Solving Multidimensional Screening Problems Using a Generalized Single Crossing Property*

William Dodds†

July 14, 2023

Abstract

This paper derives necessary and sufficient conditions for allocations to be incentive compatible in multidimensional screening problems that satisfy a generalized single crossing property. We then devise a numerical method based on these results to solve multidimensional screening problems. Importantly, our numerical method can be applied to multidimensional screening problems for which existing approaches cannot be applied. 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 implement our numerical method, our simulations highlight the importance of bunching in optimal multidimensional taxation. Finally, we prove that bunching must occur in multidimensional optimal taxation problems when the social planner has sufficiently redistributive preferences.

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. Thank you to two anonymous referees and the editor, Nicholas Yannelis, for providing many helpful comments that improved the paper. Thanks to various seminar participants at Tulane University and Virginia Tech. Thanks to Stefano Barbieri 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. Solving unidimensional screening problems is greatly simplified due to two key results: (1) if preferences satisfy the single crossing property, then local incentive compatibility is necessary and sufficient for global incentive compatibility (e.g., Mirrlees (1971)) and (2) 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. Unfortunately, neither of these two results carries over directly to general multidimensional setting in which a principal screens agents who differ on many dimensions and have many choice variables. Thus, solving multidimensional screening problems is substantially more difficult.

Broadly, there are two approaches that are commonly taken to solve multidimensional screening problems. First, one can attempt to use "first order approaches" which *assume* that first order conditions are sufficient to guarantee global incentive compatibility. Under this assumption, solutions to multidimensional screening problems can be constructed by solving a system of partial differential equations. However, when the optimal allocation features bunching (or other non-smooth properties such as agents with multiple optima), then first order approaches cannot be applied. In this case, the sole path forward (to the best of my knowledge) requires that utility take the following linear and separable form where $\mathbf{n}$ denotes an agent's type, $\mathbf{z}$ denotes the actions the agent can take, $y(\cdot)$ is an arbitrary function, and T denotes the transfer from the screening entity:

$$u(T, \mathbf{z}; \mathbf{n}) = y(\mathbf{z}) + T + \mathbf{n} \cdot v(\mathbf{z}) \tag{1}$$

When utility is given by Equation 1, we can use the classic result from Rochet (1987) which characterizes the set of incentive compatible allocations in terms of an envelope condition and a convexity constraint on indirect utility. This characterization allows us to solve multidimensional screening problems via numerical algorithms designed to solve variational calculus problems subject to a convexity constraint, e.g., Aguilera and Morin (2008), Oberman (2013), Mérigot and Oudet (2014), or Boerma, Tsyvinski and Zimin (2022).[1] However, to the best of my knowledge, there is no general method to solve screening problems when first order approaches fail and when utility does not take the above linear and separable form.

This paper contributes to the screening literature by developing a set of results which allow us to numerically solve a new class of multidimensional screening problems when first order approaches fail and when utility is not necessarily given by Equation 1. We first derive a number of results characterizing incentive compatibility in general multidimensional screening problems assuming preferences satisfy a "generalized single crossing property". Theorem 1 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 agents prefer their assigned bundle to bundles chosen by boundary individuals. Moreover, we discuss how Theorem 1 extends to non-smooth allocations via a limiting argument; this extension is important insofar as it will help us to solve multidimensional screening problems when the optimal allocation is not smooth. Theorem 1 is primarily helpful for *verifying* whether a given allocation is incentive compatible because it significantly reduces the set of incentive constraints one must check to determine incentive compatibility. However, the set of relevant incentive constraints from Theorem 1 is still too large to efficiently construct optimal allocations. Towards constructing optimal allocations, Theorem 2 derives a set of necessary conditions for incentive compatibility: first order conditions must hold almost everywhere, second order conditions must hold whenever the allocation is sufficiently smooth, and the mapping from types to actions must be globally injective whenever it is sufficiently smooth. Thus, our core theoretical contribution is to derive results which characterize incentive compatibility (under the generalized single crossing property) in multidimensional settings for which existing theory does not apply.

Our results characterizing incentive compatibility are primarily helpful insofar as they allow us to devise a method to numerically solve multidimensional screening problems that can be applied when existing methods cannot, i.e., first order methods fail and/or utility does not take the linear/separable form of Equation 1. The core idea is to maximize the objective function over the set of smooth allocations satisfying (most of) the necessary conditions in Theorem 2: first order conditions, second order conditions, and a constraint that the allocation is locally invertible. Our approach boils down to a computationally tractable optimization problem with linear and non-linear constraints. Once we have a proposed solution from this optimization problem, we can check whether it satisfies our sufficient conditions for incentive compatibility

[1]Technically, one could also use the sweeping conditions in Rochet and Chone (1998). However, this also requires utility to be given by Equation 1 and is far more difficult to apply constructively than algorithms designed to solve variational calculus problems subject to a convexity constraint.

given by Theorem 1; if so, then we have found the optimal schedule within the class of smooth allocations. Moreover, if the optimal allocation is non-smooth, then our numerical method will identify a smooth schedule that approximates the optimal non-smooth schedule arbitrarily well.[2] This turns out to be quite important as we illustrate solutions that feature non-smoothness in the form of bunching for a number of toy examples as well as a more realistic, calibrated exercise exploring optimal taxation of couples using data from the American Community Survey. Consistent with our numerical simulations, we prove theoretically in Proposition 4 that the optimal tax schedule will *always* feature non-smoothness (typically in the form of bunching) if the social planner has sufficiently redistributive preferences. While bunching was shown to be common in non-linear pricing problems by Armstrong (1996) and Rochet and Chone (1998) due to the presence of participation constraints, our numerical and theoretical results show that bunching also appears to be common in multidimensional optimal taxation problems which do not feature participation constraints.

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. 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. The key contribution that we make relative to Rochet (1987) and McAfee and McMillan (1988) is that our results allow us to solve multidimensional screening problems for 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. Finally, Carlier (2001) generalizes Rochet (1987) by characterizing incentive compatibility when utility is not necessarily linear in type, but is separable and quasi-linear in the transfer, 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. Our results are additive above and beyond Carlier (2001) primarily because

[2] This assumes that the optimal non-smooth allocation can be arbitrarily approximated by smooth allocations, which I conjecture to hold for any realistic problem.

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, (most of) our necessary conditions can be expressed using local properties of the allocation, facilitating use for numerically solving multidimensional screening problems.

The second portion of this paper develops a numerical method to solve multidimensional screening problems when existing techniques cannot be applied: when bunching occurs and utility is not given by Equation 1. We apply this method in the context of multidimensional taxation problems. The vast majority of optimal taxation papers that consider settings in which agents have multidimensional types and multidimensional choice sets appeal to first order approaches which assume *ex ante* that bunching does not occur under the optimal schedule (e.g., Mirrlees (1976), Mirrlees (1986), Kleven, Kreiner and Saez (2009), Golosov, Tsyvinski and Werquin (2014), Spiritus et al. (2022), and Krasikov and Golosov (2022)). Boerma, Tsyvinski and Zimin (2022) allow for bunching in a multidimensional taxation context by assuming the utility function takes the form of Equation 1; this allows them to use the convexity characterization of incentive compatibility from Rochet (1987). We contribute to this literature by devising a method to numerically solve multidimensional screening problems when bunching occurs (so that first order methods fail) and when utility is not given by Equation 1.

The rest of this paper proceeds as follows. Section 2 discusses the multidimensional screening environment along with existing solution techniques for multidimensional screening problems; Section 2 also introduces our generalized single crossing property. Section 3 presents our main results on incentive compatibility. Section 4 develops 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 and Generalized Single Crossing Property

In this section we present the multidimensional screening problem environment, discuss existing methods to solve such problems, and introduce our generalized single crossing property.

2.1 A General Screening Problem

We consider a general multidimensional screening problem (such as non-linear pricing with multiple goods or optimal taxation with multidimensional instruments). The screening entity (e.g., a government or a firm) contracts with agents by choosing a transfer (or tax) function $T(\mathbf{z})$ which gives agents a monetary payoff for making choices $\mathbf{z}$.

The agents that the screening entity contracts with are 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 $\mathbf{N}$ is compact with smooth boundary $\partial\mathbf{N}$. 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 agent's type $\mathbf{n}$ is private information insofar as it cannot be directly observed by the screening entity. Agents have preferences over a monetary transfer from the screening entity, T, 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.[3] For a given transfer schedule $T(\mathbf{z})$ set by the screening entity, agents choose $\mathbf{z}$ to maximize utility denoted $u(T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ where $u(T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ is a smooth utility function satisfying $u_T > 0$ and $u_{TT} \leq 0$.

The screening entity chooses to maximize an objective function (which typically depends on choices made by agents) subject to constraints (e.g., participation constraints, budget constraints, or non-negativity constraints) and the additional constraints that each agent optimizes his/her utility subject to the given transfer schedule $T(\mathbf{z})$:

$$\begin{aligned} &\max_{T(\mathbf{z})} \ \text{Objective} \\ &\text{s.t. } \mathbf{z}^* \in \operatorname*{argmax}_{z} u(T(\mathbf{z}), \mathbf{z}; \mathbf{n}) \ \forall \mathbf{n} \\ &\text{Other Constraints} \end{aligned} \tag{2}$$

Equivalently, due to the revelation principle, the screening 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 so that all types $\mathbf{n}$ satisfy the following incentive compatibility constraint:[4]

$$u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = \max_{\mathbf{n}'} \ u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}) \tag{3}$$

Hence, we can rewrite the screening entity's problem as:

$$\begin{aligned} &\max_{\mathbf{z}(\mathbf{n}), T(\mathbf{z}(\mathbf{n}))} \ \text{Objective} \\ &\text{s.t. } \mathbf{n} \in \operatorname*{argmax}_{\mathbf{n}'} u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}) \ \forall \mathbf{n} \\ &\text{Other Constraints} \end{aligned} \tag{4}$$

2.2 Methods to Solve Screening Problems

To the best of my knowledge, there are two paths that are commonly used to find solutions to screening problems. First, one can attempt to use so-called "first order approaches". First order approaches replace the global incentive compatibility constraints with first order conditions. Ignoring a substantial portion of the constraints simplifies the screening entity's problem considerably. The solution to this simplified problem can be constructed as a solution to the Euler-Lagrange equation associated with this simplified system (a non-linear partial differential

[3]We assume the dimension of the choice set is equal to the dimension of the type space; however, we discuss in Appendix A.4 how our results can be extended to cases where these dimensions differ.

[4]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 $u_T > 0$). This notation is somewhat unconventional but turns out to be helpful to prove a number of our results.

equation) or using optimal control methods (which will generate a system of partial differential equations). We give a brief overview of these methods in Appendix D. First order approaches succeed in finding solutions to screening problems when all individuals' second order conditions hold strictly and all individuals have a unique global optima (see, e.g., Assumption 1 of Spiritus et al. (2022)). However, first order approaches fail to characterize the solution when the solution to the simplified problem (that replaces global incentive constraints with first order conditions) is not globally incentive compatible. This typically occurs when the optimal allocation features *bunching*, wherein many types are assigned to the same bundle.[5] When bunching occurs, second order conditions hold only weakly.[6] And when second order conditions hold only weakly, the screening entity can almost always improve its objective function by ignoring second order conditions; in these cases, first order approaches typically generate proposed solutions for which second order conditions are violated. Thus, first order approaches fail in the presence of bunching because they do not enforce second order constraints. Moreover, we have evidence to suggest that bunching is common in multidimensional screening; for example, Armstrong (1996) and Rochet and Chone (1998) show that bunching is the norm rather than the exception in a multi-product monopolist's problem. In essence, first order approaches are helpful when applicable, but when they fail we need an alternative method to solve screening problems.

When first order approaches fail, there is, as far as I know, only one path forward and it requires utility to be given by the linear and separable form of Equation 1. In this case, we can appeal to the convexity characterization of incentive compatibility from Rochet (1987), which allows us to recast System 4 as a calculus of variations problem subject to a convexity constraint (see Rochet and Chone (1998)). And 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), Mérigot and Oudet (2014), or Boerma, Tsyvinski and Zimin (2022).

This paper is concerned with finding solutions to multidimensional screening problems when (1) first order methods fail and (2) utility is *not* given by Equation 1. We do so by developing a "generalized single crossing property" which allows us to reduce the set of relevant incentive compatibility constraints in System 4. But first, we take a slight detour to define notation.

[5] First order approaches can also fail to ensure global incentive compatibility when individuals have multiple optima, which generates 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 property holds, but extending this condition to multidimensional settings has so far been unsuccessful.

[6] Second order conditions require that $u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ is concave in $\mathbf{n}'$; when bunching occurs $\mathbf{z}(\mathbf{n}')$ is constant for some types $\mathbf{n}'$, hence $u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n})$ is only weakly concave so that second order conditions hold only weakly.

2.3 Notation

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

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

We use $\nabla_{\mathbf{x}}$ to denote the partial derivative (gradient) with respect to the vector $\mathbf{x}$. For example:

$$\nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) \equiv \nabla_{\mathbf{x}} u(T(\mathbf{z}(\mathbf{n})),\mathbf{x};\mathbf{n})|_{\mathbf{x}=\mathbf{z}(\mathbf{n})} = \begin{bmatrix} u_{z_1}(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) \\ \vdots \\ u_{z_K}(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) \end{bmatrix}$$
$$\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) = \begin{bmatrix} u_{n_1}(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) \\ \vdots \\ u_{n_K}(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) \end{bmatrix}$$

In contrast, we use $D_{\mathbf{x}}$ to denote the total derivative with respect to $\mathbf{x}$:

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

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

2.4 Generalized Single Crossing Property

We are going to restrict ourselves to a class of multidimensional screening problems that satisfy a "generalized single crossing property". This assumption places restrictions on the types of problems we can solve using the techniques in this paper. However, recall that when first order methods fail, we can currently only solve multidimensional screening problems when utility is given by Equation 1; given that our generalized single crossing condition is much more general than Equation 1, devising techniques to solve problems that satisfy our generalized single crossing condition is nonetheless a useful step forward. First, we introduce a few pieces of mathematical machinery: we need to introduce the concept of a *diffeomorphism*, we need to define P matrices, and we need to discuss the multidimensional envelope condition. We are going to rely heavily on the concept of a diffeomorphism from differential topology:

Definition 1. *A diffeomorphism is a continuously differentiable bijective function which also has a continuously differentiable inverse.*[7]

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

[7]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.

have a differentiable inverse when $x = 0$. We also define *local diffeomorphisms*:

Definition 2. *A function is a "local diffeomorphism at* $\mathbf{n}$*" if there exists an open set containing* $\mathbf{n}$ *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* $\mathbf{n}$.

Moreover, 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:

Remark 1. *A continuously differentiable function* $f(\mathbf{n})$ *is a local diffeomorphism if and only if the determinant of the Jacobian never changes sign:* $det(\nabla_{\mathbf{n}} f(\mathbf{n})) \neq 0$.

By Remark 1, we can easily check whether a function is a local diffeomorphism using the Jacobian matrix. We can also establish sufficient conditions for a mapping to be diffeomorphic in terms of properties of the Jacobian, such as Theorem 4 from Gale and Nikaido (1965):[8]

Remark 2. *A mapping* $f(\mathbf{n})$ *on a closed rectangular domain with a continuous P matrix Jacobian is a diffeomorphism.*[9]

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

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":[10]

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 is a natural multidimensional generalization of the standard unidimensional single crossing property. The set of unidimensional diffeomorphisms is equal to the set of differentiable strictly monotonic functions; hence, Assumption 1 simplifies to a monotonicity assumption on $\frac{u_z(T,z;n)}{u_T(T,z;n)}$, which is the standard single crossing property (see, e.g., Mirrlees (1971)).[11]

[8]There are other sufficient conditions for a mapping to be diffeomorphic using properties of the Jacobian, such as Theorem 6 from Gale and Nikaido (1965) which requires the Jacobian matrix to be positive quasi-definite on convex domains and Hadamard's Global Inverse Function Theorem requiring that the operator norm of the inverse Jacobian is bounded on $\mathbb{R}^n$ (see Theorem 2 of Miller (1984)).

[9]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).

[10]Our generalized single crossing property is a sort of "twist" condition. This sort of twist condition (or stronger versions of it, called "bi-twist" conditions) show up in various theory literatures: Villani (2009) uses this condition in the optimal transport literature, Chiappori, McCann and Pass (2017) use this condition in an optimal multidimensional matching context, Figalli, Kim and McCann (2011) use this condition to illustrate when multidimensional screening problems are convex, and the contemporaneous paper Spiritus et al. (2022) make an identical assumption (parts ii. and iii. of Assumption 2') in the context of optimal multidimensional taxation.

[11]In the unidimensional setting, n denotes type and z denotes the single choice variable.

To see why Assumption 1 is helpful, suppose that $T(\mathbf{z}(\mathbf{n}))$ is differentiable as a function of $\mathbf{z}$ and $\mathbf{z}(\mathbf{n})$ is a local diffeomorphism. Under these conditions, we show in Section 3 that the following first order conditions must hold for all types $\mathbf{n}$ under any incentive compatible allocation:

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

Equivalently, we have:

$$\begin{gathered} \frac{u_{z_1}(T(\mathbf{z}),\mathbf{z};\mathbf{n})}{u_T(T(\mathbf{z}),\mathbf{z};\mathbf{n})} = -T_{z_1}(\mathbf{z}) \\ \vdots \\ \frac{u_{z_K}(T(\mathbf{z}),\mathbf{z};\mathbf{n})}{u_T(T(\mathbf{z}),\mathbf{z};\mathbf{n})} = -T_{z_K}(\mathbf{z}) \end{gathered} \tag{5}$$

By Assumption 1, no two types $\mathbf{n}$ and $\mathbf{n}'$ can simultaneously solve System 5 for a given $\mathbf{z}$. Hence, if the tax schedule is differentiable at some $\mathbf{z}$, only a single type $\mathbf{n}$ can find it optimal to locate at this given $\mathbf{z}$. Previewing ahead, this logic is going to allow us to reduce the set of relevant incentive compatibility constraints in the screening entity's problem.

