Incentive Compatibility in Multidimensional Screening*

William Dodds†

September 5, 2022

Abstract

This paper derives necessary and sufficient conditions for allocations to be incentive compatible in general multidimensional screening models that satisfy a generalized single crossing property. We then derive a numerical method based on these results to solve multidimensional screening problems. We apply this method to several numerical examples in the context of multidimensional optimal taxation. In addition to illustrating how to apply our theoretical results and our numerical method, our numerical simulations highlight the importance of bunching in optimal multidimensional taxation.

Keywords: multidimensional screening, bunching, incentive compatibility, multidimensional taxation

JEL: D82, D86, H21

*The findings, interpretations, and conclusions expressed in this paper are entirely those of the author. Thanks to Stefano Barbier for providing helpful comments on this paper. This paper also benefited significantly from conversations with Ilia Krasikov and Mike Golosov. Finally, a special thanks to Katy Bergstrom who has been immensely helpful in providing feedback on this project.

†Tulane University. Email: wdodds@tulane.edu

1 Introduction

Screening problems are extremely common in economics, appearing in a wide variety of seemingly unrelated settings: nonlinear pricing (e.g., Mussa and Rosen (1978) or Armstrong (1996)), optimal taxation (e.g., Mirrlees (1971) or Saez (2001)), public procurement (e.g., Laffont and Tirole (1994)), and regulation of monopolies (e.g., Baron and Myerson (1982)). Because all of these problems can ultimately be studied within a single unifying framework of optimal screening, understanding properties of general screening problems has been a priority of the theoretical economics literature over the past 50 years.

The majority of work in the screening literature has considered the unidimensional case: agents differ on one dimension and have a single choice variable. The key result from this literature is that if preferences satisfy the single crossing property, then local incentive compatibility is necessary and sufficient for global incentive compatibility (e.g., Mirrlees (1971)). Moreover, second order conditions are typically not binding for most type spaces, which rules out bunching and allows the use of first order optimization methods to solve for the optimal mechanism. The goal of this paper is to explore the extent to which these results carry over to a general multidimensional setting in which a principal screens agents who differ on many dimensions and have many choice variables.

This paper contributes to the screening literature in two ways. First and foremost, we derive a number of results characterizing incentive compatibility in general multidimensional screening problems assuming preferences satisfy a "generalized single crossing property". First, Theorem 1 shows that incentive compatibility necessitates first order conditions to hold almost everywhere, second order conditions to hold whenever the allocation is sufficiently smooth (precisely, when the transfer schedule is differentiable and the mapping from types to actions is locally diffeomorphic), and the mapping from types to actions to be globally injective whenever it is sufficiently smooth. Second, Theorem 2 shows that if we restrict ourselves to sufficiently smooth allocations, we can derive a general necessary and sufficient condition for incentive compatibility in terms of individual first order conditions, a global injectivity condition, and a condition requiring all individuals prefer their assigned bundle to bundles chosen by boundary individuals. Moreover, we discuss how this result extends to non-smooth allocations via a limiting argument. Taken together, our results allow us to determine whether many allocations are incentive compatible in general multidimensional settings.

The second contribution of this paper is to devise a method that applies our incentive compatibility results to numerically solve multidimensional screening problems. The core idea is to maximize the objective function for a general multidimensional screening problem over the set of allocations satisfying our *necessary conditions*: first order conditions, second order conditions, and a constraint that the allocation is sufficiently smooth (i.e., locally diffeomorphic). We show that this approach boils down to an optimization problem with linear and non-linear equality and inequality constraints, which appears to be computationally tractable. Once we have a proposed solution from this procedure, we can check whether it satisfies our *sufficient conditions* for incentive compatibility; if so, then we have found the optimal schedule within the class of smooth allocations. Of course, one may question the endogenous assumption that the optimal allocation is smooth. While this is typically innocuous in unidimensional problems (as second order conditions are not binding), as illustrated in Rochet and Chone (1998) for the multiproduct monopolist setting, second order conditions often do bind in multidimensional settings, which generates bunching. Fortunately, our numerical method based on Theorems 1 and 2 is capable of handling situations in which the optimal allocation features bunching so long as the optimal (non-smooth) allocation can be approximated to arbitrary precision by smooth allocations. This turns out to be quite important as we illustrate solutions that feature bunching for a number of toy examples as well as a more realistic, calibrated exercise exploring optimal taxation of couples using data from the Current Population Survey.

Within the literature on multidimensional screening, this paper is most closely related to three papers: Rochet (1987), McAfee and McMillan (1988), and Carlier (2001), all of which characterize incentive compatible allocations in various multidimensional settings. Rochet (1987) proves that incentive compatibility is equivalent to an envelope condition and a convexity condition on utility provided that utility is separable in the transfer and all choice variables, linear in the transfer, and linear in type. The key contribution that we make relative to Rochet (1987) is that our results apply to a much wider class of utility functions: we only require that the utility function satisfy our generalized single crossing property, which does not require utility to be separable, linear in the transfer, or linear in type. McAfee and McMillan (1988) show that for the case of smooth allocations (i.e., assuming away bunching), incentive compatibility is equivalent to first and second order conditions under their own "generalized single crossing property", which differs from our generalized single crossing property. In particular, the generalized single crossing property discussed in McAfee and McMillan (1988) is much more stringent than our generalized single crossing property: it is difficult to identify any realistic utility function which satisfies the generalized single crossing property of McAfee and McMillan (1988) other than utility functions which are linear in type. Our results contribute relative to McAfee and McMillan (1988) primarily because our generalized single crossing condition holds for a much wider class of utility functions and secondarily because our results extend to allocations that feature bunching. Carlier (2001) generalizes Rochet (1987) by characterizing incentive compatibility when utility is not necessarily linear in type, but is separable and quasi-linear in consumption, via the concept of h-convexity. In contrast, our results do not require separability of the utility function, but do require a generalized single crossing condition. Secondly, our results are additive above and beyond Carlier (2001) because h-convexity constraints are inherently difficult to work with numerically because they are expressed in terms of global (rather than local) properties of the allocation. In contrast, our necessary conditions can be verified using local properties of the allocation, which enables easier use for numerically solving multidimensional screening problems.

The second portion of this paper develops a numerical methodology to solve multidimensional screening problems and then applies this methodology to multidimensional optimal taxation. While the core contribution here is developing a general numerical method to solve multidimensional screening problems, this paper is also related to a growing literature on taxation with multidimensional heterogeneity, including Mirrlees (1976), Mirrlees (1986), Kleven, Kreiner and Saez (2009), Chone and Laroque (2010), Jacquet and Lehmann (2015), Scheuer and Werning (2016), Jacquet and Lehmann (2020), Bergstrom and Dodds (2021), Spiritus et al. (2022), Boerma, Tsyvinski and Zimin (2022), and Krasikov and Golosov (2022). Many of these papers deal with settings in which agents have multiple dimensions of heterogeneity, but only one choice variable, which is a much simpler class of problems. On the other hand, Mirrlees (1976), Mirrlees (1986), Kleven, Kreiner and Saez (2009), Spiritus et al. (2022), Boerma, Tsyvinski and Zimin (2022), and Krasikov and Golosov (2022) all consider settings in which agents have multidimensional types and multidimensional choice sets. With the exception of Boerma, Tsyvinski and Zimin (2022), all of these papers assume that bunching does not occur, which allows them to use first order conditions to solve for the optimal schedule. Boerma, Tsyvinski and Zimin (2022) uses the utility function from Rochet (1987) to explore a model with multidimensional skill types and participation constraints, finding that bunching is empirically relevant. This is in line with Rochet and Chone (1998), who use the utility function from Rochet (1987) to argue that bunching is a common occurrence for a multiproduct monopolist's problem due to a tension between participation constraints and second order conditions. In our calibrated application of optimal couples taxation, we also find that bunching is empirically relevant, especially at the bottom of the income distribution. The findings from our numerical simulations are additive in that they highlight the relevance of bunching in a setting *without* participation constraints and without the linear, separable utility function of Rochet (1987).

The rest of this paper proceeds as follows. Section 2 presents the environment, introduces relevant notation, and discusses useful mathematical preliminaries. Section 3 states our main results on incentive compatibility and provides intuition by relating our results to existing results in the literature. Section 4 introduces the numerical method we use to solve multidimensional screening problems. Section 5 illustrates this method using a number of numerical examples related to optimal multidimensional taxation. Section 6 concludes.

2 Environment, Notation, Mathematical Preliminaries

In this section we present the environment and relevant notation, introduce our generalized single crossing property, and discuss a few mathematical preliminaries.

2.1 The Environment

The model consists of a population of individuals, indexed by type $\mathbf{n} = (n_1, n_2, ..., n_K) \in \mathbf{N} = N_1 \times N_2 \cdots \times N_K$. We assume that the boundary of $\mathbf{N}$, denoted $\partial\mathbf{N}$, is smooth. We assume that the distribution of types, denoted $F(n_1, n_2, ..., n_K)$ is continuously differentiable with density $f(n_1, n_2, ..., n_K)$. An individual's type $\mathbf{n}$ is private information. Individuals have preferences over a monetary transfer from the screening entity, $T(\mathbf{z})$, as well as K choice variables $\mathbf{z} = (z_1, z_2, ..., z_K) \in \mathbf{Z}$, which are all observable by the screening entity.[1] Individuals solve the following maximization problem:

$$\max_{\mathbf{z}} \; u(T(\mathbf{z}), \mathbf{z}; \mathbf{n}) \tag{1}$$

where $u(T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ is a smooth utility function satisfying $u_T > 0$ and $u_{TT} \leq 0$.

In general multidimensional screening problems (such as non-linear pricing, optimal taxation with multidimensional instruments, or regulation of multiproduct monopolies), some entity (e.g., a government or a firm) contracts with individuals by choosing a transfer (or tax) function $T(\mathbf{z})$ which gives individuals a monetary payoff for making choices $\mathbf{z}$. The entity seeks to maximize some objective function subject to constraints conditional on all individuals optimizing their own utility. Equivalently, due to the revelation principle, this entity can consider choosing allocations $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ for each type $\mathbf{n}$ subject to the constraint that the chosen allocation is incentive compatible in the sense that for all types $\mathbf{n}$:[2]

$$u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = \max_{\mathbf{n}'} \; u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$$

Hence, understanding which allocations are incentive compatible is a necessary precursor to characterizing optimality. Our first goal then is to understand which allocations $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ are incentive compatible.

[1] To begin, we assume the dimension of the choice set is equal to the dimension of the type space; however, we discuss later how our results can be extended to cases where these dimensions differ.

[2] Restricting $T(\cdot)$ to be a function of $\mathbf{z}(\mathbf{n})$ rather than $\mathbf{n}$ is WLOG as any allocation featuring $\mathbf{z}(\mathbf{n}) = \mathbf{z}(\mathbf{n}')$ and $T(\mathbf{n}) \neq T(\mathbf{n}')$ for $\mathbf{n} \neq \mathbf{n}'$ is clearly not incentive compatible (as one of the individuals $\mathbf{n}, \mathbf{n}'$ can improve utility by pretending to be the other type, yielding the same $\mathbf{z}$ and higher T, which improves utility as $u_T > 0$).

2.1.1 A Note on Notation

The arguments of the various gradients throughout the paper can be somewhat cumbersome. Subscripts denote partial derivatives with respect to a single variable, for example:

$$u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \equiv \left.\frac{\partial u(T, \mathbf{z}(\mathbf{n}); \mathbf{n})}{\partial T}\right|_{T=T(\mathbf{z}(\mathbf{n}))}$$

$$u_{n_1}(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \equiv \left.\frac{\partial u(T, \mathbf{z}; \mathbf{n})}{\partial n_1}\right|_{T=T(\mathbf{z}(\mathbf{n})), \mathbf{z}=\mathbf{z}(\mathbf{n})}$$

$\nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ denotes the partial derivative (gradient) with respect to the vector $\mathbf{z}$:

$$\nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \equiv \nabla_{\mathbf{x}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{x}; \mathbf{n})|_{\mathbf{x}=\mathbf{z}(\mathbf{n})} = \begin{bmatrix} u_{z_1}(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \\ \vdots \\ u_{z_K}(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \end{bmatrix}$$

Similarly, if we write $\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$, this denotes:

$$\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = \begin{bmatrix} u_{n_1}(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \\ \vdots \\ u_{n_K}(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \end{bmatrix}$$

In contrast, we use the following notation for total derivatives:

$$D_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$$

$$D_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = \nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) + D_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})$$

2.2 Generalized Single Crossing Property and Technical Preliminaries

Before we discuss our main results on incentive compatibility, we need to explain our "generalized single crossing property" as well as introduce a few pieces of mathematical machinery. In particular, we need to introduce the concept of a *diffeomorphism*, we need to define P matrices, and we need to discuss the multidimensional envelope condition. First, we are going to rely heavily on the concept of a diffeomorphism from differential geometry:

Definition 1. *A diffeomorphism is a continuously differentiable bijective function which also has a continuously differentiable inverse.*[3]

For example, $f(x, y) = (x^3 + x, 2y + 1)$ is a diffeomorphism from $\mathbb{R}^2 \to \mathbb{R}^2$, but $f(x) = x^3$ is not a diffeomorphism from $\mathbb{R} \to \mathbb{R}$ even though this function is bijective because it does not have a differentiable inverse when $x = 0$. We also define *local diffeomorphisms*:

Definition 2. *A function is a local diffeomorphism* ***at a given point*** *if there exists an open set containing that point where the function is differentiable, bijective, and has a continuously differentiable inverse. We refer to a function as a local diffeomorphism if it is a local diffeomorphism at all points.*

[3]Note, some authors define diffeomorphisms to be infinitely differentiable. We follow Hirsch (1988) and Encyclopedia of Mathematics (2022) in only requiring a diffeomorphism to be continuously differentiable.

Equivalently, by the inverse function theorem, a function that is continuously differentiable on an open neighborhood around a point is a local diffeomorphism at that point if and only if its derivative matrix is invertible. It is also useful to point out the well-known result that a local diffeomorphism is a diffeomorphism if and only if it is globally injective.

Next, we are going to restrict attention to a particular class of problems that satisfy what we refer to as the "generalized single crossing property":[4]

Assumption 1 (Generalized Single Crossing Property)**.** *For all $T, \mathbf{z}$, the following function is a diffeomorphism:*

$$\mathbf{n} \mapsto \frac{\nabla_{\mathbf{z}} u(T, \mathbf{z}; \mathbf{n})}{u_T(T, \mathbf{z}; \mathbf{n})} = \begin{bmatrix} \frac{u_{z_1}(T,\mathbf{z};\mathbf{n})}{u_T(T,\mathbf{z};\mathbf{n})} \\ \vdots \\ \frac{u_{z_K}(T,\mathbf{z};\mathbf{n})}{u_T(T,\mathbf{z};\mathbf{n})} \end{bmatrix}$$

Assumption 1 can be viewed as a multidimensional generalization of the standard single crossing property in the unidimensional setting. In the unidimensional setting, the standard single crossing property, which is written entirely in terms of primitives of the model, ensures that under any incentive compatible allocation, if a given n is mapped to a z, then no other type n' is mapped to the same z whenever $T(z(n))$ and $z(n)$ are differentiable.[5] As with the unidimensional single crossing property, Assumption 1 is written entirely in terms of model primitives and ensures that under any incentive compatible allocation, if a given $\mathbf{n}$ is mapped to a $\mathbf{z}$, then no other type $\mathbf{n}'$ is mapped to $\mathbf{z}$ whenever $T(\mathbf{z}(\mathbf{n}))$ is differentiable as a function of $\mathbf{z}$ at a given $\mathbf{n}$ and $\mathbf{z}(\mathbf{n})$ is a local diffeomorphism in a neighborhood around the given $\mathbf{n}$. Under these conditions, we show in Section 3 that the following first order conditions must hold for any incentive compatible allocation:

$$\begin{gathered} u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) T_{z_1}(\mathbf{z}(\mathbf{n})) + u_{z_1}(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = 0 \\ \vdots \\ u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) T_{z_K}(\mathbf{z}(\mathbf{n})) + u_{z_K}(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = 0 \end{gathered}$$

Equivalently, we have:

$$\begin{aligned} \frac{u_{z_1}(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})}{u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})} &= -T_{z_1}(\mathbf{z}(\mathbf{n})) \\ &\vdots \\ \frac{u_{z_K}(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})}{u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})} &= -T_{z_K}(\mathbf{z}(\mathbf{n})) \end{aligned} \tag{2}$$

By Assumption 1, no two types $\mathbf{n}$ and $\mathbf{n}'$ can simultaneously solve Equation 2 for a given $\mathbf{z}$, which shows that the generalized single crossing property ensures that any mapping with $\mathbf{z}(\mathbf{n}) = \mathbf{z}(\mathbf{n}')$ for $\mathbf{n} \neq \mathbf{n}'$, $T(\mathbf{z}(\mathbf{n}))$ differentiable, and $\mathbf{n} \mapsto \mathbf{z}$ a local diffeomorphism at both $\mathbf{n}$ and $\mathbf{n}'$ cannot be incentive compatible.

[4] As far as we know, this sort of "twist" condition has not been used in the optimal screening literature. However, it has been applied in the theory of optimal transport and dynamical systems (e.g., Villani (2009)) and has been used in the optimal multidimensional matching literature (e.g., Chiappori, McCann and Pass (2017)).

[5] In the unidimensional setting, n denotes type and z denotes the single choice variable. The standard single crossing property in the unidimensional setting implies that any incentive compatible allocation features monotonic $z(n)$ and *strictly* monotonic $z(n)$ whenever $T(z(n))$ and $z(n)$ are differentiable, see Lemma 1 from Bergstrom and Dodds (2021).

As an example of a utility function that satisfies Assumption 1 (yet is neither linear in $\mathbf{n}$ nor linear in T, taking us outside the realm of functions considered in Rochet (1987), McAfee and McMillan (1988), or Carlier (2001)), consider the utility function given by:

$$v(z_1 + z_2 + T) - \frac{1}{1+\theta_1}\left(\frac{z_1}{n_1}\right)^{1+\theta_1} - \frac{1}{1+\theta_2}\left(\frac{z_2}{n_2}\right)^{1+\theta_2} - \beta\frac{z_1}{n_1}\frac{z_2}{n_2} \tag{3}$$

for some increasing, concave $v(\cdot)$ and $\theta_1, \theta_2, \beta \geq 0$. This function satisfies Assumption 1 as long as $\mathbf{N}$ is a rectangular domain in $\mathbb{R}^2_{++}$, as a result of Theorem 4 from Gale and Nikaido (1965), which provides a sufficient condition for a mapping to be diffeomorphic:

Remark 1. *A mapping $\mathbf{n} \mapsto \mathbf{z}$ with a continuous P matrix Jacobian on a closed rectangular domain is a diffeomorphism.*[6]

As the concept of a P matrix will come up numerous times throughout this paper, let us define what a P matrix is:

Definition 3. *A P matrix is a square matrix whose principal minors are all positive.*

Returning to utility function 3, we have:

$$\mathbf{n} \mapsto \frac{u_{\mathbf{z}}(T,\mathbf{z};\mathbf{n})}{u_T(T,\mathbf{z};\mathbf{n})} = \begin{bmatrix} \frac{u_{z_1}(T,\mathbf{z};\mathbf{n})}{u_T(T,\mathbf{z};\mathbf{n})} \\ \frac{u_{z_2}(T,\mathbf{z};\mathbf{n})}{u_T(T,\mathbf{z};\mathbf{n})} \end{bmatrix} = \frac{1}{v'(z_1+z_2+T)} \begin{bmatrix} -\frac{z_1^{\theta_1}}{n_1^{1+\theta_1}} - \beta\frac{z_2}{n_1 n_2} \\ -\frac{z_2^{\theta_2}}{n_2^{1+\theta_2}} - \beta\frac{z_1}{n_1 n_2} \end{bmatrix}$$

Taking the partial derivative of the above with respect $\mathbf{n}$ (holding T and $\mathbf{z}$ fixed):

$$\nabla_{\mathbf{n}}\left(\frac{1}{v'(z_1+z_2+T)} \begin{bmatrix} -\frac{z_1^{\theta_1}}{n_1^{1+\theta_1}} - \beta\frac{z_2}{n_1 n_2} \\ -\frac{z_2^{\theta_2}}{n_2^{1+\theta_2}} - \beta\frac{z_1}{n_1 n_2} \end{bmatrix}\right) = \frac{1}{v'(z_1+z_2+T)} \begin{bmatrix} \frac{(1+\theta_1)z_1^{\theta_1}}{n_1^{2+\theta_1}} + \beta\frac{z_2}{n_1^2 n_2} & \beta\frac{z_2}{n_1 n_2^2} \\ \beta\frac{z_1}{n_1^2 n_2} & \frac{(1+\theta_2)z_2^{\theta_2}}{n_2^{2+\theta_2}} + \beta\frac{z_1}{n_1 n_2^2} \end{bmatrix}$$

which is a P matrix, hence diffeomorphic on any rectangular domain $\mathbf{N}$ in $\mathbb{R}^2_{++}$ by Remark 1.[7]

Next, we present the envelope condition. We say that an allocation $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ satisfies the envelope condition if the function $U(\mathbf{n}) \equiv u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ satisfies the following for any $\mathbf{n}_1$ and $\mathbf{n}_2$ and any path between these two points (denoting a path integral by $\oint$):

$$U(\mathbf{n}_1) - U(\mathbf{n}_2) = \oint_{\mathbf{n}_2}^{\mathbf{n}_1} \nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \cdot d\mathbf{n} \tag{4}$$

Finally, on a technical note, we will assume throughout that the function $\mathbf{z}(\mathbf{n}) \in L^2(\mathbf{N})$ and $U(\mathbf{n}) \in H^1(\mathbf{N})$, where $H^1(\mathbf{N})$ is the Sobolev space of functions which have finite L^2 norm and whose (weak) derivatives have finite L^2 norm.[8]

[6]Theorem 4 in Gale and Nikaido (1965) provides a condition for a mapping to be injective; however, it is a standard result that any injective local diffeomorphism is a global diffeomorphism onto its range (the determinant of a P matrix never vanishes which, together with the continuity assumption on the Jacobian, ensures any mapping with a P matrix Jacobian is a local diffeomorphism). As an aside, a slightly different result is Theorem 6 from Gale and Nikaido (1965), which tells us that a mapping $\mathbf{n} \mapsto \mathbf{z}$ with a Jacobian, J, on a convex domain such that $\frac{1}{2}(J + J^T)$ is positive (or negative) definite is also a diffeomorphism.

