Multiple Dimensions of Tax-Design,

Characterization and Implementation of the Social Optimum*

Sander Renes** Floris T. Zoutman***

November 1, 2013

Abstract

We characterize the second-best allocation and its implementation in a generalized model of optimal non-linear taxation. We show that, in two important classes of models with several tax-tools and several reasons for redistribution, the tax-tools can be described by formulas that lend themselves to direct economic interpretation. These two classes of models cover most of the existing literature. In these classes the optimal tax-system can be studied through the characterization of the second-best distortions that define the tax-rates. Four properties of these second-best distortions are shown. First, the optimal allocation can be described by a generalized version of Diamond's (1998) and Saez' (2001) formula for the optimal wedges. Second, a no-distortion at the top/bottom result holds. Third, the famous Atkinson-Stiglitz theorem that uniform indirect taxation is optimal does not hold in multi-dimensionality. Fourth, the optimal wedges generally contain as many interdependencies as the dimension of the type-space. These findings can be explained in an intuitive manner by interpreting the optimal distortions as the tools used by the planner to elicit information from the individuals.

Keywords: optimal non-linear taxation, redistribution, tax system, market implementation, price mechanism, private information.
JEL-codes: D04, D82, H21, H22, H24

---

*This paper combines the insights from Renes and Zoutman (2013a,b). The authors would like to thank Felix Bierbrauer, Eva Gavrilova, Aart Gerritsen, Yasushi Iwamoto, Bas Jacobs, Laurence Jacquet, Etienne Lehmann, John Morgan, Dominik Sachs, Dana Sisak, Dirk Schindler, Bauke Visser, Casper de Vries and Hendrik Vrijburg for useful suggestions and comments on an earlier version of this paper. Furthermore, this paper benefitted from comments and suggestions made by participants at the 2011 Nake Conference, Utrecht, the 2013 CESifo Area Conference on Public Economics, Munich, the 69th IIPF Conference, Taormina; and seminar participants at the Erasmus School of Economics, the Norwegian University of Science and Technology, the Norwegian School of Economics, the University of Konstanz and the Centre for European Economic Research. All remaining errors are our own.

**PhD. candidate at Erasmus University Rotterdam and Tinbergen Institute. Corresponding author. E-mail: srenes@ese.eur.nl. Erasmus School of Economics, Erasmus University Rotterdam, PO box 1738, 3000 DR Rotterdam, The Netherlands.

***PhD. candidate at Erasmus University Rotterdam and Tinbergen Institute. E-mail: zoutman@ese.eur.nl

1 Introduction

How should a government combine taxes on labor income with healthcare subsidies? What is the relation between capital and labor income taxes? When should housing subsidies depend on wealth and income? How can the tax-system incentivize couples to share the burden of generating income and managing the household?

The current literature on optimal taxation can only provide the answers to these questions if individuals differ in a single dimension, such as earnings ability, and play a (game that is equivalent to a) direct revelation game with a central planner. Asking these questions in full earnesty thus uncovers two aspects of the problem of optimal taxation that have so far received little attention. First, we know very little about optimal (tax-) policy when individuals differ in several unobservable dimensions like ability, preferences and health status. Second, the relationship between the direct mechanism used to identify the social optimum and the actual tax-system that would implement the social optimum in the market has not been sufficiently characterized. In this paper we take the first steps to address both issues. We first identify two important classes of situations in which a straightforward tax-schedule implements the second-best for any number of hidden characteristics and choices of the individuals. within the biggest of these classes we characterize the second-best allocation. Our results show that in this generalized model the tax schedule that creates the social optimum can be described by a set of equations that yield themselves to direct economic analysis.

In two seminal articles Mirrlees (1971, 1976) pioneered what is now the canonical approach to tax design. He characterized a social planner's second-best allocation in a static model where individuals are heterogeneous in a single unobserved characteristic and make multiple choices. In this approach the planner first has to find the second-best allocation through a direct revelation mechanism, and then has to find a tax-system that can implement this allocation. This approach to the study of taxation has yielded many fruitful results since.

Notable contributions are made, for instance, by Diamond (1998) and Saez (2001). They rewrote the solution of the second-best allocation into a formula (known as the ABC-formula) that describes the optimal distortion in terms of the difference between the marginal rates of transformation and the marginal rates of substitution. These differences (from now on: wedges) can be expressed as a function of measurable elasticities, which allows for a more intuitive explanation of the characteristics of the second-best as well as a convenient way to approach data. Unfortunately, these papers mostly study unidimensional heterogeneity, which makes it difficult to discuss the interactions between social schemes and taxation that characterize modern welfare states. Moreover, they leave the market-implementation of the second-best allocation mostly implicit, which provided the motivation for this paper.

In our model individuals differ in $p \geq 1$ hidden characteristics such as ability, health-status and patience. They make $k \geq p$ observable continuous choices concerning for instance labor income, savings, consumption and portfolio composition. Additionally, they choose how much to consume of a numeraire good. We will often refer to these choices as goods, although they can be both inputs and outputs in the production process. The assumption that $k \geq p$ is crucial to our analysis. It ensures we can rely on the revelation principle derived in Myerson (1979) since in an incentive compatible mechanism the observed choices can reveal all hidden characteristics ex-post.

We follow Mirrlees' approach and split the problem of designing an optimal tax-system into two parts, finding the optimum and identifying the tax-schedule that implements it. Relying on backward induction we start by characterizing sufficient conditions for a specific class of tax-schedules to implement the second-best. We begin by showing the need for a better characterization of the implementing tax-schedule. A simple example demonstrates that the canonical pure price tax-mechanism, as suggested by Mirrlees (1976), does not always implement the allocation. This problem could be solved by the "principle of taxation", first derived in Hammond (1979), that shows that for every incentive compatible allocation at least one non-linear tax system exists that implements it. This tax-system is a price mechanism that is strengthened by the addition of rules, prohibitions, quotas and/or prohibitive taxes. However, this result is of limited value to tax-policy makers in market economies. It requires the prohibition of a large part of the choice space, which seems unrealistic and undesirable in real-world democracies.[1] This leaves a gap in our understanding. For any second-best allocation it is unclear whether the canonical tax-implementation works, while the system that always works is unrealistic and undesirable. We bridge part of this gap by identifying two classes of models in which any second-best allocation can be implemented with a pure price mechanism, without using prohibitions.

To do so our first lemma derives the first and second order incentive constraints that a tax-schedule has to satisfy to implement a specific allocation. After solving the maximization problem of the social planner and formulating the entire tax-system, these constraints can be used to verify whether the tax-system implements the desired allocation. Unfortunately, most optimal allocations in public finance do not have a closed-form solution. Solutions may be obtained through numerical simulations, but this implies that verification of implementation can only be performed on the special cases that have been simulated. Verification of implementability is useful in such simulations, but does not provide insights in the general properties of optimal tax-systems.

We therefore continue by identifying two classes of maximization problems in which a pure price mechanism that equates marginal taxes to optimal distortions always implements the allocation. First, implementation does not have to be verified if i.) the tax system does not have an internal maximum in tax revenue, ii.) the allocation is second-best for a welfarist social planner, and iii.) there are no externalities. This result does not rely on the particular preferences of the agents. Because the planner's objective is strictly increasing in the utility of the agents, the planner does not want to hinder the agents if it can be avoided. Second, if there is a one-to-one correspondence between the type space and the choice space, then incentive compatibility and implementability constraints coincide. A second-best allocation is by definition incentive compatible, such that in this case it must be implementable. In this last case, equating taxes to optimal distortions yields a unique tax system. These two classes contain most models based on Mirrlees (1971, 1976) so that our implementation results validate the tax-systems proposed in most of the literature. Outside of these classes of problems the tax-system can require rules (such as prohibitions) to implement the second-best, which sheds new light on rules used in real-world tax and benefit systems.

[1] The name 'principle of taxation' stems from latter applications of the principle derived by Hammond to taxation by a.o. Rochet (1985), Guesnerie (1995) and Bierbrauer (2009). However, the strength of the principle is mostly in applications like the multi-product monopolist pricing problem, e.g. Armstrong (1996). In such settings it perfectly covers implementation, since the monopolist can choose what (not) to produce and sell such that no prohibitions are necessary.

In the second half of the paper we characterize the second-best allocation in a Mirrleesian optimal tax-model with multi-dimensional heterogeneity. We begin by identifying a set of necessary conditions for incentive compatibility. Using these necessary conditions for incentive compatibility we show that several well known results from the unidimensional Mirrlees model have a multi-dimensional generalization.

However, as Mirrlees (1976, p.341) already noticed "matters are not quite so simple when the population is described by a vector n". To gain more intuition for our results we suggest a reinterpretation of the model and the corresponding equilibrium in terms of screening and information. Relying on the revelation principle, the social planner has to provide each individual with the incentives to truthfully reveal his private type, such that the planner faces a multi-dimensional screening problem. The taxes, or more precisely, the optimal distortions are the tools used by the planner to gain information about the type of each individual so that he can redistribute from one type to another.

In the second-best allocation, the optimal distortion at each point in the type-space and for each good can be described by the wedges between the marginal rates of transformation and substitution with respect to the numeraire good. These optimal wedges can be characterized by a generalized version of Diamond's (1998) and Saez' (2001) *ABC*-formula, now with an additive structure over the characteristics. For each type, optimal wedges (A) increase in the quality of the signal: how much does a specific choice variable reveal about underlying characteristics; (B) increase in the welfare benefits of marginally distorting the choice for this type; and (C) decrease in the size of the tax base for which the choice is distorted.

A corollary shows that the optimal wedge on a good is zero if the marginal rate of substitution of the good is independent of all hidden characteristics. In this case the chosen level of the good does not reveal information, and hence, it is not optimal to distort the good away from Laissez-faire. The corollary implies the Atkinson-Stiglitz (A-S) theorem in case of unidimensional heterogeneity. If disutility of labor is the only aspect of utility that is not separable from type in the utility function, the labor choice is the best signal of type. Indirect taxation yields no extra information and is thus not optimal. The corollary also immediately implies the A-S theorem does not extend to multi-dimensional heterogeneity. With at least two types of heterogeneity, a single tax can never extract all available information. If a tax planner wants to separate out able types that are unwilling to work hard, from less able types that want to work hard but cannot, he will have to use more than just an income tax.

Mirrlees (1976) shows that in the case of unidimensional heterogeneity the optimal wedge on each good can be written as a function of only that good. In case of multi-dimensional heterogeneity such separable wedges are (near) impossible. We show that in general, in order to facilitate full revelation of the p underlying characteristics, the marginal tax-rates on each choice have to depend on p choices. Such interdependencies are very common in the stochastic dynamic models of the New Dynamic Public Finance (NDPF). In these inter-temporal models, the tax on labor income in any period depends on state variables, like wealth, stemming from the history of play. These state variables form a natural extension to the type-space, since they contain information about the budget or preferences of the individuals. By modeling these state variables (or history) as dimensions of the type-space our model can be used as an intermediate step between the classical Mirrleesian public finance and the NDPF. This could provide some intuition for the mathematically complicated tax-schedules found in this literature. We show, for instance, that to separate out individuals that earn little because they are wealthy from individuals that earn little because they have a low ability, the planner has to make the tax-rate on savings depend on labor income and vice versa. This finding suggests that the interdependencies in the optimal tax systems of NDPF models are (in part) due to the structure of hidden information rather than the stochastic process.

Like in the unidimensional case the optimal wedge at the "corners" of the type-distribution is zero. If an individual exists that has only extreme values of characteristics in his type, his optimal marginal wedge on all choices equals zero.[2] Intuitively, since there are no "more extreme" types, setting a wedge to separate out more extreme types yields no information to the planner. Hence for any marginal distortion the efficiency cost of the distortion is higher than the welfare gain at the end-points of the distribution.

Like most papers in the optimal tax literature, we derive the optimal allocation in a relaxed problem that takes the first-order incentive constraints into account, while assuming the second-order incentive-constraints are met.[3] It is well-known in the multi-dimensional screening literature that violation of these second-order incentive constraints can cause bunching on part of the type-space (e.g. Armstrong, 1996, Rochet and Choné, 1998). We deal with this problem in a separate subsection, showing that bunching is relatively unlikely in our model, and if it occurs, our characterization of the optimal tax-rate is still valid on part of the type-space.

The rest of the paper is organized as follows. The next section discusses related literature. Section 3 describes the model and derives the conditions for incentive compatibility that will feed in to the implementation results of section 4. Section 5 takes the incentive compatibility conditions and derives the characterization of an (implementable) second-best allocation that is then compared to several famous results from the unidimensional setting. Section 6 concludes.

2 Related Literature

Optimal taxation under multi-dimensional heterogeneity has not received much attention in the literature. Although Mirrlees (1976, sect. 4) already outlines necessary conditions for a generalized model like ours, to the best of our knowledge no further attempt has been made to characterize the solution.