While Assumption 1 does restrict the sorts of preferences our approach can handle, let us briefly discuss a number of utility functions that satisfy Assumption 1. First, utility functions of the following separable form with each $u^{(i)}$ monotonic in n_i will satisfy Assumption 1:

$$u(T,\mathbf{z};\mathbf{n}) = u_0(T,\mathbf{z}) + \sum_{i=1}^{K} u^{(i)}(\mathbf{z};n_i) \tag{6}$$

Note, in contrast to Equation 1, utility functions given by Equation 6 need not be linear in T nor linear in $\mathbf{n}$. Naturally, there are many utility functions not given by Equation 6 which also satisfy Assumption 1, such as:

$$v(z_1+z_2+T) - \frac{1}{1+\theta_1}\left(\frac{z_1}{n_1}\right)^{1+\theta_1} - \frac{1}{1+\theta_2}\left(\frac{z_2}{n_2}\right)^{1+\theta_2} - \beta\frac{z_1}{n_1}\frac{z_2}{n_2} \tag{7}$$

for some increasing, concave $v(\cdot)$ and $\theta_1,\theta_2,\beta \geq 0$. Utility function 7 satisfies Assumption 1 as long as $\mathbf{N}$ is a rectangular domain in $\mathbb{R}^2_{++}$ due to Remark 2; see Appendix B.1.

Finally, we discuss the multidimensional envelope condition. We say that an allocation $(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n}))$ satisfies the envelope condition if the indirect utility 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:

$$U(\mathbf{n}_1) - U(\mathbf{n}_2) = \int_{\mathbf{n}_2}^{\mathbf{n}_1} \nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) \cdot d\mathbf{n} \tag{8}$$

Note that the Equation 8 is *not* just the fundamental theorem of calculus because the gradient $\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n})$ is defined as the *partial* gradient representing only the derivative of u

with respect to $\mathbf{n}$ (i.e., excluding the chain rule terms involving $\mathbf{z}(\mathbf{n})$, see Section 2.3).

3 Main Results

The core difficulty with solving screening problems is that there are many incentive compatibility constraints: each type $\mathbf{n}$ must prefer their assigned allocation $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ to bundles assigned to all other types $\mathbf{n}'$. In unidimensional screening problems, we can reduce the set of relevant constraints using the following classic result (e.g., Mirrlees (1971)):

Theorem 0 (Unidimensional Incentive Compatibility). *A unidimensional allocation is incentive compatible if and only if $U(n)$ satisfies the envelope condition and $z(n)$ is (weakly) monotonic assuming preferences satisfy Assumption 1 (i.e., the single crossing property).*

In order to make multidimensional screening problems tractable, we need to reduce the set of relevant incentive constraints in some sort of similar fashion as Theorem 0. The core insight is to use the structure provided when the allocation is sufficiently smooth along with Assumption 1 to reduce the set of relevant incentive constraints. Towards this purpose, let us define "locally smooth" allocations:

Definition 4. *We say an allocation $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ is "locally smooth at $\mathbf{n}$" if $T(\mathbf{z})$ is differentiable and $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism on an open set around $\mathbf{n}$. We say an allocation is "locally smooth" if it is locally smooth $\forall \mathbf{n}$ so that $T(\mathbf{z})$ is everywhere differentiable and $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism.*

This brings us to our first main result, which is a set of necessary and sufficient conditions for an allocation to be incentive compatible *assuming* that the allocation is locally smooth:

Theorem 1. *Suppose Assumption 1 holds. Consider a locally smooth allocation $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ such that the image of $\mathbf{N}$ under the function $\mathbf{n} \mapsto (T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ is compact. Then this allocation is incentive compatible if and only if:*

1. *$U(\mathbf{n})$ satisfies the envelope condition 8*
2. *$\mathbf{n} \mapsto \mathbf{z}$ is injective*
3. *All individuals $\mathbf{n}$ satisfy:*

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

Proof. See Appendix A.1. □

Theorem 1 essentially says that if we restrict ourselves to locally smooth allocations, we can substantially reduce the set of relevant incentive compatibility constraints. For a brief sketch of the proof to Theorem 1, note that necessity of the envelope condition follows directly from Milgrom and Segal (2002). Necessity of injective $\mathbf{n} \mapsto \mathbf{z}$ results from Assumption 1 given the allocation is locally smooth (otherwise we would have multiple types $\mathbf{n}$ whose first order conditions are satisfied at the same $\mathbf{z}$).[12] The idea behind showing sufficiency is also to leverage Assumption 1. If the envelope condition 8 holds, the transfer schedule is differentiable, and $\mathbf{n} \mapsto \mathbf{z}$ is both locally diffeomorphic and injective (i.e., diffeomorphic), then 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. Hence, as long as 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.

Theorem 1 is in some ways a direct multidimensional analogue of the classic unidimensional incentive compatibility result, Theorem 0, where the injectivity requirement replaces the monotonicity requirement. But Theorem 1 also has an additional condition compared to Theorem 0: all individuals must prefer their assigned bundle to all boundary bundles. Why? Well, the "trick" behind Theorem 1 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 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.[13] 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. Theorem 1 requires additional conditions on boundary bundles compared to Theorem 0 precisely because the mean value theorem does not hold for vector valued functions.

However, there is an inconsistency with the discussion so far because Theorem 1 only applies to locally smooth allocations. Yet we motivated the need for new results about incentive compatibility so as to solve multidimensional screening problems with more complex utility functions when first order methods fail; but if the optimal allocation is locally smooth then we *can* use first order methods. Hence, we need to extend Theorem 1 to allocations which are not locally smooth:

Lemma 1. *For any convergent sequence of incentive compatible and locally smooth* $(T_j(\mathbf{z}_j(\mathbf{n})), \mathbf{z}_j(\mathbf{n}))$*, indexed by* j*, the (pointwise) 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.2. □

[12] Spiritus et al. (2022) also show that injective $\mathbf{n} \mapsto \mathbf{z}$ is necessary for incentive compatibility under Assumption 1, see their Lemma 1.

[13] 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 $x \to \infty$.

Lemma 1 shows that the results from Theorem 1 can be extended to many non-smooth allocations via a simple limiting argument: the limit of locally smooth allocations satisfying the conditions of Theorem 1 is incentive compatible. Hence, we are going to make the following assumption henceforth:

Assumption 2. *The optimal allocation is the (pointwise) limit of a sequence of locally smooth allocations; hence, the optimal allocation can be approximated arbitrarily well with locally smooth allocations.*

Assumption 2 always holds in the unidimensional case on a compact domain as a result of the fact that every (weakly) monotonic function can be arbitrarily approximated by a strictly monotonic differentiable function.[14] Assumption 2 also always holds when preferences are given by Equation 1. In this case, the set of incentive compatible allocations is the set of allocations satisfying the envelope condition with convex $U(\mathbf{n})$ while the set of locally smooth incentive compatible allocations is equal to the set of allocations satisfying the envelope condition with differentiable, strictly convex $U(\mathbf{n})$ (see Appendix C for details). And it turns out that all convex functions can be approximated arbitrarily well by differentiable, strictly convex functions (Koliha, 2004) so that Assumption 2 also holds when preferences are given by Equation 1. In more general multidimensional settings, I conjecture that Assumption 2 is satisfied for all realistic allocations, but there are likely devilish counter-examples of incentive compatible allocations which cannot be approximated by locally smooth allocations.[15]

Thus, as long as Assumption 2 holds, we can operationalize Theorem 1 (along with Lemma 1) by rewriting the optimal screening problem as follows (recognizing that if the optimal allocation is not locally smooth but rather the limit of locally smooth allocations, then the optimal allocation is the argsup rather than the argmax of System 9):

$$\begin{aligned}
&\sup_{\mathbf{z}(\mathbf{n}),T(\mathbf{z}(\mathbf{n}))} \quad \text{Objective} \\
&\text{s.t. } U(\mathbf{n}) \text{ satisfies the envelope condition 8} \\
&\mathbf{n} \mapsto \mathbf{z} \text{ is injective and locally diffeomorphic} \\
&u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}) \ \forall \mathbf{n}' \in \partial\mathbf{N}, \mathbf{n} \in \mathbf{N} \\
&\text{Other Constraints}
\end{aligned} \tag{9}$$

From a practical perspective, using Theorem 1 to simplify System 4 into System 9 allows

[14] Hence, every incentive compatible $z(n)$ can be arbitrarily approximated by a locally smooth allocation. And any associated incentive compatible $T(z)$ can also be approximated by a differentiable function because $U(n)$ is (absolutely) continuous and can therefore be arbitrarily approximated by differentiable functions by the Stone-Weierstrass Theorem. And allocations with differentiable $U(n)$ (and differentiable monotonic $z(n)$) must have differentiable $T(z)$ as we prove in Lemma 3 in Appendix A.5.

[15] For example, there are continuous bijective mappings in higher dimensions that cannot be approximated by diffeomorphisms, typically using Cantor set type constructions, see Hencl and Vejnar (2015).

us to significantly reduce the set of relevant incentive compatibility constraints when solving screening problems. Suppose we are trying to solve a screening problem on a discretized square grid of size $m \times m$. If we naively attempted to impose all incentive compatibility constraints, we would have $m^4 - m^2 = m^2(m^2 - 1)$ constraints as each of the m^2 types could potentially deviate to any of the other $m^2 - 1$ choices. Under the conditions in Theorem 1, 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 Remarks 1 and 2 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 1 reduces the number of incentive compatibility constraints we need to check from ≈ 1 trillion down to ≈ 4 billion, which is a considerable improvement.

However, while Theorem 1 is useful because it allows us to reduce the number of relevant incentive compatibility constraints by an order of magnitude, System 9 is still quite difficult to solve. For instance, as we just discussed, solving a screening problem using System 9 on a 1000×1000 grid requires us to impose ≈ 4 billion constraints, which is still likely far too many to be of practical use. Hence, Theorem 1 is most likely primarily useful for *verifying* that a given proposed solution is in fact incentive compatible rather than being useful to actually construct the optimal allocation.

In order to actually construct an optimal solution to a screening problem, we need to further reduce the number of incentive compatibility constraints. In particular, can we reduce the set of relevant incentive compatibility constraints down to just local constraints (i.e., local first and second order conditions)? Rochet (1987) and McAfee and McMillan (1988) both show that first and second order conditions are sufficient for incentive compatibility when utility is given by Equation 1 because the linearity and separability ensures that the mean value theorem *does hold* for these specific vector valued functions. But extending this sufficiency result to the multidimensional setting for general utility functions appears to be impossible because the mean value theorem does not hold for vector valued functions. Nonetheless, we can come up with a set of *necessary* conditions involving first and second order conditions that we can use to actually construct a candidate optimal allocation.

Building towards these ideas, note that when the allocation is locally smooth, the following first order condition must hold for all types $\mathbf{n}$ under any incentive compatible allocation:

$$\left\{u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})\right\} \nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n}) = 0 \tag{10}$$

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

$$FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) \equiv u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = 0 \tag{11}$$

Rewriting $\mathbf{n}$ (locally) as a function of $\mathbf{z}$ rather than the reverse, Equation 11 implies:

$$\nabla_{\mathbf{z}} T(\mathbf{z}) = -\frac{\nabla_{\mathbf{z}} u(T(\mathbf{z}), \mathbf{z}; \mathbf{n}(\mathbf{z}))}{u_T(T(\mathbf{z}), \mathbf{z}; \mathbf{n}(\mathbf{z}))} \tag{12}$$

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, because $\mathbf{n}(\mathbf{z})$ and $T(\mathbf{z})$ are both continuously differentiable in $\mathbf{z}$, Equation 12 implies that $\nabla_{\mathbf{z}} T(\mathbf{z})$ is also continuously differentiable. Hence, we have shown that:

Lemma 2. *Whenever $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ is incentive compatible and locally smooth at $\mathbf{n}$, $T(\mathbf{z}(\mathbf{n}))$ is twice continuously differentiable in $\mathbf{z}$. Hence, we can totally differentiate Equation 11:*

$$D_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) = \nabla_{\mathbf{z}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) \nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n}) + \nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) = 0 \tag{13}$$

So under the conditions of Lemma 2, the matrix of second partial derivatives of $u(T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ with respect to $\mathbf{z}$ is equal to:

$$\nabla_{\mathbf{z}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) = -\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) [\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$$

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. Thus, we have the following definition:

Definition 5. *We say that an allocation $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ satisfies the second order condition at a given $\mathbf{n}$ whenever $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) [\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1}$ is positive definite.*

This brings us to our second main result:

Theorem 2. *Suppose Assumption 1 holds. Consider an allocation $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ such that the image of $\mathbf{N}$ under the function $\mathbf{n} \mapsto (T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ is compact. The following three conditions are necessary for this allocation to be incentive compatible:*

1. *$U(\mathbf{n})$ must satisfy the envelope condition 8*
2. *At points $\mathbf{n}$ where the allocation is locally smooth then the second order condition holds (as in Definition 5)*
3. *If the allocation is locally smooth at two points $\mathbf{n}$ and $\mathbf{n}'$ then $\mathbf{z}(\mathbf{n}) \neq \mathbf{z}(\mathbf{n}')$.*

Proof. See Appendix A.3. □

Note, Theorem 2 applies to non-smooth schedules even though points (2) and (3) rely entirely on the structure generated when the allocation is locally smooth; hence, Theorem 2 places no restrictions on the portions of an allocation which are not locally smooth other than the envelope condition and the implicit restriction that any two locally smooth portions of the mapping $\mathbf{n} \mapsto \mathbf{z}$ must be globally injective. As such, there are likely some pathological allocations which are not incentive compatible yet do not violate any criteria specified in Theorem 2.

However, if the optimal allocation satisfies Assumption 2, then there is no loss in restricting ourselves to smooth allocations (because we can always construct an arbitrarily good approximation to the optimal allocation using smooth allocations). If we restrict to locally smooth allocations, we get the following Corollary that follows immediately from Theorem 2:

Corollary 2.1. *Suppose the utility function satisfies Assumption 1. Consider a locally smooth allocation* $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ *such that the image of* $\mathbf{N}$ *under the function* $\mathbf{n} \mapsto (T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ *is compact. The following three conditions are necessary for this allocation to be incentive compatible:*

1. $U(\mathbf{n})$ *must satisfy the envelope condition 8*
2. *The second order condition holds* $\forall \mathbf{n}$ *(as in Definition 5)*
3. $\mathbf{n} \mapsto \mathbf{z}$ *is injective*

Corollary 2.1 can, in principle, be used to construct potential solutions to multidimensional screening problems by maximizing the screening entity's objective subject to the reduced set of constraints that (1) the allocation is locally smooth, (2) the allocation satisfies the envelope condition, (3) the allocation satisfies second order conditions, and (4) the allocation is injective. Then, once we have a candidate solution that satisfies these necessary conditions, we can verify the potential solution using the sufficient conditions from Theorem 1. The one complication is that the set of injective functions is not necessarily easy to characterize in a simple manner; thus, in practice, we will likely have to replace the injectivity requirement with a local injectivity requirement (and then check global injectivity *ex post*) or restrict to utility functions for which local injectivity implies global injectivity. We discuss this more in Section 4.

Finally, we note that Theorem 1 and Theorem 2 can 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 Incentive Compatibility Results

It is worthwhile to mention how the results in Section 3 compare to existing results on incentive compatibility. While the details are relegated to Appendix C, the key takeaways are as follows: (1) Theorems 1 and 2 along with Lemma 1 collapse to Theorem 0 in the unidimensional setting; (2) Theorems 1 and 2 along with Lemma 1 collapse to the convexity characterization from Rochet (1987) when utility is given by Equation 1; (3) our results build on McAfee and McMillan (1988) because our generalized single crossing property applies to a broader set of utility functions; and (4) our results build on Carlier (2001) insofar as our results can be applied to solve multidimensional screening problems numerically whereas the conditions from Carlier (2001) are highly difficult to apply in practice.[16]

[16]The necessary condition in Theorem 2 only fully coincides with the necessary condition in Theorem 0 in the unidimensional case under a differentiability assumption on $z(n)$. Similarly, we require allocations to be locally

4 Solving 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? The core idea is to use (most of) the necessary conditions from Corollary 2.1 and search over all allocations such that: (1) $U(\mathbf{n})$ is differentiable and satisfies the envelope condition 8, (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).[17] We impose that $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism rather than a diffeomorphism (i.e., an injective local diffeomorphism) simply because local diffeomorphisms are easy to characterize whereas diffeomorphisms are not easily characterized in terms of local properties of the mapping.

