Robustly Optimal Income Taxation*

Maren Vairo†

January 14, 2024

Click here for the latest version

Abstract

We study the design of a tax rule to redistribute income across heterogeneous workers optimally. The social planner faces uncertainty about the possible income choices available to each type of worker, and, therefore, cannot perfectly predict the income distribution induced by a given tax rule. In the face of this uncertainty, the planner maximizes her worst-case expected payoff. We show that using a progressive tax rule is optimal regardless of the planner's preference for redistribution and that it is uniquely so under an additional richness assumption on the set of income choices that the social planner knows is available to workers. This result stands in contrast to the familiar zero-taxation-at-the-top result that arises generally (absent specific distributional assumptions) in the Bayesian optimal taxation model of Mirrlees (1971). To that extent, our robust approach to uncertainty about workers' income possibilities provides a new foundation for progressive income taxation—a feature that is prevalent in most existing tax systems—that does not rely on parametric assumptions on the distribution of workers' productivity, or on the social planner's attitude toward income inequality.

*I am deeply grateful to Eddie Dekel and Bruno Strulovici for their constant guidance and support, as well as to Piotr Dworczak, Yingni Guo, Asher Wolinsky, and Alessandro Pavan for many detailed comments and discussions at different stages of this project. For helpful comments, I thank George-Marios Angeletos, Ian Ball, Marcus Berliant, Benjamin Brooks, Meghan Busse, Modibo Camara, Joyee Deb, Peter Klibanoff, Panagiotis Kyriazis, Elliot Lipnowski, Alessandro Lizzeri, Wojciech Olszewski, Francisco Pareschi, Harry Pei, Ludvig Sinander, and Udayan Vaidya.

†Deparment of Economics, Northwestern University. Email: mvairo@u.northwestern.edu

1 Introduction

Labor income taxation is a crucial tool for governments to redistribute income and provide workers with social insurance against income shocks. Consistently with their relevance, reforms of the income tax system are constantly under debate in most OECD economies (Piketty and Saez, 2013). At the center of this debate is the so-called equity-efficiency tradeoff: A more progressive income tax is beneficial from a utilitarian perspective as it contributes to a more equitable distribution of well-being across the population. However, it may also negatively affect efficiency by distorting individuals' incentives to work. An extensive theoretical literature, dating back to Ramsey (1927) and Mirrlees (1971), has been devoted to the optimal resolution of this tradeoff.

From the perspective of the government, designing the income tax optimally is *informationally demanding*, as it requires knowledge of workers' preferences and productivity, both of which are highly idiosyncratic. The existing optimal taxation literature is predicated on the assumption that the social planner has perfect knowledge, at the aggregate level, about these details of the environment. However, as was already highlighted by Mirrlees (1971), this information is "difficult to estimate in real economies". In this paper, we relax the strong assumption that the social planner possesses all the relevant details on the primitives of the economy and study the optimal income tax when the planner is subject to this form of uncertainty.

We examine the problem of a social planner tasked with devising the labor income tax to maximize a weighted sum of utilitarian welfare and tax revenue. As in Mirrlees (1971), workers in the economy are heterogeneous with regards to their *productivity*, which we denote by θ. Roughly, θ determines the set of *income opportunities* that are available to the worker—i.e., the set of distributions over income production that the worker can choose from and the associated labor disutility costs—and we make the assumption that a worker with higher productivity faces a larger set of income opportunities (in the sense of set inclusion). Workers' preferences are assumed to be quasilinear in consumption, and separable in leisure and consumption. If the planner does not face any ambiguity about the primitives of the environment, then Mirrlees' model—or more precisely, the version of it studied by Diamond (1998) in which preferences are assumed to be quasilinear—is a special case of the economy we have just described. There, the only information constraint the social planner faces stems from the fact that workers' productivity is unobservable.

Our primary departure from this framework is that, in our model, the social planner faces unquantifiable uncertainty concerning the set of income opportunities that are available to each type of worker. To elaborate, we assume that the mapping from a worker's type θ to

the set of income opportunities that are available to him is described by a correspondence, which we refer to as the economy's *production technology*. The social planner is assumed to know the distribution of workers' productivity θ, but to have only partial knowledge regarding the production technology: She knows a minimal set of income opportunities, referred to as the *baseline technology*, that are available to a worker of type θ, but cannot rule out the possibility that workers make income choices outside of this set. Thus, the social planner entertains all possible θ-contingent sets of income opportunities so long as they are monotone, and include the known minimal set for every θ. She then formulates a tax rule designed to maximize her worst-case payoff across all these possible scenarios.

To better illustrate the environment, consider the case in which productivity is simply determined by the worker's level of education. In our model, the planner is assumed to know the distribution of levels of education in the population. However, what remains uncertain is how a worker's education influences his ability to earn income—e.g., what labor opportunities are available to him, and what is the labor disutility associated with income production. As a result, the planner cannot fully predict the income choice of a worker with type θ in response to the tax schedule that she implements. This type of uncertainty could arise if, for example, workers were subject to income shocks at the individual (e.g., disease or job loss) or the aggregate level (e.g., an economic recession). In the face of this uncertainty, we assume that the planner is unable to postulate a parametric relation between workers' education and their choices of income production, and how these respond to changes in the tax. Instead, she seeks to insure herself against these unforeseeable outcomes by choosing the tax rule that provides the highest worst-case payoff guarantee.

Our social planner thus faces two main sources of uncertainty. The first one stems from the fact that workers' productivity is either unobserved by the planner or the income tax cannot directly condition on it. In our model, as in the canonical Mirrleesian framework, the planner has a well-formed belief about how this uncertainty is distributed, and therefore, she deals with it using a Bayesian criterion. There is a second source of uncertainty resulting from the planner's limited knowledge about the economy's production technology. We assume that the nature of this uncertainty is unquantifiable, and the planner approaches it using the max-min criterion.

Our main finding (Theorem 1) is that the planner can maximize worst-case welfare by using a *progressive tax rule*, meaning a tax schedule with weakly increasing marginal tax rates. This result stands in contrast with the canonical Mirrleesian model in which the planner's only source of uncertainty pertains to the realization of workers' productivity θ. In that setting, a standard no-distortion-at-the-top argument implies that, if workers' productivity is bounded from above, the optimal tax rule involves zero marginal taxation of top incomes.

This result is often criticized due to its limited policy relevance (Diamond and Saez, 2011) and has given rise to subsequent work that has shown that zero-taxation of top incomes can be overturned under specific assumptions on the distribution of workers' productivity (Diamond, 1998; Saez, 2001). Conversely, in our model, the optimal top marginal tax rate is generally strictly positive.[1] Beyond the taxation of top incomes, we show that *the entire* tax schedule is progressive at the optimum—a result which, to the best of our knowledge, does not arise in existing versions of the Mirrlees model. To that extent, our robust approach to uncertainty about workers' income opportunities provides a foundation for progressive income taxation that does not rely on assumptions on the distribution of workers' productivity or on the social planner's preference for redistribution.

The intuition behind the optimality of tax progressivity can be better understood by first looking at the hypothetical scenario in which the planner *knows workers' types* θ—and can condition the tax on θ—and is restricted to using taxes that are an *affine* function of workers' income. In this case, the slope of the optimal type-specific affine tax rule is chosen to balance workers' welfare and tax revenue. In particular, the planner is more concerned with raising revenue relative to the worker, and thus, she needs to provide the worker with incentives to produce higher income. One would anticipate that the slope of the optimal affine tax schedule increases with θ, as it is relatively easier to incentivize a more productive worker to produce a given income level. As a result, the presence of *unobservable worker heterogeneity* introduces a force pointing to the optimality of a convex tax: In the worst case, more productive workers self-select into "higher" income choices.[2] Consequently, using a convex tax rule allows the planner to "target" more productive workers and tax them more heavily, and in that way, replicate, to a partial extent, her optimal (affine) rule in the benchmark with observable θ.

Naturally, adjustments to the tax schedule are necessary when workers' types are unobservable to accommodate their private incentives. The key steps in our analysis consist of showing that the above convexifying force remains valid and that the robustly optimal tax is indeed convex. Specifically, building on the arguments of Carroll (2015), we first show that a type-specific affine tax rule is optimal *if worker types are known by the planner*, and therefore the restriction to affine taxes in the previous paragraph is without loss of optimality. At a high level, the reason is that affine taxes serve the purpose of aligning the planner's and

[1] It is zero only in the extreme case where the planner finds it optimal to raise taxes solely through lump-sum payments. We provide conditions under which the tax schedule is not flat in Proposition 2.

[2] In our setting in which workers choose *lotteries* over income, it is, in general, not possible to rank their income choices. However, we show that in the worst case, more productive workers make income choices that lead to higher *expected after-tax income*. Unlike the Mirrleesian setting, this monotonicity does not follow from a single-crossing assumption on workers' preferences, but is instead a consequence of workers' minimal income choice sets being increasing.

workers' worst-case payoff, and, in this way, enable the planner to attain her best possible payoff guarantee *conditional on workers' type*. Second, by expanding upon the insight in the previous paragraph, we show that, in the unobservable-type setting, there is an optimal way to aggregate these type-specific affine taxes and that this aggregation renders the optimal tax convex. In doing so, our proof highlights a desirable feature of progressive taxes, which is that they allow the planner to hedge against the adversarial possibility of workers taking on excessive income risk.

These arguments are centered around the joint maximization of welfare and revenue, and they suffice to establish that the optimal tax is convex regardless of the planner's preference for redistribution. If, in addition, the planner is inequality-averse and thus her objective assigns higher marginal value to the welfare of worse-off workers—i.e., workers with a lower type—then by similar reasoning, an additional force pointing to the optimality of tax progressivity ensues. This intuition highlights that, even though we show that (weak) tax progressivity is optimal under great generality, the actual shape of the optimal tax—e.g., the extent to which marginal tax rates increase with income—is, as one would expect, sensitive to the details of the economy, such as the distribution of workers' types and the government's attitude toward inequality.

Having established progressivity as a key qualitative aspect of the optimal tax rule, we proceed to further characterize its shape. Deriving closed-form expressions for the optimal marginal tax rate is challenging: The fact that workers' income choices are random and that the planner's objective is max-min implies that the optimal control techniques applied by Mirrlees (1971) to solve the model do not apply here. Instead, we follow an approach similar in spirit to the one used by Saez (2001), which involves deriving necessary conditions for optimality by studying the effect of perturbations around the optimal tax rule. The key technical challenge that arises when doing so in our setting is that, unlike the canonical framework, we must take into account the fact that the economy's production technology is *not* fixed. Specifically, a perturbation in the income tax leads to a change in *(i)* the *production correspondence that achieves the worst-case*, and *(ii) workers' income choices* under any given technology. By extending one of the envelope theorems for arbitrary choice sets by Milgrom and Segal (2002),[3] we show that the planner's worst-case payoff is (directionally) differentiable with respect to tax perturbations and characterize the derivatives. This

[3]As we show, the planner's worst-case payoff can be computed as a constrained optimization problem over the space of income lotteries, and this problem can in turn be cast as a saddle-point problem. Milgrom and Segal (2002) provide an envelope theorem for such saddle-point problems (Theorem 5), but their result assumes continuous differentiability of the constraint with respect to the parameter, which does not always hold in our setting. To address this, we show an extension of their theorem, which holds when the constraint is not necessarily differentiable but instead is concave in the parameter. The details are in Appendix B.

approach can be applied in other settings to derive necessary conditions for optimality in the presence of ambiguity about agents' choice sets.

Following this method, we solve for the optimal affine tax, and derive conditions ensuring that affine taxes are strictly suboptimal and, therefore, the optimal tax must be strongly progressive in the sense of using a strictly higher tax rate at the top of the income distribution than at the bottom. Similarly to Saez (2001), our discussion allows to identify two types of effects of tax perturbations: A *mechanical* response, which stems from the welfare effect keeping workers' choices fixed, and a *worst-case behavioral* response which captures the change in workers' labor decisions as a consequence of a change in the tax code. Because the latter effect focuses on *adversarial* income choices by workers, its expression and interpretation are qualitatively different from its analog in the Bayesian setting.

Furthermore, by specializing to the case in which the minimal choice set that the planner contemplates corresponds to what would obtain in a classical Mirrlees economy, we provide conditions for the optimal marginal tax rates at the bottom and at the top of the income distribution.[4] In particular, we replicate a result of zero taxation of bottom incomes that arises in the canonical framework (Seade, 1977). We apply these results to a two-type economy, for which it is possible to fully solve for the optimal tax rule. Finally, in the last section of the paper, we discuss the robustness of our results to some of the modeling assumptions, such as the quasilinearity of workers' preferences and the planner's (in)ability to screen the worker's productivity.

Overall, the paper contributes to the understanding of optimal labor income taxation by incorporating the possibility that the government's information about workers is limited. Given that, in reality, the heterogeneity in workers' preferences and productivity is likely to be very rich, we believe that the environment that we study has considerable practical relevance. Moreover, our approach allows us to justify the use of increasing marginal tax rates—a feature that prevails in almost all existing tax systems, but that is *not* predicted to be optimal by the canonical optimal taxation models. To that extent, the paper also joins a growing literature establishing that simple or intuitive mechanisms are optimal when the designer faces uncertainty about the environment and evaluates that uncertainty using a max-min objective (Chung and Ely, 2007; Chassang, 2013; Carroll, 2015). More broadly and beyond the income taxation application, by building on the model by Carroll (2015), we provide a framework to study robust principal-agent contracting under moral hazard and adverse selection, in which case our main result provides a foundation for the use of concave contracts.

[4]We show that the optimal tax rule is affine at the tails of the income distribution, so we characterize the optimal constant rate for those two regions.

The rest of the paper is structured as follows. In the remainder of this section, we discuss the related literature. In Section 2, we describe the model. Section 3 contains the main result regarding the optimality of progressive taxation. There, we also show that, under certain conditions, any optimal tax rule must be progressive, and we provide sufficient conditions for the optimal tax rule to be *strongly progressive*. Section 4 discusses the special version of the model in which the baseline technology corresponds to the one assumed in the canonical Mirrlees model. We conclude in Section 5 by discussing extensions of our model, and the robustness and limitations of the analysis.

1.1 Related literature

Mirrlees (1971) introduced the canonical framework that is used to study labor income taxation. In it, a utilitarian social planner uses distortionary income taxes to optimally redistribute income across a population of workers with heterogeneous skills. The planner is constrained by the fact that she can observe and tax workers' earnings but not their skills or income choices, which gives rise to a non-trivial equity-efficiency trade-off. Mirrlees solved for the optimal non-linear tax schedule as a function of workers' earnings. Even though the solution is complex and its details are sensitive to the assumptions about the economy, substantial work has built upon the Mirrleesian model and established more general qualitative features of the optimal tax. Piketty and Saez (2013) provide an extensive review of this literature. Among the most salient findings is the result that, if the distribution of workers' skills is bounded, the marginal tax rate should be zero at the top of the income distribution (Sadka, 1976; Seade, 1977). This (in)famous result has been challenged by subsequent work by Diamond (1998) and Saez (2001), who showed that the zero-taxation-at-the-top result does not hold under the assumption that the right-tail of the distribution of workers' skills follows a Pareto distribution. There is additionally a symmetrical zero-taxation-at-the-bottom result that establishes that taxation of low earnings should be zero if all workers participate in the labor force (Seade, 1977).

Our main point of departure from the existing literature is to augment the Mirrleesian framework to allow for the possibility that the social planner faces unquantifiable uncertainty about the primitives of the economy, and uses a max-min criterion to find the optimal tax.[5,6]

[5]This approach is not to be confused with the one in the strand of the literature that studies optimal taxation in the presence of a planner with Rawlsian—commonly known as max-min—preferences (Phelps, 1973; Boadway and Jacquet, 2008). In those papers, the government has complete information about the primitives and uses the max-min criterion as a welfare aggregator.

[6]Our assumption that the government is informationally constrained shares the same motivation as the *robust control theory* of Hansen and Sargent (2001) in macroeconomics and finance, which introduces uncertainty by the government regarding the dynamic data generating process of shocks.

A second difference between our model and the one in Mirrlees (1971) is that we allow for the possibility of *stochastic output*. Formally, we dispense with the assumption that workers are restricted to making deterministic income choices, which introduces a moral hazard component in the planner's problem. It has been widely recognized that income risk is a realistic dimension that should be incorporated in models of income taxation (Mirrlees, 1974; Varian, 1980).[7] Due to the technical challenges entailed in adding moral hazard to the Mirrleesian framework, most existing attempts to study income risk shut down adverse selection from the model by assuming that workers are homogeneous. In that setting, the income taxation problem becomes one of optimally trading off incentives and insurance provision (Laffont and Martimort, 2009). In contrast, our framework jointly accommodates moral hazard and adverse selection, both plausible informational constraints faced by a real-world tax authority.

Using this approach, we provide new qualitative insights about the optimal tax code. Namely, we provide a novel foundation for the use of progressive income taxation, which does not rely on distributional assumptions about workers' types—e.g., we do not require the support of workers' skills to be unbounded. We defer a more detailed discussion of how our main finding relates to existing results about tax progressivity to Section 3.2. Here, we emphasize that, as the intuition explained in the introduction highlights, it is the *combination* of the robust approach to uncertainty and the presence of stochastic income that leads to the optimality of tax progressivity.

More broadly, our work is related to the literature on mechanism design with redistributional concerns. Recent contributions to this literature include the study of dynamic income taxation (Farhi and Werning, 2013; Golosov et al., 2016; Stantcheva, 2017; Makris and Pavan, 2021), redistribution in two-sided markets (Dworczak et al., 2021), allocation of public resources of heterogeneous quality (Akbarpour et al., 2023) and in the presence of a private market (Kang, 2023), and the taxation of externalities (Pai and Strack, 2023). In related work, Berliant and Gouveia (2022) incorporate a different robustness concern into a model of redistributive taxation with voting, by studying a setting where the government faces finite-sample uncertainty regarding the realized distribution of workers' abilities and is restricted to using taxes that balance the budget for every realization of this uncertainty.

Our paper also contributes to the growing literature on contracting and mechanism design with a worst-case objective, a comprehensive review of which is provided by Carroll (2019). Within that literature, we build on the framework introduced by Carroll (2015), who studied robust contracting under moral hazard and ambiguity about the agent's production technology. Our model differs from Carroll's in that the designer faces both moral

[7] See Chapter 11 in Tuomala (2016) for a review of this line of work.

hazard and adverse selection; the latter arising due to the heterogeneity in agents' sets of baseline actions. A second, albeit less important difference, is that the principal's objective in our problem is a weighted sum of agents' welfare and revenue, whereas the principal in Carroll (2015) is concerned only with revenue maximization. We expand on that paper's result that linear contracts are optimal, by showing that the introduction of adverse selection justifies the use of *concave contracts*.[8,9] Recent work has augmented the model of Carroll (2015) along other dimensions by studying teamwork (Dai and Toikka, 2022), common agency (Marku et al., 2022), uncertainty about the agent's beliefs (Rosenthal, 2022), more general conditions for the optimality of linear contracts (Walton and Carroll, 2022), the access to past data about the agent's actions (Antic and Georgiadis, 2023), and the role of random mechanisms (Kambhampati, 2023).

Even though we focus on the application to income taxation, our results apply more broadly to models of robust contracting with moral hazard and adverse selection. In this vein, Carroll and Meng (2016) provide a foundation for linear contracts in a model with moral hazard where the agent holds private information about an income shock, and the principal faces ambiguity over the distribution of shocks. The nature of ambiguity in their paper differs considerably from ours, leading to qualitatively different results about the optimal contract.

We also relate to the work by Garrett (2014), who studies robust contracting for cost-based procurement à la Laffont and Tirole (1986). Like ours, Garrett's model involves the principal using a combination of Bayesian mechanism design with a max-min objective. In both cases, the optimal mechanism is shaped by the principal's need to jointly hedge against the worst-case scenario and to optimally resolve the standard Bayesian trade-off between efficiency and information rents given to the agent. We show that a progressive tax achieves this objective in our setting, whereas in Garrett (2014), this leads to the optimality of a different class of contracts commonly known as fixed-price cost-reimbursement.

2 Model

Preliminaries. In what follows, for a metric space X we use $\Delta(X)$ to denote the set of Borel probability distributions on X, and we equip $\Delta(X)$ with the topology of weak convergence. For any $x \in X$, we use δ_x to denote the Dirac measure on $\{x\}$. For any $F \in \Delta(X)$, we use $\text{supp}(F)$ to denote the support of F. Any space of functions from X to

[8]Here, we define a contract as the payment made *to* the agent. The counterpart of a concave contract is a convex (progressive) tax.

[9]Barron et al. (2020) provide a different foundation for concave contracts by studying a moral hazard environment where the agent may deliberately garble production and the principal has a Bayesian objective.

$\mathbb{R}$ is equipped with the sup norm.

2.1 General Framework

There is a social planner (she), and a non-atomic unit mass of workers (he, for the singular) indexed by i. Each worker i produces income $y_i \in Y \equiv [0, \overline{y}]$, where $\overline{y} \in \mathbb{R}_{++}$ and Y is the set of possible income realizations in the economy. The social planner designs an income tax rule, mapping income realizations into payments made by the worker. Formally, a tax rule is a function $T : Y \to \mathbb{R}$. We restrict attention to a class of tax rules which we refer to as feasible. A tax rule T is said to be *feasible* if it is continuous[10] and satisfies $y - T(y) \geq 0$.[11] In line with the rest of the optimal taxation literature, we restrict attention to deterministic tax rules.[12] We let $\overline{\mathbf{T}}$ denote the set of feasible tax rules. A worker's consumption, as a function of his earned income, equals after-tax income $y - T(y)$. Our formulation assumes that a worker's earned income is *observable* by the social planner.

Workers' labor decision. Conditional on the tax rule T, workers make labor decisions to maximize their utility. Workers' preferences are assumed to be quasilinear in consumption, and additively separable in consumption and labor. Their income choice consists of a pair (F, ϕ) where $F \in \Delta(Y)$ is the worker's stochastic earned income, and $\phi \in \mathbb{R}_+$ is the labor disutility associated with his income choice. As a result, given the tax rule imposed by the social planner, a worker's payoff from making income choice (F, ϕ) is equal to $\mathbb{E}_F[y - T(y)] - \phi$.

We now describe the economy's production technology—i.e., the set of income choices (F, ϕ) available to each worker. We assume that workers are heterogeneous with respect to how much income they can produce and their labor disutility, and that this heterogeneity is captured by workers' *types*. Specifically, worker i's *type* is described by the random variable $\theta_i \in \Theta$, where Θ is a Borel subset of $\mathbb{R}$. Types are independently and identically distributed (i.i.d.) across workers according to the probability distribution $\mu \in \Delta(\Theta)$.[13] We denote $\underline{\theta} \equiv \inf \Theta$ and $\overline{\theta} \equiv \sup \Theta$. A worker with type θ is endowed with the compact set of income

[10] Our results extend to the case in which only lower semi-continuity of T is required: A minor extra verification in the proof of Theorem 1 establishes that any lower semi-continuous tax is weakly dominated by one that is continuous, and thus the restriction to continuous taxes is without loss of optimality.

[11] The analysis goes through if a different lower bound on consumption is used instead. Moreover, it is possible to show that a lower bound on consumption arises endogenously at the optimum, provided that the social planner values the welfare of a worker with arbitrarily low consumption more than she values raising an extra dollar of revenue.

[12] This may entail a loss in optimality: We can apply an argument similar to Kambhampati (2023) to provide conditions under which the planner can do strictly better by using a random mechanism.

[13] In our analysis, we make use of the Exact Law of Large Numbers for a continuum of i.i.d. random variables (Sun, 2006; Duffie and Sun, 2012) which implies that the aggregate distribution of workers' types is almost surely equal to μ.

choices $M(\theta) \subseteq \Delta(Y) \times \mathbb{R}_+$. We refer to the correspondence $M : \Theta \rightrightarrows \Delta(Y) \times \mathbb{R}_+$ as the economy's *production technology*.

Given the tax rule T, the income choice of a worker of type θ is described by

$$V_w(T|M(\theta)) \equiv \max_{(F,\phi)\in M(\theta)} \{\mathbb{E}_F[y - T(y)] - \phi\}.$$

The following example illustrates that the canonical model by Mirrlees (1971) can be described using our terminology, and therefore is a special case of our environment.

Example 1 (Mirrlees economy)**.** Suppose that $\theta_i > 0$ represents worker i's *wage*, and that workers choose how many units of labor to provide, denoted by $l \in [0, 1]$. Assume that $\Theta \subseteq Y$. Given his wage θ and his income choice l, a worker's earned income is *deterministic* and equal to $y = l \times \theta$. The continuous, increasing function $\Phi : [0, 1] \to \mathbb{R}_+$ governs workers' labor disutility. Assuming quasilinearity of preferences over consumption, the utility of a worker of type θ who faces the tax rule T and chooses to earn y dollars of income (by providing y/θ units of labor) is equal to $y - T(y) - \Phi(y/\theta)$. In this setting, the production technology is equal to

$$M(\theta) = \{(\delta_y, \phi) \in \Delta(Y) \times [0, \Phi(1)] : y \leq \theta, \phi \geq \Phi(y/\theta)\}.^{14}$$

Social planner's information. The social planner faces unquantifiable uncertainty about workers' production technology: She only has limited knowledge regarding the correspondence M. All of the other features of the environment, including the distribution μ, are assumed to be known by the planner. Formally, we define a correspondence $M^0 : \Theta \to \Delta(Y) \times \mathbb{R}_+$, which we refer to as the *baseline technology*, representing the set of income choices that the planner *knows* are available to each type. We assume the social planner entertains any production technology M belonging to the *plausible set*, as defined in Definition 1.

Definition 1 (Plausible technologies)**.** The plausible set of technologies $\mathcal{M}$ consists of all correspondences $M : \Theta \rightrightarrows \Delta(Y) \times \mathbb{R}_+$ such that: *(i)* M is continuous and compact-valued, *(ii)* M is monotone—i.e., $\theta' > \theta$ implies $M(\theta') \supseteq M(\theta)$—, and *(iii)* $M(\theta) \supseteq M^0(\theta)$ for all $\theta \in \Theta$.

Intuitively, other than knowing that M satisfies some regularity conditions, we assume that the social planner only knows that: *(i)* there is a *minimal set of income opportunities*

[14] Unlike the standard formulation of the model, we allow the worker to choose (δ_y, ϕ) satisfying $\phi > \Phi(y/\theta)$. This modification is irrelevant, given that in equilibrium, workers always choose (δ_y, ϕ) satisfying $\phi = \Phi(y/\theta)$. We introduce it to be consistent with our maintained assumption that $M(\cdot)$ is monotone.

that are available to each worker, and this minimal set may be contingent on the worker's type θ, and *(ii)* a worker with a *higher type* has access to a *larger set of income choices*. We make the following assumption on M^0.

Assumption 1 (Baseline technology)**.** M^0 is continuous, compact-valued and monotone. Moreover, the following hold

(i) *Free option*: there exists $F \in \Delta(Y)$ such that $(F, 0) \in M^0(\underline{\theta})$, and

(ii) *Non-triviality:* for a strictly positive measure of θ, there exists $(F, \phi) \in M^0(\theta)$ such that $\mathbb{E}_F[y] - \phi > 0$.

Continuity and compact-valuedness of M and M^0 are mild technical assumptions that ensure that the respective optimization problems faced by the workers and the social planner are well-defined. Monotonicity captures the idea that more productive workers are able to either earn higher income, or to earn the same income as a lower type at a lower labor disutility. Interestingly, our results remain unaffected if monotonicity is dropped from Definition 1. Thus, it is irrelevant whether or not the planner believes that the ranking of workers that M^0 induces is preserved under all realizations of the production technology. In contrast, the assumption that M^0 is monotone plays a key role. The "free option" assumption implies a lower bound on workers' equilibrium utility, and "non-triviality" rules out the uninteresting situation in which the first-best at the baseline M^0 is to have all workers produce zero income.

Observe that Assumption 1 is satisfied if we set M^0 to be defined as in the Mirrlees economy of Example 1; if we add in the example the assumption that non-triviality holds for the most productive worker in the economy—this is a special version of our model that we will study in Section 4.

In this framework, the timing of events can be summarized as follows: *(i)* the social planner offers a feasible tax rule T; *(ii)* workers learn T, θ_i and their choice set $M(\theta_i)$[15]; *(iii)* each worker $i \in [0, 1]$ makes an income choice in $M(\theta_i)$; *(iv)* workers' income $y_i \in Y$ is realized, and is observed by the planner; and *(v)* worker i makes the tax payment $T(y_i)$ and consumes $y_i - T(y_i)$.

2.2 Social planner's problem

We now turn to describing the social planner's problem of finding the optimal tax rule. Her objective is a weighted sum of average workers' welfare and tax revenue, taking as given

[15]It is irrelevant whether or not workers know the structure of the entire correspondence M.

workers' equilibrium income choices. Specifically, for a given plausible production technology $M \in \mathcal{M}$ and a feasible tax rule $T \in \overline{\mathbf{T}}$, let $(F_\theta, \phi_\theta) \in M(\theta)$ be the equilibrium income choice of a worker with type θ.[16] The planner values type θ's worker-welfare according to $\omega(\theta)W(\mathbb{E}_{F_\theta}[y - T(y)] - \phi_\theta)$, where $\omega : \Theta \to \mathbb{R}_{++}$ is a measurable function representing the social welfare weight associated with a worker of type θ, and $W : \mathbb{R} \to \mathbb{R}_+$ is a utilitarian welfare function. We assume that W is strictly increasing and continuously differentiable.[17]

The second component in the social planner's objective is *expected revenue*, which we assume for simplicity that she values linearly with a constant marginal value that we denote by $\alpha \in \mathbb{R}_{++}$. We discuss the role of this assumption and provide an interpretation for α as the government's marginal value of public funds in the following subsection. We make the following assumption regarding the relative weights of workers' welfare and revenue in the planner's objective.

Assumption 2 (α-bound)**.** There exists $\overline{C} \in \mathbb{R}_+$ such that $c > \overline{C}$ implies that $\mathbb{E}[\omega(\theta)]W'(c) \leq \alpha$.

Finally, we assume that, when indifferent, workers break ties in favor of the income choice that maximizes tax revenue.[18] As a result, the planner's payoff when she offers the tax rule $T \in \overline{\mathbf{T}}$ and given a production technology $M \in \mathcal{M}$ is given by[19]

$$V_P(T|M) \equiv \int_\Theta \{\omega(\theta)W(V_w(T|M(\theta))) + \alpha \max_{(F_\theta,\phi_\theta)\in M(\theta)} \mathbb{E}_{F_\theta}[T(y)]\}\, d\mu(\theta) \tag{2.1}$$

$$s.t. \quad \mathbb{E}_{F_\theta}[y - T(y)] - \phi_\theta = V_w(T|M(\theta)). \tag{2.2}$$

We refer to $V_P(T|M)$ as *total welfare*, to its first component as *total worker-welfare*, and to its second component as *total revenue*, and oftentimes omit the word "total" for conciseness.

The planner's problem is to choose a feasible tax rule that maximizes her worst-case

[16] Here, we are implicitly imposing that any two workers i and i' who have the same type make the same income choice. Below, we introduce the assumption that workers break ties in favor of the social planner. Under that assumption, allowing for the possibility that different workers with the same type make different income choices would be outcome-irrelevant.

[17] The assumption that the planner strictly values workers' welfare is introduced solely to simplify the exposition. Theorem 1 continues to hold if the planner's objective is purely to maximize revenue—i.e., if either $\omega(\theta) = 0$ almost everywhere or/and $W(\cdot)$ is constant.

[18] This assumption ensures that the maximum in the planner's problem is attained. If the worker were instead to break ties adversarially, the supremum in the planner's problem is the same, but is not necessarily attained by some tax rule.

[19] The conditions in the definition of $\mathcal{M}$ ensure that the integral in (2.1) is well-defined for all $M \in \mathcal{M}$ and all feasible T.

payoff across all plausible M's. Formally, the planner's problem is

$$\sup_{T \in \overline{\mathbf{T}}} \inf_{M \in \mathcal{M}} V_P(T|M). \tag{2.3}$$

We refer to a tax rule as being *worst-case optimal* if it is feasible and attains the value in (2.3).

2.3 Discussion of the model

In this section, we provide an interpretation of the model and its assumptions. We come back to this discussion in Section 5, where we comment on the robustness of our results.

Income risk. One point of departure of our model relative to the standard Mirrleesian framework is that workers' income is stochastic. The interpretation of this assumption is that workers' occupations may intrinsically involve income risk. Examples of sources of uncertainty in workers' labor income abound: e.g., bonus-based compensation schemes, entrepreneurial activity, or freelance work are a few salient ones.[20] Therefore, we view this assumption not as a limitation but as a realistic feature of the model. As mentioned in the discussion of the literature, it has been recognized by Mirrlees (1974) that income risk is an important aspect of workers' labor decisions that the original framework in Mirrlees (1971) does not accommodate. Aside from the technical difficulties stemming from the combination of moral hazard and adverse selection, allowing for the possibility that workers' income is random poses the additional challenge of having to take a stance on, first, what set of income lotteries is available to each worker, and second, on workers' labor disutility as a function of these lotteries. In the absence of a well-founded assumption on preferences and technology over income risk, our max-min approach arises as a natural, non-parametric way to study the problem.

Assumptions on workers' payoffs Our assumption that workers' preferences are quasilinear in consumption has two implications: First, it implies that workers are risk-neutral, and second, that there are no income effects in workers' income choices. Under a Bayesian objective, the optimal income taxation problem with quasilinear preferences was first studied by Atkinson (1995) and Diamond (1998), and has ever since been commonly used by theoretical and empirical work. As in those papers, this (restrictive) assumption enhances the model's tractability and allows us to make progress in studying the question at hand.