The most straightforward way of decreasing the complexities of multi-dimensionality is by restricting attention to a discrete type-space. Armstrong and Rochet (1999) show that with multi-dimensionality the mechanism designer can no longer assume that only the local, downward incentive constraints are binding. He has to make sure all incentive constraints are met. In a similar discrete model with bi-dimensional heterogeneity Cremer et al. (2001) show that if types differ in wealth and ability indirect taxes are optimally used to tax savings, such that the Atkinson-Stiglitz theorem fails in this case.[4] In our paper we will show that the Atkinson-Stiglitz theorem has to fail in any setting with several sources of heterogeneity. Kleven et al. (2009) study the taxation of couples where the types of both partners are random variables and the couples maximize a joint utility function. They

[2]This result is reminiscent of the theorem derived in Golosov et al. (2011) where the wedge at the bottom and top is zero if the stochastic process allows agents to be located at the extremes.

[3]A good introduction to this technique can be found in Wilson (1996).

[4]Saez (2002) shows a similar result in a continuous setting

show that even if the types of primary and secondary earners in a couple are independent of each other, the marginal tax rate of the secondary earner depends on the income of the primary earner and vice versa. We will show that this kind of interdependencies are typical for tax-systems that deal with multi-dimensional heterogeneity.

Several papers studying multi-dimensional screening with continuous choices have focused on settings where the number of choices is smaller than the number of characteristics, $k < p$ (e.g. Lewis and Sappington, 1988; Pass, 2012). In that case the revelation principle can only be applied after the dimension of the type-space has been reduced to the dimension of the choice-space. The dimensional reduction of the type-space can lead to ill-behaved final allocations. This is demonstrated in Choné and Laroque (2010) where the reduction leads to a welfarist social planner that wants to set negative marginal income taxes on part of the allocation. We restrict our attention to the case where the choice space is big enough to contain all information in the type-space in an incentive compatible allocation, $k \geq p$, such that we can apply the revelation principle to the problem directly.

With $k \geq p$ Rochet and Choné (1998) provide a characterization of the solution to a large class of multi-dimensional screening problems with both participation and incentive constraints, and a mechanism designer that maximizes monetary revenue. They show that a solution obtained by the first-order approach consistently violates second-order incentive constraints at the bottom of the type-space because of interactions between incentive and participation constraints. In our model there is no participation constraint, since there is no clear outside option. Although we cannot exclude the possibility of bunching, our results obtained by the first-order approach are valid for the (sub-)set of types that separate and we describe this subset. Furthermore, since in our setting the central planner does not care about revenue, his goals are directly opposed to the goals of the mechanism designer in Rochet and Choné (1998) and the characterization of the equilibrium in their paper does not extend to our setting. Despite these differences, the incentive constraints are similar in both models and we find that the solutions to the models share important characteristics.

The issue of how the second-best allocation can be implemented, that is, how a tax-system can be constructed that causes individuals to choose their assigned bundles on the market, is often left implicit in the optimal taxation literature. Given that the second-best allocation is (first and second-order) incentive compatible, the problems in the implementation are mostly caused by joint or double deviations. In the implementation individuals might choose to deviate on several choices simultaneously and create a choice pattern that is not assigned to any type in the direct mechanism. Since this choice pattern does not exist in the direct mechanism, incentive compatibility does not rule out profitable joint deviations.

Joint deviations are not unique to models of taxation, they are an issue in any mechanism design problem where the individuals' choice space is multi-dimensional. A particular class of joint deviations, known as unbalanced or skewed bidding, has been the studied extensively in the literature on procurement auctions, and operations research management.[5] The unbalanced bidding literature focuses on a principal that needs to procure several goods in a single contract, but who is uncertain about the exact quantities required at the time of the procurement procedure.[6] In the principal's first-best all goods

[5]See for an overview Cattell et al. (2007) and Renes (2011).

[6]For simplicity we focus on procurement auctions in this review. However, these are simply reverse

are acquired from the cheapest firm at zero profit for the firm. However, the principal cannot observe the firm's cost structure. The principal therefore uses an reverse auction to select his supplier. The bidding firms state their price for each individual good and these are multiplied with score weights, yielding a scalar score. The firm with the lowest score wins. Using the expected quantities as score weights would seem to select the cheapest supplier. Unfortunately, if the expected quantities are slightly misestimated by the principal, the bidding firm has a profitable joint deviation. By asking more for the goods that are under-weighted in the score rule and less for the over-weighted goods, the bidder can increase expected payment while keeping his score constant. Even if the bidders do not wish to increase any price in isolation, such joint deviations are profitable. For risk neutral bidders the optimal bid contains infinite prices, yielding unbounded profit and risk. Renes (2011) studies mechanisms to prevent skewed bidding but finds no general solution when the principal is committed to accepting the bid with the lowest score. He notes that legal rules in the US allow the government to reject unbalanced bids, creating a solution to the problem through prohibitions. Ewerhart and Fieseler (2003) study the optimal score-rule under unidimensional firm heterogeneity. Using the restriction that unit prices have to be weakly positive, and thus prohibiting a large part of the choice space of bidders, they recoup a version of the revelation principle and are able to determine a second-best allocation. Both solutions are logically equivalent to the principle of taxation. In both cases a large set of joint deviations or skewed bids are effectively prohibited to solve the implementation issues. We will show that a welfarist central planner does not necessarily face such implementation issues, which suggests that the reliance on rules is necessary because of the differences between the goals of the individuals and the central planner.

Recently the NDPF generalized the Mirrlees-model to a setting where a single-dimensional hidden characteristic follows a stochastic dynamic process.[7] Kocherlakota (2005), in line with the principle of taxation, shows that in such models the optimal tax system generally contains prohibitive tax rates on specific combinations of choices. Even in the case of an i.i.d. stochastic process Albanesi and Sleet (2006) find that excessive savings choices should be prohibited by, for example, setting a borrowing limit. This ensures the joint deviation of first saving too much and then working too little is not optimal. Because of these interdependencies, where the tax on labor income has to depend on savings, implementation is very difficult. Unfortunately, the intuition behind these interdependencies is not very well developed. One reason for this relative lack in intuition is the complexity of the models. As was shown by Pavan et al. (2010), adding dynamic stochastic processes to the type-space makes the model more complicated in two (interconnected) ways. First, the dynamic stochastic process of the type-space adds relationships between choices in different periods via an information channel. The information an individual gains about his type in the current period can influence the probability distribution over his future types. In that case the hidden information in the current period provides information about the hidden information that will be learned in the future. Secondly, the type becomes a multi-dimensional construct through the state variables. Decisions in the past, say the amount an individual saved, influence the choices in subsequent periods. Therefore these state variables have to become part of the type-space. Given this twofold change it is hard to identify the cause of the differences between the NDPF and the clas-

auctions and all issues encountered in procurement are also encountered in sale auctions. See Athey and Levin (2001) for an example of joint deviations in an sale auction.

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

sical Mirrleesian models. Since our model has the multi-dimensional type-space without the dynamic information channel, we can isolate the different causes. We also show that there are important classes of problems in the static setting where joint deviations are not an issue and prohibitive tax-rates are not necessary for implementation. Fortunately these classes can readily be related to the NDPF models.

3 The model

Assume there is a social planner that maximizes social welfare (SW) in an economy. The economy is populated by a unit mass of individuals that are characterized by a twice-differentiable utility-function:

$$u(\mathbf{x}, y, \mathbf{n})$$

Where $\mathbf{x} \in \mathbf{X} \subseteq \mathcal{R}^k$ denotes a vector of decision variables, $y \in Y \subseteq \mathcal{R}$ a numeraire decision variable, and $\mathbf{n} \in \mathbf{N} \subseteq \mathcal{R}^k$ denotes the type of an individual. The utility function is assumed to be increasing and concave in the numeraire: $u_y > 0, u_{yy} \leq 0$ for any value of $(\mathbf{x}, y, \mathbf{n})$, such that we have non-satiation of the utility function everywhere. Decision 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. 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 (e.g. we call individual effective labor supply a good as well). Each element $n_j \in \mathbf{n}$ is referred to as a characteristic, the full vector $\mathbf{n}$ is the type, and the space $\mathbf{N}$ will be referred to as the type-space. For technical convenience we assume $\mathbf{N}$ is an open convex set. Let $\mathbf{n}$ follow a multi-dimensional differentiable cumulative distribution function $F(\mathbf{n})$, with $F : \mathbf{N} \to [0, 1]$ and probability density $f(\mathbf{n}) \geq 0 \quad \forall n \in \mathbf{N}$ both defined on the closure of N. For simplicity we assume probability density is strictly positive on the interior of $\mathbf{N}$.[8] Each characteristic is assumed to denote some independent aspect of the individuals, such that no characteristic can be found as a deterministic function of the other characteristics. His type is private information to each individual. Our model may be interpreted as either static or dynamic, depending on whether decisions in $\{\mathbf{x}, y\}$ occur in the same period or in different periods. However, we do assume that their full type $\mathbf{n}$ and the mechanism are revealed to the agents before the first period.[9]

It is assumed that $k \geq p \geq 1$, such that there are at least as many decision variables in $\mathbf{x}$ as characteristics in $\mathbf{n}$. This assumption allows the direct application of the revelation principle in our analysis (see Myerson, 1979), since the bundle of choices is big enough to contain all information in the type-space in an incentive compatible allocation.

The social planner is assumed to maximize a concave sum of the individuals' utility:

$$SW = \int_{\mathbf{N}} W(u(\mathbf{x}, y, \mathbf{n}))\, dF(\mathbf{n}), \tag{1}$$

$$W' > 0, W'' \leq 0, \tag{2}$$

[8] See Hellwig (2010b) for a treatment of unidimensional incentive problems where the type-distribution may have holes and mass points.

[9] For an example how this model can be used to study intertemporal settings see section 5.3, or Golosov et al. (2013) for a more complete example.

where $W(\cdot)$ is a Bergson-Samuelson welfare function. We assume the social planner commits to the allocation he offers such that he cannot alter the allocation after types are revealed.[10] We assume that redistribution is welfare increasing because of (at least) one of two reasons. First, concavity in the utility functions of the individuals would imply that individuals with higher utility have lower marginal utility of consumption of each good. Second, $W'' < 0$ would imply the social planner gives higher welfare weights to individuals with lower utility.

The social planner is bound by the economy's resource constraint:

$$\int_{\mathbf{N}} y\left(\mathbf{n}\right) dF\left(\mathbf{n}\right) + R \leq \int_{\mathbf{N}} q\left(\mathbf{x}\left(\mathbf{n}\right)\right) dF\left(\mathbf{n}\right), \tag{3}$$

where R denotes exogenous government expenditure and $q\left(\cdot\right)$ is the economy's production of y as a function of the decision variables in $\mathbf{x}$. A partial derivative q_{x_i} may be either positive or negative depending on whether decision variable x_i is an input, respectively an output variable of the production process. We assume the production technology exhibits diminishing marginal returns such that $q_{x_i x_i} \leq 0$ for all decisions x_i. Together with the non-satiation of the utility function this guarantees that an interior solution will be reached in laissez-faire.

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

3.1 Incentive compatibility

Before we go to the problem faced by the social planner, we need to consider the problem of the individuals in our economy. In particular, we derive the conditions under which an allocation is incentive compatible. The incentive compatibility constraints will subsequently be used to verify implementability and solve for the optimal allocation. In a direct mechanism the social planner offers bundles $\{\mathbf{x}^*\left(\mathbf{m}\right), y^*\left(\mathbf{m}\right)\}$ for all $\mathbf{m} \in \mathbf{N}$. Each individual selects a bundle $\{\mathbf{x}\left(\mathbf{m}\right), y\left(\mathbf{m}\right)\}$ by sending a message $\mathbf{m} \in \mathbf{N}$ to the social planner. Function $\mathbf{x}^*$ maps from the message space to the choice-space, $\mathbf{x}^* : \mathbf{N} \rightarrow \mathbf{X},$ and y^* maps from the message space to the numeraire commodity space, $y^* : \mathbf{N} \rightarrow Y$. An allocation $\{\mathbf{x}^*\left(\mathbf{m}\right), y^*\left(\mathbf{m}\right)\}$ is incentive compatible if each individual truthfully reveals all his unobserved characteristics and receives the bundle designed for him in a direct mechanism.

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

$$\mathbf{n} = \arg\max_{\mathbf{m}} u\left(\mathbf{x}^*\left(\mathbf{m}\right), y^*\left(\mathbf{m}\right), \mathbf{n}\right) \quad \forall \mathbf{n} \in \mathbf{N}. \tag{4}$$

and in addition satisfies equation (3).

[10] See Roberts (1984) for a discussion on the issue of commitment.