We can further simplify the screening entity's problem by using a few standard tricks. Instead of considering the screening entity as maximizing their objective over $(\mathbf{z}(\mathbf{n}), T(\mathbf{z}(\mathbf{n})))$, we can consider the screening entity as choosing $(\mathbf{z}(\mathbf{n}), U(\mathbf{n}))$, implicitly defining $T(\mathbf{z}(\mathbf{n}))$ via $U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$.[18] Moreover, the screening entity only actually gets to choose $\mathbf{z}(\mathbf{n})$ along with 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 $\underline{\mathbf{n}}$. Additionally, because we assume $U(\mathbf{n})$ is differentiable, we can express the envelope condition in differentiable form, but this requires us to include an additional constraint that the gradient of $U(\mathbf{n})$ forms a conservative vector field. Putting all of this together, our general technique for solving multidimensional screening problems under Assumptions 1 and 2 is to solve:[19]

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

System 14 is an optimization problem with linear and non-linear constraints which can be solved to find the global supremum of the objective function over the set of locally smooth allocations. Importantly, the optimal allocation may be the argsup of System 14 (rather than the

smooth in order for Theorem 2 to fully coincide with the necessary condition from Rochet (1987) when utility is given by Equation 1. See Appendix C for further explanation.

[17] If $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $U(\mathbf{n})$ is differentiable, then Lemma 3 in Appendix A.5 shows that $T(\mathbf{z})$ must be differentiable.

[18] The mapping $U(\mathbf{n}) \mapsto T(\mathbf{z}(\mathbf{n}))$ defined by $U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ is bijective because $u_T > 0$.

[19] WLOG, we assume $\det(\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})) > 0$ rather than $\det(\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})) < 0$.

argmax) if the optimal allocation is not locally smooth (e.g., features bunching).[20] Conceptually, System 14 is simply maximizing the screening entity's objective subject to *both* individual first and second order conditions.[21] This explicit inclusion of second order conditions is the key difference relative to first order approaches. By ignoring second order conditions, first order approaches will typically find an allocation that is not incentive compatible when the optimal allocation features bunching. In contrast, our approach can approximate bunching within the set of incentive compatible allocations by finding an allocation where $\det(\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n}))$ is everywhere non-zero yet arbitrarily small so that the mapping $\mathbf{n} \mapsto \mathbf{z}$ is arbitrarily close to being non-invertible; hence, many different types choose almost exactly the same $\mathbf{z}$. In this case, the numerical solution to System 14 will be a locally smooth allocation that approximates bunching as closely as desired by setting an arbitrarily low optimization tolerance.

Once we have a proposed solution from System 14, we can verify that the solution is incentive compatible by appealing to Theorem 1 and verifying: (1) $\mathbf{n} \mapsto \mathbf{z}$ is injective, which can often be checked via Remark 2, and (2) all individuals prefer their assigned bundles to boundary bundles. If these conditions are satisfied then the proposed solution is globally optimal among locally smooth allocations.

4.1 Separable Utility

It turns out that we can strengthen our results and simplify System 14 substantially if we restrict ourselves to utility functions which have the nice separable form:

$$u(T, \mathbf{z}; \mathbf{n}) = u^{(0)}(T, \mathbf{z}) + \sum_{i}^{K} u^{(i)}(z_i, n_i) \tag{15}$$

When utility is given by Equation 15, we can prove the following two propositions:

Proposition 1. *Consider any utility function given by Equation 15 on a rectangular domain* $\mathbf{N}$ *for which Assumption 1 holds. 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.6. □

Proposition 2. *Consider any utility function given by Equation 15. 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.7. □

[20] When the optimal allocation is not locally smooth, the numerical solution to System 14 will be a locally smooth allocation under which the value of the objective function is made close (i.e., within some optimization tolerance) to the supremum of the objective function.

[21] Technically, System 14 also ensures that $\mathbf{n} \mapsto \mathbf{z}$ is locally smooth (so that $\det(\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})) > 0$) and that $\mathbf{n} \mapsto \mathbf{z}$ is consistent with individual maximization (so that $\nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ is conservative).

Proposition 1 allows us to drop the constraint that $\det(\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})) > 0$ from System 14 and Proposition 2 allows us to drop the conservative requirement from System 14, yielding a simpler system when utility is given by Equation 15:[22]

$$\begin{aligned}
&\sup_{\mathbf{z}(\mathbf{n}),U(\underline{\mathbf{n}})} \quad \text{Objective} \\
&\text{s.t. } \nabla_{\mathbf{n}}U(\mathbf{n}) = \nabla_{\mathbf{n}}u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) \\
&\nabla_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n});\mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1} \text{ is positive definite} \\
&U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n}) \\
&\text{Other Constraints}
\end{aligned} \tag{16}$$

Provided utility is given by Equation 15, if we find a solution to System 16 then all we need to do is check that all individuals prefer their assigned bundles to boundary bundles;[23] if so, we have found the optimal allocation within the class of allocations which can be approximated by locally smooth allocations.

4.2 Relationship to Previous Numerical Methods

It is worthwhile to reiterate how our numerical method should be viewed relative to existing methods for solving multidimensional screening problems. Using a first order approach is typically the easiest option when second order conditions never bind and bunching does not occur. As long as the resulting partial differential equation from the Euler-Lagrange equation is not extraordinarily complex, finite difference or finite element methods can be used to solve these sorts of problems. When the optimal allocation features bunching or is otherwise non-smooth (e.g., features individuals with multiple optima) *and* utility is given by Equation 1 (i.e., linear and separable in type $\mathbf{n}$), then methods which solve variation calculus problems subject to convexity constraints will almost certainly be the fastest numerical method. Such methods have been developed by, for example, Aguilera and Morin (2008), Oberman (2013), Mérigot and Oudet (2014). Boerma, Tsyvinski and Zimin (2022) improves upon these methods using a Legendre transformation to further improve speed and scalability.

When first order approaches cannot be applied (e.g., due to bunching) and utility is *not* given by Equation 1, then the numerical method outlined in Section 4 is, as far as I know, the only algorithm which can be applied to solve multidimensional screening problems.[24] Thus, we believe developing a new numerical method to solve multidimensional screening problems is an

[22] As pointed out by Spiritus et al. (2022) and Krasikov and Golosov (2022), maximizing an objective function over the vector field $\mathbf{z}(\mathbf{n})$ using first order approaches is problematic because the proposed solution may not satisfy the necessary conservative vector field requirement on $\nabla_{\mathbf{n}}u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n})$. Proposition 2 ensures that by also including the second order constraint that $\nabla_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n});\mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1}$ is positive definite $\forall\mathbf{n}$, we *can* maximize over $\mathbf{z}(\mathbf{n})$ and be ensured a solution where $\nabla_{\mathbf{n}}u(T(\mathbf{z}(\mathbf{n})),\mathbf{z}(\mathbf{n});\mathbf{n})$ is a conservative vector field.

[23] Any allocation that solves System 16 necessarily features diffeomorphic $\mathbf{n} \mapsto \mathbf{z}$ by Proposition 1.

[24] Note, the utility function discussed in Section 4.1 is *separable* in type $\mathbf{n}$ but *not necessarily linear* in type $\mathbf{n}$, distinguishing it from utility given by Equation 1.

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. We also prove Proposition 4, which shows that the optimal tax schedule is non-smooth (i.e., features bunching and/or individuals with multiple optima) whenever the social planner has sufficiently redistributive preferences.

5.1 Optimal Multidimensional Taxation Objective

We will consider the following optimal taxation problem, where agents 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 $\mathbf{n}$.[25] The government chooses the function $T(\mathbf{z})$ to maximize a welfare function subject to a revenue constraint that total transfers in society 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 $F(\mathbf{n})$. Appealing to the revelation principle and formulating this problem as in System 4 the government solves:[26]

$$
\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}), T(\mathbf{z}(\mathbf{n}))} \int_{\mathbf{N}} W\left(U(\mathbf{n}), \mathbf{n}\right) dF(\mathbf{n}) \\
&\text{s.t.} \int_{\mathbf{N}} T(\mathbf{z}(\mathbf{n})) dF(\mathbf{n}) \leq 0 \\
&\mathbf{n} \in \underset{\mathbf{n}'}{\text{argmax}}\; u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}) \; \forall \mathbf{n} \\
&U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})
\end{aligned}
\tag{17}
$$

System 17 can represent many different multidimensional optimal taxation problems, such as: (1) joint taxation of couples (where $K = 2$, z_1 and z_2 represent the two labor incomes), (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 and 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 15. First, we consider utility

[25] 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.

[26] We prove and discuss an existence result for the optimal taxation problem given by System 17 in Appendix B.2, but leave this out of the main text for the sake of brevity and streamlining.

function 18:[27]

$$u(T, \mathbf{z}; \mathbf{n}) = z_1 + z_2 + T + n_1 \frac{z_1^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2^{1+\theta_2}}{1+\theta_2} \tag{18}$$

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

Applying our method to utility function 18 is useful as a test case because utility function 18 takes the form of Equation 1. Hence, we can check that the solutions from our numerical methods align with the solutions from numerical methods designed to solve variational calculus problems with convexity constraints. In particular, when utility is given by Equation 18, we can express $T(\cdot)$ as a function of $U(\mathbf{n})$ and $\nabla_{\mathbf{n}} U(\mathbf{n})$ (see Remark 6 in Appendix D); moreover, we can replace the incentive compatibility constraints with a convexity constraint as a result of Rochet (1987):[28]

$$\begin{aligned}
&\max_{U(\mathbf{n})} \int_{\mathbf{N}} W\left(U(\mathbf{n}), \mathbf{n}\right) f(\mathbf{n}) d\mathbf{n} \\
&\text{s.t. } \int_{\mathbf{N}} T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})) dF(\mathbf{n}) \leq 0 \\
&U(\mathbf{n}) \text{ is convex}
\end{aligned} \tag{19}$$

A number of algorithms have been developed to solve variational calculus problems subject to convexity constraints like System 19, e.g., Aguilera and Morin (2008), Oberman (2013), Mérigot and Oudet (2014), or Boerma, Tsyvinski and Zimin (2022).

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) because utility is *not* of the form in Equation 1; hence, methods based on convexity of $U(\mathbf{n})$ no longer apply and first order approaches will also fail to recover the optimal allocation if it features bunching:

$$u(T, \mathbf{z}; \mathbf{n}) = z_1 + z_2 + T + n_1 \frac{z_1^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2^{1+\theta_2}}{1+\theta_2} - \frac{n_1^2}{2\alpha} z_1 - \frac{n_2^2}{2\alpha} z_2 \tag{20}$$

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 households differ in terms of n_1, n_2, which govern disutility over generating income. We can nest utility function 18 within utility function 20 if we set $\alpha = \infty$. Hence, we will just discuss how to apply our method to utility function 20. First, note:

[27] Note, for those familiar with optimal taxation models, the parametrization may seem a bit strange. For instance, a more conventional parametrization would be to have $u(T, l_1, l_2; m_1, m_2) = m_1 l_1 + m_2 l_2 + T - \frac{l_1^{1+\theta_1}}{1+\theta_1} - \frac{l_2^{1+\theta_2}}{1+\theta_2}$, 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).

[28] We prove in Appendix B.3 that $\mathbf{z}(\mathbf{n})$ must be continuous (i.e., individuals cannot have multiple optima) when preferences are given by utility function 18.

Proposition 3. *Utility function 20 (and hence utility function 18) satisfies Assumption 1 on rectangular domains.*

Proof. See Appendix A.8. $\square$

Let us now discuss how to apply the numerical method outlined in Section 4 for utility function 20. Because utility function 20 takes the form of Equation 15, we can solve System 16. We need to first ensure that $U(\mathbf{n})$ satisfies the envelope condition so that $\forall \mathbf{n}$:

$$\nabla_{\mathbf{n}} U(\mathbf{n}) = \nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) = \begin{bmatrix} \frac{z_1(\mathbf{n})^{1+\theta_1}}{1+\theta_1} - \frac{n_1}{\alpha} z_1(\mathbf{n}) \\ \frac{z_2(\mathbf{n})^{1+\theta_2}}{1+\theta_2} - \frac{n_2}{\alpha} z_2(\mathbf{n}) \end{bmatrix}$$

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 $\forall \mathbf{n}$:

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1} = \begin{bmatrix} z_1(\mathbf{n})^{\theta_1} - \frac{n_1}{\alpha} & 0 \\ 0 & z_2(\mathbf{n})^{\theta_2} - \frac{n_2}{\alpha} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) \end{bmatrix}^{-1} \tag{21}$$

Equation 21 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:

$$\left(z_1^{\theta_1}(\mathbf{n}) - \frac{n_1}{\alpha}\right) \frac{\partial z_1}{\partial n_2}(\mathbf{n}) = \left(z_2(\mathbf{n})^{\theta_2} - \frac{n_2}{\alpha}\right) \frac{\partial z_2}{\partial n_1}(\mathbf{n}) \tag{22}$$

And in order for Equation 21 to have all positive principal minors we must have:

$$\begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) \end{bmatrix} \tag{23}$$

is a P matrix (i.e., all principal minors are positive so that $\frac{\partial z_1}{\partial n_1}, \frac{\partial z_2}{\partial n_2}, \frac{\partial z_1}{\partial n_1}\frac{\partial z_2}{\partial n_2} - \frac{\partial z_1}{\partial n_2}\frac{\partial z_2}{\partial n_1} > 0$).[29] Hence, for utility function 20, the government's problem is given by:

$$\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}), U(\underline{\mathbf{n}})} \int_{\mathbf{N}} W(U(\mathbf{n}), \mathbf{n}) dF(\mathbf{n}) \\
&\text{s.t. } \int_{\mathbf{N}} T(\mathbf{z}(\mathbf{n})) dF(\mathbf{n}) \leq 0 \\
&\nabla_{\mathbf{n}} U(\mathbf{n}) = \begin{bmatrix} \frac{z_1(\mathbf{n})^{1+\theta_1}}{1+\theta_1} - \frac{n_1}{\alpha} z_1(\mathbf{n}) \\ \frac{z_2(\mathbf{n})^{1+\theta_2}}{1+\theta_2} - \frac{n_2}{\alpha} z_2(\mathbf{n}) \end{bmatrix} \\
&\frac{\partial z_1}{\partial n_1}(\mathbf{n}) > 0, \frac{\partial z_2}{\partial n_2}(\mathbf{n}) > 0, \frac{\partial z_1}{\partial n_1}(\mathbf{n})\frac{\partial z_2}{\partial n_2}(\mathbf{n}) - \frac{\partial z_1}{\partial n_2}(\mathbf{n})\frac{\partial z_2}{\partial n_1}(\mathbf{n}) > 0 \\
&\left(z_1^{\theta_1}(\mathbf{n}) - \frac{n_1}{\alpha}\right) \frac{\partial z_1}{\partial n_2}(\mathbf{n}) = \left(z_2(\mathbf{n})^{\theta_2} - \frac{n_2}{\alpha}\right) \frac{\partial z_2}{\partial n_1}(\mathbf{n}) \\
&T(\mathbf{z}(\mathbf{n})) = U(\mathbf{n}) - \left[n_1 \frac{z_1(\mathbf{n})^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2(\mathbf{n})^{1+\theta_2}}{1+\theta_2} - \frac{n_1^2}{2\alpha} z_1(\mathbf{n}) - \frac{n_2^2}{2\alpha} z_2(\mathbf{n})\right] - z_1(\mathbf{n}) - z_2(\mathbf{n})
\end{aligned} \tag{24}$$

[29] The fact that Equation 21 has positive principal minors (i.e., is a P matrix) if and only if Equation 23 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.

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.[30] Once we have a candidate solution for System 24, we verify that this solution is in fact incentive compatible using Theorem 1. If we solve the problem on a rectangular domain and utility is given by Equation 18 or 20, Proposition 1 ensures that the solution from System 24 features diffeomorphic $\mathbf{n} \mapsto \mathbf{z}$; in this case, we must only check that $\forall \mathbf{n}$ we have $u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \geq u(T(\mathbf{z}(\mathbf{n}')), \mathbf{z}(\mathbf{n}'); \mathbf{n}) \ \forall \ \mathbf{n}' \in \partial \mathbf{N}$.

5.3 Four Illustrative Simulations

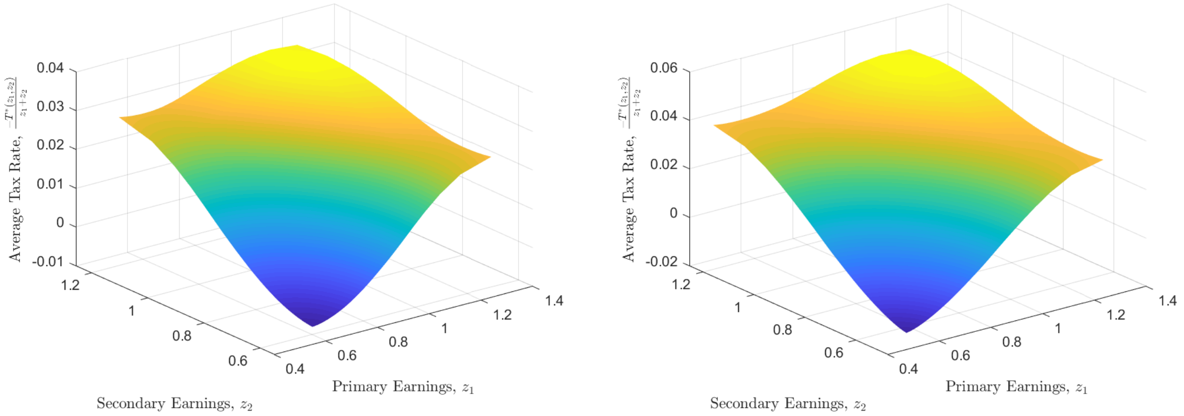
Next, we work through four toy examples, which are meant to illustrate the above simulation method rather than closely depict reality. For each of the two utility functions 18 and 20, we consider two scenarios, which feature different welfare weights, $\psi(\mathbf{n})$. We suppose that the social welfare function is weighted utilitarian: $W(U(\mathbf{n}), \mathbf{n}) = \psi(\mathbf{n})U(\mathbf{n})$. Welfare weights $\psi(\mathbf{n})$ are chosen so that the marginal social welfare gain from increasing utility is decreasing with $\mathbf{n}$ (similar to, for example, Lockwood and Weinzierl (2016)). In both examples we consider a rectangular domain of $[-6, -0.5]^2$ with normally distributed $f(\mathbf{n})$ with a small positive co-variance between n_1 and n_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.[31] These welfare weights $\psi(\mathbf{n})$ imply that the government cares more about increasing consumption for low income households than high income households. For our first set of simulations, we assume that the government has weak redistributive preferences. We set $\beta = 1$, which, given the range of $\mathbf{n}$ we consider, means that the marginal social welfare gain of giving the lowest income household a dollar is about 3 times higher than giving the highest income household a dollar under zero taxes (the highest income household earns about 2.5 times as much as the lowest income household).

