Random redistribution and fiscal discrimination*

Stéphane Gauthier†
PSE, University of Paris 1 and Institute for Fiscal Studies

Guy Laroque‡
Sciences-Po, UCL and Institute for Fiscal Studies

October 13, 2023

Abstract

We examine optimal random nonlinear income taxation in a Mirrlees economy with a continuum of risk-averse agents whose utility functions are quasilinear in labor. Bunching in the deterministic optimum is necessary for socially beneficial random redistribution. In a parametric example featuring a Rawlsian redistribution motive, redistribution through lotteries dominates the optimal deterministic policy. Fiscal discrimination that is unattainable in the deterministic case due to bunching becomes feasible in the stochastic case. Beneficial randomizations favor the population that is most, rather than least, risk-averse. They operate through a reversal of incentives from a downward incentive pattern in the deterministic case to an upward pattern in the stochastic case, aligning incentives with the redistribution motive.

JEL classification numbers: C61, D82, D86, H21.

Keywords: redistribution, asymmetric information, bunching, random taxation.

*We are very grateful to Francis Bloch, Ben Brooks, Piotr Dworczak, Marc Fleurbaey, Deniz Kattwinkel, David Martimort, Nicolas Werquin, Floris Zoutman and participants to the Public Finance and Mechanism Design session of the 2023 conference of the Society for the Advancement of Economic Theory in Paris for feedback on our work. We are responsible for all remaining errors.

†stephane.gauthier@univ-paris1.fr; PSE, 48 bd Jourdan, 75014 Paris, France.

‡g.laroque@ucl.ac.uk; UCL, Drayton House, 30 Gordon St, London WC1H 0AX, UK.

1 Introduction

In a first-best environment the government is assumed to have the ability to observe the innate traits of individuals, which serve as a basis for the design of redistribution policies. Financial assistance, such as food stamps, housing subsidies or cash benefits can be provided to the poor, disabled persons or those with qualifying medical conditions preventing gainful activity. The spectrum of policies available to the government is more limited when some relevant characteristics of individuals are not publicly known. Extra resources then have to be spent to perform a suitable targeting of income support to those in need, implying a balance between efficiency and equity concerns. A fundamental insight of Mirrlees (1971) is to provide us with a formal representation of the additional costs from asymmetric information. When relevant traits are privately known to agents, the government must also take into account incentive constraints to meet self-selection among those genuinely in need of support, while deterring claims from those who do not require assistance.

The existing literature shows that the burden of asymmetric information typically falls on those in need, rather than on those in more favorable situations. Indeed the typical response to asymmetric information in redistribution is to reduce the amount of assistance. The lower aid enables the government to target those in need as beneficiaries of assistance; others are discouraged.

Several ideas have been explored to expand possible redistribution through improved targeting of assistance. Most of them involve introducing some form of ordeal mechanism, in the spirit of Nichols and Zeckhauser (1982). A better targeting can be achieved by subjecting vulnerable populations to challenging tasks or stressful situations. The government may for instance rely on time-consuming, possibly shameful queuing to distribute essential goods to low-income households. It may also implement unnecessarily complex and lengthy application processes to prove eligibility for benefits, or impose additional conditions after admission to continue receiving benefits, such as requiring beneficiaries to regularly send their children to school or undergo health check-ups. The social usefulness of these complementary schemes depends on balancing the direct cost borne by the targeted population and the benefits from relaxed incentive constraints that discourage undue claims.