[20] See Vereshchagina and Hopenhayn (2009) for a theoretical justification of workers' decision to take on income risk, even in the absence of a risk premium.

We discuss in more detail what happens when the workers' utility of consumption is allowed to be strictly concave in Section 5.

Information structure. Our modeling of the social planner's uncertainty about workers' income possibilities can be interpreted as follows. Workers are heterogeneous with regard to their productivity, which in our model is represented by the type θ. It is assumed that the planner knows that workers are heterogeneous along this dimension, and knows how this heterogeneity is distributed in the population. This assumption is plausible if we interpret θ as either workers' education level or wages. However, the planner does not perfectly know the mapping between workers' types, and how much income they can produce and at what cost. This uncertainty may arise due to workers facing individual or aggregate shocks affecting their labor opportunities.

The critical piece of information that the planner has is that there is a minimal set of choices available to each worker, as described by the baseline technology M^0. A natural candidate for this minimal set consists of income choices made by the worker in the past. In the interesting case, this minimal set is non-constant in the worker's type θ: We can show that, if M^0 were constant in θ, then the problem of finding the optimal tax rule reduces to a single-type problem.[21] As a result, the baseline technology plays a dual role in the model: First, it gives a lower bound on workers' equilibrium payoffs, and second, it dictates the extent to which workers are effectively heterogeneous. A consequence of these two features is that we can rule out trivial situations in which the worst-case scenario is one where all workers are highly unproductive and zero aggregate production is the only feasible choice.

In addition, the uncertainty faced by the planner is rich: Any monotone correspondence, so long as it contains the baseline and satisfies mild regularity conditions, is plausible. In Section 5, we argue that our main result continues to hold under an alternative information structure in which the uncertainty about M is considerably smaller. We also discuss the possibility that the planner may face ambiguity about θ—i.e., she does not know μ.

Assumptions on the social planner's objective. Our assumptions on how the planner values workers' welfare are standard. In particular, our specification implies that the planner's marginal value for a worker's utility has two components: An exogenous one ($\omega(\theta)$) that describes the extent to which the planner intrinsically cares about the welfare of a worker with type θ, and an endogenous one ($W'(\cdot)$) which specifies the marginal value of changing the welfare of a worker of any type as a function of the worker's equilibrium utility.[22] In this

[21] This is a priori not obvious given that θ may directly enter the planner's objective through $\omega(\theta)$.

[22] So long as the planner's objective is increasing in workers' welfare, our results continue to hold under more general valuations by the planner, such as the one proposed by Saez and Stantcheva (2016).

setting, given that workers' equilibrium utilities increase with θ, inequality aversion by the planner would be captured if, for example, W is concave and ω is decreasing. However, our main result (Theorem 1) does not rely on any assumptions about the planner's preference for redistribution.

The assumption that the planner values revenue linearly, on the other hand, differs from the standard Mirrleesian model: There, the social planner faces a budget constraint and revenue does not directly enter her objective. In that setting, society's marginal value for revenue is endogenously described by the shadow value associated with the budget constraint. However, under the max-min approach, it is not immediate how to incorporate a hard budget constraint to the problem because whether or not the constraint holds depends on workers' income choices and thus on the technology M, which the planner does not know at the time of designing the tax code. Instead of using the conservative approach of requiring the budget to balance *for every technology* M, we assume that the government tolerates running a budget deficit but strictly values raising more funds. Allowing the government to run a deficit is, in our opinion, a more realistic assumption. However, as we explain next, this point of departure from the canonical framework is *not* a critical force driving our main result.

For simplicity, we assume that the government's marginal value for public funds is captured by the constant $\alpha > 0$.[23] This assumption is not crucial and our main result (Theorem 1) continues to hold if the planner values total revenue using any non-decreasing function.[24] We may interpret α as the value of devoting an extra dollar of tax revenue to fund alternative projects. To ensure that the problem is well-defined, Assumption 2 requires that the value for revenue is not too low relative to how much the planner values making lump-sum payments to the average worker. To provide a simple interpretation of this assumption, consider the case in which W is the identity function, in which case the assumption requires that $\mathbb{E}[\omega(\theta)] \leq \alpha$. Intuitively, this says that the planner weakly prefers to devote any extra dollar of revenue that she raises to other projects instead of lump-sum redistributing it to workers.[25] As an implication, we can rule out the situation in which the planner finds it optimal to give every worker in the economy an infinite lump-sum payment, which in turn allows us to establish the existence of an optimal tax rule. In the case in which W is not linear, then Assumption 2 would be satisfied if we make a standard Inada-type assumption, whereby the marginal value of giving workers an extra dollar converges to zero if workers'

[23] The same approach is used, in a different setting, by Akbarpour et al. (2023).

[24] For example, a hard budget constraint can be incorporated by assuming that the value for revenue is zero if the budget is balanced, and $-\infty$ if it is not. More generally, we could set the planner's value for revenue to capture any non-linear costs of government debt.

[25] In Mirrlees (1971), there are no other alternative uses for public funds and therefore this condition is satisfied with equality.

after-tax income is arbitrarily high.

Available tax rules. We have simplified the social planner's problem by restricting to tax rules that are deterministic and only depend on realized income. We believe that this restriction is in line with real-life tax systems, but in principle, entails a loss in optimality. In particular, because workers have private information about their types and the production correspondence, the planner could benefit from screening this information by offering a menu of tax rules. We explore this possibility in Section 5. Finally, in keeping with the classical taxation literature, we have assumed that the government observes workers' realized income and can directly tax it. However, an alternative interpretation of randomness in workers' income allows to somewhat relax this assumption, if we regard the noise in income as a reduced-form way to capture workers' tax evasion practices.

3 Main Results

In this section, we state our main result regarding the optimality of progressive taxation (Theorem 1), and provide an intuitive discussion of its proof. We discuss its implications in relation to the literature in Section 3.2. Following this, Proposition 1 establishes a sense in which progressivity is necessary for optimality, and Proposition 2 provides conditions under which the optimal tax is strongly progressive.

3.1 Optimality of progressive taxation

In what follows, we use the term *progressive* to refer to a tax rule $T(y)$ which is convex. Theorem 1 establishes that there exists an optimal tax that is progressive, and that features non-decreasing tax payments and non-decreasing consumption. These latter two monotonicity properties are also satisfied by the optimal tax that arises in the canonical model of Mirrlees (1971). On the other hand, as we argue in Section 3.2, the convexity property is novel and does not arise in the Bayesian framework.

Theorem 1 (Progressivity)**.** *There exists a worst-case optimal T such that: $T(y)$ is non-decreasing and convex, and $y - T(y)$ is non-decreasing.*

The formal proof of the theorem is in Appendix A. In this section, we give an intuitive discussion of the proof by proceeding in two steps. First, we provide a simple characterization of the worst-case payoff of the planner (Lemma 1). Second, building on this characterization, we give a graphical description (Figure 3.1) of the other main part of the argument in the

proof of Theorem 1, which consists of constructing a welfare-improving convex version of an arbitrary tax function.

Characterizing the worst case. As a key step towards establishing the main result, Lemma 1 describes the worst-case payoff for the planner under a given tax rule $T \in \overline{\mathbf{T}}$, which we denote by

$$V_P(T) \equiv \inf_{M \in \mathcal{M}} V_P(T|M).$$

There, we show that the problem of finding the worst-case payoff across correspondences $M \in \mathcal{M}$ can be reduced to one of solving a simple constrained revenue minimization problem, and that this problem can be solved separately for each $\theta \in \Theta$.

Lemma 1 (Worst-case)**.** *For any feasible T,*

$$V_P(T) = \int_\Theta [\omega(\theta) W(V_w(T|M^0(\theta))) + \alpha \mathbb{E}_{F_\theta}[T(y)]] \, d\mu(\theta), \tag{3.1}$$

where F_θ is defined by one of the following two cases

(i) If $V_w(T|M^0(\theta)) < \max_{y \in Y}\{y - T(y)\}$:

$$F_\theta \in \arg\min_{F \in \Delta(Y)} \mathbb{E}_F[T(y)], \quad s.t. \quad \mathbb{E}_F[y - T(y)] \geq V_w(T|M^0(\theta)),$$

(ii) If $V_w(T|M^0(\theta)) = \max_{y \in Y}\{y - T(y)\}$:

$$F_\theta \in \arg\max_{(F,0) \in M^0(\theta)} \mathbb{E}_F[T(y)], \quad s.t. \quad \mathbb{E}_F[y - T(y)] = V_w(T|M^0(\theta)).$$

The proof is in Appendix A. According to the first term in (3.1), given the tax rule T, the utility of a worker with type θ in the worst case is equal to the lower bound on their equilibrium payoff as given by $V_w(T|M^0(\theta))$.[26] The intuition for this part of the result is quite simple: If workers' welfare was strictly higher than the welfare-lower-bound $V_w(T|M^0(\theta))$, it would be possible to increase their labor cost $\phi \in \mathbb{R}_+$ for every income lottery in their choice set. This constant shift strictly reduces workers' welfare without affecting their incentives to produce income, and hence overall reduces the planner's objective.

The second term in (3.1) describes worst-case revenue by distinguishing two cases. The first case, which, as we argue below, is the salient one, pertains to worker types who cannot attain maximal consumption at zero cost under T and $M^0(\theta)$. Loosely speaking, Lemma

[26] The fact that $M(\theta) \supseteq M^0(\theta)$ for all $\theta \in \Theta$ and $M \in \mathcal{M}$ implies that $V_w(T|M(\theta)) \geq V_w(T|M^0(\theta))$.

1 states that under the worst-case scenario, these workers' income choice (F, ϕ) minimizes their expected tax payment, with ϕ defined so that the worker is indifferent between choosing (F, ϕ) and choosing the best element in $M^0(\theta)$. This choice clearly gives a lower bound on the planner's revenue, given that it is always the case that workers' income choice satisfies $\mathbb{E}_F[y - T(y)] \geq V_w(T|M^0(\theta))$. In the proof, we construct a sequence of correspondences under which workers' income choice becomes arbitrarily close to (F, ϕ), and in that way, we show that this lower bound on the planner's payoff is attained in the limit by some plausible technology.[27]

The second case in Lemma 1 concerns workers whose equilibrium payoff is constant in $M(\theta)$, and equal to the highest possible utility that can be attained in the economy. Given that, when indifferent, workers make the income choice that maximizes tax revenue, the adversarial scenario for these types is attained when they face $M^0(\theta)$ since this minimizes their ability to break ties favorably. As a result, their worst-case income choice is $(F_\theta, 0) \in M^0(\theta)$, with F_θ as defined in the lemma. This case only arises when the consumption rule $y - T(y)$ has multiple maximizers. In Theorem 1, we show that the optimal consumption rule is monotone and concave, so on-path this case becomes relevant only if the tax rule involves full taxation of higher incomes and workers can produce high income at zero cost.

In summary, the worst-case correspondence is effectively $M(\theta) = M^0(\theta) \cup \{(F_\theta, 0)\}$ with F_θ defined as in Lemma 1, with the following two qualifications that we formally deal with in the proof: One, we need to modify $M(\theta)$ to ensure that it is plausible (i.e., continuous and monotone); and two, as mentioned above, $M(\theta)$ need not attain the worst-case payoff in (3.1), but, as we show, there is a perturbed version of it that approximates it arbitrarily well.

Lemma 1 reduces the objective of the planner in two important ways. First, we have transformed the problem of finding the worst-case payoff under T, which was originally a problem of finding the infimum across plausible correspondences, to a much simpler one which consists of finding, for every $\theta \in \Theta$, the distribution F_θ that satisfies the conditions in the Lemma. Second, there are no interactions across types, and therefore the worst-case scenario can be computed independently for every $\theta \in \Theta$. As a result, in the most complicated of the two cases (Case (i)), the worst-case can be computed by pointwise solving a simple prior-free constrained information design problem.[28]

[27] The reason why the lower bound is not necessarily attained stems from the fact that workers break ties in favor of the social planner. As a result, we need to perturb (F, ϕ) so as to ensure that the worker is strictly willing to choose it over choosing an element of $M^0(\theta)$.

[28] Le Treust and Tomala (2019) and Doval and Skreta (2023) develop techniques for solving this type of problem.

Construction of a welfare-improving convex tax. In the proof of Theorem 1, we consider an arbitrary feasible tax T, and use it to construct a convex tax $\overline{T}$ that satisfies $V_P(\overline{T}) \geq V_P(T)$. Here, we informally discuss this construction for the case with binary types. To that end, suppose that $\Theta = \{\underline{\theta}, \overline{\theta}\}$ with $\underline{\theta} < \overline{\theta}$. Suppose that the social planner uses a tax rule $T(y)$, with associated consumption rule $c(y) = y - T(y)$ as the one depicted in the left-panel of Figure 3.1. The shaded area in the figure consists of all the pairs $(\mathbb{E}_F[y], \mathbb{E}_F[c(y)])$ that can be attained by some income choice $F \in \Delta(Y)$. Formally, it is the convex hull of the graph of $c(y)$.

According to Lemma 1, in the worst case, a worker with type $\theta \in \{\underline{\theta}, \overline{\theta}\}$ makes an income choice that minimizes expected tax revenue subject to the constraint that expected consumption has to be weakly greater than $V_w(T|M^0(\theta))$.[29] It follows from well-known arguments (Aumann and Maschler, 1995; Kamenica and Gentzkow, 2011) that, for every $\theta \in \Theta$, the revenue minimizing pair $(\mathbb{E}_{F_\theta}[y], \mathbb{E}_{F_\theta}[c(y)])$ is a point in the graph of the upper concave envelope of $c(y)$, as depicted by the green line in the figure. If, as the example in the figure illustrates, the constraint in the revenue minimization problem binds, then the exact location of $(\mathbb{E}_{F_\theta}[y], \mathbb{E}_{F_\theta}[c(y)])$ is pinned down by the intersection between this upper concave envelope and the horizontal line passing through $V_w(T|M^0(\theta))$.

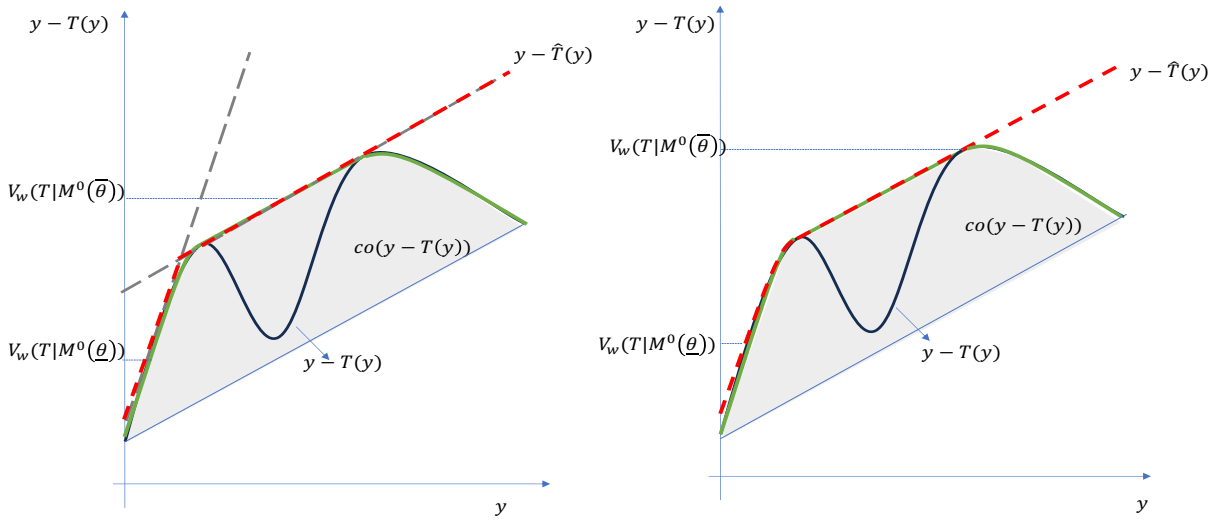

Figure 3.1: An arbitrary consumption rule (solid black line), and a concave consumption rule that dominates it (dashed red line) under binary types (left-panel) and under a continuum of types (right-panel).

Thus, by construction, for each $\theta \in \Theta$, $(\mathbb{E}_{F_\theta}[y], \mathbb{E}_{F_\theta}[c(y)])$ belongs to the boundary of the convex hull of the graph of $c(y)$. Building on the arguments in Carroll (2015), the

[29] In Figure 3.1, the consumption rule has a unique global maximizer, which implies that Case (ii) in Lemma 1 does not apply. Our formal proof of Theorem 1 accounts for this case when it arises.

supporting hyperplane to this set passing through the point $(\mathbb{E}_{F_\theta}[y], \mathbb{E}_{F_\theta}[c(y)])$ defines an affine consumption rule that leads to weakly higher revenue *conditional on type* θ. These θ-specific consumption rules are depicted by the dashed gray lines in the left-panel of Figure 3.1.

In order to construct a consumption rule that outperforms the original one, we seek to aggregate these two affine functions into a single one that outperforms c. We show that this goal is attained by taking the pointwise minimum between the two lines, as given by the red dashed function in the figure. The resulting consumption (tax) function is concave (convex), and gives weakly higher revenue than the starting point. Also, as the figure shows, the new consumption rule is pointwise higher than the original one, thereby improving workers' welfare at the same time. With more than two types, we apply an analogous argument and take the pointwise infimum over the family of θ-specific affine consumption rules that support the point $(\mathbb{E}_{F_\theta}[y], \mathbb{E}_{F_\theta}[c(y)])$. The right-panel in Figure 3.1 shows this construction for the case in which the type space is an interval.

3.2 Discussion and implications of Theorem 1

In this section, we discuss the intuition behind the optimality of tax progressivity and contrast this finding with the canonical optimal taxation model. The intuition behind the result can be conveyed through the graphical argument provided in Figure 3.1 for the model with binary types. The argument highlights two forces that point toward the optimality of a convex tax schedule. The first one stems from workers' (worst-case) incentives to make risky income choices. As illustrated in the picture, strict concavity in the tax schedule implies that, in the worst case, workers will take on income risk which drives down revenue. Intuitively, the concavity in T creates a wedge between the planner and the workers, by making the planner risk-averse (because tax revenue is concave) and the workers risk-loving (because after-tax income is convex). Further, whether or not a worker's income choice is random does not affect the worker-welfare component of the planner's objective because, by Lemma 1, this is pinned down by the lower bound on workers' payoffs $V_w(T|M^0(\theta))$. As a result, the planner should anticipate this adversarial income risk, by substituting T with a suitable convex version of it, as the one constructed in the proof of the theorem.

A similar logic is identified by Barron et al. (2020) in a Bayesian moral hazard model. In their model, the production technology is known by the principal, and workers are assumed to be able to garble any output realization at zero cost. The latter feature gives rise to a "no gaming" constraint in the principal's problem, whereby the principal should anticipate the agent's gambles by concavifying the contract that she offers. In our model, uncertainty is

considerably richer, which implies that the principal needs to hedge against other ways—not necessarily involving mean-preserving spreads—in which the agent may alter the distribution of output.[30]

The second force stems from heterogeneity in workers' productivity. As shown in Figure 3.1, the slope of the dominating type-specific affine tax is increasing in type. Moreover, in the worst case, a more productive worker makes an income choice that leads to higher expected income. Thus, by offering a convex tax schedule, the planner is able to target more productive workers with higher marginal tax rates which, as was just argued, is beneficial for revenue.

The above intuition was framed in terms of optimal tax revenue. The reason is that most of the worst-case forces act through the revenue component of the planner's objective: According to Lemma 1, the worker-welfare component is fixed at the value that arises under the baseline technology. However, as Theorem 1 demonstrates, any non-progressive tax rule can be substituted by a progressive one that simultaneously improves *workers' welfare and revenue*, so there is actually no tension between the two objectives of the planner in the argument. Crucially, our arguments hold regardless of the planner's preference for redistribution across workers, and of how she values workers' welfare relative to tax revenue. It also does not rely on any assumptions in the distribution of workers' types μ.

Welfare-improving tax reform. By the above logic, starting from any non-progressive tax, we can use a construction similar to the one in the proof of Theorem 1 to propose a tax reform that leads to a weak improvement in both (worst-case) workers' welfare and tax revenue. Furthermore, this welfare-improving tax reform is *independent of the details of the environment*, such as the distribution μ and the choice of the baseline technology $M^0(\theta)$.

To state the result formally, let $T(y)$ be the prevailing tax system and put $c(y) = y - T(y)$. Let $\tilde{c}(y)$ be the concavification of $c(y)$—i.e., $\tilde{c}$ is the smallest concave function that majorizes c[31]—and put $\tilde{T}(y) = y - \tilde{c}(y)$. By construction, $\tilde{T}(y)$ is a progressive—i.e., convex—tax. We refer to the policy of substituting $T(y)$ with $\tilde{T}(y)$ as a *basic progressive tax reform of* T.

Theorem 1* (Tax reform)**.** *For any feasible tax T, the basic progressive tax reform of T leads to a weak improvement in both (worst-case) workers' welfare and revenue.*

The proof is in Appendix A. Theorem 1* offers a simple, detail-free tax policy that improves upon any non-progressive tax system. We refer to it as a *basic* reform because

[30] In their Online Appendix, Walton and Carroll (2022) examine the case in which the principal's robust concern only involves the agent's garbling of output, and show that this setting leads to the optimality of a concave contract as well.

[31] Formally, $\tilde{c}(y) = \sup\{c : (y, c) \in \text{co}(C)\}$, where $C = \{(y, c(y)) : y \in Y\}$ and $\text{co}(C)$ is the convex hull of C. Graphically, for a given c, $\tilde{c}$ is depicted as the green line in Figure 3.1.

it is, in a sense, a minimal way in which the government can substitute the prevailing tax policy with one that leads to better worst-case outcomes in all dimensions. An alternative, detail-free tax reform that outperforms the basic one is to substitute $c(y)$ by its *monotone concavification*—i.e., the smallest non-decreasing concave function that majorizes c. Still, it is worth noting that the distinction between these two alternative tax reforms is relevant only in the case in which the prevailing $c(y)$ is non-monotone, which is an implausible scenario.[32]

In terms of our graphical argument in Figure 3.1, the basic progressive tax reform of T entails substituting $c(y)$ with the function depicted as the green line. By the arguments provided there, it is possible to further improve workers' welfare and revenue by instead using the alternative consumption rule that we construct in the proof of Theorem 1 (depicted by the dashed red line), which incorporates the information provided by the knowledge of the baseline technology M^0.

Relationship to results in canonical Mirrlees model. We now contrast our finding with the most salient results about tax progressivity in the standard Mirrleesian framework. If the distribution of workers' productivity is assumed to be bounded, then it is well-known that the optimal tax is flat at the top of the income distribution (Sadka, 1976; Seade, 1977). This result follows from a standard no-distortion-at-the-top-argument which we illustrate in Figure 3.2 below. There, $c(y)$ represents an after-tax-income schedule in which the marginal tax rate faced by the highest income produced in the economy ($\hat{y}$) is strictly positive. By moving to a tax schedule such as the one depicted by $\tilde{c}(y)$, where the tax rate above income $\tilde{y}$ is strictly lower than that under $c(y)$, the planner is able to strictly improve welfare for a positive measure of types and raise more tax revenue. Consequently, the tax schedule associated to $c(y)$ cannot be optimal.

The reason why this argument fails in our model is that the perturbation of $c(y)$ that we just described does not preserve concavity of consumption. To elaborate, in our setting, if the tax schedule associated with $\tilde{c}(y)$ were implemented, then in the worst case, the income choice of workers on the higher end of the income distribution would be random as depicted in the figure, and thus will typically not lead to an improvement in revenue. This observation follows from the logic highlighted above: Concavity in the tax schedule implies that, under the worst-case technology, workers find it optimal to take on income risk at a very low labor disutility cost, which in turn drives down expected revenue. Thus, by the argument provided in Theorem 1, $\tilde{c}(y)$ is itself dominated by a progressive tax such as the one illustrated by $\hat{c}(y)$.

Subsequent work by Diamond (1998) and Saez (2001) has challenged the zero-marginal-

[32] We are not aware of any existing tax systems where after-tax income is non-monotone in pre-tax income.

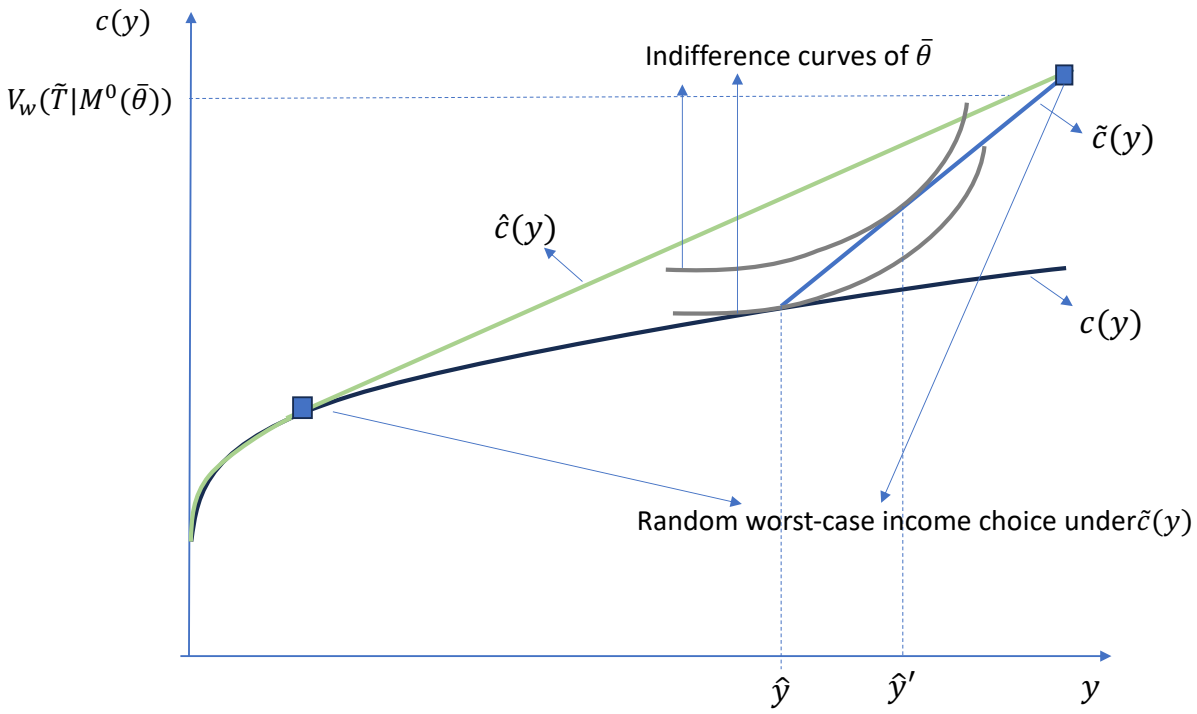

Figure 3.2: A classical no distortion at the top argument, and its unraveling in our setting.

rate-at-the-top result. By assuming that the distribution of workers' productivity is unbounded and that its right tail follows a Pareto distribution, they show that the optimal top rate is strictly positive. The result also relies on the assumption that the social planner is inequality-averse and that the marginal value that she places on the welfare of top earners is negligible. Our result relies on no such assumptions, and instead is a consequence of the robust approach to uncertainty about workers' income choices. Moreover, in those models, the entire tax schedule is typically *not* progressive, with regions in which the marginal tax rate is strictly decreasing in income. Again, when introduced in our environment, such a non-convex tax rule would induce worst-case income risk, and would thereby be dominated by a suitable convex version of it.

3.3 Is progressivity necessary?

Given that the optimal tax need not be unique, we now turn to the question of whether or not tax progressivity is necessary for optimality. The answer to this question depends on the choice of the relevant set of income realizations over which convexity of T is required. We show in Proposition 1 that, at the optimum, T must be convex over the range of incomes that may arise in equilibrium, in a sense that we make precise below. To that extent, the proposition offers an endogenous (T-dependent) range of incomes over which tax progres-

sivity must be satisfied at the optimum. Using this result, Corollary 1 provides a condition on the primitives of the model that ensures that this endogenous range coincides with the full income domain Y. Further, in Section 4, we study the special case in which M^0 is defined similarly to the Mirrlees economy from Example 1. There, Corollary 2 states a very different assumption that, together with the Mirrleesian assumption on M^0, implies that tax progressivity is necessary for optimality.

For any feasible T, let $\mathcal{F}^0_\theta(T) \subseteq \Delta(Y)$ be the set of optimal income choices for type θ under T and $M^0(\theta)$ subject to favorable tie-breaking, i.e.,

$$\mathcal{F}^0_\theta(T) \equiv \underset{F:(F,\phi)\in M^0(\theta), \mathbb{E}_F[y-T(y)]-\phi=V_w(T|M^0(\theta))}{\arg\max} \mathbb{E}_F[T(y)],$$

and let $\underline{\mathcal{F}}_\theta(T) \subseteq \Delta(Y)$ be their income choices under the worst-case scenario as defined by F_θ in Lemma 1. Let

$$\mathcal{F}^0(T) \equiv \left\{ \int_\Theta F^0_\theta \, d\mu(\theta) : F^0_\theta \in \mathcal{F}^0_\theta(T), \forall \theta \in \Theta \right\} \subseteq \Delta(Y),$$

$$\underline{\mathcal{F}}(T) \equiv \left\{ \int_\Theta \underline{F}_\theta \, d\mu(\theta) : \underline{F}_\theta \in \underline{\mathcal{F}}_\theta(T), \forall \theta \in \Theta \right\} \subseteq \Delta(Y)$$

be the sets of aggregate income distributions that may arise under, respectively, M^0 and the worst-case scenario. Observe that all of the elements in $\mathcal{F}^0(T)$ are outcome-equivalent, in that they lead to the same distribution over worker-welfare and revenue, and the same is true of the elements of $\underline{\mathcal{F}}(T)$. Let $c(y) = y - T(y)$ be the consumption rule associated with $T(y)$, let $\overline{c}(y)$ be defined as in (A.9)—i.e., the dominating concave consumption rule depicted in Figure 3.1—, and let $\tilde{c}(y)$ be the concavification of $c(y)$.

Proposition 1 (Necessity)**.** *For any worst-case optimal tax rule T and any $F \in \mathcal{F}^0(T) \cup \underline{\mathcal{F}}(T)$, it holds that $c(y) = \tilde{c}(y) = \overline{c}(y)$ F-almost everywhere.*

The proof is in Appendix A. As a consequence of Proposition 1 and the definition of $\tilde{c}$, any optimal tax rule must be convex in a neighborhood of every $y \in (0, \overline{y})$ that may arise with positive probability under either the baseline or the worst-case technology. The following Corollary shows that this set of income realizations covers the entire range of income possibilities Y if there is a positive measure of types for whom $M^0(\theta)$ satisfies the following full-support condition: Say that $M \subseteq \Delta(Y) \times \mathbb{R}_+$ satisfies the *full-support condition* if for all $(F, \phi) \in M$, F has full support on Y.[33]

[33]The full-support condition ensures that the optimal contract has to be linear in Carroll (2015).

Corollary 1 (Full support)**.** *If $M^0(\theta)$ satisfies the full-support condition for a positive measure of θ, then any worst-case optimal tax rule is convex on Y.*

Proof. The result follows from Proposition 1 and the fact that, if the full-support condition holds for a positive measure of θ, then $\text{supp}(F^0) = Y$ for all $F^0 \in \mathcal{F}^0(T)$ and every feasible T. □

Intuitively, if there is a positive measure of types θ who can only make full-support income choices under the baseline M^0 and if $\overline{c}(y) > c(y)$ for a non-degenerate interval of incomes, then $V_w(\overline{c}|M^0(\theta)) > V_w(c|M^0(\theta))$ for any such θ. By the arguments given in the proof of Theorem 1, $\overline{c}$ allows the planner to raise weakly higher revenue than c while yielding a strict worker-welfare improvement over c, and therefore c cannot be optimal. In Section 4 we provide a different condition which, conversely, holds when workers make deterministic income choices at the baseline. In any case, the two conditions that we provide should be seen as examples: Proposition 1 allows to devise different assumptions on the primitives (some more natural than others) to ensure that any $y \in Y$ may arise with positive probability in equilibrium, and thus that any optimal tax rule must be convex on Y.

3.4 When is the optimal policy strongly progressive?

In this section, we provide conditions under which the optimal tax rule (within the convex class) is strongly progressive, in the sense of using a marginal tax rate at the top of the income distribution that is strictly higher than that at the bottom. Formally, for any convex tax rule T, we let $T'(y-)$ and $T'(y+)$ denote the left- and right-directional derivatives. We define strong progressivity as follows.

Definition 2 (Strong progressivity)**.** A tax rule $T(y)$ is strongly progressive if it is convex and $T'(0+) < T'(\overline{y}-)$.

Henceforth, we maintain the following assumption, which states that the social planner is weakly inequality-averse.[34]

Assumption 3 (Inequality-averse planner)**.** $\omega(\cdot)$ is non-increasing and $W(\cdot)$ is concave.

Let $\mathcal{F}^0_\theta(\tau) \subseteq \Delta(Y)$ be the set of type-θ's income choices under $M^0(\theta)$ and an affine tax rule with slope τ, and let $F^0_\theta(\tau) \in \mathcal{F}^0_\theta(\tau)$ and $y^0_\theta(\tau) = \mathbb{E}_{F^0_\theta(\tau)}[y]$.[35] Let $\phi^0_\theta(\tau)$ be such that

[34] The only role that Assumption 3 plays in Proposition 2 is through the concavity of W, which ensures that second-order conditions are satisfied in the problem of finding the optimal affine tax rule. The assumption on $\omega(\cdot)$ is not needed here, but is used later in Section 4 so for concreteness we state them together.