For $\beta = 1$ and both utility functions 18 and 20, we find that the solution to System 24 does not feature any bunching (and could therefore, in principle, be computed using a first order approach). Moreover, the solution to System 24 is incentive compatible by Theorem 1 because all individuals prefer their assigned bundle to boundary bundles. 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 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 gain of giving \$1 to the lowest income household is only about 3 times higher

[30] Our simulation algorithm uses Matlab's "fmincon" algorithm to solve for the optimal schedule given the non-linear equality and inequality constraints in System 24.

[31] $f(\mathbf{n})$ is joint normal with mean matrix $\mu = \begin{bmatrix} -5.5 \\ -5.5 \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} 5 & 0.5 \\ 0.5 & 5 \end{bmatrix}$.

than giving \$1 to the highest income household; the maximum average tax rate is 4% (5% for utility function 20) and the maximum marginal tax rate is about 13% (17% for utility function 20).



(a) Utility Function 18

(b) Utility Function 20

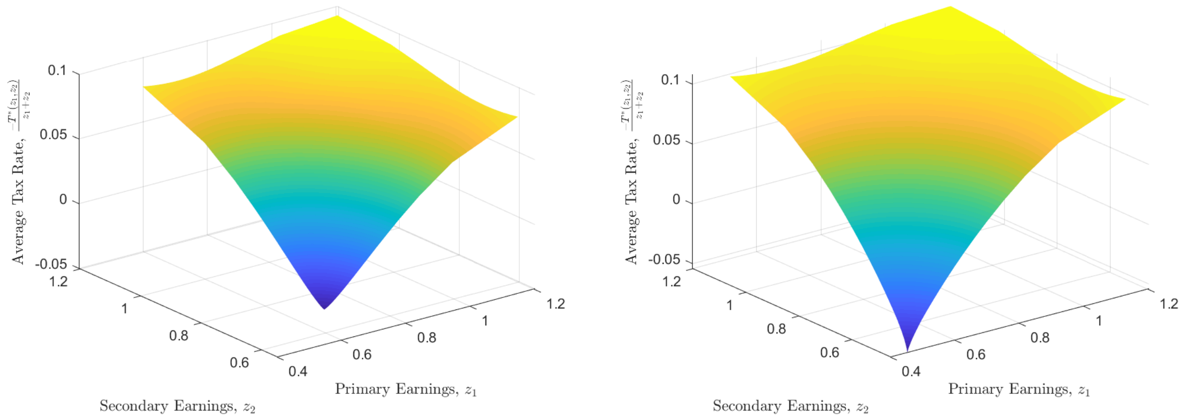
Figure 1: Optimal Average Tax Rates, Weak Redistributive Preferences

Note: This figure shows the optimal average tax schedule $\frac{-T^*(z_1,z_2)}{z_1+z_2}$ under weak redistributive preferences. We set $\theta_1 = \theta_2 = 3$ (corresponding to a compensated taxable income elasticity of 1/3 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 $\Sigma = \begin{bmatrix} 5 & 0.5 \\ 0.5 & 5 \end{bmatrix}$. $\psi(\mathbf{n}) = e^{-U_0(\mathbf{n})}$ 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. Results for utility function 18 are shown in panel 1a and results for utility function 20 are shown in panel 1b.

For our second set of simulations, we assume that the social planner has stronger redistributive preferences: we set $\beta = 10.9$ (and $\beta = 8.1$ for utility function 20), 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.[32] The optimal tax schedule from solving System 24 is shown in Figure 2. Tax rates are much higher: for utility function 18 the maximum average tax rate is around 9% (11% for utility function 20) and the maximum marginal tax rate is around 34% (36% for utility function 20).

With stronger redistributive preferences, we find that it is optimal for the government to set a tax schedule which induces "bunching" at the bottom of the income distribution (for both utility functions 18 and 20), 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 income distribution, implying that

[32] 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, 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.

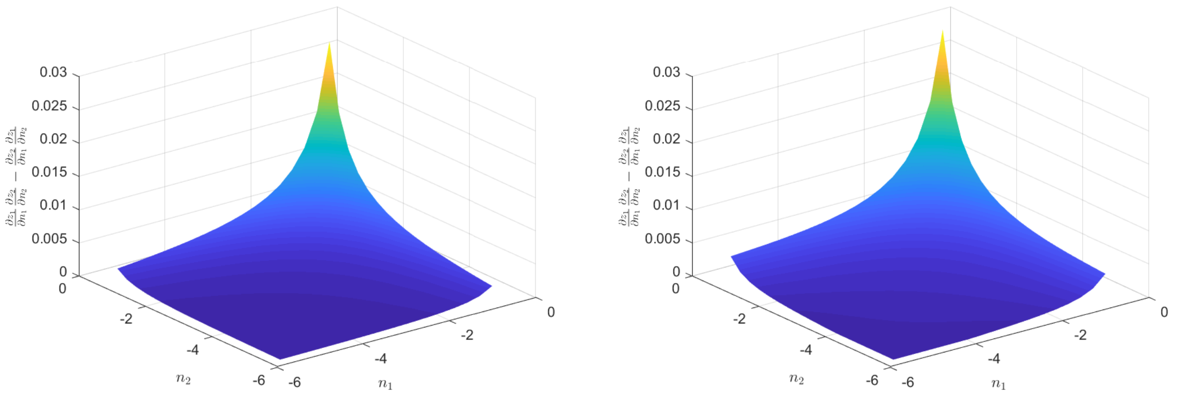


(a) Utility Function 18 (b) Utility Function 20

Figure 2: Optimal Average Tax Rates: Strong Redistributive Preferences

Note: This figure shows the optimal average tax schedule $\frac{-T^*(z_1,z_2)}{z_1+z_2}$ under strong redistributive preferences. Parameters are the same as in Figure 1 other than welfare weights, which are chosen so that $\psi(\mathbf{n}) = e^{-\beta U_0(\mathbf{n})}$ where $U_0(\mathbf{n})$ is optimal utility under zero taxes, $\beta = 10.9$ for utility function 18, and $\beta = 8.1$ for utility function 20. This sets the marginal social welfare of giving the lowest income household a dollar to be roughly 100,000 times higher than giving the highest income household a dollar for both utility functions. Results for utility functions 18 are shown in panel 2a and results for utility function 20 are shown in panel 2b.

$\mathbf{n} \mapsto \mathbf{z}$ is not invertible near the bottom of the income distribution.[33] Because of the presence of bunching and the fact that utility function 20 does not take the simple linear/separable form of Equation 1, constructing an optimal solution to this screening problem falls outside the realm of currently available methods.



(a) Utility Function 18 (b) Utility Function 20

Figure 3: Jacobian Determinant, Solution to System 24 for Strong Redistributive Preferences

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 24 for utility functions 18 (panel 3a) and 20 (panel 3b) under the same parameters as in Figure 2.

The above numerical examples suggest that bunching occurs in multidimensional optimal

[33]As discussed, our numerical procedure to compute optimal schedules when utility is given by Equation 18 or 20 enforces that $\mathbf{n} \mapsto \mathbf{z}$ is a diffeomorphism because second order conditions holding strictly implies 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.

taxation problems when society has sufficiently strong redistributive preferences. It turns out that we can formalize this idea theoretically. First, let us suppose that our problem satisfies the following regularity conditions:[34]

Assumption 3. *Suppose that* $\mathbf{N}$ *is a compact rectangle and preferences are given by Equation 15. Also suppose that* $\frac{\partial u^{(0)}(T,\mathbf{z})}{\partial T}$ *is bounded away from both 0 and* ∞ *and that* $-\left(\frac{\partial u^{(0)}}{\partial z_i}+\frac{\partial u^{(i)}}{\partial z_i}\right)/\left(\frac{\partial^2 u^{(i)}}{\partial n_i \partial z_i}\right)$ *is bounded from below for all* $\mathbf{n}, \mathbf{z}, T$.

Under Assumption 3, we can prove the following:[35]

Proposition 4. *Suppose Assumption 3 holds. Let* $B_\epsilon(\underline{\mathbf{n}})$ *denote the* ϵ *ball around the lowest type* $\underline{\mathbf{n}}$. *When* $dim(\mathbf{N}) \geq 2$, *first order approaches will fail to characterize the optimal tax schedule whenever social preferences are sufficiently redistributive in the sense that* $\frac{\int_{B_\epsilon(\underline{\mathbf{n}})}\psi(\mathbf{n})dF(\mathbf{n})}{\int_{\mathbf{N}}\psi(\mathbf{n})dF(\mathbf{n})} \approx 1$. *If, additionally, utility is given by Equation 1, then bunching will necessarily occur.*

Proof. See Appendix A.9. □

To the best of my knowledge, Proposition 4 is the first result deriving exogenous conditions under which first order approaches will necessarily fail and/or bunching is guaranteed to occur in a multidimensional optimal taxation setting. As discussed in Section 2.2, first order approaches fail to generate an incentive compatible allocation if the optimal allocation features bunching *or* if individuals have multiple optima. When utility is given by Equation 1, we can infer that bunching must occur for sufficiently redistributive welfare weights because the optimal allocation never features individuals with multiple optima when utility is given by Equation 1 (see Proposition 6 of Rochet and Chone (1998) or Appendix B.3 for a proof). However, when utility is given by the more general Equation 15 and welfare weights are sufficiently redistributive, we suspect that bunching will typically occur given that first order methods fail, but we cannot rule out the possibility that first order methods fail due to individuals with multiple optima.

Bunching was shown to be robust in the context of multiproduct monopolists by both Armstrong (1996) and Rochet and Chone (1998) due to the presence of participation constraints. While optimal multidimensional taxation problems typically do not feature participation constraints (as we assume that individuals cannot leave the jurisdiction), the intuition for why bunching occurs turns out to be similar. For those familiar with screening problems in the context of multiproduct monopolists, suppose utility is given by Equation 1 and consider the limiting case when the social planner cares only about the utility of the lowest type $\underline{\mathbf{n}}$. In this case, the optimal taxation problem is equivalent to maximizing tax revenue subject to a

[34]Note that both utility functions 18 and 20 satisfy Assumption 3; see Appendix A.9.
[35]Thanks to the two anonymous referees who suggested I derive this result and provided guidance on the intuitive relationship to Armstrong (1996).

participation constraint (Boadway and Jacquet, 2008); hence, the logic of Armstrong (1996) ensures that bunching must occur at the bottom of the distribution (the proof of Proposition 4 is conceptually similar to the proof of Proposition 1 in Armstrong (1996)). For those more familiar with screening problems in the context of multidimensional optimal taxation, consider the following heuristic derivation of Proposition 4. First, note that if first order approaches characterize the optimum, then everyone will move smoothly in response to a perturbation of the tax schedule. So, consider a perturbation that increases marginal tax rates at the bottom of the distribution and average tax rates for everyone above the bottom. If we only care directly about the utility of those right near the bottom of the distribution, then the direct welfare effects can be made arbitrarily small (because changes in marginal tax rates have second order impacts on utility). The budgetary impacts of this reform consist solely of elasticity effects (which reduce revenue) along with the effects of raising average tax rates for individuals higher up in the income distribution (which increase revenue). The key insight is that the elasticity effects can also be made arbitrarily small because they only affect individuals within an ϵ neighborhood of the lowest type and this set has volume of order ϵ^K for $K = \dim(\mathbf{N})$. When $K \geq 2$, these effects can be made arbitrarily small relative to the effect of raising average tax rates for higher income individuals; hence, the total impact of this perturbation on the government's Lagrangian must be positive. Therefore it can not be optimal for all individuals to move smoothly in response to this perturbation (i.e., we have individuals either bunching or with multiple optima).[36]

Finally, to confirm that our method (i.e., solving System 24) is generating the correct optimal schedules, we can check that our results for utility function 18 match with the results computed using the method of Aguilera and Morin (2008), which utilizes the convexity characterization of incentive compatibility from Rochet (1987).[37] We show the difference in average tax rates between our method and the method of Aguilera and Morin (2008) in Figure 9 in Appendix E.3; they are very close and the small differences between the solutions shrink with the grid size, suggesting they are simply numerical noise. The biggest drawback of our numerical method is speed. Solving System 24 is computationally slower than using algorithms based on first order approaches or convexity constraints (e.g., Aguilera and Morin (2008) or Boerma, Tsyvinski and Zimin (2022)). Moreover, System 24 seems difficult to scale up to much larger grids.[38] Hence, these methods are almost certainly preferable from a computational perspective *when applicable*. The benefit of using the method proposed in this paper is solely to solve screening

[36]Loosely speaking, the conditions requiring $\frac{\partial u^{(0)}(T,\mathbf{z})}{\partial T}$ and $-\left(\frac{\partial u^{(0)}}{\partial z_i} + \frac{\partial u^{(i)}}{\partial z_i}\right) / \left(\frac{\partial^2 u^{(i)}}{\partial n_i \partial z_i}\right)$ to be bounded ensure that the individual elasticity effects remain bounded as social preferences become more and more redistributive.

[37]The method of Aguilera and Morin (2008) essentially boils down to non-linear semidefinite programming: solve System 19 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).

[38]However, I strongly conjecture that scalability could be improved by more efficient programmers.

problems for which existing algorithms cannot be applied, e.g., those that (1) feature bunching and (2) do not use a utility function given by Equation 1.

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 for bunching (e.g., Spiritus et al. (2022) and Krasikov and Golosov (2022) solve cases where there is no bunching so that the tax schedule can be found using first order methods) and (3) to allow for utility to take more general forms than Equation 1 (e.g., Boerma, Tsyvinski and Zimin (2022) allow for bunching, but utility is given by Equation 1). We use income data from the American Community Survey (ACS) in 2019 and suppose utility over consumption, c, labor supply for men, l_1, and labor supply for women, l_2, is given by:

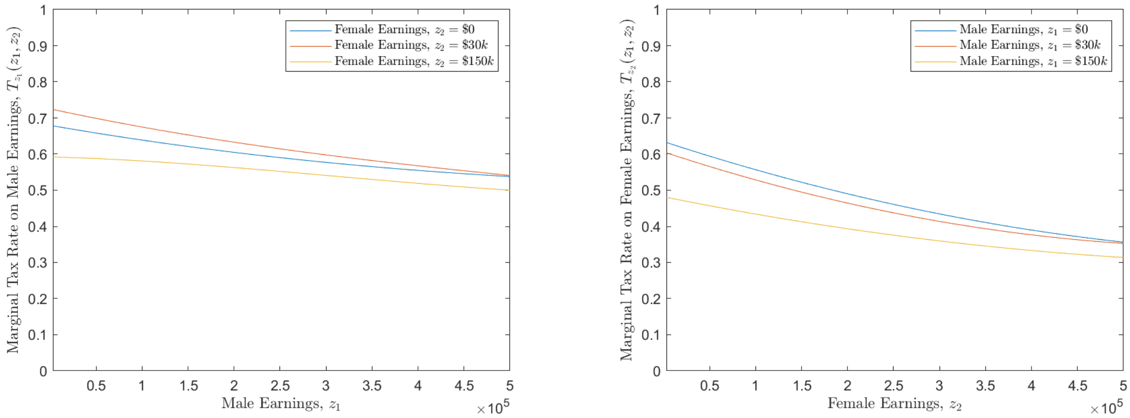
$$\begin{aligned} u(c, l_1, l_2; n_1, n_2) &= \log(c) - \frac{l_1^{1+\theta_1}}{1+\theta_1} - \frac{l_2^{1+\theta_2}}{1+\theta_2} - \frac{l_1}{\alpha_1} - \frac{l_2}{\alpha_2} \\ c &= n_1 l_1 + n_2 l_2 + T \end{aligned} \tag{25}$$

The government sets taxes as function of men's income $z_1 = n_1 l_1$ and women's income $z_2 = n_2 l_2$. Using this change of variables, we show in Appendix E how to express utility function 25 in the form $u(T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ discussed throughout this paper. We assume log utility over consumption, $\log(c)$, consistent with the findings of Chetty (2006). The parameters $n_1, n_2 > 0$ capture differences in labor productivity for men and women. Note that this utility function is neither linear in T nor linear in $\mathbf{n}$, which, to the best of my knowledge, takes us outside of the realm of any previous results on incentive compatibility.[39] Note that disutility of labor supply 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. 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.[40] $\theta_1, \theta_2, \alpha_1$, and α_2 are constant for all households and are chosen to match four moments: the compensated taxable income elasticities for men and women (0.2 and 1, respectively, taken from Blomquist and Selin (2010)) along with the percentages of men and women who do not work

[39] Utility function 25 satisfies Assumption 1; see Appendix E.

[40] Similarly, Saez (2001) calibrates a unidimensional $f(n)$ to match the empirical income distribution.