In these examples ordeal is usually taken as deterministic, i.e., pain comes for sure. In this paper, we are interested into a specific form of ordeal, which is to impose random noisy transfers to risk averse recipients. It is known from Weiss (1976), Stiglitz (1981), Stiglitz (1982) or Brito, Hamilton, Slutsky, and Stiglitz (1995) that deterministic redistribution sometimes is socially dominated. In Lang (2017), Ederer, Holden, and Meyer (2018) or Lang (2023), randomization limits gaming social rules by blurring incentives. Vague standards or legal uncertainty deter firms to undertake strategies detrimental to the society; for instance, hospitals may be discouraged from selecting healthy but less costly to treat patients if they are not aware of the exact amount of compensation they will receive. In public finance, income lotteries may follow from random noise in taxes, because of e.g., administration errors or non-comprehensive auditing. In a mechanism design approach in line with Dworczak, Kominers, and Akbarpour (2021), randomness amounts to rely on quotas and rationing in the allocation of certain goods or services, as limits on market transactions, along with other constraints on transferability, lead some agents to engage in trade with strictly interior probabilities (with some risk of being rationed). In this vein, random labor and before-tax income variations based on employment status can be induced by the minimum wage and the risk of unemployment; they can also be due to randomness in occupations when, e.g., students who apply for medical training in the Netherlands are accepted by draw. But perhaps the most explicit randomizations concern situations where the government relies on random allocation of after-tax income and consumption goods for redistributive purposes. Income tests for housing subsidies make support 'available only for an accidentally or arbitrarily selected few' in Tobin (1971) while randomness improves selectivity in housing access for low-income populations in Weitzman (1977).

Solutions to optimization programs that we use in economics to characterize optimal policies may involve randomness because of failures of convexity assumptions. In the presence of asymmetric information, these failures come from the fact that the allocation intended for a given agent influences both her utility and the utility of those who are willing to mimic her, thus appearing in both sides of the incentive constraints. Randomizing over allocations located on the frontier of the set delimited by the constraints leads to allocations on the convex hull of the constraints of the deterministic program. The deterministic and random constraint sets differ in the presence of nonconvexities, which makes it possible for some allocations within the convex hull to outperform the best deterministic alternative.[1]

The general flavor of the economic argument seems very simple to grasp. Suppose that the government would like to redistribute income to low-skilled in a population of risk-averse workers. Redistribution is potentially limited if the government observes neither skill nor the exact amount of labor, as high-skilled then might reduce labor effort to enjoy higher transfers. Ordeal from randomness in the after-tax income designed for low-skilled is detrimental to their welfare, but this also expands the scope of possible redistribution by discouraging risk-averse high-skilled from relaxing effort. A deterministic optimum obtains if the welfare cost incurred by those facing noise overcomes the gain from expanded scope of redistribution. This is more likely to happen if high-skilled do not suffer much from the income noise. Hellwig (2007) indeed shows that the government should rely on deterministic redistribution if risk aversion decreases with labor productivity, i.e., risk aversion is higher among low-skilled individuals than among high-skilled individuals. This leaves us with little hope for randomized taxes to improve the welfare of the poor, who are usually found more risk averse than the rich.

Although the above argument for useful randomization sounds highly intuitive, it does not accord with a puzzling parametric example in Strausz (2006). In this example, a regulatory authority would like to implement a first-best policy that violates incentive compatibility, fitting the familiar pattern with high production cost (inefficient) firms ready to mimic low cost (efficient) firms. The profit of inefficient firms displays greater concavity, implying greater risk aversion to random production requirements that would be set by

[1] See also Pavlov (2011), Gauthier and Laroque (2014), Pycia and Unver (2015) or Gauthier and Laroque (2017) for related approaches.

the authority. Nevertheless the second-best regulatory policy involves a random option designed for the inefficient firms, which are the most rather than the less risk-averse firms.

Our paper provides an example in the same vein as Strausz (2006) in an optimal taxation setup with a continuum of risk-averse taxpayers. The example puts forward a new channel for socially useful random policies. It combines bunching in the deterministic optimum and a reversal of incentives.