Let:

$$V(\mathbf{n}) \equiv \max_{\mathbf{m}} u(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}) \tag{5}$$

denote the indirect utility function as a function of type. In an incentive compatible allocation $V(\cdot)$ satisfies:

$$V(\mathbf{n}) = u(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n})$$

This equation simply states that maximized utility equals the utility function under optimal choices. Let $\mathbf{s}(\mathbf{x}, y, \mathbf{n}) \equiv -\frac{u_{\mathbf{x}}(\mathbf{x},y,\mathbf{n})}{u_y(\mathbf{x},y,\mathbf{n})}$ denote the vector of shadow prices, such that element i of $\mathbf{s}$ denotes the marginal rate of substitution for decision variable x_i with respect to the numeraire y. Proposition 1 below largely follows Mirrlees (1976) and McAfee and McMillan (1988), it establishes the first and second-order conditions for incentive compatibility.

Proposition 1 *An allocation* $\{\mathbf{x} = \mathbf{x}^*(\mathbf{n}), y = y^*(\mathbf{n})\}$ $\forall \mathbf{n} \in \mathbf{N}$ *is incentive compatible if:*

for $\forall \mathbf{n} \in \mathbf{N}$ *it holds that*

$$y^{*\prime}(\mathbf{n}) = \mathbf{s}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n})^T \mathbf{x}^{*\prime}(\mathbf{n}) \tag{6}$$

$$\mathbf{x}^{*\prime}(\mathbf{n})^T s_{\mathbf{n}} \lessdot 0 \tag{7}$$

where the inequality sign, $\lessdot$*, signifies negative definiteness of the matrix.*

Through the envelope theorem a fully equivalent set of conditions can be derived:

$$V'(\mathbf{n}) = u_{\mathbf{n}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \tag{8}$$

$$u_{\mathbf{nn}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) - V''(\mathbf{n}) \lessdot 0 \tag{9}$$

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

Equation (6) states that all individuals should be indifferent between truth-telling and mimicking at the margin for all characteristics. For each row j the left hand-side of the equation denotes the gain in y as a consequence of marginally changing the reported characteristic n_j. The right-hand side denotes the utility loss in $\mathbf{x}$ measured in units of y for the same change. Therefore, equation (6) states that in equilibrium the marginal cost of mimicking equals the marginal benefits for all characteristics. Equation (7) is the usual second-order condition derived by Mirrlees (1976). If the marginal rate of substitution for decision variable x_i is increasing (decreasing) in characteristic n_j, $(s_i)_{n_j} > 0$ ($(s_i)_{n_j} < 0$), and the allocated amount of the good is also increasing (decreasing) in the characteristic, $(x_i^*)'_{n_j} > 0$ ($(x_i^*)'_{n_j} < 0$), then the allocation induces self-selection. It implies higher (lower) quantities of the good are assigned to people with a stronger (weaker) preference for the good.

The alternative formulations of the first and second order incentive compatibility constraints, equations (8) and (9) respectively, also have a very intuitive interpretation. Equation (8) can be obtained by taking the difference between the two utility profiles $u(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{m})$ and $V(\mathbf{m})$ and minimizing this difference. By definition the minimum difference of 0 is obtained if the first-order conditions of optimality are met. The equation thus identifies the type for which the bundle $\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}$ is an optimum. The second-order condition (9) is a mimicking constraint. Consider an individual with exogenous characteristics $\mathbf{n}+\varepsilon$ where ε is a p-vector with all elements equal to ϵ. Suppose in addition that he is considering whether he should truthfully report his type, or mimic an individual with exogenous characteristics $\mathbf{n}$. A Taylor-approximation of the utility from truth-telling is:

$$\begin{aligned} u\left(\mathbf{x}^*\left(\mathbf{n}+\varepsilon\right), y^*\left(\mathbf{n}+\varepsilon\right), \mathbf{n}+\varepsilon\right) &= V\left(\mathbf{n}+\varepsilon\right) \\ &= V\left(\mathbf{n}\right)+V'\left(\mathbf{n}\right)\varepsilon+\frac{1}{2}\varepsilon^T V''\left(\mathbf{n}\right)\varepsilon+\ldots. \end{aligned}$$

The utility gained by mimicking type $\mathbf{n}$ is given by:

$$u\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}+\varepsilon\right)=V\left(\mathbf{n}\right)+u_{\mathbf{n}}^T\varepsilon+\frac{1}{2}\varepsilon^T u_{\mathbf{nn}}\varepsilon+\ldots.$$

Subtracting the second from the first yields:

$$\begin{aligned} \Delta u\left(\varepsilon\right) &= \left(V'\left(\mathbf{n}\right)-u_{\mathbf{n}}^T\right)\varepsilon+\frac{1}{2}\varepsilon^T\left(V''\left(\mathbf{n}\right)-u_{\mathbf{nn}}\right)\varepsilon+\ldots \\ &= \frac{1}{2}\varepsilon^T\left(V''\left(\mathbf{n}\right)-u_{\mathbf{nn}}\right)\varepsilon+\ldots, \end{aligned}$$

where $\Delta u\left(\varepsilon\right)$ denotes the extra utility of truth-telling and the second equality follows from (8). Now take the limit as $\varepsilon\to\mathbf{0}_p$, then $\Delta u\left(\varepsilon\right)$ is positive as long as $V''\left(\mathbf{n}\right)-u_{\mathbf{nn}}$ is positive definite. Intuitively, the first-order condition (8) makes individuals indifferent between truth-telling and mimicking at the margin. The second-order condition (9) ensures that they strictly prefer truth-telling over mimicking any non-adjacent type.

4 Market implementation

In this section we aim to find the properties of a tax system that implements a feasible, incentive compatible allocation in the market. Therefore, we have to go beyond the direct mechanism that identifies the second-best and study the choice problem of agents in a market. Agents maximize their utility function (5) subject to their budget constraint in the market:

$$y\leq q\left(\mathbf{x}\right)-T\left(\mathbf{x}\right), \tag{10}$$

Where the tax function T, like the production function q, maps from the goods-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 $T(\cdot)$. We assume the tax system is twice differentiable, without any other a priori restrictions. The tax function may be fully non-linear and can contain arbitrarily complex interdependencies.

By Walras' law if the economies resource constraint, (3), and the agents' budget constraints, (10), are simultaneously satisfied, the government's budget constraint must also be satisfied. Therefore, if the tax system is successful in implementing a feasible allocation, we do not need to check whether the government's budget is balanced.

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 implements an allocation* $\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}$ *if each agent selects the bundle on the market that was assigned to him in the allocation:*

$$\begin{aligned}\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\} &= \arg\max_{\mathbf{x},y} \{u(\mathbf{x},y,\mathbf{n}) : y = q(\mathbf{x}) - T(\mathbf{x}), \mathbf{x} \in \mathbf{X}, y \in Y\} \\ \forall \quad &\mathbf{n} \in \mathbf{N}\end{aligned} \tag{11}$$

The difficulty of implementability can be understood by comparing definitions 1 and 2. In a direct mechanism the agents maximize their utility by sending the optimal p-dimensional messages. On the market the agents maximize their utility by choosing k goods. Since $k \geq p$ the agents can deviate in more directions on the market than in the direct mechanism. That is, the market allows the agents to create new bundles that were not assigned to any type in the direct mechanism. Such a strategy is 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. If a tax-system allows for profitable joint deviations, it cannot satisfy definition 2.

Mirrlees (1976, sect. 3) defined the tax-schedule that has become the canonical implementation for unidimensional heterogeneity ($p = 1$). The Mirrleesian implementation has two properties. First, marginal taxes are equated to optimal distortions such that $T_i'(\mathbf{x}^*(n)) = q_{x_i}(\mathbf{x}^*(n), y^*(n), n) - s_i(\mathbf{x}^*(n), y^*(n), n)$ for all goods. This is a very intuitive property for a tax system since it sets the market prices of all goods equal to the individuals' shadow prices. In section 4.2 we will show that this is a property of any optimal tax system. Second, in the Mirrleesian implementation $T_i'(\mathbf{x}^*(n))$ equals $T_i'(x_i^*(n))$. That is, the optimal tax on good i depends only on the consumption of good i. In this case the tax is separable. This property makes taxing joint deviations at prohibitive levels impossible. To define a joint deviation the vector of choices has to be compared to the vectors that constitute the allocation. A separable tax-system works on each element of the vector separately so can never identify joint deviations. Therefore, it can never prohibit them. As a result it is unclear if and when the Mirrleesian tax-schedule implements the desired allocation. In the next section we will illustrate this problem through a simple example in which incentive compatibility is not sufficient to guarantee implementability through a Mirrleesian tax schedule.

4.1 Failure of a pure price system: A simple example

Figures 1 and 2 give a clear example where a Mirrleesian implementation does not achieve the desired result.[11] The figures describe a situation with two goods in $\mathbf{x}$ and one exogenous characteristic, $k = 2$ and $p = 1$. In this example each type could represent a perfectly symmetric couple that maximizes a joint utility function. The spouses have to decide how much time each of the partners 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 the couple consumes, $y^*(n)$, as a function of the shared type. Since there is only one hidden characteristic, the bundles assigned to the types by this allocation form a line in $X_1 \times X_2 \times Y$ space. This line is represented by the black lines in figures 1 and 2. The lines show that the government wants the couple to share the burden equally. The dots represent the bundle of one particular couple. In figure 1 the vertical axis denotes the consumption bundle such

[11]The mathematics behind this example can be found in the appendix.

that the hyper-plane shows the budget constraint of individuals in the unique Mirrleesian implementation where taxes are equated to distortions and the tax system is separable.

In figure 2 the vertical axis shows the utility level. The surface now represents the utility function of the couple, with the assigned bundle at the dot, for all combinations $\{x_1, x_2, y\}$ that satisfy their budget constraint with equality. 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. Note that incentive compatibility implies that the bundle is either a maximum or a saddle-point in the couple's utility, we will use this observation in the next sub-section. In the market the couple can deviate from the allocation (line) while satisfying their budget constraint by using a joint deviation. In figure 2 all such deviations are located on the surface, and can yield much higher utility since the couple prefers having a single earner with a stay-at-home mom/dad. So although the allocation is incentive compatible, the Mirrleesian tax schedule does not implement it because the goals of the individuals and the government are too far apart.

The fact that Mirrlees' tax system cannot implement the allocation does not mean that the allocation cannot be implemented at all. Clearly, the government could levy prohibitively high tax rates for all off-allocation choices to prevent these deviations. This results in tax-systems that do not resemble realistic price systems. These systems use interdependencies to increase tax-rates to levels that will stop any agent from choosing certain combinations, while bundles arbitrarily close to some of these prohibited bundles can and are chosen. It is therefore more natural to interpret these tax-rates as the prohibitions or rules they are equivalent to. In this sense, the need to use interdependencies with prohibitive taxes signals that the limits of the price mechanism as a means to influence behavior are reached. We therefore say that in this situation implementation can only be reached through a non-price mechanism, for instance rules and prohibitions. Because such non-price mechanisms are very restrictive and are not very desirable, we will show that some problems can be solved with less restrictive implementations. We identify two important classes of such problems in the next subsections.

4.2 Implementation through Pure Price Mechanisms

In this section we first define the necessary and sufficient conditions for the tax implementation of an allocation. These conditions can be a useful test to verify whether a specific tax-system implements a specific allocation. However, in order to perform the test one first needs to derive the entire allocation and tax-system. In many cases an explicit, closed form solution for the allocation that we are interested in is not available. Numerical solutions are available, but these describe only special cases by definition. One can verify that the specific tax-system implements the specific allocation under analysis, but the solutions cannot be used to say anything about tax-systems in general. To overcome this problem we need to classify problems where the second-best allocation can be implemented through a straightforward tax-system, without creating profitable joint deviations. We identify two of these classes of maximization problems and describe their characteristics. In these two classes we will show that a tax system that equates marginal tax rates to optimal distortions implements the desired allocation without off-allocation prohibitions. In these classes of problems the solution satisfies the implementation con-

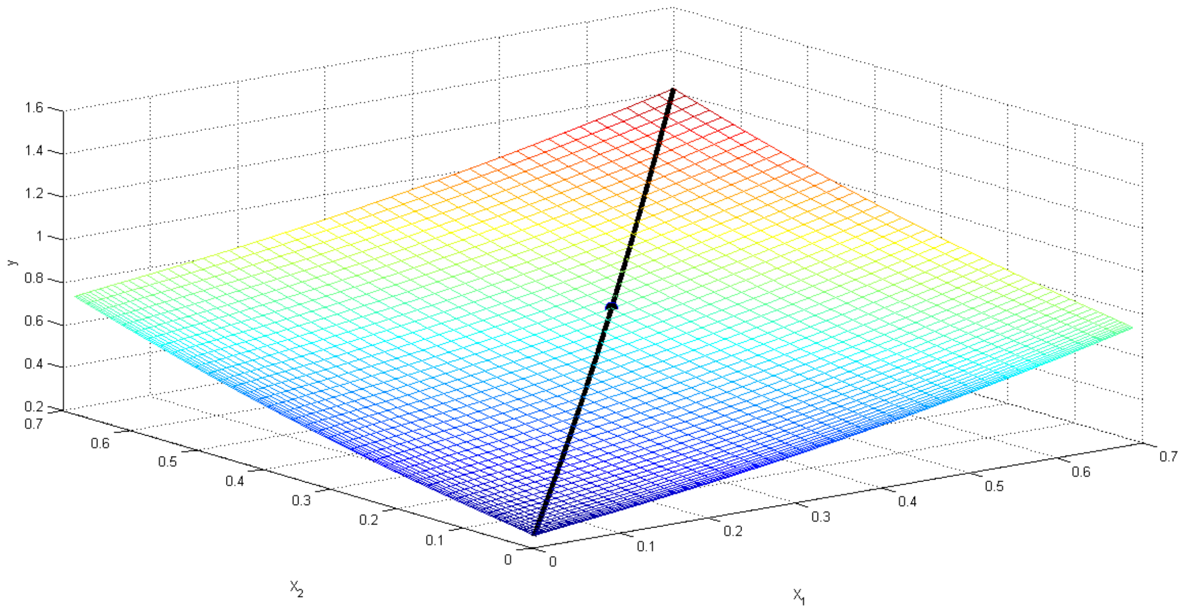

Figure 1: An optimal allocation and a budget constraint.