[7]Assuming we can restrict attention to $z_1, z_2 \geq 0$ and either $z_1 > 0$ or $z_2 > 0$.

[8]This is a standard assumption in mutlidimensional screening, e.g., Rochet and Chone (1998) or Carlier (2001).

3 Main Results

Next we present our main results characterizing incentive compatibility. The core insight is to use the structure provided when the allocation is sufficiently smooth (i.e., $T(\mathbf{z})$ is differentiable and $\mathbf{z}(\mathbf{n})$ is a local diffeomorphism) along with the generalized single crossing property (Assumption 1) to make statements about incentive compatibility. Building towards these ideas, note that when $T(\mathbf{z})$ is differentiable and $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism, the following first order condition must hold for all types $\mathbf{n}$ under any incentive compatible allocation:

$$\{u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})\nabla_{\mathbf{z}}T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}}u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})\} \nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n}) = 0 \tag{5}$$

Or, using the fact that $\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})$ is invertible when $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism:

$$FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) \equiv u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})\nabla_{\mathbf{z}}T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}}u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = 0 \tag{6}$$

Rewriting $\mathbf{n}$ (locally) as a function of $\mathbf{z}$ rather than the reverse, Equation 6 implies:

$$\nabla_{\mathbf{z}}T(\mathbf{z}) = -\frac{\nabla_{\mathbf{z}}u(T(\mathbf{z}), \mathbf{z}; \mathbf{n}(\mathbf{z}))}{u_T(T(\mathbf{z}), \mathbf{z}; \mathbf{n}(\mathbf{z}))} \tag{7}$$

The right hand side is differentiable in $\mathbf{z}$, which implies that $\nabla_{\mathbf{z}}T(\mathbf{z})$ is differentiable (and hence continuous) in $\mathbf{z}$. But then, if $\mathbf{n}(\mathbf{z})$ and $T(\mathbf{z})$ are both continuously differentiable in $\mathbf{z}$, Equation 7 implies that $\nabla_{\mathbf{z}}T(\mathbf{z})$ is also continuously differentiable. Hence, we have shown that:

Lemma 1. *$T(\mathbf{z}(\mathbf{n}))$ is twice continuously differentiable in $\mathbf{z}$ under any incentive compatible allocation at all $\mathbf{n}$ for which $T(\mathbf{z}(\mathbf{n}))$ is differentiable and $\mathbf{z}(\mathbf{n})$ is a local diffeomorphism at $\mathbf{n}$.*

Under the conditions of Lemma 1, we can totally differentiate Equation 6:

$$D_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) = \nabla_{\mathbf{z}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n}) + \nabla_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) = 0 \tag{8}$$

So under the conditions of Lemma 1, the matrix of second partial deriviatves of $u(T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ with respect to $\mathbf{z}$, denoted $\nabla_{\mathbf{z}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})$, is equal to $-\nabla_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1}$. Hence, the second order condition that $u(T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ is concave in $\mathbf{z}$ around $\mathbf{z}(\mathbf{n})$ is equivalent to $\nabla_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1}$ being positive semi-definite. This second order condition on the matrix $\nabla_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1}$ will be important for our general results on incentive compatibility.[9] This brings us to our first main result:

Theorem 1. *Suppose Assumption 1 holds. Consider an allocation $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ such that $\mathbf{N}$ is compact and the image of $\mathbf{N}$ under the function $\mathbf{n} \mapsto (T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ is compact. In order for this allocation to be incentive compatible: (1) $U(\mathbf{n})$ must satisfy the envelope condition 4, (2) at points $\mathbf{n}$ where $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z}(\mathbf{n}))$ is differentiable then $\nabla_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1}$ must be positive definite, and (3) whenever $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z}(\mathbf{n}))$ is differentiable at two points $\mathbf{n}$ and $\mathbf{n}'$ then $\mathbf{z}(\mathbf{n}) \neq \mathbf{z}(\mathbf{n}')$.*

Proof. See Appendix A.1. □

[9] Using $\nabla_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1}$ rather than $\nabla_{\mathbf{z}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})$ will turn out to be helpful later on especially when utility is separable in $T(\mathbf{z})$ and $\mathbf{z}$ because it allows us to remove explicit dependence on $T(\mathbf{z})$ from the second order condition.

Theorem 1 says that if the envelope condition 4 does not hold, or if the second order condition does not hold at points $\mathbf{n}$ where $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable, or if the allocation is not globally injective across the set of points for which $\mathbf{n} \mapsto \mathbf{z}$ is locally diffeomorphic and $T(\mathbf{z})$ is differentiable, then the allocation is not incentive compatible. One simple way to apply Theorem 1 in practice is to check the Jacobian matrix of $\mathbf{n} \mapsto \mathbf{z}$:

Corollary 1.1. *Suppose Assumption 1 is satisfied. Suppose $\mathbf{n} \mapsto \mathbf{z}$ is locally diffeomorphic and $T(\mathbf{z}(\mathbf{n}))$ is differentiable in $\mathbf{z}$ at two points for which the determinant of the Jacobian, $det(\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n}))$, has different signs. Then then this allocation is not incentive compatible.*

Proof. See Appendix A.2. □

Points (2) and (3) of Theorem 1 rely entirely on the structure generated when the allocation is sufficiently smooth (i.e., $\mathbf{n} \mapsto \mathbf{z}$ is locally diffeomorphic and the transfer schedule is differentiable); hence, Theorem 1 places no restrictions on the portions of an allocation which are not sufficiently smooth (other than satisfying the envelope condition and the implicit restriction that any two sufficiently smooth portions of the mapping $\mathbf{n} \mapsto \mathbf{z}$ must be globally injective). Hence, there are likely some pathological allocations which are not incentive compatible yet do not violate any criteria specified in Theorem 1. However, this is likely inconsequential in practice, as it often seems sensible to restrict ourselves to reasonably smooth allocations when solving multidimensional screening problems numerically. For instance, the following Corollary follows immediately from Theorem 1 if we restrict ourselves to sufficiently smooth allocations:

Corollary 1.2. *Suppose the utility function satisfies Assumption 1. Consider an allocation $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ such that $T(\mathbf{z})$ is differentiable and $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism on compact domain $\mathbf{N}$. Then this allocation is incentive compatible only if (1) $U(\mathbf{n})$ satisfies the envelope condition 4, (2) $\nabla_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n});\mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1}$ is positive definite $\forall \mathbf{n}$, and (3) $\mathbf{n} \mapsto \mathbf{z}$ is injective.*

But if we are willing to restrict ourselves to consider allocations for which $T(\mathbf{z})$ is differentiable and $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism along with the limits of such allocations, we can actually devise a necessary *and sufficient* condition for incentive compatibility:

Theorem 2. *Part (a): Suppose Assumption 1 holds. Consider an allocation $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ such that $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism with compact domain $\mathbf{N}$ and $T(\mathbf{z})$ is differentiable. Then this allocation is incentive compatible if and only if (1) $U(\mathbf{n})$ satisfies the envelope condition 4, (2) $\mathbf{n} \mapsto \mathbf{z}$ is injective, and (3) all individuals $\mathbf{n}$ satisfy:*

$$u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}) \ \forall \mathbf{n}' \in \partial \mathbf{N}$$

Part (b): For any uniformly convergent sequence $(T_j(\mathbf{z}_j(\mathbf{n})), \mathbf{z}_j(\mathbf{n}))$, indexed by j, satisfying the above properties, the limit $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n})) = \lim_{j\to\infty}(T_j(\mathbf{z}_j(\mathbf{n})), \mathbf{z}_j(\mathbf{n}))$ also yields an incentive compatible allocation.

Proof. See Appendix A.3. □

Part (a) of Theorem 2 essentially says that if we restrict ourselves to sufficiently smooth mechanisms, then we only need to ensure that the envelope condition is satisfied, the allocation is injective, and that everyone prefers their assigned bundle to bundles assigned to individuals on the boundary; hence, this reduces the set of relevant incentive compatibility constraints in a meaningful way. Necessity of these conditions follows from Theorem 1; the idea behind showing sufficiency is to leverage Assumption 1. If the envelope condition 4 holds, the transfer schedule is differentiable, and $\mathbf{n} \mapsto \mathbf{z}$ is both locally diffeomorphic and injective (i.e., diffeomorphic), Assumption 1 ensures that no $\mathbf{n} \neq \mathbf{n}'$ can have a local optima at $(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'))$. This means that $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ is the only critical point for each type $\mathbf{n}$. Given that $u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ is a continuous function of $\mathbf{n}'$ on a compact domain $\mathbf{N}$, it has a global maximum that must either be at the sole critical point or the boundary. Given that the sole critical point is preferred to all bundles chosen by individuals on the boundary ($\mathbf{n}' \in \partial\mathbf{N}$), the sole critical point must be the global maximum.

Note, Theorem 2 does *not* require us to check second order conditions: first order conditions holding (via the envelope condition 4) and all individuals preferring their assigned bundle to all boundary bundles implies that local second order conditions must hold strictly $\forall \mathbf{n}$:[10]

Remark 2. *Any allocation satisfying the conditions of Theorem 2 has second order conditions holding strictly so that:*

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}), \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1} \text{ is positive definite}$$

However, one may naturally be concerned that the optimal allocation features non-differentiable $T(\mathbf{z})$ and/or $\mathbf{n} \mapsto \mathbf{z}$ which is not a local diffeomorphism. For example, allocations which feature bunching typically entail non-differentiable $T(\mathbf{z})$ and $\mathbf{n} \mapsto \mathbf{z}$ which are not locally diffeomorphic. Rochet and Chone (1998) show that bunching will necessarily occur under certain circumstances in particular multi-product monopolist problems; hence, it is important that our theory can be applied when the optimal allocation is non-smooth. Fortunately, Part (b) of Theorem 2 shows that our theory can be extended to many such allocations via a simple limiting argument: the uniform limit of allocations with locally diffeomorphic $\mathbf{n} \mapsto \mathbf{z}$ and differentiable $T(\mathbf{z})$ satisfying conditions (1), (2), and (3) from Theorem 2 is incentive compatible.

From a practical perspective, Theorem 2 allows us to drastically reduce the set of relevant incentive compatibility constraints when solving screening problems. Suppose we are trying to solve a screening problem on a square grid of size $m \times m$. If we naively attempted to impose all incentive compatibility constraints, we would have m^4 constraints as each of the m^2 points has m^2 constraints. Under the conditions in Theorem 2, we only need to check m^2 first order conditions plus $m^2 \times 4(m-1)$ boundary conditions plus ensure that $\mathbf{n} \mapsto \mathbf{z}$ is injective. Fortunately, we can often utilize Remark 1 and Corollary 1.1 to check the injectivity of $\mathbf{n} \mapsto \mathbf{z}$ in terms of m^2 conditions on the Jacobian of $\mathbf{n} \mapsto \mathbf{z}$. This order of magnitude is important: if we solve a two dimensional screening problem on a 1000×1000 grid, Theorem 2 reduces the number of incentive compatibility constraints we need to check from ≈ 1 trillion down to ≈ 4 billion, which is a considerable improvement.

While Theorem 2 is perhaps elegant from a mathematical perspective in the sense that it allows us to reduce the number of relevant incentive compatibility constraints by an order of magnitude, it would be great if we could further reduce the number of incentive compatibility constraints. In particular, can we find a way to further reduce the set of incentive compatibility constraints down to only local constraints (i.e., local first and second order incentive compatibility constraints)? Unfortunately, the answer to this question appears to be, in general, no.

[10] By the fact that the allocation is incentive compatible, second order conditions must hold weakly so that $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}), \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ is positive semidefinite. But $\det[\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}), \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}] \neq 0$ by Assumption 1 and the fact $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism (see discussion in point (2) of Appendix A.1 for a proof of this fact), so that $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}), \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ must be positive definite.

Why? Well, the "trick" behind Theorem 2 is essentially to use Assumption 1 to ensure that each type $\mathbf{n}$ has a unique critical point for their utility maximization problem when the allocation is sufficiently smooth. However, even if second order conditions hold so that this unique critical point is a local maximum, it need not be the case that the unique local maximum of a multivariable function is the *global* maximum.[11] Thus, in general, we cannot dispense with the condition that individuals prefer their assigned bundle to boundary bundles. This contrasts to the unidimensional setting: if an individual has a unique critical point for their utility maximization problem which is a local maximum, then it is the global maximum; this follows essentially from the mean value theorem. Similarly, Rochet (1987) and McAfee and McMillan (1988) both show that first and second order conditions are sufficient for incentive compatibility when utility is linear and separable in type because the linearity and separability ensures that the mean value theorem *does hold* for these specific vector valued functions. But extending this result to the multidimensional setting for general utility functions appears to be impossible precisely because the mean value theorem does not hold for vector valued functions.

Finally, we note that Theorem 1 and Theorem 2 can sometimes be extended to settings in which $\dim(\mathbf{n}) \neq \dim(\mathbf{z})$ by restricting to appropriate subspaces, see Appendix A.4.

3.1 Relationship to Previous Results

For completeness, we will now discuss in detail how our results compare with previous incentive compatibility results. However, the material in this section is not crucial for understanding the main ideas, so readers in a hurry can skip to Section 4.

First, we discuss the relationship to the well known incentive compatibility result in one dimension, which states that if the single crossing property holds, an allocation is incentive compatible if and only if $z(n)$ is non-decreasing and $U(n)$ satisfies the envelope condition, see, for example, Mirrlees (1971). In the unidimensional setting, n denotes type and z denotes the single choice variable. Assumption 1 is equivalent to the standard single crossing property in one dimension if utility is given by $u(c, z/n)$ and $c = z + T$ where T is the transfer function (i.e., the negative tax function). Because the function $c \mapsto T$ is bijective conditional on any given z, it is WLOG to consider the government as choosing the function $c(z(n))$ as opposed to $T(z(n))$. In this case, Assumption 1 requires that for all c, z:[12]

$$\frac{\partial \left(\frac{u_2(c, \frac{z}{n})}{n u_1(c, \frac{z}{n})} \right)}{\partial n} > 0$$

And this is exactly the standard single crossing property, see, e.g., Mirrlees (1971). Theorem 1 shows that if the second order condition does not hold when $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and the transfer schedule is differentiable, then the allocation is not incentive compatible. In the unidimensional case, this implies that if $z'(n) < 0$ on some neighborhood, then the allocation

[11] For example, the function $f(x, y) = -x^2 - y^2(1-x)^3$ has a unique critical point (a local maximum) at $(0, 0)$, which is clearly not a global maximum as this function grows arbitrarily large as $x \to \infty$ when $y \neq 0$.

[12] Technically, Assumption 1 tells us that:

$$\frac{\partial \left(\frac{u_2(c, \frac{z}{n})}{n u_1(c, \frac{z}{n})} \right)}{\partial n} > 0 \text{ or } \frac{\partial \left(\frac{u_2(c, \frac{z}{n})}{n u_1(c, \frac{z}{n})} \right)}{\partial n} < 0$$

But if $\frac{\partial \left(\frac{u_2(c, \frac{z}{n})}{n u_1(c, \frac{z}{n})} \right)}{\partial n} < 0$, then $\frac{\partial \left(\frac{u_2(c, -\frac{z}{-n})}{n u_1(c, -\frac{z}{-n})} \right)}{\partial (-n)} > 0$ and we can just relabel $m \equiv -n$, at which point an allocation is incentive compatible if and only if $z(m)$ is non-decreasing and satisfies the envelope condition.

is not incentive compatible.[13] [14] Hence, in the unidimensional setting, Theorem 1 boils down to the statement that non-monotonic $z(n)$ is not incentive compatible *as long as* $z(n)$ and $T(z)$ are sufficiently smooth. In contrast, the standard unidimensional incentive compatibility result does not require differentiability to conclude that a given allocation is not incentive compatible (i.e., the standard result shows that any decreasing $z(n)$ is not incentive compatible). This highlights how our approach is able to say more about the multidimensional case via the added structure of differentiability, but relying on differentiability comes at the cost of not being able to ascertain whether some non-differentiable allocations are incentive compatible.

On the flip side, Theorem 2 tells us in the unidimensional case that any continuous allocation is incentive compatible if $z(n)$ is non-decreasing and $U(n)$ satisfies the envelope condition, which is (essentially) the standard unidimensional sufficient condition. To see why, first note that the set of unidimensional local diffeomorphisms (i.e., functions whose derivative never vanishes) is simply the set of monotonic functions. Then note that:

Remark 3. *Consider a unidimensional setting with a utility function satisfying the single crossing property. If $T(z)$ is differentiable, $n \mapsto z$ is locally diffeomorphic (i.e., monotonic), and $U(n)$ satisfies the envelope condition, then all individuals prefer their assigned bundle to boundary bundles if and only if $z'(n) > 0$.*

Proof. See Appendix A.5. □

Hence, in the unidimensional case, Part (a) of Theorem 2 tells us that any strictly increasing differentiable monotonic function $z(n)$ with differentiable $U(n)$ satisfying the envelope condition is incentive compatible.[15] Part (b) of Theorem 2 then tells us that the uniform limit of any such functions is incentive compatible. But note that the uniform limit of strictly monotonic functions is necessarily weakly monotonic; moreover, *any* bounded (weakly) continuous monotonic function on an interval can be expressed as the uniform limit of strictly monotonic differentiable functions.[16] Thus, in the unidimensional case the sufficient condition in Theorem 2 reduces to saying that any non-decreasing continuous $z(n)$ with $U(n)$ satisfying the envelope condition is incentive compatible.[17]

Next, it's useful to discuss how our results relate to the other case which has been studied extensively in the literature: the model of Rochet (1987) and Rochet and Chone (1998), wherein

[13] The second order condition requires $FOC_n[z'(n)]^{-1}$ to be a positive definite matrix (i.e., positive), where FOC_n is the partial derivative of $FOC(z(n), n) \equiv u_T(T(z(n)), z(n); n)T'(z(n)) + u_z(T(z(n)), z(n); n)$ with respect to n holding $z(n)$ fixed. FOC_n is necessarily positive by the single crossing property, so $FOC_n[z'(n)]^{-1}$ is positive if and only if $z'(n) > 0$. Hence, if $z'(n) < 0$ then we have $FOC_n[z'(n)]^{-1} < 0$, which means the allocation is not incentive compatible by Theorem 1.

[14] Theorem 1 also says a unidimensional allocation is not incentive compatible if two different types n and n' are mapped to the same z for which $n \mapsto z$ is locally diffeomorphic and the transfer schedule is differentiable at both n and n'; this also implies that there must be some region of non-monotonicity in the mapping $n \mapsto z$.

[15] Lemma 2 below shows that locally diffeomorphic $z(n)$ and differentiable $U(n)$ satisfying the envelope condition implies the transfer schedule $T(z)$ will be differentiable.

[16] This follows because (1) every continuous monotonic function can be approximated arbitrarily well (uniformly) by monotonic polynomials (DeVore and Yu, 1985), and (2) every monotonic polynomial on an interval is approximated arbitrarily well by a strictly monotonic polynomial (by adding ϵx for small ϵ).

[17] Hence, Theorem 2 is *slightly* weaker than the standard sufficient condition because it requires $z(n)$ to be continuous. Bergstrom and Dodds (2021) provide a condition on primitives in the unidimensional setting which ensures that the optimal $z(n)$ is continuous; under this condition Theorem 2 coincides with the standard sufficient condition.

individual preferences are quasi-linear in consumption and linear in type $\mathbf{n}$:

$$u(T, \mathbf{z}; \mathbf{n}) = y(\mathbf{z}) + T + \mathbf{n} \cdot v(\mathbf{z}) \tag{9}$$

with $v(\mathbf{z})$ increasing in $z_1, z_2, ..., z_K$. Rochet (1987) showed that if utility is given by 9 and $\mathbf{N}$ is a convex subset of $\mathbb{R}^K$, then $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ is incentive compatible if and only if $U(\mathbf{n})$ satisfies the envelope condition 4 and $U(\mathbf{n})$ is convex in $\mathbf{n}$. Theorem 1 says that $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ must be positive definite if $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable. When utility is given by Equation 9, we have that (we omit the $\mathbf{n}$ argument of $\mathbf{z}(\mathbf{n})$ and $\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})$ in the matrices below for brevity):

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1} = \begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1} \tag{10}$$

On the other hand, convex functions are twice differentiable almost everywhere (Alexandrov's Theorem), which means one can calculate the Hessian matrix (w.r.t $\mathbf{n}$) of $U(\mathbf{n})$ a.e. as:[18]

$$\begin{bmatrix} v_{z_1}(\mathbf{z}) \frac{\partial z_1}{\partial n_1} & \cdots & v_{z_1}(\mathbf{z}) \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ v_{z_K}(\mathbf{z}) \frac{\partial z_K}{\partial n_1} & \cdots & v_{z_K}(\mathbf{z}) \frac{\partial z_K}{\partial n_K} \end{bmatrix} = \begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix} \tag{11}$$

Appendix A.6 shows that Equation 10 is positive definite if and only if Equation 11 is positive definite. Hence, Theorem 1 requires Equation 10 (and therefore Equation 11) to be positive definite whenever $\mathbf{n} \mapsto \mathbf{z}$ is locally diffeomorphic and $T(\mathbf{z})$ is differentiable, which implies that any incentive compatible allocation has $U(\mathbf{n})$ strictly convex at all points for which $\mathbf{n} \mapsto \mathbf{z}$ is locally diffeomorphic and $T(\mathbf{z})$ is differentiable. However, Theorem 1 is slightly weaker than Rochet (1987)'s necessary condition because it places no condition on the convexity of $U(\mathbf{n})$ when $\mathbf{n} \mapsto \mathbf{z}$ is not locally diffeomorphic or $T(\mathbf{z})$ is non-differentiable at a given $\mathbf{n}$ other than the implicit global constraint that if the allocation is a local diffeomorphism (and $T(\mathbf{z})$ is differentiable) at two points $\mathbf{n}$ and $\mathbf{n}'$, then $\mathbf{z}(\mathbf{n}) \neq \mathbf{z}(\mathbf{n}')$.