[35] This choice is independent of the intercept of T.

$(F^0_\theta(\tau), \phi^0_\theta(\tau))$ is optimal under $M^0(\theta)$ and τ. Let $w_\theta(t,\tau) = \omega(\theta)W'(-t+(1-\tau)y^0_\theta(\tau)-\phi^0_\theta(\tau))$ which is the endogenous Pareto weight of type θ under an affine tax rule. Proposition 2 describes the worst-case optimal tax rule within the affine class, and provides a sufficient condition under which this tax rule is strictly suboptimal. For simplicity, we drop the dependence of the above objects with respect to (t,τ), which are understood to be evaluated at their optimal values (t^*,τ^*).

Proposition 2 (Affine tax)**.** *The tax rule $T_a(y) = t^* + \tau^* y$ where (t^*,τ^*) satisfy*

$$\int_\Theta (w_\theta - \alpha)\, d\mu(\theta) \leq 0, \quad \textit{with equality if } t^* < 0, \tag{3.2}$$

$$\int_\Theta (w_\theta - \alpha) y^0_\theta \, d\mu(\theta) \leq 0 \textit{ and } \int_\Theta \phi^0_\theta \, d\mu(\theta) = 0, \quad \textit{if } \tau^* = 1, \tag{3.3}$$

$$\tau^* = \max\left\{1 - \sqrt{\frac{\alpha \int_\Theta \phi^0_\theta \, d\mu(\theta)}{\int_\Theta (\alpha - w_\theta) y^0_\theta \, d\mu(\theta)}}, 0\right\}, \quad \textit{if } \tau^* < 1. \tag{3.4}$$

is worst-case optimal within the class of affine tax rules.

If $\tau^* < 1$ and, for some $\tilde{y} \in Y$, it holds that

$$\int_\Theta \left[\left(w_\theta + \alpha \frac{\tau^*}{1-\tau^*}\right)\mathbb{E}_{F^0_\theta}[\min\{\tilde{y} - y, 0\}] - \alpha \mathbf{I}(y^0_\theta > \tilde{y}) \frac{\tilde{y} - y^0_\theta + \phi^0_\theta/(1-\tau^*)}{1-\tau^*}\right] d\mu(\theta) > 0, \tag{3.5}$$

then there exists a strongly progressive tax rule that yields strictly higher worst-case welfare.

The proof is in Appendix A. We begin by discussing the features of the worst-case optimal affine tax. Condition (3.2), which pins down the lump-sum payment $t^* = T(0)$, is intuitive: If $t^* < 0$ (workers receive a positive lump-sum payment), then the transfer amount has to be such that the marginal benefit of redistributing an extra dollar—which is given by the average social welfare weight—is equal to the marginal cost—which is represented by the marginal value of public funds, α. Otherwise, the non-negativity constraint on consumption is binding and optimal lump-sum payments are equal to zero. An analogous condition arises in the canonical model if we substitute α for the Lagrange multiplier associated with the government's budget constraint.

Next, consider the optimal marginal rate τ^*. For conciseness, we focus on discussing the most interesting case in which $\tau^* \in (0,1)$. Applying Lemma 1, under an an affine tax rule with rate $\tau \in (0,1)$ and intercept t, the worst-case income choice of a worker with type θ,

which we denote by $y_\theta(\tau) \equiv \mathbb{E}_{F_\theta}[y]$, is such that

$$V_w(T|M^0(\theta)) = (1-\tau)y^0_\theta(\tau) - t - \phi^0_\theta(\tau) = (1-\tau)y_\theta(\tau) - t \iff y_\theta(\tau) = y^0_\theta(\tau) - \frac{\phi^0_\theta(\tau)}{1-\tau},$$

and worst-case consumption is equal to $V_w(T|M^0(\theta))$.

Following the discussion in Saez (2001), we can informally decompose the first-order effect of changing the tax rate into two parts. First, the *mechanical* effect captures the fact that, keeping workers' income choice fixed, their worst-case consumption and tax payment changes proportionally to $y^0_\theta(\tau)$. This results in the following mechanical effect on total welfare

$$\int_\Theta (\alpha - w_\theta(t,\tau)) y^0_\theta(\tau)\, d\mu(\theta).$$

Second, there is a *worst-case behavioral* response, whereby workers' worst-case income choice is affected by the change in τ. According to the definition of $y_\theta(\tau)$, the first-order effect on type θ's income production is equal to $-\phi^0_\theta(\tau)/(1-\tau)^2$.[36] Thus, the value of the change in income production due to a marginal increase in the tax rate is equal to

$$-\alpha \int_\Theta \frac{\phi^0_\theta(\tau)}{(1-\tau)^2}\, d\mu(\theta).$$

Setting the sum of the two effects equal to zero gives the tax rate in (3.4).

We note that the expression for the worst-case behavioral response is qualitatively different—i.e., depends on different primitives—compared to its analog in the canonical framework studied by Saez (2001). Contrary to that setting, the worst-case approach taken here implies that the production technology is not kept fixed as we change the tax rate, and therefore the behavioral effect is not described in terms of the labor supply elasticity. Instead, in our model, the (worst-case) behavioral response is governed by the *level* of workers' labor disutility under the baseline technology. Intuitively, $\phi^0_\theta(\tau)$ captures how much scope there is for adversarially lowering revenue by adding a new low-cost income choice to the worker's baseline choice set $M^0(\theta)$.

By applying a similar logic, we can derive condition (3.5), which rules out the possibility that the affine tax in Proposition 2 is fully optimal. For the sake of exposition, suppose momentarily that workers' optimal income choice under T_a and $M^0(\theta)$—i.e., $\mathcal{F}^0_\theta(\tau^*)$—is single-valued and deterministic for all $\theta \in \Theta$.[37] Consider the perturbation around $T_a(y)$

[36] We show in the proof of Proposition 2 that the behavioral effect through a change in $(y^0_\theta(\tau), \phi^0_\theta(\tau))$ is of second order.

[37] The Mirrlees economy studied in the next section conforms to this assumption. Our formal proof of Proposition 2 does not rely on this assumption.

given by $T^{\varepsilon}(y) \equiv T_a(y) + \varepsilon \max\{y - \tilde{y}, 0\}$, with $\tilde{y} \in (\inf_{\theta\in\Theta} y_\theta^0(\tau^*), \sup_{\theta\in\Theta} y_\theta^0(\tau^*))$ and $\varepsilon > 0$. $T^{\varepsilon}(y)$ is a strongly progressive perturbation that involves increasing the tax rate by ε above the income level $\tilde{y}$. We show in Appendix B that the planner's objective is directionally differentiable with respect to ε and that the marginal effect on total welfare as $\varepsilon \downarrow 0$ is given by the left-hand-side of (3.5). Under the simplifying assumptions made here, that condition reduces to

$$\int_\Theta \mathbb{I}(y_\theta^0(\tau^*) > \tilde{y}) \left((\alpha - w_\theta(t^*, \tau^*))(y_\theta^0(\tau^*) - \tilde{y}) - \alpha \frac{\phi_\theta^0(\tau^*)}{(1-\tau^*)^2} \right) d\mu(\theta) > 0. \tag{3.6}$$

Intuitively, an increase in the tax rate above $\tilde{y}$ only affects workers whose baseline income $y_\theta^0(\tau^*)$ is above this cutoff. Within that region of the income distribution, a very similar intuition to the affine case yields that the mechanical welfare effect is

$$\Pr(y_\theta^0(\tau^*) > \tilde{y}) \mathbb{E}[(\alpha - w_\theta(t^*, \tau^*))(y_\theta^0(\tau^*) - \tilde{y}) | y_\theta^0(\tau^*) > \tilde{y}],$$

and the worst-case behavioral one is

$$-\alpha \Pr(y_\theta^0(\tau^*) > \tilde{y}) \mathbb{E}[\phi_\theta^0(\tau^*)/(1-\tau^*)^2 | y_\theta^0(\tau^*) > \tilde{y}].$$

Thus, if condition (3.6) is satisfied, $T^{\varepsilon}(y)$ is a strongly progressive tax that yields a strict improvement over $T_a(y)$.

Recall from our discussion of the proof of Theorem 1 that an affine tax is optimal in the benchmark with a single type. Thus, a minimal amount of heterogeneity with respect to θ is required for (3.5) to hold. Indeed, the definition of τ^* ensures that (3.5) is never satisfied whenever μ is degenerate.

To further illustrate the intuition behind (3.6), consider the special case in which $M^0(\theta)$ is defined as in Example 1, in which case $y_\theta^0(\tau^*)$ is increasing in θ. Then, condition (3.6) is always satisfied if τ^* is bounded away from 1, and there exists a cutoff $\tilde{\theta}$ such that: $\mathbb{E}[\phi_\theta^0(\tau^*) | \theta > \tilde{\theta}]$ is arbitrarily close to zero, and $\mathbb{E}[\alpha - w_\theta(t^*, \tau^*) | \theta > \tilde{\theta}] > 0$—i.e. if the cost of income production is arbitrarily low for the most skilled workers in the economy and the government's objective assigns relatively low social welfare weight to these workers. The first condition holds precisely when the (worst-case) income elasticity of the most skilled workers is extremely low. In this case, the planner can tax these workers heavily while incentivizing them to produce high income. This feature is consistent with our intuition in Section 3.2 that increasingness (with respect to θ) in the slope of the revenue-maximizing type-specific affine tax pushes toward *strict* convexity of the optimal tax.

A natural question is whether the negation of (3.5) is sufficient to ensure that an affine tax rule is optimal. Unfortunately, the perturbation argument that we provide in the proof does not enable us to derive sufficient conditions for the optimality of $T_a(y)$. The reason is that the planner's objective is neither differentiable with respect to all perturbations nor concave in T. Interestingly, in the binary-type model that we study in Section 4.1, condition (3.5) is indeed necessary and sufficient to rule out optimality of an affine tax.

4 Application: Mirrlees economy

In this section, we study the special case of our model in which the baseline technology is described by a *Mirrlees economy*, as defined in Example 1. To that end, we maintain Assumptions 2 and 3, and substitute Assumption 1 with the following stronger requirement:

Assumption 4 (Mirrlees economy)**.** $M^0(\theta) = \{(\delta_{\{y\}}, \phi) : y \in Y, \phi \in [\Phi(y,\theta), \Phi(\overline{y},\theta)]\}$, where $\Phi : Y \times \Theta \to \mathbb{R}_+$ is continuously differentiable, strictly increasing and strictly convex in y, and satisfies strict single crossing in (y,θ). Moreover, $\lim_{y\to\overline{y}} \Phi_y(y,\underline{\theta}) > 1$, and for all $\theta \in \Theta$, $\Phi(0,\theta) = 0$ and $\lim_{y\to 0} \Phi_y(y,\theta) = 0$.

$M^0(\theta)$ as defined in Assumption 4 satisfies Assumption 1, and introduces some additional restrictions. First, workers' baseline income choices are restricted to be deterministic. Second, single-crossing of Φ implies that, in equilibrium, income choices under M^0 are non-decreasing in θ. Third, strict convexity of $\Phi(\cdot,\theta)$ implies that the worker's optimal income choice when facing $M^0(\theta)$ and any convex tax rule is unique. Fourth, the assumption that $\lim_{y\to 0} \Phi_y(y,\theta) = 0$ implies *no bunching at zero*: So long as $T'(0+) < 1$, all workers are willing to choose $y > 0$ in the baseline. This is a standard assumption that simplifies the analysis by allowing us to focus on labor supply responses to taxes only through the intensive margin—i.e., how much labor to provide. And fifth, the assumption that $\lim_{y\to\overline{y}} \Phi_y(y,\underline{\theta}) > 1$ implies that the upper bound on income $\overline{y}$ does not bind for the lowest type. Observe that the possibility of an increasing type-dependent upper bound on how much income workers can earn, as the one that arises in Example 1, is not ruled out by Assumption 4: It can be captured through additional functional form assumptions on Φ.

Let $y^0_\theta(T)$ be type θ's optimal income choice under T and $M^0(\theta)$ (applying favorable tie-breaking when needed). Let $y_\theta(T) = \min\{y \in Y : y - T(y) = V_w(T|M^0(\theta))\}$, which is well-defined given continuity of $T(y)$, and the fact that for all $\theta \in \Theta$, $-T(0) \leq V_w(T|M^0(\theta))$ and $\max_{y\in Y}\{y - T(y)\} \geq V_w(T|M^0(\theta))$. As before, put $w_\theta(T) = \omega(\theta)W'(V_w(T|M^0(\theta)))$. In what follows, whenever it is clear, we ignore the dependence of these objects on T. Observe

that, under our assumptions, $y_\theta(T)$ is continuous and non-decreasing in θ, and so is $y^0_\theta(T)$ whenever it is single-valued (which always holds if T is convex).

As anticipated in Section 3.3, we begin by providing conditions under which any optimal tax rule must be progressive in this environment. Let y^0_θ be any selection from $y^0_\theta(T)$. The result below is a corollary of Proposition 1.

Corollary 2 (Necessity of progressivity - Mirrlees)**.** *If Assumption 4 holds and if μ is atomless with full support on $[\underline{\theta}, \overline{\theta}]$, then any worst-case optimal tax rule must be convex on $[y^0_{\underline{\theta}}, y^0_{\overline{\theta}}]$.*

Proof. Under Assumption 4, for any feasible tax rule, workers' income choice under $M^0(\theta)$ is single-valued except for at most countably many θs. Thus, y^0_θ is almost everywhere continuous and non-decreasing in θ. Under the assumption that μ has full support on $[\underline{\theta}, \overline{\theta}]$, for any $F \in \mathcal{F}^0(T)$, it holds that $[y^0_{\underline{\theta}}, y^0_{\overline{\theta}}] \subseteq \text{supp}(F)$, and thus the result follows from Proposition 1. □

The conditions that we provide in Corollary 2, ensuring that any optimal tax rule has to be convex in the relevant domain, are qualitatively different than those in Corollary 1: Corollary 2 requires that heterogeneity across workers is rich and does *not* require that workers face income risk at the baseline, whereas the reverse is true for the conditions given in Corollary 1. In both cases, the conditions ensure that the set of income choices that may arise with positive probability in equilibrium under the baseline M^0 is sufficiently rich.

We now turn to further characterize the optimal tax rule. Appealing to Theorem 1, from now on, we restrict attention to tax rules that are convex, non-decreasing, and that lead to non-decreasing consumption. As mentioned above, under convexity of T, workers' baseline income choice $y^0_\theta(T)$ is unique for all $\theta \in \Theta$. Our next result characterizes the optimal tax schedule at the bottom of the income distribution.

Proposition 3 (Lump-sum transfer and bottom rate)**.** *There exists a worst-case optimal T such that $T(0) \in \mathbb{R}_-$ is determined by*

$$\int_\Theta w_\theta \, d\mu(\theta) - \alpha \leq 0, \quad \text{with equality if } T(0) < 0. \tag{4.1}$$

Moreover, there exists a cutoff $y_l \in Y$ with $y_l > y_{\underline{\theta}} > 0$ such that T is affine on $[0, y_l]$, and the marginal tax rate ($\underline{\tau}$) below y_l satisfies:

(i) If $\int_\Theta w_\theta \, d\mu(\theta) - \alpha = 0$, then $\underline{\tau} = 0$.

(ii) Otherwise,

$$\underline{\tau} \in \left[\max\left\{1 - \sqrt{\frac{\alpha \int_\Theta \mathbb{I}(y_\theta \leq y_l)\Phi(y_\theta^0, \theta)\, d\mu(\theta)}{\int_\Theta (\alpha - w_\theta) \min\{y_\theta^0, y_l\}\, d\mu(\theta)}}, 0\right\}, \max\left\{1 - \sqrt{\frac{\alpha \int_\Theta \Phi(\theta, y_\theta^0)\, d\mu(\theta)}{\int_\Theta (\alpha - w_\theta) y_\theta^0\, d\mu(\theta)}}, 0\right\}\right]. \tag{4.2}$$

The proof, which is in Appendix A, relies on using perturbations around an optimal T, and applying our results on the directional-differentiability of $V_P(T)$ with respect to these perturbations which we establish in Appendix B. The first part of Proposition 3 describes the optimal lump-sum transfer to workers. At the optimum, this transfer is defined analogously to the affine tax from Proposition 2, by balancing the marginal gains and costs from transferring an extra unit of income to every worker while still satisfying non-negativity of consumption.

The second part characterizes the optimal tax rate at the bottom of the income distribution. To describe this part of the result, we first observe that our construction of a dominating tax rule in the proof of Theorem 1 (see also Figure 3.1) implied that it is without loss of optimality to restrict attention to tax rules that have a constant marginal rate below $y_{\underline{\theta}}$. In the proof of Proposition 3 we further establish that in fact, this feature holds over the larger range given by $[0, y_l]$ with $y_l > y_{\underline{\theta}}$. We also show that the marginal tax rate is strictly less than one on $[y_{\underline{\theta}}, y_{\overline{\theta}}]$ which in turn implies, under our assumption of no-bunching at zero income, that $y_{\underline{\theta}} > 0$.

We show that, if $\int_\Theta w_\theta\, d\mu(\theta) - \alpha = 0$, then the optimal bottom tax rate is equal to zero. A result analogous to this one holds in the canonical framework as well. In Mirrlees (1971), the budget constraint always binds and therefore the average social welfare weight is equal to the (endogenous) marginal value of public funds, and therefore the analog of (4.1) holds with equality in that model. In the absence of extensive-margin labor supply responses, which is a consequence of the assumption of no bunching at zero income in Assumption 4, then a zero-taxation-at-the-bottom result obtains in that model (Seade, 1977). Thus, our model replicates the finding of zero taxation of low incomes, while challenging the less reasonable one of zero taxation of top incomes.

In the complementary case in which $\int_\Theta w_\theta\, d\mu(\theta) - \alpha < 0$, the planner may benefit from taxing workers at the bottom of the income distribution. The proof in the Appendix derives the expression for the optimal rate $\underline{\tau}$, which depends on the endogenous tax rates used above y_l (see equation (A.25)). This dependence holds because the tax rate used for lower incomes affects the tax payment over the entire income distribution. Absent a closed-form solution for the entire tax schedule, we use this expression to derive the bounds on $\underline{\tau}$ in Proposition

3.[38] Observe that, if $y_l = y^0_{\overline{\theta}}$, which is the case in which the entire tax schedule is affine, then the lower and upper bounds in (4.2) coincide, and we obtain that $\underline{\tau} = \tau^*$ with τ^* being the tax rate under the optimal affine tax rule from Proposition 2. In the case in which $y_l < y^0_{\overline{\theta}}$ (the optimal tax is strongly progressive), there is a gap between the two bounds. In fact, we show in Section 4.1 that, when the type space is binary, either: The optimal tax rule is affine and $\underline{\tau}$ is equal to the upper bound in (4.2), or it is strongly progressive and $\underline{\tau}$ is equal to the lower bound in (4.2). In this sense, the bounds in Proposition 3 are tight.

We now turn to characterizing the optimal tax rate at the top of the income distribution. An identical argument to the one in Proposition 3, invoking the proof of Theorem 1, implies that there is an optimal tax rule which is affine above $y_{\overline{\theta}} < y^0_{\overline{\theta}}$. Proposition 4 provides a characterization of the constant tax rate above this cutoff, which uses the following bounds as a function of $y \in Y$

$$\overline{\tau}_*(y) \equiv 1 - \sqrt{\frac{\alpha \int_\Theta \mathbb{I}(y_\theta > y)\Phi(y^0_\theta, \theta)\, d\mu(\theta)}{\int_\Theta \mathbb{I}(y_\theta > y)(\alpha - w_\theta)(y^0_\theta - y^0_{\theta^*(y)})\, d\mu(\theta)}}, \tag{4.3}$$

$$\overline{\tau}^*(y) \equiv \max\left\{1 - \sqrt{\frac{\alpha \int_\Theta \mathbb{I}(y_\theta > y)\Phi(y^0_\theta, \theta)\, d\mu(\theta)}{\int_\Theta \mathbb{I}(y^0_\theta > y)(\alpha - w_\theta)(y^0_\theta - y)\, d\mu(\theta)}}, 0\right\}, \tag{4.4}$$

where $\theta^*(y) \equiv \min(\{\theta \in \Theta : y_\theta = y\} \cup \{\underline{\theta}\})$.

Proposition 4 (Top rate)**.** *There exists a worst-case optimal T and a cutoff $y_u \in Y$ with $y_u < y^0_{\overline{\theta}}$ such that T is affine on $[y_u, \overline{y}]$ with constant rate $\overline{\tau}$. Moreover, if $\underline{\tau} < \overline{\tau}$, then $\overline{\tau} \in [\overline{\tau}_*(y_u), \overline{\tau}^*(y_u)]$*

The proof is in Appendix A. Proposition 4 provides bounds for the optimal top tax rate for the case in which the optimal policy is strongly progressive (otherwise, the optimal constant tax rate is characterized by Proposition 2). As before, as $y_u \downarrow 0$, the interval $[\overline{\tau}_*(y_u), \overline{\tau}^*(y_u)]$ degenerates into the optimal constant tax rate from Proposition 2. Moreover, in the binary-type case that we study next, the two endpoints coincide, and thus Proposition 4 exactly pins down the top tax rate.

[38] To see that these bounds are well-defined, note that, the fact that $\int_\Theta w_\theta\, d\mu(\theta) - \alpha < 0$, and that both $(\alpha - w_\theta)$ and $\min\{y^0_\theta, y_l\}$ are non-decreasing in θ implies that

$$\int_\Theta (\alpha - w_\theta) y^0_\theta\, d\mu(\theta) \geq \int_\Theta (\alpha - w_\theta) \min\{y^0_\theta, y_l\}\, d\mu(\theta) \geq \int_\Theta (\alpha - w_\theta)\, d\mu(\theta) \int_\Theta \min\{y^0_\theta, y_l\}\, d\mu(\theta) > 0.$$

4.1 Binary types

In this section, we further specialize the model to the case in which the type space is binary. Although very stylized, this version of the model allows us to derive closed-form solutions for the optimal tax schedule and to perform comparative statics on the model's primitives. The analysis for any finite-type space is similar. A similar analysis would also apply if the type space were continuous but the planner was restricted to using a piecewise affine tax rule with *finitely* many income brackets (as is always observed in practice).

Suppose that $\Theta = \{\theta_1, \theta_2\}$ with $\theta_1 < \theta_2$. We refer to type θ_1 as the 'low type' and to θ_2 as the 'high type'. We write $p_j = \Pr(\theta = \theta_j)$, and assume that $p_j \in (0,1)$ for all $j = 1, 2$. For simplicity, in this section, we focus on the case with exogenous social welfare weights, by setting $W'(c) = 1$ for all $c \in \mathbb{R}$.[39] In this setting, Assumption 2 reduces to $\mathbb{E}[\omega(\theta)] \leq \alpha$. For any $\tau \in [0,1]$, let $y_i^0(\tau)$ be the optimal income choice of type θ_i under $M^0(\theta)$ and a constant tax rate equal to τ.

Let τ^* be the tax rate under the optimal affine tax rule, as described in Proposition 2. Under the assumptions in this section, we show in Proposition 5 that τ^* is implicitly defined by

$$\tau^* = \begin{cases} 0, & \text{if } \mathbb{E}[\omega(\theta)] = \alpha, \\ \max\left\{0, 1 - \sqrt{\frac{\alpha[p_1\Phi(y_1^0(\tau^*),\theta_1)+p_2\Phi(y_2^0(\tau^*),\theta_2)]}{p_1(\alpha-\omega(\theta_1))y_1^0(\tau^*)+p_2(\alpha-\omega(\theta_2))y_2^0(\tau^*)}}\right\}, & \text{if } \mathbb{E}[\omega(\theta)] < \alpha. \end{cases} \tag{4.5}$$

As depicted in Figure 3.1 (and formally shown in the proof of Proposition 5), when the type space is binary, it is without loss of optimality to focus on tax rules that are piecewise affine with at most one kink. Formally, we focus on tax rules of the form

$$T(y) = \begin{cases} t + \tau_1 y, & \text{if } y \leq \tilde{y} \\ t + \tau_1 \tilde{y} + \tau_2(y - \tilde{y}), & \text{if } y > \tilde{y}, \end{cases} \tag{4.6}$$

with $t \in \mathbb{R}_-$, $\tilde{y} \in Y$, and $0 \leq \tau_1 \leq \tau_2 \leq 1$ have to be defined optimally.

Proposition 5 characterizes the optimal tax rule. We rely on the following condition, which we discuss below, that allows us to distinguish the case in which the optimal tax rule

[39] With endogenous social welfare weights, we can obtain implicit formulas for the tax rates which coincide with the ones provided here if one substitutes $\omega(\theta)$ with the endogenous weight w_θ. However, having exogenous social welfare weights considerably simplifies the comparative statics exercises.

is affine from the one in which it is strongly progressive:

$$(\omega(\theta_2) - \alpha)(y_2^0(\tau^*) - y_1^0(\tau^*)) + \frac{\alpha\Phi(y_2^0(\tau^*), \theta_2)}{(1-\tau^*)^2} \geq 0. \tag{4.7}$$

Proposition 5 (Binary types)**.** *If* (4.7) *is satisfied, then an affine tax rule with rate* τ^* *given by* (4.5) *is worst-case optimal.*

Otherwise, a piecewise affine tax rule with two income brackets is worst-case optimal. The marginal tax rates are τ_1 *for* $y \leq y_1^0$ *and* $\tau_2 > \tau_1$ *for* $y > y_1^0$*, where* y_j^0 *equals type* θ_j*'s income choice under* $(T, M^0(\theta_j))$*.* (τ_1, τ_2) *is given by*

$$\tau_1 = \begin{cases} 0, & \text{if } \mathbb{E}[\omega(\theta)] = \alpha, \\ \max\left\{0, 1 - \sqrt{\frac{p_1\alpha\Phi(y_1^0,\theta_1)}{(\alpha - \mathbb{E}[\omega(\theta)])y_1^0}}\right\}, & \text{if } \mathbb{E}[\omega(\theta)] < \alpha, \end{cases} \qquad \tau_2 = 1 - \sqrt{\frac{\alpha\Phi(y_2^0, \theta_2)}{(\alpha - \omega(\theta_2))(y_2^0 - y_1^0)}}.$$

In this case, workers' income choices under M^0 *satisfy* $y_1^0 < y_1^0(\tau_1)$*, and* $y_2^0 = y_2^0(\tau_2)$*.*

The proof is in Appendix A. According to Proposition 5, condition (4.7) is necessary and sufficient for the optimality of an affine tax rule. It is a special case of the condition given in Proposition 2, with $\tilde{y} = y_1^0(\tau^*)$. Essentially, this condition ensures that it is not beneficial to perturb the optimal affine tax by slightly increasing the tax rate above the baseline income level of the low type, $y_1^0(\tau^*)$.

Figure 4.1 depicts the optimal tax schedule for a case in which (4.7) is violated and therefore the optimal tax is strongly progressive. At the optimum, the kink is placed at the baseline income choice of the lower type, i.e., $\tilde{y} = y_1^0$. The intuition for this result can be understood as follows: Increasing the kink above y_1^0 only leads to a flat reduction in the tax paid by type θ_2 (see (4.6)), which the planner values according to $p_2(\omega(\theta_2) - \alpha) \leq 0$. As a consequence of this feature of the optimal tax, the income choice of the lower type is distorted in equilibrium: Even if there is zero taxation below y_1^0 (i.e., if at the optimum $\tau_1 = 0$), type θ_1 would like to produce more income if the marginal tax rate was everywhere constant and equal to τ_1.[40] As we show in the proof, the distortion in y_1^0 is proportional to the probability of the high type, and therefore disappears in the absence of worker heterogeneity.

As in Section 3.4, we can interpret the expressions for τ_1 and τ_2 in light of the mechanical and worst-case behavioral effects of perturbing the marginal taxes. A perturbation of the

[40] Specifically, the low type's income choice under both $M^0(\theta)$ and under the worst-case are distorted *relative to the worker's first-best welfare.* The fact that, as the Proposition states, $y_1^0 < y_1^0(\tau_1)$ implies that y_1^0 is distorted when workers face $M^0(\theta)$. Further, by Lemma 1, workers' worst-case income choice gives them utility that is (almost) equal to $V_w(T|M^0(\theta))$. Thus, worst-case income choices are strictly suboptimal as well.

tax rate above y_1^0 leads to a mechanical effect that is proportional to $y_2^0 - y_1^0$ (the part of the high type's income that is taxed at rate τ_2). Specifically, the mechanical welfare effect from perturbing τ_2 is $p_2(\alpha - \omega(\theta_2))(y_2^0 - y_1^0)$. The behavioral effect represents the reduction in worst-case income production by type θ_2, which leads to a loss in total welfare of $p_2\alpha\frac{\Phi(y_2^0,\theta_2)}{(1-\tau_2)^2}$. Setting the total welfare effect equal to zero gives the expression for τ_2. By a similar logic, a small increase in τ_1 leads to a mechanical increase of the tax payment made by *all workers*, with an effect on total welfare which is equal to $(\alpha - \mathbb{E}[\omega(\theta)])y_1^0$. On the other hand, the change only affects the income choice of the low type, and therefore the behavioral welfare loss of increasing τ_1 is given by $p_1\alpha\frac{\Phi(y_1^0,\theta_1)}{(1-\tau_1)^2}$.

Another pattern illustrated in Figure 4.1, which holds generally, is that the optimal tax schedule *does not* coincide with the pointwise minimum of the type-specific optimal affine tax.[41] This is in contrast to the intuition that would appear to emerge from the proof of Theorem 1, and highlights the role played by the Bayesian component in the planner's objective: Even though the 'inf' part of the planner's problem can be computed type-by-type irrespective of the distribution μ (Lemma 1), the tax rule attaining the 'sup' trades off having to tax different types using the same tax schedule and is therefore shaped by the distribution of types.

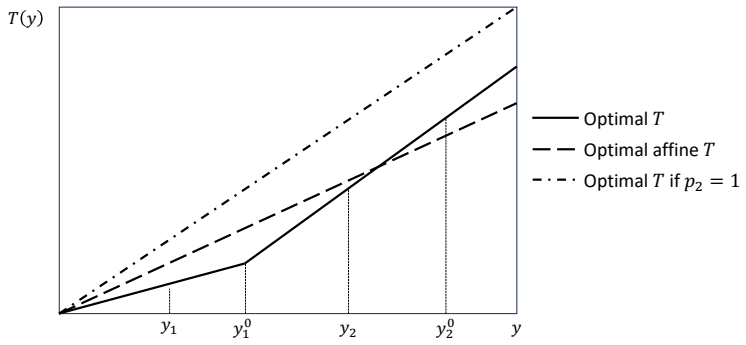

Figure 4.1: Optimal tax schedule under binary types.

Comparative statics. We conclude this section by studying the comparative statics of the optimal tax with respect to the model's primitives. The proofs are in Appendix C. Our results refer to the effects on the optimal τ_1, τ_2 and $\tilde{y}$ as defined in (4.6).

Our first result, Corollary 3, states that the optimal tax rate at the top converges to full taxation if the high type's marginal cost of producing income converges to zero. This result stands in stark contrast with zero taxation at the top in the canonical framework. It follows

[41] The optimal tax rate conditional on type θ_j is given by $\tau_j^* = \max\left\{0, 1 - \sqrt{\frac{\alpha\Phi(y_s^0(\tau_j),\theta_j)}{(\alpha-\omega(\theta_j))y_j^0(\tau_j)}}\right\}$ if $\omega(\theta_j) < \alpha$, and $\tau_j^* = 0$ if $\omega(\theta_j) = \alpha$. The problem is not well-defined if $\omega(\theta_j) > \alpha$.

from the assumption that the planner is inequality averse (Assumption 3) together with our recurring intuition that, if a worker is extremely productive, the planner is able to tax him heavily and in that way raise higher revenue with little allocative distortion.

Corollary 3 (Full taxation at the top)**.** *Suppose that* $\Phi(y, \theta_2) = a\phi(y)$*, where* $\phi : Y \to \mathbb{R}_+$ *is strictly increasing and strictly convex and* $a \in \mathbb{R}_{++}$*. For every* $\varepsilon > 0$*, there exists* $a^* > 0$ *such that* $a < a^*$ *implies that at the worst-case optimum* $\tau_2 > 1 - \varepsilon$*.*

Corollary 4 describes how the optimal tax changes with the planner's value for revenue, α. As expected, the optimal tax burden is higher when the planner's value for revenue increases. This result is graphically depicted in Figure 4.2. The Figure also illustrates that tax progressivity as measured by $\tau_2 - \tau_1$ is not necessarily monotone in α.

Corollary 4 (Comparative statics for α)**.** *At the worst-case optimum,* τ_1 *and* τ_2 *are non-decreasing in* α *(regardless of whether or not they are equal), and if* $\tau_1 < \tau_2$ *then* $\tilde{y}$ *is non-increasing in* α*.*

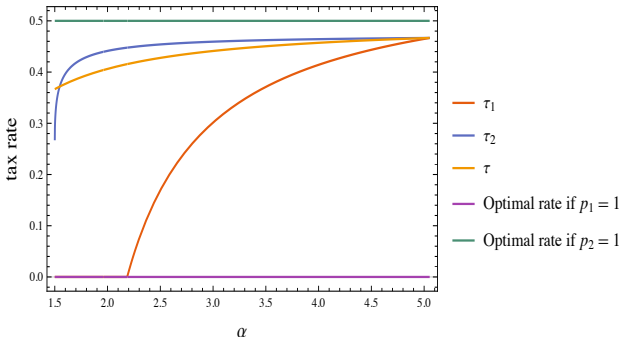

Figure 4.2: Tax rates as a function of α, in a region where strong progressivity is optimal ((4.7) does not hold). Parameter values: $p_1 = 0.75$, $\omega(\theta_1) = 2$, $\omega(\theta_2) = 0$.