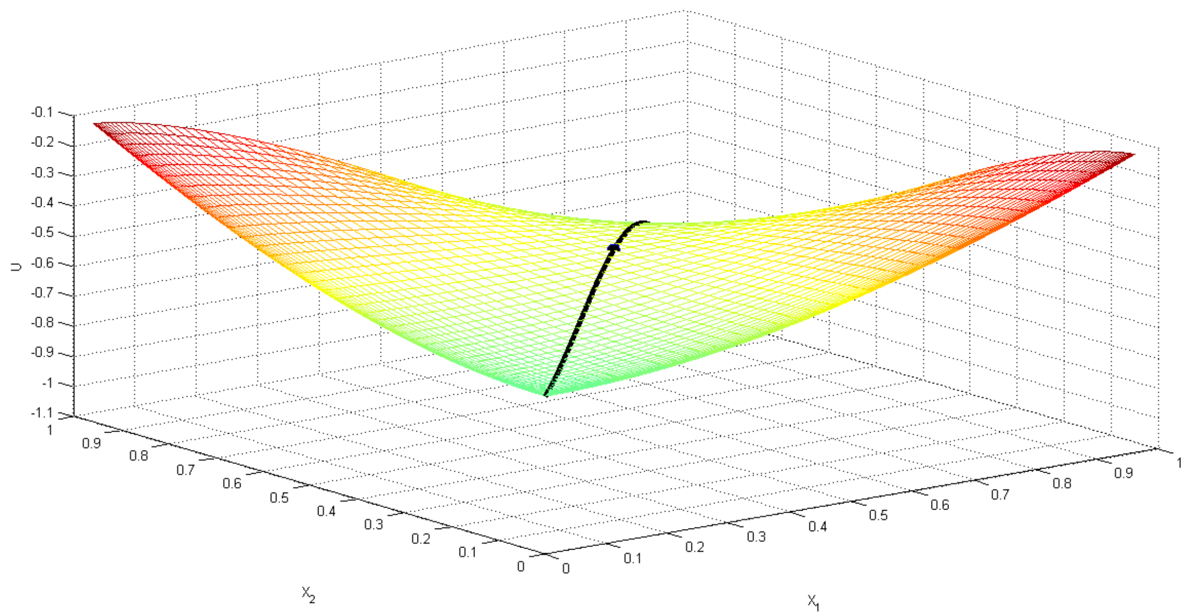

Figure 2: Utility of a couple faced with the budget constraint in Figure 1.

straints, no matter how the parameters of the model are calibrated or determined. Since the optimal tax system in these settings is a pure price mechanism that can be studied by focusing on the optimal distortions alone, we can make more general statements about the optimal tax system by studying these distortions. The first class of problems is characterized by a planner that maximizes a concave sum of individual utility. The second class of problems is characterized by a bijective relationship between the choice and the type-space. These two classes of models contain the canonical models of Mirrlees (1971, 1976) and most models based on those pioneering papers. This shows the two classes cover most of the current literature on the classical Mirrlees model of optimal taxation.

4.2.1 A useful lemma: implementation conditions

To show our main results, we first need to define more clearly what we mean by implementation. Lemma 1 derives the general conditions under which a differentiable pure price tax-system implements an allocation, by formally solving the problem of definition 2.

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

$$y^*(\mathbf{n}) = q(\mathbf{x}^*(\mathbf{n})) - T(\mathbf{x}^*(\mathbf{n})), \tag{12}$$

ii.)

$$T_i'(\mathbf{x}^*(\mathbf{n})) = q_{x_i}(\mathbf{x}^*(n), y^*(n), n) - s_i(\mathbf{x}^*(n), y^*(n), n), \tag{13}$$

iii.)

$$-\frac{\partial \mathbf{s}(\mathbf{x}, y(u, \mathbf{x}, \mathbf{n}), \mathbf{n})}{\partial \mathbf{x}} + q''(\mathbf{x}^*(\mathbf{n})) - T''(\mathbf{x}^*(\mathbf{n})) \lessdot 0. \tag{14}$$

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

Equation (12) 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 $y^*(\mathbf{n})$ if they choose their assigned quantities $\mathbf{x}$, and vice versa. Equation (13) is the first order condition for a market implementation. It states that marginal taxes are equated to marginal distortions. 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 (13) 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 off the allocation. In our example of figures 1 and 2, this translates to a tax-system that is fully defined on the line, but undefined everywhere else.[12]

Equation (14) states that the indifference curve of any linear combination of x_i's with respect to y should be more convex than the budget constraint for the same linear

[12]The problem of how to define the tax-rate off the allocation has been asked elsewhere. Mirrlees (1976) suggested making the tax-schedule separable. This leads to a unique implementation as in figure 1. Similarly, Hellwig (2007) focuses on "canonical" tax-schedules, which he defines as being continuous at end-points of the type-space. Although not necessarily unique, this does narrow down the class of tax-schedules.

combination of x_i's. This condition differs 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 if the indifference curve is convex, it is automatically more convex than the budget constraint. 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.

Since the conditions derived are both necessary and sufficient, they can be used to verify whether or not a specific tax system implements an allocation, after both the allocation and the tax-schedule have been determined. Because we restrict ourselves to differentiable tax systems it is technically impossible to prohibit joint deviations in the spirit of Hammond (1979). If the conditions of lemma 1 are met, the allocation is implemented through a price mechanism alone.

The next two subsections provide our main results on implementation, and characterize two situations in which we can guarantee that the tax-schedule defined by the distortions on the allocation can implement the allocation, without the need of verifying implementation afterwards.

4.3 Pure price mechanisms: Second-best of a welfarist planner

As we have seen in figure 2 a pure price tax system may sometimes place agents on utility saddle-points, causing them to deviate from the desired allocation if given the possibility. In the next proposition we show that, if there are no externalities, an allocation that places individuals in saddle-points allows Pareto-improvements and thus cannot be the optimum of a welfarist planner.

Proposition 2 *If an allocation maximizes a welfare function* $SW = \int W(u)\, dF(\mathbf{n})$*, subject to the incentive compatibility constraints and the resource constraint, and* $W' \geq 0$ *, then any tax schedule* $T(\mathbf{x})$ *that does not have an internal maximum in tax-revenue in any linear combination of the* $\mathbf{x}$*'s on the allocation and satisfies (12) and (13) has to satisfy (14).*

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

Intuitively, any tax schedule that does not satisfy (14) allows for at least one deviation that increases the utility of at least one agent. In addition, since the first-order conditions (13) combined with the violation of (14) imply that the agent is located in a saddle-point, the exact opposite deviation must increase his utility by approximately the same amount. This can easily be seen in figure 2. The agents' utility increases as much if he moves to the right off the allocation as when he moves to the left. Provided tax revenue is not maximized in the allocation, tax revenue must weakly increase either for the deviation to the right, or for the opposite deviation to the left. In figure 1 the tax schedule is monotone and hence such a deviation exists. This deviation increases tax-revenue, individual utility and welfare. It forms a Pareto improvement over the allocation, while it does not violate the resource constraint since it increases tax-revenue. The allocation in figure 1 could therefore not have been second-best for a welfarist social planner. Since such a deviation is possible in any allocation that places individuals in a saddle-point, saddle points cannot be part of the second best of a welfarist planner. Combing the observations made we can prove the implementability of the second best allocation of a welfarist social planner. By construction the second-best allocation forms a fixed point in individual utilities. Incentive compatibility guarantees that the allocation cannot be minimum in individual utility and a Pareto argument rules out saddle-points. Therefore, the second-best of a welfarist social planner has to form a maximum in the individuals problem. This second-best must therefore be implementable by a tax-system satisfying equations (12) and (13).

This proof breaks down in the presence of externalities, internalities or the more general non-welfarist approach to taxation. With externalities the deviation of any agent can influence the utility of other agents, such that it is unclear when a deviation from the saddle-point entails a Pareto-improvement. This implies that implementability has to be checked through lemma 1 in this case, since the goals of the individuals and the planner might differ too much for implementation through prices.

In practice, the restriction that a tax schedule does not contain an internal maximum in revenue is rather weak. As can be seen from lemma 1, the first derivative of the tax system is defined by the optimal distortion on a good. As such, if none of the distortions change sign from positive to negative the tax system does not contain internal maxima in tax revenue. Most models in public finance satisfy this criterion trivially because each good is either taxed or subsidized over the entire domain.

4.3.1 Mirrleesian Implementation

By proposition 2 a convex or monotonic tax system can implement an allocation provided it is optimal to a welfarist planner. In Mirrlees (1976) the planner is welfarist and the second-best is characterized by the equations (12) and (13) such that a separable tax-system can implement this allocation provided distortions do not change sign from positive to negative and there is only one-dimensional heterogeneity. This is summarized in the next corollary

Corollary 1 *If $p = 1$ and the conditions of proposition 2 are met, the Mirrleesian tax system can implement the second-best.*

Proof. The proof follows from propositions 2 and the separability of the optimal wedges in Mirrlees (1976) ■

Although our tax-schedule is closely related to the Mirrleesian implementation, we have to be careful with separability of the tax-schedule. Due to the nature of second-best distortions, separability of the tax-schedule can only occur in the case of unidimensional heterogeneity. If heterogeneity is multi-dimensional, a separable tax can never separate out all dimensions of the type-space, such that the distortions have to be non-separable by definition (see prop. 4). The separability found in Mirrlees (1976, sect. 3) is therefore lost in the case of multi-dimensional heterogeneity, and the Mirrleesian implementation can therefore only be used in a setting with unidimensional heterogeneity.

4.4 Pure price mechanisms: A bijective allocation

The combination of (12) and (13) defines the tax schedule on the allocation. If the allocation perfectly covers the choice space this pure price tax-schedule must implement the allocation. Proposition 3 provides a sufficient condition for such a unique tax implementation to exist.

Proposition 3 *If the mapping* $\mathbf{x}^*(\mathbf{n})$ *is bijective, then the tax implementation described by equations (12) and (13) is the unique differentiable tax-schedule that implements the second-best allocation.*

Proof. proof in appendix ■

Note that bijectiveness of the mapping $\mathbf{x}^*(\mathbf{n})$ is a rather strict requirement. It requires $\mathbf{x}$ and $\mathbf{n}$ to be of the same dimension and both be similarly (un)bounded, such that incentive compatibility and implementability coincide. In this situation every choice in the market corresponds to the choice of a unique type in the direct mechanism, and every type in the direct mechanism to a unique bundle on the market. Since all types prefer their own bundle over the bundles of the other types and all bundles are assigned to a type, all types prefer their bundle above all other bundles.

The allocation derived in Mirrlees (1971) is an example of a bijective allocation, provided the ability distribution is unbounded. In the direct mechanism all ability types are assigned a specific gross income level x. Mirrlees shows that if ability is continuously distributed in $\mathcal{R}_+$, the second-best allocation assigns all gross income levels to a specific ability type without bunching. Hence, the function $x^*(n)$, mapping ability to gross income, is bijective. Then by definition incentive compatibility and implementability coincide, thus incentive compatibility is enough to ensure double deviations are not profitable even in case of multiple dimensions of heterogeneity. When the choice-space of the individuals is larger than the type-space, implementability conditions cover a larger space than incentive compatibility conditions. In this case, since the choice-space is larger than the type-space, (double) deviations outside of the allocation are not ruled out by incentive compatibility. The principle of taxation derived in Hammond (1979) extends the underlying idea of proposition (3) to this situation. If all choices outside the allocation are prohibited the allocation must be bijective in the remaining, non-prohibited choice space. That is, the space that is not prohibited by the mechanism designer has a one-to-one correspondence with the type space. Hence, implementability is guaranteed. However, as we have noted before, this restriction of the choice-space requires the planner to use prohibitions conditioned on the entire vector of choices and results in a rather unrealistic tax system.