Next, we show that for utility function 9, Theorem 2 tells us that allocations with continuous $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ and convex $U(\mathbf{n})$ such that the envelope condition holds are incentive compatible. First, note that any allocation satisfying the conditions of Part (a) of Theorem 2 must have the envelope condition 4 holding and differentiable, strictly convex $U(\mathbf{n})$. This is because any allocation satisfying the conditions of Part (a) of Theorem 2 has second order conditions holding strictly (see Remark 2) so that Equation 10 (and therefore Equation 11) is positive definite. Moreover, any allocation satisfying the envelope condition 4 everywhere with differentiable, strictly convex $U(\mathbf{n})$ satisfies the conditions required in Theorem 2 because: (1) $\mathbf{n} \mapsto \mathbf{z}$ is locally diffeomorphic as Equation 11 is positive definite (so that $\det[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})] \neq 0$), (2) $\mathbf{n} \mapsto \mathbf{z}$ is injective because if two types $\mathbf{n}, \mathbf{n}'$ are mapped to the same $\mathbf{z}$, then $U(\mathbf{n}) - U(\mathbf{n}')$ is a linear function of $\mathbf{n} - \mathbf{n}'$, hence not strictly convex, and (3) all individuals prefer their assigned bundle to boundary bundles by a mean value theorem argument (see McAfee and McMillan (1988)).[19] Part (b) of Theorem 2 then strengthens this relationship by telling us that the uniform limit of allocations with differentiable, strictly convex $U(\mathbf{n})$ with the envelope condition 4

[18] Note, we have used the envelope condition 4 in calculating this Hessian.

[19] Also, $T(\mathbf{z})$ is differentiable by Lemma 2 below.

holding is incentive compatible. Hence, for utility function 9, Part (b) of Theorem 2 tells us that allocations with continuous $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ and convex $U(\mathbf{n})$ such that the envelope condition holds are incentive compatible.[20] When utility is given by Equation 9, optimal $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ is always continuous; thus, the sufficient condition in Theorem 2 coincides with the sufficient criterion from Rochet (1987).[21]

Our results are also related to McAfee and McMillan (1988), who show that first and second order conditions are necessary and sufficient to characterize multidimensional incentive compatibility if preferences satisfy a sort of single crossing property and one restricts attention to smooth allocations. The key difference is that their single crossing property is very restrictive as they require $\mathbf{n} \mapsto \frac{u_{\mathbf{z}}(T,\mathbf{z};\mathbf{n})}{u_T(T,\mathbf{z};\mathbf{n})}$ to satisfy a mean value theorem for vector-valued functions. Matkowski (2012) shows that only linear functions of $\mathbf{n}$ or functions that are linear in $\mathbf{n}$ with a common non-linear component satisfy this property; in practice, it appears to be quite difficult to find any realistic utility function which satisfies their single crossing property *other than* the case where $u(T, \mathbf{z}; \mathbf{n})$ is linear and separable in type so that utility is given by:

$$u(T, \mathbf{z}; \mathbf{n}) = w(\mathbf{z}, T) + \mathbf{n} \cdot v(\mathbf{z}) \tag{12}$$

In contrast, there appear to be many reasonable utility functions which satisfy our generalized single crossing property such as Equation 3 and Equation 21 used in our numerical simulations later on. These utility functions do not satisfy the generalized single crossing property from McAfee and McMillan (1988). A second difference is that the results in McAfee and McMillan (1988) only pertain to *differentiable* allocations, which means their results cannot be applied to situations in which bunching occurs. In contrast, our results can be applied to non-differentiable allocations (because we only require smoothness in *portions* of the allocation in Theorem 1 and via the limiting argument in Theorem 2).

Finally, we compare our results to Carlier (2001), who characterizes incentive compatibility when preferences are separable and quasi-linear in consumption using the notion of h-convexity. The first difference between the present paper and Carlier (2001) is that our results do not require quasi-linearity or separability; however, we do require a generalized single crossing property whereas Carlier (2001) does not require any sort of single crossing condition. In this sense, our results are complementary. Secondly, our necessary conditions (Theorem 1 and Corollary 1.2) can be checked using local properties of the allocation (note injectivity of $\mathbf{n} \mapsto \mathbf{z}$ can often be checked via Remark 1 and Corollary 1.1); this local description aids greatly in solving screening problems numerically. In contrast, Carlier (2001)'s characterization of incentive compatibility is based on *global* h-convexity constraints which are arguably less intuitive and cannot (to my knowledge) be expressed in terms of local properties of the allocation, making them less useful for numerically solving optimal screening problems.

4 Solving Multidimensional Screening Problems Numerically

The next question we want to address is: how can we use the results from Section 3 to actually solve multidimensional screening problems in practice? Suppose we want to optimize some objective function subject to incentive compatibility constraints (and potentially other constraints,

[20]The fact that every convex function can be approximated with a differentiable, strictly convex function results from the facts that (1) every convex function can be approximated arbitrarily well using a strictly convex function, (2) we assume $U(\mathbf{n}) \in H^1(\mathbf{N})$, and (3) the set of infinitely differentiable functions over $\mathbf{N}$, $C^\infty(\mathbf{N})$, is dense in the Sobolev space $H^1(\mathbf{N})$; see Theorem 8.7 of Brezis (2011).

[21]Proposition 6 of Rochet and Chone (1998) states this fact without proof; we thus prove in Appendix A.12 that $\mathbf{z}(\mathbf{n})$, and therefore $T(\mathbf{z}(\mathbf{n}))$, must be continuous when preferences are given by utility function 9.

such as participation constraints, budget constraints, or non-negativity constraints):

$$\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}),T(\mathbf{z}(\mathbf{n}))} \quad \text{Objective} \\
&\text{s.t. } \mathbf{n} \in \operatorname*{argmax}_{\mathbf{n}'} u(T(\mathbf{z}(\mathbf{n}')),\mathbf{z}(\mathbf{n}');\mathbf{n}) \ \forall \mathbf{n} \\
&\text{Other Constraints}
\end{aligned} \tag{13}$$

One path to solving System 13 might be to simply apply Theorem 2 and search over all allocations with differentiable $T(\mathbf{z})$ and diffeomorphic $\mathbf{n} \mapsto \mathbf{z}$ that satisfy the envelope condition and for which all individuals prefer their assigned bundle to all boundary bundles. However, because there are fewer second order condition constraints than there are constraints that all individuals prefer their assigned bundle to all boundary bundles, we will instead appeal to Corollary 1.2 and search over all allocations such that: (1) $U(\mathbf{n})$ is differentiable and satisfies the envelope condition 4, (2) second order conditions are satisfied, and (3) $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism (i.e., $\det(\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n}))$ never vanishes). We don't need to impose that $T(\mathbf{z})$ is differentiable because of the following Lemma:

Lemma 2. *Consider any potential allocation $(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n}))$ which yields differentiable $U(\mathbf{n})$ satisfying the envelope condition 4. If $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism, then $T(\mathbf{z})$ is twice continuously differentiable.*

Proof. See Appendix A.7. □

Then, we can check to ensure that any proposed optimal allocation satisfying the above criteria of Corollary 1.2 satisfies the sufficient conditions of Theorem 1: $\mathbf{n} \mapsto \mathbf{z}$ is injective and all individuals prefer their assigned bundles to boundary bundles. This allows us to state our general technique to solve multidimensional screening problems (WLOG, we assume $\det(\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})) > 0$ rather than $\det(\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})) < 0$):

$$\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}),T(\mathbf{z}(\mathbf{n}))} \quad \text{Objective} \\
&\text{s.t. } U(\mathbf{n}_1) - U(\mathbf{n}_2) = \oint_{\mathbf{n}_2}^{\mathbf{n}_1} \nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) \cdot d\mathbf{n} \ \forall \mathbf{n}_1,\mathbf{n}_2 \\
&\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n});\mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1} \text{ is positive definite} \\
&\det(\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})) > 0 \\
&U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) \\
&\text{Other Constraints}
\end{aligned} \tag{14}$$

Note, we can further simplify System 14 in two ways. First, instead of maximizing over $(\mathbf{z}(\mathbf{n}),T(\mathbf{z}(\mathbf{n})))$, let us maximize over $(\mathbf{z}(\mathbf{n}),U(\mathbf{n}))$.[22] Second, note that we only actually have a choice over utility at single point, $U(\underline{\mathbf{n}})$ as utility at all other levels is determined by the envelope condition given utility at a given point $\underline{\mathbf{n}}$. Hence, we can instead simply think about choosing the function $\mathbf{z}(\mathbf{n})$ and the scalar $U(\underline{\mathbf{n}})$, and defining $U(\mathbf{n})$ by the envelope condition. In doing so, we must incorporate the fact that $U(\mathbf{n})$ must be a function so that

$$\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n})$$

[22] We define $T(\mathbf{z}(\mathbf{n}))$ implicitly by $U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n})$.

forms a conservative vector field. Hence, we can recast System 14 as:

$$
\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}),U(\underline{\mathbf{n}})} \quad \text{Objective} \\
&\text{s.t. } U(\mathbf{n}) = U(\underline{\mathbf{n}}) + \oint_{\underline{\mathbf{n}}}^{\mathbf{n}} \nabla_{\mathbf{s}} u(T(\mathbf{z}(\mathbf{s})), \mathbf{z}(\mathbf{s}); \mathbf{s}) \cdot d\mathbf{s} \\
&\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1} \text{ is positive definite} \\
&\det(\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})) > 0 \\
&U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \\
&\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \text{ is conservative} \\
&\text{Other Constraints}
\end{aligned} \tag{15}
$$

System 15 is an optimization problem with linear and non-linear constraints. If we find a global maximizer for this problem, then we know that it is globally optimal among smooth allocations by Theorem 2 if: (1) $\mathbf{n} \mapsto \mathbf{z}$ is injective, which can be checked via Remark 1, and (2) all individuals prefer their assigned bundles to boundary bundles.

Finally, it is very important to mention that the approach we have outlined to solve multidimensional screening problems assumes that the optimal allocation is smooth ($\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable). But what if the optimal allocation is not smooth (e.g., features bunching)? As long as the true optimal schedule can be approximated arbitrarily well by smooth functions (which I conjecture is true for most realistic scenarios), then solving System 15 will simply find a smooth allocation which is arbitrarily close to the optimal allocation.

4.1 Separable Utility

It turns out that we can strengthen our results and simplify System 15 substantially if we restrict ourselves to utility functions which have the nice separable form:

$$
u(T, \mathbf{z}; \mathbf{n}) = u^{(0)}(T, \mathbf{z}) + \sum_{i}^{K} u^{(i)}(z_i, n_i) \tag{16}
$$

When utility is given by Equation 16, we can prove the following two propositions:

Proposition 1. *Consider any utility function given by Equation 16 on a rectangular domain* $\mathbf{N}$*. Suppose* $\frac{\partial^2 u^{(i)}(z_i,n_i)}{\partial z_i \partial n_i} > 0 \; \forall i$*. Any allocation for which* $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ *is positive definite* $\forall \mathbf{n}$ *has diffeomorphic* $\mathbf{n} \mapsto \mathbf{z}$*.*

Proof. See Appendix A.8. □

Proposition 2. *Consider any utility function given by Equation 16. Any allocation for which* $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ *is positive definite* $\forall \mathbf{n}$ *generates a conservative vector field* $\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$*.*

Proof. See Appendix A.9. □

Propositions 1 and 2 both use the fact that utility function 16 will always yield a matrix $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})$ that is diagonal; this additional structure is the key feature that allows us to prove these statements. Proposition 1 implies that when utility is given by Equation 16, any allocation satisfying the envelope condition 4 and the second order condition everywhere will satisfy our necessary conditions for incentive compatibility given by Theorem 1. Moreover, Proposition 1 allows us to replace the conditions from Theorem 2 that $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and that $\mathbf{n} \mapsto \mathbf{z}$ is injective with the simpler requirement that the second order condition is satisfied $\forall \mathbf{n}$:

Corollary 2.1. *Suppose Assumption 1 holds and utility is given by Equation 16 with positive cross partials as in Proposition 1. Consider an allocation* $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ *on compact rectangular domain* $\mathbf{N}$*. The following three conditions are sufficient for incentive compatibility: (1)* $U(\mathbf{n})$ *satisfies the envelope condition 4, (2)* $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ *is positive definite* $\forall \mathbf{n}$*, and (3)* $u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})\ \forall \mathbf{n}' \in \partial \mathbf{N}$.

Moreover, the uniform limit of such allocations is also incentive compatible (as in Theorem 2).

Proof. Lemma 2 ensures $T(\mathbf{z})$ is differentiable and Proposition 1 ensures $\mathbf{n} \mapsto \mathbf{z}$ is a diffeomorphism (i.e., an injective local diffeomorphism). Corollary 2.1 then follows immediately from Theorem 2. ☐

When the conditions for Corollary 2.1 are satisfied, we can simplify System 15 substantially:

$$\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}), U(\underline{\mathbf{n}})} \ \text{Objective} \\
&\text{s.t. } U(\mathbf{n}) = U(\underline{\mathbf{n}}) + \oint_{\underline{\mathbf{n}}}^{\mathbf{n}} \nabla_{\mathbf{s}} u(T(\mathbf{z}(\mathbf{s})), \mathbf{z}(\mathbf{s}); \mathbf{s}) \cdot d\mathbf{s} \\
&\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1} \text{ is positive definite} \\
&U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \\
&\text{Other Constraints}
\end{aligned} \tag{17}$$

If we find a solution to System 17 all we need to do is check that all individuals prefer their assigned bundles to boundary bundles; if so, we have found the optimal allocation within the class of smooth allocations ($\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable).

4.2 Relationship to Previous Numerical Methods

To the best of my knowledge, there are two paths that are commonly used to solve multidimensional screening problems numerically. First, one can attempt to use so-called "first order approaches" that replace the incentive compatibility constraints with first order conditions and hope that the optimal allocation is interior (so that second order conditions are not binding). We outline two different first order methods (one based on the Euler-Lagrange equation and one based on optimal control) in Appendix B. However, there are two potential issues with these first order approaches. First, the resulting partial differential equation(s) is often quite difficult to solve. Second, and more importantly, if the optimal allocation is not interior (so that second order conditions bind) then first order approaches fail to generate the optimal schedule. Typically, allocations fail to be interior when the optimal allocation features bunching, wherein many types are assigned to the same bundle.[23] Rochet and Chone (1998) show that this phenomenon

[23]Allocations can also fail to be interior when some individuals have multiple optima, which result in discontinuities in $\mathbf{n} \mapsto \mathbf{z}$. When the type and action space are both one dimensional, Bergstrom and Dodds (2021) provide a condition which rules out multiple optima under any optimal allocation as long as the single crossing

is the norm rather than the exception in a multi-product monopolists problem; we show below that bunching also appears to be quite common in optimal multidimensional taxation.

When first order approaches fail, there is, as far as we know, only one path forward and it requires utility to be given by Equation 9 (i.e., linear and separable in type $\mathbf{n}$). When utility takes this simple form, we can appeal to the convexity characterization of incentive compatibility from Rochet (1987), which turns the mechanism design problem into a calculus of variations problem subject to a convexity constraint. Then we can use numerical algorithms designed to solve variational calculus problems subject to a convexity constraint, e.g., Aguilera and Morin (2008), Oberman (2013), or Mérigot and Oudet (2014).[24]

Hence, the numerical method outlined in Section 4 is, as far as we know, the first algorithm designed to solve multidimensional screening problems which can be applied even when the optimal allocation features bunching and when utility is not linear and separable in type (the method devised in this paper can be applied whenever Assumption 1 holds).[25] Thus, we believe developing a new numerical method to solve multidimensional screening problems is an important contribution of this paper.

5 Application: Optimal Multidimensional Taxation

Next, we will illustrate how to apply our incentive compatibility results to a particular class of screening problems: optimal multidimensional taxation.

5.1 Problem Setup

Individuals make (observable) choices $\mathbf{z}$ given characteristics $\mathbf{n}$ to maximize utility $u(T, \mathbf{z}; \mathbf{n})$, which depends on the transfer $T(\mathbf{z})$, choices $\mathbf{z}$, and type $\mathbf{n}$ (see Problem 1).[26] We assume that $u_T(T, \mathbf{z}; \mathbf{n}) > 0$. The government chooses the function $T(\mathbf{z})$ to maximize a welfare function subject to a revenue constraint that total transfers in society plus exogenous per-capita government expenditures, E, must not exceed zero. The government cannot observe $\mathbf{n}$ for any individual, but can observe both $\mathbf{z}$ as well as the distribution of types $F(\mathbf{n})$. We can express this as a mechanism design problem wherein the government chooses functions $\mathbf{z}(\mathbf{n})$ and $T(\mathbf{z}(\mathbf{n}))$ and

property holds, but extending this condition to multidimensional settings does not appear to be possible other than in the case where preferences are given by utility function 19. We prove in Appendix A.12 that individuals with multiple optima cannot ever be socially optimal when preferences are given by utility function 19. In theory, the schedule could also not feature bunching or multiple optima, but $\mathbf{n} \mapsto \mathbf{z}$ could simply be pathological, e.g., Hencl (2011).

[24] Technically, a second path forward would be to use the sweeping conditions in Rochet and Chone (1998). However, this also requires utility to be given by Equation 9 and is more difficult to apply numerically than the algorithms designed to solve variational calculus problems subject to a convexity constraint.

[25] Note, the utility function discussed in Section 4.1 is *separable* in type $\mathbf{n}$ but *not necessarily linear* in type $\mathbf{n}$.

[26] If $T(\mathbf{z})$ is positive, then the government transfers money to the individual if $T(\mathbf{z})$ is negative, the government taxes money away from the individual.

individuals choose an $\mathbf{n}'$ to report to the government, which determines their $\mathbf{z}$ and T:[27]

$$\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}),T(\mathbf{z}(\mathbf{n}))} \int_{\mathbf{N}} W\left(U(\mathbf{n}),\mathbf{n}\right) dF(\mathbf{n}) \\
&\text{s.t. } \int_{\mathbf{N}} [T(\mathbf{z}(\mathbf{n}))+E] dF(\mathbf{n}) \leq 0 \\
&\mathbf{n} \in \underset{\mathbf{n}'}{\operatorname{argmax}}\; u(T(\mathbf{z}(\mathbf{n}')),\mathbf{z}(\mathbf{n}');\mathbf{n}) \;\forall \mathbf{n} \\
&U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n})
\end{aligned} \tag{18}$$

This general framework can capture many different applications of multidimensional optimal taxation. For instance, one may use this framework to analyze (1) joint taxation of couples (where $K = 2$, z_1 and z_2 represent the labor income of each individual), (2) taxation of earnings and hours worked (where $K = 2$ and z_1 represents earnings and z_2 represents hours worked), (3) tax-preferred consumption, such as mortgage payments in the United States (where z_1 captures earnings, $z_2, ..., z_K$ capture spending on various goods which are tax-preferred), or (4) joint income and capital taxation (where $K = 2$, z_1 represents labor income, and z_2 represents capital income).

5.2 Illustration of Method

We will use two utility functions to illustrate how to apply our theoretical results as well as our numerical solution technique. In order to utilize the simplifications discussed in Section 4.1, both of these utility functions will be of the form of Equation 16. First, we will apply this method to utility function 19:[28]

$$u(T,\mathbf{z};\mathbf{n}) = z_1 + z_2 + T + n_1 \frac{z_1^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2^{1+\theta_2}}{1+\theta_2} \tag{19}$$

with $n_1, n_2 < 0$ and $\theta_1, \theta_2 > 0$.

Applying our method to utility function 19 will be useful as a test case because we can check that the solutions from our numerical methods align with the solutions which arise from numerical methods designed to solve variational calculus problems with convexity constraints. In particular, when utility is given by Equation 19, we can express $T(\cdot)$ as a function of $U(\mathbf{n})$ and $\nabla_{\mathbf{n}} U(\mathbf{n})$ (see Remark 5 in Appendix B); moreover, we can replace the incentive compatibility constraints with a convexity constraint as a result of Rochet (1987):

$$\begin{aligned}
&\max_{U(\mathbf{n})} \int_{\mathbf{N}} W\left(U(\mathbf{n}),\mathbf{n}\right) f(\mathbf{n}) d\mathbf{n} \\
&\text{s.t. } \int_{\mathbf{N}} [T(U(\mathbf{n}),\nabla_{\mathbf{n}} U(\mathbf{n}))+E] dF(\mathbf{n}) \leq 0 \\
&U(\mathbf{n}) \text{ is convex}
\end{aligned} \tag{20}$$

[27]One may be interested in *existence* of a solution to the optimal taxation problem given by System 18. We prove and discuss such an existence result in Appendix A.11, but leave this out of the main text for the sake of brevity and streamlining.