A similar conclusion is obtained when carrying out comparative statics with respect to the planner's preference for redistribution, as measured by the Pareto weights $\omega(\cdot)$. An increase in $\omega(\theta_2)$ leads to a decrease in both τ_1 and τ_2, and therefore an overall downward shift in the optimal tax schedule $T(y)$. To understand these comparative statics, observe that the mechanical effect of slightly increasing τ_1 and τ_2 is, respectively $(\alpha - \mathbb{E}[\omega(\theta)])y_1^0$ and $p_2(\alpha - \omega(\theta_2))(y_2^0 - y_1^0)$, both of which are decreasing in $\omega(\theta_2)$.[42] Consequently, a higher $\omega(\theta_2)$ reduces the relative gain from raising revenue and therefore favors the use of lower marginal taxes over the entire income distribution.

As depicted in Figure 4.3, the overall effect of changing $\omega(\theta_2)$ on $\tau_2 - \tau_1$ is ambiguous. However, we show in Corollary 5 that, if the planner's *average* value for workers' welfare is

[42] The behavioral effect is constant in $\omega(\cdot)$.

sufficiently high (in the sense of $\mathbb{E}[\omega(\theta)]$ being high enough), then the intuitive comparative statics hold and the progressivity of the tax $(\tau_2 - \tau_1)$ increases with the degree of inequality aversion $(\omega(\theta_1) - \omega(\theta_2))$. In particular, if as in the canonical Mirrleesian model, we set $\mathbb{E}[\omega(\theta)] = \alpha$, then Corollary 5 implies that indeed progressivity (locally) increases with the planner's preference for redistribution.

Corollary 5 (Comparative statics for $\omega(\theta_2)$)**.** *At the worst-case optimum, τ_1 and τ_2 are non-increasing in $\omega(\theta_2)$ (regardless of whether or not they are equal). Moreover, there exists a cutoff $\omega \in (0, \alpha)$ such that if $\mathbb{E}[\omega(\theta)] \in (\omega, \alpha]$, then $\tau_2 - \tau_1$ is non-increasing in $\omega(\theta_2)$.*

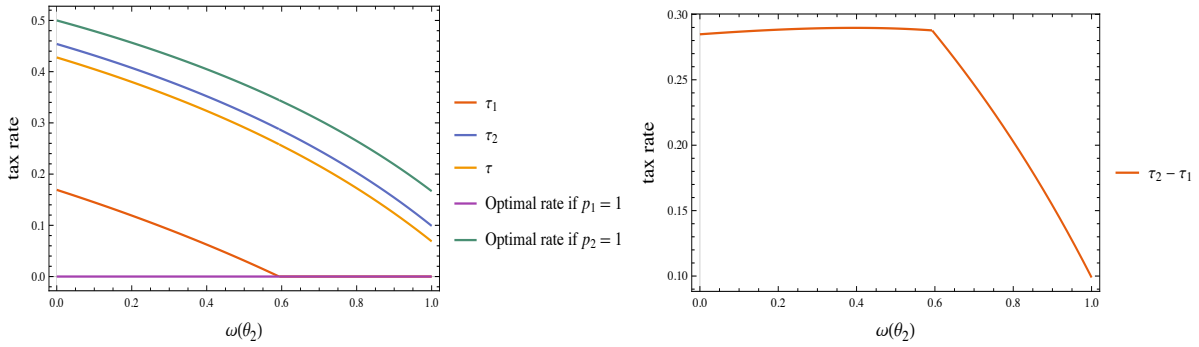

Figure 4.3: Tax rates as a function of $\omega(\theta_2)$, in a region where strong progressivity is optimal ((4.7) does not hold). Parameter values: $p_1 = 0.75$, $\omega(\theta_1) = 2$, $\alpha = 2.5$.

5 Discussion and Extensions

We studied the design of optimal labor income taxes in the presence of informational constraints by the government. Using a robust approach to uncertainty, we established that the optimal tax rule is progressive, and that this result holds independently of the government's attitudes toward inequality and of the details of the distribution of workers' types. Our model can be interpreted more generally as an instance of robust principal-agent contracting under moral hazard and adverse selection, in which case our main result provides a foundation for the use of concave contracts.

To maintain tractability, our model relies on a set of stylized assumptions regarding workers' utility functions, the planner's information structure, and the set of available mechanisms. These assumptions enable us to show a key qualitative aspect of the optimal tax. To conclude, we discuss the extent to which our results remain valid when relaxing these assumptions.

Uncertainty about the distribution of workers' productivity. Given the paper's motivation, it is natural to question the assumption that the planner knows the distribution of worker types μ. This assumption plays no role in our main result: The proof of Theorem 1 relies on constructing a progressive tax rule that dominates type by type, and thereby does not rely on the details about μ. Therefore, progressivity continues to be a qualitative desideratum for income taxes in a model where, on top of uncertainty about workers' income possibilities, the planner faces uncertainty about the distribution of workers' heterogeneity.

Of course, as the results that follow Theorem 1 have illustrated, the actual shape of the tax rule depends on the details of μ and the knowledge that the planner has about it. In the paper, we have focused on the case in which μ is known. A straightforward extension of our main arguments can be used to show that, in the opposite case in which the planner has no knowledge about μ other than its support, and her objective is the worst-case welfare over M and μ, then the optimal tax is affine and coincides with the optimal tax in the case in which μ is known to be the Dirac measure on $\{\underline{\theta}\}$. In practice, the government is likely to have partial knowledge about μ, and this can be accommodated in our framework.

Other structures of uncertainty about the production technology. The uncertainty faced by the social planner about the production technology M is very rich. Our main result continues to hold if the planner instead only considers a simpler class of correspondences, which are those of the form $M^0(\theta) \cup \{(F, \phi)\}$ with $(F, \phi) \in \Delta(Y) \times \mathbb{R}_+$. Intuitively, this would involve assuming that the planner entertains the possibility that workers have a single income opportunity added to their baseline choice set. As we showed in Lemma 1, in the worst case, this additional income choice minimizes revenue and does not improve the worker's welfare relative to the baseline. Additionally, the structure of the problem remains unaffected if instead of minimizing over the entire set of distributions $\Delta(Y)$, the worst case was computed across the smaller space of all distributions on Y with *binary* support. As a result, our findings do not require workers' income to be drawn from highly complicated lotteries.

Overall, the set of plausible technologies that we have assumed in the paper is a natural starting point that has received considerable attention in the robust contracting literature. There are certainly other information structures that the planner may face in reality and that might lead to different insights. For example, one might consider the case in which the planner faces *a small amount of uncertainty*, whereby she only contemplates correspondences that belong to some neighborhood of the baseline M^0. If, as in Section 4, one were to further assume that M^0 satisfies the typical assumptions of a Mirrlees economy, then we would expect the results that arise in that model to be similar to the ones that arise in the canonical framework.

Screening. We assumed that the planner is restricted to offering a single tax rule for the entire population of workers. This simplifying assumption is in many cases realistic,[43] but can entail a loss in optimality. Here, we look into what happens if the planner can offer workers a menu of tax rules, and each worker knowing θ and $M(\theta)$ self-selects into his most preferred tax.

Suppose that the planner can offer any compact menu of tax rules $\mathcal{T} \subseteq \overline{\mathbf{T}}$. When facing $\mathcal{T}$ and the choice set $M(\theta)$, a worker of type θ chooses a tax and an income choice to maximize

$$V_w(\mathcal{T}|M(\theta)) = \max_{T\in\mathcal{T},(F,\phi)\in M(\theta)} \{\mathbb{E}_F[y - T(y)] - \phi\}.$$

Given workers' income and tax choices, we let $V_P(\mathcal{T}|M)$ denote total welfare under the menu $\mathcal{T}$ and the correspondence $M \in \mathcal{M}$ defined in the same way as in Section 2, and $V_P(\mathcal{T}) = \inf_{M\in\mathcal{M}} V_P(\mathcal{T}|M)$. Our next result establishes that, without loss of optimality, the planner can restrict attention to menus whose cardinality is at most the cardinality of Θ. This result is not trivial, given that in principle the planner may use taxes to screen for the technology M. It generalizes Theorem 4 in Carroll (2015), which states that screening does not help the principal whenever θ is degenerate.

Proposition 6 (Upper bound on $\mathcal{T}$)**.** *For every feasible menu of tax rules $\mathcal{T}$, there exists a function $f : \Theta \to \mathcal{T}$ and a feasible menu $\overline{\mathcal{T}} = \bigcup_{\theta\in\Theta}\{f(\theta)\} \subseteq \mathcal{T}$ such that $V_P(\overline{\mathcal{T}}) \geq V_P(\mathcal{T})$.*

The proof is in Appendix A. To put Proposition 6 in plain words, suppose that $\Theta = \{\theta_1, \theta_2\}$. Then, our result says that it is without loss of optimality for the planner to use a menu of taxes with at most two elements. According to the proof of the proposition, the planner may benefit from offering more than one tax only so long as different worker types are willing to choose different elements of the menu under the baseline M^0. As a consequence, if M^0 is constant (but workers are still ex-ante heterogeneous due to the social welfare weight $\omega(\theta)$), there is an optimal menu which is a singleton. Giving a more general characterization of the condition under which the principal can strictly gain from screening by offering a menu of taxes remains an interesting open question.

Next, we show that tax progressivity remains optimal even if the planner has the ability to screen: Any menu of taxes is dominated, under the worst-case criterion, by an alternative menu that contains exclusively convex taxes. Again, the proof is in Appendix A.

Proposition 7 (Menu of progressive taxes)**.** *For every menu of tax rules $\mathcal{T}$, there exists a menu $\overline{\mathcal{T}}$ such that T is convex for every $T \in \overline{\mathcal{T}}$ and $V_P(\overline{\mathcal{T}}) \geq V_P(\mathcal{T})$.*

[43] Existing systems typically do not entail the government eliciting workers' private information through the tax code. A more common practice, which can be easily accommodated by our model, is to condition taxes on observables.

Workers' risk aversion. The assumption that workers are risk-neutral is admittedly strong. The role that it plays in our main result is that it enables to sustain random income choices by workers, which in turn motivates the planner to hedge against worst-case income risk by offering a progressive tax. We can generalize our main result to the case in which workers are risk-averse by establishing that the optimal *worker-welfare schedule* is progressive (in the sense of being concave in income).

To that end, suppose that workers value consumption according to the continuous, strictly increasing, and concave utility function $\tilde{u} : \mathbb{R}_+ \to \mathbb{R}$. The other aspects of the model are as described in Section 2. Let $\underline{u} \equiv \tilde{u}(0)$. Any tax rule $T(y)$ gives rise to a *utility contract* $u : Y \to \mathbb{R}$ given by $\tilde{u}(y - T(y))$. Since $\tilde{u}$ is invertible, any utility contract u uniquely pins down the associated tax rule T that implements it. So the problem where the planner optimizes over feasible tax rules is equivalent to one where the planner optimizes over continuous utility contracts $u : Y \to \mathbb{R}$ subject to the constraint that $u(y) \geq \underline{u}$ for all $y \in Y$. Proposition 8 states that, if workers are risk-averse, then the analog of Theorem 1 holds in utility space.

Proposition 8 (Risk-aversion)**.** *If $\tilde{u}$ is concave on $\mathbb{R}_+$, then there exists an optimal utility contract u which is concave.*

According to Proposition 8, there is an optimal tax rule that is progressive in the sense that the marginal *utility* that workers extract from their earned income decreases with their earnings. Given that it is workers' utility for consumption (not consumption itself) that the planner cares about, this notion is arguably a valid way to judge the progressivity of the tax system.

Alternatively, our model can be deemed as an approximation to settings in which workers have relatively low risk aversion. Under that interpretation, it is possible to extend our proof of Proposition 8 to show that, as workers' risk aversion vanishes, the optimum can be approximated arbitrarily well by a progressive tax rule. In particular, a commonly made assumption that has gained empirical support is that high-income earners have low risk aversion (Ogaki and Zhang, 2001). In light of this, our findings can be used to justify the implementation of progressive taxes, specifically *at the top* of the income distribution.

References

Akbarpour, M., P. Dworczak, and S. D. Kominers (2023): "Redistributive allocation mechanisms," Available at SSRN 3609182.

Aliprantis, C. D. and K. C. Border (2006): Infinite Dimensional Analysis, Springer.

Antic, N. and G. Georgiadis (2023): "Robust Contracts: A Revealed Preference Approach," Available at SSRN 4281652.

Atkinson, A. B. (1995): Public economics in action: the basic income/flat tax proposal, Clarendon Press.

Aumann, R. J. and M. Maschler (1995): Repeated games with incomplete information, MIT press.

Barron, D., G. Georgiadis, and J. Swinkels (2020): "Optimal contracts with a risk-taking agent," Theoretical Economics, 15, 715–761.

Berliant, M. and M. Gouveia (2022): "On the Political Economy of Nonlinear Income Taxation," .

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

Carroll, G. (2015): "Robustness and linear contracts," American Economic Review, 105, 536–563.

——— (2019): "Robustness in mechanism design and contracting," Annual Review of Economics, 11, 139–166.

Carroll, G. and D. Meng (2016): "Robust contracting with additive noise," Journal of Economic Theory, 166, 586–604.

Chassang, S. (2013): "Calibrated incentive contracts," Econometrica, 81, 1935–1971.

Chung, K.-S. and J. C. Ely (2007): "Foundations of dominant-strategy mechanisms," The Review of Economic Studies, 74, 447–476.

Dai, T. and J. Toikka (2022): "Robust incentives for teams," Econometrica, 90, 1583–1613.

Diamond, P. and E. Saez (2011): "The case for a progressive tax: From basic research to policy recommendation," Journal of Economic Perspectives, 25, 165–190.

Diamond, P. A. (1998): "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates," American Economic Review, 83–95.

Doval, L. and V. Skreta (2023): "Constrained information design," Mathematics of Operations Research.

Duffie, D. and Y. Sun (2012): "The exact law of large numbers for independent random matching," Journal of Economic Theory, 147, 1105–1139.

Dworczak, P., S. D. Kominers, and M. Akbarpour (2021): "Redistribution through markets," Econometrica, 89, 1665–1698.

Farhi, E. and I. Werning (2013): "Insurance and taxation over the life cycle," Review of Economic Studies, 80, 596–635.

Fenchel, W. and D. W. Blackett (1953): Convex cones, sets, and functions, Princeton University, Department of Mathematics, Logistics Research Project.

Garrett, D. F. (2014): "Robustness of simple menus of contracts in cost-based procurement," Games and Economic Behavior, 87, 631–641.

Golosov, M., M. Troshkin, and A. Tsyvinski (2016): "Redistribution and social insurance," American Economic Review, 106, 359–386.

Hansen, L. P. and T. J. Sargent (2001): "Robust control and model uncertainty," American Economic Review, 91, 60–66.

Kambhampati, A. (2023): "Randomization is optimal in the robust principal-agent problem," Journal of Economic Theory, 207, 105585.

Kamenica, E. and M. Gentzkow (2011): "Bayesian persuasion," American Economic Review, 101, 2590–2615.

Kang, Z. Y. (2023): "The Public Option and Optimal Redistribution," Tech. rep., Working Paper.

Laffont, J.-J. and D. Martimort (2009): "The theory of incentives: the principal-agent model," in The theory of incentives, Princeton university press.

Laffont, J.-J. and J. Tirole (1986): "Using cost observation to regulate firms," Journal of political Economy, 94, 614–641.

Le Treust, M. and T. Tomala (2019): "Persuasion with limited communication capacity," Journal of Economic Theory, 184, 104940.

Luenberger, D. G. (1997): Optimization by vector space methods, John Wiley & Sons.

Makris, M. and A. Pavan (2021): "Taxation under learning by doing," Journal of Political Economy, 129, 1878–1944.

Marku, K., S. Ocampo Díaz, and J.-B. Tondji (2022): "Robust contracts in common agency," Tech. rep., Research Report.

Milgrom, P. and I. Segal (2002): "Envelope theorems for arbitrary choice sets," Econometrica, 70, 583–601.

Milgrom, P. and C. Shannon (1994): "Monotone comparative statics," Econometrica: Journal of the Econometric Society, 157–180.

Mirrlees, J. (1974): "Notes on welfare economics, information, and uncertainty. M. Balch, D. McFadden, S. Wu, eds. Essays on Economic Behavior Under Uncertainty," .

Mirrlees, J. A. (1971): "An exploration in the theory of optimum income taxation," The review of economic studies, 38, 175–208.

Ogaki, M. and Q. Zhang (2001): "Decreasing relative risk aversion and tests of risk sharing," Econometrica, 69, 515–526.

Pai, M. and P. Strack (2023): "Taxing Externalities Without Hurting the Poor," Available at SSRN 4180522.

Phelps, E. S. (1973): "Taxation of wage income for economic justice," The Quarterly Journal of Economics, 87, 331–354.

Piketty, T. and E. Saez (2013): "Optimal labor income taxation," in Handbook of public economics, Elsevier, vol. 5, 391–474.

Ramsey, F. P. (1927): "A Contribution to the Theory of Taxation," The economic journal, 37, 47–61.

Rockafellar, R. T. (1997): Convex analysis, vol. 11, Princeton university press.

Rosenthal, M. (2022): "Simple Incentives and Diverse Beliefs," Available at SSRN 4266430.

Sadka, E. (1976): "On income distribution, incentive effects and optimal income taxation," The review of economic studies, 43, 261–267.

Saez, E. (2001): "Using elasticities to derive optimal income tax rates," The review of economic studies, 68, 205–229.

Saez, E. and S. Stantcheva (2016): "Generalized social marginal welfare weights for optimal tax theory," American Economic Review, 106, 24–45.

Seade, J. K. (1977): "On the shape of optimal tax schedules," Journal of public Economics, 7, 203–235.

Stantcheva, S. (2017): "Optimal taxation and human capital policies over the life cycle," Journal of Political Economy, 125, 1931–1990.

Sun, Y. (2006): "The exact law of large numbers via Fubini extension and characterization of insurable risks," Journal of Economic Theory, 126, 31–69.

Tuomala, M. (2016): Optimal redistributive taxation, Oxford University Press.

Varian, H. R. (1980): "Redistributive taxation as social insurance," Journal of public Economics, 14, 49–68.

Vereshchagina, G. and H. A. Hopenhayn (2009): "Risk taking by entrepreneurs," American Economic Review, 99, 1808–1830.

Walton, D. and G. Carroll (2022): "A general framework for robust contracting models," Econometrica, 90, 2129–2159.

Appendix

A Omitted Proofs

Proof of Lemma 1

Let

$$r^\theta(T) = \min_{F\in\Delta(Y)} \mathbb{E}_F[T(y)], \quad s.t. \quad \mathbb{E}_F[y - T(y)] \geq V_w(T|M^0(\theta)). \tag{A.1}$$

We make use of the following claim.

Claim 1. $\omega(\theta)W(V_w(T|M^0(\theta))) + \alpha r^\theta(T) = \tilde{v}^\theta(T)$, where

$$\tilde{v}^\theta(T) \equiv \min_{(F,\phi)\in\Delta(Y)\times\mathbb{R}_+} \{\omega(\theta)W(\mathbb{E}_F[y - T(y)] - \phi)) + \alpha\mathbb{E}_F[T(y)]\}, \tag{A.2}$$

$$s.t. \quad \mathbb{E}_F[y - T(y)] - \phi \geq V_w(T|M^0(\theta)). \tag{A.3}$$

Moreover, if $(\tilde{F}_\theta, \tilde{\phi}_\theta)$ is a solution to (A.2),(A.3), it holds that $\mathbb{E}_{\tilde{F}_\theta}[y - T(y)] - \tilde{\phi}_\theta = V_w(T|M^0(\theta))$.

Proof. In any solution to the problem (A.2),(A.3), it must be that $\mathbb{E}_F[y - T(y)] - \phi = V_w(T|M^0(\theta))$. This is because, otherwise, it is possible to increase ϕ which strictly reduces the value of the objective while still satisfying the constraint. Therefore, we can write

$$\tilde{v}^\theta(T) = \min_{F\in\Delta(Y),\phi\in\mathbb{R}_+} \{\omega(\theta)W(V_w(T|M^0(\theta))) + \alpha\mathbb{E}_F[T(y)], \quad s.t. \quad \mathbb{E}_F[y - T(y)] - \phi \geq V_w(T|M^0(\theta))\}.$$

Let $M^*_\theta \subseteq \Delta(Y) \times \mathbb{R}_+$ be the set of income choices that attains the minimum for type θ in this program. For any $(F_\theta, \phi_\theta) \in M^*_\theta$, it holds that $\mathbb{E}_{F_\theta}[y - T(y)] \geq V_w(T|M^0(\theta))$, and thus $r^\theta(T) \leq \mathbb{E}_{F_\theta}[T(y)]$. On the other hand, letting F^*_θ be an argmin for $r^\theta(T)$, we can define $(F^*_\theta, \mathbb{E}_{F^*_\theta}[y - T(y)] - V_w(T|M^0(\theta))) \in \Delta(Y) \times \mathbb{R}_+$, which is feasible in (A.2). Hence, $r^\theta(T) \geq \mathbb{E}_{F_\theta}[T(y)]$ for F_θ such that $(F_\theta, \phi_\theta) \in M^*_\theta$. This establishes that $\tilde{v}^\theta(T) = \omega(\theta)W(V_w(T|M^0(\theta))) + \alpha r^\theta(T)$ as desired. □

We also refer to the following claim, which is a direct consequence of monotonicity and continuity of $M^0(\theta)$, and of Berge's Maximum Theorem.

Claim 2. For any feasible T, the functions $\theta \to r^\theta(T)$ and $\theta \to V_w(T|M^0(\theta))$ are continuous and non-decreasing.

Let us now use the above result to show the statement of the Lemma. Fix a correspondence $M \in \mathcal{M}$. Let $(\tilde{F}_\theta, \tilde{\phi}_\theta)$ be type θ's choice under $M(\theta)$, and let F_θ be defined as in

Lemma 1. First, if θ is such that $V_w(T|M^0(\theta)) = \max_{y\in Y}\{y - T(y)\}$, then type θ can secure maximal consumption at zero cost under $M^0(\theta)$. Therefore, under any M, type θ's choice $\tilde{F}_\theta$ will be supported on $Y^* \equiv \arg\max_{y\in Y}\{y - T(y)\}$, and $V_w(T|M^0(\theta)) = V_w(T|M(\theta))$. Since the worker breaks ties in favor of the social planner and $M(\theta) \supseteq M^0(\theta)$ it follows that $\mathbb{E}_{\tilde{F}_\theta}[T(y)] \geq \mathbb{E}_{F_\theta}[T(y)]$.

Next, consider θ such that $V_w(T|M^0(\theta)) < \max_{y\in Y}\{y - T(y)\}$. Since $M(\theta) \supseteq M^0(\theta)$, it holds that $(\tilde{F}_\theta, \tilde{\phi}_\theta)$ satisfies (A.3). Thus, $\omega(\theta)W(\mathbb{E}_{\tilde{F}_\theta}[y - T(y)] - \tilde{\phi}_\theta)) + \alpha\mathbb{E}_{\tilde{F}_\theta}[T(y)] \geq \tilde{v}^\theta(T)$. Combining these two points, applying Claim 1, and taking expectations with respect to θ we have that

$$V_P(T|M) \geq \int_\Theta \{\omega(\theta)W(V_w(T|M^0(\theta))) + \alpha\mathbb{E}_{F_\theta}[T(y)]\}\, d\mu(\theta).$$

Since M was arbitrary, the inequality holds if we substitute $V_P(T|M)$ for $V_P(T)$.

For the other direction, let $y^* \in Y^*$. For $\varepsilon \in (0,1)$, define

$$\underline{M}(\theta) \equiv \{((1-\varepsilon)F + \varepsilon\delta_{y^*}, (1-\varepsilon)\phi) : \mathbb{E}_F[T(y)] \leq r^\theta(T), \mathbb{E}_F[y - T(y)] - \phi \leq V_w(T|M^0(\theta))\},$$

and consider $M^\varepsilon \in \mathcal{M}$ defined by $M^\varepsilon(\theta) = M^0(\theta) \cup \underline{M}(\theta)$.[44] We show that the planner's payoff under M^ε approaches the value in the Lemma as $\varepsilon \to 0$.

Claim 3. For any feasible T,

$$\lim_{\varepsilon\downarrow 0} V_P(T|M^\varepsilon) = \int_\Theta \{\omega(\theta)W(V_w(T|M^0(\theta))) + \alpha\mathbb{E}_{F_\theta}[T(y)]\}\, d\mu(\theta)$$

Proof. Consider first θ such that $V_w(T|M^0(\theta)) < y^* - T(y^*)$, and therefore $\mathbb{E}_{\tilde{F}_\theta}[y - T(y)] - \tilde{\phi}_\theta < y^* - T(y^*)$ for all $(\tilde{F}_\theta, \tilde{\phi}_\theta) \in M^*_\theta$. Type θ's optimal choice under $M^\varepsilon(\theta)$ is to choose $((1-\varepsilon)F + \varepsilon\delta_{y^*}, (1-\varepsilon)\phi) \in \underline{M}(\theta)$ such that $\mathbb{E}_F[y - T(y)] - \phi = V_w(T|M^0(\theta))$, which gives him a payoff of $(1-\varepsilon)V_w(T|M^0(\theta)) + \varepsilon(y^* - T(y^*)) > V_w(T|M^0(\theta))$. By definition of $r^\theta(T)$ and Claim 1, the only pairs (F,ϕ) satisfying $\mathbb{E}_F[y - T(y)] - \phi = V_w(T|M^0(\theta))$ and $\mathbb{E}_F[T(y)] \leq r^\theta(T)$ are such that $(F,\phi) \in M^*_\theta$, and therefore the worker chooses an element of M^*_θ under $M^\varepsilon(\theta)$.

On the other hand, if $V_w(T|M^0(\theta)) = y^* - T(y^*)$, then as argued before, type θ's choice takes the form $(F, 0)$ with F fully supported on Y^*. If there are several income choices in $M^\varepsilon(\theta)$ of this form, he picks the one that maximizes the planner's revenue. Let $\tilde{M}_\theta \subset M^\varepsilon(\theta)$ be the set of choices for type θ that satisfy these two conditions. For ε sufficiently small, it

[44] Monotonicity and continuity of M^ε follow from Claim 2.

must be that $\tilde{M}_\theta \cap M^0(\theta)$ is non-empty. If not, there exists $(F,0) \in \underline{M}(\theta)$ with F supported on Y^* that gives the planner strictly higher revenue than F_θ (with F_θ defined as in Lemma 1). This contradicts the definition of $r^\theta(T)$ as minimizing revenue subject to the constraint that $\mathbb{E}_F[y - T(y)] = y^* - T(y^*)$. Therefore, for sufficiently small ε, type θ's choice under $M^\varepsilon(\theta)$ is $(F_\theta, 0)$ with F_θ as defined in Lemma 1.

The fact that $V_w(T|M^0(\theta))$ is continuous and non-decreasing in θ implies that there is a cutoff $\tilde{\theta} \in \Theta \cup \{+\infty\}$ such that $V_w(T|M^0(\theta)) = y^* - T(y^*)$ if and only if $\theta \geq \tilde{\theta}$. Combining the above and letting $(\tilde{F}_\theta, \tilde{\phi}_\theta) \in M^*_\theta$, we have

$$
\begin{aligned}
&V_P(T) \leq \lim_{\varepsilon \downarrow 0} V_P(T|M^\varepsilon) = \\
&\lim_{\varepsilon \downarrow 0} \int_{\underline{\theta}}^{\tilde{\theta}} \{\omega(\theta) W((1-\varepsilon)(\mathbb{E}_{\tilde{F}_\theta}[y - T(y)] - \tilde{\phi}_\theta) + \varepsilon(y^* - T(y^*))) + \alpha((1-\varepsilon)\mathbb{E}_{\tilde{F}_\theta}[T(y)] + \varepsilon T(y^*))\}\, d\mu(\theta) \\
&+ \int_{\tilde{\theta}}^{\overline{\theta}} \{\omega(\theta) W(\mathbb{E}_{F_\theta}[y - T(y)]) + \alpha \mathbb{E}_{F_\theta}[T(y)]\}\, d\mu(\theta)
\end{aligned}
$$

For all $\varepsilon \in (0,1)$ and $\theta \leq \tilde{\theta}$, it holds that

$$
\begin{aligned}
&|\omega(\theta) W((1-\varepsilon)(\mathbb{E}_{\tilde{F}_\theta}[y - T(y)] - \tilde{\phi}_\theta) + \varepsilon(y^* - T(y^*))) + \alpha((1-\varepsilon)\mathbb{E}_{\tilde{F}_\theta}[T(y)] + \varepsilon T(y^*))| \\
&\qquad \leq \omega(\theta) W(y^* - T(y^*)) + \alpha \max_{y \in Y} |T(y)|.
\end{aligned}
$$

Thus, by the Dominated Convergence Theorem,

$$
\begin{aligned}
&V_P(T) \leq \\
&\int_{\underline{\theta}}^{\tilde{\theta}} \lim_{\varepsilon \downarrow 0} \{\omega(\theta) W((1-\varepsilon)(\mathbb{E}_{\tilde{F}_\theta}[y - T(y)] - \tilde{\phi}_\theta) + \varepsilon(y^* - T(y^*))) + \alpha((1-\varepsilon)\mathbb{E}_{\tilde{F}_\theta}[T(y)] + \varepsilon T(y^*))\}\, d\mu(\theta) \\
&+ \int_{\tilde{\theta}}^{\overline{\theta}} \{\omega(\theta) W(\mathbb{E}_{F_\theta}[y - T(y)]) + \alpha \mathbb{E}_{F_\theta}[T(y)]\}\, d\mu(\theta) \\
&= \int_{\underline{\theta}}^{\tilde{\theta}} \{\omega(\theta) W(\mathbb{E}_{\tilde{F}_\theta}[y - T(y)] - \tilde{\phi}_\theta) + \alpha \mathbb{E}_{\tilde{F}_\theta}[T(y)]\}\, d\mu(\theta) \\
&+ \int_{\tilde{\theta}}^{\overline{\theta}} \{\omega(\theta) W(\mathbb{E}_{F_\theta}[y - T(y)]) + \alpha \mathbb{E}_{F_\theta}[T(y)]\}\, d\mu(\theta) \\
&= \int_{\Theta} \{\omega(\theta) W(V_w(T|M^0(\theta)))) + \alpha \mathbb{E}_{F_\theta}[T(y)]\}\, d\mu(\theta),
\end{aligned}
$$

where the final equality follows from Claim 1. □

This establishes (3.1).

Proof of Theorem 1

We begin by showing that, for every feasible T, there exists an alternative feasible $\overline{T}$ which is non-decreasing and convex such that $V_P(T) \leq V_P(\overline{T})$. Then, we show that a worst-case optimal tax exists. Let T be a feasible tax rule, and write $c(y) = y - T(y)$ which is the associated consumption rule.

Step 1: θ-specific dominating affine rules. We begin by constructing an affine tax rule for every θ that dominates the original one, conditional on the worker's type.

Let F_θ be defined as in Lemma 3.1. Let

$$A \equiv \text{co}\{(c(y), y - c(y)) \in \mathbb{R}^2 : y \in Y\}, \quad B_\theta \equiv \{(u, v) \in \mathbb{R}^2 : u > V_w(T|M^0(\theta)), v < E_{F_\theta}[T(y)]\},$$

where $\text{co}(A)$ stands for the convex hull of the set A. For every $\theta \in \Theta$, the definition of F_θ implies that A and B_θ are disjoint. Since the two sets are convex, by the Separating Hyperplane Theorem, there exists $(\kappa_\theta, \beta_\theta, \lambda_\theta) \in \mathbb{R}^3$ with $(\beta_\theta, \lambda_\theta) \neq 0$ such that

$$\kappa_\theta + \lambda_\theta c(y) - \beta_\theta (y - c(y)) \leq 0, \quad \forall y \in Y \tag{A.4}$$

$$\kappa_\theta + \lambda_\theta u - \beta_\theta v \geq 0, \quad \forall (u, v) \in B_\theta. \tag{A.5}$$

Observe that $(\mathbb{E}_{F_\theta}[c(y)], \mathbb{E}_{F_\theta}[T(y)])$ belongs to the closure of B_θ, and thus

$$\kappa_\theta + \lambda_\theta \mathbb{E}_{F_\theta}[c(y)] - \beta_\theta \mathbb{E}_{F_\theta}[T(y)] = 0. \tag{A.6}$$

By (A.5) and the definition of B_θ, we have that $\lambda_\theta \geq 0$ and $\beta_\theta \geq 0$. By (A.4), $\kappa_\theta \leq -(\lambda_\theta + \beta_\theta)c(0) \leq 0$. Moreover, if $\beta_\theta = 0$, then (A.4) and (A.5) imply that $V_w(T|M^0(\theta)) \geq \max_{y \in Y} c(y)$. Thus, $\beta_\theta > 0$ for all θ such that $V_w(T|M^0(\theta)) < \max_{y \in Y} c(y)$.

