1}^{2} \left(\frac{X_i}{n}\right)^\alpha - \frac{1}{\beta}\left(\frac{X_1 X_2}{n^2}\right)^\beta,
\end{aligned}
$$

where utility is assumed to be logarithmic in consumption of Y, CES with parameter $\alpha > 1$ in each spouse's hours worked, and CES with parameter β in the interaction of hours worked. γ_Y and γ_X are scaling parameters which allows one to weight the relative importance of each of the three terms in the utility function.

Each couple is assumed to have one child. Their utility function does not weight the utility of the child.[14] The child does not make any decisions but its utility is affected by the decisions of his parents. This creates a classic externality where the preferences of the child are not included in the transactions made by the parents, but do affect total welfare. The child is assumed to receive utility of consumption of the household. In addition, it receives disutility from ±labor, since leisure time of each parent is spend with the child. Finally, the child enjoys spending time with both of his parents, and therefore receives utility if the parents do not specialize but share the task of rearing him. We capture this by letting the child's utility increase in $l_1 l_2$. We assume the following utility function for the child's utility function:

$$
U_c\left(Y, l_1, l_2, l_1 l_2\right) = \delta_Y \log\left(Y\right) - \frac{\delta_l}{\xi} \sum_{i=1}^{2} l_i^\xi + \frac{1}{\zeta}\left(l_1 l_2\right)^\zeta,
$$

where $\xi > 1$ is the parameter for CES disutility of labor, ζ the parameter for utility from the interaction of hours worked, and δ_Y and δ_l scaling parameters.

In the simulations we take a discrete approximation of our model, such that n follows a discrete distribution with J possible outcomes $n_{\min}, \ldots, n_{\max}$.[15] We assume each group has similar size f such that the mass of a particular type is equal to $g = \frac{1}{J}$. We assume

---

[14]Note that the main intuition carries over to the case where the parents take account of some, but not all of the utility of the child.

[15]Note that our model formally does not apply to discrete models of taxation. However, as we increase the number of types, we can approximate the case fior continuous type with increasing precision.

the welfare function of the planner is a concave increasing function of joint family utility $U = U_P + U_C$. Hence, social welfare can be written as:

$$SW = \frac{1}{f} \sum_{j=1}^{J} \frac{1}{1-\nu} U_j^{1-\nu}, \tag{13}$$

where ν measures the rate of relative inequality aversion.

We normalize all prizes to one such that the resource constraint of the economy can be written as:

$$0 \geq \frac{1}{f} \sum_{j=1}^{J} Y^j - X_1 - X_2,$$

$$0 \geq \sum_{j=1}^{J} Y^j - X_1^j - X_2^j. \tag{14}$$

In the direct mechanism, the planner is assumed to offer bundles $\left(X_1^j, X_2^j, Y^j \right)$ to each household j on the basis of a their reported type. The allocation is incentive compatible if each couple truthfully reports their type to the planner:

$$j = \arg \max_{k=1,\ldots,J} U_p \left(Y^k, \frac{X_1^k}{n^j}, \frac{X_2^k}{n^j}, \frac{X_1^k X_2^k}{\left(n^j \right)^2} \right) \quad \forall \quad j = 1, \ldots, J. \tag{15}$$

Note that incentive compatibility only depends on the utility of the parents, and not of the children since they do not take any decision. In the simulations we take the first-order approach and calculate the optimum under the condition that each couple j has at least as high utility by telling the truth as by mimicking type $j - 1$:

$$U_p \left(Y^j, \frac{X_1^j}{n^j}, \frac{X_2^j}{n^j}, \frac{X_1^j X_2^j}{\left(n^j \right)^2} \right) \geq U_p \left(Y^{j-1}, \frac{X_1^{j-1}}{n^j}, \frac{X_2^{j-1}}{n^j}, \frac{X_1^{j-1} X_2^{j-1}}{\left(n^j \right)^2} \right) \quad \forall \quad j = 2, \ldots, J. \tag{16}$$

At the end of the simulation we check whether the allocation calculated under the first-order approach also satisfies (15) and in the reported simulation this is indeed the case. The optimal allocation was calculated numerically by maximizing equation (13) subject to (16) and (14).

A.4.2 Market Implementation

The market implementation is the standard Mirrlees implementation as described in the end of section 3. That is, the budget constraint of the households is given by:

$$Y = X_1 - X_2 - T(X_1) - T(X_2),$$

and marginal tax rates are found by equating taxes to the optimal wedges in the second-best allocation, while assuming separability of the marginal tax rates.

A.4.3 Parametrization

For the figures in the main text we used the following parameterization: $\alpha = \beta = \xi = \zeta = 1.5$. In addition, we chose $\gamma_Y = \gamma_l = \delta_Y = \delta_l = 1$. Finally, the planner is assumed to be utilitarian such that $\nu = 0$. This yields the following set of equations:

$$
\begin{aligned}
SW &= \frac{1}{f} \sum_{j=1}^{J} \frac{1}{1-\nu} U_j^{1-\nu} \\
U &= U_p + U_c \\
U_p &= \log(Y) - \frac{1}{1.5} \sum_{i=1}^{2} \left(\frac{X_i}{n}\right)^{1.5} - \frac{1}{1.5} \left(\frac{X_1 X_2}{n^2}\right)^{1.5}, \\
U_c &= \log(Y) - \frac{1}{1.5} \sum_{i=1}^{2} \left(\frac{X_i}{n}\right)^{1.5} + \frac{1}{1.5} \left(\frac{X_1 X_2}{n^2}\right)^{1.5}, \\
0 &\geq \sum_{j=1}^{J} Y^j - X_1^j - X_2^j.
\end{aligned}
$$

We programmed this optimization problem in Matlab. Codes are available upon request.