We consider a Rawlsian government that only values the agents with the lowest utility. If incentive compatibility issues could be dealt with the first-order approach that neglects the possibility of bunching, the best deterministic redistribution policy would involve socially disfavored types envying the option designed for socially favored types. However, the strong redistribution motive underlying Rawlsian criteria leads to a severe form of bunching where many different taxpayers, including those socially favored, enjoy the same income transfers and eventually earn the same after-tax income. Thus, though government would like to discriminate some taxpayers, this is prevented for incentive reasons; no redistribution relying on deterministic tax tools is possible among them.

The uniform treatment of the agents in the deterministic optimum with bunching blurs the pattern of incentives. Indeed every agent could then be viewed as willing to mimic no other agent, as well as envied by the others. In this razor's edge situation, small tax randomizations are enough to allow the government to exploit heterogeneity in risk aversion in a way that reverses the pattern of incentives. Indeed we exhibit a special design of random transfers from the deterministic optimum that makes the agents considered by the Rawlsian government now envying the treatment of those with lower social importance, a feature reminiscent to countervailing incentives. That is, tax randomization allows for some fiscal discrimination by aligning incentives with the social desire for redistribution. A similar reversal occurs in Strausz (2006), but not in Hellwig (2007) where the same structure of incentives prevails both in the deterministic and stochastic cases. Actually it should be clear that the disappointing outcome for pro-poor policies in Hellwig (2007) precisely relies on the fact that high-skilled types continue to envy the low skilled once random noise is introduced into the tax system. Our example shows that this is not a general property.

The gain from the aligned incentives however comes with a cost, as the favored by the government have to face randomness. In a particular specification of our model the gain from alignment overcomes the cost, and so redistribution should involve a random income for the socially favored (lowest utility) agents, though they display the highest risk aversion. This provides an incentive-based justification for the risk of unemployment induced by the minimum wage, and other forms of rationing commonly used in social assistance and housing policies.

The paper proceeds as follows. Our setup is described in Section 2. Section 3 characterizes the role played by bunching in the deterministic optimum, which is exploited in the parametric example in Section 4. Concluding comments are in Section 5.

2 General framework

A government wants to redistribute income between a continuum of agents in a population of total unit size. Heterogeneity across agents is characterized by θ, a real parameter taking values in $\Theta = \left[\theta^{\inf}, \theta^{\sup}\right]$, which is referred to as the type of the agent. It has cumulative distribution function $F : \Theta \to [0, 1]$ associated with positive density $f : \Theta \to \mathbb{R}_{++}$.

The preferences of a type θ agent are represented by the quasilinear utility function

$$u(c, \theta) - y \tag{1}$$

when she earns before-tax income y and pays $y - c$ as tax. The after-tax income c is also her consumption. Earning y requires providing an effort, hence the disutility cost.

The function u is increasing, differentiable everywhere in c and θ, and strictly concave in c, so that every agent is risk-averse. Its first derivative $u'_c(c, \theta)$ with respect to c coincides with the marginal rate of substitution of consumption for before-tax income for a type θ agent. We assume that lower values of θ signal a higher importance given to marginal increase in consumption: the utility function $u(c, \theta)$ satisfies the Spence-Mirrlees condition that the cross-derivative $u''_{c\theta}(c, \theta)$ is negative for all (c, θ). The formulation (1), where θ is an argument of the function u, suggests to interpret this parameter as related to consumption tastes. However, following Lollivier and Rochet (1983), θ could also measure labor productivity or skill of the agent. Indeed we expect that low-skilled, presumably poor agents who have lower consumption, value more consumption than the rich.

The government designs a redistribution policy between these agents. A policy is defined by a menu $(\tilde{c}(\theta), \tilde{y}(\theta))$ of after and before-tax income lotteries. The menu is feasible if aggregate consumption falls below aggregate production,

$$\int_{\Theta} \mathbb{E}\left[\tilde{c}(\theta) - \tilde{y}(\theta)\right] \, \mathrm{d}F(\theta) \leq 0. \tag{2}$$

The government is assumed to know the distribution of types, but does not observe the value of θ. This value remains private information to the agent. Therefore the government must also ensure that every agent chooses the income pair designed for her. This is satisfied if the incentive constraints

$$\mathbb{E}\left[u(\tilde{c}(\theta), \theta) - \tilde{y}(\theta)\right] \geq \mathbb{E}[u(\tilde{c}(\tau), \theta) - \tilde{y}(\tau)] \tag{3}$$

hold for all (θ, τ) in $\Theta \times \Theta$.