[28]Note, for those familiar with optimal taxation models, the parametrization may seem a bit strange. For instance, for the first utility function, a more natural parametrization might be to have $u(T, l_1, l_2; m_1, m_2) = m_1 l_1 + m_2 l_2 + T(m_1 l_1, m_2 l_2) - \frac{l_1^{1+\theta_1}}{1+\theta_1} - \frac{l_2^{1+\theta_2}}{1+\theta_2}$, where m_1, m_2 represent labor productivities and l_1, l_2 represent the two labor supply choices. A simple change of variables to $z_1 = m_1 l_1$ and $z_2 = m_2 l_2$ and redefining $n_1 = -m_1^{-(1+\theta_1)}$, $n_2 = -m_2^{-(1+\theta_2)}$ shows that these parametrizations are isomorphic; we do this change of variables so as to get our utility function into the form of Rochet (1987).

A number of algorithms have been developed to solve variational calculus problems subject to convexity constraints like System 20, e.g., Aguilera and Morin (2008), Oberman (2013), or Mérigot and Oudet (2014).

The second utility function we consider is quadratic in $\mathbf{n}$ rather than linear in $\mathbf{n}$, which takes us outside of the realm of problems covered by the incentive compatibility results in Rochet (1987) or McAfee and McMillan (1988):[29]

$$u(T, \mathbf{z}; \mathbf{n}) = z_1 + z_2 + T + n_1 \frac{z_1^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2^{1+\theta_2}}{1+\theta_2} - \frac{n_1^2}{2\alpha} z_1 - \frac{n_2^2}{2\alpha} z_2 \tag{21}$$

with scaling parameter $\alpha > 0$, $n_1, n_2 < 0$, and $\theta_1, \theta_2 > 0$. Note, we assume that $\alpha, \theta_1, \theta_2$ are homogeneous across the population. Both of these utility functions can, for example, be interpreted as taxation of couples, where the two individuals have different disutilities over generating income n_1, n_2. We can nest utility function 19 within utility function 21 if we set $\alpha = \infty$. Hence, we will just discuss how to apply our method to utility function 21. First, note:

Proposition 3. *Utility function 21 (and hence utility function 19) satisfies Assumption 1 on rectangular domains.*

Proof. See Appendix A.10. □

Let us now discuss how to apply the numerical method outlined in Section 4 for utility function 21. Note that Utility function 21 satisfies the conditions of Corollary 2.1. Hence, we need to first ensure that $U(\mathbf{n})$ satisfies the envelope condition so that $\forall \mathbf{n}$:

$$U(\mathbf{n}) = U(\underline{\mathbf{n}}) + \oint_{\underline{\mathbf{n}}}^{\mathbf{n}} \nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \cdot d\mathbf{s} = U(\underline{\mathbf{n}}) + \oint_{\underline{\mathbf{n}}}^{\mathbf{n}} \begin{bmatrix} \frac{z_1(\mathbf{s})^{1+\theta_1}}{1+\theta_1} - \frac{s_1}{\alpha} z_1(\mathbf{s}) \\ \frac{z_2(\mathbf{s})^{1+\theta_2}}{1+\theta_2} - \frac{s_2}{\alpha} z_2(\mathbf{s}) \end{bmatrix} \cdot d\mathbf{s}$$

We also need to ensure that the chosen allocation satisfies the second order condition $\forall \mathbf{n}$, i.e., that the following is positive definite $\forall \mathbf{n}$:

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1} = \begin{bmatrix} z_1(\mathbf{n})^{\theta_1} - \frac{n_1}{\alpha} & 0 \\ 0 & z_2(\mathbf{n})^{\theta_2} - \frac{n_2}{\alpha} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) \end{bmatrix}^{-1} \tag{22}$$

But Equation 22 is positive definite if and only if it is symmetric and has all positive principal minors. In order for $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ to be symmetric we require that:[30]

$$\left(z_1^{\theta_1}(\mathbf{n}) - \frac{n_1}{\alpha}\right) \frac{\partial z_1}{\partial n_2}(\mathbf{n}) = \left(z_2(\mathbf{n})^{\theta_2} - \frac{n_2}{\alpha}\right) \frac{\partial z_2}{\partial n_1}(\mathbf{n}) \tag{23}$$

[29]Note, this utility function is not outside the realm of functions considered by Carlier (2001); however, as discussed in Section 3.1, the characterization of Carlier (2001) uses global (rather than local) properties of the allocation, making them difficult to implement numerically (Carlier (2001) does not compute any numerical solutions of multidimensional screening problems).

[30]By Proposition 2, we know that for utility function 21, satisfying the second order condition everywhere ensures that $\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ is a conservative vector field. This is also evident by direct computation noting that $\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ is conservative if and only if Equation 23 holds.

And in order for Equation 22 to have all positive principal minors we must have:

$$\begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) \end{bmatrix} \tag{24}$$

is a P matrix (i.e., all principal minors are positive so that $\frac{\partial z_1}{\partial n_1}, \frac{\partial z_2}{\partial n_2}, \frac{\partial z_1}{\partial n_1}\frac{\partial z_2}{\partial n_2} - \frac{\partial z_1}{\partial n_2}\frac{\partial z_2}{\partial n_1} > 0$).[31]

Hence, we solve the following numerical optimization problem:

$$\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}),U(\underline{\mathbf{n}})} \int_{\mathbf{N}} W(U(\mathbf{n}),\mathbf{n})dF(\mathbf{n}) \\
&\text{s.t. } \int_{\mathbf{N}} [T(\mathbf{z}(\mathbf{n}))+E]dF(\mathbf{n}) \leq 0 \\
&U(\mathbf{n}) = U(\underline{\mathbf{n}}) + \oint_{\underline{\mathbf{n}}}^{\mathbf{n}} \begin{bmatrix} \frac{z_1(\mathbf{s})^{1+\theta_1}}{1+\theta_1} - \frac{s_1}{\alpha} z_1(\mathbf{s}) \\ \frac{z_2(\mathbf{s})^{1+\theta_2}}{1+\theta_2} - \frac{s_2}{\alpha} z_2(\mathbf{s}) \end{bmatrix} \cdot d\mathbf{s} \\
&\frac{\partial z_1}{\partial n_1}(\mathbf{n}) > 0, \frac{\partial z_2}{\partial n_2}(\mathbf{n}) > 0, \frac{\partial z_1}{\partial n_1}(\mathbf{n})\frac{\partial z_2}{\partial n_2}(\mathbf{n}) - \frac{\partial z_1}{\partial n_2}(\mathbf{n})\frac{\partial z_2}{\partial n_1}(\mathbf{n}) > 0 \\
&\left(z_1^{\theta_1}(\mathbf{n}) - \frac{n_1}{\alpha}\right)\frac{\partial z_1}{\partial n_2}(\mathbf{n}) = \left(z_2(\mathbf{n})^{\theta_2} - \frac{n_2}{\alpha}\right)\frac{\partial z_2}{\partial n_1}(\mathbf{n}) \\
&T(\mathbf{z}(\mathbf{n})) = U(\mathbf{n}) - \left[n_1\frac{z_1(\mathbf{n})^{1+\theta_1}}{1+\theta_1} + n_2\frac{z_2(\mathbf{n})^{1+\theta_2}}{1+\theta_2} - \frac{n_1^2}{2\alpha}z_1(\mathbf{n}) - \frac{n_2^2}{2\alpha}z_2(\mathbf{n})\right] - z_1(\mathbf{n}) - z_2(\mathbf{n})
\end{aligned} \tag{25}$$

Ultimately, this is a fairly straight-forward optimization problem with non-linear inequality and equality constraints. For moderate sized grids (e.g., 40x40), this problem can be solved within a few hours using a standard version of Matlab on a laptop.[32] Once we have a candidate solution for System 25, we simply check whether it satisfies the conditions of Theorem 2. Given that we will solve this problem on a rectangular domain, we can appeal to Proposition 1 which ensures that the the second order conditions holding everywhere implies that $\mathbf{n} \mapsto \mathbf{z}$ is a diffeomorphism. Hence, any allocation that solves System 25 necessarily features diffeomorphic $\mathbf{n} \mapsto \mathbf{z}$. As a result, our solution method when utility is given by Equation 19 or 21 simply boils down to solving System 25 and checking that $\forall \mathbf{n}$ we have $u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ $\forall$ $\mathbf{n}' \in \partial\mathbf{N}$. Theorem 2 will then ensure that the proposed solution is incentive compatible.

5.3 Four Illustrative Simulations

Next, we work through four toy examples, which are meant to illustrate the above simulation method rather than closely depict reality. For each of the two utility functions 19 and 21, we consider two scenarios, which differ on the chosen distribution of types $f(\mathbf{n})$ and welfare weights $\psi(\mathbf{n})$. We suppose that $W(U(\mathbf{n}),\mathbf{n}) = \psi(\mathbf{n})U(\mathbf{n})$ so that the marginal social welfare gain from increasing utility for each type $\mathbf{n}$ is constant. Welfare weights $\psi(\mathbf{n})$ are chosen so that marginal social welfare gain from increasing utility is decreasing with $\mathbf{n}$.[33] In both examples we consider a rectangular domain of $[-6,-0.5]^2$, $\theta_1 = \theta_2 = 3$ (corresponding to a compensated taxable income elasticity of 1/3 with no taxes), $\alpha = 50$, and $\psi(\mathbf{n}) = e^{-\beta U_0(\mathbf{n})}$ where $U_0(\mathbf{n})$ is optimal utility under zero taxes. These welfare weights $\psi(\mathbf{n})$ imply that the government cares more

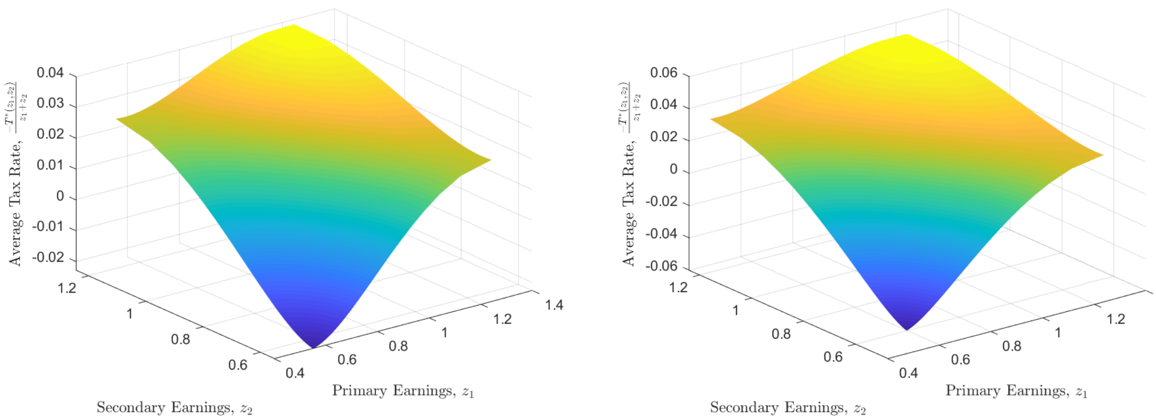
[31] The fact that Equation 22 has positive principal minors (i.e., is a P matrix) if and only if Equation 24 is a P matrix follows from the fact that the product of a diagonal matrix with positive entries and a P matrix is a P matrix and the fact that a matrix is a P matrix if and only if its inverse is a P matrix. These are standard results; see Theorem 3.1 of Tsatsomeros (2004) for a proof.

[32] Our simulation algorithm uses Matlab's "fmincon" method to solve for the optimal schedule given the non-linear equality and inequality constraints in system 25.

[33] E.g., Lockwood and Weinzierl (2016) consider such a social welfare function in an optimal taxation problem.

about increasing consumption for low income households than high income households. For our first set of input data, we consider uniform $f(\mathbf{n})$ and $\beta = 1$, which, given the range of $\mathbf{n}$ we consider, means that the marginal social welfare of giving the lowest income household a dollar is about 3 times higher than giving the highest income household a dollar (the highest income household earns about 2.5 times as much as the lowest income household).

For both utility functions 19 and 21, we find that the solution to System 25 is interior in the sense that it does not feature any bunching; moreover, the solution is such that all individuals prefer their assigned bundle to boundary bundles. Hence, by Theorem 2, we know that the proposed solution is incentive compatible. Optimal average tax rates, $\frac{-T^*(z_1,z_2)}{z_1+z_2}$, for these two utility functions are displayed in Figure 1 (recall that $T^*(z_1, z_2)$ is the optimal transfer given at income (z_1, z_2) so that $-T^*(z_1, z_2)$ is the optimal tax at income (z_1, z_2)). Note, for both utility functions, the tax schedule is not overly progressive because marginal social welfare of giving \$1 to the lowest income household is only about 3 times higher than giving \$1 to the highest income household; the maximum average tax rate is 4% (5% for utility function 21) and the maximum marginal tax rate is about 13% (17% for utility function 21).



(a) Utility Function 19

(b) Utility Function 21

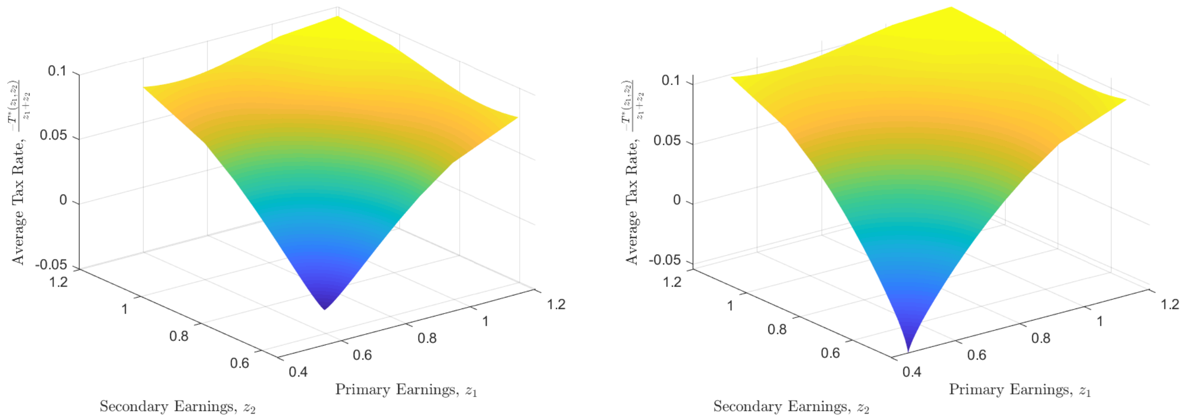
Figure 1: Optimal Average Tax Rates for First Set of Input Data

Note: This figure shows the optimal average tax schedule $\frac{-T^*(z_1,z_2)}{z_1+z_2}$ for utility functions 19 (panel 1a) and 21 (panel 1b). $\theta_1 = \theta_2 = 3$ (corresponding to a compensated taxable income elasticity of 0.33 with no taxes) and $\alpha = 50$. $f(\mathbf{n})$ is uniform on a rectangular domain of $[-6, -0.5]^2$ and $\psi(\mathbf{n}) = e^{-U_0(\mathbf{n})}$ where $U_0(\mathbf{n})$ is optimal utility under zero taxes: this means the marginal social welfare of giving the lowest income household a dollar is roughly 3 times higher than giving the highest income household a dollar for both utility functions.

For our second choice of $f(\mathbf{n})$ and $\psi(\mathbf{n})$, we consider normally distributed $f(\mathbf{n})$ with a small positive covariance between n_1 and n_2. We still assume $\psi(\mathbf{n}) = e^{-\beta U_0(\mathbf{n})}$ where $U_0(\mathbf{n})$ is optimal utility under zero taxes, but now we pick $\beta = 10.9$ (and $\beta = 8.1$ for utility function 21), which, given the range of $\mathbf{n}$ we consider, means that the marginal social welfare of giving the lowest income household a dollar is roughly 100,000 times higher than giving the highest income household a dollar (recall the highest income household earns about 2.5 times as much as the lowest income household).[34] The optimal tax schedule from solving System 25 is shown in Figure 2. Tax rates are much higher: for utility function 19 the maximum average tax rate is around 9% (11% for utility function 19) and the maximum marginal tax rate is around 34%

[34]We set the marginal social welfare of giving the lowest income household a dollar to be many thousands of times higher than for the highest income household so as to generate more substantial bunching in the optimal solution. However, some bunching occurs even if the marginal social welfare of giving a dollar to the lowest income household is only a few hundred times higher than of giving a dollar to the highest income household.

(36% for utility function 19).



(a) Utility Function 19

(b) Utility Function 21

Figure 2: Optimal Average Tax Rates for Second Set of Input Data

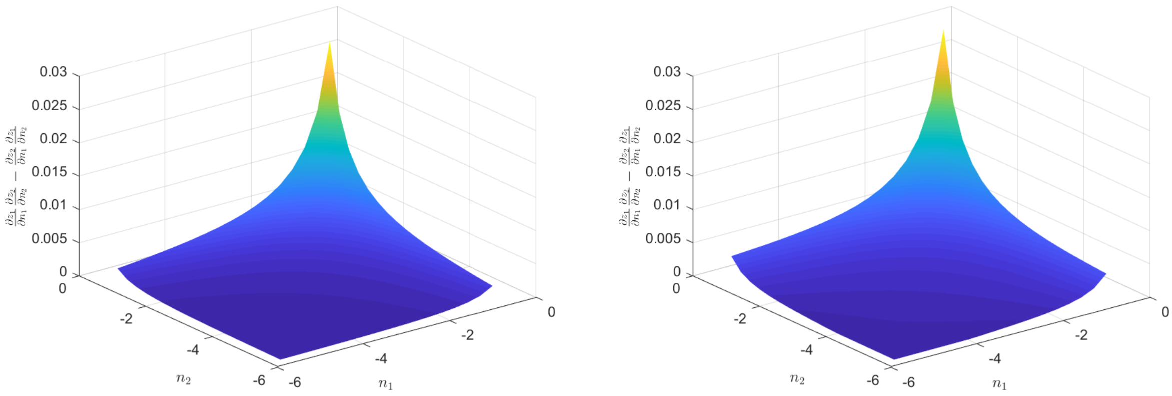
Note: This figure shows the optimal average tax schedule $\frac{-T^*(z_1,z_2)}{z_1+z_2}$ for utility functions 19 (panel 2a) and 21 (panel 2b). $\theta_1 = \theta_2 = 3$ (corresponding to a compensated taxable income elasticity of 0.33 with no taxes) and $\alpha = 50$. $f(\mathbf{n})$ is joint normal on a rectangular domain of $[-6, -0.5]^2$ with mean matrix $\mu = \begin{bmatrix} -5.5 \\ -5.5 \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} 5 & 0.5 \\ 0.5 & 5 \end{bmatrix}$. $\psi(\mathbf{n}) = e^{-\beta U_0(\mathbf{n})}$ where $U_0(\mathbf{n})$ is optimal utility under zero taxes with $\beta = 10.9$ for utility function 19 and $\beta = 8.1$ for utility function 21, so that the marginal social welfare of giving the lowest income household a dollar is roughly 100,000 times higher than giving the highest income household a dollar for both utility functions.

For this choice of $f(\mathbf{n}), \psi(\mathbf{n})$ and both utility functions 19 and 21, we find that it is optimal for the government to set a tax schedule which induces "bunching" at the bottom of the income distribution, so that optimal $\mathbf{n} \mapsto \mathbf{z}$ is not a diffeomorphism. This can be observed by looking at the determinant of the Jacobian for the optimal mapping $\mathbf{n} \mapsto \mathbf{z}$; we see in Figure 3 that this Jacobian determinant is 0 at the bottom of the skill distribution, implying that $\mathbf{n} \mapsto \mathbf{z}$ is not invertible near the bottom of the income distribution.[35] This bunching phenomenon was shown to be robust in the context of a multiproduct monopolist by Rochet and Chone (1998) due to a tension between participation constraints and second order conditions; our numerical examples show that bunching also appears to be simple to generate in the context of multidimensional optimal taxation problems which do not feature participation constraints (as we make the standard assumption that individuals cannot leave the country). Through trial and error, it appears bunching occurs whenever (a) welfare weights for low income households are sufficiently large relative to high income households and/or (b) the density of types has sufficient curvature. This seems roughly consistent with the results on bunching in unidimensional settings discussed in Simula and Trannoy (2020).

To confirm that solutions from our method (i.e., solving System 25) are generating the correct optimal schedule, we can check that our results for utility function 19 match with the results computed using the method of Aguilera and Morin (2008), which utilizes the convexity characterization of incentive compatibility from Rochet (1987).[36] We show the difference in

[35] As discussed, our numerical procedure to compute optimal schedules when utility is given by Equation 19 or 21 enforces that $\mathbf{n} \mapsto \mathbf{z}$ is a diffeomorphism because second order conditions holding imply that $\mathbf{n} \mapsto \mathbf{z}$ is a diffeomorphism by Proposition 1. Thus, with our numerical procedure there is never exact bunching; our algorithm instead simply approximates bunching arbitrarily well. For instance, the Jacobian determinant is on the order of 10^{-17} in Figure 3.

[36] The method of Aguilera and Morin (2008) essentially boils down to non-linear semidefinite programming:



(a) Utility Function 19

(b) Utility Function 21

Figure 3: Jacobian Determinant, Solution to System 25 for Second Set of Input Data

Note: This figure shows the Jacobian determinant $\frac{\partial z_1}{\partial n_1}\frac{\partial z_2}{\partial n_2} - \frac{\partial z_2}{\partial n_1}\frac{\partial z_1}{\partial n_2}$ from solving System 25 for utility functions 19 (panel 3a) and 21 (panel 3b). $\theta_1 = \theta_2 = 3$ (corresponding to a compensated taxable income elasticity of 0.33 with no taxes) and $\alpha = 50$. $f(\mathbf{n})$ is joint normal with mean matrix $\mu = \begin{bmatrix} -5.5 \\ -5.5 \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} 5 & 0.5 \\ 0.5 & 5 \end{bmatrix}$ on a rectangular domain of $[-6, -0.5]^2$ and set $\psi(\mathbf{n}) = e^{-\beta U_0(\mathbf{n})}$ where $U_0(\mathbf{n})$ is optimal utility under zero taxes with $\beta = 10.9$ for utility function 19 and $\beta = 8.1$ for utility function 21, so that the marginal social welfare of giving the lowest income household a dollar is roughly 100,000 times higher than giving the highest income household a dollar for both utility functions.

average tax rates between our method and the method of Aguilera and Morin (2008) in Figure 9 in Appendix C.3; they are very close and the small differences between the solutions shrink with the grid size, suggesting they are simply numerical noise.