Moreover, we show that, if $\beta_\theta > 0$, then

$$\mathbb{E}_{F_\theta}[T(y)] = \frac{\kappa_\theta + \lambda_\theta V_w(T|M^0(\theta))}{\beta_\theta}. \tag{A.7}$$

If $\lambda_\theta = 0$, then (A.7) follows immediately from (A.6). If $\lambda_\theta > 0$ and $\beta_\theta > 0$, then (A.5) implies that $\kappa_\theta + \lambda_\theta V_w(T|M^0(\theta)) - \beta_\theta \mathbb{E}_{F_\theta}[T(y)] \geq 0$; while on the other hand we have that $\mathbb{E}_{F_\theta}[c(y)] \geq V_w(T|M^0(\theta))$ and $\kappa_\theta + \lambda_\theta \mathbb{E}_{F_\theta}[c(y)] - \beta_\theta \mathbb{E}_{F_\theta}[T(y)] = 0$. The two assertions together with $\lambda_\theta > 0$ imply that $V_w(T|M^0(\theta)) = \mathbb{E}_{F_\theta}[c(y)]$ and thus (A.7) holds.

For each $\theta \in \Theta$, we define the following non-negative affine consumption rule

$$c_\theta(y) = \frac{\beta_\theta y - \kappa_\theta}{\beta_\theta + \lambda_\theta}. \tag{A.8}$$

Step 2: Constructing a dominating tax rule. We now use c_θ to construct a consumption rule that outperforms c. Let

$$\bar{c}(y) = \inf_{\theta \in \Theta} c_\theta(y), \tag{A.9}$$

and $\overline{T}(y) = y - \bar{c}(y)$. $\overline{T}$ satisfies the properties of Theorem 1.[45] By (A.4), $c_\theta(y) \geq c(y)$ for all $\theta \in \Theta$ and $y \in Y$, and hence $\bar{c}(y) \geq c(y)$. It follows that $V_w(\overline{T}|M^0(\theta)) \geq V_w(T|M^0(\theta))$ for all θ, and thus worst-case worker-welfare is higher under $\overline{T}$.

We now show that revenue is also higher under $\overline{T}$ for every $\theta \in \Theta$. To that end, consider first $\theta \in \Theta$ such that $\beta_\theta > 0$. For any $\tilde{F} \in \Delta(Y)$ satisfying $\mathbb{E}_{\tilde{F}}[\bar{c}(y)] \geq V_w(\overline{T}|M^0(\theta))$ (by Lemma 3.1 it suffices to restrict attention to income choices satisfying this condition), we can write

$$\mathbb{E}_{\tilde{F}}[y - \bar{c}(y)] \geq \mathbb{E}_{\tilde{F}}[y - c_\theta(y)] = \frac{\kappa_\theta + \lambda_\theta \mathbb{E}_{\tilde{F}}[c_\theta(y)]}{\beta_\theta} \geq \frac{\kappa_\theta + \lambda_\theta V_w(\overline{T}|M^0(\theta))}{\beta_\theta} \tag{A.10}$$

$$\geq \frac{\kappa_\theta + \lambda_\theta V_w(T|M^0(\theta))}{\beta_\theta} = \mathbb{E}_{F_\theta}[T(y)], \tag{A.11}$$

where the final equality follows from (A.7).

Second, suppose that $\beta_\theta = 0$, and thus $\mathbb{E}_{F_\theta}[c(y)] = V_w(T|M^0(\theta)) = -\kappa_\theta/\lambda_\theta$. Since $c(y) \leq c_\theta(y) = -\kappa_\theta/\lambda_\theta$ for all y, it follows that $c(y) = -\kappa_\theta/\lambda_\theta$ for all y in the support of F_θ. Further, since $c_{\theta'}(y) \geq c(y)$ for all $\theta' \in \Theta$ and $y \in Y$, and equality holds under θ and y in the support of F_θ, we have that $\bar{c}(y) = -\kappa_\theta/\lambda_\theta = c(y)$ for all y in the support of F_θ. Moreover, for every $y \in Y$, $\bar{c}(y) \leq c_\theta(y) = -\kappa_\theta/\lambda_\theta$. Thus, under $\bar{c}$ and any M, the worker's payoff is maximized by choosing $(F_\theta, 0) \in M^0(\theta)$, and this choice yields revenue equal to $\mathbb{E}_{F_\theta}[T(y)] = \mathbb{E}_{F_\theta}[\overline{T}(y)]$. Given that the worker breaks indifference in favor of the planner, the revenue from type θ under M and $\bar{c}$ is at least $\mathbb{E}_{F_\theta}[T(y)]$.

In sum, $\overline{T}$ yields an improvement over T in terms of both workers' welfare and aggregate revenue, and thus $V_P(\overline{T}) \geq V_P(T)$.

Step 3: Existence. This concludes the argument that, for every feasible T, there is a $\overline{T}$ which is *(i)* non-decreasing and *(ii)* convex, that satisfies $V_P(\overline{T}) \geq V_P(T)$. Moreover, our construction shows that it is without loss of optimality to restrict to $T(y)$ such that *(iii)*

[45] This follows from the fact that $\bar{c}(y)$, as the pointwise infimum of a family of affine functions, is concave.

$y - T(y)$ is non-decreasing. We further show in Lemma 2 that $y - T(y)$ is without loss bounded above by $\overline{C} + \overline{y}$, where $\overline{C}$ is defined in Assumption 2.

Lemma 2. *For any feasible T such that $y - T(y)$ is non-decreasing and that satisfies $T(0) < -\overline{C}$, there exists $\varepsilon > 0$ and a tax rule $T^\varepsilon(y) = T(y) + \varepsilon$ such that $V_P(T^\varepsilon) \geq V_P(T)$.*

Proof. Take any feasible T such that $c(y) = y - T(y)$ is non-decreasing and $T(0) < -\overline{C}$, and consider a perturbation of the form $T^\varepsilon(y) = T(y) + \varepsilon$ with $\varepsilon > 0$. Since a constant shift in $T(y)$ does not affect workers' worst-case income choice as defined in Lemma 1, the function $\varepsilon \to V_P(T^\varepsilon)$ is differentiable, and we can write

$$\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon} = -\int_\Theta \omega(\theta) W'(V_w(T|M^0(\theta)) - \varepsilon)\, d\mu(\theta) + \alpha. \tag{A.12}$$

By Assumption 2, there exists $\overline{C} > 0$ such that $c > \overline{C}$ implies that $\mathbb{E}[\omega(\theta)]W'(c) \leq \alpha$. Moreover, by Assumption 1 and monotonicity of $c(y)$, we have that $V_w(T|M^0(\theta)) \geq c(0)$ for all θ. So, if $T(0) < -\overline{C}$ and $\varepsilon > 0$ is small enough, we have

$$-\int_\Theta \omega(\theta) W'(V_w(T|M^0(\theta)) - \varepsilon)\, d\mu(\theta) + \alpha \geq 0$$

Thus, T^ε with $\varepsilon > 0$ sufficiently small is a weak improvement over T. □

The proof is completed by showing that the maximum exists, which we do in the following lemma.

Lemma 3. *There exists a worst-case optimal tax rule.*

Proof. Let $\mathbf{T}$ be the set of continuous functions from Y to $[-\overline{C}, \overline{y}]$ that satisfy conditions *(i)* to *(iii)* from above. The arguments above, together with Lemma 2, imply that it suffices to show that $\max_{T \in \mathbf{T}} V_P(T)$ is well-defined. The family $\mathbf{T}$ is uniformly bounded and equicontinuous, and is thus compact by the Arzelà-Ascoli Theorem. Next, we establish that the objective is upper semi-continuous.

Claim 4. $V_P(T)$ is upper semi-continuous on $\mathbf{T}$.

Proof. Let $T^k \in \mathbf{T}$ be a sequence of tax rules converging to $T^\infty \in \mathbf{T}$. Fix a correspondence $M \in \mathcal{M}$ and let $(F_\theta^k, \phi_\theta^k) \in M(\theta)$ be type θ's chosen action under T^k and M. Without loss, $V_P(T^k)$ converges (if not, replace T^k by a subsequence). By compact-valuedness of M, without loss, for every $\theta \in \Theta$, $(F_\theta^k, \phi_\theta^k)$ converges to an element of $M(\theta)$, which we denote by $(F_\theta^\infty, \phi_\theta^\infty)$. By continuity, $(F_\theta^\infty, \phi_\theta^\infty)$ is optimal for type θ under T^∞ and M. Since the

worker breaks indifference in favor of the planner, the planner's payoff under T^∞ and M is bounded below by what obtains if all types choose $(F_\theta^\infty, \phi_\theta^\infty)$. Thus,[46]

$$V_P(T^\infty|M) \geq \int_\Theta \{\omega(\theta)W(\mathbb{E}_{F_\theta^\infty}[y - T^\infty(y)] - \phi_\theta^\infty) + \alpha\mathbb{E}_{F_\theta^\infty}[T^\infty(y)]\}\, d\mu(\theta) \tag{A.13}$$

$$= \lim_{k\to\infty} V_P(T^k|M) \geq \lim_{k\to\infty} V_P(T^k). \tag{A.14}$$

Since M was arbitrary, $V_P(T^\infty) \geq \lim_{k\to\infty} V_P(T^k)$, which establishes the result. □

Existence of the maximum follows from the fact that the objective is upper semi-continuous and the choice set compact. □

Applying Lemma 3 concludes the proof of the theorem.

Proof of Theorem 1*

Take any feasible $T \in \overline{\mathbf{T}}$. The result follows from applying Claim 8 below to the singleton menu $\{T\}$.

Proof of Proposition 1

Let $T(y)$ be an optimal tax rule and $c(y) = y - T(y)$ the associated consumption rule. Set $\overline{c}(y)$ to be the consumption rule defined in (A.9), and $\tilde{c}(y)$ to be the concavification of c, and write $\overline{T}(y) = y - \overline{c}(y)$ and $\tilde{T}(y) = y - \tilde{c}(y)$. Let $Y^* = \arg\max_{y\in Y} c(y)$, and $y^* \in Y^*$. We make use of the following claim.

Claim 5. $c(y) \leq \tilde{c}(y) \leq \overline{c}(y)$.

Proof. Recall that $\tilde{c}(y)$ is equal to the pointwise infimum of the collection of affine functions which majorize $c(y)$ (Rockafellar, 1997). The first inequality follows immediately from this definition of $\tilde{c}(y)$. Moreover, for each $\theta \in \Theta$, $c_\theta(y)$, as defined in (A.8), is one such affine function. Then, $\overline{c}(y)$ as defined in (A.9) obtains from taking the pointwise infimum over a *smaller* family of affine functions than the one used in the definition of $\tilde{c}(y)$, and therefore $\overline{c}(y) \geq \tilde{c}(y)$. □

[46] The equality in (A.13) uses the Dominated Convergence Theorem, which can be applied due to the fact that for each $k \in \mathbb{N}$ and $\theta \in \Theta$:

$$|\omega(\theta)W(\mathbb{E}_{F_\theta^k}[y - T^k(y)] - \phi_\theta^k) + \alpha\mathbb{E}_{F_\theta^k}[T^k(y)]| \leq \omega(\theta)W(\overline{C} + \overline{y}) + \alpha\overline{y}.$$

Take any $F^0_\theta \in \mathcal{F}^0_\theta(T)$ for each $\theta \in \Theta$, and set $\phi^0_\theta \in \mathbb{R}_+$ to be such that $(F^0_\theta, \phi^0_\theta) \in M^0(\theta)$ and $\mathbb{E}_{F^0_\theta}[c(y)] - \phi^0_\theta = V_w(T|M^0(\theta))$. Let $\tilde{\theta} = \min(\{\theta \in \Theta : V_w(T|M^0(\theta)) = c(y^*)\} \cup \{+\infty\})$, and $F^0 = \int_\Theta F^0_\theta \, d\mu(\theta)$. As shown in Theorem 1, $\theta < \tilde{\theta}$ implies that $\beta_\theta > 0$. Moreover, for all $\theta \geq \tilde{\theta}$, it holds that F^0_θ is supported on a subset of Y^* and that $\mathcal{F}^0_\theta(T) = \underline{\mathcal{F}}_\theta(T)$ (by Lemma 1). Since $c(y^*) = \overline{c}(y^*)$ for all $y^* \in Y^*$ (also shown in Theorem 1), it follows that $c(y) = \tilde{c}(y) = \overline{c}(y)$ F^0_θ- and F_θ-almost everywhere for all $\theta \geq \tilde{\theta}$, and $F^0_\theta \in \mathcal{F}^0_\theta$ and $F_\theta \in \underline{\mathcal{F}}_\theta$. If $\tilde{\theta} = \underline{\theta}$, then the proof is concluded.

Suppose now that $\tilde{\theta} > \underline{\theta}$, so that there is a positive measure of types with $\beta_\theta > 0$. Applying (A.11) and using the fact that $\overline{T}$ yields weakly higher payoff than T *conditional on each* θ, we can write

$$V_P(\overline{T}) - V_P(T) \geq$$
$$\int_{\underline{\theta}}^{\tilde{\theta}} \left[\omega(\theta)[W(V_w(\overline{T}|M^0(\theta))) - W(V_w(T|M^0(\theta)))] + \alpha \frac{\lambda_\theta[V_w(\overline{T}|M^0(\theta)) - V_w(T|M^0(\theta))]}{\beta_\theta} \right] d\mu(\theta) \geq$$
$$\int_{\underline{\theta}}^{\tilde{\theta}} \omega(\theta)[W(\mathbb{E}^0_{F_\theta}[\overline{c}(y)] - \phi^0_\theta) - W(\mathbb{E}^0_{F_\theta}[c(y)] - \phi^0_\theta)] \, d\mu(\theta),$$

where the second inequality uses the fact that $V_w(\overline{T}|M^0(\theta)) - V_w(T|M^0(\theta)) \geq 0$ for all θ and that $V_w(\overline{T}|M^0(\theta)) \geq \mathbb{E}^0_{F_\theta}[\overline{c}(y)] - \phi^0_\theta$. Given that W is strictly increasing and $\omega(\theta) > 0$ for all $\theta \in \Theta$, optimality of T requires that $\mathbb{E}_{F^0_\theta}[\overline{c}(y)] = \mathbb{E}_{F^0_\theta}[c(y)]$ almost everywhere in $\theta < \tilde{\theta}$. Combining everything and applying Claim 5, we obtain that $\overline{c}(y) = \tilde{c}(y) = c(y)$ F^0-almost everywhere.

Next, fix $F_\theta \in \underline{\mathcal{F}}_\theta$ for all $\theta \in \Theta$. It remains to be shown that $\int_{\underline{\theta}}^{\tilde{\theta}} \int_Y [\overline{c}(y) - c(y)] \, dF_\theta(y) \, d\mu(\theta) = 0$. We know from the previous paragraph that $V_w(\overline{T}|M^0(\theta)) = V_w(T|M^0(\theta)) < \max_{y \in Y} c(y) \leq \max_{y \in Y} \overline{c}(y)$ almost everywhere in $\theta < \tilde{\theta}$. Moreover, optimality of T requires that $\int_{\underline{\theta}}^{\tilde{\theta}} [r^\theta(\overline{T}) - r^\theta(T)] \, d\mu(\theta) = 0$. We thus have

$$\int_{\underline{\theta}}^{\tilde{\theta}} r^\theta(\overline{T}) \, d\mu(\theta) \leq \int_{\underline{\theta}}^{\tilde{\theta}} \mathbb{E}_{F_\theta}[\overline{T}(y)] \, d\mu(\theta) \leq \int_{\underline{\theta}}^{\tilde{\theta}} \mathbb{E}_{F_\theta}[T(y)] \, d\mu(\theta) = \int_{\underline{\theta}}^{\tilde{\theta}} r^\theta(T) \, d\mu(\theta) = \int_{\underline{\theta}}^{\tilde{\theta}} r^\theta(\overline{T}) \, d\mu(\theta),$$

where the first inequality follows from the definition of $r^\theta(\overline{T})$ and the fact that $\mathbb{E}_{F_\theta}[y - \overline{T}(y)] \geq \mathbb{E}_{F_\theta}[y - T(y)] \geq V_w(T|M^0(\theta)) = V_w(\overline{T}|M^0(\theta))$ almost everywhere in θ, and the second inequality from the fact that $\overline{T}(y) \leq T(y)$ for all $y \in Y$. Hence, it must be that $\mathbb{E}_{F_\theta}[\overline{T}(y)] = \mathbb{E}_{F_\theta}[T(y)]$ almost everywhere in θ, which together with Claim 5 yields the result.

Proof of Proposition 2

We begin by finding an optimal tax rule within the affine class, which we denote by $T_a(y) = t^* + \tau^* y$. Without loss of optimality $\tau^* \in [0,1]$: An analogous argument to the one in the proof of Theorem 1 shows that, starting from an affine tax with $\tau \notin [0,1]$, it is possible to construct a dominating affine tax rule that satisfies this property.

Lemma 4. *The tax rule $T_a(y) = t^* + \tau^* y$ where (τ^*, t^*) satisfy*

$$\int_\Theta (w_\theta(t^*,\tau^*) - \alpha)\, d\mu(\theta) \leq 0, \quad \textit{with equality if } t^* < 0,$$

$$\int_\Theta (w_\theta(t^*,\tau^*) - \alpha) y_\theta^0(\tau^*)\, d\mu(\theta) \leq 0 \textit{ and } \int_\Theta \phi_\theta^0(\tau^*)\, d\mu(\theta) = 0, \quad \textit{if } \tau^* = 1,$$

$$\tau^* = \max\left\{1 - \sqrt{\frac{\alpha \int_\Theta \phi_\theta^0(\tau^*)\, d\mu(\theta)}{\int_\Theta (\alpha - w_\theta(t^*,\tau^*)) y_\theta^0(\tau^*)\, d\mu(\theta)}}, 0\right\}, \quad \textit{if } \tau^* < 1.$$

is optimal within the class of affine tax rules.

Proof. We begin by computing the planner's payoff from any affine tax $T(y) = t + \tau y$.

Claim 6. If $T(y) = t + \tau y$ with $\tau \in [0,1)$, then

$$V_P(T) = \int_\Theta \left[\omega(\theta) W(V_w(T|M^0(\theta))) + \alpha \frac{\tau V_w(T|M^0(\theta)) + t}{1-\tau}\right] d\mu(\theta).$$

If $\tau = 1$, then

$$V_P(T) = \int_\Theta \left[\omega(\theta) W(V_w(T|M^0(\theta))) + \alpha \max_{(F,0)\in M^0(\theta)} (\tau \mathbb{E}_F[y] + t)\right] d\mu(\theta).$$

Proof. Suppose first that $\tau = 1$, in which case the consumption rule $y - T(y)$ is constant in y. By *free option* in Assumption 1, we have that $V_w(T|M^0(\theta)) = \max_{y\in Y}\{y - T(y)\} = -t$ for all $\theta \in \Theta$. The result then follows immediately from Case (ii) in Lemma 3.1.

Otherwise for an affine T with $\tau \in [0,1)$, the resulting consumption rule $y - T(y)$ is strictly increasing. As a result, $V_w(T|M^0(\theta)) = \max_{y\in Y}\{y - T(y)\} \iff (\delta_{\overline{y}}, 0) \in M^0(\theta)$. For any such θ, $(\delta_{\overline{y}}, 0)$ is the unique income choice $(F,\phi) \in \Delta(Y) \times \mathbb{R}_+$ satisfying $\mathbb{E}_F[y - T(y)] \geq V_w(T|M^0(\theta))$. Therefore, the distinction between the two cases in Lemma 3.1 is vacuous, and we have

$$V_P(T) = \int_\Theta [\omega(\theta) W(V_w(T|M^0(\theta)) + \alpha r^\theta(T)]\, d\mu(\theta),$$

where $r^\theta(T)$ is defined by (A.1).

Next, we show that $r^\theta(T) = \frac{\tau V_w(T|M^0(\theta))-t}{1-\tau}$ for all $\tau \in [0,1)$ and $\theta \in \Theta$. The result is immediate if $\tau = 0$, since then revenue is constant and equal to $-t$. Next, suppose that $\tau \in (0,1)$. We argue that, for any $F_\theta \in \Delta(Y)$ that is a solution for $r^\theta(T)$, it must be that $\mathbb{E}_{F_\theta}[y - T(y)] = V_w(T|M^0(\theta))$. Suppose towards a contradiction that F_θ is a solution for $r^\theta(T)$ and that $\mathbb{E}_{F_\theta}[y - T(y)] > V_w(T|M^0(\theta))$. Since $V_w(T|M^0(\theta)) \geq -T(0)$ and $y - T(y)$ is strictly increasing, it cannot be that F_θ assigns probability one to $y = 0$. Then, for $\eta \in (0,1)$ arbitrarily small, consider $\tilde{F}_\theta = (1-\eta)F_\theta + \eta\delta_0$. If η is sufficiently small, $\mathbb{E}_{\tilde{F}_\theta}[y - T(y)] \geq V_w(T|M^0(\theta))$. Moreover, since $T(y)$ is strictly increasing and F_θ is not degenerate on $y = 0$, we have that $\mathbb{E}_{\tilde{F}_\theta}[T(y)] < \mathbb{E}_{F_\theta}[T(y)]$, contradicting F_θ being revenue-minimizing. Thus,

$$r^\theta(T) = \mathbb{E}_{F_\theta}[T(y)] = \frac{\tau\mathbb{E}_{F_\theta}[y - T(y)] + t}{1-\tau} = \frac{\tau V_w(T|M^0(\theta)) + t}{1-\tau}, \forall \theta \in \Theta,$$

which establishes the Claim. □

Claim 6 implies that the planner's objective under an affine tax with rate $\tau \in [0,1]$ can be written as

$$V_P(T) = \int_\Theta \max_{(F_\theta,\phi_\theta)\in M^0(\theta)} \left[\omega(\theta)W(-t + (1-\tau)\mathbb{E}_{F_\theta}[y] - \phi_\theta) + \alpha\frac{\tau(-t + (1-\tau)\mathbb{E}_{F_\theta}[y] - \phi_\theta) + t}{1-\tau}\right] d\mu(\theta), \tag{A.15}$$

with the interpretation that, if $\tau = 1$, then $\int_\Theta \frac{\tau(-t+(1-\tau)\mathbb{E}_{F_\theta}[y]-\phi_\theta)+t}{1-\tau}\, d\mu(\theta) = t$ if $\int_\Theta \phi_\theta\, d\mu(\theta) = 0$, and $\int_\Theta \frac{\tau(-t+(1-\tau)\mathbb{E}_{F_\theta}[y]-\phi_\theta)+t}{1-\tau}\, d\mu(\theta) = -\infty$ if $\int_\Theta \phi_\theta\, d\mu(\theta) > 0$.

By Lemma 2, (t^*, τ^*) must maximize (A.15) on $[-\overline{C}, 0] \times [0,1]$. Since the planner's objective is continuous in (t,τ), the maximum exists. Given any profile $\mathbf{C} = \{(F^0_\theta, \phi^0_\theta)\}_{\theta\in\Theta}$ such that $(F^0_\theta, \phi^0_\theta) \in M^0(\theta)$ for all θ and setting $y^0_\theta = \mathbb{E}_{F^0_\theta}[y]$, the optimal $t(\mathbf{C})$ and $\tau(\mathbf{C})$ solve the following first-order necessary conditions

$$\int_\Theta [w_\theta(t(\mathbf{C}), \tau(\mathbf{C})) - \alpha]\, d\mu(\theta) \leq 0, \quad \text{with equality if } t < 0, \tag{A.16}$$

$$\int_\Theta \left((w_\theta(t(\mathbf{C}), \tau(\mathbf{C})) - \alpha)y^0_\theta + \alpha\frac{\phi^0_\theta}{(1-\tau(\mathbf{C}))^2}\right) d\mu(\theta) \geq 0, \text{ if } \tau(\mathbf{C}) \in [0,1), \text{ with equality if } \tau(\mathbf{C}) > 0, \tag{A.17}$$

$$\int_\Theta (w_\theta(t(\mathbf{C}), \tau(\mathbf{C})) - \alpha)y^0_\theta\, d\mu(\theta) \leq 0 \text{ and } \int_\Theta \phi^0_\theta\, d\mu(\theta) = 0, \forall\theta \in \Theta, \text{ if } \tau(\mathbf{C}) = 1. \tag{A.18}$$

Concavity of W ensures that the above conditions are also sufficient.

Given this, the profile $\mathbf{C}$ is chosen optimally so as to solve

$$\max_{\mathbf{C}} \int_{\Theta} \left[\omega(\theta) W(-t(\mathbf{C}) + (1-\tau(\mathbf{C}))\mathbb{E}_{F_\theta}[y] - \phi_\theta) + \alpha \frac{\tau(\mathbf{C})(-t(\mathbf{C}) + (1-\tau(\mathbf{C}))\mathbb{E}_{F_\theta}[y] - \phi_\theta) + t(\mathbf{C})}{1-\tau(\mathbf{C})} \right] d\mu(\theta),$$

$$s.t. \quad (F_\theta, \phi_\theta) \in M^0(\theta) \quad \forall \theta, \quad \mathbf{C} = \{(F_\theta, \phi_\theta)\}_{\theta \in \Theta}. \tag{A.19}$$

Conditions (A.16) to (A.19) describe the optimal affine tax rule, and imply the description in Lemma 4. □

Next, suppose that the tax rule $T_a(y)$ from Lemma 4 is optimal. Let w_θ be type-θ's social welfare weight under $T_a(y)$ as defined above (dropping the arguments in the function for conciseness). Let $\mathcal{F}_\theta^0 \subseteq \Delta(Y)$ be the set of income distributions chosen by type θ under T_a and $M^0(\theta)$ and pick $F_\theta^0 \in \mathcal{F}_\theta^0$, and let $\underline{\mathcal{F}}_\theta \subseteq \Delta(Y)$ be his revenue minimizing income choices as defined in Lemma 1.

Suppose that $\tau^* < 1$. It has to be the case that T_a is not strictly dominated by any feasible perturbation of the form

$$T^\varepsilon(y) \equiv T(y) + \varepsilon D(y),$$

where $\varepsilon \in \mathbb{R}$ and $D(y)$ is a continuous function. In Appendix B, we show that $V_P(T^\varepsilon)$ is directionally differentiable at $\varepsilon = 0$.

Here, we consider a perturbation with $D(y) = -\min\{\tilde{y} - y, 0\}$ with $\tilde{y} > 0$. Observe that $T^\varepsilon(y)$ is strongly progressive. The fact that $\tilde{y} > 0$ ensures that non-negativity of consumption is still satisfied under T^ε for $\varepsilon > 0$ sufficiently small, and thus T^ε is feasible. Applying Lemma OA.1, we then have the following necessary condition for optimality of the affine tax rule

$$\left. \frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +} \right|_{\varepsilon=0} =$$

$$\int_{\Theta} \left[(w_\theta - \alpha) \max_{F \in \mathcal{F}_\theta^0} \mathbb{E}_F[\min\{\tilde{y} - y, 0\}] + \alpha \frac{\max_{F \in \mathcal{F}_\theta^0} \mathbb{E}_F[\min\{\tilde{y} - y, 0\}] - \min\{\tilde{y} - y_\theta, 0\}}{1 - \tau^*} \right] d\mu(\theta) \leq 0.$$

The derivative is bounded below by

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} \geq \int_\Theta \left[(w_\theta - \alpha)\mathbb{E}_{F_\theta^0}[\min\{\tilde{y}-y,0\}] + \alpha \frac{\mathbb{E}_{F_\theta^0}[\min\{\tilde{y}-y,0\}] - \min\{\tilde{y}-y_\theta,0\}}{1-\tau^*}\right] d\mu(\theta)$$
$$= \int_\Theta \left[\left(w_\theta + \alpha\frac{\tau^*}{1-\tau^*}\right)\mathbb{E}_{F_\theta^0}[\min\{\tilde{y}-y,0\}] - \alpha\frac{\min\{\tilde{y}-y_\theta^0+\phi_\theta^0/(1-\tau^*),0\}}{1-\tau^*}\right] d\mu(\theta)$$
$$\geq \int_\Theta \left[\left(w_\theta + \alpha\frac{\tau^*}{1-\tau^*}\right)\mathbb{E}_{F_\theta^0}[\min\{\tilde{y}-y,0\}] - \alpha\mathbf{I}(y_\theta^0 > \tilde{y})\frac{\tilde{y}-y_\theta^0+\phi_\theta^0/(1-\tau^*)}{1-\tau^*}\right] d\mu(\theta).$$

Then, if condition (3.5) is satisfied, we have that $\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} > 0$, which contradicts optimality of T_a.

Proof of Proposition 3

Let $T \in \mathbf{T}$ be an optimal tax, and write $c(y) = y - T(y)$. Let $w_\theta = \omega(\theta)W'(V_w(T|M^0(\theta))$ be the endogenous social welfare weight of type θ. Let $y_\theta = \min\{y : y - T(y) = V_w(T|M^0(\theta))\}$, and let $y_\theta^0 = \arg\max_{y\in Y}\{c(y) - \Phi(y,\theta)\}$.

Affineness at the tails follows from the proof of Theorem 1. Specifically, if T is not affine to the left of $y_{\underline{\theta}}$, we can substitute it by $\overline{T}$ as defined in the proof of Theorem 1 for a weak improvement. Below, we will further show that in fact the optimal tax is affine to the left of $y_l > y_{\underline{\theta}}$.

Step 1: optimal $T(0)$. As in the proof of Lemma 2, to rule out that the social planner strictly benefits from shifting $T(y)$ by a constant amount, it must be that

$$\int_\Theta (w_\theta - \alpha)\, d\mu(\theta) \leq 0, \quad \text{with equality if } T(0) < 0. \tag{A.20}$$

Step 2: full taxation is suboptimal. We now rule out $T'(0+) = 1$ as being optimal. Suppose towards a contradiction that $T'(0+) = 1$ is optimal. Then, $y_\theta^0 = 0$ for all θ and total revenue is equal to $T(0)$. If the planner instead uses the tax rule $\tilde{T}(y) = T(0) + (1-\varepsilon)y$, where $\varepsilon > 0$ is arbitrarily small, workers' welfare strictly increases. In particular, our no-bunching-at-zero assumption implies that $y_\theta^0(\tilde{T}) > 0$ for all $\theta \in \Theta$. Moreover, applying Claim 6, total revenue under this tax rule is

$$\frac{(1-\varepsilon)V_w(\tilde{T}|M^0(\theta)) + T(0)}{\varepsilon} > T(0),$$

where the inequality follows from the fact that $V_w(\tilde{T}|M^0(\theta)) > -T(0)$. This contradicts

optimality of $T(y)$.

Therefore, $c(y)$ is not everywhere flat, and workers' income choices under $M^0(\theta)$ are strictly positive. By concavity of c, there exists $\tilde{y} \in Y$ such that $c(y)$ is strictly increasing if and only if $y \leq \tilde{y}$. Observe that workers' optimality then requires that $y^0_\theta \in (0, \tilde{y}]$ for all $\theta \in \Theta$. As a consequence of $y^0_\theta > 0$, we have that $\Phi(y^0_\theta, \theta) > 0$ for all θ, and we can apply Lemma 9 to differentiate the planner's objective with respect to the type of perturbations defined in Appendix B. We now use this approach to derive necessary conditions that have to be satisfied by an optimal $T(y)$.

Step 3: optimal bottom rate. Let $\underline{\lambda}$ be the consumption rate at the bottom, i.e. $c'(y_{\underline{\theta}}-) = \underline{\lambda}$, and let $y_l \geq y_{\underline{\theta}} > 0$ be the highest $y \in Y$ such that $c'(y-) = \underline{\lambda}$. We use the following lemma.

Lemma 5. *If T is optimal and $\underline{\lambda} < 1$, then it is differentiable at y_θ almost everywhere in θ.*

Proof. Suppose that T is optimal. Consider the perturbation $T^\varepsilon(y) = T(y) - \varepsilon y$. By Step 2 and concavity of c, it holds that $c'(0+) > 0$, and therefore this perturbation does not violate non-negativity of consumption if $|\varepsilon| < c'(0+)$. Applying again the results in Section B, the principal's objective is directionally differentiable with respect to this perturbation. Then, Lemma OA.2 gives the following necessary condition for optimality:[47]

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} = \int_\Theta \left[(w_\theta - \alpha) y^0_\theta + \alpha \frac{y^0_\theta - y_\theta}{c'(y_\theta +)}\right] d\mu(\theta) \leq 0, \tag{A.21}$$

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon -}\right|_{\varepsilon=0} = \int_\Theta \left[(w_\theta - \alpha) y^0_\theta + \alpha \frac{y^0_\theta - y_\theta}{c'(y_\theta -)}\right] d\mu(\theta) \geq 0, \tag{A.22}$$

Given that c is concave and that $\Phi(y^0_\theta, \theta) > 0$ implies that $y^0_\theta > y_\theta$ for all $\theta \in \Theta$, the two conditions can hold only if $c'(y_\theta +) = c'(y_\theta -)$ almost everywhere in θ. □

Suppose first that $\underline{\lambda} < 1$, and consider the perturbation $T^\varepsilon(y) = T(y) - \varepsilon \min\{y, \tilde{y}\}$ with $\tilde{y} \in (y_{\underline{\theta}}, y^0_{\underline{\theta}})$ and $\varepsilon > 0$. Applying the results in Appendix B and Lemma OA.3, this perturbation gives the following necessary condition for optimality of T

$$\left.\frac{\partial V_P(T^\varepsilon)}{\varepsilon +}\right|_{\varepsilon=0} = \int_\Theta \left((w_\theta - \alpha)\tilde{y} + \alpha \frac{\tilde{y} - \min\{y_\theta, \tilde{y}\}}{c'(y_\theta)}\right) d\mu(\theta) \leq 0, \tag{A.23}$$

Observe that the above is strictly positive if (A.20) holds with equality. Therefore, $\underline{\lambda} = 1$ and $y_l \geq y^0_{\underline{\theta}}$ when $\int_\Theta w_\theta \, d\mu(\theta) - \alpha = 0$, as stated in Proposition 3. More generally, $\underline{\lambda} = 1$ if (A.23) is violated for some $\tilde{y} \in (y_{\underline{\theta}}, y^0_{\underline{\theta}})$.