5 The second-best allocation

Now that we have established the conditions for incentive compatibility and implementability, we can turn our attention to the social planner. In both our classes of implementable second-best allocations we find that the allocation is implementable if it is given through the first-order conditions of incentive compatibility (13) and the budget-constraint (12). That is, the allocation needs to be an internal maximum in the problem of the individuals that exactly meets their budget constraints. Consequently, and like most papers in the optimal taxation literature, we will use the first-order approach. We analyze a relaxed version of the maximization problem that assumes the second-order incentive constraints are met in optimum. This can be verified ex-post through equation (7) or equivalently (9). We will return to the problem of violations of the second-order constraints in section 5.4. To ensure we do not suffer additional implementation problems we assume that the social planner is welfarist and there are no externalities, such that we remain within the class of models described in proposition 2.

5.1 Solving the Lagrangian

The first-order approach allows us to write the problem of the social planner as maximizing social welfare subject to the first-order incentive compatibility constraint (8) and feasibility constraint (3):

$$
\begin{aligned}
& \max_{V(\mathbf{n}),\mathbf{x}^*(\mathbf{n}),y^*(\mathbf{n})} \int_{\mathbf{N}} W\left(V\left(\mathbf{n}\right)\right) dF\left(\mathbf{n}\right), \quad s.t. && (15)\\
0 & \geq R + \int_{\mathbf{N}} \left(y^*(\mathbf{n}) - q(\mathbf{x}^*\left(\mathbf{n}\right))\right) dF\left(\mathbf{n}\right), \\
V'\left(\mathbf{n}\right) & = u_{\mathbf{n}}\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)^T, && (16)\\
V\left(\mathbf{n}\right) & = u\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right), && (17)
\end{aligned}
$$

where maximized utility $V\left(\mathbf{n}\right)$ is explicitly modeled as a choice variable. The last constraint, (17), guarantees that maximized utility is equal to the value of the utility function on the allocation. The Lagrangian to this problem is given by:

$$
\mathcal{L} = \int_{\mathbf{N}} \left[\left(W\left(V\right) - \lambda\left(R + y^* - q(\mathbf{x}^*)\right)\right) f + \theta^T\left(V'^T - u_{\mathbf{n}}\right) + \eta\left(u - V\right)\right] d\mathbf{n},
$$

where λ is the Lagrangian multiplier associated with the resource constraint, $\theta\left(\mathbf{n}\right)$ is a p-column vector of Lagrangian multipliers for the set of local incentive compatibility constraints, and $\eta\left(\mathbf{n}\right)$ is the Lagrangian multiplier that ensures maximized utility equals the utility function for each type. Note that s, f, F, θ, η, u and their derivatives depend on $\mathbf{n}$, but for clarity of exposition this notation is suppressed. We let $\partial\mathbf{N}$ denote the boundary of $\mathbf{N}$ and $d\mathbf{s}$ the outward unit surface normal vector to the boundary of $\mathbf{N}$. Through the divergence theorem (or multi-dimensional integration by parts) we can rewrite the Lagrangian as:

$$
\begin{aligned}
\mathcal{L} & = \int_{\mathbf{N}} \left[\left(W\left(V\right) - \lambda\left(R + y^* - q(\mathbf{x}^*)\right)\right) f - V\sum_{j=1}^{p}\frac{\partial\theta_j}{\partial n_j} - \theta^T u_{\mathbf{n}} + \eta\left(u - V\right)\right] d\mathbf{n} \\
& \quad + \int_{\partial\mathbf{N}} V(\theta^T \cdot d\mathbf{s})\mathbf{d}\partial\mathbf{N} && (18)
\end{aligned}
$$

Then, assuming the functions V and θ are smooth, this function can be maximized pointwise on the interior and boundary of the type-space.

On the interior of the type-space the first-order conditions with respect to variables $\mathbf{x}$, y and V are:

$$
\begin{aligned}
\frac{\partial\mathcal{L}}{\partial y} & = 0 : -\lambda f - u_{y\mathbf{n}}\theta + \eta u_y = 0, && (19)\\
\frac{\partial\mathcal{L}}{\partial\mathbf{x}} & = \mathbf{0}_k : \lambda q'^T f - u_{\mathbf{x}\mathbf{n}}\theta + \eta u_{\mathbf{x}} = \mathbf{0}_k, && (20)\\
\frac{\partial\mathcal{L}}{\partial V} & = 0 : W' f - \sum_{j=1}^{p}\frac{\partial\theta_j}{\partial n_j} - \eta = 0. && (21)
\end{aligned}
$$

The next proposition uses these first-order conditions to derive an elasticity-based formulation of the optimal wedge in the spirit of Diamond (1998) and Saez (2001). The discussion of this result is postponed to the next section where we compare our implicit solution of the multi-dimensional case to that of the unidimensional case.

Proposition 4 *The optimal wedge on good i for type $\mathbf{n}$ can be described by the following formula:*

$$\begin{aligned} \frac{q_{x_i}\left(\mathbf{x}^*\left(\mathbf{n}\right)\right)-s_i\left(\mathbf{x}^*\left(\mathbf{n}\right),y^*\left(\mathbf{n}\right),\mathbf{n}\right)}{s_i\left(\mathbf{x}^*\left(\mathbf{n}\right),y^*\left(\mathbf{n}\right),\mathbf{n}\right)} &= \sum_{j=1}^{p} A_{ij}\left(\mathbf{n}\right)B_{ij}\left(\mathbf{n}\right)C_{ij}\left(\mathbf{n}\right) \\ \forall \quad i &= 1,\ldots,k; \mathbf{n}\in\mathbf{N}, \end{aligned} \tag{22}$$

where:

$$\begin{aligned} A_{ij}\left(\mathbf{n}\right) &\equiv \varepsilon_{x_i n_j}\left(\mathbf{n}\right) = -\frac{\partial s_i\left(\mathbf{x}^*\left(\mathbf{n}\right),y^*\left(\mathbf{n}\right),\mathbf{n}\right)}{\partial n_j}\frac{n_j}{s_i\left(\mathbf{x}^*\left(\mathbf{n}\right),y^*\left(\mathbf{n}\right),\mathbf{n}\right)}, \\ B_{ij}\left(\mathbf{n}\right) &= \theta_j\left(\mathbf{n}\right)\frac{u_y\left(\mathbf{x}^*\left(\mathbf{n}\right),y^*\left(\mathbf{n}\right),\mathbf{n}\right)}{\lambda}, \\ C_{ij}\left(\mathbf{n}\right) &= \frac{1}{n_j f\left(\mathbf{n}\right)}. \end{aligned} \tag{23}$$

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

5.1.1 Boundary conditions, no distortion result

The boundary conditions can be found by differentiating equation (18) with respect to $V(\cdot)$ at the boundary of the type space ∂N. Only the final term of equation (18) depends on the boundary and hence, we get that:

$$\theta_j\left(\underline{n}_j\right)=\theta_j\left(\overline{n}_j\right)=0, \tag{24}$$

where $\underline{n}_j$ $(\overline{n}_j)$ represents the type that has the lowest (highest) value for characteristic j. There are at most 2^p types that have either the highest or the lowest value for all of their characteristics. Take any specific corner type $\mathbf{n}_{\lrcorner}^{\urcorner}$, then corollary 2 establishes that the optimal wedge on all goods for this type equals zero. Note that these types may or may not exist depending on the type-distribution, that is, for any type $\mathbf{n}_{\lrcorner}^{\urcorner}$ the distribution might be such that $f\left(\mathbf{n}_{\lrcorner}^{\urcorner}\right)$ is zero.

Corollary 2 *The optimal wedge for any type $\mathbf{n}_{\lrcorner}^{\urcorner}$ equals zero if the types exist.*

Proof. From the boundary conditions it follows that $\theta_j\left(\underline{n}_j\right)=\theta_j\left(\overline{n}_j\right)=0$ for all $j=1,\ldots,p$. The optimal wedge at the corner types can be found by taking the limit of equation (22) if $\mathbf{n}$ goes to a $\mathbf{n}_{\lrcorner}^{\urcorner}$.:

$$\lim_{\mathbf{n}\to\mathbf{n}_{\lrcorner}^{\urcorner}}\frac{q_{x_i}-s_i}{s_i} = \lim_{\mathbf{n}\to\mathbf{n}_{\lrcorner}^{\urcorner}}\sum_{j=1}^{p}\varepsilon_{x_i n_j}\frac{u_y\theta_j\left(\left(\mathbf{n}\right)\right)/\lambda}{n_j f\left(\mathbf{n}\right)}$$

which equals 0 provided $f\left(\mathbf{n}_{\lrcorner}^{\urcorner}\right)$ does not equal zero. ■

Corollary (2) shows that the no-distortion at the top and the bottom result, as derived in Sadka (1976), remains valid in a multi-dimensional framework. In this multi-dimensional space there are just a few extra corners to take care off. This is not a surprising result as Golosov et al. (2011) derives a similar result in a model where the type-space was extended with state variables stemming from dynamic uncertainty. In all these cases the transversality conditions of the problem guarantee the outcome. If there are no individuals of more extreme type, taxes cannot pry information from them. In terms of our motivating examples, if an individual is the healthiest, most able person around, it is unlikely that another individual would want to mimic his consumption and labor choice. This choice is not worth it for any-one else to begin with, so the social planner does not have to distort the choices of the healthiest, most able person to stop mimicking. While any distortion would come at an efficiency cost. By not distorting the choices, the surplus generated by this individual is maximized, allowing the planner the biggest amount of surplus to extract rents from. Such that efficiency requires that these choices are not distorted at the margin.

5.1.2 Finding the optimal allocation

To solve the maximization problem of the planner the second-best allocation has to be found. This can be done by combining the results in the last sub-sections in three steps. First, the set of equations (16, 17) can be used to solve for $\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}$ as an (implicit) function of $V'(\mathbf{n}), V(\mathbf{n})$ and $\mathbf{n}$:

$$\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\} = \psi(V'(\mathbf{n}), V(\mathbf{n}), \mathbf{n}).$$

Second, by means of this equation and (19) and (20) we can solve for $\{\theta(\mathbf{n}), \eta(\mathbf{n})\}$ as an explicit function of $\mathbf{x}^*(\mathbf{n})$, $y^*(\mathbf{n})$ and $\mathbf{n}$, and hence as an (implicit) function of $V'(\mathbf{n}), V(\mathbf{n})$ and $\mathbf{n}$:

$$\{\theta(\mathbf{n}), \eta(\mathbf{n})\} = \hat{\phi}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})) = \phi(V'(\mathbf{n}), V(\mathbf{n}), \mathbf{n}).$$

Finally, if we substitute this result into the last first-order condition (21) it becomes a second-order partial differential equation, that can be integrated numerically under the boundary conditions (24) and (3). The solution provides us with $V'(\mathbf{n})$ and $V(\mathbf{n})$ which can subsequently be used to find the allocation $\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}$. The optimal wedges can be found by substituting the solution $\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}$ into the ABC-formula (22).

Solving this problem is by no means trivial. An analytical solution clearly does not exist, and even solving these partial differential equations numerically is difficult. As we will discuss in section 5.4, the second-order incentive constraints can add even more complications. Solving the partial-differential equations also does not fully identify the tax-system. The solution "only" characterizes the optimal wedges at every point of the type-space. How to create a tax-schedule from this allocation was covered in the first part of this paper.

5.2 Comparing ABCs

We will now use the ABC-formula (22) to compare our outcomes to the unidimensional case. As in the unidimensional case, the left-hand side of equation (22) represents the optimal wedge on good i for type $\mathbf{n}$, while this distortion is broken down into different factors of interest on the right-hand side.