(13.5% and 20%, respectively, taken from ACS data).[41] 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 \$2 million per year than for households who earn \$0 per year. Figure 4 shows marginal tax rates over the income distribution for men and women conditional on a given level of spousal income.



(a) Marginal Tax Schedules for Men

(b) Marginal Tax Schedules for Women

Figure 4: Optimal Marginal Tax Schedules for Couples

Note: This figure shows the optimal marginal tax rates for men and women conditional on spousal earnings. We assume utility is given by Equation 25. $f(\mathbf{n})$ is calibrated to match the joint income distribution from the ACS 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 ACS data), and the fraction of women who do not work (20%, from ACS 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.

As far as marginal tax rates, top earning women are typically taxed at a lower marginal tax rate than top earning men because women 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 35% whereas the optimal marginal tax rate for men earning \$500,000 per year whose wife does not work is around 60%. Interestingly, we find that the tax schedule features portions of both negative jointness ($\frac{\partial^2 T(z_1,z_2)}{\partial z_1 \partial z_2} < 0$) and positive jointness ($\frac{\partial^2 T(z_1,z_2)}{\partial z_1 \partial z_2} > 0$).

We find that optimal average tax rates are around 40-50% for relatively high income couples (i.e., those with combined income over \$200,000 per year).[42] 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

[41] We provide more details about the calibration in Appendix E.

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

\$70,000 per year. Most importantly, the optimal solution features substantial bunching at the bottom of the distribution whereby many households do not work at all (this is consistent with Proposition 4). 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).[43]

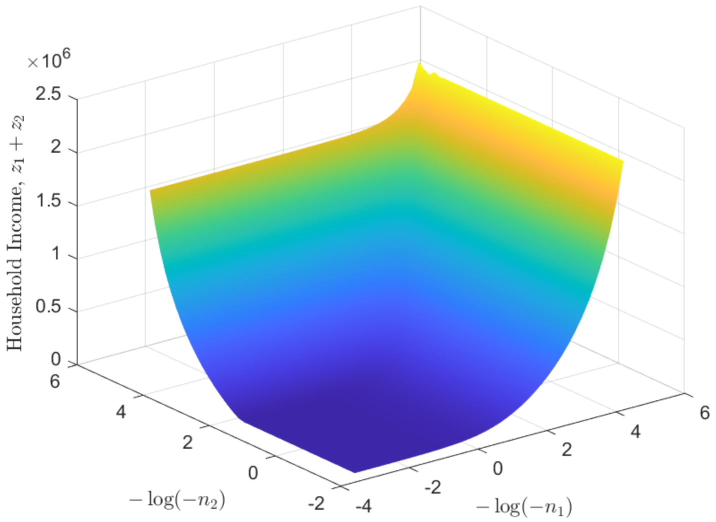


Figure 5: Optimal Total Household Income Across the Type Distribution

Note: This figure shows the optimal household income by type, assuming utility is given by Equation 25. 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 ACS 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 ACS data), and the fraction of women who do not work (20%, from ACS 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.

Because the optimal solution features bunching and utility is not given by Equation 1, both first order approaches and methods designed to exploit convexity constraints cannot be used to solve for the optimal schedule. Hence, our method provides a new pathway to solving these sorts of multidimensional optimal taxation problems. Ultimately, we hope that this basic simulation using ACS data not only highlights the methods developed in this paper via a somewhat realistic application, but also can be used as a starting point towards even more involved work on multidimensional taxation.

6 Conclusion

This paper has developed a set of results which allow us to solve certain multidimensional screening problems when first order approaches fail and when utility does not take the simple

[43]Figure 8 in Appendix E.3 shows the corresponding Jacobian determinant for this simulation.

linear/separable form of Equation 1. We have derived necessary conditions and sufficient conditions for incentive compatibility in multidimensional screening problems 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. We find that bunching in the optimal allocation appears to be relatively common and we prove that first order approaches will fail when solving optimal tax problems when society has sufficiently redistributive preferences. We apply this method to data from the ACS to better understand optimal couples taxation, again finding that it is optimal to have significant bunching in the form of many households that do not work.

Looking forward, we believe that the 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 the importance of bunching in 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.

Boadway, Robin, and Laurence Jacquet. 2008. "Optimal marginal and average income taxation under maximin." *Journal of Economic Theory*, 143(1): 425–441.

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

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."

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

Figalli, Alessio, Young-Heon Kim, and Robert J. McCann. 2011. "When is multidimensional screening a convex program?" *Journal of Economic Theory*, 146(2): 454–478.

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

Golosov, Mikhail, Aleh Tsyvinski, and Nicolas Werquin. 2014. "A Variational Approach to the Analysis of Tax Systems." *NBER Working Paper 20780*, 48.

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

Heinonen, Juha. 2005. *Lectures on Lipschitz analysis.* University of Jyväskylä.

Hencl, Stanislav, and Benjamin Vejnar. 2015. "Sobolev homeomorphism that cannot be approximated by diffeomorphisms in $W^{1,1}$." *Archive for Rational Mechanics and Analysis*, 219(1): 183–202.

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

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.

Koliha, J. J. 2004. "Approximation of convex functions." *Real Analysis Exchange*, 29(1): 465.

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. 2002. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica*, 70: 583–601.

Miller, John D. 1984. "Some global inverse function theorems." *Journal of Mathematical Analysis and Applications*, 100(2): 375–384.

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.

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 Appendix: Proofs

A.1 Proof of Theorem 1

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 8 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:

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

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 $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$.

As far as necessity, it should be clear that 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. The proof to Theorem 2 in Appendix A.3 shows that the envelope condition is necessary for incentive compatibility by Milgrom and Segal (2002) and that if $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable everywhere then injectivity of $\mathbf{n} \mapsto \mathbf{z}$ is necessary for incentive compatibility by Assumption 1.

A.2 Proof of Lemma 1

If we have a sequence of incentive compatible $(T_j(\mathbf{z}(\mathbf{n})), \mathbf{z}_j(\mathbf{n}))$, we know that $\forall \mathbf{n}, \mathbf{n}'$:

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

Where we have WLOG written $T(\cdot)$ as a function of $\mathbf{n}$ rather than $\mathbf{z}(\mathbf{n})$. Taking limits and passing through the continuous function $u(T, z; \mathbf{n})$, we see that:

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

Hence, the limiting allocation is incentive compatible. Finally, we know that an allocation

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

with multiple types $\mathbf{n}$ with the same $\mathbf{z}$ yet different T cannot be incentive compatible. Hence, it must be the case that we can express $T(\mathbf{n})$ as $T(\mathbf{z}(\mathbf{n}))$, which means $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n})) = \lim_{j\to\infty}(T_j(\mathbf{z}_j(\mathbf{n})), \mathbf{z}_j(\mathbf{n}))$ is incentive compatible.

A.3 Proof of Theorem 2

(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 (2002)) implies that the following envelope condition holds for all paths between $\mathbf{n}_1$ and $\mathbf{n}_2$:[45]

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

(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 2 tells us that the tax schedule $T(\mathbf{z})$ is twice continuously differentiable at $\mathbf{z}(\mathbf{n})$ so that Equation 13 holds. Under any incentive compatible allocation, we know that $u(T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ must be concave in $\mathbf{z}$ around $\mathbf{z}(\mathbf{n})$.[46] Hence, $\nabla_{\mathbf{z}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})$ must be negative semi-definite. Equation 13 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 11, 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:

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n}) = u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) \left[\nabla_{\mathbf{n}} \left(\frac{\nabla_{\mathbf{z}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})}{u_T(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})}\right)\right] \tag{26}$$

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

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

[45] To apply Corollary 1 from Milgrom and Segal (2002) 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.

[46] 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.

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

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

by Equation 26 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 $\mathbf{n}'$:

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

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

Proposition 5. *Let $\dim(\mathbf{N}) > \dim(\mathbf{Z})$. 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:*

1. *The envelope condition 8 holds $\forall\ \mathbf{n} = (\mathbf{n}^{(1)},\mathbf{n}^{(2)})$*
2. *$\mathbf{n}^{(1)} \mapsto \mathbf{z}$ is bijective $\forall \mathbf{n}^{(2)}$*
3. *$u\left(T\left(\mathbf{z}\left(\mathbf{n}^{(1)},\mathbf{n}^{(2)}\right)\right),\mathbf{z}\left(\mathbf{n}^{(1)},\mathbf{n}^{(2)}\right);\mathbf{n}^{(1)},\mathbf{n}^{(2)}\right) \geq u\left(T\left(\mathbf{z}\left(\mathbf{n}^{(1)\prime},\mathbf{n}^{(2)}\right)\right),\mathbf{z}\left(\mathbf{n}^{(1)\prime},\mathbf{n}^{(2)}\right);\mathbf{n}^{(1)},\mathbf{n}^{(2)}\right)$ $\forall \mathbf{n}^{(1)\prime} \in \partial\mathbf{N}^{(1)}$ 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 1 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 1 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 $\forall \mathbf{n}^{(2)}$.[47] Thus, we know that

[47] Note: in order to apply Theorem 1 when $\dim(\mathbf{N}) > \dim(\mathbf{Z})$, we require that $\mathbf{n}^{(1)} \mapsto \mathbf{z}$ is *surjective* onto the set of chosen incomes conditional on any $\mathbf{n}_2$. 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.

each type $\left(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}\right)$ prefers his/her assigned bundle to the bundles assigned to all other types types $\left(\mathbf{n}^{(1)'}, \mathbf{n}^{(2)'}\right)$.

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. If Assumption 1 holds on this restricted subset, we can immediately apply Theorem 1. $\square$

Proposition 6. *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 envelope condition 8 holds $\forall\ \mathbf{n} = (\mathbf{n}^{(1)}, \mathbf{n}^{(2)})$*
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 $(\mathbf{n}^{(1)'}, \mathbf{n}^{(2)})$ such that the allocation $\mathbf{n}^{(1)} \mapsto \mathbf{z}$ is a local diffeomorphism and $T(\mathbf{z})$ is differentiable we must have $\mathbf{z}(\mathbf{n}^{(1)}, \mathbf{n}^{(2)}) \neq \mathbf{z}(\mathbf{n}^{(1)'}, \mathbf{n}^{(2)})$.*

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 hold, then Theorem 2 provides necessary conditions for incentive compatibility.

Proof. When $\dim(\mathbf{N}) > \dim(\mathbf{Z})$, the above follows immediately from Theorem 2: 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 $\left(\mathbf{n}^{(1)'}, \mathbf{n}^{(2)}\right)$. When $\dim(\mathbf{N}) < \dim(\mathbf{Z})$, we can immediately apply Theorem 2 (restricting to the subset $\tilde{\mathbf{Z}}$) as long as Assumption 1 holds on $\tilde{\mathbf{Z}}$. $\square$

A.5 Lemma 3: Differentiability of $T(\mathbf{z})$

Lemma 3. *Consider any allocation $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ which yields differentiable $U(\mathbf{n})$ satisfying the envelope condition 8. If $\mathbf{n} \mapsto \mathbf{z}$ is a local diffeomorphism, then $T(\mathbf{z})$ is twice continuously differentiable.*

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:

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

By the envelope condition 8, we then know that:

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

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.6 Proof of Proposition 1

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:

$$\nabla_{\mathbf{n}}FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1} = \begin{bmatrix} \frac{\partial^2 u^{(1)}(z_1,n_1)}{\partial z_1 \partial n_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial^2 u^{(K)}(z_K,n_K)}{\partial z_K \partial n_K} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) & \cdots \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \tag{27}$$

By Assumption 1, we know that $\frac{\partial u^{(i)}(z_i,n_i)}{\partial z_i}$ is monotonic in n_i $\forall i$. Therefore, WLOG let us assume that $\frac{\partial^2 u^{(i)}(z_i,n_i)}{\partial z_i \partial n_i} > 0$ $\forall i$. Then Equation 27 has positive principal minors if and only if:

$$\begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) & \cdots \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{28}$$

is a P matrix (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, we know that Equation 28 is a P matrix (i.e., has all positive principal minors). On a rectangular domain, Remark 2 then implies that $\mathbf{n} \mapsto \mathbf{z}$ is a diffeomorphism.

A.7 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 show 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:

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

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

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1} = \begin{bmatrix} \frac{\partial^2 u^{(1)}(z_1,n_1)}{\partial z_1 \partial n_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial^2 u^{(K)}(z_K,n_K)}{\partial z_K \partial n_K} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1}(\mathbf{n}) & \frac{\partial z_1}{\partial n_2}(\mathbf{n}) & \cdots \\ \frac{\partial z_2}{\partial n_1}(\mathbf{n}) & \frac{\partial z_2}{\partial n_2}(\mathbf{n}) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}^{-1} \tag{30}$$

is symmetric. But Equation 30 is symmetric if and only if Equation 29 is symmetric by the following Lemma:

Lemma 4. *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 $VA = [VA]^{\mathcal{T}} = A^{\mathcal{T}} V^{\mathcal{T}} = A^{\mathcal{T}} V$. Thus:[48]

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

□

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

is clearly symmetric; hence, Lemma 4 implies that if Equation 30 is symmetric, then so is Equation 29.

A.8 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 2.

A.9 Proof of Proposition 4

First, note that because welfare weights are only unique up to a normalization, we can normalize $\int_{\mathbf{N}} \psi(\mathbf{n}) dF(\mathbf{n}) = 1$. Suppose then, to the contrary, that for all welfare weights with $\int_{B_\epsilon(\mathbf{n})} \psi(\mathbf{n}) dF(\mathbf{n}) \approx 1$, the optimal allocation can be found via first order approaches.

[48] Note, the first if and only if statement follows by taking inverses on both sides of $VA = A^{\mathcal{T}} V$ and then left and right multiplying both sides by V.

When utility is given by Equation 15, the following set of equations define a function $(T, \mathbf{z}) \mapsto (U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))$ for each type $\mathbf{n}$:

$$U_{n_i}(\mathbf{n}) = \frac{\partial U(\mathbf{n})}{\partial n_i} = \frac{\partial u^{(i)}(z_i; n_i)}{\partial n_i} \tag{31}$$

$$U(\mathbf{n}) = u^{(0)}(T, \mathbf{z}) + \sum_{i}^{K} u^{(i)}(z_i; n_i) \tag{32}$$

Assumption 1 ensures $\frac{\partial^2 u^{(i)}(z_i;n_i)}{\partial n_i \partial z_i} \neq 0 \ \forall i$. From Equations 31 and 32, it should be clear by the implicit function theorem that if $\frac{\partial^2 u^{(i)}(z_i;n_i)}{\partial n_i \partial z_i} \neq 0$, then the mapping $z_i \mapsto U_{n_i}(\mathbf{n})$ is a diffeomorphism. Moreover, because $\frac{\partial u^{(0)}}{\partial T}$ is assumed to be greater than zero, the mapping $T \mapsto U(\mathbf{n})$ is also a diffeomorphism, for any fixed value of $\mathbf{z}$. Thus, $(T, \mathbf{z}) \mapsto (U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))$ is a diffeomorphism.

Let us then consider the government's Lagrangian as a function of $(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))$, with some Lagrange multiplier λ (note, this is all entirely analogous to Rochet and Chone (1998)):

$$L(\mathbf{n}, U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})) = \int_{\mathbf{N}} \left[\psi(\mathbf{n}) U(\mathbf{n}) - \lambda T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))\right] dF(\mathbf{n})$$

Now, if a given utility profile $U(\mathbf{n})$ is optimal and can be found via the first order approach, then any small perturbation that changes the utility profile from $U(\mathbf{n})$ to $U(\mathbf{n}) + \eta \tilde{U}(\mathbf{n})$ must have a net zero effect on the government's Lagrangian so that $\forall \tilde{U}(\mathbf{n})$:

$$\left. \frac{\partial L(\mathbf{n}, U(\mathbf{n}) + \eta \tilde{U}(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}) + \eta \nabla_{\mathbf{n}} \tilde{U}(\mathbf{n}))}{\partial \eta} \right|_{\eta=0} = 0$$

Now, let us suppose that almost all of the mass of the $\psi(\mathbf{n})$ function lies within the ϵ^2 ball so that $\int_{B_{\epsilon^2}(\underline{\mathbf{n}})} \psi(\mathbf{n}) dF(\mathbf{n}) = 1 - \delta$ for some small ϵ and δ. Now consider a perturbation which sets $\nabla_{\mathbf{n}} \tilde{U}(\mathbf{n}) = -[\epsilon, \epsilon, ..., \epsilon]$ on the ϵ ball $B_\epsilon(\underline{\mathbf{n}})$ and 0 elsewhere, defining $\tilde{U}(\mathbf{n})$ everywhere via a line integral as: $\tilde{U}(\mathbf{n}) = \int_0^1 \nabla_{\mathbf{n}} \tilde{U}(\mathbf{n}(s)) \cdot (\mathbf{n} - \underline{\mathbf{n}}) ds$ where $\mathbf{n}(s)$ is a parametrization of the line between $\underline{\mathbf{n}}$ and $\mathbf{n}$. Let us now compute the derivative of the government's Lagrangian in the direction of this $\tilde{U}(\mathbf{n})$:

$$\begin{aligned} &\left. \frac{\partial L(\mathbf{n}, U(\mathbf{n}) + \eta \tilde{U}(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}) + \eta \nabla_{\mathbf{n}} \tilde{U}(\mathbf{n}))}{\partial \eta} \right|_{\eta=0} = -\lambda \int_{B_\epsilon(\underline{\mathbf{n}})} \sum_{i=1}^{K} \frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U_{n_i}(\mathbf{n})} (-\epsilon) dF(\mathbf{n}) \\ &+ \int_{B_{\epsilon^2}(\underline{\mathbf{n}})} \left[\psi(\mathbf{n}) - \lambda \frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})}\right] \tilde{U}(\mathbf{n}) dF(\mathbf{n}) + \int_{\mathbf{N} \setminus B_{\epsilon^2}(\underline{\mathbf{n}})} \left[\psi(\mathbf{n}) - \lambda \frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})}\right] \tilde{U}(\mathbf{n}) dF(\mathbf{n}) \end{aligned} \tag{33}$$

Next, let us analyze the order of the effects on the right hand side of Equation 33. The first term is evaluated on a ball of radius ϵ, which necessarily has volume proportional to ϵ^K for $K = \dim(\mathbf{N})$. Note, that the ϵ ball has volume proportional to ϵ in dimension 1, which is why the proposition will not hold in the unidimensional case. As long as the terms $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U_{n_i}(\mathbf{n})}$ are all bounded from below by some number M_1 (we will return to this point at the end of the

proof), then:
$$\lambda \int_{B_\epsilon(\underline{\mathbf{n}})} \sum_{i=1}^{K} \frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U_{n_i}(\mathbf{n})} \epsilon dF(\mathbf{n}) > M_1 c_1 \epsilon^{K+1}$$
where c_1 is a positive constant that depends on λ, K, and $F(\mathbf{n})$ (we will show at the end of the proof that $c_1 > 0$ because $\lambda > 0$ and $c_1 \not\to 0$ as $\epsilon \to 0$ given that $\lambda \not\to 0$ as $\epsilon \to 0$).