5.4 A More Realistic Example: Optimal Taxation of Couples

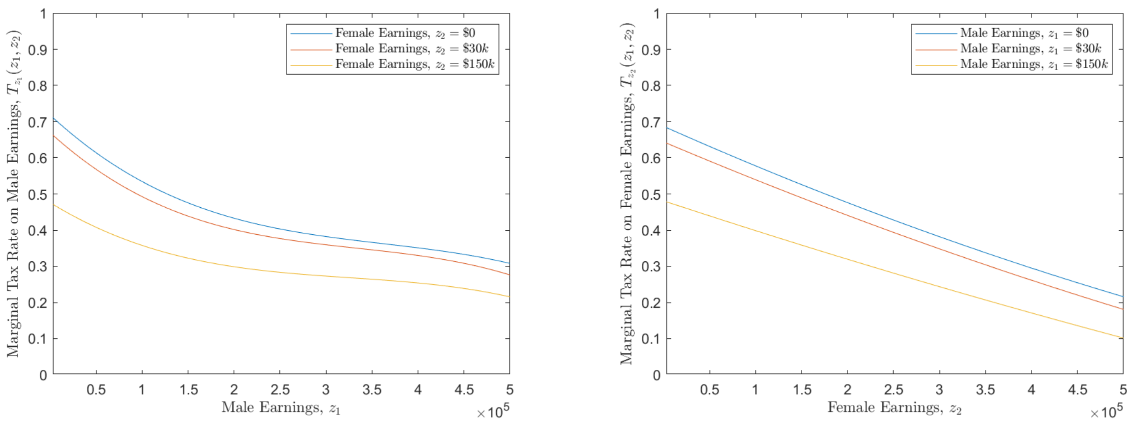
The examples in Section 5.3 are not intended to be realistic: they are merely meant to showcase how to apply our theoretical results from Section 3 and how to use our novel numerical method to solve multidimensional screening problems. Next, we show how these methods can be applied to a somewhat more realistic, calibrated setting: optimal taxation of couples. Optimal taxation of couples has been studied in the public finance literature (e.g., Kleven, Kreiner and Saez (2009), Spiritus et al. (2022), or Krasikov and Golosov (2022)). Our contribution here is to (1) allow the problem to be fully multidimensional (e.g., Kleven, Kreiner and Saez (2009) assume that women's labor supply is dichotomous and that all women have the same productivity) and (2) to allow generally for bunching (e.g., Spiritus et al. (2022) and Krasikov and Golosov (2022) assume there is no bunching so that the tax schedule can be found using first order methods). We use income data from the Current Population Survey in 2019 and suppose utility is again quadratic in type $\mathbf{n}$:

$$u(T, \mathbf{z}; \mathbf{n}) = \log(z_1 + z_2 + T) + n_1 \frac{z_1^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2^{1+\theta_2}}{1+\theta_2} - \frac{n_1^2}{2\alpha_1} z_1 - \frac{n_2^2}{2\alpha_2} z_2 \tag{26}$$

We assume utility over consumption, $z_1 + z_2 + T$, is $\log(z_1 + z_2 + T)$, consistent with the findings of Chetty (2006) on the curvature of utility over consumption. The parameters $n_1, n_2 < 0$ capture differences in disutility of producing income: larger (i.e., less negative) values capture

solve System 20 by finding the utility function $U^*(\mathbf{n})$ which numerically optimizes welfare subject to the budget constraint and the convexity constraint (i.e., the discrete Hessian matrix of $U^*(\mathbf{n})$ is positive semi-definite).

lower disutility of labor (loosely, we can think of n_1, n_2 as being related to productivity as more productive individuals have to exert less effort to generate a given level of income). Note that this utility function is neither linear in consumption nor linear in type, which, to the best of my knowledge, takes us outside of the realm of any previous results on incentive compatibility.[37] Disutility of generating income is given by: $D(z_1, z_2; n_1, n_2) = n_1 \frac{z_1^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2^{1+\theta_2}}{1+\theta_2} - \frac{n_1^2}{2\alpha_1} z_1 - \frac{n_2^2}{2\alpha_2} z_2$. Hence, the utility function is augmented from the standard iso-elastic form to include additional terms which capture positive marginal disutility of labor supply even when labor supply is zero (i.e., $\partial D(z_1, z_2; n_1, n_2)/\partial z_i|_{z_1=z_2=0} \neq 0$ for $i = 1, 2$). This leads to some individuals optimally choosing to not work, which is an empirically relevant modification. The distribution of types $f(n_1, n_2)$ is calibrated to match the empirical joint income distribution of couples.[38] We choose $\theta_1, \theta_2, \alpha_1$, and α_2 to match four moments: the compensated taxable income elasticity for men (0.2, taken from Blomquist and Selin (2010)), the compensated taxable income elasticity for women (1, also taken from Blomquist and Selin (2010)), the percentage of men who do not work (13.5%, from CPS data), and the fraction of women who do not work (20%, from CPS data).[39] We assume that the government has no exogenous expenditure requirements. Finally, we suppose $W(U(\mathbf{n}), \mathbf{n}) = \psi(\mathbf{n})U(\mathbf{n})$ with $\psi(\mathbf{n})$ decreasing in $\mathbf{n}$ so that the government desires to redistribute to those with high disutilities of generating income. For purposes of illustration, we choose welfare weights $\psi(\mathbf{n})$ so that $\psi(\mathbf{n})$ is $\approx$ 10,000 times higher for households earning \$1 million per year than for households who earn \$0 per year. Figure 4 shows marginal tax rates over the income distribution for males and females conditional on a given level of male (or female) income.



(a) Marginal Tax Schedules for Males

(b) Marginal Tax Schedules for Females

Figure 4: Optimal Marginal Tax Schedules for Couples

Note: This figure shows the optimal marginal tax rates for males and females conditional on spousal earnings. We assume utility is given by Equation 26. $f(\mathbf{n})$ is calibrated to match the joint income distribution from the CPS and $\theta_1, \theta_2, \alpha_1, \alpha_2$ are chosen to match four moments: the compensated taxable income elasticity for men (0.2, taken from Blomquist and Selin (2010)), the compensated taxable income elasticity for women (1, also taken from Blomquist and Selin (2010)), the percentage of men who do not work (13.5%, from CPS data), and the fraction of women who do not work (20%, from CPS data). The social welfare function is given by $W(U(\mathbf{n}), \mathbf{n}) = \psi(\mathbf{n})U(\mathbf{n})$ with welfare weights $\psi(\mathbf{n})$ chosen so that $\psi(\mathbf{n})$ is $\approx 10,000$ times higher for the lowest income household than for the highest income household.

[37] Utility function 26 satisfies Assumption 1 as the Jacobian matrix of $\mathbf{n} \mapsto \frac{\nabla_{\mathbf{z}} u(T, \mathbf{z}; \mathbf{n})}{u_T(T, \mathbf{z}; \mathbf{n})}$ equals:

$$(z_1 + z_2 + T) \begin{bmatrix} z_1^{\theta_1} - \frac{n_1}{\alpha} & 0 \\ 0 & z_2^{\theta_2} - \frac{n_2}{\alpha} \end{bmatrix}$$

which is a P matrix given that $(z_1 + z_2 + T), z_1, z_2 > 0$ and $n_1, n_2 < 0$.

[38] This is similar to Saez (2001) who calibrates a unidimensional $f(n)$ to match the empirical income distribution.

[39] We provide more detail about the calibration in Appendix C.

As far as marginal tax rates, top earning females are typically taxed at a lower marginal tax rate than top earning males because females have higher elasticities, on average, than men. For example, the optimal marginal tax rate for women earning $500,000 per year whose husband does not work is around 21% whereas the optimal marginal tax rate for men earning $500,000 per year whose wife does not work is around 31%. We find that at most incomes (z_1, z_2) the tax schedule features negative jointness: $\frac{\partial^2 T(z_1,z_2)}{\partial z_1 \partial z_2} < 0$, but there are also portions of the distribution where this is not the case.

We find that optimal average tax rates are around 30-40% for relatively high income couples (i.e., those with combined income over $200,000 per year).[40] This is combined with large transfers for low income couples (i.e., those making combined incomes less than $20,000 per year): we find it is optimal to transfer around $50,000 per year to these households. This should be interpreted as government provision of public goods as well as direct transfers via the tax system. This large benefit is fully taxed away roughly by the point a household earns $70,000 per year. The optimal solution features substantial bunching at the bottom of the distribution whereby many households do not work at all. This can be seen in Figure 5, which shows total household income across the type distribution (note we plot household income against $(-\log(-n_1), -\log(-n_2))$ to compress the type distribution for readability); there are a substantial number of households with a combined income of $0 (which means both members do not work as income cannot be negative).[41]

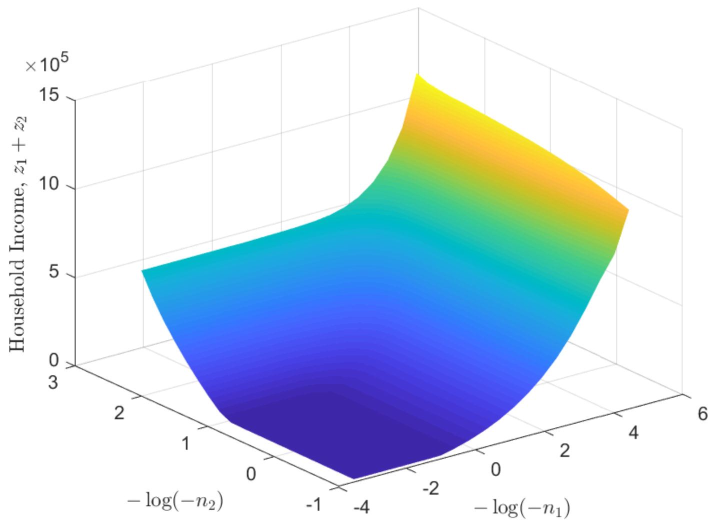


Figure 5: Optimal Total Household Income Across the Type Distribution

Note: This figure shows the optimal household income by type, assuming utility is given by Equation 26. We plot household income against $(-\log(-n_1), -\log(-n_2))$ to compress the type distribution for readability. $f(\mathbf{n})$ is calibrated to match the joint income distribution from the CPS and $\theta_1, \theta_2, \alpha_1, \alpha_2$ are chosen to match four moments: the compensated taxable income elasticity for men (0.2, taken from Blomquist and Selin (2010)), the compensated taxable income elasticity for women (1, also taken from Blomquist and Selin (2010)), the percentage of men who do not work (13.5%, from CPS data), and the fraction of women who do not work (20%, from CPS data). The social welfare function is given by $W(U(\mathbf{n}), \mathbf{n}) = \psi(\mathbf{n})U(\mathbf{n})$ with welfare weights $\psi(\mathbf{n})$ chosen so that $\psi(\mathbf{n})$ is $\approx 10,000$ times higher for the lowest income household than for the highest income household.

Ultimately, we hope that this basic simulation using CPS data not only highlights the methods developed in this paper via a somewhat realistic application, but also can be used as a

[40] The average tax rate surface (analogous to Figures 1 and 2) is shown in Appendix C.3.

[41] Figure 8 in Appendix C shows the corresponding Jacobian determinant for this simulation.

starting point towards even more involved work on multidimensional taxation.

6 Conclusion

This paper has derived necessary conditions and sufficient conditions for incentive compatibility in the context of a general multidimensional screening problem assuming a generalized single crossing property. We then used these results to derive a novel numerical method to solve multidimensional screening problems, illustrating the method with a number of numerical examples in the context of optimal multidimensional taxation, finding that bunching in the optimal allocation appears to be relatively easy to generate. We apply this method to data from the CPS to better understand optimal couples taxation, finding that it is optimal to have significant bunching in the form of unemployment at the bottom of the income distribution.

Looking forward, we believe that results in this paper can be used to better understand other multidimensional screening problems in the areas of, for example, non-linear pricing or public procurement. We think our analysis of optimal multidimensional taxation suggests that bunching behavior is perhaps more relevant for optimal taxation than previously believed. Thus, a more thorough investigation into importance of bunching for multidimensional settings is an important area for further work.

References

Aguilera, Néstor E., and Pedro Morin. 2008. "Approximating optimization problems over convex functions." *Numerische Mathematik*, 111(1): 1–34.

Armstrong, Mark. 1996. "Multiproduct nonlinear pricing." *Econometrica*, 64: 51–75.

Baron, David P., and Roger B. Myerson. 1982. "Regulating a monopolist with unknown costs." *Econometrica*, 50(4): 911.

Basov, Suren. 2001. "Hamiltonian approach to multi-dimensional screening." *Journal of Mathematical Economics*, 36(1): 77–94.

Bergstrom, Katy, and William Dodds. 2021. "Optimal Taxation with Multiple Dimensions of Heterogeneity." *Journal of Public Economics*, 200: 104442.

Blomquist, Sören, and Håkan Selin. 2010. "Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates." *Journal of Public Economics*, 94: 878–889.

Boerma, Job, Aleh Tsyvinski, and Alexander Zimin. 2022. "Bunching and taxing multidimensional skills." *SSRN Electronic Journal.*

Brezis, Haim. 2011. *Functional analysis, Sobolev spaces and partial differential equations.* Springer.

Carlier, Guillaume. 2001. "A general existence result for the principal-agent problem with adverse selection." *Journal of Mathematical Economics*, 35(1): 129–150.

Chetty, Raj. 2006. "A New Method of Estimating Risk Aversion." *American Economic Review*, 96(5): 1821–1834.

Chiappori, Pierre-Andre, Robert McCann, and Brendan Pass. 2017. "Multidimensional matching."

Chone, Philippe, and Guy Laroque. 2010. "Negative Marginal Tax Rates and Heterogeneity." *American Economic Review*, 100: 2532–2547.

DeVore, Ronald A., and Xiang Ming Yu. 1985. "Pointwise estimates for monotone polynomial approximation." *Constructive Approximation*, 1(1): 323–331.

Encyclopedia of Mathematics. 2022. "Diffeomorphism." In *Encyclopedia of Mathematics.* EMS Press.

Gale, David, and Hukukane Nikaido. 1965. "The Jacobian Matrix and Global Univalence of Mappings." *Math. Annalen*, 159: 81–93.

Grabovsky, Yury, and Lev Truskinovsky. 2010. "Roughening instability of broken extremals." *Archive for Rational Mechanics and Analysis*, 200(1): 183–202.

Hencl, Stanislav. 2011. "Sobolev homeomorphism with zero jacobian almost everywhere." *Journal de Mathématiques Pures et Appliquées*, 95(4): 444–458.

Hirsch, Morris W. 1988. *Differential topology.* Springer.

Ito, Kazufumi. 2020. *Functional Analysis and Optimization.*

Jacquet, Laurence, and Etienne Lehmann. 2015. "Optimal Income Taxation when Skills and Behavioral Elasticities are Heterogeneous." *CESifo Working Paper Series 5265, CESifo.*

Jacquet, Laurence, and Etienne Lehmann. 2020. "Optimal Income Taxation with Composition Effects." *Journal of the European Economic Association.*

Kinderlehrer, David, and Guido Stampacchia. 1980. *An introduction to variational inequalities and their applications.*

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez. 2009. "The optimal income taxation of couples." *Econometrica*, 77(2): 537–560.

Krasikov, Ilia, and Mikhail Golosov. 2022. "Multidimensional Screening in Public Finance: The Optimal Taxation of Couples."

Laffont, Jean Jacques, and Jean Tirole. 1994. *A theory of incentives in procurement and Regulation.* MIT Press.

Lockwood, Benjamin B., and Matthew Weinzierl. 2016. "De Gustibus non est Taxandum: Heterogeneity in preferences and optimal redistribution." *Journal of Public Economics*, 124: 74–80.

Matkowski, Janusz. 2012. "Mean-value theorem for vector-valued functions." *Mathematica Bohemica*, 137(4): 415–423.

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

Milgrom, Paul, and Ilya Segal. 2016. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica*, 70: 583–601.

Mirrlees, James. 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies*, 38: 175–208.

Mirrlees, James. 1976. "Optimal Tax Theory: A Synthesis." *Journal of Public Economics*, 6(4): 327–358.

Mirrlees, James. 1986. "The Theory of Optimal Taxation." *Handbook of Mathematical Economics.*

Mussa, Michael, and Sherwin Rosen. 1978. "Monopoly and product quality." *Journal of Economic Theory*, 18(2): 301–317.

Mérigot, Quentin, and Édouard Oudet. 2014. "Handling convexity-like constraints in variational problems." *SIAM Journal on Numerical Analysis*, 52(5): 2466–2487.

Oberman, Adam M. 2013. "A numerical method for variational problems with convexity constraints." *SIAM Journal on Scientific Computing*, 35(1).

Rochet, Jean-Charles. 1987. "A necessary and sufficient condition for rationalizability in a quasi-linear context." *Journal of Mathematical Economics*, 16(2): 191–200.

Rochet, Jean-Charles, and Philippe Chone. 1998. "Ironing, sweeping, and multidimensional screening." *Econometrica*, 66(4): 783.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies*, 68: 205–229.

Scheuer, Florian, and Ivan Werning. 2016. "Mirrlees meets Diamond-Mirrlees."

Simula, Laurent, and Alain Trannoy. 2020. "Bunching and rank-dependent optimal income taxation." *SSRN Electronic Journal.*

Spiritus, Kevin, Etienne Lehmann, Sander Renes, and Floris Zoutman. 2022. "Optimal taxation with multiple incomes and types." *SSRN Electronic Journal.*

Tsatsomeros, Michael J. 2004. "Generating and Detecting Matrices with Positive Principal Minors." *Focus on Computational Neurobiology*, 115–132. USA:Nova Science Publishers, Inc.

Udriste, Constantin. 2009. "Simplified multitime maximum principle." *Balkan Journal of Geometry and Its Applications*, 14(1): 102–119.

Villani, Cedric. 2009. *Optimal Transport Old and New.* Springer Berlin Heidelberg.

A Proofs Appendix

A.1 Proof of Theorem 1

(1) We know that for any incentive compatible allocation, $U(\mathbf{n})$ is equal to:

$$U(\mathbf{n}) \equiv u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = \max_{\mathbf{n}'} \; u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$$

Because $u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ is differentiable in $\mathbf{n}$ $\forall \mathbf{n}'$, the envelope theorem (Corollary 1, Milgrom and Segal (2016)) implies that the following envelope condition holds for all paths between $\mathbf{n}_1$ and $\mathbf{n}_2$:[42]

$$U(\mathbf{n}_1) - U(\mathbf{n}_2) = \oint_{\mathbf{n}_2}^{\mathbf{n}_1} \nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \cdot d\mathbf{n}$$

(2) If $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z}(\mathbf{n}))$ is differentiable at a point $\mathbf{n}$, then Lemma 1 tells us that the tax schedule $T(\mathbf{z})$ is twice continuously differentiable at $\mathbf{z}(\mathbf{n})$ so that Equation 8 holds. Under any incentive compatible allocation, we know that $u(T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ must be concave in $\mathbf{z}$ around $\mathbf{z}(\mathbf{n})$.[43] Hence, $\nabla_{\mathbf{z}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})$ must be negative semi-definite. Equation 8 then implies that for any incentive compatible allocation, $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ is positive semi-definite. In order to show that $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ is positive definite, we simply need to show that $\det[\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}] \neq 0$.

We know that $\det([\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}) \neq 0$ because $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism. We claim that $\det(\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})) \neq 0$ by Assumption 1. Using Equation 6, we have:

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) = \nabla_{\mathbf{n}} \left(u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \right) \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{n}} \left(\nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \right)$$

Using the fact that $u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = 0$, it is straightforward to verify that:

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) = u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \left[\nabla_{\mathbf{n}} \left(\frac{\nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})}{u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})} \right) \right] \tag{27}$$

But

$$\nabla_{\mathbf{n}} \left(\frac{\nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})}{u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})} \right)$$

is simply the Jacobian of the mapping $\mathbf{n} \mapsto \frac{\nabla_{\mathbf{z}} u(T, \mathbf{z}; \mathbf{n})}{u_T(T, \mathbf{z}; \mathbf{n})}\big|_{T=T(\mathbf{z}(\mathbf{n})), \mathbf{z}=\mathbf{z}(\mathbf{n})}$, which is a diffeomorphism by Assumption 1. Hence:

$$\det \left[\nabla_{\mathbf{n}} \left(\frac{\nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})}{u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})} \right) \right] \neq 0$$

because the determinant of the Jacobian of a diffeomorphism never vanishes. This implies then that:

$$\det \left[\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) \right] \neq 0$$

[42] To apply Corollary 1 from Milgrom and Segal (2016) we also need that the gradient $\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ is bounded $\forall \mathbf{n}, \mathbf{n}'$. But this holds because we assume the domain $\mathbf{N}$ is compact and that the set of all assigned $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ is compact. Because $u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ is assumed continuously differentiable in $\mathbf{n}$, we have that $\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ is a continuous function on a compact domain, which implies its image is also compact and therefore bounded.