Our A-term is a measure of the informational value of good x_i. Intuitively, if the elasticity $\varepsilon_{x_i n_j}$ is large, it means that the preference (MRS) for choice i strongly increases in characteristic j. Hence, x_i is a very strong signal of characteristic j and therefore the optimal wedge is high. Our A-term is more general than the one derived in, Diamond (1998), Saez (2001) and Jacquet et al. (forthcoming) because we use a more general utility function. In Diamond (1998) and Saez (2001) the utility-function is of the form: $u(y,l) = u\left(y, \frac{x}{n}\right)$, where y is consumption, n is productivity and $x = nl$ is effective labor supply (or labor income). Therefore, their A-term can be directly related to the (Frisch) labor supply elasticity. Note that in the notation used in Saez (2001) our C-term is included in the A-term. In Jacquet et al. (forthcoming) the assumed utility function is $u^1(y) + u^2(x,n)$. Due to separability of the utility function, the A-term is related to the elasticity of the absolute preference for x instead of the marginal rate of substitution of x with respect to y. In all cases, however, the intuition behind the equation remains the same. Taxes are used to elicit information from the individuals. If the preference for work is a strong indicator of type, taxes should respond strongly to labor income (i.e. high marginal tax-rate on labor income).

The B and C-term can be interpreted as scale parameters that determine the size of the optimal distortion. The B-term represents the redistributive benefits of distorting choice i for characteristic j. θ_j Is the Lagrangian multiplier of incentive compatibility constraint j. Hence, it represents the welfare cost of separating out type $\mathbf{n}$ in characteristic j. In equilibrium this should equal the marginal welfare benefit of making the allocation marginally less incentive compatible in choice j (i.e. increase the distortion on choice j). By multiplying θ_j with u_y and dividing through λ the welfare gain for such a redistribution is expressed in the numeraire unit. An increase in the marginal welfare benefit of distortion (higher θ_j) logically increases the optimal distortion.

The C-term is related to the size of the tax-base that is distorted by the wedge. The denominator represents the size of the tax base with respect to characteristic j. The larger this tax base is, the bigger the incidence of the distortion and hence, the larger the efficiency cost associated with the distortion. Efficiency implies that the size of the distortion is inversely related to the incidence of the distortion, as the C-term clearly shows. In unidimensional cases, the C-term is often multiplied by $1 - F(n)$ to make it proportional to the (measurable) inverse hazard-rate of the ability distribution. This is corrected for by dividing the A or B-term through the same factor. In a single-dimensional distribution of types, such fractions have an intuitive interpretation as conditional means. Unfortunately, this interpretation is lost when the distribution is multi-dimensional.

The biggest difference between the unidimensional and our multi-dimensional ABC-formula is that in the later case one has to sum over all characteristics to get the optimal wedge for a good i. That the optimal distortion with a multi-dimensional type-space can be described with a formula that resembles the unidimensional description so clearly seems quite remarkable. The similarities are a clear testament to the strength of the ABC-formula as a description of the equilibrium in the Mirrleesian tax model. Especially the additive nature of the wedges over the different dimensions of the type space is surprising. It appears to indicate that wedges used to identify types in one dimension can be treated separately from wedges used to identify the other dimensions of the type-space. However, this appearance is deceiving. Intuitively, it seems unlikely that one can separate out individuals on wealth and income independently of each other. Individuals will treat monetary wealth and monetary income as substitutes in their budgets, thus creating interdependence in choices. This relation will lead to interdependencies between wedges in the second-best. In the optimum the marginal tax on income will have to depend on wealth if wealth and income are independent aspects of the type-space Similarly if the planner wants to redistribute more toward the unhealthy individuals with a given ability, than to similarly able individuals with good health, marginal tax-rates will have to depend on income and health. These complexities are discussed in the next subsection.

5.3 Complexity of the tax-schedule, between Atkinson-Stiglitz and the NDPF

The Atkinson-Stiglitz theorem (see Atkinson and Stiglitz (1976), from now on A-S) states that indirect taxation is superfluous in a setting with unidimensional heterogeneity and a utility function that is weakly separable in consumption and the type-space. The utility function in Jacquet et al. (forthcoming), $u^1(x_1) + u^2(x_2, n)$, is a clear example. Each individuals' choice of x_2 is the best signal of type, since the marginal preference for x_2 depends on the type. Taxing the choice of x_1 does not provide with any information that cannot better be gained by taxing x_2. This result has been generalized to more general utility functions and non-optimal taxation in Hellwig (2010a) and Laroque (2005). This result has not gone unchallenged in literature. Several authors have shown that the result does not hold if, for instance, separability is weakened (e.g. Boadway and Pestieau, 2011), unemployment insurance plays a role (Lee, 2011), several generations of agents are involved (e.g.Kopczuk, 2013), or, in a problem closely related to ours, under multi-dimensional (discrete) heterogeneity (e.g. Cremer et al., 2001; Saez, 2002). The next corollary generalizes the A-S theorem to a situation with multiple hidden characteristics. The generalization is not straightforward, however, since income taxation alone is never optimal in case of multi-dimensional heterogeneity. Instead it turns the interpretation of the result upside down: only goods that provide no first-order information are not taxed.

Corollary 3 *The optimal wedge on good i is zero if the marginal rate of substitution for x_i does not directly depend on any characteristic.*

Proof. If the marginal rate of substitution, s_i, is independent of all characteristics n_j, then $\varepsilon_{x_i n_j} = 0 \quad \forall x_i, n_j$, such that all A_{ij} are zero and the optimal wedge on x_i is zero by equation (22). ■

In general, corollary 3 indicates that the minimum amount of tax-rates of the optimal tax-system is equal to p. However, it can happen that not all goods are taxed in the second-best allocation. Since we assumed that there are more dimensions of choice than there are dimensions in the type space, all types can be separated out without taxing every choice separately.

The original A-S theorem can easily be derived from corollary 3 by realizing that with a unidimensional type-space the separability assumptions made in A-S imply $\varepsilon_{x_i n_j} = 0$ for all but a single x_i, namely income. The ABC-formula (22) then implies that in an optimal allocation only the income choice is distorted, and hence only income taxes are necessary. Intuitively, if the preference for a choice is not directly influenced by any characteristic, the choice does not provide first order information and distorting it is not efficient.

The NDPF approached the complexity question from a different angle than A-S. Instead of identifying the one necessary rate on a single choice, this literature has identified the necessity of using tax-schedules that depend on a complex interplay between choices made in (all) previous periods of play. Our model does not have any of the dynamic stochastic processes that characterizes the NDPF models. If we assume, however, that the type space consists of two iid variables, like wealth and ability, we have a model that perfectly describes the inter-period problem of Albanesi and Sleet (2006). This way we can get quite close to NDPF models through a multi-dimensional type space alone and isolate the effects of increasing the type-space from the effects dynamic stochastics.

The combination of the two aspects of the type space determines the budget-set of the individual and thus his labor and consumption choices. If the tax-system is used to separate out all types, i.e. separate out individuals in both ability and wealth, we need to need to tax more than one choice. To prevent redistribution toward low income individuals with high wealth, the redistribution will have to be conditioned on both wealth and labor income.

To capture several time periods the definition of the NDPF type has to be expanded. In any game with a finite number of periods P, in any period τ the entire history of an individual can be captured in the vector $\mathbf{n}^P = \{n_1, .., n_\tau, E(n_{\tau+1}), ..E(n_P)\}$. In many cases, the joint distribution of the history $\mathcal{F} = F_1 \circ .. \circ F_P$ can actually be determined in the first period. Unfortunately, as the type description $\mathbf{n}^P$ shows there is a problem with the incentive structure. In our model the entire vector is known to the individual, so that he can plan his entire choice vector $\mathbf{x}$ with full knowledge. In a dynamic stochastic setting much information gets revealed only later, so that there are less (profitable) deviations possible. As Pavan et al. (2010) notes, this implies the incentive constraints are more binding in our model than in the NDPF-models. Expressing those models in terms of our model, should provide a lower bound to the maximized welfare of a NDPF model.

Corollary 3 Does not say anything yet about the relationships between different tax tools. By the same reasoning as before, it seems unlikely that the tax on income and the tax on wealth are unrelated. In fact, we can show the multi-dimensionality of the type-space adds more complexity to the tax-schedule than just more marginal tax-rates. These complexities of the tax-schedule can be shown most straightforwardly if we assume the allocation is very nicely behaved.

Assumption 1 *In the optimal allocation,* $\mathbf{x}^*(\mathbf{n})$*, has a* p*-rank Jacobian for all* $\mathbf{n}$*.*

Corollary 4 *If assumption 1 holds, the optimal wedge on good* i *can be written as a function of* p *choices.*

Proof. If assumption 1 holds, the constant rank theorem allows us to create a local inverted mapping $(x^*)^{\leftarrow} : \mathbf{X} \to \mathbf{N}$ for all $\mathbf{n} \in \mathbf{N}$ such that an inverted mapping exists on the allocation. This inverse mapping must depend on p choices, since the allocation is of rank p everywhere on the type-space. Then we can write the optimal wedge as a function of p choices. ■

If the allocation satisfies assumption 1 it has a reversible mapping $(x^*)^{\leftarrow} : \mathbf{X} \to \mathbf{N}$ and all n_j influence the composition of $\mathbf{x}^*$, i.e. they contain useful information. If all wedges are functions of p choices, then by construction so are all marginal taxes (see proposition 2). In general, therefore, even if the allocation is extremely well behaved (assumption 1 creates a one-to-one relationship with the type-space and a form of monotonicity), a separable tax system does not exist if $p > 1$. If the allocation is less well-behaved we can hardly hope to find simpler tax-schedules.

There are several mathematical reasons for the existence of these interdependencies. First, most probability density functions, $f(\mathbf{n})$, are a function of all n such that equation

(22) depends on all characteristics. More importantly, the partial differential equations (21), which solve for the θ's, are a function of indirect utility. In equilibrium, indirect utility always has to be a function of all $\mathbf{n}$ in order to fulfill the incentive compatibility constraint (16). The solution to the set of partial differential equations in equation (22) therefore depends on all characteristics (it is a p rank system of equations). This will also be the case in the multi-dimensional type-space of the NDPF models, where state-variables stemming from the history of play (like savings in our example) form a natural extension of the type-space. If one interprets the NDPF models as the multi-dimensional screening problems they entail, one directly sees that at least part of the interdependence found in the optimal tax-schedules is due to the multi-dimensional nature of the information possessed by the individuals. To pry this information from the individuals, the marginal tax on labor income will have to depend on wealth (related choices) and vice versa even in a static model.

5.4 On bunching

In most multi-dimensional screening problems the second-best allocation contains bunching at the lower end(s) of the type distribution. Bunching occurs if the solution to the relaxed problem violates the second order conditions (9) on part of the type-space. In this case certain types would prefer the bundle of another type over the one assigned to them. In our examples wealthy, highly able individuals might prefer the low labor output the planner intents for low ability individuals and consume a lot of leisure. In this paper we have so far worked from the relaxed problem that ignores the second-order constraints and thus have avoided bunching.

If bunching occurs, the first-order constraints used in proposition (4) do not hold for the types that bunch. Instead, a solution has to be sought that defines which set of individuals bunches and where they bunch. For all types that separate out the allocation is perfectly described by proposition (4) and the second-order conditions (9) have to be satisfied.

Assume that bunching exists on some set of consumers that has a non-zero measure (in $F(\mathbf{n})$). Denote by $\mathbf{N}_S$ the region of $\mathbf{N}$ where full separation is optimal and by $\mathbf{N}_B$ its complement, the region where bunching is optimal. The next proposition describes the separating region of the type-space.

Proposition 5 *If $\mathbf{N}_S$ exists, it is a single convex set that extends from the upper boundary of the type-space.*

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

The proof of proposition (5) shows that the incentive structure of screening problems in general ensures that if bunching occurs because of the second order incentive constraints, it happens at lower end of the type-space. Individuals on the boundary between the separating and bunching region must be indifferent between selecting a separating and a bunching bundle. While the second-order conditions of (9) guarantee that the utility profile of truthtelling (V) increases at least as much in all characteristics as the utility profile of the bunching (u). In most screening problems, like auctions (Zheng, 2000), non-linear pricing (Armstrong, 1996), and the general class of screening models of Rochet and Choné (1998) (further R-C), it is guaranteed that bunching exists in the optimal solution. The general argument for this bunching is that the planner can extract more money from the high types than from the low types, such that he distorts the allocation for the low-types to prevent mimicking by the high types. R-C show that this causes a problem with outside options, since these limit the possibilities of the planner to distort the lowest allocations. Because our setting does not have an explicit outside option and the welfare function of the planner depends positively on the surplus of the individuals and not on tax revenue, the conclusion that bunching is always optimal cannot be supported by revenue based arguments. Whether or not bunching occurs in a specific tax setting will therefore depend on the specification of the problem. If it occurs, bunching happens at the bottom of the type-space like in all other classes of multi-dimensional screening models.

This pattern of low-types bunching is also what we see in modern social schemes. In the Netherlands for instance, it does not matter why you get welfare, every recipient that is unable to earn a minimum standard has the same basic entitlement. The only deviations from this basic amount are based on observable choices and characteristics. A family with children will get different amounts than a family without children, the chronically ill or disabled can receive some compensation for medical bills, while individuals with large amounts of savings do not get any welfare assistance.