Let $\tilde{V}(\theta) = \mathbb{E}\left[u(\tilde{c}(\theta), \theta) - \tilde{y}(\theta)\right]$ denote the expected indirect utility of type θ. The government is utilitarian. Its social welfare objective is

$$\int_{\Theta} \tilde{V}(\theta) \, \mathrm{d}G(\theta), \tag{4}$$

where the social weights embodied in $G(\cdot)$ are non-negative and normalized so that they sum up to 1. In the polar Benthamite utilitarian case, $F(\theta) = G(\theta)$ for all θ, every agent is given the same social importance. Differences between the two distributions capture social redistribution desires. If, for instance, $G(\theta) \geq F(\theta)$ for all θ, then agents with low types are given higher social importance.

An optimal redistribution policy is a menu $(\tilde{c}(\theta), \tilde{y}(\theta))$ that maximizes the objective (4) subject to the feasibility constraint (2) and the incentive constraints (3). The policy is deterministic if it is made of degenerate lotteries $(c(\theta), y(\theta))$ where every type θ earns the before-tax income $y(\theta)$ with certainty and consumes $c(\theta)$ with certainty (the absence of a tilde mark applies to deterministic options). We are interested into the circumstances where some agents face non-degenerate lotteries in the optimal redistribution policy.

3 Discrimination through randomness

The quasilinear form of the utility in (1) implies that replacing the lottery $\tilde{y}(\theta)$ with the sure outcome $y(\theta) = \mathbb{E}[\tilde{y}(\theta)]$ affects neither the constraints nor the objective. There is accordingly no loss to restrict to the case where every type θ earns before-tax $y(\theta)$ with certainty.

Following Laffont and Martimort (2002) we define the optimal 'relaxed' redistribution policy as a menu $(\tilde{c}(\theta), y(\theta))$ maximizing the social objective (4) subject to the feasibility constraint (2) and the necessary first-order conditions for a truthful report,

$$\tilde{V}'(\theta) = \mathbb{E}\left[u'_\theta(\tilde{c}(\theta), \theta)\right]$$

for all θ. The optimal relaxed redistribution policy coincides with the optimal redistribution policy if it meets the incentive constraints (3), but not otherwise. Summing up the first-order conditions yields

$$\tilde{V}(\theta) = \tilde{V}(\theta^{\inf}) + \int_{\theta^{\inf}}^{\theta} \mathbb{E}\left[u'_\theta(\tilde{c}(z), z)\right] \mathrm{d}z.$$

Replacing $y(\theta)$ with $\mathbb{E}\left[u(\tilde{c}(\theta), \theta)\right] - \tilde{V}(\theta)$ in the feasibility constraint (2) gives the indirect utility $\tilde{V}(\theta^{\inf})$ of the least type $\theta^{\inf}$, after using the integration by parts formula,

$$\tilde{V}(\theta^{\inf}) = \int_{\theta^{\inf}}^{\theta^{\sup}} \left(\mathbb{E}\left[u(\tilde{c}(\theta), \theta) - \tilde{c}(\theta)\right] - m(\theta)\mathbb{E}\left[u'_\theta(\tilde{c}(\theta), \theta)\right]\right) \mathrm{d}F(\theta),$$

with $m(\theta) = (1 - F(\theta))/f(\theta)$ being the Mills ratio. Reintroducing the expression of $\tilde{V}(\theta)$ into the objective (4), and using once again the integration by parts formula, the menu $(\tilde{c}(\theta))$ of after-tax income lotteries in an optimal relaxed redistribution program maximizes

$$\int_{\theta^{\inf}}^{\theta^{\sup}} \mathbb{E}\left[W(\tilde{c}(\theta), \theta)\right] \mathrm{d}F(\theta) \tag{5}$$

where $W(c,\theta) = u(c,\theta) - c - m(\theta)u'_\theta(c,\theta)$ represents the virtual contribution of type θ to social welfare when she faces after-tax income c with certainty.