The second term on the right hand side of Equation 33 is evaluated on a ball of radius ϵ^2. Note that $\int_{B_{\epsilon^2}(\underline{\mathbf{n}})} \psi(\mathbf{n}) dF(\mathbf{n}) = 1 - \delta$ and that $\forall \mathbf{n} \in B_{\epsilon^2}(\underline{\mathbf{n}})$, there exists some unit vector $\mathbf{v}$ such that $\mathbf{n} = \underline{\mathbf{n}} + \rho \mathbf{v}$ for some $\rho < \epsilon^2$. Thus, $\tilde{U}(\mathbf{n}) = \int_0^1 -[\epsilon, \epsilon, ..., \epsilon] \cdot \rho \mathbf{v} ds = -\epsilon\rho > -\epsilon^3$. Given that $\tilde{U}(\mathbf{n}) < 0$, as long as $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})}$ is positive (we will return to this point later) then we can conclude:
$$\int_{B_{\epsilon^2}(\underline{\mathbf{n}})} \left[\psi(\mathbf{n}) - \lambda \frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})}\right] \tilde{U}(\mathbf{n}) dF(\mathbf{n}) > \int_{B_{\epsilon^2}(\underline{\mathbf{n}})} \psi(\mathbf{n}) \tilde{U}(\mathbf{n}) dF(\mathbf{n}) > -\epsilon^3(1-\delta)$$
For the third and final term on the right hand side of Equation 33, note that $\int_{\mathbf{N} \setminus B_{\epsilon^2}(\underline{\mathbf{n}})} \psi(\mathbf{n}) dF(\mathbf{n}) = \delta$. Next, we use the fact that $\nabla_{\mathbf{n}} \tilde{U}(\mathbf{n})$ is zero outside of $B_{\epsilon^2}(\underline{\mathbf{n}})$ so that for all $\mathbf{n} \in \mathbf{N} \setminus B_{\epsilon^2}(\underline{\mathbf{n}})$ there exists a unit vector $\mathbf{v}$ such that $\tilde{U}(\mathbf{n}) = \int_0^1 -[\epsilon, \epsilon, ..., \epsilon] \cdot (\rho\epsilon\mathbf{v}) ds = -\rho\epsilon^2$ where $\epsilon \le \rho \le 1$ for $\mathbf{n} \in B_\epsilon(\underline{\mathbf{n}}) \setminus B_{\epsilon^2}(\underline{\mathbf{n}})$ and $\rho = 1$ for $\mathbf{n} \in \mathbf{N} \setminus B_\epsilon(\underline{\mathbf{n}})$. Thus:
$$\int_{\mathbf{N} \setminus B_{\epsilon^2}(\underline{\mathbf{n}})} \psi(\mathbf{n}) \tilde{U}(\mathbf{n}) dF(\mathbf{n}) > -\delta\epsilon^2$$
Hence, we can control this term to be as small as desired, relative to ϵ^2, by shrinking δ. Finally, as long as $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})}$ is positive and bounded away from zero by some number M_2 (we will return to this point at the end of the proof):[49]
$$-\int_{\mathbf{N} \setminus B_{\epsilon^2}(\underline{\mathbf{n}})} \lambda \frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})} \tilde{U}(\mathbf{n}) dF(\mathbf{n}) > M_2 c_2 \epsilon^2 \tag{34}$$
for a positive constant c_2 that depends on λ and $F(\mathbf{n})$.

For sufficiently small ϵ and δ, the term given by Equation 34 is arbitrarily large relative to all other terms in Equation 33. Hence, for sufficiently small ϵ and δ, Equation 34 is approximately (up to terms of order higher than ϵ^2):
$$\left.\frac{\partial L(\mathbf{n}, U(\mathbf{n}) + \eta\tilde{U}(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}) + \eta\nabla_{\mathbf{n}} \tilde{U}(\mathbf{n}))}{\partial \eta}\right|_{\eta=0} \approx -\int_{\mathbf{N} \setminus B_{\epsilon^2}(\underline{\mathbf{n}})} \lambda \frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})} \tilde{U}(\mathbf{n}) dF(\mathbf{n}) > 0$$
But this violates the first order condition for the optimal transfer schedule. This means that the optimal schedule cannot be found via the first order approach.

Finally, we return to the three technical details that we previously swept under the rug: (1) $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U_{n_i}(\mathbf{n})}$ is bounded from below, (2) $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})}$ is positive and bounded away from zero, and (3) λ is positive and bounded away from zero. More specifically, we need the above three conditions to hold for all arbitrarily small ϵ because the limiting arguments above require

[49] Technically, M_2 could also depend on ϵ because we have only ensured that $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})}$ is bounded from below. But, this is not a problem because if $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})}$ is bounded away from zero then $M_2 \not\to 0$ as $\epsilon \to 0$.

that the constants $M_1, c_1, M_2, c_2 \not\to 0$ as $\epsilon \to 0$. Let us consider the function $T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))$. Given that $z_i \mapsto U_{n_i}(\mathbf{n})$ is a diffeomorphism, we can express $T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))$ implicitly via:

$$U(\mathbf{n}) = u^{(0)}(T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})), \mathbf{z}(\nabla_{\mathbf{n}} U(\mathbf{n}))) + \sum_i^K u^{(i)}(z_i(U_{n_i}(\mathbf{n})); n_i) \tag{35}$$

Given that $\frac{\partial u^{(0)}}{\partial T}$ is assumed to be bounded away from zero, we can implicitly differentiate with respect to $U_{n_i}(\mathbf{n})$ to yield (omitting function arguments for clarity/brevity and recognizing that $\mathbf{z}(\nabla_{\mathbf{n}} U(\mathbf{n})) = [z_1(U_{n_1}(\mathbf{n})), z_2(U_{n_2}(\mathbf{n})), ..., z_K(U_{n_K}(\mathbf{n}))]$):

$$\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U_{n_i}(\mathbf{n})} = -\frac{\frac{\partial u^{(0)}}{\partial z_i} + \frac{\partial u^{(i)}}{\partial z_i}}{\frac{\partial u^{(0)}}{\partial T}} \frac{\partial z_i}{\partial U_{n_i}(\mathbf{n})} \tag{36}$$

Next, note that (by Equation 31):

$$\frac{\partial z_i}{\partial U_{n_i}(\mathbf{n})} = \left(\frac{\partial U_{n_i}(\mathbf{n})}{\partial z_i}\right)^{-1} = \left(\frac{\partial^2 u^{(i)}}{\partial n_i \partial z_i}\right)^{-1}$$

Hence, we can rewrite Equation 36 as:

$$\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U_{n_i}(\mathbf{n})} = -\frac{\frac{\partial u^{(0)}}{\partial z_i} + \frac{\partial u^{(i)}}{\partial z_i}}{\frac{\partial u^{(0)}}{\partial T}} \left(\frac{\partial^2 u^{(i)}}{\partial n_i \partial z_i}\right)^{-1}$$

Hence, if we have that $\frac{\partial u^{(0)}}{\partial T}$ is bounded away from zero and $-\left(\frac{\partial u^{(0)}}{\partial z_i} + \frac{\partial u^{(i)}}{\partial z_i}\right) / \left(\frac{\partial^2 u^{(i)}}{\partial n_i \partial z_i}\right)$ is bounded from below for all $T, \mathbf{z}, \mathbf{n}$, we know that $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U_{n_i}(\mathbf{n})}$ is also bounded from below.

For instance, for preferences given by Equation 18 and Equation 20, we have, respectively:

$$\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U_{n_i}(\mathbf{n})} = \frac{1}{z_i(U_{n_i})^{\theta_i}} + n_i$$

and

$$\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U_{n_i}(\mathbf{n})} = \frac{1 + n_i z_i(U_{n_i})^{\theta_i} - \frac{n_i^2}{2\alpha}}{z_i(U_{n_i})^{\theta_i} - \frac{n_i}{\alpha}}$$

One can easily verify that for $z_i \geq 0$ and $\mathbf{n} \in \mathbf{N}$ (with compact $\mathbf{N}$ such that all elements of $\mathbf{N}$ are strictly less than zero) we have that $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U_{n_i}(\mathbf{n})} = -\left(\frac{\partial u^{(0)}}{\partial z_i} + \frac{\partial u^{(i)}}{\partial z_i}\right) / \left(\frac{\partial^2 u^{(i)}}{\partial n_i \partial z_i}\right)$ is bounded from below (note, we have $\frac{\partial u^{(0)}}{\partial T} = 1$ because utility is linear in T).

Next, let us turn to showing that $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})}$ is positive and bounded away from zero. Implicitly differentiating Equation 35, we get that:

$$\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})} = \left(\frac{\partial u^{(0)}}{\partial T}\right)^{-1}$$

which is positive and bounded away from zero as long as $\frac{\partial u^{(0)}}{\partial T}$ is bounded away from ∞ for all $\mathbf{z}$ and T, which we assume. Note, with linear utility in T, $\frac{\partial u^{(0)}}{\partial T} = 1$ so preferences given by Equation 18 and Equation 20 both satisfy this property.

As far as λ, if the first order approach holds we can derive an expression for λ by perturbing utility in the direction of $\tilde{U}(\mathbf{n}) = 1$. The impact on the Lagrangian is:

$$\int_{\mathbf{N}} \left[\psi(\mathbf{n}) - \lambda \frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})} \right] dF(\mathbf{n}) = 0$$

Hence, given $\int_{\mathbf{N}} \psi(\mathbf{n}) dF(\mathbf{n}) = 1$ and $\frac{\partial T(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))}{\partial U(\mathbf{n})}$ is positive and bounded from above because $\frac{\partial u^{(0)}}{\partial T}$ is bounded away from 0 for all $\mathbf{n}$, we know that λ is positive and bounded away from 0.

Finally, if utility is given by Equation 1, then if the optimal utility profile $U(\mathbf{n})$ is strictly convex, first order methods can identify the optimal $U(\mathbf{n})$ (Rochet and Chone, 1998). Hence, the preceding argument implies that $U(\mathbf{n})$ cannot be strictly convex when $\int_{B_\epsilon(\underline{\mathbf{n}})} \psi(\mathbf{n}) dF(\mathbf{n}) \approx 1$; it must have portions which are only weakly convex so that for some $\mathbf{n}_1$, $\mathbf{n}_2$, $\alpha \in (0, 1)$ we have:

$$U(\alpha \mathbf{n}_1 + (1 - \alpha)\mathbf{n}_2) = \alpha U(\mathbf{n}_1) + (1 - \alpha)U(\mathbf{n}_2)$$

But weakly convex functions must be linear between $\mathbf{n}_1$ and $\mathbf{n}_2$.[50] When utility is given by Equation 1, $U(\mathbf{n})$ is linear if and only if $\mathbf{z}$ is constant between $\mathbf{n}_1$ and $\mathbf{n}_2$ so that bunching occurs.[51] This completes the proof of the proposition.

[50] This is a standard fact, but we prove it nonetheless. Set $\mathbf{n}' = \alpha \mathbf{n}_1 + (1-\alpha)\mathbf{n}_2$. Let $l(\beta)$ be the linear function defined on $[0, 1]$ with $l(\beta) = \beta U(\mathbf{n}_1) + (1 - \beta)U(\mathbf{n}_2)$. Now $U(\beta \mathbf{n}_1 + (1 - \beta)\mathbf{n}_2) \leq l(\beta)$ on $[0, 1]$ because $U(\cdot)$ is convex. Because $U(\cdot)$ is convex on the line segment between $\mathbf{n}'$ and $\mathbf{n}_2$, $U(\gamma \mathbf{n}' + (1-\gamma)\mathbf{n}_2) \leq \gamma U(\mathbf{n}') + (1-\gamma)U(\mathbf{n}_2)$ for $\gamma \in [0, 1]$. Plugging in the value of $\mathbf{n}'$ we see that $U(\beta \mathbf{n}_1 + (1 - \beta)\mathbf{n}_2) \geq l(\beta)$ on $[0, \alpha]$. Analogous reasoning shows $U(\beta \mathbf{n}_1 + (1 - \beta)\mathbf{n}_2) \geq l(\beta)$ on $[\alpha, 1]$. Hence, $U(\cdot)$ is linear between $\mathbf{n}_1$ and $\mathbf{n}_2$.

[51] Recall that the gradient $\nabla_{\mathbf{n}} U(\mathbf{n}) = [v_{z_1}(\mathbf{z}), v_{z_2}(\mathbf{z}), ..., v_{z_K}(\mathbf{z})]$. Given that we assume $v_{z_i}(\mathbf{z}) \neq 0$, if $\nabla_{\mathbf{n}} U(\mathbf{n})$ is constant along a given segment, $\mathbf{z}$ must also be constant.

B Online Appendix: Additional Proofs

B.1 Utility Function 7 Satisfies Assumption 1

For utility function 7, we have:

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

The Jacobian of this mapping is found by taking the gradient with respect to $\mathbf{n}$:

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

If $\mathbf{N}$ in $\mathbb{R}^2_{++}$ and we can restrict attention to $z_1, z_2 \geq 0$ and either $z_1 > 0$ or $z_2 > 0$ we have:

$$\frac{1}{v'(z_1+z_2+T)}\left(\frac{(1+\theta_1)z_1^{\theta_1}}{n_1^{2+\theta_1}} + \beta\frac{z_2}{n_1^2 n_2}\right) > 0, \quad \frac{1}{v'(z_1+z_2+T)}\left(\frac{(1+\theta_2)z_2^{\theta_2}}{n_2^{2+\theta_2}} + \beta\frac{z_1}{n_1 n_2^2}\right) > 0$$

$$\det\left[\nabla_{\mathbf{n}}\left(\frac{1}{v'(z_1+z_2+T)}\begin{bmatrix} -\frac{z_1^{\theta_1}}{n_1^{1+\theta_1}} - \beta\frac{z_2}{n_1 n_2} \\ -\frac{z_2^{\theta_2}}{n_2^{1+\theta_2}} - \beta\frac{z_1}{n_1 n_2} \end{bmatrix}\right)\right] = \frac{\frac{(1+\theta_1)z_1^{\theta_1}}{n_1^{2+\theta_1}}\frac{(1+\theta_2)z_2^{\theta_2}}{n_2^{2+\theta_2}} + \frac{(1+\theta_1)z_1^{\theta_1}}{n_1^{2+\theta_1}}\beta\frac{z_1}{n_1 n_2^2} + \frac{(1+\theta_2)z_2^{\theta_2}}{n_2^{2+\theta_2}}\beta\frac{z_2}{n_1^2 n_2}}{v'(z_1+z_2+T)^2} > 0$$

Hence, for utility function 7, the mapping $\mathbf{n} \mapsto \frac{u_{\mathbf{z}}(T,\mathbf{z};\mathbf{n})}{u_T(T,\mathbf{z};\mathbf{n})}$ has a continuous P matrix Jacobian. Remark 2 then ensures utility function 7 satisfies Assumption 1 on any rectangular domain $\mathbf{N}$ in $\mathbb{R}^2_{++}$.

B.2 Existence of a Solution to Problem 17

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

Proposition 7. *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* $(U(\mathbf{n}), \nabla_{\mathbf{n}}U(\mathbf{n})) \mapsto (T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$. *The optimal taxation problem has a solution 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 for some $\phi(\mathbf{n}) \in L^1(\mathbf{N})$

$$\int_{\mathbf{N}} \phi(\mathbf{n})U(\mathbf{n})dF(\mathbf{n}) \to -\infty \text{ as } U(\mathbf{n}) \to -\infty \text{ on a positive measure set}$$

Proof. We are going to argue that any maximizing sequence of functions $U_j^*(\mathbf{n})$ which satisfy the budget constraint must be bounded. To begin, suppose $U_j^*(\mathbf{n}) \in W^{1,\infty}$, which is the Sobolev space of bounded functions with bounded weak derivatives (hence the equations $\nabla_{\mathbf{n}} U(\mathbf{n}) = \nabla_{\mathbf{n}} u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ and $U(\mathbf{n}) = u(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$ do in fact define an a.e. correspondence $(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})) \mapsto (T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n})))$. Given that $\mathbf{N}$ is bounded and convex, the functions in $W^{1,\infty}$ coincide with the Lipschitz continuous functions, see Theorem 4.1 of Heinonen (2005). First, let us recall the definition of the $W^{1,\infty}$ norm:

$$||U||_{W^{1,\infty}} = ||U(\mathbf{n})||_\infty + \text{ess sup}_{\mathbf{n}\in\mathbf{N}} ||\nabla_{\mathbf{n}} U(\mathbf{n})||$$

where $||U(\mathbf{n})||_\infty$ represents the essential supremum norm and $||\nabla_{\mathbf{n}} U(\mathbf{n})||$ represents the operator norm.