6 Concluding remarks

Although significant progress has been made in multi-dimensional mechanism design, the equilibrium in a multi-dimensional Mirrleesian optimal tax model had so far not been characterized. Simultaneously, the implementation of the second-best allocation has been largely left implicit even in the unidimensional optimal taxation literature. In this paper we characterize both the multi-dimensional second-best and its implementation for a large class of models.

Our model can be used to study the relationship between several tax-tools while taking the multi-dimensional nature of the hidden information and the nature of the tax-tools seriously. Our characterization of the second-best shows that the government should search for consumption patterns that provide as much information on the underlying types as possible. More importantly, the multi-dimensionality in the type-space generally forces the government to make the redistributive taxes depend on several observable choices to separate out different aspects of the hidden types. It might not be optimal to separate out types everywhere in the type-space, in which case some bunching occurs at the lower end of the type-space. This prescription fits reasonably well with the tax-systems observed in welfare states. The lowest earning individuals get welfare assistance, or income subsidies, creating a bunch at the lower end of the income distribution. Most assistance programs are conditioned on (the absence of) wealth, to make sure that no abuse occurs. This is the kind of interdependencies between underlying characteristics (wealth and ability) our model predicts. Many welfare states also subsidies medical expenses or housing for a large group of people. In theory the government could directly transfer the required money, rather than paying part of the price. A direct transfer, however, would make it impossible for the government to find out whether or not you are in need of health-care, i.e. the government could not determine your hidden type through a direct transfer, but can do so through the subsidy. Multi-dimensional hidden information forces the government to use a subsidy that depends on income and expenses (or other observables) rather than direct transfers to create differentiated assistance.

Since the equilibrium in our model depends on solving a set of partial differential equations for which no general solution exists, we can only characterize the equilibrium trough a set of necessary conditions. These conditions strongly limit the possible outcomes, but can never give a full description of the second-best. The next step in this line of research clearly is to find specific, realistic and relevant settings and simulate the model. This is, however, a difficult step. The multi-dimensional heterogeneity sets strong requirements on the optimization algorithms. While the problem of implementation might add further difficulties. Propositions 2 and 3 show that a relatively simple tax-schedule, a (separable) pure price mechanism, can implement the allocation found in most problems studied in the existing literature, as long as the equilibrium is described by the first order conditions. If bunching occurs these conditions no longer describe the second-best for the types that do not separate themselves out, such that a different set of tax-tools (rules) is required. However, once these difficulties have been overcome the model presented in this paper can be used to provide a more precise insight in the optimal relation between the income tax-system and the myriad of social schemes like healthcare subsidies, housing subsidies, and welfare that characterize modern welfare states.

Proposition 2 highlights a unique feature of the Mirrleesian optimal tax model. Unlike the design problem of auctioneers and monopolists, the maximization of the central planner is quite closely aligned with that of the agents he faces. A relatively broad class of tax systems implements the second-best of a welfarist planner because the objective of the planner is strictly increasing in agent's utility. This means the planner can let the agents maximize their utility with relatively little restrictions, irrespective of the actual utility function of individuals. This alignment also has interesting effects on the restrictions that are required for implementation. Necessary restrictions will often contain interdependencies, increasing the complexity of the tax-schedule relatively quickly. These interdependencies, like wealth tests on income assistance in welfare states, are necessary to prevent rational people from taking advantage of subsidies that are not targeted at them.

In fields such as monopoly pricing and auction theory the objectives of the principal and the agents are opposed. An increase in a monopolist's profits (at fixed quantities) automatically comes at the expense of the consumers. As such, implementation will generally not be possible through a pure price mechanism, consequently bunching, prohibitions, restrictions on quantity or more generally restrictions in the choice space will be quite prevalent (see also Armstrong (1996), Renes (2011), Rochet and Choné (1998)).

Since the model presented in this paper contains the multi-dimensional type-space that is also found in the NDPF, it could also provide a convenient middle ground between the complex stochastic dynamics in these models, and the known intuitions in the classical Mirrlees model. The problems of joint (or double) deviation and the interdependencies in the optimal wedges that plague NDPF models are, for instance, also prevalent in our setting, but can be traced much more conveniently. Our results indicate at least part of the difficulties in the NDPF literature are due to the structure of hidden information. We might therefore, be able to gain some intuition for the complicated NDPF tax-schedules from multi-dimensional screening models in general, and our model in particular. In fact, the discussion in section 5.3 already suggests that our findings might be generalized to dynamic stochastic settings. This would allow an elasticity approach and a new focus on implementation in these models as well.

References

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

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 Pierre Pestieau (2011) 'Indirect taxes for redistribution: Should necessity goods be favored?' CESifo Working Paper Series Munich

Cattell, D.W., P.A. Bowen, and A.P. Kaka (2007) 'Review of unbalanced bidding models in construction.' *Journal of construction engineering and management* 133(8), 562–573

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

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

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* pp. 569–581

Golosov, M., M. Troshkin, and A. Tsyvinski (2011) 'Optimal dynamic taxes.' Technical Report, National Bureau of Economic Research

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)

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

Hellwig, Martin F. (2007) 'A contribution to the theory of optimal utilitarian income taxation.' *Journal of Public Economics* 91(7), 1449–1477

_ (2010a) 'A generalization of the atkinson–stiglitz (1976) theorem on the undesirability of nonuniform excise taxation.' *Economics Letters* 108(2), 156–158

_ (2010b) 'Incentive problems with unidimensional hidden characteristics: A unified approach.' *Econometrica* 78(4), 1201–1237

Jacquet, Laurence, Etienne Lehmann, and Bruno Van Der Linden (forthcoming) 'The optimal marginal tax rates with both extensive and intensive responses.' *Journal of Economic Theory*

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)

Kopczuk, Wojciech (2013) 'Incentive effects of inheritances and optimal estate taxation.' Technical Report, National Bureau of Economic Research

Laroque, Guy R. (2005) 'Indirect taxation is superfluous under separability and taste homogeneity: A simple proof.' *Economics Letters* 87(1), 141–144

Lee, Chul-in (2011) 'Commodity taxation in welfare states.' *Economic Inquiry* 49(1), 194–211

Lewis, Tracy R, and David EM Sappington (1988) 'Regulating a monopolist with unknown demand and cost functions.' *RAND Journal of Economics* pp. 438–457

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

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

Pass, Brendan (2012) 'Convexity and multi-dimensional screening for spaces with different dimensions.' *Journal of Economic Theory* 147(6), 2399–2418

Pavan, Alessandro, Ilya Segal, and Juuso Toikka (2010) 'Dynamic mechanism design: Incentive compatibility, profit maximization and information disclosure.' SSRN

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

Renes, Sander, and Floris T. Zoutman (2013a) 'As easy as abc? multi-dimensional screening in public finance.' Rotterdam. mimeo

_ (2013b) 'When a price is enough, a toolbox for optimal tax-design.' Rotterdam. mimeo

Roberts, Kevin (1984) 'The theoretical limits to redistribution.' *Review of Economic Studies* 51(2), 177–195

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

Sadka, Efraim (1976) 'On income distribution, incentive effects and optimal income taxation.' *Review of Economic Studies* 43(2), 261–267

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

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

Wilson, Robert (1996) 'Nonlinear pricing and mechanism design.' *Handbook of computational economics* 1, 253–293

Zheng, Charles Zhoucheng (2000) *Optimal auction in a multidimensional world,* vol. 1282 (Center for Mathematical Studies in Economics and Management Science, Northwestern University)

A Appendix

As before, our bookkeeping definitions: 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 second derivatives $\phi''(\cdot)$ are 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} are of dimension $a^i \times 1$ and the matrix of second derivatives $\psi_{z^i z^j}$ are 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-face, scalars are in normal-face.

A.1 Proof of proposition 1

Individuals maximize a twice differentiable utility function

$$u(\mathbf{x}, y, \mathbf{n})$$

By sending a messages $m \in \mathbf{N}$ to the social planner, who will assign each individual a bundle $\{\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m})\}$ from the allocation $\{\mathbf{x} = \mathbf{x}^*(\mathbf{n}), y = y^*(\mathbf{n})\}$ $\forall \mathbf{n} \in \mathbf{N}$.

Proof. The first order condition for incentive compatibility is given by:

$$\begin{aligned} \mathbf{0}_p &= \frac{\partial u(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n})}{\partial \mathbf{m}}|_{\mathbf{m}=\mathbf{n}}, \\ &= \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{x}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) + u_y(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) y^{*\prime}(\mathbf{n})^T, \end{aligned} \tag{25}$$

where $\mathbf{0}_p$ denotes a p-column vector of zeros. This can be rewritten to:

$$y^{*\prime}(\mathbf{n}) = s(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n})^T \mathbf{x}^{*\prime}(\mathbf{n}).$$

Proofing equation (6)

We can derive (8) from (5) using the envelope theorem:

$$\begin{aligned} V'(\mathbf{n}) &= \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{x}} + u_y y^*(\mathbf{n})^T + u_{\mathbf{n}}^T, \\ V'(\mathbf{n}) &= u_{\mathbf{n}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n})^T, \end{aligned} \tag{26}$$

where the latter equality follows from the first-order conditions.

The second-order conditions of a maximum are :

$$\frac{\partial^2 u(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n})}{\partial \mathbf{m}^2}|_{\mathbf{m}=\mathbf{n}} \lessdot 0, \tag{27}$$

Where $\lessdot 0$ denotes the negative definiteness of the matrix.
Taking the derivative of (26) toward $\mathbf{m}$ gives:

$$\begin{aligned} \frac{\partial^2 u(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n})}{\partial \mathbf{m}^2} &= \left(u_{\mathbf{x}}(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n})^T \otimes I_p\right) \mathbf{x}^{*\prime\prime}(\mathbf{m}) \\ &\quad + \mathbf{x}^{*\prime}(\mathbf{m})^T u_{\mathbf{xx}}(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}) \mathbf{x}^{*\prime}(\mathbf{m}) \\ &\quad + u_{yy}(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}) y^{*\prime}(\mathbf{m}) y^{*\prime}(\mathbf{m})^T \\ &\quad + u_y(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}) y^{*\prime\prime}(\mathbf{m}). \end{aligned} \tag{28}$$

where $\otimes$ denotes the Kronecker product.

To simplify this expression we take the total derivative of the first order condition (25):

$$D_{\mathbf{n}} \mathbf{0}_p = D_{\mathbf{n}} \left[\mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{x}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) + u_y(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) y^{*\prime}(\mathbf{n})^T\right] \tag{29}$$

$$\begin{aligned} \mathbf{0}_{p \times p} = &\left(u_{\mathbf{x}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n})^T \otimes I_p\right) \mathbf{x}^{*\prime\prime}(\mathbf{n})^T + \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{xx}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \mathbf{x}^{*\prime}(\mathbf{n}) + \\ &u_{yy}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) y^{*\prime}(\mathbf{n}) y^{*\prime}(\mathbf{n})^T + u_y(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) y^{*\prime\prime}(\mathbf{n}) + \\ &\mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{xn}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) + y^{*\prime}(\mathbf{n})^T u_{y\mathbf{n}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}), \end{aligned} \tag{30}$$

Now combine equations (29,27,28) to get the following expression:

$$\begin{aligned} 0 &\gtrdot \left(u_{\mathbf{x}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n})^T \otimes I_p\right) \mathbf{x}^{*\prime\prime}(\mathbf{n})^T + \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{xx}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \mathbf{x}^{*\prime}(\mathbf{n}) + \\ &\quad u_{yy}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) y^{*\prime}(\mathbf{n}) y^{*\prime}(\mathbf{n})^T + u_y(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) y^{*\prime\prime}(\mathbf{n}) - \mathbf{0}_{p \times p} \\ 0 &\gtrdot \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{xn}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) + y^{*\prime}(\mathbf{n})^T u_{y\mathbf{n}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}). \end{aligned} \tag{31}$$

Then partially differentiate the vector of shadow prices with respect to $\mathbf{n}$ to get:

$$\begin{aligned} s_{\mathbf{n}} &= \frac{-u_{\mathbf{xn}}u_y + u_{\mathbf{x}}u_{y\mathbf{n}}}{(u_y)^2} \\ u_{\mathbf{xn}} &= -s_{\mathbf{n}}u_y - su_{y\mathbf{n}}, \end{aligned}$$

and substitute this result and (6) into (31) to yield

$$\begin{aligned} 0 &\gtrdot \mathbf{x}^{*\prime}(\mathbf{n})^T(-s_{\mathbf{n}}u_y - su_{y\mathbf{n}}) + y^{*\prime}(\mathbf{n})^T u_{y\mathbf{n}} \\ 0 &\lessdot \mathbf{x}^{*\prime}(\mathbf{n})^T s_{\mathbf{n}}u_y \Leftrightarrow \\ 0 &\lessdot \mathbf{x}^{*\prime}(\mathbf{n})^T s_{\mathbf{n}}, \end{aligned}$$

where the final inequality, equation (7), follows from the fact that $u_y > 0$.