[47] The fact that $\underline{\lambda} < 1$ implies that $\hat{y} = 0$.

Henceforth, we focus on deriving $\underline{\lambda}$ in the case in which (A.23) is satisfied for all $\tilde{y} \in (y_{\underline{\theta}}, y^0_{\underline{\theta}})$, which in turn implies that $\int_\Theta w_\theta \, d\mu(\theta) - \alpha < 0$. To do that, we now consider the perturbation $T^\varepsilon(y) = T(y) - \varepsilon \min\{y, y_l\}$, which is feasible (i.e., satisfies non-negativity of consumption) if ε is sufficiently small. If $\underline{\lambda} < 1$, by Lemmas 5 and OA.4, the objective is differentiable with respect to this perturbation at $\varepsilon = 0$ with

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} = \int_\Theta \left((w_\theta - \alpha) \min\{y^0_\theta, y_l\} + \alpha \frac{\min\{y^0_\theta, y_l\} - \min\{y_\theta, y_l\}}{c'(y_\theta)} \right) d\mu(\theta) =$$

$$\int_\Theta (w_\theta - \alpha) \min\{y^0_\theta, y_l\} \, d\mu(\theta) + \alpha \int_\Theta \mathbb{I}(y_\theta \leq y_l) \frac{\Phi(y^0_\theta, \theta) - \max\left\{0, \int_{y_l}^{y^0_\theta} c'(y)\, dy\right\}}{\underline{\lambda}^2} \, d\mu(\theta), \tag{A.24}$$

where the second equality follows from the definition of y_θ. Thus, a necessary condition for optimality of T is that $\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} = 0$. This in turn implies that $y_l > y_{\underline{\theta}}$: Otherwise, $y_l = y_{\underline{\theta}}$ and (A.24) becomes

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} = \int_\Theta (w_\theta - \alpha) \min\{y^0_\theta, y_l\} \, d\mu(\theta) \leq \int_\Theta (w_\theta - \alpha) \, d\mu(\theta) \int_\Theta \min\{y^0_\theta, y_l\} \, d\mu(\theta) < 0,$$

in contradiction with optimality of T.

Moreover, observe that, under the maintained assumption that $\int_\Theta w_\theta \, d\mu(\theta) - \alpha < 0$, it holds that

$$\int_\Theta (w_\theta - \alpha) \min\{y^0_\theta, y_l\} \, d\mu(\theta) \leq \int_\Theta (w_\theta - \alpha) \, d\mu(\theta) \int_\Theta \min\{y^0_\theta, y_l\} \, d\mu(\theta) < 0,$$

where the first inequality follows from the fact that w_θ is non-increasing in θ and $\min\{y^0_\theta, y_l\}$ is non-decreasing in θ. Combining everything, we obtain the following necessary condition for optimality when $\underline{\lambda} < 1$

$$\underline{\lambda} = \sqrt{\frac{\alpha \int_\Theta \mathbb{I}(y_\theta \leq y_l) \left[\Phi(y^0_\theta, \theta) - \max\left\{0, \int_{y_l}^{y^0_\theta} c'(y)\, dy\right\}\right] d\mu(\theta)}{\int_\Theta (\alpha - w_\theta) \min\{y^0_\theta, y_l\} \, d\mu(\theta)}} \tag{A.25}$$

$$\implies \underline{\lambda} \leq \sqrt{\alpha \frac{\int_\Theta \mathbb{I}(y_\theta \leq y_l) \Phi(y^0_\theta, \theta) \, d\mu(\theta)}{\int_\Theta (\alpha - w_\theta) \min\{y^0_\theta, y_l\} \, d\mu(\theta)}}$$

To obtain the lower bound on $\underline{\lambda}$, consider the perturbation $T^\varepsilon(y) = T(y) - \varepsilon y$. Applying (A.21) and (A.22), we obtain the following necessary condition for optimality of T:

$$\frac{\partial V_P(T^\varepsilon)}{\varepsilon+} = \frac{\partial V_P(T^\varepsilon)}{\varepsilon-} = \int_\Theta \left[(w_\theta - \alpha) y^0_\theta + \alpha \frac{y^0_\theta - y_\theta}{c'(y_\theta)} \right] d\mu(\theta) = 0 \tag{A.26}$$

Recall that the definition of y_θ and concavity of c imply that almost everywhere in θ:

$$\Phi(y^0_\theta, \theta) = \int_{y_\theta}^{y^0_\theta} c'(y)\, dy \leq c'(y_\theta)(y^0_\theta - y_\theta).$$

Thus, the following weaker condition must be satisfied

$$\int_\Theta \left[(w_\theta - \alpha) y^0_\theta + \alpha \frac{\Phi(y^0_\theta, \theta)}{(c'(y_\theta))^2} \right] d\mu(\theta) \leq 0,$$

which by concavity of c in turn requires that

$$\int_\Theta \left[(w_\theta - \alpha) y^0_\theta + \alpha \frac{\Phi(y^0_\theta, \theta)}{\underline{\lambda}^2} \right] d\mu(\theta) \leq 0 \iff \underline{\lambda} \geq \sqrt{\frac{\alpha \int_\Theta \Phi(\theta, y^0_\theta)\, d\mu(\theta)}{\int_\Theta (\alpha - w_\theta) y^0_\theta\, d\mu(\theta)}}.$$

Proof of Proposition 4

By an analogous argument to the one in Proposition 3, there is an optimal tax rule which is affine to the right of $y_u \leq y_{\overline{\theta}}$. The claim that $y_{\overline{\theta}} < y^0_{\overline{\theta}}$ follows from the fact that, as shown in Proposition 3, $y^0_{\overline{\theta}} > 0$, and therefore $\Phi(y^0_{\overline{\theta}}, \theta) > 0$ for all $\theta \in \Theta$.

Let $\overline{\tau} \equiv T'(y_{\overline{\theta}}+)$, and let y_u be the lowest $y \in Y$ such that $T'(y+) = \overline{\tau}$. Put $\overline{\lambda} \equiv 1 - \overline{\tau}$. We proceed again by studying the effects of certain perturbations around the optimal tax rule. Consider first the perturbation $T^\varepsilon(y) = T(y) + \varepsilon \max\{y - y_u, 0\}$. According to Lemma OA.5, if $\overline{\lambda} < \underline{\lambda}$, the following condition is necessary for optimality of T

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} = \int_\Theta \mathbb{I}(y^0_\theta > y_u) \left[-(w_\theta - \alpha)(y^0_\theta - y_u) - \alpha \frac{(y^0_\theta - y_u) - \max\{y_\theta - y_u, 0\}}{1 - T'(y_\theta+)} \right] d\mu(\theta) \leq 0, \tag{A.27}$$

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon-}\right|_{\varepsilon=0} = \int_\Theta \mathbb{I}(y^0_\theta > y_u) \left[-(w_\theta - \alpha)(y^0_\theta - y_u) - \alpha \frac{(y^0_\theta - y_u) - \max\{y_\theta - y_u, 0\}}{1 - T'(y_\theta-)} \right] d\mu(\theta) \geq 0. \tag{A.28}$$

We begin by noting that

$$\int_\Theta -\mathbb{I}(y_\theta^0 > y_u)(w_\theta - \alpha)(y_\theta^0 - y_u)\, d\mu(\theta) \geq \int_\Theta -(w_\theta - \alpha)\, d\mu(\theta) \int_\Theta \mathbb{I}(y_\theta^0 > y_u)(y_\theta^0 - y_u)\, d\mu(\theta) \geq 0.$$

If equality holds, then (A.28) is strictly negative, and thus it must be that $\underline{\lambda} = \overline{\lambda}$, i.e. the optimal tax is affine. In that case, the characterization from Proposition 2 applies.

From now on, we focus on the case with $\int_\Theta \mathbb{I}(y_\theta^0 > y_u)(w_\theta - \alpha)(y_\theta^0 - y_u)\, d\mu(\theta) < 0$. (A.28) gives the following necessary condition for optimality of $\overline{\lambda} < \underline{\lambda}$

$$\int_\Theta \mathbb{I}(y_\theta^0 > y_u)\left[-(w_\theta - \alpha)(y_\theta^0 - y_u) - \alpha \frac{\mathbb{I}(y_\theta > y_u)\Phi(y_\theta^0, \theta)}{\overline{\lambda}^2} \right] d\mu(\theta) \geq 0$$
$$\iff \overline{\lambda} \geq \sqrt{\frac{\alpha \int_\Theta \mathbb{I}(y_\theta > y_u)\Phi(y_\theta^0, \theta)\, d\mu(\theta)}{\int_\Theta \mathbb{I}(y_\theta^0 > y_u)(\alpha - w_\theta)(y_\theta^0 - y_u)\, d\mu(\theta)}}.$$

To derive the upper bound, let θ^* be the smallest $\theta \in \Theta$ such that $y_\theta \geq y_u$, which is well defined given that $y_u \leq y_{\overline{\theta}}$. Consider now the perturbation $T^\varepsilon(y) = T(y) + \varepsilon \max\{y - y_{\theta^*}^0, 0\}$. By definition, $y_{\theta^*}^0 > y_u$. According to Appendix B and Lemma OA.6, the planner's objective is directionally differentiable with respect to this perturbation, with the following condition being necessary for optimality

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon = 0} = \int_\Theta \mathbb{I}(y_\theta > y_u)\left[-(w_\theta - \alpha)(y_\theta^0 - y_{\theta^*}^0) - \alpha \frac{(y_\theta^0 - y_{\theta^*}^0) - \max\{y_\theta - y_{\theta^*}^0, 0\}}{\overline{\lambda}} \right] d\mu(\theta) \leq 0,$$

which in turn requires that the following weaker condition is satisfied

$$\int_\Theta \mathbb{I}(y_\theta > y_u)\left[-(w_\theta - \alpha)(y_\theta^0 - y_{\theta^*}^0) - \alpha \frac{y_\theta^0 - y_\theta}{\overline{\lambda}} \right] d\mu(\theta) =$$
$$\int_\Theta \mathbb{I}(y_\theta > y_u)\left[-(w_\theta - \alpha)(y_\theta^0 - y_{\theta^*}^0) - \alpha \frac{\Phi(y_\theta^0, \theta)}{\overline{\lambda}^2} \right] d\mu(\theta) \leq 0$$
$$\iff \overline{\lambda} \leq \sqrt{\frac{\alpha \int_\Theta \mathbb{I}(y_\theta > y_u)\Phi(y_\theta^0, \theta)\, d\mu(\theta)}{\int_\Theta \mathbb{I}(y_\theta > y_u)(\alpha - w_\theta)(y_\theta^0 - y_{\theta^*}^0)\, d\mu(\theta)}}.$$

Proof of Proposition 5

We use the following lemma.

Lemma 6. *Under the Assumptions of Section 4.1,*

$$\max_{T\in\mathbf{T}} V_P(T) =$$
$$\max_{y_1^0<y_2^0}\Bigg\{\max_{(\tau_1,\tau_2)\in[0,1]^2,\tilde{y}\in Y}\Bigg\{p_1\left(\omega(\theta_1)-\alpha+\frac{\alpha}{1-\tau_1}\right)[(1-\tau_1)y_1^0-\Phi(y_1^0,\theta_1)]$$
$$+p_2\left[\omega(\theta_2)((1-\tau_1)\tilde{y}+(1-\tau_2)(y_2^0-\tilde{y})-\Phi(y_2^0,\theta_2))+\alpha\left(\left(y_2^0-\frac{\Phi(y_2^0,\theta_2)}{1-\tau_2}\right)\tau_2-(\tau_2-\tau_1)\tilde{y}\right)\right],$$
$$s.t.\quad 0\le\tau_1\le\tau_2<1,\quad \textit{and } \tau_1<\tau_2 \implies y_1^0\le\tilde{y}\le y_2^0-\Phi(y_2^0,\theta_2)/(1-\tau_2)\}\Bigg\}\Bigg\}$$

Proof. For any feasible tax rule T, Lemma 1 implies that the planner's objective is non-decreasing in $V_w(T|M^0(\theta))$ for all $\theta\in\Theta$. Therefore, the worker and the social planner are aligned in terms of the optimal income choice under T and $M^0(\theta)$, and we can write

$$V_P(T)=\max_{y_i^0\in Y}\sum_{i=1}^{2}p_i[\omega(\theta_i)(y_i^0-T(y_i^0)-\Phi(y_i^0,\theta_i))+\alpha r^{\theta_i}(T;y_i^0)],$$

with

$$r^{\theta_i}(T;y_i^0)=\min_{F\in\Delta(Y)}\mathbb{E}_F[T(y)],\quad s.t.\quad \mathbb{E}_F[y-T(y)]\ge y_i^0-T(y_i^0)-\Phi(y_i^0,\theta_i). \tag{A.29}$$

As a result, we can take (y_1^0,y_2^0) as fixed and solve for the optimal tax rule (the inner problem in Lemma 6), and as a second step optimize over (y_1^0,y_2^0) (the outer problem).

Under binary types, the dominating consumption rule $\bar{c}(y)$ defined in (A.9) is piecewise affine with at most one point of discontinuity in its derivative. Therefore, the problem can be reduced to one of finding the optimal intercept of the tax rule $T(0)$, the two relevant marginal tax rates (τ_1,τ_2), and the location of the kink $\tilde{y}$. (A.9) also implies that it is without loss to restrict to $0\le\tau_1\le\tau_2$, and Proposition 4 implies that $\tau_2<1$. As a result, at the optimum, $y_i^0>0$ for all $i\in\{1,2\}$ and, by single-crossing of Φ, it holds that $y_1^0\le y_2^0$. On the other hand, the assumption that $\mathbb{E}[\omega(\theta)]\le\alpha$ implies that either the objective is constant in $T(0)$ (if equality holds), or that it is strictly increasing and therefore at the optimum $T(0)=0$ (if the inequality is strict). Either way, at the optimum, $(\mathbb{E}[\omega(\theta)]-\alpha)T(0)=0$.

Once we restrict attention to T being convex, it follows that $r^{\theta_i}(T;y_i^0)$ is attained by a deterministic income, which we denote by $y_i(y_i^0)$, which satisfies

$$y_i(y_i^0)-T(y_i(y_i^0))=y_i^0-T(y_i^0)-\Phi(y_i^0,\theta_i).$$

Further, without loss, if $\tau_1 < \tau_2$ it holds that $y_1(y_1^0) \leq \tilde{y} < y_2(y_2^0)$. If not, an analogous argument to Theorem 1 shows that T is (weakly) dominated by an affine tax rule.

Given the above simplifications, for a fixed $(y_1^0, y_2^0) \in Y^2$ satisfying $y_1^0 \leq y_2^0$, the planner's problem can be written as

$$\max_{\tau_1,\tau_2,\tilde{y}} \Bigg\{ p_1 \left(\omega(\theta_1) - \alpha + \frac{\alpha}{1-\tau_1} \right) [\mathbb{I}(\tilde{y} \geq y_1^0)(1-\tau_1)y_1^0 + \mathbb{I}(\tilde{y} < y_1^0)((1-\tau_1)\tilde{y} + (1-\tau_2)(y_1^0 - \tilde{y})) - \Phi(y_1^0, \theta_1)]$$
$$+ p_2 \left[(\omega(\theta_2) - \alpha)((1-\tau_1)\tilde{y} + (1-\tau_2)(y_2^0 - \tilde{y}) - \Phi(y_2^0, \theta_2)) + \alpha \left(y_2^0 - \frac{\Phi(y_2^0, \theta_2)}{1-\tau_2} \right) \right] \Bigg\},$$
$$s.t. \quad \tau_1 \leq \tau_2 < 1, \quad \tau_1 < \tau_2 \implies \tilde{y} \leq y_2^0 - \Phi(y_2^0, \theta_2)/(1-\tau_2).$$

Finally, we argue that without loss of optimality, $y_1^0 \leq \tilde{y}$ whenever the tax rule satisfies $\tau_1 < \tau_2$. This follows from the fact that $\omega(\theta_2) \leq \alpha$, which implies that whenever $\tau_1 < \tau_2$, the planner's objective is weakly decreasing in $\tilde{y} \in [y_1^0, \overline{y}]$. As a consequence, it also holds that $y_1^0 \leq y_2^0 - \Phi(y_2^0, \theta_2)/(1-\tau_2)$ whenever $\tau_1 < \tau_2$, and therefore $y_1^0 < y_2^0$. If otherwise $\tau_1 = \tau_2 < 1$, strict single-crossing of Φ together with the fact that the tax rule is smooth implies that $y_1^0 < y_2^0$.

Combining these results, we obtain the program in Lemma 6. □

We now turn to solving the program in Lemma 6. First, we find the optimal $\tilde{y} \in Y$, for a given (τ_1, τ_2). If $\tau_2 < \tau_1$, the objective is non-increasing in $\tilde{y}$ on $[y_1^0, y_2^0 - \Phi(y_2^0, \theta_2)/(1-\tau_2)]$, and therefore $\tilde{y} = y_1^0$ is optimal. If $\tau_2 = \tau_1$, then the objective is constant in $\tilde{y}$, so we can without loss set $\tilde{y} = y_1^0$ as well. Therefore, the problem becomes

$$\max_{y_1^0 < y_2^0, 0 \leq \tau_1 \leq \tau_2 < 1} \Bigg\{ p_1 \left(\omega(\theta_1) - \alpha + \frac{\alpha}{1-\tau_1} \right) [(1-\tau_1)y_1^0 - \Phi(y_1^0, \theta_1)]$$
$$+ p_2 \left[\omega(\theta_2)((1-\tau_1)y_1^0 + (1-\tau_2)(y_2^0 - y_1^0) - \Phi(y_2^0, \theta_2)) + \alpha \left(\left(y_2^0 - \frac{\Phi(y_2^0, \theta_2)}{1-\tau_2} \right) \tau_2 - (\tau_2 - \tau_1)y_1^0 \right) \right] \Bigg\}. \tag{O_{bin}}$$

The maximum exists due to the objective being continuous and the constraint set compact.[48] We next show that the maximizer is unique.

Claim 7. The arg max in (O_{bin}) is unique.

Proof. Suppose, toward a contradiction, that $(y_1^0, y_2^0, \tau_1, \tau_2)$ and $(y_1^{0\prime}, y_2^{0\prime}, \tau_1', \tau_2')$ are optimal with $(y_1^0, y_2^0, \tau_1, \tau_2) \neq (y_1^{0\prime}, y_2^{0\prime}, \tau_1', \tau_2')$. If $\tau_i \neq \tau_i'$ for some i, take $\beta \in (0,1)$ and consider

[48] To ensure compactness, the strict inequalities in the constraints can be substituted by weak inequalities without affecting the solution to the problem.

choosing $(y_1^0, y_2^0, \beta\tau_1 + (1-\beta)\tau_1', \beta\tau_2 + (1-\beta)\tau_2')$ which is feasible for (O_{bin}). One can readily check that, by strict concavity of the objective with respect to τ_i for each $i = 1, 2$, it holds that $(y_1^0, y_2^0, \beta\tau_1 + (1-\beta)\tau_1', \beta\tau_2 + (1-\beta)\tau_2')$ gives a strictly higher value than $(y_1^0, y_2^0, \tau_1, \tau_2)$, contradicting optimality of the latter. An analogous argument applies if $\tau_i = \tau_i'$ for all $i = 1, 2$, and $y_i^0 \neq y_i^{0\prime}$ for some $i = 1, 2$. □

Let $\gamma \in \mathbb{R}_+$ be the Lagrange multiplier associated with the constraint $\tau_2 - \tau_1 \geq 0$. The first-order conditions are

$$-\mathbb{E}[\omega(\theta) - \alpha]y_1^0 - p_1 \frac{\alpha\Phi(y_1^0, \theta_1)}{(1-\tau_1)^2} - \gamma \leq 0, \quad = 0 \text{ if } \tau_1 > 0, \tag{A.30}$$

$$-p_2(\omega(\theta_2) - \alpha)(y_2^0 - y_1^0) - p_2 \frac{\alpha\Phi(y_2^0, \theta_2)}{(1-\tau_2)^2} + \gamma = 0, \tag{A.31}$$

$$p_1\left(\omega(\theta_1) - \alpha + \frac{\alpha}{1-\tau_1}\right)(1 - \tau_1 - \Phi_y(y_1^0, \theta_1)) + p_2(\omega(\theta_2) - \alpha)(\tau_2 - \tau_1) = 0 \tag{A.32}$$

$$1 - \tau_2 - \Phi_y(y_2^0, \theta_2) \geq 0, \quad = 0 \text{ if } y_2^0 < \overline{y}. \tag{A.33}$$

If $\omega(\theta_2) = \alpha$, then under our assumptions it must be that $\omega(\theta_1) = \omega(\theta_2) = \alpha$. (A.30) and (A.31) then imply that $\tau_1 = \tau_2 = 0$. Next suppose that $\omega(\theta_2) < \alpha$. We consider the cases with $\tau_1 < \tau_2$ and with $\tau_1 = \tau_2$ separately.

Case 1: $\tau_1 = \tau_2 \equiv \tau$. In this case, we have

$$\tau = \begin{cases} 0, & \text{if } \mathbb{E}[\omega(\theta)] = \alpha, \\ \max\left\{0, 1 - \sqrt{\frac{\alpha[p_1\Phi(y_1^0,\theta_1) + p_2\Phi(y_2^0,\theta_2)]}{p_1(\alpha - \omega(\theta_1))y_1^0 + p_2(\alpha - \omega(\theta_2))y_2^0}}\right\}, & \text{if } \mathbb{E}[\omega(\theta)] < \alpha, \end{cases} \tag{A.34}$$

where we have used the fact that monotonicity of $\omega(\theta_i)$ and y_i^0 with respect to i implies that $p_1(\alpha - \omega(\theta_1))y_1^0 + p_2(\alpha - \omega(\theta_2))y_2^0 \geq (\alpha - \mathbb{E}[\omega(\theta)])(p_1 y_1^0 + p_2 y_2^0) > 0$ if $\alpha > \mathbb{E}[\omega(\theta)]$, and therefore τ is well-defined.

Let $y_i^0(\tau)$ be such that $\Phi_y(y_i^0(\tau), \theta_i) = 1 - \tau$. If (and only if)

$$(\omega(\theta_2) - \alpha)(y_2^0(\tau) - y_1^0(\tau)) + \frac{\alpha\Phi(y_2^0(\tau), \theta_2)}{(1-\tau)^2} \geq 0, \tag{A.35}$$

then the solution to (A.30)-(A.33) satisfies $\gamma \geq 0$, and thus $\tau_1 = \tau_2 = \tau$. If (A.35) does not hold, then the optimum must feature $\tau_1 < \tau_2$, which is the case that we analyze next.

Case 2: $\tau_1 < \tau_2$. In this case, $\gamma = 0$ and (A.30) and (A.31) give:

$$\tau_1 = \begin{cases} 0, & \text{if } \mathbb{E}[\omega(\theta)] = \alpha, \\ \max\left\{0, 1 - \sqrt{\frac{p_1 \alpha \Phi(y_1^0, \theta_1)}{(\alpha - \mathbb{E}[\omega(\theta)]) y_1^0}}\right\}, & \text{if } \mathbb{E}[\omega(\theta)] < \alpha, \end{cases} \tag{A.36}$$

$$\tau_2 = 1 - \sqrt{\frac{\alpha \Phi(y_2^0, \theta_2)}{(\alpha - \omega(\theta_2))(y_2^0 - y_1^0)}}, \tag{A.37}$$

which, together with equations (A.32) and (A.33), give the unique solution for the optimal $(y_1^0, y_2^0, \tau_1, \tau_2)$ when (A.35) is violated.

Proof of Proposition 6

Fix a menu of tax rules $\mathcal{T}$. Let $T_\theta^0 \in \mathcal{T}$ and $(F_\theta^0, \phi_\theta^0) \in M^0(\theta)$ be type θ's optimal choice (of tax and income) under the menu $\mathcal{T}$ and the baseline technology M^0. Let $\mathcal{T}_0(\mathcal{T}) = \bigcup_{\theta \in \Theta} \{T_\theta^0\}$. We will show that $V_P(\mathcal{T}_0(\mathcal{T})) \geq V_P(\mathcal{T})$.

To that end, fix $M \in \mathcal{M}$, and let $(F_\theta, \phi_\theta) \in M(\theta)$ and $T_\theta \in \mathcal{T}_0(\mathcal{T})$ be type θ's choice when facing $M(\theta)$ and $\mathcal{T}_0(\mathcal{T})$. Put $T'_\theta \in \arg\min_{T \in \mathcal{T}} \mathbb{E}_{F_\theta}[T(y)]$. Consider the following correspondence $M' \in \mathcal{M}$

$$M'(\theta) \equiv M^0(\theta) \cup \left(\bigcup_{\theta' \leq \theta} \{(F_{\theta'}, \phi_{\theta'} + \mathbb{E}_{F_{\theta'}}[T_{\theta'}(y) - T'_{\theta'}(y)])\} \right).$$

We argue that $V_P(\mathcal{T}|M') \leq V_P(\mathcal{T}_0(\mathcal{T})|M)$. To that end, let us characterize workers' choices under $\mathcal{T}$ and M'. Observe that, if a worker of type θ chooses $(F_{\theta'}, \phi_{\theta'} + \mathbb{E}_{F_{\theta'}}[T_{\theta'}(y) - T'_{\theta'}(y)]) \in M'(\theta)$ for some $\theta' \leq \theta$, the optimal associated tax choice in $\mathcal{T}$ is $T'_{\theta'}$. Thus, the workers' payoff from making this income choice is

$$\mathbb{E}_{F_{\theta'}}[y - T'_{\theta'}(y)] - \phi_{\theta'} - \mathbb{E}_{F_{\theta'}}[T_{\theta'}(y) - T'_{\theta'}(y)] = \mathbb{E}_{F_{\theta'}}[y - T_{\theta'}(y)] - \phi_{\theta'} = V_w(\mathcal{T}_0(\mathcal{T})|M(\theta')).$$

By monotonicity of M, it follows that the optimal choice for type θ within this class is given by setting $\theta' = \theta$, which gives him a payoff of $V_w(\mathcal{T}_0(\mathcal{T})|M(\theta))$.

Suppose first that $V_w(\mathcal{T}_0(\mathcal{T})|M(\theta)) = V_w(\mathcal{T}|M^0(\theta))$. Then, both when facing $\mathcal{T}$ and $M'(\theta)$, and when facing $\mathcal{T}_0(\mathcal{T})$ and $M(\theta)$, the worker's indirect utility is the same. When facing $\mathcal{T}$ and $M'(\theta)$, the worker can achieve his optimal utility by either picking the best income choice in $M^0(\theta)$ and the tax T_θ^0, or by picking $(F_\theta, \phi_\theta + \mathbb{E}_{F_\theta}[T_\theta(y) - T'_\theta(y)])$ and the

tax T'_θ. In both cases this choice leads to lower tax revenue than what obtains under $\mathcal{T}_0(\mathcal{T})$ and $M(\theta)$: In the former case this holds because of the worker's favorable tie-breaking, and in the latter case because of the definition of T'_θ which implies that $\mathbb{E}_{F_\theta}[T'_\theta(y)] \leq \mathbb{E}_{F_\theta}[T_\theta(y)]$. Therefore, total welfare *conditional on type* θ such that $V_w(\mathcal{T}_0(\mathcal{T})|M(\theta)) = V_w(\mathcal{T}|M^0(\theta))$ is the same under $\mathcal{T}$ and M', and under $\mathcal{T}_0(\mathcal{T})$ and M.

Second, suppose that $V_w(\mathcal{T}_0(\mathcal{T})|M(\theta)) > V_w(\mathcal{T}|M^0(\theta))$. Then, the worker's uniquely optimal choice when facing $M'(\theta)$ and $\mathcal{T}$ is $(F_\theta, \phi_\theta + \mathbb{E}_{F_\theta}[T_\theta(y) - T'_\theta(y)])$ and the tax T'_θ. As argued above, this gives him a payoff of $\mathbb{E}_{F_\theta}[y - T_\theta(y)] - \phi_\theta = V_w(\mathcal{T}_0(\mathcal{T})|M(\theta))$, which by assumption is strictly greater than his optimal payoff when choosing optimally an element of $M^0(\theta)$. Moreover, expected tax revenue from the worker's optimal choice under $\mathcal{T}$ and $M'(\theta)$ is $\mathbb{E}_{F_\theta}[T'_\theta(y)] \leq \mathbb{E}_{F_\theta}[T_\theta(y)]$, where the inequality follows from the definition of T'_θ.

Combining everything and taking expectation with respect to θ we have

$$V_P(\mathcal{T}) \leq V_P(\mathcal{T}|M') \leq \int_\Theta [\omega(\theta)W(V_w(\mathcal{T}_0(\mathcal{T})|M(\theta))) + \alpha\mathbb{E}_{F_\theta}[T_\theta(y)]]\, d\mu(\theta) = V_P(\mathcal{T}_0(\mathcal{T})|M).$$

Since M was arbitrary, it follows that $V_P(\mathcal{T}) \leq V_P(\mathcal{T}_0(\mathcal{T}))$.

Proof of Proposition 7

To show that, without loss of optimality, every element of $\mathcal{T}$ is convex, take any feasible menu of taxes $\mathcal{T} \subseteq \overline{\mathbf{T}}$. For each $T \in \mathcal{T}$, let $\tilde{T}_T(y) = y - \text{cav}(y - T(y))$, where for any $f : Y \to \mathbb{R}$, $\text{cav}f(y)$ stands for the concavification of f. Let $c_T(y) = y - T(y)$ and $\tilde{c}_T(y) = y - \tilde{T}_T(y)$. Consider the menu of convex taxes given by $\tilde{\mathcal{T}} = \bigcup_{T\in\mathcal{T}} \{\tilde{T}_T\}$.

For any feasible menu $\mathcal{T}$ and any $M \in \mathcal{M}$, let

$$R_\theta(\mathcal{T}|M) \equiv \max_{T\in\mathcal{T},(F,\phi)\in M(\theta):\mathbb{E}_F[y-T(y)]-\phi=V_w(\mathcal{T}|M(\theta))} \mathbb{E}_F[T(y)],$$

which is tax revenue from type θ under the menu $\mathcal{T}$ and the technology M. We use the following Claim.

Claim 8. For every $M \in \mathcal{M}$, there exists $M' \in \mathcal{M}$ such that, for all $\theta \in \Theta$, $V_w(\tilde{\mathcal{T}}|M(\theta)) = V_w(\mathcal{T}|M'(\theta))$ and $R_\theta(\tilde{\mathcal{T}}|M) = R_\theta(\mathcal{T}|M')$.

Proof. We follow an approach similar to Proposition S-5 in Walton and Carroll (2022).[49] For any $T \in \mathcal{T}$, let $l_T : Y \to Y$ and $u_T : Y \to Y$ denote the endpoints of the relevant intervals

[49] This approach can also be used to show convexity of the optimal singleton tax in Theorem 1. However, our original approach allows us to derive further characteristics of the optimal tax which are used later in the paper, like monotonicity and piecewise linearity in the finite-type case.

of concavification of c_T. Specifically, l_T and u_T are defined so that $l_T(y) \leq y \leq u_T(y)$ for all $y \in Y$, $\tilde{c}_T$ is affine on $[l_T(y), u_T(y)]$, and $\tilde{c}_T$ and c_T coincide at points $l_T(y)$ and $u_T(y)$. Fix any $M \in \mathcal{M}$, and let $\tilde{T}_\theta \in \tilde{\mathcal{T}}$ and $(F_\theta, \phi_\theta) \in M(\theta)$ be the tax and income chosen by type θ under the menu $\tilde{\mathcal{T}}$. Let $T_\theta \in \mathcal{T}$ be such that $\tilde{T}_{T_\theta} = \tilde{T}_\theta$. Let Y_θ be the random variable with distribution F_θ, and consider the income lottery F'_θ and associated random variable Y'_θ whose distribution conditional on Y_θ is given by

$$\Pr(Y'_\theta = l_{T_\theta}(y)|Y_\theta = y) = \frac{u_{T_\theta}(y) - y}{u_{T_\theta}(y) - l_{T_\theta}(y)}, \quad \Pr(Y'_\theta = u_{T_\theta}(y)|Y_\theta = y) = \frac{y - l_{T_\theta}(y)}{u_{T_\theta}(y) - l_{T_\theta}(y)},$$

and we set $\Pr(Y'_\theta = y|Y_\theta = y) = 1$ whenever $u_{T_\theta}(y) - l_{T_\theta}(y) = 0$. Observe that, for all $y \in Y$, $\mathbb{E}[Y'_\theta|Y_\theta = y] = y$ and therefore F'_θ is a mean-preserving spread of F_θ. Moreover, since $\tilde{c}_{T_\theta}$ is affine on $[l_{T_\theta}(y), u_{T_\theta}(y)]$, it holds that $\mathbb{E}[\tilde{T}_\theta(Y'_\theta)|Y_\theta = y] = \tilde{T}_\theta(y)$ for all $y \in Y$, and thus $\mathbb{E}[\tilde{T}_\theta(Y'_\theta)] = \mathbb{E}[\tilde{T}_\theta(Y_\theta)]$.