Next, note that as $||U||_{W^{1,\infty}} \to \infty$ 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:

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

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 $||U||_{W^{1,\infty}} \to \infty$, we know we must have $U(\mathbf{n}) \to -\infty$ on a positive measure set. But then $\int_{\mathbf{N}} \phi(\mathbf{n}) U(\mathbf{n}) dF(\mathbf{n}) \to -\infty$, which clearly cannot be optimal as the government can always choose to set taxes equal to zero with $T(\mathbf{z}) = 0$, which we assume will yield welfare greater than $-\infty$.[52]

Hence, any maximizing sequence $U_j^*(\mathbf{n})$ of $\int_{\mathbf{N}} \phi(\mathbf{n}) U(\mathbf{n}) dF(\mathbf{n})$ that satisfies the budget constraint must be bounded so that $\exists M_1, M_2$ such that $||U_j^*(\mathbf{n})||_\infty \leq M_1$ and $\text{ess sup}_{\mathbf{n}\in\mathbf{N}} ||\nabla_{\mathbf{n}} U_j^*(\mathbf{n})|| \leq M_2$ for sufficiently large j. Next, we use the fact that the set of incentive compatible allocations satisfying the budget constraint is a complete subspace of $W^{1,\infty}(\mathbf{N})$, which is weak* sequentially compact by Banach-Alaoglu. This is because $W^{1,\infty}(\mathbf{N})$ is a complete subspace of $L^\infty(\mathbf{N})^{K+1}$, which is the dual of the separable space $L^1(\mathbf{N})^{K+1}$. Hence, there is a subsequence of the bounded sequence $U_j^*(\mathbf{n})$ that converges (in the weak* topology) to some $U^*(\mathbf{n}) \in W^{1,\infty}(\mathbf{N})$ so

[52] $\int_{\mathbf{N}} \phi(\mathbf{n}) U(\mathbf{n}) dF(\mathbf{n}) \to -\infty$ as $||U(\mathbf{n})||_\infty \to -\infty$ if welfare weights are decreasing with type or are smooth and bounded away from zero.

that for all $g(\mathbf{n}) \in L^1(\mathbf{N})$:

$$\int_{\mathbf{N}} g(\mathbf{n}) U_j^*(\mathbf{n}) dF(\mathbf{n}) \to \int_{\mathbf{N}} g(\mathbf{n}) U^*(\mathbf{n}) dF(\mathbf{n})$$

Finally, by the fact that we assume weights are smooth so that $\phi(\mathbf{n}) \in L^1(\mathbf{N})$, we know that $U^*(\mathbf{n})$ satisfies:

$$\int_{\mathbf{N}} \phi(\mathbf{n}) U^*(\mathbf{n}) dF(\mathbf{n}) = \lim_{j\to\infty} \int_{\mathbf{N}} \phi(\mathbf{n}) U_j^*(\mathbf{n}) dF(\mathbf{n}) = \sup_{U(\mathbf{n})\in W^{1,\infty}(\mathbf{N})} \int_{\mathbf{N}} \phi(\mathbf{n}) U(\mathbf{n}) dF(\mathbf{n})$$

□

Proposition 7, 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. Hence, we use coercivity of the budget constraint to show that any maximizing sequence will be bounded; the rest of the proof is similar to standard existence proofs, e.g., Kinderlehrer and Stampacchia (1980). The minor additional technical difference is that we use weak* sequential compactness rather than weak convergence because we show that a solution exists in the space of bounded Lipschitz function $W^{1,\infty}$ (which is not reflexive, but which has a separable pre-dual) rather than the space of square integrable functions with square integrable weak derivatives $W^{1,2}$ (which is reflexive).

Remark 3. *As an example application of Proposition 7, 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:*[53]

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

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* $U \to -\infty$*. Moreover, for smooth weights bounded away from zero on the compact set* $\mathbf{N}$*,* $\int_{\mathbf{N}} \phi(\mathbf{n}) U(\mathbf{n}) dF(\mathbf{n}) \to -\infty$ *as* $U(\mathbf{n}) \to -\infty$ *on any positive measure set so that the conditions for Proposition 7 are satisfied.*

B.3 No Jumping for Utility Function 18

Proposition 8. *Suppose that utility is given by:*

$$u(T, \mathbf{z}; \mathbf{n}) = \sum_{i=1}^K z_i(\mathbf{n}) + T + n_1 \frac{z_1^{1+\theta_1}}{1+\theta_1} + n_2 \frac{z_2^{1+\theta_2}}{1+\theta_2} + ... + n_K \frac{z_K^{1+\theta_K}}{1+\theta_K} \tag{37}$$

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)):

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

Lemma 5. *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:*

$$\nabla_{\mathbf{n}} U^*(\mathbf{n}) = v(\mathbf{z}^*(\mathbf{n})) \tag{38}$$

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 System 17 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 5 to convert the global incentive compatibility constraints into a convexity constraint, we can write the government's maximization problem as:

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

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-Erdmann corner condition along the discontinuity surface:[54]

$$\left[L_3(\mathbf{n}, U^*, \nabla^-_{\mathbf{n}} U^*) - L_3(\mathbf{n}, U^*, \nabla^+_{\mathbf{n}} U^*) \right] \cdot \left[\nabla^+_{\mathbf{n}} U^* - \nabla^-_{\mathbf{n}} U^* \right] = 0 \tag{39}$$

However, note that the second derivative matrix of L w.r.t. $\nabla_{\mathbf{n}} U$ is a negative diagonal matrix

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

(recall $\theta_i > 0$ and $\frac{\partial U}{\partial n_i} = \frac{z_i^{1+\theta_i}}{1+\theta_i} > 0$):

$$L_{33}(\mathbf{n}, U, \nabla_{\mathbf{n}} U) = \lambda f(\mathbf{n}) \begin{bmatrix} -\theta_1 \left((1+\theta_1) \frac{\partial U}{\partial n_1} \right)^{\frac{1}{1+\theta_1} - 2} & & \\ & -\theta_2 \left((1+\theta_2) \frac{\partial U}{\partial n_2} \right)^{\frac{1}{1+\theta_2} - 2} & \\ & & -\theta_K \left((1+\theta_K) \frac{\partial U}{\partial n_K} \right)^{\frac{1}{1+\theta_K} - 2} \end{bmatrix}$$

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

Strict concavity of L in $\nabla_{\mathbf{n}} U$ implies that:

$$\left[L_3(\mathbf{n}, U^*, \nabla_{\mathbf{n}}^- U^*) - L_3(\mathbf{n}, U^*, \nabla_{\mathbf{n}}^+ U^*) \right] \cdot \left[\nabla_{\mathbf{n}}^+ U^* - \nabla_{\mathbf{n}}^- U^* \right] > 0 \tag{40}$$

Hence, Equation 40 implies that the Weierstrass-Erdmann corner condition 39 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 because a function is convex if and only if it remains convex when restricted to a line segment).[55] 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 $\Delta U(n_1, n_2)$).

[55]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.

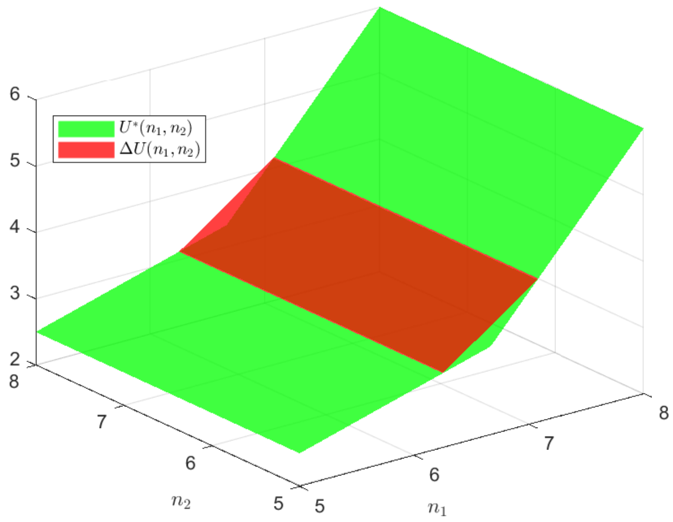


Figure 6: Illustration of Perturbation to $U^*(\mathbf{n})$

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:

$$\Delta \int_{\mathbf{N}} L(\mathbf{n}, U, \nabla_{\mathbf{n}} U) =$$
$$\int_{\Sigma} \int_{\sigma-\epsilon \mathbf{p}(\sigma)}^{\sigma+\epsilon'(\sigma)\mathbf{p}(\sigma)} \Bigg\{ L\left(\mathbf{n}, \int_{\sigma-\epsilon\mathbf{p}(\sigma)}^{\mathbf{n}} \left[\frac{1}{2}\nabla^+_{\mathbf{n}} U^*(\sigma) + \frac{1}{2}\nabla^-_{\mathbf{n}} U^*(\sigma) - \nabla_{\mathbf{n}} U^*(\mathbf{s})\right] \cdot \mathbf{p}(\sigma) d\mathbf{s}, \frac{1}{2}\nabla^+_{\mathbf{n}} U^*(\sigma) + \frac{1}{2}\nabla^-_{\mathbf{n}} U^*(\sigma)\right)$$
$$- L(\mathbf{n}, U^*(\mathbf{n}), \nabla_{\mathbf{n}} U^*(\mathbf{n}))\Bigg\} d\mathbf{n} d\sigma \tag{41}$$

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 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:[56]

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

The final inequality follows by Equation 40. 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 $z^*(\mathbf{n})$. □

C Online Appendix: Relationship to Previous Incentive Compatibility Results

We discuss in detail how our results compare with previous incentive compatibility results.

C.1 Relation to Existing Results on Unidimensional Incentive Compatibility

First, we discuss the relationship to the well known incentive compatibility result in one dimension, Theorem 0. 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

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

this case, Assumption 1 requires that for all c, z:[57]

$$\frac{\partial\left(\frac{u_2\left(c,\frac{z}{n}\right)}{nu_1\left(c,\frac{z}{n}\right)}\right)}{\partial n} > 0$$

And this is exactly the standard single crossing property, see, e.g., Mirrlees (1971).

In the unidimensional case, Theorem 1 combined with Lemma 1 tells us that any allocation is incentive compatible if $z(n)$ is non-decreasing and $U(n)$ satisfies the envelope condition, which is 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 4. *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 differentiable), 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. When $T(z)$ is differentiable, $n \mapsto z$ is locally diffeomorphic, and $U(n)$ satisfies the envelope condition, the proof to Theorem 1 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 60), 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 5), which in turn implies that $z'(n) > 0$ (see footnote 60). $\square$

Hence, in the unidimensional case, Theorem 1 tells us that any strictly increasing differentiable monotonic function $z(n)$ with differentiable $U(n)$ satisfying the envelope condition is incentive compatible.[58] Lemma 1 then tells us that the limit of such functions is incentive compatible. But note that the limit of strictly monotonic functions is necessarily weakly monotonic; moreover, *any* (weakly) monotonic function can be expressed as the (pointwise) limit of strictly monotonic differentiable functions.[59] Thus, in the unidimensional case the sufficient condition in Theorem 1 combined with Lemma 1 reduces to saying that any non-decreasing $z(n)$ with $U(n)$ satisfying

[57] Technically, Assumption 1 tells us that $\frac{\partial\left(\frac{u_2\left(c,\frac{z}{n}\right)}{nu_1\left(c,\frac{z}{n}\right)}\right)}{\partial n} > 0$ or $\frac{\partial\left(\frac{u_2\left(c,\frac{z}{n}\right)}{nu_1\left(c,\frac{z}{n}\right)}\right)}{\partial n} < 0$. But if $\frac{\partial\left(\frac{u_2\left(c,\frac{z}{n}\right)}{nu_1\left(c,\frac{z}{n}\right)}\right)}{\partial n} < 0$, then $\frac{\partial\left(\frac{u_2\left(c,\frac{z}{n}\right)}{nu_1\left(c,\frac{z}{n}\right)}\right)}{\partial(-n)} > 0$ 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.

[58] Lemma 3 in Appendix A.5 below shows that the transfer schedule $T(z)$ must also be differentiable when $z(n)$ is differentiable and monotonic and when $U(n)$ is differentiable and satisfies the envelope condition.

[59] For instance, any monotonic $f(x)$ is the limit of the sequence $\{f_j(x)\}$ with $f_j(x) = j\int_x^{x+1/j} f(t)dt + x/j$.

the envelope condition is incentive compatible.

On the flip side, Theorem 2 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 is not incentive compatible.[60] [61] Hence, in the unidimensional setting, Theorem 2 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.

C.2 Relation to Existing Results when Preferences are Given by Equation 1

Next, it is useful to discuss how our results relate to the other case which has been studied extensively: the model of Rochet (1987) and Rochet and Chone (1998), wherein preferences are linear in T and $\mathbf{n}$, given by Equation 1, expanded as follows:

$$u(T, \mathbf{z}; \mathbf{n}) = y(\mathbf{z}) + T + \mathbf{n} \cdot v(\mathbf{z}) = y(\mathbf{z}) + T + \sum_{i=1}^{K} n_i v^{(i)}(\mathbf{z})$$

Rochet (1987) showed that if utility is given by Equation 1 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 8 and $U(\mathbf{n})$ is convex in $\mathbf{n}$. 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 ($\mathcal{T}$ denotes the matrix transpose and 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):[62]

$$\begin{bmatrix} \sum_i v_{z_i}^{(1)}(\mathbf{z}) \frac{\partial z_i}{\partial n_1} & \cdots & \sum_i v_{z_i}^{(1)}(\mathbf{z}) \frac{\partial z_i}{\partial n_K} \\ & \vdots & \\ \sum_i v_{z_i}^{(K)}(\mathbf{z}) \frac{\partial z_i}{\partial n_1} & \cdots & \sum_i v_{z_i}^{(K)}(\mathbf{z}) \frac{\partial z_i}{\partial n_K} \end{bmatrix} = \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{\mathcal{T}} \begin{bmatrix} v_{z_1}^{(1)} & \cdots & v_{z_1}^{(K)} \\ & \vdots & \\ v_{z_K}^{(1)} & \cdots & v_{z_K}^{(K)} \end{bmatrix} \quad (42)$$

On the other hand, Theorem 2 says that $\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}}\mathbf{z}(\mathbf{n})]^{-1}$ must be positive definite

[60] 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 2.

[61] Theorem 2 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 $n \mapsto z$.

[62] Note, we have used the envelope condition 8 in calculating this Hessian.

wherever the allocation is locally smooth. When utility is given by Equation 1, we have that:

$$\nabla_{\mathbf{n}} FOC(\mathbf{z}(\mathbf{n}); \mathbf{n})[\nabla_{\mathbf{n}} \mathbf{z}(\mathbf{n})]^{-1} = \begin{bmatrix} v_{z_1}^{(1)} & \cdots & v_{z_1}^{(K)} \\ & \vdots & \\ v_{z_K}^{(1)} & \cdots & v_{z_K}^{(K)} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1} \tag{43}$$

Proposition 9. *Equation 42 is positive definite if and only if Equation 43 is positive definite.*

Proof. First, note that:

$$\begin{aligned} & \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{\mathcal{T}} \begin{bmatrix} v_{z_1}^{(1)} & \cdots & v_{z_1}^{(K)} \\ & \vdots & \\ v_{z_K}^{(1)} & \cdots & v_{z_K}^{(K)} \end{bmatrix} \\ = & \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{\mathcal{T}} \begin{bmatrix} v_{z_1}^{(1)} & \cdots & v_{z_1}^{(K)} \\ & \vdots & \\ v_{z_K}^{(1)} & \cdots & v_{z_K}^{(K)} \end{bmatrix} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial z_1}{\partial n_1} & \cdots & \frac{\partial z_1}{\partial n_K} \\ & \vdots & \\ \frac{\partial z_K}{\partial n_1} & \cdots & \frac{\partial z_K}{\partial n_K} \end{bmatrix} \end{aligned}$$

The claim follows because a matrix A is positive definite if and only if $B^{\mathcal{T}}AB$ is positive definite for all invertible matrices B.[63] □

Hence, Theorem 2 requires Equation 43 (and therefore Equation 42 by Proposition 9) to be positive definite whenever the allocation is locally smooth, 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 2 is slightly weaker than Rochet (1987)'s necessary condition when utility is given by Equation 1 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 1, Theorem 1 combined with Lemma 1 tells us that allocations with $(T(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}))$ and convex $U(\mathbf{n})$ such that the envelope condition holds are incentive compatible. First, note that any allocation satisfying the conditions of Theorem 1 must have the envelope condition 8 holding and differentiable, strictly convex $U(\mathbf{n})$. This is because any allocation satisfying the conditions of Theorem 1 has second order conditions holding strictly:[64]

[63] A matrix A of dimension K is positive definite if and only if $\mathbf{x}^{\mathcal{T}}A\mathbf{x} > 0 \forall \mathbf{x} \in \mathbb{R}^K \neq 0$. But if B is invertible then the linear mapping $B\mathbf{x}$ is a bijection from $\mathbb{R}^K$ into $\mathbb{R}^K$. Hence, for any $\mathbf{y} = B\mathbf{x}$ we have $\mathbf{x}^{\mathcal{T}}B^{\mathcal{T}}AB\mathbf{x} = \mathbf{y}^{\mathcal{T}}A\mathbf{y} > 0$.