[43] If not, $u(T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ is increasing in the direction of some $\mathbf{z}$, which means $u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ is increasing in the direction of some $\mathbf{n}'$ by the fact that $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism.

by Equation 27 because $u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) > 0$. But then we know that:

$$\det[\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}] = \det\left[\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})\right] \det([\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}) \neq 0$$

(3) Suppose that an allocation is such that two points $\mathbf{n}$ and $\mathbf{n}'$ are mapped to the same $\mathbf{z}$ at which $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z}(\mathbf{n}))$ is differentiable. Hence, the FOC must be satisfied for both $\mathbf{n}$ and $\mathbf{n}'$:

$$u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = 0$$

$$u_T(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}') \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n}')) + \nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}') = 0$$

But Assumption 1 then implies that $\nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n})) \neq \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n}'))$, which means that the proposed allocation requires two individuals $\mathbf{n}, \mathbf{n}'$ with $\mathbf{z}(\mathbf{n}) = \mathbf{z}(\mathbf{n}')$ to face different marginal transfer rates. Clearly, this cannot be achieved with any $T(\mathbf{z}(\mathbf{n}))$, implying that the allocation is not incentive compatible.

A.2 Proof of Corollary 1.1

If $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z}(\mathbf{n}))$ is differentiable in $\mathbf{z}$ at two points $\mathbf{n}$ and $\mathbf{n}'$, then we know that the tax schedule is twice continuously differentiable in $\mathbf{z}$ there by Lemma 1 so that Equation 8 holds. Assumption 1 tells us that $\det\left[\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})\right]$ never changes sign (see discussion in point (2) of Appendix A.1). Hence, if $\det[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]$ has different signs at $\mathbf{n}$ and $\mathbf{n}'$, then we know $\det[\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}]$ has different signs at $\mathbf{n}$ and $\mathbf{n}'$, which means $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ cannot be positive definite at both $\mathbf{n}$ and $\mathbf{n}'$. Applying Theorem 1 then yields the desired conclusion.

A.3 Proof of Theorem 2

First, we show that if $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable everywhere, then the envelope condition 4 holding, $\mathbf{n} \mapsto \mathbf{z}$ being injective, and all types $\mathbf{n}$ satisfying $u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})\ \forall \mathbf{n}' \in \partial \mathbf{N}$ is sufficient for incentive compatibility.

Because $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable everywhere, all individuals $\mathbf{n}$ must have:

$$\left[u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})\right] \nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n}) = 0$$

Because $\mathbf{n} \mapsto \mathbf{z}$ is injective, Assumption 1 tells us that for all $\mathbf{n}, \mathbf{n}'$:

$$u_T(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}) \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n}')) + \nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}) \neq 0$$

Because $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism, we also know that:[44]

$$\left[u_T(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}) \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n}')) + \nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})\right] \nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n}') \neq 0$$

Hence, we know that type $\mathbf{n}$ has no other critical points $(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'))$. Because $u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ is a continuous function of $\mathbf{n}'$ and $\mathbf{N}$ is compact, the global maximum for type $\mathbf{n}$ can occur only at $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ or on the boundary $\partial \mathbf{N}$. By assumption, all $\mathbf{n}$ satisfy: $u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})\ \forall \mathbf{n}' \in \partial \mathbf{N}$. Hence, the global maximum must occur at $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$.

[44] Otherwise, $\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n}')$ would have a non-zero eigenvector, which would mean that it is not invertible, violating the fact that the Jacobian matrix of a local diffeomorphism is everywhere invertible.

Next, we show that if $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable everywhere, then the envelope condition 4 holding, $\mathbf{n} \mapsto \mathbf{z}$ being injective, and all types $\mathbf{n}$ satisfying $u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ for $\mathbf{n}' \in \partial\mathbf{N}$ is necessary for incentive compatibility. Corollary 1.2 shows that if $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable everywhere, then the envelope condition holding and $\mathbf{n} \mapsto \mathbf{z}$ being injective are both necessary for incentive compatibility. All types $\mathbf{n}$ satisfying $u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ for $\mathbf{n}' \in \partial\mathbf{N}$ is definitionally necessary for incentive compatibility, which proves necessity of these conditions.

Finally, if we have a sequence of $(T_j(\mathbf{z}(\mathbf{n})), \mathbf{z}_j(\mathbf{n}))$ that all satisfy the conditions of Theorem 2 (and hence are incentive compatible), we know that $\forall \mathbf{n}, \mathbf{n}'$:

$$u(T_j(\mathbf{z}_j(\mathbf{n})), \mathbf{z}_j(\mathbf{n}); \mathbf{n}) \geq u(T_j(\mathbf{z}_j(\mathbf{n}')), \mathbf{z}_j(\mathbf{n}'); \mathbf{n})$$

Taking limits and passing through the continuous functions $u(T, z; \mathbf{n})$ and $T(\mathbf{z})$, we see that:[45]

$$u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$$

A.4 Incentive Compatibility when $\dim(\mathbf{N}) \neq \dim(\mathbf{Z})$

Proposition 4. *Suppose $dim(\mathbf{N}) > dim(\mathbf{Z})$ and suppose we can split the domain $\mathbf{N}$ into some $\mathbf{N}^{(1)}$ and $\mathbf{N}^{(2)}$ with $dim\left(\mathbf{N}^{(1)}\right) = dim(\mathbf{Z})$ and, for each $\mathbf{n}^{(2)} \in \mathbf{N}^{(2)}$, Assumption 1 holds for $\mathbf{n}^{(1)} \in \mathbf{N}^{(1)}$, i.e.:*

$$\mathbf{n}^{(1)} \mapsto \frac{\nabla_{\mathbf{z}} u\left(T, \mathbf{z}; \mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)}{u_T\left(T, \mathbf{z}; \mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)}$$

is a diffeomorphism $\forall \mathbf{n}^{(2)}$.

In order for this allocation to be incentive compatible: (1) the following envelope condition must hold for all paths between $\mathbf{n}^{(1)\prime}$ and $\mathbf{n}^{(1)}$ and all $\mathbf{n}^{(2)}$:

$$U\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right) - U\left(\mathbf{n}^{(1)\prime}, \mathbf{n}^{(2)}\right) = \oint_{\mathbf{n}^{(1)\prime}}^{\mathbf{n}^{(1)}} \nabla_{\mathbf{n}^{(1)}} u\left(T\left(\mathbf{z}\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)\right), \mathbf{z}\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right); \mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right) \cdot d\mathbf{n}^{(1)} \quad (28)$$

(2) for all $(\mathbf{n}^{(1)}, \mathbf{n}^{(2)})$ such that the allocation $\mathbf{n}^{(1)} \mapsto \mathbf{z}$ is a local diffeomorphism at some $\mathbf{n}^{(1)}$ and $T(\mathbf{z})$ is differentiable, $\nabla_{\mathbf{n}^{(1)}} FOC\left(\mathbf{z}\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right); \mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right) \left[\nabla_{\mathbf{n}^{(1)}} z\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)\right]^{-1}$ must be positive definite, (3) for all $(\mathbf{n}^{(1)}, \mathbf{n}^{(2)})$ and $(\mathbf{n}^{(1)\prime}, \mathbf{n}^{(2)})$ such that the allocation $\mathbf{n}^{(1)} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable we must have $\mathbf{z}(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}) \neq \mathbf{z}(\mathbf{n}^{(1)\prime}, \mathbf{n}^{(2)})$. Similarly, if $dim(\mathbf{N}) < dim(\mathbf{Z})$, then if there exists a subset $\tilde{\mathbf{Z}} \subset \mathbf{Z}$ with $dim(\mathbf{N}) = dim(\tilde{\mathbf{Z}})$ such that the conditions of Assumption 1 and Theorem 1 hold after replacing $\mathbf{z} \in \mathbf{Z}$ with $\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}$, then the allocation is not incentive compatible.

Proof. When $\dim(\mathbf{N}) > \dim(\mathbf{Z})$, the above follows immediately from Theorem 1: if any of the stated conditions fail to hold, then some individual $\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)$ prefers a bundle chosen by some type $\left(\mathbf{n}^{(1)\prime}, \mathbf{n}^{(2)}\right)$.

When $\dim(\mathbf{N}) < \dim(\mathbf{Z})$, we know that $\mathbf{n} \mapsto \mathbf{z}$ definitionally cannot be diffeomorphic, but $\mathbf{n} \mapsto \mathbf{z}$ can be diffeomorphic if we just restrict the set of $\mathbf{z}$'s to an appropriate subset, allowing us to again apply Theorem 1. ☐

[45] Note, we require the limit to be uniform so that we can apply the uniform limit theorem to ensure that $T(\mathbf{z}(\mathbf{n})) = \lim_{j\to\infty} T_j(\mathbf{z}(\mathbf{n}))$ is continuous.

Proposition 5. *Suppose dim*$(\mathbf{N}) >$ *dim*$(\mathbf{Z})$*, but suppose we can split the domain* $\mathbf{N}$ *into some* $\mathbf{N}^{(1)}$ *and* $\mathbf{N}^{(2)}$ *with dim* $\left(\mathbf{N}^{(1)}\right) =$ *dim*$(\mathbf{Z})$ *such that for each* $\mathbf{n}^{(2)} \in \mathbf{N}^{(2)}$*, Assumption 1 holds after replacing* $\mathbf{n}$ *with* $\mathbf{n}^{(1)}$ *(as in Proposition 5). Suppose we consider an allocation such that* $\mathbf{n}^{(1)} \mapsto \mathbf{z}$ *is a local diffeomorphism* $\forall \mathbf{n}^{(2)}$ *and* $T(\mathbf{z})$ *is differentiable. Then sufficient conditions for incentive compatibility are: the envelope condition 28 holds* $\forall \left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)$*, and* $\mathbf{n}^{(1)} \mapsto \mathbf{z}$ *is bijective* $\forall \mathbf{n}^{(2)}$*, and* $u\left(T\left(\mathbf{z}\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)\right), \mathbf{z}\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right); \mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right) \geq u\left(T\left(\mathbf{z}\left(\mathbf{n}^{(1)\prime}, \mathbf{n}^{(2)}\right)\right), \mathbf{z}\left(\mathbf{n}^{(1)\prime}, \mathbf{n}^{(2)}\right); \mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)$ $\forall \mathbf{n}^{(1)\prime} \in \partial \mathbf{N}^{(1)}$ *and* $\forall \left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)$*. Similarly, if dim*$(\mathbf{N}) <$ *dim*$(\mathbf{Z})$*, then if there exists a subspace* $\tilde{\mathbf{Z}} \in \mathbf{Z}$ *with dim*$(\mathbf{N}) =$ *dim*$(\tilde{\mathbf{Z}})$ *such that the conditions of Assumption 1 and Theorem 2 hold after replacing* $\mathbf{z} \in \mathbf{Z}$ *with* $\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}$*, then the allocation is incentive compatible.*

Proof. When $\dim(\mathbf{N}) > \dim(\mathbf{Z})$, under the conditions listed above Theorem 2 tells us that each type $\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)$ prefers his/her assigned bundle to the bundles assigned to all other types types $\left(\mathbf{n}^{(1)\prime}, \mathbf{n}^{(2)}\right)$. Moreover, some type $\left(\mathbf{n}^{(1)\prime}, \mathbf{n}^{(2)}\right)$ chooses every $\mathbf{z}$ that is chosen by any $\left(\mathbf{n}^{(1)\prime}, \mathbf{n}^{(2)\prime}\right)$ because we now require that $\mathbf{n}^{(1)} \mapsto \mathbf{z}$ is bijective $\forall \mathbf{n}^{(2)}$.[46] Thus, we know that each type $\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)$ prefers his/her assigned bundle to the bundles assigned to all other types types $\left(\mathbf{n}^{(1)\prime}, \mathbf{n}^{(2)\prime}\right)$.

When $\dim(\mathbf{N}) < \dim(\mathbf{Z})$, we know that $\mathbf{n} \mapsto \mathbf{z}$ definitionally cannot be diffeomorphic, but $\mathbf{n} \mapsto \mathbf{z}$ can be diffeomorphic if we just restrict the set of $\mathbf{z}$'s to an appropriate subset, which allows us to again just apply Theorem 2. □

A.5 Proof of Remark 3

First, suppose $T(z)$ is differentiable, $n \mapsto z$ is locally diffeomorphic (i.e., monotonic), $U(n)$ satisfies the envelope condition, and $z'(n) > 0$. When $T(z)$ is differentiable, $n \mapsto z$ is locally diffeomorphic (i.e., monotonic), and $U(n)$ satisfies the envelope condition, the proof to Theorem 2 shows that each individual has a unique critical point for their utility maximization problem. But we also know that all individuals' second order conditions hold strictly when $z'(n) > 0$ (see footnote 13), implying the unique critical point is a local maximum. The mean value theorem implies that a local maximum which is the unique critical point of a real differentiable function is a global maximum. Hence, all individuals prefer their assigned bundle to boundary bundles.

On the other hand, if all individuals prefer their bundle to boundary bundles, we know that their second order condition holds strictly (see Remark 2), which in turn implies that $z'(n) > 0$ (see footnote 13).

[46]Hence, in order to apply Theorem 2 to settings wherein $\dim(\mathbf{N}) > \dim(\mathbf{Z})$, we require that for every $\mathbf{n}_2$, we have that $\mathbf{n}^{(1)} \mapsto \mathbf{z}$ is *surjective* onto the set of chosen incomes. While this limits the applicability of the result to some degree, it may often be possible to artificially enlarge the domain $\mathbf{N}$ so that this condition holds.

A.6 Proof that 10 is positive definite $\iff$ 11 is positive definite

We will show that:

$$V\left[\nabla_{\mathbf{n}}\mathbf{z}\right]^{-1} \equiv \begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1} \text{ is positive definite}$$

$$\iff V\nabla_{\mathbf{n}}\mathbf{z} \equiv \begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix} \text{ is positive definite}$$

We first prove that $V\left[\nabla_{\mathbf{n}}\mathbf{z}\right]^{-1}$ has all principal minors positive (i.e., is a P matrix) if and only if $V\nabla_{\mathbf{n}}\mathbf{z}$ has all principal minors positive. Then we will prove that $V\left[\nabla_{\mathbf{n}}\mathbf{z}\right]^{-1}$ is symmetric if and only if $V\nabla_{\mathbf{n}}\mathbf{z}$ is symmetric. First, note that:

$$\begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1}$$

is a P matrix (i.e., has positive principal minors) if and only if:

$$\begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1}$$

is a P matrix. The if statement follows immediately from the fact that the product of a P matrix and a diagonal matrix with positive entries is a P matrix.[47] The only if statement follows immediately as well because:

$$\begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1} = \begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix}^{-1} \left(\begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1} \right)$$

Next, we use the fact that:

$$\begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix} \text{ is a P matrix} \iff \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1} \text{ is a P matrix}$$

Hence we have:

$$\begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1} \text{ is a P matrix}$$

$$\iff \begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix} \text{ is a P matrix}$$

[47]This is a standard result; see Theorem 3.1 of Tsatsomeros (2004) for a proof.

Finally, we need to show that:

$$V\left[\nabla_{\mathbf{n}}\mathbf{z}\right]^{-1} \equiv \begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1} \text{ is symmetric}$$

$$\iff V\nabla_{\mathbf{n}}\mathbf{z} \equiv \begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix} \text{ is symmetric}$$

This follows essentially because the matrix V is symmetric. Lemma 3 in Appendix A.9 shows that the product VA of two invertible matrices V and A with V symmetric is symmetric if and only if VA^{-1} is symmetric. Hence, we know that:

$$\begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1} \text{ is symmetric}$$

$$\iff \begin{bmatrix} v_{z_1}(\mathbf{z}) & \cdots & 0 \\ & \vdots & \\ 0 & \cdots & v_{z_K}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix} \text{ is symmetric}$$

This proves the claim.

A.7 Proof of Lemma 2

By assumption, $U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ is differentiable in $\mathbf{n}$. Because $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism, we can express $\mathbf{n}$ locally as a function of $\mathbf{z}$ rather than the reverse to infer that $U(\mathbf{n}(\mathbf{z})) = u(T(\mathbf{z}), \mathbf{z}; \mathbf{n}(\mathbf{z}))$ is differentiable in $\mathbf{z}$. Because $u(T, \mathbf{z}; \mathbf{n})$ is continuously differentiable in all of its arguments and $u_T(T, \mathbf{z}; \mathbf{n}) > 0$, the implicit function theorem tells us that $T(\mathbf{z})$ is differentiable as a function of $\mathbf{z}$. As a result, we can differentiate both sides of $U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ to yield:

$$D_{\mathbf{n}}U(\mathbf{n}) = \nabla_{\mathbf{n}}u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) + \left\{u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})\nabla_{\mathbf{z}}T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}}u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})\right\}\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})$$

By the envelope condition 4, we then know that:

$$\left\{u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})\nabla_{\mathbf{z}}T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}}u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})\right\}\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n}) = 0$$

Given that $\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})$ is invertible (by the local diffeomorphism assumption), the following first order condition must also hold for each individual $\mathbf{n}$:

$$u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})\nabla_{\mathbf{z}}T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}}u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = 0$$

Rewriting $\mathbf{n}$ as a function of $\mathbf{z}$ rather than the reverse, the above first order condition defines:

$$\nabla_{\mathbf{z}}T(\mathbf{z}) = -\frac{\nabla_{\mathbf{z}}u(T(\mathbf{z}), \mathbf{z}; \mathbf{n}(\mathbf{z}))}{u_T(T(\mathbf{z}), \mathbf{z}; \mathbf{n}(\mathbf{z}))}$$

The right hand side is continuous in $\mathbf{z}$, which implies that $\nabla_{\mathbf{z}}T(\mathbf{z})$ is continuous in $\mathbf{z}$. As $\mathbf{n}(\mathbf{z})$ and $T(\mathbf{z})$ are both continuously differentiable in $\mathbf{z}$, we must have that $\nabla_{\mathbf{z}}T(\mathbf{z})$ is also continuously differentiable, which means that $T(\mathbf{z})$ is twice continuously differentiable.

A.8 Proof of Proposition 1

Note that when utility is given by $u(T, \mathbf{z}; \mathbf{n}) = u^{(0)}(T, \mathbf{z}) + \sum_i^K u^{(i)}(z_i, n_i)$, we have:

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1} = \begin{bmatrix} \frac{\partial^2 u^{(1)}(z_1, n_1)}{\partial z_1 \partial n_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial^2 u^{(K)}(z_K, n_K)}{\partial z_K \partial n_K} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) & \cdots \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \tag{29}$$

Now, Equation 29 has positive principal minors if and only if:

$$\begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) & \cdots \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{30}$$

is a P matix (i.e., all principal minors are positive). This if and only if follows because the product of a diagonal matrix with positive entries and a P matrix is a P matrix and the fact that a matrix is a P matrix if and only if its inverse is a P matrix. These are standard results; see Theorem 3.1 of Tsatsomeros (2004) for a proof.

Hence, if $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ is positive definite (and hence has positive principal minors), we know that Equation 30 is a P matrix. On a rectangular domain, Remark 1 then implies that $\mathbf{n} \mapsto \mathbf{z}$ is a diffeomorphism.

A.9 Proof of Proposition 2

In order for $\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ to be a conservative vector field, we must have that cross partials are equal, i.e., that the Jacobian of this vector field is symmetric. When $u(T, \mathbf{z}; \mathbf{n}) = u^{(0)}(T, \mathbf{z}) + \sum_i^K u^{(i)}(z_i, n_i)$, the Jacobian, J, is given by:

$$J(\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})) = \begin{bmatrix} \frac{\partial^2 u^{(1)}(z_1, n_1)}{\partial z_1 \partial n_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial^2 u^{(K)}(z_K, n_K)}{\partial z_K \partial n_K} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) & \cdots \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{31}$$

Now, if $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ is positive definite, we know that

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1} = \begin{bmatrix} \frac{\partial^2 u^{(1)}(z_1, n_1)}{\partial z_1 \partial n_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial^2 u^{(K)}(z_K, n_K)}{\partial z_K \partial n_K} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) & \cdots \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \tag{32}$$

is symmetric. But Equation 32 is symmetric if and only if Equation 31 is symmetric by the following Lemma:

Lemma 3. *The product VA of two invertible matrices V and A with V symmetric is symmetric if and only if VA^{-1} is symmetric.*

Proof. Let I be the identity matrix and $\mathcal{T}$ represent the transpose operator. Symmetry of VA and V means that $VA = [VA]^{\mathcal{T}} = A^{\mathcal{T}}V^{\mathcal{T}} = A^{\mathcal{T}}V$ (the last equality follows because V is symmetric. But then we have:

$$VA = [VA]^{\mathcal{T}} \iff VA = A^{\mathcal{T}}V \iff [A^{\mathcal{T}}]^{-1}V = VA^{-1} \iff [A^{-1}]^{\mathcal{T}}V = VA^{-1} \iff [VA^{-1}]^{\mathcal{T}} = VA^{-1}$$

□

Because the matrix

$$\begin{bmatrix} \frac{\partial^2 u^{(1)}(z_1,n_1)}{\partial z_1 \partial n_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial^2 u^{(K)}(z_K,n_K)}{\partial z_K \partial n_K} \end{bmatrix}$$

is symmetric, Lemma 3 implies that if Equation 32 is symmetric, then so is Equation 31. Hence, if $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n});\mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1}$ is positive definite and utility is of the form $u(T,\mathbf{z};\mathbf{n}) = u^{(0)}(T,\mathbf{z}) + \sum_i^K u^{(i)}(z_i,n_i)$, then $\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n})$ is a conservative vector field.

A.10 Proof of Proposition 3

The Jacobian matrix of $\mathbf{n} \mapsto \frac{\nabla_{\mathbf{z}} u(T,\mathbf{z};\mathbf{n})}{u_T(T,\mathbf{z};\mathbf{n})}$ equals:

$$\begin{bmatrix} z_1^{\theta_1} - \frac{n_1}{\alpha} & 0 \\ 0 & z_2^{\theta_2} - \frac{n_2}{\alpha} \end{bmatrix}$$

which is a P matrix as long as $z_1, z_2 > 0$ and $n_1, n_2 < 0$. Hence, the claim follows by Remark 1.