Given this, consider the correspondence $M' \in \mathcal{M}$ defined by

$$M'(\theta) \equiv M(\theta) \cup \left(\bigcup_{\theta' \leq \theta} \{(F'_{\theta'}, \phi_{\theta'})\} \right).$$

Let us derive workers' choices when they face M' and the menu $\mathcal{T}$. By the argument in the previous paragraph, $V_w(\tilde{\mathcal{T}}|M(\theta)) \leq V_w(\tilde{\mathcal{T}}|M'(\theta)) = \mathbb{E}[\tilde{T}_\theta(Y_\theta)] - \phi_\theta = V_w(\tilde{\mathcal{T}}|M(\theta))$, and thus $V_w(\tilde{\mathcal{T}}|M'(\theta)) = V_w(\tilde{\mathcal{T}}|M(\theta))$. Moreover, since $\tilde{c}_T(y) \geq c_T(y)$ for all $T \in \mathcal{T}$ and $y \in Y$, it holds that $V_w(\mathcal{T}|M'(\theta)) \leq V_w(\tilde{\mathcal{T}}|M'(\theta))$. Since $\tilde{c}_{T_\theta}$ and c_{T_θ} coincide on the support of F'_θ, we have that $V_w(\mathcal{T}|M'(\theta) \geq \mathbb{E}[c_{T_\theta}(Y'_\theta)] - \phi_\theta = \mathbb{E}[\tilde{c}_{T_\theta}(Y_\theta)] - \phi_\theta = V_w(\tilde{\mathcal{T}}|M'(\theta))$, and thus (F'_θ, ϕ_θ) and T_θ are optimal for the worker under M' and $\mathcal{T}$. This also implies that $V_w(\mathcal{T}|M'(\theta)) = V_w(\tilde{\mathcal{T}}|M(\theta))$ as stated in the claim.

Moreover, for any choice $(F, \phi) \in M'(\theta)$ and $T \in \mathcal{T}$ which is optimal for the worker under M' and $\mathcal{T}$, it must be that

$$V_w(\mathcal{T}|M'(\theta)) = \mathbb{E}_F[c_T(y)] - \phi \leq \mathbb{E}_F[\tilde{c}_T(y)] - \phi \leq V_w(\tilde{\mathcal{T}}|M'(\theta)) = V_w(\mathcal{T}|M'(\theta)),$$

and thus $\mathbb{E}_F[c_T(y)] = \mathbb{E}_F[\tilde{c}_T(y)]$. This, together with the fact that $c_T(y) \leq \tilde{c}_T(y)$ for all $y \in Y$, in turn requires that $\tilde{c}_T(y) = c_T(y)$ for all $y \in \text{supp}(F)$, and thus $\mathbb{E}_F[T(y)] = \mathbb{E}_F[\tilde{T}_T(y)]$ for any worker-optimal choice under M' and $\mathcal{T}$. The fact that the worker breaks indifference in favor of the planner and that she chooses (F_θ, ϕ_θ) when facing $M(\theta)$ and $\tilde{\mathcal{T}}$, implies that (F'_θ, ϕ_θ) and T_θ is a revenue-maximizing income and tax choice among those that are optimal

for the worker under M' and $\mathcal{T}$, and therefore

$$R_\theta(\mathcal{T}|M') = \mathbb{E}_{F'_\theta}[T_\theta(y)] = \mathbb{E}_{F'_\theta}[\tilde{T}_\theta(y)] = \mathbb{E}_{F_\theta}[\tilde{T}_\theta(y)] = R_\theta(\tilde{\mathcal{T}}|M).$$

□

Applying Claim 8,

$$V_P(\mathcal{T}) \leq V_P(\mathcal{T}|M') = \int_\Theta [\omega(\theta)W(V_w(\mathcal{T}|M'(\theta))) + \alpha R_\theta(\mathcal{T}|M')]\, d\mu(\theta) = \int_\Theta [\omega(\theta)W(V_w(\tilde{\mathcal{T}}|M(\theta))) + \alpha R_\theta(\tilde{\mathcal{T}}|M)]\, d\mu(\theta) = V_P(\tilde{\mathcal{T}}|M).$$

Since M is arbitrary, this establishes that $V_P(\tilde{\mathcal{T}}) \geq V_P(\mathcal{T})$ as desired.

Proof of Proposition 8

We begin by stating a version of Lemma 1 that allows for workers' risk aversion. The proof is identical to Lemma 1 so we omit it.

Lemma 7. *For any feasible T,*

$$V_P(T) = \int_\Theta [\omega(\theta)W(V_w(T|M^0(\theta))) + \alpha \mathbb{E}_{F_\theta}[T(y)]]\, d\mu(\theta), \tag{A.38}$$

where F_θ is defined by:

(i) If $V_w(T|M^0(\theta)) < \max_{y\in Y} \tilde{u}(y - T(y))$:

$$F_\theta \in \arg\min_{F\in\Delta(Y)} \mathbb{E}_F[T(y)], \quad s.t. \quad \mathbb{E}_F[\tilde{u}(y - T(y))] \geq V_w(T|M^0(\theta)),$$

(ii) If $V_w(T|M^0(\theta)) = \max_{y\in Y} \tilde{u}(y - T(y))$:

$$F_\theta \in \arg\max_{(F,0)\in M^0(\theta)} \mathbb{E}_F[T(y)], \quad s.t. \quad \mathbb{E}_F[\tilde{u}(y - T(y))] = V_w(T|M^0(\theta)).$$

The rest of the proof consists of modifying the arguments in the proof of Theorem 1. To avoid repetition, here we only focus on the arguments that are not exactly the same as in the original proof. Let T be any feasible tax rule, and let $u(y) = \tilde{u}(y - T(y))$ be the associated

utility contract. To construct the relevant separating hyperplanes, consider the sets

$$A = \text{co}\{(u(y), y - \tilde{u}^{-1}(u(y))) : y \in Y\}, \quad B_\theta = \{(u, v) : u > V_w(T|M^0(\theta)), v < \mathbb{E}_{F_\theta}[T(y)]\},$$

where F_θ is defined as in Lemma 7. For every $\theta \in \Theta$, A and B_θ are convex and disjoint, and thus the Separating Hyperplane Theorem implies that we can construct a θ-specific utility contract, u_θ, implicitly defined by:

$$\kappa_\theta + \lambda_\theta u_\theta(y) - \beta_\theta(y - \tilde{u}^{-1}(u_\theta(y))) = 0, \tag{A.39}$$

where $(\kappa_\theta, \beta_\theta, \lambda_\theta) \in \mathbb{R} \times \mathbb{R}^2_+$ satisfy for all $\theta \in \Theta$

$$\kappa_\theta + \lambda_\theta u(y) - \beta_\theta(y - \tilde{u}^{-1}(u(y))) \leq 0, \quad \forall y \in Y, \tag{A.40}$$

$$\kappa_\theta + \lambda_\theta \mathbb{E}_{F_\theta}[u(y)] - \beta_\theta \mathbb{E}_{F_\theta}[y - \tilde{u}^{-1}(u(y))] = 0. \tag{A.41}$$

Claim 9. If $\tilde{u}$ is concave on $\mathbb{R}_+$, then for all $\theta \in \Theta$, u_θ is continuous, non-decreasing and concave, and satisfies for all $y \in Y$, $u_\theta(y) \geq u(y)$.

Proof. The Claim is immediate if $\beta_\theta = 0$, so suppose that $\beta_\theta > 0$. First, we begin by showing that (A.39) has a unique solution $u_\theta(y)$ for each $y \in Y$ and $\theta \in \Theta$. By (A.40), for any $y \in Y$, there exist $u \geq \overline{u}$ (e.g., $u = u(y)$) such that $\kappa_\theta + \lambda_\theta u - \beta_\theta(y - \tilde{u}^{-1}(u)) \leq 0$. On the other hand, setting $u = \tilde{u}(c)$ where $c \in \mathbb{R}_+$ is arbitrarily large, we have that

$$\kappa_\theta + \lambda_\theta u - \beta_\theta(y - \tilde{u}^{-1}(u)) = \kappa_\theta + \lambda_\theta \tilde{u}(c) - \beta_\theta(y - c) \geq 0,$$

where the inequality follows from the fact that $\beta_\theta > 0$ and $c \in \mathbb{R}_+$ is arbitrarily large. Hence, by continuity, there exists $u \geq \underline{u}$ that satisfies (A.39) with equality. Uniqueness of the solution follows from the fact that the left-hand-side of (A.39) is strictly increasing in $u_\theta(y)$.

Second, to show that $u_\theta(y)$ is non-decreasing, take $y, y' \in Y$ such that $y' > y$. (A.39) implies that

$$\lambda_\theta(u_\theta(y') - u_\theta(y)) + \beta_\theta(\tilde{u}^{-1}(u_\theta(y')) - \tilde{u}^{-1}(u_\theta(y))) = \beta_\theta(y' - y) > 0, \tag{A.42}$$

where the inequality follows from the fact that $\beta_\theta > 0$. Given that $(\lambda_\theta, \beta_\theta) \neq (0, 0)$, $\lambda_\theta \geq 0$ and $\tilde{u}^{-1}$ is strictly increasing, (A.42) requires that $u_\theta(y') \geq u_\theta(y)$.

Third, we show that $u_\theta(y)$ is concave. Suppose first that $\lambda_\theta > 0$ and suppose toward a contradiction that there is $y, y' \in Y$ with $y \neq y'$ and $\gamma \in (0, 1)$ such that $\gamma u_\theta(y) + (1 -$

$\gamma)u_\theta(y') > u_\theta(y_\gamma)$, where $y_\gamma = \gamma y + (1-\gamma)y'$. By the definition of $u_\theta(y)$, we have

$$u_\theta(y_\gamma) - \gamma u_\theta(y) - (1-\gamma)u_\theta(y') = \frac{\beta_\theta}{\lambda_\theta}[\gamma\tilde{u}^{-1}(u_\theta(y)) + (1-\gamma)\tilde{u}^{-1}(u_\theta(y')) - \tilde{u}^{-1}(u_\theta(y_\gamma))] >$$
$$\frac{\beta_\theta}{\lambda_\theta}[\gamma\tilde{u}^{-1}(u_\theta(y)) + (1-\gamma)\tilde{u}^{-1}(u_\theta(y')) - \tilde{u}^{-1}(\gamma u_\theta(y) + (1-\gamma)u_\theta(y'))] \geq 0,$$

where the first inequality follows from the assumption that $\gamma u_\theta(y) + (1-\gamma)u_\theta(y') > u_\theta(y_\gamma)$ and the fact that $\tilde{u}^{-1}$ is strictly increasing, and the second inequality from convexity of $\tilde{u}^{-1}$. This contradicts $\gamma u_\theta(y) + (1-\gamma)u_\theta(y') > u_\theta(y_\gamma)$. Second, if $\lambda_\theta = 0$, then

$$u_\theta(y) = \tilde{u}(-\kappa_\theta/\beta_\theta + y),$$

which is concave by concavity of $\tilde{u}(y)$.

Fourth, we show that $u_\theta(y)$ is continuous. The fact that $\tilde{u}$ is strictly increasing implies that it is almost everywhere differentiable. Let $C \subset (0, +\infty)$ be an open set in which $\tilde{u}$ is differentiable. (A.39) is strictly increasing in $u_\theta(y)$, and therefore by the Implicit Function Theorem, u_θ is differentiable (and therefore continuous) on C. By concavity of u_θ this is enough to ensure that u_θ is continuous on Y (see Theorem 5.27 in Aliprantis and Border (2006)).

Finally, for each $y \in Y$, (A.39) and (A.40) imply that

$$\lambda_\theta(u_\theta(y) - u(y)) + \beta_\theta[\tilde{u}^{-1}(u_\theta(y)) - \tilde{u}^{-1}(u(y))] \geq 0,$$

which is only possible if $u_\theta(y) \geq u(y)$. □

Let $\overline{u}(y) = \inf_{\theta\in\Theta} u_\theta(y)$, and $\overline{T}(y) = y - \tilde{u}^{-1}(\overline{u}(y))$. By Claim 9, $\overline{u}$ is feasible, non-decreasing, concave, and satisfies $V_w(\overline{T}|M^0(\theta)) \geq V_w(T|M^0(\theta))$ for all $\theta \in \Theta$. An analogous argument to the one in the proof of Theorem 1, shows that $\overline{T}$ yields weakly higher expected revenue than T conditional on every $\theta \in \Theta$. Thus, $V_P(\overline{T}) \geq V_P(T)$. Existence of the maximum can also be established analogously to Theorem 1.

B Differentiability of worst-case welfare with respect to tax perturbations

B.1 Envelope theorem for problems with concave parametrized constraints

In this section, we provide an extension of the envelope theorem for problems with paramatrized constraints by Milgrom and Segal (2002) (henceforth, MS), that does not require the constraint to be differentiable—as assumed in their Theorem 5 and Corollary 5—, but instead requires that the constraint is concave with respect to the parameter.

As MS, we are concerned with a constrained optimization problem of the form

$$V(t) = \sup_{x \in X : g(x,t) \geq 0} f(x,t), \quad X^*(t) = \{x \in X : g(x,t) \geq 0, f(x,t) = V(t)\}, \tag{B.1}$$

where X is a convex compact set in a normed linear space, and $f : X \times [\underline{t}, \bar{t}] \to \mathbb{R}$ and $g : X \times [\underline{t}, \bar{t}] \to \mathbb{R}^k$ and $\underline{t}, \bar{t} \in \mathbb{R}$ satisfy $\underline{t} < \bar{t}$. We define the Lagrangian of this problem as $L : X \times \mathbb{R}^k_+ \times [\underline{t}, \bar{t}] \to \mathbb{R}$

$$L(x, y, t) = f(x,t) + \sum_{i=1}^{k} y_i g_i(x,t),$$

and we let $Y^*(t)$ be the set of solutions to the dual program,

$$Y^*(t) = \arg\min_{y \in \mathbb{R}^k_+} \sup_{x \in X} L(x, y, t).$$

Lemma 8. *Suppose that f and g are continuous and concave in x, $f_t(x,t)$ is continuous in (x,t), $g(x,t)$ is continuous and concave in t for all $x \in X$ with directional derivatives with respect to t that are continuous in x, and there exists $\hat{x} \in X$ such that $g(\hat{x}, t) \gg 0$ for all $t \in [\underline{t}, \bar{t}]$. Then, V is directionally differentiable and its directional derivatives equal:*

$$V'(t+) = \max_{x \in X^*(t)} \min_{y \in Y^*(t)} \frac{\partial L(x,y,t)}{\partial t+} = \min_{y \in Y^*(t)} \max_{x \in X^*(t)} \frac{\partial L(x,y,t)}{\partial t+}, \quad \text{for } t < \bar{t}$$
$$V'(t-) = \min_{x \in X^*(t)} \max_{y \in Y^*(t)} \frac{\partial L(x,y,t)}{\partial t-} = \max_{y \in Y^*(t)} \min_{x \in X^*(t)} \frac{\partial L(x,y,t)}{\partial t-}, \quad \text{for } t > \underline{t}.$$

Proof. By an analogous argument to Corollary 5 in MS, it holds that $X^*(t) \times Y^*(t)$ is non-

empty and equal to the saddle-set of the Lagrangian, and thus

$$V(t) = \max_{x \in X} \min_{y \in \mathbb{R}^k_+} L(x, y, t) = \min_{y \in \mathbb{R}^k_+} \max_{x \in X} L(x, y, t).$$

Fix $t_0 \in [\underline{t}, \bar{t})$, by definition of the saddle point, for any selection $(x(t), y(t)) \in X^*(t) \times Y^*(t)$ and any $t > t_0$, we can write

$$\frac{L(x(t_0), y(t), t) - L(x(t_0), y(t), t_0)}{t - t_0} \leq \frac{V(t) - V(t_0)}{t - t_0} \leq \frac{L(x(t), y(t_0), t) - L(x(t), y(t_0), t_0)}{t - t_0}. \tag{B.2}$$

The fact that $g_i(x, \cdot)$ is concave implies that it is directionally differentiable and thus, by the generalized Mean Value Theorem,

$$\frac{\partial g_i(x(t_0), t'(t))}{\partial t+} \leq \frac{g_i(x(t_0), t) - g(x(t_0), t_0)}{t - t_0} \leq \frac{\partial g_i(x(t_0), t'(t))}{\partial t-}.$$

for some $t'(t) \in [t_0, t]$. Moreover, concavity of $g_i(x, \cdot)$ implies that

$$\frac{g_i(x(t), t) - g_i(x(t), t_0)}{t - t_0} \leq \frac{\partial g_i(x(t), t_0)}{\partial t_0+}.$$

Further, differentiability of $f(x, \cdot)$ and the Mean Value Theorem imply that

$$\frac{f(x(t_0), t) - f(x(t_0), t_0)}{t - t_0} = f_t(x(t_0), t''(t)), \quad \frac{f(x(t), t) - f(x(t), t_0)}{t - t_0} = f_t(x(t), t'''(t)),$$

for some $t''(t), t'''(t) \in [t_0, t]$.

Combining all of this, we have

$$\max_{x \in X^*(t_0)} \left\{ f_t(x, t''(t)) + \sum_{i=1}^k y_i(t) \frac{\partial g_i(x, t'(t))}{\partial t+} \right\} \leq \frac{V(t) - V(t_0)}{t - t_0} \leq \min_{y \in Y^*(t_0)} \left\{ f_t(x(t), t'''(t)) + \sum_{i=1}^k y_i \frac{\partial g_i(x(t), t_0)}{\partial t+} \right\}.$$

And, in the limit,

$$\liminf_{t\downarrow t_0} \max_{x\in X^*(t_0)} \left\{ f_t(x,t''(t)) + \sum_{i=1}^{k} y_i(t) \frac{\partial g_i(x,t'(t))}{\partial t+} \right\} \geq \liminf_{t\downarrow t_0} \max_{x\in X^*(t_0)} \left\{ f_t(x,t''(t)) + \sum_{i=1}^{k} y_i(t) \frac{\partial g_i(x,t_0)}{\partial t+} \right\}$$
$$\geq \min_{y\in Y^*(t_0)} \max_{x\in X^*(t_0)} \left\{ f_t(x,t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x,t_0)}{\partial t+} \right\},$$

where the first inequality follows from lower semi-continuity of the directional derivatives of $g(x,\cdot)$ with respect to t which in turn follows from concavity of $g(x,\cdot)$ (see Theorem 33 in Fenchel and Blackett (1953)), and the last inequality follows from the fact that the Lagrangian is continuous, and hence the saddle set is upper hemicontinuous in t by Berge's Maximum Theorem.[50]

And similarly,

$$\limsup_{t\downarrow t_0} \min_{y\in Y^*(t_0)} \left\{ f_t(x(t),t'''(t)) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x(t),t_0)}{\partial t+} \right\} \leq$$
$$\max_{x\in X^*(t_0)} \min_{y\in Y^*(t_0)} \left\{ f_t(x,t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x,t_0)}{\partial t+} \right\},$$

where the inequality follows again from upper hemicontinuity of the saddle set.

Given the above, taking the limits inferior and superior in (B.2), we have

$$\min_{y\in Y^*(t_0)} \max_{x\in X^*(t_0)} \left\{ f_t(x,t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x,t_0)}{\partial t+} \right\} \leq \liminf_{t\downarrow t_0} \frac{V(t)-V(t_0)}{t-t_0}$$
$$\leq \limsup_{t\downarrow t_0} \frac{V(t)-V(t_0)}{t-t_0} \leq \max_{x\in X^*(t_0)} \min_{y\in Y^*(t_0)} \left\{ f_t(x,t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x,t_0)}{\partial t+} \right\}$$

We also know that

$$\min_{y\in Y^*(t_0)} \max_{x\in X^*(t_0)} \left\{ f_t(x,t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x,t_0)}{\partial t+} \right\} \geq \max_{x\in X^*(t_0)} \min_{y\in Y^*(t_0)} \left\{ f_t(x,t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x,t_0)}{\partial t+} \right\}.$$

Hence,

$$V'(t_0+) = \min_{y\in Y^*(t_0)} \max_{x\in X^*(t_0)} \left\{ f_t(x,t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x,t_0)}{\partial t+} \right\} = \max_{x\in X^*(t_0)} \min_{y\in Y^*(t_0)} \left\{ f_t(x,t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x,t_0)}{\partial t+} \right\}.$$

[50] To ensure compactness of the choice set in B.4, we can bound above the relevant choices for y in the same way as in MS.

The argument for $V'(t-)$ is analogous, which gives

$$V'(t-) = \max_{y \in Y^*(t_0)} \min_{x \in X^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t-} \right\} = \min_{x \in X^*(t_0)} \max_{y \in Y^*(t_0)} \left\{ f_t(x, t_0) + \sum_{i=1}^{k} y_i \frac{\partial g_i(x, t_0)}{\partial t-} \right\}.$$

□

B.2 Directional derivatives of $V_P(T)$

Take any $T \in \mathbf{T}$. By convexity, T is directionally differentiable. We let $T'(y+)$ and $T'(y-)$ denote respectively its right and left derivatives. Consider the perturbation

$$T^\varepsilon(y) = T(y) + \varepsilon D(y),$$

where $D : Y \to \mathbb{R}$ is continuous and $\varepsilon \in \mathbb{R}$. Let

$$M_\theta(\varepsilon) = \arg\max_{(F,\phi) \in M^0(\theta)} \{\mathbb{E}_F[y - T^\varepsilon(y)] - \phi\}, \quad M^*_\theta(\varepsilon) = \arg\max_{(F,\phi) \in M_\theta(\varepsilon)} \mathbb{E}_F[T^\varepsilon(y)].$$

Lemma 9. *For any $T \in \mathbf{T}$ such that either (i) $y - T(y)$ is strictly increasing or (ii) $\phi > 0$ for some $(F, \phi) \in M^*_\theta(0)$ almost everywhere in θ, $V_P(T^\varepsilon)$ is directionally differentiable at $\varepsilon = 0$.*

The expression for the directional derivatives can be found in the proof below.

Proof. We use the expression for $V_P(T)$ derived in Lemma 1, and compute the directional derivative at each value of the integrand. We do so by applying Lemma 8 and some results in MS.

Claim 10. For all $\theta \in \Theta$, the function $\varepsilon \to V_w(T^\varepsilon | M^0(\theta))$ is convex.

Proof. Take $\varepsilon, \varepsilon' \in \mathbb{R}$ and $\gamma \in [0, 1]$, and let $\varepsilon_\gamma = \gamma\varepsilon + (1 - \gamma)\varepsilon'$. Applying the definition,

$$\begin{aligned} &\gamma V_w(T^\varepsilon | M^0(\theta)) + (1 - \gamma) V_w(T^{\varepsilon'} | M^0(\theta)) = \\ &\gamma \max_{(F,\phi) \in M^0(\theta)} \{\mathbb{E}_F[y - T^\varepsilon(y)] - \phi\} + (1 - \gamma) \max_{(F,\phi) \in M^0(\theta)} \{\mathbb{E}_F[y - T^{\varepsilon'}(y)] - \phi\} \geq \\ &\max_{(F,\phi) \in M^0(\theta)} \{\mathbb{E}_F[y - \gamma T^\varepsilon(y) - (1 - \gamma) T^{\varepsilon'}(y)] - \phi\} = V_w(T^{\varepsilon_\gamma} | M^0(\theta)). \end{aligned}$$

□

It follows from Claim 10 and the chain rule that worst-case total worker-welfare (see (3.1)) is directionally differentiable with respect to ε. Its directional derivative can be computed applying Corollary 4 in MS. We state its expression in the following Claim.

Claim 11. For all $\theta \in \Theta$, the function $\varepsilon \to V_w(T^\varepsilon | M(\theta))$ is directionally differentiable with derivatives

$$\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial \varepsilon +} = \max_{(F,\phi)\in M_\theta^*(\varepsilon)} \mathbb{E}_F[-D(y)],$$
$$\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial \varepsilon -} = \min_{(F,\phi)\in M_\theta^*(\varepsilon)} \mathbb{E}_F[-D(y)].$$

Next, we turn to the revenue component of (3.1). Under condition (i) or (ii) in the lemma we have that, for all $\theta \in \Theta$, worst-case revenue is determined by Case (i) in Lemma 1. In particular, if there is a positive measure of types such that $V_w(T|M^0(\theta)) = \max_{y\in Y}\{y - T(y)\}$ and thus $\phi = 0$ for all $(F,\phi) \in M_\theta^*(0)$, strict monotonicity of T implies that there exists a unique $(F, 0)$ that attains $V_w(T|M^0(\theta))$, and thus the favorable tie-breaking rule does not play a role. Therefore, for ε small enough to ensure that $T^\varepsilon(y)$ is strictly increasing in y, worst-case revenue under T^ε for any type θ can be written as

$$R^\theta(\varepsilon) = \min_{F\in\Delta(Y)} \mathbb{E}_F[T^\varepsilon(y)], \quad s.t. \quad \mathbb{E}_F[y - T^\varepsilon(y)] \geq V_w(T^\varepsilon|M^0(\theta)). \tag{B.3}$$

Let $y_\theta \equiv \min\{y \in Y : V_w(T|M^0(\theta)) = y - T(y)\}$, which is well defined given that $T(y)$ is continuous, $V_w(T|M^0(\theta)) \geq -T(0)$ and $V_w(T|M^0(\theta)) \leq \overline{y} - T(\overline{y})$. Let $\hat{y} \in Y$ be the highest y such that $T'(y-) = 0$, and set $\hat{y} = 0$ if $T'(0+) > 0$.

Claim 12. For all θ such that $V_w(T|M^0(\theta)) < \max_{y\in Y}\{y - T(y)\}$, $R^\theta(\varepsilon)$ is directionally differentiable at $\varepsilon = 0$ with

$$\begin{aligned}
\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} &= -\min_{\lambda\in\Lambda_\theta^*(0)} \max_{F\in X_\theta^*(0)} \left\{ -(1+\lambda)\mathbb{E}_F[D(y)] - \lambda \left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial \varepsilon +}\right|_{\varepsilon=0} \right\} \\
&= -\max_{F\in X_\theta^*(0)} \min_{\lambda\in\Lambda_\theta^*(0)} \left\{ -(1+\lambda)\mathbb{E}_F[D(y)] - \lambda \left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial \varepsilon +}\right|_{\varepsilon=0} \right\}, \\
\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon -}\right|_{\varepsilon=0} &= -\max_{\lambda\in\Lambda_\theta^*(0)} \min_{F\in X_\theta^*(0)} \left\{ -(1+\lambda)\mathbb{E}_F[D(y)] - \lambda \left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial \varepsilon -}\right|_{\varepsilon=0} \right\} \\
&= -\min_{F\in X_\theta^*(0)} \max_{\lambda\in\Lambda_\theta^*(0)} \left\{ -(1+\lambda)\mathbb{E}_F[D(y)] - \lambda \left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial \varepsilon -}\right|_{\varepsilon=0} \right\},
\end{aligned}$$

and

$$\Lambda_\theta^*(0) = \left[\frac{T'(y_\theta-)}{1-T'(y_\theta-)}, \frac{T'(y_\theta+)}{1-T'(y_\theta+)}\right],$$
$$X_\theta^*(0) = \begin{cases} \{F \in \Delta(Y) : \mathbb{E}_F[T(y)] = T(y_\theta), \mathbb{E}_F[y] = y_\theta\}, & \text{if } y_\theta \geq \hat{y}, \\ \{F \in \Delta([0,\hat{y}]) : \mathbb{E}_F[y] \geq y_\theta\}, & \text{if } y_\theta < \hat{y}. \end{cases}$$

Proof. The fact that $V_w(T|M^0(\theta)) < \max_{y\in Y}\{y - T(y)\}$, implies that, for ε sufficiently small, the constraint set in (B.3) has non-empty interior. Therefore, we can write (B.3) as the following saddle-point problem (Luenberger, 1997)

$$R^\theta(\varepsilon) = -\min_{\lambda\in\mathbb{R}_+}\max_{F\in\Delta(Y)} L(F,\lambda,\varepsilon), \tag{B.4}$$

where $L(F,\lambda,\varepsilon) = -\mathbb{E}_F[T^\varepsilon(y)] + \lambda(\mathbb{E}_F[y - T^\varepsilon(y)] - V_w(T^\varepsilon|M^0(\theta)))$.

By Claim 10, the conditions in Lemma 8 are satisfied. Thus, $\varepsilon \to R^\theta(\varepsilon)$ is directionally differentiable with

$$\begin{aligned}
\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} &= -\min_{\lambda\in\Lambda_\theta^*(0)}\max_{F\in X_\theta^*(0)}\left\{-(1+\lambda)\mathbb{E}_F[D(y)] - \lambda\left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial\varepsilon+}\right|_{\varepsilon=0}\right\} \\
&= -\max_{F\in X_\theta^*(0)}\min_{\lambda\in\Lambda_\theta^*(0)}\left\{-(1+\lambda)\mathbb{E}_F[D(y)] - \lambda\left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial\varepsilon+}\right|_{\varepsilon=0}\right\}, \\
\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon-}\right|_{\varepsilon=0} &= -\max_{\lambda\in\Lambda_\theta^*(0)}\min_{F\in X_\theta^*(0)}\left\{-(1+\lambda)\mathbb{E}_F[D(y)] - \lambda\left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial\varepsilon-}\right|_{\varepsilon=0}\right\} \\
&= -\min_{F\in X_\theta^*(0)}\max_{\lambda\in\Lambda_\theta^*(0)}\left\{-(1+\lambda)\mathbb{E}_F[D(y)] - \lambda\left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial\varepsilon-}\right|_{\varepsilon=0}\right\},
\end{aligned}$$

where $X_\theta^*(0)\times\Lambda_\theta^*(0)$ is the saddle set of $L(F,\lambda,\varepsilon)$.

To complete the proof, we compute the saddle set $X_\theta^*(0)\times\Lambda_\theta^*(0)$. Let $\Lambda_\theta = \left[\frac{T'(y_\theta-)}{1-T'(y_\theta-)}, \frac{T'(y_\theta+)}{1-T'(y_\theta+)}\right]$, and

$$X_\theta = \begin{cases} \{F \in \Delta(Y) : \mathbb{E}_F[T(y)] = T(y_\theta), \mathbb{E}_F[y] = y_\theta\}, & \text{if } y_\theta \geq \hat{y}, \\ \{F \in \Delta([0,\hat{y}]) : \mathbb{E}_F[y] \geq y_\theta\}, & \text{if } y_\theta < \hat{y}. \end{cases}$$

Take any $(F^*,\lambda^*) \in X_\theta \times \Lambda_\theta$, and let $\tau \in [T'(y_\theta-), T'(y_\theta+)]$ be such that $\lambda^* = \tau/(1-\tau)$. Suppose first that $y_\theta \geq \hat{y}$. Then, for all $\lambda \in \mathbb{R}$,

$$L(F^*,\lambda,0) = -\mathbb{E}_{F^*}[T(y)] = L(F^*,\lambda^*,0).$$

Also, for each $F \in \Delta(Y)$ with mean $y_F \in Y$, we have

$$L(F, \lambda^*, 0) = -\mathbb{E}_F[T(y)] + \frac{\tau}{1-\tau}(\mathbb{E}_F[y - T(y)] - V_w(T|M^0(\theta))) \leq$$
$$-T(y_F) + \frac{\tau}{1-\tau}(y_F - T(y_F) - V_w(T|M^0(\theta))) \leq -T(y_\theta) + \frac{\tau}{1-\tau}(y_\theta - T(y_\theta) - V_w(T|M^0(\theta))) =$$
$$-T(y_\theta) = L(F^*, \lambda^*, 0),$$

where the first inequality follows from convexity of T, and the second inequality follows from the definition of τ which implies that $-T(y_F)+\frac{\tau}{1-\tau}(y_F-T(y_F)-V_w(T|M^0(\theta)))$ is maximized at $y_F = y_\theta$. Thus, any $(F^*, \lambda^*) \in X_\theta \times \Lambda_\theta$ is a saddle point of the Lagrangian.

To show uniqueness, let $F \in \Delta(Y)$ be any solution to (B.3) (with $\varepsilon = 0$), and let $y_F \in Y$ be its mean. We argue that F must satisfy the constraint in (B.3) with equality. If not, $y_F - T(y_F) \geq \mathbb{E}_F[y - T(y)] > V_w(T|M^0(\theta)) = y_\theta - T(y_\theta)$, and montonicity of $y - T(y)$ implies that $y_F > y_\theta$. Moreover, the fact that $y_\theta \geq \hat{y}$ implies that $T(y_F) > T(0)$. Thus, we have that $-\mathbb{E}_F[T(y)] \leq -T(y_F) < -(1-\varepsilon)T(y_F) - \varepsilon T(0)$ for $\varepsilon \in (0,1)$. Also, for ε sufficiently small, $(1-\varepsilon)(y_F - T(y_F)) - \varepsilon T(0) \geq (1-\varepsilon)\mathbb{E}_F[y-T(y)] - \varepsilon T(0) \geq V_w(T|M^0(\theta))$. These last two statements are in contradiction with F being a solution for (B.3). As a result, $\mathbb{E}_F[y - T(y)] = y_\theta - T(y_\theta)$ for all $F \in X^*_\theta(0)$. Moreover, convexity of T implies that the optimal value of the Lagrangian is attained by a deterministic F, and therefore $\mathbb{E}_F[T(y)] = T(y_F)$ at the optimum. Combining these two facts, we have that any optimal F satisfies $y_F - T(y_F) = y_\theta - T(y_\theta)$. Further, the fact that $y_\theta - T(y_\theta) = V_w(T|M^0(\theta)) < \overline{y} - T(\overline{y})$, together with concavity and monotonicity of $y - T(y)$, implies that $y - T(y)$ is strictly increasing a neighborhood of y_θ. Thus, $y_F - T(y_F) = y_\theta - T(y_\theta)$ if and only if $y_F = y_\theta$. This concludes the proof that $X^*_\theta(0) = X_\theta$.