In the optimal deterministic relaxed redistribution policy, every type θ gets for sure

$$c^*(\theta) = \arg\max_c W(c,\theta).$$

Suppose that this policy coincides with the optimal deterministic redistribution policy, so that there is no bunching in the deterministic optimum. We have then $W(c^*(\theta),\theta) \geq W(c,\theta)$ for all c and θ. It follows that $W(c^*(\theta),\theta) \geq \mathbb{E}[W(\tilde{c},\theta)]$ for all $\tilde{c}$ and θ. In particular the inequality holds true if for all θ we set $\tilde{c}$ equal to the lottery $\tilde{c}(\theta)$ that maximizes (5). This yields:

Lemma 1. *A random redistribution policy is socially useless if the optimal deterministic redistribution policy coincides with the optimal deterministic relaxed policy, i.e., the optimal deterministic redistribution policy involves no bunching.*

Proof. The argument given above leads to

$$\int_{\theta^{\inf}}^{\theta^{\sup}} \mathbb{E}[W(\tilde{c}(\theta),\theta)]\,\mathrm{d}F(\theta) \leq \int_{\theta^{\inf}}^{\theta^{\sup}} W(c^*(\theta),\theta)\,\mathrm{d}F(\theta).$$

for every menu $(\tilde{c}(\theta))$. To conclude the proof, observe that social welfare in an optimal redistribution policy cannot be greater than welfare in an optimal relaxed policy (which appears in the left-hand side of the above inequality). ∎

Lemma 1 shows that socially useful randomness in redistribution obtains only if the incentive constraints associated with a deterministic policy require bunching, with different types of agents facing the same income option. The government would like to rely on deterministic discrimination by giving with certainty $c^*(\theta)$ to type θ agents, as this quantity typically varies with θ. By Lemma 1, if a random allocation of consumption is to be socially useful by improving upon deterministic redistribution, then some type has to prefer an option designed for another type of agent in the relaxed menu $(c^*(\theta))$. Incentive requirements are not met. Actually every quantity $c^*(\tau)$ designed for types τ close to θ will then be preferred to $c^*(\theta)$ by type θ. The government solves this problem by giving the same consumption to each type. This suggests a possible role for randomness as a discriminatory tool for the government. Random transfers are beneficial only if a non-uniform treatment is socially desirable, but cannot be achieved relying on deterministic tools.

4 An example of useful randomization

We consider a Rawlsian government, i.e., the social objective only refers to the type of agents with the least utility, facing a population of agents whose preferences are represented by

$$u(c,\theta) - y = \ln(c+\theta) - y, \tag{6}$$

and we set $\theta^{\inf} = 0$.

4.1 Deterministic optimum

Let $V(\theta)$ denote the indirect utility $\ln(c(\theta)+\theta)-y(\theta)$ of type θ when she faces the pair $(c(\theta),y(\theta))$ with certainty. The incentive constraints (3) simplify to

$$V(\theta)=\ln(c(\theta)+\theta)-y(\theta)\geq\ln(c(\tau)+\theta)-y(\tau) \tag{7}$$

for all θ and τ. They are satisfied if and only if, for all θ,

$$V'(\theta)=\frac{1}{c(\theta)+\theta} \tag{8}$$

and $c(\theta)$ is non-increasing in θ.