A.11 Existence of a Solution to Problem 18

In this Appendix, we prove the following existence result for the optimal multidimensional taxation problem:

Proposition 6. *The equations $\nabla_{\mathbf{n}} U(\mathbf{n}) = \nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n})$ and $U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n})$ define an a.e. correspondence $(U(\mathbf{n}),\nabla_{\mathbf{n}} U(\mathbf{n})) \mapsto (T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n}))$. There exists a solution to Problem 18 if for a.e. $\mathbf{n}$ and any selection from the correspondence $(U(\mathbf{n}),\nabla_{\mathbf{n}} U(\mathbf{n})) \mapsto (T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n}))$:*

$$-T(U(\mathbf{n}),\nabla_{\mathbf{n}} U(\mathbf{n})) \to -\infty \text{ as } ||\nabla_{\mathbf{n}} U(\mathbf{n})|| \to \infty \text{ and/or } U(\mathbf{n}) \to \infty$$

and

$$-T(U(\mathbf{n}),\nabla_{\mathbf{n}} U(\mathbf{n})) \not\to \infty \text{ as } U(\mathbf{n}) \to -\infty$$

and

$$W(U(\mathbf{n}),\mathbf{n}) \to -\infty \text{ as } U(\mathbf{n}) \to -\infty$$

Proof. We are going to argue that any maximizing sequence of functions $U_j^*(\mathbf{n})$ which satisfy the budget constraint must be bounded. First, let us recall the definition of the H^1 norm:

$$|U|_{H^1} = \int_{\mathbf{N}} U^2(\mathbf{n}) + ||\nabla_{\mathbf{n}} U(\mathbf{n})||^2 d\mathbf{n}$$

Next, note that as $|U|_{H^1} \to \infty$ we have either (1) $U(\mathbf{n}) \to \infty$ on a positive measure set, and/or (2) $U(\mathbf{n}) \to -\infty$ on a positive measure set, and/or (3) $||\nabla_{\mathbf{n}} U(\mathbf{n})|| \to \infty$ on a positive measure set. If $U(\mathbf{n}) \to \infty$ and/or $||\nabla_{\mathbf{n}} U(\mathbf{n})|| \to \infty$ on a positive measure set, we know that the budget constraint must not be satisfied. Defining:

$$BC \equiv \int_{\mathbf{N}} \{-T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})) - E\} f(\mathbf{n}) d\mathbf{n}$$

we know that, under the stated assumptions, $BC \to -\infty$ as $U(\mathbf{n}) \to \infty$ or $||\nabla_{\mathbf{n}} U(\mathbf{n})|| \to \infty$ on a positive measure set. Note, this holds true even if we also have $U(\mathbf{n}) \to -\infty$ on a positive measure set because:

$$-T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})) \to -\infty \text{ as } ||\nabla_{\mathbf{n}} U(\mathbf{n})|| \to \infty \text{ and/or } U(\mathbf{n}) \to \infty$$

and

$$-T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})) \not\to \infty \text{ as } U(\mathbf{n}) \to -\infty$$

Moreover, if neither $U(\mathbf{n}) \to \infty$ or $||\nabla_{\mathbf{n}} U(\mathbf{n})|| \to \infty$ on a positive measure set, yet $|U|_{H^1} \to \infty$, we know we must have $U \to -\infty$ on a positive measure set. But then by assumption, $\int_{\mathbf{N}} W(U(\mathbf{n}), \mathbf{n}) f(\mathbf{n}) d\mathbf{n} \to -\infty$, which clearly cannot be optimal as the government can always choose to set taxes equal to zero with $T(\mathbf{z}) = 0$, which will yield welfare greater than $-\infty$.

Hence, any maximizing sequence $U_j^*(\mathbf{n})$ of $\int_{\mathbf{N}} W(U(\mathbf{n}), \mathbf{n})) f(\mathbf{n}) d\mathbf{n}$ that satisfies the budget constraint must be bounded. The rest of the proof is standard because the Sobolev space $H^1(\mathbf{N})$ is closed (and the subspaces of functions which satisfy the global incentive compatibility constraints and the budget constraint are both closed, implying the intersection of these spaces is also closed) and the functional $\int_{\mathbf{N}} W(U, \mathbf{n}) f(\mathbf{n}) d\mathbf{n}$ is continuous. Essentially, we extract a weakly convergent subsequence from the bounded maximizing sequence and use the continuity of the functional to pass the limit inside to show that the functional evaluated at the limit of the minimizing sequence is at least as small as the infimum, proving the claim (see, e.g., Kinderlehrer and Stampacchia (1980) or Ito (2020), Chapter 3.6).

□

Proposition 6, while similar in flavor to results in Rochet and Chone (1998) and Basov (2001), is not exactly the same because the functional to maximize, $W(U, \mathbf{n})$, is not coercive in the H^1 norm. Hence, we use coercivity of the budget constraint to show that any maximizing sequence will be bounded; the rest of the proof is identical to standard existence proofs, e.g., Kinderlehrer and Stampacchia (1980).

Remark 4. *As an example application of Proposition 6, suppose that* $\mathbf{N} \subseteq (-\infty, 0)^K$ *and*

$$u(T, \mathbf{z}; \mathbf{n}) = \log\left(\sum_{i=1}^{K} z_i + T\right) + \sum_{i=1}^{K} n_i \frac{z_i^{1+\theta_i}}{1+\theta_i}$$

with $z_1, z_2, ..., z_K \geq 0$. *Then we have:*[48]

$$-T(U, \nabla_{\mathbf{n}} U) = \sum_{i=1}^{K} \left((1+\theta_i)\frac{\partial U}{\partial n_i}\right)^{\frac{1}{1+\theta_i}} - \exp\left(U - \sum_{i=1}^{K} n_i \frac{\partial U}{\partial n_i}\right) \to -\infty \text{ as } ||\nabla_{\mathbf{n}} U|| \to \infty \text{ or } U \to \infty$$

[48]The limit uses the fact that $\frac{\partial u}{\partial n_i} > 0 \; \forall i$.

And note that we have $-T(U, \nabla_{\mathbf{n}} U) \to \sum_{i=1}^{K} \left((1+\theta_i) \frac{\partial U}{\partial n_i} \right)^{\frac{1}{1+\theta_i}}$ *as* $U \to -\infty$*. If we have:*

$$\int_{\mathbf{N}} W\left(U(\mathbf{n}), \mathbf{n}\right) dF(\mathbf{n}) = \int_{\mathbf{N}} \psi(\mathbf{n}) U(\mathbf{n}) dF(\mathbf{n})$$

then $\int_{\mathbf{N}} W\left(U(\mathbf{n}), \mathbf{n}\right) dF(\mathbf{n}) \to -\infty$ *as* $U \to -\infty$*, so that the conditions for Proposition 6 are satisfied.*

A.12 No Jumping for Utility Function 19

We prove the following slightly more general result which implies no jumping in the two dimensional case when utility is given by Equation 19:

Proposition 7. *Suppose that utility is given by:*

$$u(T, \mathbf{z}; \mathbf{n}) = \sum_{i=1}^{K} z_i(\mathbf{n}) + T + n_1 \frac{z_1^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2^{1+\theta_2}}{1+\theta_2} + ... + n_K \frac{z_K^{1+\theta_K}}{1+\theta_K} \tag{33}$$

with $\theta_i > 0 \forall i$*. Further suppose that* $\mathbf{N}$ *is a convex subset of* $\mathbb{R}^K$*. Then the optimal tax schedule yields a continuous allocation* $\mathbf{z}^*(\mathbf{n})$*.*

Proof. We begin by stating the following Lemma (adapted from Rochet (1987)):

Lemma 4. *If* $\mathbf{N}$ *is a convex subset of* $\mathbb{R}^k$ *and utility takes the form:*

$$u(T, \mathbf{z}; \mathbf{n}) = y(\mathbf{z}) + T + \mathbf{n} \cdot v(\mathbf{z})$$

then $\mathbf{z}^*(\mathbf{n})$ *is incentive compatible if and only if* $U^*(\mathbf{n})$ *is a convex (and hence continuous) function with:*

$$\nabla_{\mathbf{n}} U^*(\mathbf{n}) = v(\mathbf{z}^*(\mathbf{n})) \tag{34}$$

holding a.e. $\mathbf{n}$*.*

Now, suppose that $\mathbf{z}^*(\mathbf{n})$ has a discontinuity. By injectivity of the function:

$$\nabla_{\mathbf{n}} U(\mathbf{n}) = \left[\frac{z_1^{1+\theta_1}}{1+\theta_1}, \frac{z_2^{1+\theta_2}}{1+\theta_2}, ..., \frac{z_K^{1+\theta_K}}{1+\theta_K} \right]$$

we know that this implies that $\nabla_{\mathbf{n}} U^*(\mathbf{n})$ also has a discontinuity. We will show that it can never be optimal to have a discontinuous $\nabla_{\mathbf{n}} U^*(\mathbf{n})$. Towards a contradiction, suppose that optimal $U^*(\mathbf{n})$ has a discontinuity over a surface Σ with normal vector $\mathbf{p}$ pointing from the arbitrarily chosen $''-''$ side into the other $''+''$ side. Let $\nabla^+_{\mathbf{n}} U^*(\mathbf{n})$ denote the gradient on the "+" side and $\nabla^-_{\mathbf{n}} U^*(\mathbf{n})$ denote the gradient on the "-" side.

Next, we can actually transform our problem into a classical calculus of variations problem by writing $z_i(\mathbf{n}) = [(1+\theta_i) \frac{\partial U}{\partial n_i}(\mathbf{n})]^{\frac{1}{1+\theta_i}}$ and $\sum_{i=1}^{K} z_i(\mathbf{n}) + T(\mathbf{z}(\mathbf{n})) = U(\mathbf{n}) - \mathbf{n} \cdot \nabla_{\mathbf{n}} U(\mathbf{n})$. Turning maximization problem 43 into a Lagrangian by adjoining the budget constraint with Lagrange multiplier λ, expressing $\mathbf{z}(\mathbf{n})$ and $T(\mathbf{z}(\mathbf{n}))$ as functions of U and $\nabla_{\mathbf{n}} U$ (omitting arguments of

$U(\mathbf{n})$ and $\nabla_{\mathbf{n}} U(\mathbf{n})$), and using Lemma 4 to convert the global incentive compatibility constraints into a convexity constraint, we can write the government's maximization problem as:

$$\max_U \int_{\mathbf{N}} L(\mathbf{n}, U, \nabla_{\mathbf{n}} U) = \int_{\mathbf{N}} \left\{ W(U, \mathbf{n}) + \lambda \left[\sum_{i=1}^{K} \left[(1+\theta_i) \frac{\partial U}{\partial n_i} \right]^{\frac{1}{1+\theta_i}} - (U - \mathbf{n} \cdot \nabla_{\mathbf{n}} U) - E \right] \right\} f(\mathbf{n})$$

s.t. U is convex

To simplify ideas, first consider the case when $U^*(\mathbf{n})$ happens to be *strictly* convex. In this case, $U^*(\mathbf{n})$ solves the relaxed problem ignoring the convexity constraint, which implies that this function is a stationary point of this functional within the class of all piece-wise smooth functions. Hence, if this function has a discontinuity in its gradient (i.e., a kink) then it must satisfy the classical Weierstrass-Erdman corner condition along the discontinuity surface (omitting the first two arguments of L):[49]

$$\left[L_3(\mathbf{n}, U^*, \nabla^-_{\mathbf{n}} U^*) - L_3(\mathbf{n}, U^*, \nabla^+_{\mathbf{n}} U^*) \right] \cdot \left[\nabla^+_{\mathbf{n}} U^* - \nabla^-_{\mathbf{n}} U^* \right] = 0 \tag{35}$$

However, note that:

$$L_3(\mathbf{n}, U, \nabla_{\mathbf{n}} U) = \lambda f(\mathbf{n}) \begin{bmatrix} \left((1+\theta_1) \frac{\partial U}{\partial n_1} \right)^{\frac{1}{1+\theta_1} - 1} + n_1 \\ \left((1+\theta_2) \frac{\partial U}{\partial n_2} \right)^{\frac{1}{1+\theta_2} - 1} + n_2 \\ \vdots \\ \left((1+\theta_K) \frac{\partial U}{\partial n_K} \right)^{\frac{1}{1+\theta_K} - 1} + n_K \end{bmatrix}$$

But this implies $L_3(\mathbf{n}, U, \nabla_{\mathbf{n}} U)$ is strictly concave in $\nabla_{\mathbf{n}} U = \left[\frac{\partial U}{\partial n_1}, \frac{\partial U}{\partial n_2}, ..., \frac{\partial U}{\partial n_K} \right]$.

Concavity of L_3 in $\nabla_{\mathbf{n}} U$ implies that:

$$\left[L_3(\mathbf{n}, U^*, \nabla^-_{\mathbf{n}} U^*) - L_3(\mathbf{n}, U^*, \nabla^+_{\mathbf{n}} U^*) \right] \cdot \left[\nabla^+_{\mathbf{n}} U^* - \nabla^-_{\mathbf{n}} U^* \right] > 0 \tag{36}$$

Hence, Equation 36 implies that the Weierstrass-Erdman corner condition 35 cannot be satisfied over any discontinuity surface, which in turn implies that the problem has continuous gradient.

The proof for the case where $U^*(\mathbf{n})$ may be weakly convex (i.e., linear over some portion of the space) is more difficult because not every perturbation maintains convexity (i.e., there are some directions in which we cannot perturb $U^*(\mathbf{n})$ because they would violate the convexity condition). Thus, we need to devise our own perturbation to the schedule which maintains convexity yet still yields a contradiction. Towards this purpose, we will consider a particular perturbation to $U^*(\mathbf{n})$.

Consider a surface that equals $U^*(\mathbf{n})$ at $\sigma - \epsilon \mathbf{p}(\sigma)$ for each $\sigma \in \Sigma$ and a small ϵ. Now suppose this surface has gradient $\frac{1}{2}[\nabla^+_{\mathbf{n}} U^*(\sigma) + \nabla^-_{\mathbf{n}} U^*(\sigma)]$ along the segment $[\sigma - \epsilon \mathbf{p}(\sigma), \sigma + \epsilon'(\sigma)\mathbf{p}(\sigma)]$, intersecting $U^*(\mathbf{n})$ again at $\sigma + \epsilon'(\sigma)\mathbf{p}(\sigma)$ for small function $\epsilon'(\sigma)$. Next, note that this new surface is convex. $\nabla^+_{\mathbf{n}} U^*(\sigma) \cdot \mathbf{v}$ is increasing in all directions $\mathbf{v}$ as is $\nabla^-_{\mathbf{n}} U^*(\sigma) \cdot \mathbf{v}$ (this follows

[49] See, for example Grabovsky and Truskinovsky (2010).

because a function is convex if and only if it remains convex when restricted to a line segment).[50] Hence, $\frac{1}{2}[\nabla^+_{\mathbf{n}} U^*(\sigma) + \nabla^-_{\mathbf{n}} U^*(\sigma)] \cdot \mathbf{v}$ is also increasing in all directions $\mathbf{v}$, which implies that this function is convex. Next, consider the pointwise maximum of this surface with $U^*(\mathbf{n})$. This new function is also convex as the pointwise maximum of two convex functions is convex. Graphically, we are perturbing the utility schedule to look like Figure 6, where we have labeled the perturbation surface $\Delta U(n_1, n_2)$ (and the perturbed schedule is the pointwise maximum of $U^*(n_1, n_2)$ and $\Delta U(n_1, n_2)$).

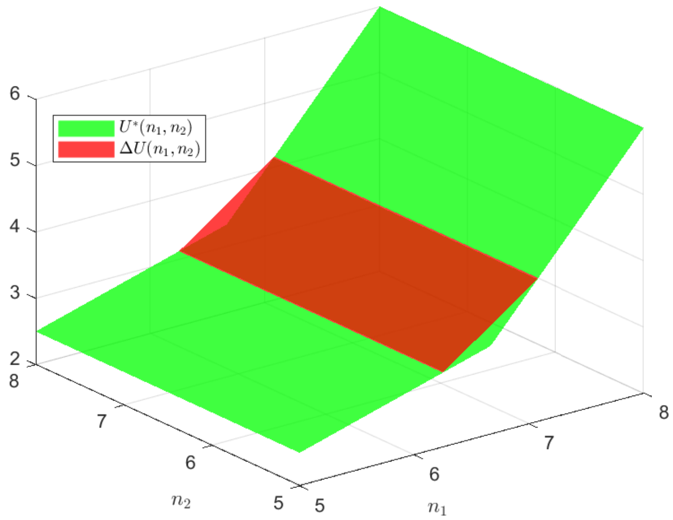


Figure 6: Illustration of Perturbation to $U^*(\mathbf{n})$

Note: This figure shows an example of the perturbation we consider to the optimal utility schedule $U^*(\mathbf{n}) = U^*(n_1, n_2)$. The perturbation surface is $\Delta U(n_1, n_2)$ and the perturbed schedule is given by the pointwise maximum of $U^*(n_1, n_2)$ and $\Delta U(n_1, n_2)$.

Next, we consider the impact of this perturbation on the government's Lagrangian:

$$\Delta \int_{\mathbf{N}} L(\mathbf{n}, U, \nabla_{\mathbf{n}} U) =$$
$$\int_{\Sigma} \int_{\sigma - \epsilon \mathbf{p}(\sigma)}^{\sigma + \epsilon'(\sigma)\mathbf{p}(\sigma)} \left\{ L \left(\mathbf{n}, \oint_{\sigma - \epsilon \mathbf{p}(\sigma)}^{\mathbf{n}} \left[\frac{1}{2} \nabla^+_{\mathbf{n}} U^*(\sigma) + \frac{1}{2} \nabla^-_{\mathbf{n}} U^*(\sigma) - \nabla_{\mathbf{n}} U^*(\mathbf{s}) \right] \cdot \mathbf{p}(\sigma) d\mathbf{s}, \frac{1}{2} \nabla^+_{\mathbf{n}} U^*(\sigma) + \frac{1}{2} \nabla^-_{\mathbf{n}} U^*(\sigma) \right) \right.$$
$$\left. - L(\mathbf{n}, U^*(\mathbf{n}), \nabla_{\mathbf{n}} U^*(\mathbf{n})) \right\} d\mathbf{n} d\sigma \tag{37}$$

where the impact on U is calculated via the fundamental theorem of calculus as the line integral of the gradient of utility between $\sigma - \epsilon \mathbf{p}(\sigma)$ and $\mathbf{n}$, noting that definitionally $\mathbf{p}(\mathbf{s}) = \mathbf{p}(\sigma)$ along the orthogonal line between $\sigma - \epsilon \mathbf{p}(\sigma)$ and $\mathbf{n}$.

Splitting up the inner integral into the line segments $[\sigma - \epsilon \mathbf{p}(\sigma), \sigma]$ and $[\sigma, \sigma + \epsilon'(\sigma)\mathbf{p}(\sigma)]$, dividing by ϵ, taking limits as $\epsilon \to 0$, and using the rectangle approximation yields the following

[50] Note that $\nabla^+_{\mathbf{n}} U^*(\sigma) \cdot \mathbf{v}$ and $\nabla^-_{\mathbf{n}} U^*(\sigma) \cdot \mathbf{v}$ are constant (hence weakly increasing) if we move orthogonal to Σ and increasing if we move along Σ.

derivative:

$$
\begin{aligned}
&\frac{1}{\epsilon}\Delta \int_{\mathbf{N}} L(\mathbf{n}, U, \nabla_{\mathbf{n}} U) \\
&\to \int_{\Sigma} \Bigg\{ \frac{\epsilon}{\epsilon}\Bigg(L_3(\sigma, U^*(\sigma), \nabla^-_{\mathbf{n}} U^*(\sigma)) \cdot \left[\frac{1}{2}\nabla^+_{\mathbf{n}} U^*(\sigma) + \frac{1}{2}\nabla^-_{\mathbf{n}} U^*(\sigma) - \nabla^-_{\mathbf{n}} U^*(\sigma)\right] \\
&+ L_2(\sigma, U^*(\sigma), \nabla^-_{\mathbf{n}} U^*(\sigma))\epsilon \left[\frac{1}{2}\nabla^+_{\mathbf{n}} U^*(\sigma) + \frac{1}{2}\nabla^-_{\mathbf{n}} U^*(\sigma) - \nabla^-_{\mathbf{n}} U^*(\sigma)\right] \cdot \mathbf{p}(\sigma)\Bigg) \\
&+ \frac{\epsilon'(\sigma)}{\epsilon}\Bigg(L_3(\sigma, U^*(\sigma)\nabla^+_{\mathbf{n}} U^*(\sigma)) \cdot \left[\frac{1}{2}\nabla^+_{\mathbf{n}} U^*(\sigma) + \frac{1}{2}\nabla^-_{\mathbf{n}} U^*(\sigma) - \nabla^+_{\mathbf{n}} U^*(\sigma)\right] \\
&+ L_2(\sigma, U^*(\sigma), \nabla^+_{\mathbf{n}} U^*(\sigma))\epsilon \left[\frac{1}{2}\nabla^+_{\mathbf{n}} U^*(\sigma) + \frac{1}{2}\nabla^-_{\mathbf{n}} U^*(\sigma) - \nabla^+_{\mathbf{n}} U^*(\sigma)\right] \cdot \mathbf{p}(\sigma)\Bigg) \Bigg\} d\sigma
\end{aligned}
$$

$\epsilon'(\sigma)/\epsilon \to 1$ as $\epsilon \to 0$ (and the L_2 terms go to zero as they are of order ϵ^2) so we are left with:[51]

$$
\begin{aligned}
&\frac{1}{\epsilon}\Delta \int_{\mathbf{N}} L(\mathbf{n}, U, \nabla_{\mathbf{n}} U) \\
&\to \int_{\Sigma} \Bigg\{ L_3(\sigma, U^*(\sigma), \nabla^-_{\mathbf{n}} U^*(\sigma)) \cdot \left[\frac{1}{2}\nabla^+_{\mathbf{n}} U^*(\sigma) + \frac{1}{2}\nabla^-_{\mathbf{n}} U^*(\sigma) - \nabla^-_{\mathbf{n}} U^*(\sigma)\right] \\
&+ L_3(\sigma, U^*(\sigma), \nabla^+_{\mathbf{n}} U^*(\sigma)) \cdot \left[\frac{1}{2}\nabla^+_{\mathbf{n}} U^*(\sigma) + \frac{1}{2}\nabla^-_{\mathbf{n}} U^*(\sigma) - \nabla^+_{\mathbf{n}} U^*(\sigma)\right] \Bigg\} d\sigma \\
&= \int_{\Sigma} \frac{1}{2}\left[L_3(\sigma, U^*(\sigma), \nabla^-_{\mathbf{n}} U^*(\sigma)) - L_3(\sigma, U^*(\sigma), \nabla^+_{\mathbf{n}} U^*(\sigma))\right] \cdot \left[\nabla^+_{\mathbf{n}} U^*(\sigma) - \nabla^-_{\mathbf{n}} U^*(\sigma)\right] d\sigma > 0
\end{aligned}
$$

The final inequality follows by Equation 36. But this means that from the supposed optimal schedule $U^*(\mathbf{n})$, we have found a welfare improving perturbation, which is a contradiction. Hence, it can never be optimal to have a discontinuous $\nabla_{\mathbf{n}} U^*(\mathbf{n})$, which implies that it cannot be optimal to have discontinuous $z^*(\mathbf{n})$. □

B First Order Approaches to Solving Multidimensional Screening Problems

B.1 First Order Approach I: Euler-Lagrange Equation

Perhaps the most fundamental first order approach to multidimensional screening involves the Euler-Lagrange equation. The idea is to use the envelope condition 4 to express $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ as a function of $(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))$ in order to rewrite the optimization problem solely in terms of $(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))$.[52] Then we can derive the Euler-Lagrange equation associated to this calculus of variations problem, which will in general be a complicated second order PDE.