An equivalent expression can be derived by totally differentiating, equation (26) with respect to $\mathbf{n}$:

$$\begin{aligned} V''(\mathbf{n}) &= Du_{\mathbf{n}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \\ &= \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{xn}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) + y^{*\prime}(\mathbf{n})^T u_{y\mathbf{n}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \\ &\quad + u_{\mathbf{nn}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \end{aligned}$$

Now combine this last expression with (31) to get the final equation:

$$V''(\mathbf{n}) - u_{\mathbf{nn}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \gtrdot 0.$$

■

A.2 Example

In figure 1 and 2, welfare function and resource constraint are equal to:

$$\begin{aligned} u &= log(y) - \frac{1}{1.5}\left(\frac{x_1}{n}\right)^{1.5} - \frac{1}{1.5}\left(\frac{x_2}{n}\right)^{1.5} \\ W &= \int [u(n) + E(x_1.x_2)]\, dF(n) \\ E(x_1, x_2) &= \frac{1}{1.5}\left(\left(\frac{x_1(n)}{n}\right)^{1.5} + \left(\frac{x_2(n)}{n}\right)^{1.5} - \left(\frac{x_1(n)}{n}\right)^{1.5} * \left(\frac{x_2(n)}{n}\right)^{1.5}\right) \\ \int_N y(n)dF(n) &= \int_N x_1(n) + x_2(n)dF(n) \end{aligned}$$

We assume that the type-space is unidimensional and the types are uniformly distributed over a closed interval on the real line. The first-order approach to this problem yields the allocation shown in the figures 1 2. This second-best allocation can only be implemented by the central planner if he uses interdependencies to map out the off-allocation consumption choices/coordinates. The planner can than determine what off-allocation points have to be taxed prohibitively to ensure that each individual prefers his own bundle over any other choice.

A.3 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:

$$\mathbf{0} = u_{\mathbf{x}} + (q' - T')^T u_y \tag{32}$$

which directly implies equations (12) and (13).

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

$$u_{\mathbf{xx}} + \left(2u_{\mathbf{x}y} + u_{yy}\left(q'(\mathbf{x}^*) - T'(\mathbf{x}^*)\right)^T\right)\left(q'(\mathbf{x}^*) - T'(\mathbf{x}^*)\right) + u_y\left(q''(\mathbf{x}^*) - T''(\mathbf{x}^*)\right) \lessdot 0 \tag{33}$$

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

$$\frac{\partial \mathbf{s}(\mathbf{x}, y(u, \mathbf{x}, \mathbf{n}), \mathbf{n})}{\partial \mathbf{x}} = -\frac{\left(u_{\mathbf{xx}} + 2u_{\mathbf{x}y}\mathbf{s}^T\right) - u_{yy}\mathbf{s}\mathbf{s}^T}{u_y} \tag{34}$$

Now combining (13) with (34) allows us to simplify (33) and obtain the final condition:

$$\begin{aligned} -\left(\frac{\partial \mathbf{s}(\mathbf{x}, y(u, \mathbf{x}, \mathbf{n}), \mathbf{n})}{\partial \mathbf{x}} + q''(\mathbf{x}^*) - T''(\mathbf{x}^*)\right) u_y &\lessdot 0 \Leftrightarrow \\ -\frac{\partial \mathbf{s}(\mathbf{x}, y(u, \mathbf{x}, \mathbf{n}), \mathbf{n})}{\partial \mathbf{x}} + q''(\mathbf{x}^*) - T''(\mathbf{x}^*) &\lessdot 0 \end{aligned}$$

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

A.4 Proof to proposition 2

Suppose on the contrary that (14) is not satisfied for some agent of type $\mathbf{n}$. Consider a deviation from the second-best allocation $\alpha\Delta\mathbf{x}$ where $\alpha > 0$ and $\Delta\mathbf{x}$ is a $k \times 1$ vector with length one. The utility gain of such a deviation can be approximated by a second-order Taylor expansion:

$$\begin{aligned} u(\mathbf{x}^*(\mathbf{n}) + \alpha\boldsymbol{\Delta}\mathbf{x}, q(\mathbf{x}^*(\mathbf{n}) + \alpha\Delta\mathbf{x}) - T(\mathbf{x}^*(\mathbf{n}) + \alpha\Delta\mathbf{x}), \mathbf{n}) - u^* &= \\ \left(u_{\mathbf{x}}^T + u_y(q' - T')\right)\alpha\boldsymbol{\Delta}\mathbf{x} &+ \\ \frac{1}{2}\alpha^2\boldsymbol{\Delta}\mathbf{x}^T\left(u_{\mathbf{xx}} + \left(2u_{\mathbf{x}y} + u_{yy}(q' - T')^T\right)(q' - T') + u_y(q'' - T'')\right)\boldsymbol{\Delta}\mathbf{x} &= \\ \frac{1}{2}\alpha^2\boldsymbol{\Delta}\mathbf{x}^T\left(u_{\mathbf{xx}} + \left(2u_{\mathbf{x}y} + u_{yy}(q' - T')^T\right)(q' - T') + u_y(q'' - T'')\right)\boldsymbol{\Delta}\mathbf{x} &= \\ \frac{1}{2}\alpha^2 u_y\boldsymbol{\Delta}\mathbf{x}^T\left(-\frac{\partial \mathbf{s}(\mathbf{x}, y(u, \mathbf{x}, \mathbf{n}), \mathbf{n})}{\partial \mathbf{x}} + q''(\mathbf{x}^*) - T''(\mathbf{x}^*)\right)\boldsymbol{\Delta}\mathbf{x} & \end{aligned}$$

where the first order terms equal zero by the first-order condition (13). Due to symmetry of the matrix of second order conditions for sufficiently small α the deviation strategy $\alpha\Delta\mathbf{x}$

and $-\alpha\Delta\mathbf{x}$ yield approximately the same utility. In addition, if $\left(-\frac{\partial \mathbf{s}(\mathbf{x},y(u,\mathbf{x},\mathbf{n}),\mathbf{n})}{\partial \mathbf{x}}+q''\left(\mathbf{x}^*\right)-T''\left(\mathbf{x}^*\right)\right)$ is not negative semi-definite there is at least one deviation strategy $\Delta\hat{\mathbf{x}}$ which yields a positive utility gain. The change in tax revenue due to such a deviation can also be found by means of a second-order Taylor expansion:

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

The first-order term will always be non-negative for either strategy $-\Delta\hat{\mathbf{x}}$ or $\Delta\hat{\mathbf{x}}$. If for either choice it's positive, the first-order term dominates the second-order term for sufficiently small α and hence, the deviation results in higher tax revenue. If the first-order term is zero, we need to consider the second-order term. If it's negative apparently the tax schedule contains an internal maximum on the allocation in $\Delta\hat{\mathbf{x}}$ which violates our assumption. Therefore, if the first-order term is zero the second term must be non-negative. Hence, tax revenue always weakly increases in either $-\Delta\hat{\mathbf{x}}$ or $\Delta\hat{\mathbf{x}}$. Therefore, one of these deviations must be a Pareto-improvement and we run into a contradiction. If a Pareto-improvement over the allocation can be found within a particular implementation, then the original allocation could not have been second-best.

A.5 Proof to proposition 3

Equations (12) and (13) uniquely define the tax schedule for $\mathbf{x}^*\left(\mathbf{n}\right)$ on its domain $\mathbf{N}$. If $\mathbf{x}^*\left(\mathbf{n}\right)$ is bijective there is an unique inverse mapping $\mathbf{n}^*\left(\mathbf{x}\right)$ for all $\mathbf{x}\in\mathbf{X}$. Therefore, equations (12) and (13) define the tax schedule for $\mathbf{x}^*\left(\mathbf{n}^*\left(\mathbf{x}\right)\right)$ on its domain $\mathbf{x}\in\mathbf{X}$. Hence, the tax schedule is defined on the entire choice space. Note that we do not need to check for second-order conditions (14) in this case, because we have assumed that the allocation $\mathbf{x}^*\left(\mathbf{n}\right)$ is (second-order) incentive compatible for all $\mathbf{n}\in\mathbf{N}$. Therefore, the unique tax schedule that implements this allocation must also be implementable.

A.6 proof of proposition 4

Proof. Starting from the first order conditions (20), (19) and (21). First solving (19) for yields η:

$$\eta=\frac{\lambda f+u_{y\mathbf{n}}\theta}{u_y}.$$

Now substitute this expression into (20) and simplify to get the desired equation:

$$\begin{aligned}
\lambda q'^T f-u_{\mathbf{xn}}\theta+\frac{\lambda f+u_{y\mathbf{n}}\theta}{u_y}u_{\mathbf{x}} &= \mathbf{0}_k,\\
\lambda q_{x_i} f-u_{x_i\mathbf{n}}\theta+\frac{\lambda f+u_{y\mathbf{n}}\theta}{u_y}u_{x_i} &= \mathbf{0}_k,\\
\lambda f\left(q_{x_i}-s_i\right) &= u_{x_i\mathbf{n}}\theta+u_{y\mathbf{n}}\theta\mathbf{s},\\
q_{x_i}-s_i &= \left(u_{x_i\mathbf{n}}+u_{y\mathbf{n}}s_i\right)\frac{\theta}{\lambda f},\\
q_{x_i}-s_i &= \sum_{j=1}^{p}-\frac{\partial s_i}{\partial n_j}\theta_j\frac{u_y}{\lambda f},\\
\frac{q_{x_i}-s_i}{s_i} &= \sum_{j=1}^{p}\varepsilon_{x_i n_j}\times\theta_j\frac{u_y}{\lambda}\times\frac{1}{n_j f}.
\end{aligned}$$

■

A.7 Proof of proposition 5

The proof of the proposition is done in two steps. First we show the separating region is convex and then we show that it is unique and at the top of the type-space.

Lemma 2 *The separating region $\mathbf{N}_S$ is convex or empty.*

Proof. Define the types α, β, γ s.t. $\alpha, \gamma \in \mathbf{N}_S$ and $\exists k \in (0,1)$ s.t. $(1-k)\,\alpha + k\gamma = \beta$. Denote by $\{\widetilde{x}, \widetilde{y}\}$ the solution to the full problem, and by $u(\{\overline{x}, \overline{y}\}, \mathbf{n}) = \max_{\{x,y\}} u(x, y, \mathbf{n} | \{\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}\}^{\leftarrow} \subseteq \mathbf{N}_B)$ the bunching choice that delivers type $\mathbf{n}$ the highest utility from the set of bunching allocations. Recall that $\frac{\partial u(\mathbf{x}, y, \mathbf{n})}{\partial n_j} > 0 \quad \forall n_j \in \mathbf{N}$.
Then define the function $L(\mathbf{n}) = u(\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}, \mathbf{n}) - u(\{\overline{x}, \overline{y}\}, \mathbf{n})$. By individual rationality we know that $0 < L(\alpha)$,$L(\gamma)$**.** Equation (9) implies that L is convex, and continuity of u implies L is piece-wise continuous. For any $0 < k < 1$ it must be then be the case that $0 < L(\ \beta)$, and β must also be part of the separating set. ■

Lemma 3 *Bunching occurs below the boundary $\partial N_S \cap \partial N_B$, and only one such boundary exists.*

Proof. Since $V(\mathbf{n})$ is continuous, individual rationality requires any type at $\mathbf{b} \in \partial N_S \cap \partial N_B$, $u(\{\mathbf{x}^*(\mathbf{b}), y^*(\mathbf{b})\}, \mathbf{b}) = u(\{\overline{x}, \overline{y}\}, \mathbf{b})$. While on the separating set equation (8) guarantees that the first derivative of u has to be equal to the first-order derivative of v. By equation (9), however, the second-order derivative on the optimal allocation has to be higher than the second derivative of the utility function on $\{\overline{x}, \overline{y}\}$, such that the utility profile of the separating and the bunching region cross only once for each type, and they cross at $\{\mathbf{x}^*(\mathbf{b}), y^*(\mathbf{b})\}$ for type **b.** For any type $\mathbf{g}$, with $b_j \leq g_j\ j \in \{1, .., p\}$ and at least one inequality strict, the utility profile of the optimal allocation, $V(\mathbf{n})$, has to be higher than $u(\{\overline{x}, \overline{y}\}, \mathbf{g})$. For any type $\mathbf{g}$ therefore (8) holds and the allocation found above the boundary $\partial N_S \cap \partial N_B$ induces separation. Simultaneously, below the boundary equation (9) cannot hold, which (together with (8) and continuity of V) implies that bunching yields a higher utility, such that bunching occurs there. ■

Together lemmas (2) and (3) imply the result.