Finally, by the previous paragraph, any $\lambda \in \Lambda^*(0)$ has to be such that $y_\theta \in \arg\max_{y \in Y} \{-T(y) + \lambda[y - T(y) - V_w(T|M^0(\theta))]\}$. By convexity of T, this is achieved if and only if $\lambda \in \Lambda_\theta$, and thus $\Lambda^*_\theta(0) = \Lambda_\theta$ as desired.

It remains to consider the case in which $y_\theta < \hat{y}$. By definition of $\hat{y}$, $T(y)$ is constant on $[0, \hat{y}]$, and by monotonicity and convexity of T it holds that $\min_{y \in Y} T(y) = T(\hat{y})$. Any $F \in X_\theta$ satisfies $\mathbb{E}_F[T(y)] = T(\hat{y})$, and

$$\mathbb{E}_F[y - T(y)] = \mathbb{E}_F[y] - T(\hat{y}) \geq y_\theta - T(\hat{y}) = y_\theta - T(y_\theta) = V_w(T|M^0(\theta)),$$

and is therefore optimal for (B.3). Hence, $X_\theta \subseteq X^*_\theta(0)$.

On the other hand, the definition of $\hat{y}$ implies that $T(y) > T(\hat{y})$ for all $y > \hat{y}$, and thus any F that is optimal for (B.3) has to be supported on a subset of $[0, \hat{y}]$. Moreover,

any $F \in \Delta([0,\hat{y}])$ that satisfies $\mathbb{E}_F[y] < y_\theta$ is not feasible—i.e., it violates $\mathbb{E}_F[y - T(y)] \geq V_w(T|M^0(\theta))$. These two observations imply that $X_\theta \supseteq X^*_\theta(0)$. Additionally, $y_\theta < \hat{y}$ implies that $\hat{y} - T(\hat{y}) > V_w(T|M^0(\theta))$, which implies that the constraint in (B.3) is slack, and thus by complementary slackness, $\Lambda^*_\theta(0) = \{0\} = \{T'(y_\theta)\} = \Lambda_\theta$. □

It remains to consider the case where $V_w(T|M^0(\theta)) = \max_{y\in Y}\{y - T(y)\} = \overline{y} - T(\overline{y})$. In this case, condition (i) in Lemma 9 must hold and thus, for ε sufficiently small, $y - T(y) - \varepsilon D(y)$ is strictly increasing. Given that $(\delta_{\overline{y}}, 0) \in M^0(\theta)$ by assumption, the unique income choice maximizing type θ's payoff (for any $M(\theta)$) is $(\delta_{\overline{y}}, 0)$. As a result, we obtain the following Claim.

Claim 13. For all θ such that $V_w(T|M^0(\theta)) = \max_{y\in Y}\{y - T(y)\}$, $R^\theta(\varepsilon)$ is differentiable at $\varepsilon = 0$ with

$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} = \Delta(\overline{y}).$$

The above arguments and the chain rule can be applied to compute the directional derivative for $V_P(T^\varepsilon)$, which is given by

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon *}\right|_{\varepsilon=0} = \int_\Theta \left[\omega(\theta)W'(V_w(T|M^0(\theta)))\left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial \varepsilon *}\right|_{\varepsilon=0} + \alpha\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon *}\right|_{\varepsilon=0}\right] d\mu(\theta), \quad * \in \{+,-\}. \tag{B.5}$$

□

C Comparative Statics for Binary Types

Proof of Corollary 3

Suppose that $\Phi(\cdot, \theta_2)$ is as assumed in the corollary (and all of the other assumptions in Section 4.1 are satisfied). Specializing the conditions for optimality (A.31) and (A.33) to this setting gives

$$-p_2(\omega(\theta_2) - \alpha)(y_2^0 - y_1^0) - p_2\frac{\alpha a \phi(y_2^0)}{(1-\tau_2)^2} + \gamma = 0, \tag{A.19'}$$

$$1 - \tau_2 - a\phi'(y_2^0) \geq 0, \quad = 0 \text{ if } y_2^0 < \overline{y}. \tag{A.21'}$$

The solution to the system of equations yields

$$\tau_1 < \tau_2 = 1 - \sqrt{\frac{\alpha a \phi(y_2^0(\tau_2))}{(\alpha - \omega(\theta_2))(y_2^0(\tau_2) - y_1^0)}},$$

which converges to 1 as $a \to 0$.

Proof of Corollary 4

We consider two cases separately.

Case 1: (4.7) does not hold. In this region of parameter values, we have that $\tau_1 < \tau_2$ is optimal, and thus the objective of the planner is

$$\begin{aligned} V(x_1^0, \tau_1, \tau_2) &= p_1 \left(\omega(\theta_1) - \alpha + \frac{\alpha}{1 - \tau_1} \right) [-(1 - \tau_1) x_1^0 - \Phi(-x_1^0, \theta_1)] \\ &+ p_2 \Bigg[\omega(\theta_2)(-(1 - \tau_1) x_1^0 + (1 - \tau_2)(y_2^0(\tau_2) + x_1^0) - \Phi(y_2^0(\tau_2), \theta_2)) \\ &+ \alpha \left(\left(y_2^0(\tau_2) - \frac{\Phi(y_2^0(\tau_2), \theta_2)}{1 - \tau_2} \right) \tau_2 + (\tau_2 - \tau_1) x_1^0 \right) \Bigg], \end{aligned}$$

which is the objective in (O_{bin}), evaluated at $y_2^0 = y_2^0(\tau_2)$ (which always holds at the optimal tax rule according to (A.33)), and where we have made the change of variables $x_1^0 = -y_1^0$. The gradient of V is:

$$\nabla V(x_1^0, \tau_1, \tau_2) = \begin{pmatrix} -p_1 \left(\omega(\theta_1) - \alpha + \frac{\alpha}{1 - \tau_1} \right) (1 - \tau_1 - \Phi_y(-x_1^0, \theta_1)) + p_2(\alpha - \omega(\theta_2))(\tau_2 - \tau_1) \\ (\mathbb{E}[\omega(\theta)] - \alpha) x_1^0 - \frac{p_1 \alpha \Phi(-x_1^0, \theta_1)}{(1 - \tau_1)^2} \\ p_2 \left((\alpha - \omega(\theta_2))(y_2^0(\tau_2) + x_1^0) - \frac{\alpha \Phi(y_2^0(\tau_2), \theta_2)}{(1 - \tau_2)^2} \right). \end{pmatrix}$$

Let $(x_1^{0*}, \tau_1^*, \tau_2^*)$ be optimal.

Claim 14. *V is supermodular in a neighborhood of $(x_1^{0*}, \tau_1^*, \tau_2^*)$.*

Proof. It suffices to show that all the cross-partial derivatives of V are non-negative. For $i = 1, 2, 3$, let $\nabla V(x_1^0, \tau_1, \tau_2)_i$ denote the ith row of $\nabla V(x_1^0, \tau_1, \tau_2)$. We examine each row separately:

- $\nabla V(x_1^0, \tau_1, \tau_2)_1$ is non-decreasing in τ_1 in a neighborhood of $(x_1^{0*}, \tau_1^*, \tau_2^*)$. To see this, we

compute the cross-partial derivative:

$$\frac{\partial \nabla V(x_1^0, \tau_1, \tau_2)_1}{\partial \tau_1} = -(\alpha - \mathbb{E}[\omega(\theta)]) + \frac{p_1 \alpha \Phi_y(-x_1^0, \theta_1)}{(1-\tau_1)^2}$$

If $\alpha = \mathbb{E}[\omega(\theta)]$, then it is immediate that $\frac{\partial \nabla V(x_1^{*0}, \tau_1^*, \tau_2^*)_1}{\partial \tau_1} > 0$. Next, suppose that $\alpha > \mathbb{E}[\omega(\theta)]$, in which case we have

$$-(\alpha - \mathbb{E}[\omega(\theta)]) + \frac{p_1 \alpha \Phi_y(-x_1^{0*}, \theta_1)}{(1-\tau_1^*)^2} \geq -(\alpha - \mathbb{E}[\omega(\theta)]) - \frac{\Phi_y(-x_1^{0*}, \theta_1)}{\Phi(-x_1^{0*}, \theta_1)}(\alpha - \mathbb{E}[\omega(\theta)])x_1^{0*} > 0, \tag{C.1}$$

where the first inequality follows from (A.36), and the second one follows from strict convexity of $\Phi(\cdot, \theta_1)$, and the fact that $\Phi(0, \theta_1) = 0$ and that $\alpha > \mathbb{E}[\omega(\theta)]$ by assumption. Also, $\nabla V(x_1^0, \tau_1, \tau_2)_1$ is non-decreasing everywhere in τ_2 due to $\alpha - \omega(\theta_2) \geq 0$.

- $\nabla V(x_1^0, \tau_1, \tau_2)_2$ is constant in τ_2, and strictly increasing (locally) in x_1^0 (by symmetry of the cross-partial derivative).
- Also by symmetry, $\nabla V(x_1^0, \tau_1, \tau_2)_3$ is non-decreasing with respect to τ_1 and x_1^0.

□

Next, we establish increasing differences of the objective with respect to (x_1^0, τ_1, τ_2) and α. Suppose that $\tau_1^* > 0$. Differentiating $\nabla V(x_1^0, \tau_1, \tau_2)_i$ with respect to α gives

$$\begin{aligned}
\frac{\partial \nabla V(x_1^{0*}, \tau_1^*, \tau_2^*)_1}{\partial \alpha} &= -p_1 \frac{\tau_1^*}{1-\tau_1^*}(1 - \tau_1^* - \Phi_y(-x_1^{0*}, \theta_1)) + p_2(\tau_2^* - \tau_1^*) = \\
&\qquad \frac{p_2(\omega(\theta_2)\tau_1^* + \omega(\theta_1)(1-\tau_1^*))(\tau_2^* - \tau_1^*)}{\alpha \tau_1^* + \omega(\theta_1)(1-\tau_1^*)} > 0, \\
\frac{\partial \nabla V(x_1^{0*}, \tau_1^*, \tau_2^*)_2}{\partial \alpha} &= -x_1^{0*} - \frac{p_1 \Phi(-x_1^{0*}, \theta_1)}{(1-\tau_1^*)^2} = \frac{-E[\omega(\theta)]x_1^{0*}}{\alpha} > 0, \\
\frac{\partial \nabla V(x_1^{0*}, \tau_1^*, \tau_2^*)_3}{\partial \alpha} &= p_2\left(y_2^0(\tau_2^*) + x_1^{0*} - \frac{\Phi(y_2^0(\tau_2^*), \theta_2)}{(1-\tau_2^*)^2}\right) = \frac{p_2 \omega(\theta_2)(y_2^0(\tau_2^*) + x_1^{0*})}{\alpha} > 0,
\end{aligned}$$

where we have substituted in (A.30), (A.31) and (A.32).

It follows from Theorem 5 in Milgrom and Shannon (1994) that $(x_1^{0*}, \tau_1^*, \tau_2^*)$ is increasing with respect to α. If $\tau_1^* = 0$, then an analogous argument yields the result.

Case 2: (4.7) holds with strict inequality. Here, $\tau_1^* = \tau_2^* = \tau^*$, and the planner's objective is

$$V(\tau) = \sum_{i=1}^{2} p_i \left(\omega(\theta_i) + \frac{\alpha\tau}{1-\tau} \right) ((1-\tau) y_i^0(\tau) - \Phi(y_i^0(\tau), \theta_i)),$$

By continuity, it suffices to show that τ^* is strictly in the region in which $\tau^* > 0$. The cross partial derivative of $V(\tau)$ with respect to (τ, α) at τ^* is

$$\frac{\partial V'(\tau^*)}{\partial \alpha} = \sum_{i=1}^{2} p_i \left(y_i^0(\tau^*) - \frac{\Phi(y_i^0(\tau^*), \theta_i)}{(1-\tau^*)^2} \right) = \sum_{i=1}^{2} p_i \frac{\omega(\theta_i) y_i^0(\tau^*)}{\alpha} > 0.$$

Therefore, applying the univariate Implicit Function Theorem, we obtain that τ^* is non-decreasing in α in the region where $\tau_1^* = \tau_2^* = \tau^* > 0$.

By Berge's Maximum Theorem, the (unique) maximizer in (O_{bin}) is continuous in α, and therefore monotonicity of (τ_1^*, τ_2^*) continues to hold at the point in which (4.7) holds with equality.

Proof of Corollary 5

Case 1: (4.7) does not hold. Using the definitions in the proof of Corollary 4, we have

$$\frac{\partial \nabla V(x_1^{0*}, \tau_1^*, \tau_2^*)_1}{\partial \omega(\theta_2)} = -p_2(\tau_2^* - \tau_1^*) < 0, \quad \frac{\partial \nabla V(x_1^{0*}, \tau_1^*, \tau_2^*)_2}{\partial \omega(\theta_2)} = p_2 x_1^0 < 0,$$
$$\frac{\partial \nabla V(x_1^{0*}, \tau_1^*, \tau_2^*)_3}{\partial \omega(\theta_2)} = -p_2(y_2^0(\tau_2^*) + x_1^{0*}) < 0.$$

Therefore, the objective satisfies increasing differences with respecto to (x_1^0, τ_1, τ_2) and $-\omega(\theta_2)$. Applying Claim 14 and Theorem 5 in Milgrom and Shannon (1994), it follows that $(x_1^{0*}, \tau_1^*, \tau_2^*)$ is decreasing with respect to $\omega(\theta_2)$ in the region of parameters that violates (4.7).
Case 2: (4.7) holds with strict inequality. In this case, we have

$$\frac{\partial V'(\tau^*)}{\partial \omega(\theta_2)} = -p_2 y_2^0(\tau^*) < 0,$$

and therefore τ^* is decreasing in $\omega(\theta_2)$ in this region as well. Monotonicity at the point in which (4.7) holds with equality follows again from continuity of the maximizer. This establishes the first part of the Corollary involving the monotonicity of τ_1 and τ_2.

To show the second part of the result, consider the region of parameters such that $\mathbb{E}[\omega(\theta)]$

is arbitrarily close to α.[51] In that region, we have that $\tau_1^* = 0$ and τ_2^* is weakly decreasing in $\omega(\theta_2)$ (strictly so unless $\tau_2^* = \tau_1^* = 0$). Therefore, it holds that $\tau_2^* - \tau_1^*$ decreases with $\omega(\theta_2)$ in that region.

[51] To ensure existence of the optimal tax, we continue to restrict attention to the case in which $\mathbb{E}[\omega(\theta)] \leq \alpha$.

Online Appendix

OA.1 Computation of Directional Derivatives of $V_P(T)$

In this online appendix, we collect the computation of all the directional derivatives used in the proofs in Appendix A. To do so, fix any $T \in \mathbf{T}$ and let $D(y)$ be a continuous function. Let $\hat{y} \in Y$ be the highest y such that $T'(y-) = 0$, and set $\hat{y} = 0$ if $T'(0+) > 0$. We write

$$T^{\varepsilon}(y) = T(y) + \varepsilon D(y),$$

with $\varepsilon \in \mathbb{R}$. We let $\mathcal{F}_{\theta}^{0} \subseteq \Delta(Y)$ be the set of optimal income lotteries of a worker with type θ under T and $M^{0}(\theta)$. Throughout this section, we focus on computing the directional derivative of tax revenue $R^{\theta}(\varepsilon)$ as defined in Section B for any $\theta \in \Theta$ satisfying $V_w(T|M^{0}(\theta)) < \max_{y \in Y}\{y - T(y)\}$. The remaining components in the directional derivative of $V_P(T^{\varepsilon})$ follow immediately from Claims 11, 12 and 13, and equation (B.5).

Lemma OA.1. *If $T(y) = t + \tau y$ with $t \leq 0$ and $\tau \in [0, 1)$, and if $D(y) = -\min\{\tilde{y} - y, 0\}$ for some $\tilde{y} \in (0, \overline{y}]$, then*

$$\left.\frac{\partial V_P(T^{\varepsilon})}{\partial \varepsilon +}\right|_{\varepsilon = 0} =$$

$$\int_{\Theta} \left[(w_{\theta} - \alpha) \max_{F \in \mathcal{F}_{\theta}^{0}} \mathbb{E}_F[\min\{\tilde{y} - y, 0\}] + \alpha \frac{\max_{F \in \mathcal{F}_{\theta}^{0}} \mathbb{E}_F[\min\{\tilde{y} - y, 0\}] - \min\{\tilde{y} - y_{\theta}, 0\}}{1 - \tau} \right] d\mu(\theta).$$

Proof. The fact that $\tau < 1$ implies that $y - T(y)$ is strictly increasing and thus Lemma 9 applies. We compute the directional derivative of $R^{\theta}(\varepsilon)$ in the case in which $V_w(T|M^{0}(\theta)) < \overline{y} - T(\overline{y})$ using the results in Claims 11 and 12. For any such θ, $\Lambda_{\theta}^{*}(0) = \{\tau/(1-\tau)\}$. Fix $(F_{\theta}^{0}, \phi_{\theta}^{0}) \in M^{0}(\theta)$ to be an optimal income choice for type θ under $M^{0}(\theta)$ and the tax T, and put $y_{\theta}^{0} = \mathbb{E}_{F_{\theta}^{0}}[y]$. Then, under an affine tax, $y_{\theta} = y_{\theta}^{0} - \phi_{\theta}^{0}/(1-\tau)$.

If $\tau = 0$, then $\hat{y} = \overline{y}$, and thus

$$X_{\theta}^{*}(0) = \{F \in \Delta(Y) : \mathbb{E}_F[y] \geq y_{\theta}\}.$$

And hence by Claim 12,

$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} = -\max_{F\in\Delta(Y)}\{\mathbb{E}_F[\min\{\tilde{y}-y,0\}], \quad s.t. \quad \mathbb{E}_F[y]\geq y_\theta\} =$$
$$-\max_{y\in Y}\{\min\{\tilde{y}-y,0\}, \quad s.t. \quad y\geq y_\theta\} = -\min\{\tilde{y}-y_\theta,0\},$$

where the second equality follows from concavity of $y\to\min\{\tilde{y}-y,0\}$.

Second if $\tau\in(0,1)$, then $\hat{y}<0$, and thus

$$X^*_\theta(0) = \{F\in\Delta(Y):\mathbb{E}_F[y]=y_\theta\}.$$

Applying Claim 12,

$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} = -\frac{1}{1-\tau}\max_{F\in\Delta(Y):\mathbb{E}_F[y]=y_\theta}\mathbb{E}_F[\min\{\tilde{y}-y,0\}] + \frac{\tau}{1-\tau}\left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial\varepsilon+}\right|_{\varepsilon=0} =$$
$$-\frac{1}{1-\tau}\min\{\tilde{y}-y_\theta,0\} + \frac{\tau}{1-\tau}\left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial\varepsilon+}\right|_{\varepsilon=0},$$

where the second equality follows again from concavity of $y\to\min\{\tilde{y}-y,0\}$. □

The remainder of the results in this section refer to Section 4, and therefore rely on Assumption 4 which we henceforth maintain. Recall that, under this assumption, workers' optimal income choice under $M^0(\theta)$ and any $T\in\mathbf{T}$ is unique and deterministic. We denote this choice by $y^0_\theta\in Y$. Moreover, uniqueness of y^0_θ implies that $V_w(T^\varepsilon|M^0(\theta))$ is differentiable at $\varepsilon=0$, with

$$\left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial\varepsilon}\right|_{\varepsilon=0} = -D(y^0_\theta).$$

Lemma OA.2. *Suppose that Assumption 4 holds. If $T\in\mathbf{T}$ is such that $T'(0+)<1$ and if $D(y)=-y$, then*

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial\varepsilon+}\right|_{\varepsilon=0} = \int_\Theta\left[(w_\theta-\alpha)y^0_\theta + \alpha\frac{y^0_\theta-\max\{y_\theta,\hat{y}\}}{1-T'(y_\theta+)}\right]d\mu(\theta),$$
$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial\varepsilon-}\right|_{\varepsilon=0} = \int_\Theta\left[(w_\theta-\alpha)y^0_\theta + \alpha\frac{y^0_\theta-y_\theta}{1-T'(y_\theta-)}\right]d\mu(\theta).$$

Proof. The fact that $T'(0+)<1$ and Assumption 4 imply that $y^0_\theta>0$ for all $\theta\in\Theta$, and therefore $\Phi(y^0_\theta,\theta)>0$ for all $\theta\in\Theta$. Thus, the conditions in Lemma 9 apply. Moreover,

$V_w(T|M^0(\theta)) < \max_{y\in Y}\{y - T(y)\}$, and thus the derivation of the directional derivative of $V_P(T^\varepsilon)$ follows from Claims 11 and 12.

To compute the directional derivative of $R^\theta(T^\varepsilon)$, consider first the case in which $y_\theta \geq \hat{y}$. In this case, applying Claims 11 and 12,

$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} = -\min_{\lambda\in\Lambda^*_\theta(0)}\max_{F\in X^*_\theta(0)}\left\{(1+\lambda)\mathbb{E}_F[y] - \lambda\left.\frac{\partial V_w(T^\varepsilon|M^0(\theta))}{\partial\varepsilon+}\right|_{\varepsilon=0}\right\} =$$
$$-\min_{\lambda\in\Lambda^*_\theta(0)}\{(1+\lambda)y_\theta - \lambda y^0_\theta\} = -\frac{y_\theta}{1-T'(y_\theta+)} + \frac{T'(y_\theta+)y^0_\theta}{1-T'(y_\theta+)},$$

where the last equality follows from the fact that $y_\theta < y^0_\theta$ and $T'(y_\theta+) \geq T'(y_\theta-)$ by convexity of T.

By an analogous argument, we have that

$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon-}\right|_{\varepsilon=0} = -\frac{y_\theta}{1-T'(y_\theta-)} + \frac{T'(y_\theta-)y^0_\theta}{1-T'(y_\theta-)}.$$

If $y_\theta < \hat{y}$, $\Lambda^*_\theta(0) = \{0\}$ and thus

$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} = -\max_{F\in X^*_\theta(0)}\mathbb{E}_F[y] = -\hat{y}, \quad \left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon-}\right|_{\varepsilon=0} = -\min_{F\in X^*_\theta(0)}\mathbb{E}_F[y] = -y_\theta.$$

Substituting these results into equation (B.5) gives the expressions in the lemma. □

Lemma OA.3. *Suppose that Assumption 4 holds. If $T \in \mathbf{T}$ is such that $0 < T'(0+) < 1$ and if $D(y) = -\min\{y, \tilde{y}\}$ for some $\tilde{y} \in (y_{\underline{\theta}}, y^0_{\underline{\theta}})$, then*

$$\left.\frac{\partial V_P(T^\varepsilon)}{\varepsilon+}\right|_{\varepsilon=0} = \int_\Theta \left((w_\theta - \alpha)\tilde{y} + \alpha\frac{\tilde{y} - \min\{y_\theta, \tilde{y}\}}{1-T'(y_\theta+)}\right) d\mu(\theta).$$

Proof. The fact that $T'(0+) < 1$ and Assumption 4 imply that $y^0_\theta > 0$ for all $\theta \in \Theta$, and therefore $\Phi(y^0_\theta, \theta) > 0$ for all $\theta \in \Theta$. Thus, the conditions in Lemma 9 apply. Moreover, $V_w(T|M^0(\theta)) < \max_{y\in Y}\{y - T(y)\}$, and thus the derivation of the directional derivative of $V_P(T^\varepsilon)$ follows from Claims 11 and 12. Also, $T'(0+) > 0$ implies that $\hat{y} = 0$. Then, by Claim

12, for all $\theta \in \Theta$ we have

$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} = -\min_{\lambda\in\Lambda^*_\theta(0)} \max_{F\in X^*_\theta(0)} \{(1+\lambda)\mathbb{E}_F[\min\{y,\tilde{y}\}] - \lambda \min\{y^0_\theta,\tilde{y}\}\} =$$
$$-\min_{\lambda\in\Lambda^*_\theta(0)} \{(1+\lambda)\min\{y_\theta,\tilde{y}\} - \lambda\tilde{y}\} = -\frac{\min\{y_\theta,\tilde{y}\}}{1-T'(y_\theta+)} + \frac{T'(y_\theta+)\tilde{y}}{1-T'(y_\theta+)},$$

where the first equality follows from concavity of $y \to \min\{y,\tilde{y}\}$. Substituting this expression into equation (B.5) give the result in the lemma. □

Let y_l be the highest $y \in Y$ such that $T'(y-) = T'(0+)$, and set $y_l = 0$ if $T'(y-) > T'(0+)$ for all $y > 0$.

Lemma OA.4. *Suppose that Assumption 4 holds. If $T \in \mathbf{T}$ is such that $0 < T'(0+) < 1$ and if $D(y) = -\min\{y, y_l\}$, then*

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} = \int_\Theta \left((w_\theta-\alpha)\min\{y^0_\theta,y_l\} + \alpha\frac{\min\{y^0_\theta,y_l\}-\min\{y_\theta,y_l\}}{1-T'(y_\theta+)}\right) d\mu(\theta),$$
$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon-}\right|_{\varepsilon=0} = \int_\Theta \left((w_\theta-\alpha)\min\{y^0_\theta,y_l\} + \alpha\frac{\min\{y^0_\theta,y_l\}-\min\{y_\theta,y_l\}}{1-T'(y_\theta-)}\right) d\mu(\theta).$$

Proof. We focus on the case in which $y_l > 0$. The case in which $y_l = 0$ is analyzed in Lemma OA.2. We will show that, for all $\theta \in \Theta$, $\mathbb{E}_F[\min\{y,y_l\}]$ is constant in $F \in X^*_\theta(0)$ and equal to $\min\{y_\theta, y_l\}$. Consider first θ such that $y_\theta \le y_l$. $T(y)$ is affine on $[0,y_l]$ and, by the definition of y_l, $T'(y-) > T'(y_l-)$ for all $y > y_l$. Thus, any $F \in \Delta(Y)$ satisfying $\mathbb{E}_F[T(y)] = T(y_\theta)$ must be supported on some subset of $[0,y_l]$. Hence, for any $F \in X^*_\theta(0)$ it holds that $\mathbb{E}_F[\min\{y,y_l\}] = \mathbb{E}_F[y] = y_\theta$. Second, consider θ such that $y_\theta > y_l$. By an analogous reasoning, any $F \in \Delta(Y)$ that satisfies $\mathbb{E}_F[T(y)] = T(y_\theta)$ has to be supported on some subset of $[y_l,\overline{y}]$. Then, for any $F \in X^*_\theta(0)$ it holds that $\mathbb{E}_F[\min\{y,y_l\}] = y_l$.

Applying this result, we have that for all $\theta \in \Theta$

$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon+}\right|_{\varepsilon=0} = -\min_{\lambda\in\Lambda^*_\theta(0)} \{(1+\lambda)\min\{y_\theta,y_l\} - \lambda\min\{y^0_\theta,y_l\}\} = -\frac{\min\{y_\theta,y_l\}}{1-T'(y_\theta+)} + \frac{T'(y_\theta+)\min\{y^0_\theta,y_l\}}{1-T'(y_\theta+)},$$
$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon-}\right|_{\varepsilon=0} = -\max_{\lambda\in\Lambda^*_\theta(0)} \{(1+\lambda)\min\{y_\theta,y_l\} - \lambda\min\{y^0_\theta,y_l\}\} = -\frac{\min\{y_\theta,y_l\}}{1-T'(y_\theta-)} + \frac{T'(y_\theta-)\min\{y^0_\theta,y_l\}}{1-T'(y_\theta-)}.$$

The result in the lemma obtains again from substituting this into equation (B.5). □

Let y_u be the lowest $y \in Y$ such that $T'(y+) = T'(\overline{y}-)$, and set $y_u = \overline{y}$ if $T'(y+) < T'(\overline{y}-)$

for all $y < \overline{y}$.

Lemma OA.5. *Suppose that Assumption 4 holds. If $T \in \mathbf{T}$ is such that $T'(0+) < 1$ and if $D(y) = \max\{y - y_u, 0\}$, then*

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} = \int_\Theta \mathbb{I}(y_\theta^0 > y_u)\left[-(w_\theta - \alpha)(y_\theta^0 - y_u) - \alpha \frac{(y_\theta^0 - y_u) - \max\{y_\theta - y_u, 0\}}{1 - T'(y_\theta +)}\right] d\mu(\theta),$$
$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon -}\right|_{\varepsilon=0} = \int_\Theta \mathbb{I}(y_\theta^0 > y_u)\left[-(w_\theta - \alpha)(y_\theta^0 - y_u) - \alpha \frac{(y_\theta^0 - y_u) - \max\{y_\theta - y_u, 0\}}{1 - T'(y_\theta -)}\right] d\mu(\theta).$$

Proof. Suppose that $y_u < y_{\overline{\theta}}^0$ (otherwise, the perturbation is irrelevant). We will show that, for all $\theta \in \Theta$, $\mathbb{E}_F[\max\{y - y_u, 0\}]$ is constant in $F \in X_\theta^*(0)$ and equal to $\max\{y_\theta - y_u, 0\}$. Consider first θ such that $y_\theta \geq y_u$. $T(y)$ is affine on $[y_u, \overline{y}]$ and, by the definition of y_u, $0 \leq T'(y+) < T'(y_u+)$ for all $y < y_u$. Thus, any $F \in X_\theta^*(0)$ must satisfy $\mathbb{E}_F[T(y)] = T(y_\theta)$, which in turn implies that F must be supported on some subset of $[y_u, \overline{y}]$. Hence, for any $F \in X_\theta^*(0)$ it holds that $\mathbb{E}_F[\max\{y - y_u, 0\}] = \mathbb{E}_F[y - y_u] = y_\theta - y_u$.

Second, consider θ such that $y_\theta < y_u$. If $y_\theta \geq \hat{y}$, then by an analogous reasoning, any $F \in \Delta(Y)$ that satisfies $\mathbb{E}_F[T(y)] = T(y_\theta)$ has to be supported on some subset of $[0, y_u]$. Then, for any $F \in X_\theta^*(0)$ it holds that $\mathbb{E}_F[\max\{y - y_u, 0\}] = 0$. If otherwise, $y_\theta < \hat{y}$, then the definition of $X_\theta^*(0)$ and the fact that $\hat{y} \leq y_u$ implies that $X_\theta^*(0) \subseteq \Delta([0, y_u])$, and thus for any $F \in X_\theta^*(0)$ it holds that $\mathbb{E}_F[\max\{y - y_u, 0\}] = 0$.

Then, applying Claim 12, for all $\theta \in \Theta$,

$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} = \frac{\max\{y_\theta - y_u, 0\}}{1 - T'(y_\theta +)} - \frac{T'(y_\theta +)\max\{y_\theta^0 - y_u, 0\}}{1 - T'(y_\theta +)},$$
$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon -}\right|_{\varepsilon=0} = \frac{\max\{y_\theta - y_u, 0\}}{1 - T'(y_\theta -)} - \frac{T'(y_\theta -)\max\{y_\theta^0 - y_u, 0\}}{1 - T'(y_\theta -)}.$$

The result in the lemma obtains again from substituting this into equation (B.5). □

Lemma OA.6. *Suppose that Assumption 4 holds. If $T \in \mathbf{T}$ is such that $T'(0+) < 1$ and $y_u < y_{\overline{\theta}}^0$, and if $D(y) = \max\{y - \tilde{y}, 0\}$ for some $\tilde{y} \in (y_u, y_{\overline{\theta}}^0)$, then*

$$\left.\frac{\partial V_P(T^\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} = \int_\Theta \mathbb{I}(y_\theta^0 > \tilde{y})\left[-(w_\theta - \alpha)(y_\theta^0 - \tilde{y}) - \alpha \frac{(y_\theta^0 - \tilde{y}) - \max\{y_\theta - \tilde{y}, 0\}}{1 - T'(y_\theta +)}\right] d\mu(\theta).$$

Proof. Suppose first that $y_\theta \geq \hat{y}$. Applying Claim 12,

$$\left.\frac{\partial R^\theta(\varepsilon)}{\partial \varepsilon +}\right|_{\varepsilon=0} = -\min_{\lambda \in \Lambda_\theta^*(0)} \max_{F \in X_\theta^*(0)} \{-(1+\lambda)\mathbb{E}_F[\max\{y-\tilde{y},0\}] + \lambda \max\{y_\theta^0 - \tilde{y}, 0\}\} =$$

$$-\min_{\lambda \in \Lambda_\theta^*(0)} \{-(1+\lambda)\max\{y_\theta - \tilde{y}, 0\} + \lambda \max\{y_\theta^0 - \tilde{y}, 0\}\} = \frac{\max\{y_\theta - \tilde{y}, 0\}}{1 - T'(y_\theta +)} - \frac{T'(y_\theta +)\max\{y_\theta^0 - \tilde{y}, 0\}}{1 - T'(y_\theta +)}.$$

If $y_\theta < \hat{y}$,

$$-\max_{F \in X_\theta^*(0)} \{-\mathbb{E}_F[\max\{y-\tilde{y},0\}]\} = -\max_{y \in [y_\theta, \hat{y}]} \{-\max\{y-\tilde{y},0\}\} = \max\{y_\theta - \tilde{y}, 0\}$$

The result in the lemma obtains again from substituting this into equation (B.5). □