Remark 5. *As an example of this approach, suppose that* $\mathbf{N} \subseteq (-\infty, 0)^K$ *and*

$$
u(T, \mathbf{z}; \mathbf{n}) = \sum_{i=1}^{K} z_i + T + \sum_{i=1}^{K} n_i \frac{z_i^{1+\theta_i}}{1+\theta_i} \tag{38}
$$

[51] The limiting perturbation is symmetric around Σ, which is why $\epsilon'(\sigma)/\epsilon \to 1$.

[52] This of course requires that the mapping between $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ and $(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))$ is bijective, which is naturally not always the case.

with $z_1, z_2, ..., z_K \geq 0$. *Then the envelope condition tells us that:*

$$U_{n_i}(\mathbf{n}) = u_{n_i}(T, \mathbf{z}; \mathbf{n}) = \frac{z_i^{1+\theta_i}}{1+\theta_i}$$

Hence, this yields that:

$$z_i(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})) = [(1+\theta_i) U_{n_i}(\mathbf{n})]^{\frac{1}{1+\theta_i}} \tag{39}$$

Moreover, we can use Equations 38 and 39 to write:

$$T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})) = -\sum_{i=1}^{K} ((1+\theta_i) U_{n_i}(\mathbf{n}))^{\frac{1}{1+\theta_i}} + U(\mathbf{n}) - \sum_{i=1}^{K} n_i U_{n_i}(\mathbf{n})$$

Hence, we can rewrite Equation 18 as (appending the budget constraint on to the objective function with Lagrange multiplier λ):

$$\begin{aligned} &\max_{U(\mathbf{n}), \lambda} \int_{\mathbf{N}} \left\{ W\left(U(\mathbf{n}), \mathbf{n}\right) + \lambda[-T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})) - E] \right\} f(\mathbf{n}) d\mathbf{n} \\ &\mathbf{n} \in \underset{\mathbf{n}'}{\operatorname{argmax}}\ u(T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})), \mathbf{z}(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})); \mathbf{n})\ \forall \mathbf{n} \end{aligned} \tag{40}$$

Now, if the solution $U^*(\mathbf{n})$ is interior in the sense that $U^*(\mathbf{n}) + \epsilon \tilde{U}(\mathbf{n})$ is incentive compatible for any perturbation function $\tilde{U}(\mathbf{n})$ and sufficiently small ϵ, then the solution to the above variational calculus problem is given by the Euler-Lagrange equation:

$$\frac{\partial L(U, \nabla_{\mathbf{n}} U, \mathbf{n})}{\partial U} - \sum_{i=1}^{K} \frac{\partial}{\partial n_i} \left(\frac{\partial L(U, \nabla_{\mathbf{n}} U, \mathbf{n})}{\partial U_{n_i}} \right) = 0 \tag{41}$$

where

$$L(U, \nabla_{\mathbf{n}} U, \mathbf{n}) = W\left(U, \mathbf{n}\right) f(\mathbf{n}) + \lambda[-T(U, \nabla_{\mathbf{n}} U) - E] f(\mathbf{n})$$

along with associated boundary condition, where $\mathbf{p}$ is the outward pointing normal to the boundary $\partial \mathbf{N}$:

$$\left(\frac{\partial L(U, \nabla_{\mathbf{n}} U, \mathbf{n})}{\partial U_{n_i}} \right) \cdot \mathbf{p} = 0 \tag{42}$$

Unfortunately, Equations 41 and 42 typically have no known analytical solution and are moreover are a difficult system of partial differential equations to solve. And if the optimal utility function $U(\mathbf{n})$ is not interior (e.g., it features bunching), then we cannot use this approach.

B.2 First Order Approach II: Optimal Control

A second potential approach might be to consider using optimal control theory.[53] Mirrlees (1976) suggested this approach for multidimensional optimal taxation as did Basov (2001) in the context of more general multidimensional screening problems. Unfortunately, optimal control methods often cannot be applied to multidimensional screening due to the inability to apply the fundamental lemma of calculus of variations. To see why, it is helpful to do a change of variables (as is standard in this literature, e.g., Mirrlees (1971)) and consider the government as

[53] A number of the points raised in this section developed out of conversations with Ilia Krasikov and Mike Golosov as well as with Etienne Lehmann.

choosing the functions $\mathbf{z}(\mathbf{n})$ and $U(\mathbf{n})$, with the transfer function (now expressed as a function of U, $\mathbf{z}$, and $\mathbf{n}$) determined implicitly via $U = u(T(U, \mathbf{z}, \mathbf{n}), \mathbf{z}; \mathbf{n})$.[54]

$$
\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}),U(\mathbf{n})} \int_{\mathbf{N}} W\left(U(\mathbf{n}), \mathbf{n}\right) dF(\mathbf{n}) \\
&\text{s.t.} \quad \int_{\mathbf{N}} [T(\mathbf{z}(\mathbf{n})) + E] dF(\mathbf{n}) \leq 0 \\
&\mathbf{n} \in \operatorname*{argmax}_{\mathbf{n}'} u(T(U(\mathbf{n}'), \mathbf{z}(\mathbf{n}'), \mathbf{n}'), \mathbf{z}(\mathbf{n}'); \mathbf{n}) \ \forall \mathbf{n} \\
&U(\mathbf{n}) = u(T(U(\mathbf{n}), \mathbf{z}(\mathbf{n}), \mathbf{n}), \mathbf{z}(\mathbf{n}); \mathbf{n})
\end{aligned} \tag{43}
$$

To solve System 43, we will consider $\mathbf{z}(\mathbf{n})$ as control variables, and $U(\mathbf{n})$ as the state variable governed by the envelope condition 4 (which plays the role of the equation of motion). We define the *Hamiltonian* of this problem as:[55]

$$
\mathcal{H}(\mathbf{z}, U, \phi, \mathbf{n}, \lambda) = \left\{W\left(U, \mathbf{n}\right) + \lambda \left[-T(U, \mathbf{z}, \mathbf{n}) - E\right]\right\} f(\mathbf{n}) + \phi(\mathbf{n}) \cdot \nabla_{\mathbf{n}} u(T(U, \mathbf{z}, \mathbf{n}), \mathbf{z}; \mathbf{n}) \tag{44}
$$

where $\phi(\mathbf{n})$ is a vector of *costate variables*. Now, if the solution $\mathbf{z}^*(\mathbf{n})$ is interior in the sense that $\mathbf{z}^*(\mathbf{n}) + \epsilon\tilde{\mathbf{z}}(\mathbf{n})$ is incentive compatible when coupled with a suitable perturbation to the utility function $U(\mathbf{n}) + \epsilon\tilde{U}(\mathbf{n})$, then we can apply a multidimensional analogue to Pontryagin's Maximum Principle to characterize the optimal solution (e.g., Udriste (2009)). Unfortunately, in general, the solution $\mathbf{z}^*(\mathbf{n})$ will not be interior due to the requirement that $\nabla_{\mathbf{n}} u(T, \mathbf{z}; \mathbf{n})|_{\mathbf{z}=\mathbf{z}(\mathbf{n}), T=T(\mathbf{z}(\mathbf{n}))}$ is a conservative vector field; hence, it is likely often the case that perturbations to $\mathbf{z}^*(\mathbf{n})$ will generate a vector field $\nabla_{\mathbf{n}} u(T, \mathbf{z}; \mathbf{n})|_{\mathbf{z}=\mathbf{z}^*(\mathbf{n}), T=T(\mathbf{z}^*(\mathbf{n}))}$ that is *not* conservative. In other words, System 43 does not impose the requirement that $\nabla_{\mathbf{n}} u(T, \mathbf{z}; \mathbf{n})|_{\mathbf{z}=\mathbf{z}(\mathbf{n}), T=T(\mathbf{z}(\mathbf{n}))}$ forms a conservative vector field. Thus, optimal control approaches are not valid in many multidimensional screening settings.

C Simulations Appendix

C.1 Calibration for Section 5.4

We use income data from the 2019 Current Population Survey for married heterosexual couples both of whom are under the age of 65. The calibration exercise searches over the space of four parameters $\theta_1, \theta_2, \alpha_1$, and α_2. For each choice of these four parameters, we calibrate the distribution of types $f(n_1, n_2)$ to match the empirical joint income distribution of couples. We assume that $f(n_1, n_2)$ is log-normal and choose the parameters of the log-normal distribution to best match the observed income distribution.[56] Then given these four parameters and the corresponding calibrated log-normal distribution $f(n_1, n_2)$, we calculate the sum of squares between the true and calibrated values of four statistics: the median compensated taxable income elasticity for men (0.2, taken from Blomquist and Selin (2010)), the median compensated taxable income elasticity for women (1, also taken from Blomquist and Selin (2010)), the percentage of men who do not work (13.5%, from CPS data), and the fraction of women who do not work (20%, from CPS data). Finally, we search over the space of these four parameters $\theta_1, \theta_2, \alpha_1$, and α_2 to minimize this sum of squares.

[54]This is WLOG as long as the mapping $T \mapsto U$ is bijective. Now, $T \mapsto U$ is injective conditional on a $\mathbf{z}$ because $u_T(T, \mathbf{z}; \mathbf{n}) > 0$ and we also typically assume $T \mapsto U$ is surjective onto $\mathbb{R}$ (i.e., any utility can be reached with a sufficiently small or large transfer).

[55]Notationally, it's important to remember that U represents the utility schedule (as a function of $\mathbf{n}$) that the government chooses and u represents the utility function of the individual problem: $u(T, \mathbf{z}; \mathbf{n})$.

[56]Technically, we assume that $f(-n_1, -n_2)$ is log-normal because $n_1, n_2 < 0$.

C.2 Simulations for Utility Function 26

Simulations when utility is given by Equation 26 are nearly identical as when utility is given by Equation 21 after a change of variables: $d = \log(z_1 + z_2 + T)$. Then, we consider agents as maximizing:

$$u(d, \mathbf{z}; \mathbf{n}) = d + n_1 \frac{z_1^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2^{1+\theta_2}}{1+\theta_2} - \frac{n_1^2}{2\alpha} z_1 - \frac{n_2^2}{2\alpha} z_2$$

which is exactly the same as Equation 21. The only change to our optimization problem is that we need to account for the change of variables from $z_1 + z_2 + T$ to d when computing the government's budget constraint. Hence, we solve:

$$\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}), U(\underline{\mathbf{n}})} \int_{\mathbf{N}} W(U(\mathbf{n}), \mathbf{n}) dF(\mathbf{n}) \\
&\text{s.t. } \int_{\mathbf{N}} [\exp(d(\mathbf{n})) - z_1 - z_2 + E] dF(\mathbf{n}) \leq 0 \\
&U(\mathbf{n}) = U(\underline{\mathbf{n}}) + \oint_{\underline{\mathbf{n}}}^{\mathbf{n}} \begin{bmatrix} \frac{z_1(\mathbf{s})^{1+\theta_1}}{1+\theta_1} - \frac{s_1}{\alpha} z_1(\mathbf{s}) \\ \frac{z_2(\mathbf{s})^{1+\theta_2}}{1+\theta_2} - \frac{s_2}{\alpha} z_2(\mathbf{s}) \end{bmatrix} \cdot d\mathbf{s} \\
&\frac{\partial z_1}{\partial n_1}(\mathbf{n}) > 0, \frac{\partial z_2}{\partial n_2}(\mathbf{n}) > 0, \frac{\partial z_1}{\partial n_1}(\mathbf{n}) \frac{\partial z_2}{\partial n_2}(\mathbf{n}) - \frac{\partial z_1}{\partial n_2}(\mathbf{n}) \frac{\partial z_2}{\partial n_1}(\mathbf{n}) > 0 \\
&\left(z_1^{\theta_1}(\mathbf{n}) - \frac{n_1}{\alpha} \right) \frac{\partial z_1}{\partial n_2}(\mathbf{n}) = \left(z_2(\mathbf{n})^{\theta_2} - \frac{n_2}{\alpha} \right) \frac{\partial z_2}{\partial n_1}(\mathbf{n}) \\
&d(\mathbf{n}) = U(\mathbf{n}) - \left[n_1 \frac{z_1(\mathbf{n})^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2(\mathbf{n})^{1+\theta_2}}{1+\theta_2} - \frac{n_1^2}{2\alpha} z_1(\mathbf{n}) - \frac{n_2^2}{2\alpha} z_2(\mathbf{n}) \right] - z_1 - z_2
\end{aligned} \tag{45}$$

By Corollary 2.1 any solution to System 45 will have diffeomorphic $\mathbf{n} \mapsto \mathbf{z}$. Moreover, we know that any solution to System 45 satisfies the envelope condition 4 everywhere; hence, if we confirm that $\forall \mathbf{n}$ we have $u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ for $\mathbf{n}' \in \partial \mathbf{N}$, Theorem 2 ensures that the allocation is incentive compatible.

C.3 Additional Simulation Figures

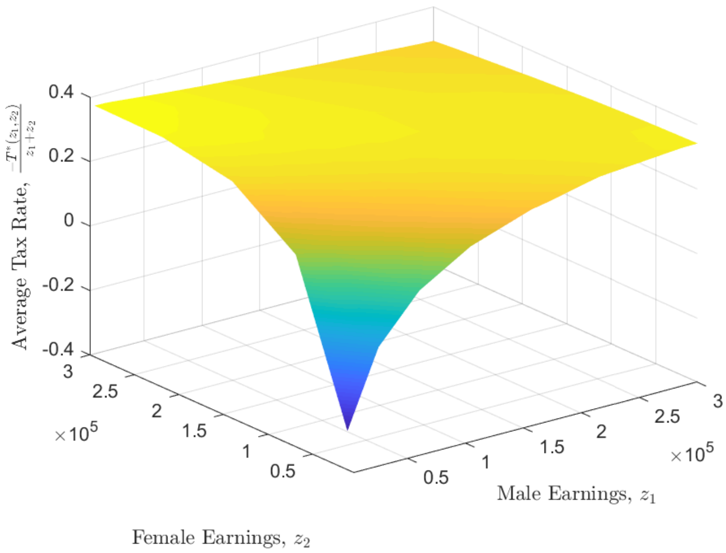


Figure 7: Optimal Average Tax Schedule for Couples

Note: This figure shows the optimal average tax rates for couples, assuming utility is given by Equation 26. $f(\mathbf{n})$ is calibrated to match the joint income distribution from the CPS and $\theta_1, \theta_2, \alpha_1, \alpha_2$ are chosen to match four moments: the compensated taxable income elasticity for men (0.2, taken from Blomquist and Selin (2010)), the compensated taxable income elasticity for women (1, also taken from Blomquist and Selin (2010)), the percentage of men who do not work (13.5%, from CPS data), and the fraction of women who do not work (20%, from CPS data). The social welfare function is given by $W(U(\mathbf{n}), \mathbf{n}) = \psi(\mathbf{n})U(\mathbf{n})$ with welfare weights $\psi(\mathbf{n})$ chosen so that $\psi(\mathbf{n})$ is $\approx 10,000$ times higher for the lowest income household than for the highest income household.

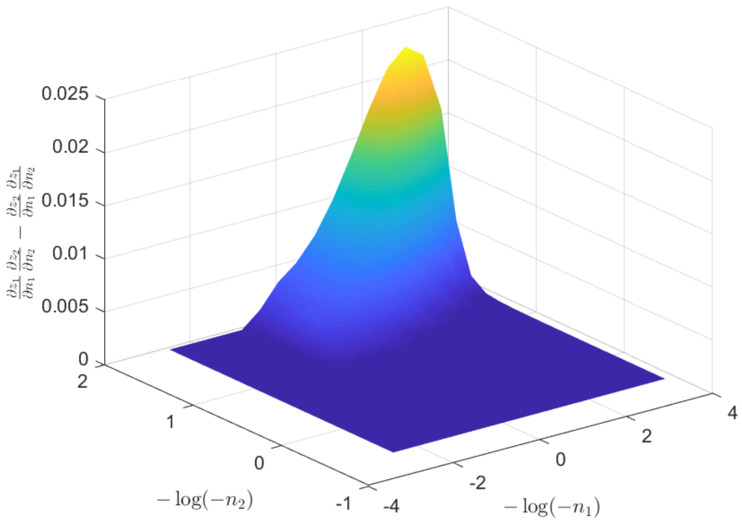
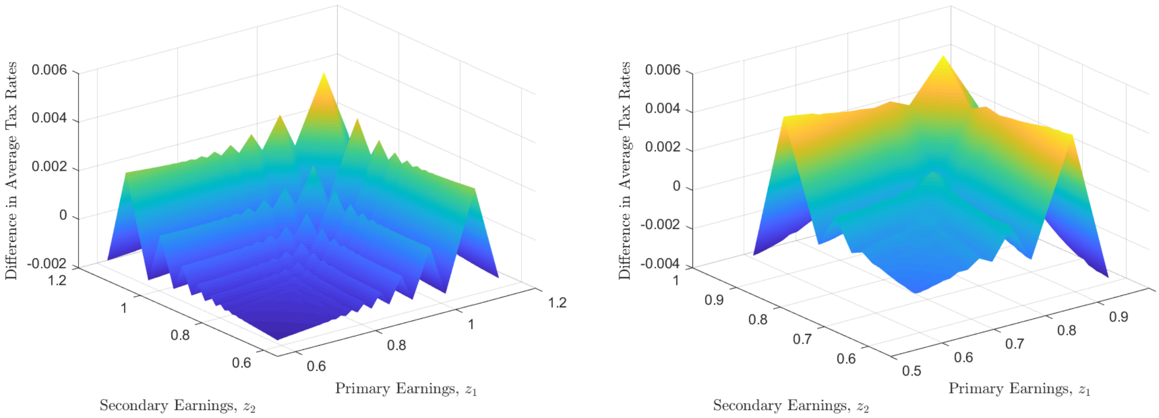


Figure 8: Jacobian Determinant, Couples Taxation Using CPS Data and Log Utility Over Consumption

Note: This figure shows the Jacobian determinant $\frac{\partial z_1}{\partial n_1}\frac{\partial z_2}{\partial n_2} - \frac{\partial z_2}{\partial n_1}\frac{\partial z_1}{\partial n_2}$ assuming utility is given by Equation 26. We plot the Jacobian determinant against $(-\log(-n_1), -\log(-n_2))$ to compress the type distribution for readability. $f(\mathbf{n})$ is calibrated to match the empirical joint income distribution of couples from the CPS and $\theta_1, \theta_2, \alpha_1, \alpha_2$ are chosen to match four moments: the compensated taxable income elasticity for men (0.2, taken from Blomquist and Selin (2010)), the compensated taxable income elasticity for women (1, also taken from Blomquist and Selin (2010)), the percentage of men who do not work (13.5%, from CPS data), and the fraction of women who do not work (20%, from CPS data). The social welfare function is given by $W(U(\mathbf{n}), \mathbf{n}) = \psi(\mathbf{n})U(\mathbf{n})$ with welfare weights $\psi(\mathbf{n})$ chosen so that $\psi(\mathbf{n})$ is $\approx 10,000$ times higher for the lowest income household than for the highest income household.



(a) First Set of Input Data

(b) Second Set of Input Data

Figure 9: Differences in Average Taxes, Our Method vs. Method of Aguilera and Morin (2008)

Note: This figure shows the differences in the average tax rates computed via our method described in Section 5.2 and the method of Aguilera and Morin (2008) for utility functions 19. Panel 9a shows this difference for the first set of input data from Section 5.3 and panel 9b shows this difference for the second set of input data from Section 5.3. Averaged over all individuals, the mean absolute difference between the average tax rates computed via the two different methods is 0.14 percentage points (panel 9a) and 0.31 percentage points (panel 9b).