[64] 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.3 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.

Remark 5. *Any allocation satisfying the conditions of Theorem 1 has second order conditions holding strictly so that:*

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

Hence, Equation 43 (and therefore Equation 42 by Proposition 9) is positive definite. Moreover, any allocation satisfying the envelope condition 8 everywhere with differentiable, strictly convex $U(\mathbf{n})$ satisfies the conditions required in Theorem 1 because: (1) $\mathbf{n} \mapsto \mathbf{z}$ is locally diffeomorphic as Equation 42 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)).[65] Lemma 1 then strengthens this relationship by telling us that the limit of allocations with differentiable, strictly convex $U(\mathbf{n})$ with the envelope condition 8 holding is incentive compatible. Hence, for utility function 1, Lemma 1 tells us that allocations with convex $U(\mathbf{n})$ such that the envelope condition holds are incentive compatible.[66] Thus, Theorem 1 combined with Lemma 1 coincides with the sufficient criterion from Rochet (1987).

C.3 Relation to McAfee and McMillan (1988)

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:

$$u(T, \mathbf{z}; \mathbf{n}) = w(\mathbf{z}, T) + \mathbf{n} \cdot v(\mathbf{z}) \tag{44}$$

In contrast, there appear to be many reasonable utility functions which satisfy our generalized single crossing property such as Equation 7 and Equation 20 used in our numerical simulations later on. These utility functions do not satisfy the generalized single crossing property from McAfee and McMillan (1988).

[65] Also, $T(\mathbf{z})$ is differentiable by Lemma 3 below.

[66] It is well-known that every convex function can be arbitrarily approximated with a differentiable strictly convex function (Koliha, 2004).

C.4 Relation to Carlier (2001)

Finally, our results are related 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 2 and Corollary 2.1) can be checked using local properties of the allocation (note injectivity of $\mathbf{n} \mapsto \mathbf{z}$ can often be checked via Remarks 1 and 2); 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 intractable for numerically solving optimal screening problems.

D Online Appendix: First Order Approaches to Solving Multidimensional Screening Problems

D.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 8 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 $(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n}))$.[67] 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 6. *As an example of this approach, suppose that* $\mathbf{N} \subseteq (-\infty, 0)^K$ *and*

$$u(T, \mathbf{z}; \mathbf{n}) = \sum_{i=1}^{K} z_i + T + \sum_{i=1}^{K} n_i \frac{z_i^{1+\theta_i}}{1+\theta_i} \tag{45}$$

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:

$$z_i(U(\mathbf{n}), \nabla_{\mathbf{n}} U(\mathbf{n})) = [(1+\theta_i) U_{n_i}(\mathbf{n})]^{\frac{1}{1+\theta_i}} \tag{46}$$

Moreover, we can use Equations 45 and 46 to write:

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

[67] 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. Assumption 4' of Spiritus et al. (2022) discusses this point in more detail.

Hence, we can rewrite Equation 17 as (appending the budget constraint on to the objective function with Lagrange multiplier λ):

$$
\begin{aligned}
&\max_{U(\mathbf{n}),\lambda} \int_{\mathbf{N}} \{W\left(U(\mathbf{n}),\mathbf{n}\right) + \lambda[-T(U(\mathbf{n}),\nabla_{\mathbf{n}}U(\mathbf{n}))]\} f(\mathbf{n})d\mathbf{n} \\
&\mathbf{n} \in \operatorname*{argmax}_{\mathbf{n}'} u(T(U(\mathbf{n}),\nabla_{\mathbf{n}}U(\mathbf{n})),\mathbf{z}(U(\mathbf{n}),\nabla_{\mathbf{n}}U(\mathbf{n}));\mathbf{n})\ \forall \mathbf{n}
\end{aligned} \tag{47}
$$

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:

$$
\frac{\partial L(U,\nabla_{\mathbf{n}}U,\mathbf{n})}{\partial U} - \sum_{i=1}^{K} \frac{\partial}{\partial n_i}\left(\frac{\partial L(U,\nabla_{\mathbf{n}}U,\mathbf{n})}{\partial U_{n_i}}\right) = 0 \tag{48}
$$

where

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

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

$$
\left(\frac{\partial L(U,\nabla_{\mathbf{n}}U,\mathbf{n})}{\partial U_{n_i}}\right)\cdot \mathbf{p} = 0 \tag{49}
$$

Unfortunately, Equations 48 and 49 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.

D.2 First Order Approach II: Optimal Control

A second potential approach might be to consider using optimal control theory.[68] 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 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 $U = u(T(U,\mathbf{z},\mathbf{n}),\mathbf{z};\mathbf{n})$.[69]

$$
\begin{aligned}
&\max_{\mathbf{z}(\mathbf{n}),U(\mathbf{n})} \int_{\mathbf{N}} W\left(U(\mathbf{n}),\mathbf{n}\right)dF(\mathbf{n}) \\
&\text{s.t.} \int_{\mathbf{N}} T(\mathbf{z}(\mathbf{n}))dF(\mathbf{n}) \le 0 \\
&\mathbf{n} \in \operatorname*{argmax}_{\mathbf{n}'} u(T(U(\mathbf{n}'),\mathbf{z}(\mathbf{n}'),\mathbf{n}'),\mathbf{z}(\mathbf{n}');\mathbf{n})\ \forall \mathbf{n} \\
&U(\mathbf{n}) = u(T(U(\mathbf{n}),\mathbf{z}(\mathbf{n}),\mathbf{n}),\mathbf{z}(\mathbf{n});\mathbf{n})
\end{aligned} \tag{50}
$$

[68] 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.

[69] 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).

To solve System 50, we will consider $\mathbf{z}(\mathbf{n})$ as control variables, and $U(\mathbf{n})$ as the state variable governed by the envelope condition 8 (which plays the role of the equation of motion). We define the *Hamiltonian* of this problem as:[70]

$$\mathcal{H}(\mathbf{z}, U, \phi, \mathbf{n}, \lambda) = \{W(U, \mathbf{n}) + \lambda[-T(U, \mathbf{z}, \mathbf{n})]\} f(\mathbf{n}) + \phi(\mathbf{n}) \cdot \nabla_{\mathbf{n}} u(T(U, \mathbf{z}, \mathbf{n}), \mathbf{z}; \mathbf{n}) \tag{51}$$

where $\phi(\mathbf{n})$ is a vector of *costate variables.* One can apply a multidimensional analogue to Pontryagin's Maximum Principle to characterize the optimal solution (e.g., Udriste (2009)). The chief complexity is that the usual one dimensional Maximum Principle must be augmented with an additional *integrability condition* mandating that the vector field formed by $\nabla_{\mathbf{n}} u(T(U, \mathbf{z}, \mathbf{n}), \mathbf{z}; \mathbf{n})$ is conservative. This additional complexity appears to substantially complicate use of Hamiltonian optimization in general multidimensional settings.

E Online Appendix: Simulations

E.1 Calibration for Section 5.4

We use income data from the 2019 American Community 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. 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 ACS data), and the fraction of women who do not work (20%, from ACS 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.

E.2 Simulations for Utility Function 25

Simulations when utility is given by Equation 25 are similar to the case when utility is given by Equation 20 after a change of variables: $z_1 = n_1 l_1$, $z_2 = n_2 l_2$, $d = \log(z_1 + z_2 + T)$, $m_1 = -\frac{1}{n_1^{1+\theta_1}}$, $m_2 = -\frac{1}{n_2^{1+\theta_2}}$. Then, we consider agents as maximizing (note $m_1, m_2 < 0$):

$$u(d, \mathbf{z}; \mathbf{n}) = d + m_1 \frac{z_1^{1+\theta_1}}{1+\theta_1} + m_2 \frac{z_2^{1+\theta_2}}{1+\theta_2} - \frac{(-m_1)^{\frac{1}{1+\theta_1}}}{\alpha_1} z_1 - \frac{(-m_2)^{\frac{1}{1+\theta_2}}}{\alpha_2} z_2 \tag{52}$$

[70]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: $u(T, \mathbf{z}; \mathbf{n})$.

which is similar to Equation 20 except now utility is not quadratic in m_1 and m_2, but rather includes a linear term and a term of $-m_1$ to the power of $\frac{1}{1+\theta_1}$ and a term of $-m_2$ to the power of $\frac{1}{1+\theta_2}$. Note that Equation 52 satisfies Assumption 1 on rectangular domains as the Jacobian matrix of $\mathbf{m} \mapsto \frac{\nabla_{\mathbf{z}} u(d,\mathbf{z};\mathbf{m})}{u_d(d,\mathbf{z};\mathbf{m})}$ equals:

$$\begin{bmatrix} z_1^{\theta_1} + \frac{1}{1+\theta_1}\frac{(-m_1)^{\frac{1}{1+\theta_1}-1}}{\alpha_1} & 0 \\ 0 & z_2^{\theta_2} + \frac{1}{1+\theta_2}\frac{(-m_2)^{\frac{1}{1+\theta_2}-1}}{\alpha_2} \end{bmatrix}$$

which is a P matrix given that $z_1, z_2 > 0$ and $m_1, m_2 < 0$. The only material 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:

$$\begin{aligned}
&\max_{\mathbf{z}(\mathbf{m}),U(\underline{\mathbf{m}})} \int_{\mathbf{M}} W(U(\mathbf{m}),\mathbf{m})dF(\mathbf{m}) \\
&\text{s.t.} \int_{\mathbf{M}} [\exp(d(\mathbf{m})) - z_1(\mathbf{m}) - z_2(\mathbf{m})]dF(\mathbf{m}) \leq 0 \\
&U(\mathbf{m}) = U(\underline{\mathbf{m}}) + \int_{\underline{\mathbf{m}}}^{\mathbf{m}} \begin{bmatrix} \frac{z_1(\mathbf{s})^{1+\theta_1}}{1+\theta_1} + \frac{1}{1+\theta_1}\frac{(-s_1)^{\frac{1}{1+\theta_1}-1}}{\alpha_1} z_1(\mathbf{s}) \\ \frac{z_2(\mathbf{s})^{1+\theta_2}}{1+\theta_2} + \frac{1}{1+\theta_2}\frac{(-s_2)^{\frac{1}{1+\theta_2}-1}}{\alpha_2} z_2(\mathbf{s}) \end{bmatrix} \cdot d\mathbf{s} \\
&\frac{\partial z_1}{\partial m_1}(\mathbf{m}) > 0, \frac{\partial z_2}{\partial m_2}(\mathbf{m}) > 0, \frac{\partial z_1}{\partial m_1}(\mathbf{m})\frac{\partial z_2}{\partial m_2}(\mathbf{m}) - \frac{\partial z_1}{\partial m_2}(\mathbf{m})\frac{\partial z_2}{\partial m_1}(\mathbf{m}) > 0 \\
&\left(z_1^{\theta_1}(\mathbf{m}) + \frac{1}{1+\theta_1}\frac{(-m_1)^{\frac{1}{1+\theta_1}-1}}{\alpha_1}\right)\frac{\partial z_1}{\partial m_2}(\mathbf{n}) = \left(z_2(\mathbf{n})^{\theta_2} + \frac{1}{1+\theta_2}\frac{(-m_2)^{\frac{1}{1+\theta_2}-1}}{\alpha_2}\right)\frac{\partial z_2}{\partial m_1}(\mathbf{m}) \\
&d(\mathbf{m}) = U(\mathbf{m}) - \left[m_1\frac{z_1(\mathbf{m})^{1+\theta_1}}{1+\theta_1} + m_2\frac{z_2(\mathbf{m})^{1+\theta_2}}{1+\theta_2} - \frac{(-m_1)^{\frac{1}{1+\theta_1}}}{\alpha_1}z_1(\mathbf{m}) - \frac{(-m_2)^{\frac{1}{1+\theta_2}}}{\alpha_2}z_2(\mathbf{m})\right] - z_1(\mathbf{m}) - z_2(\mathbf{m})
\end{aligned} \tag{53}$$

By Proposition 1 any solution to System 53 will have diffeomorphic $\mathbf{m} \mapsto \mathbf{z}$. Moreover, we know that any solution to System 53 satisfies the envelope condition 8 everywhere; hence, if we confirm that $\forall\mathbf{m}$ we have $u(T(\mathbf{z}(\mathbf{m})),\mathbf{z}(\mathbf{m});\mathbf{m}) \geq u(T(\mathbf{z}(\mathbf{m}')),\mathbf{z}(\mathbf{m}');\mathbf{m})$ for $\mathbf{m}' \in \partial\mathbf{M}$, Theorem 1 ensures that the allocation is incentive compatible.

E.3 Additional Simulation Figures

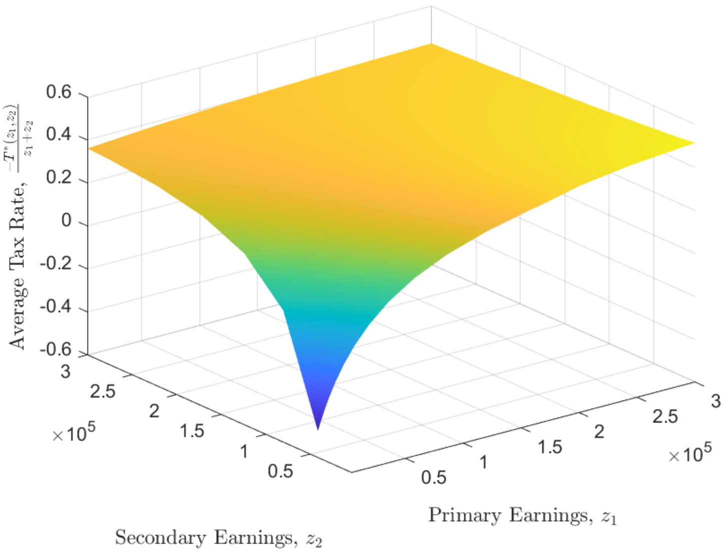


Figure 7: Optimal Average Tax Schedule for Couples

Note: This figure shows the optimal average tax rates for couples, assuming utility is given by Equation 25. $f(\mathbf{n})$ is calibrated to match the joint income distribution from the ACS 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 ACS data), and the fraction of women who do not work (20%, from ACS 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.

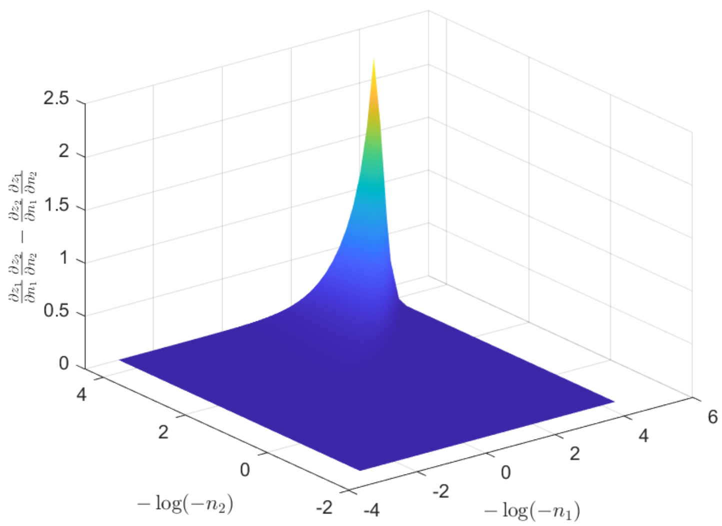
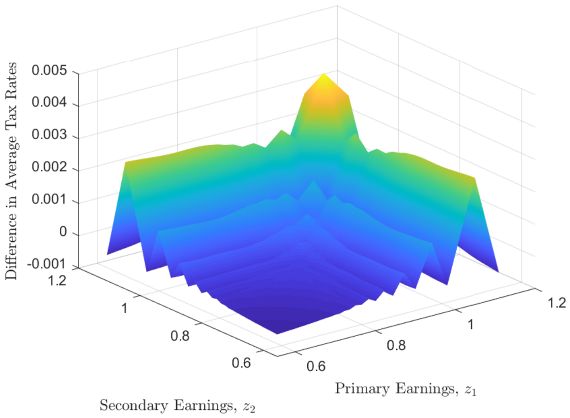
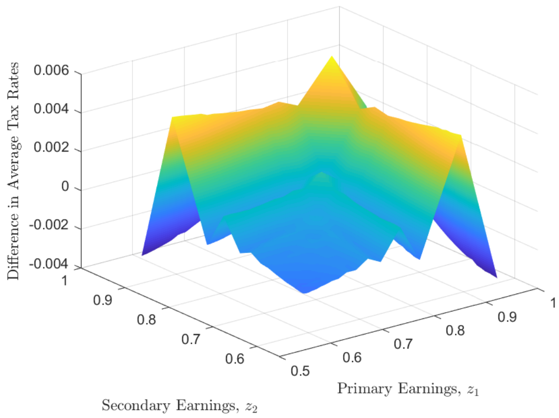


Figure 8: Jacobian Determinant, Couples Taxation Using ACS Data and Log Utility Over Consumption

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 25. 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 ACS 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 ACS data), and the fraction of women who do not work (20%, from ACS 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) Weaker Redistributive Preferences



(b) Stronger Redistributive Preferences

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 18. Panel 9a shows this difference for weaker redistributive preferences from Section 5.3 and panel 9b shows this difference for the stronger redistributive preferences 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).