Incentive considerations imply an increasing indirect utility with type, $V'(\theta)\geq 0$ for all θ. As a result, the Rawlsian government only values the least type θ^{inf} of agents. It has however to incur a cost to ensure self-selection among higher types. Indeed type θ has to receive the informational rent $\ln(c(\tau)+\theta)-\ln(c(\tau)+\tau)$ so that she does not mimic type τ. This rent is positive for $\tau<\theta$, i.e., incentives go downward, with high types envying lower neighboring types.

Using (5) the optimal deterministic relaxed redistribution policy maximizes

$$V(\theta^{\text{inf}})=\int_{\theta^{\text{inf}}}^{\theta^{\text{sup}}}\left[\ln(c(\theta)+\theta)-c(\theta)-\frac{m(\theta)}{c(\theta)+\theta}\right]\mathrm{d}F(\theta). \tag{9}$$

The after-tax income $c^*(\theta)$ designed for type θ that maximizes pointwise $V(\theta^{\text{inf}})$ is a non-negative root of the first-order condition $(c^*(\theta)+\theta)^2-(c^*(\theta)+\theta)-m(\theta)=0$. Since the Mills ratio $m(\theta)$ is non-negative, there is only one possible such root,

$$c^*(\theta)=\frac{1+(1+4m(\theta))^{1/2}}{2}-\theta. \tag{10}$$

This is an increasing quantity if and only if

$$m'(\theta)>(1+4m(\theta))^{1/2}. \tag{11}$$

The above inequality cannot be met for standard probability distributions, as they display a decreasing Mills ratio, $m'(\theta)\leq 0$. It may however be satisfied for well-chosen log-logistic, Weibull, and variants of Weibull distributions such as generalized or power generalized Weibull commonly used in econometric models for duration data (see Section 4.5). Where (11) holds true, the quantity $c^*(\theta)$ violates the second-order monotonicity requirements for incentive compatibility.

The main argument in favor of tax randomization can be easily grasped in the scenario where the Rawlsian government to equalize income across all agents by setting $y(\theta)=y^*$ and $c^*(\theta)=c^*$ for all θ, so that (11) is satisfied for nearly all types. This alternative can be implemented in a decentralized setup through the use of a progressive income tax schedule with two brackets, where before-tax income is taxed at a given low rate if falling below y^* while it is taxed a higher enough tax rate otherwise. More general schedules are examined in Section 4.4. In this polar case, feasibility requires $y^* = c^*$, so that there is no redistribution at all in the optimal deterministic policy. The utility of the least type, $V(\theta^{\inf}) = \ln(c^* + \theta^{\inf}) - c^*$, is maximized by setting $c^* = 1 - \theta^{\inf} = 1$ (since $\theta^{\inf} = 0$), which yields a level of social welfare $V(\theta^{\inf}) = -1$.

4.2 A reversal of incentives

We now randomize after-tax income, which becomes $\tilde{c}(\theta) = c^* + \tilde{\varepsilon}(\theta)$ where type-specific realizations of the random variable $\tilde{\varepsilon}(\theta)$ stand close to 0. We restrict our attention to perturbations where a greater noise comes with a greater income transfer. The mean and variance of $\tilde{\varepsilon}(\theta)$ are parameterized so that $\mathbb{E}[\tilde{\varepsilon}(\theta)] = \mathrm{var}[\tilde{\varepsilon}(\theta)] = \lambda v(\theta)$, with λ a positive real number close to 0, and $v(\theta) \geq 0$ a (rescaled) variance bounded from above.

Relying on lotteries allows the government to take advantage of agents' risk aversion. The (second-order Taylor expansion of the) expected utility of type θ when she chooses the lottery $\tilde{c}(\tau)$ designed for some type τ writes

$$\mathbb{E}[u(c^* + \tilde{\varepsilon}(\tau), \theta)] \simeq u(c^*, \theta) + \lambda u_c'(c^*, \theta)\left(1 - \frac{A(c^*, \theta)}{2}\right) v(\tau), \tag{12}$$

where

$$A(c, \theta) = -\frac{u_{cc}''(c, \theta)}{u_c'(c, \theta)} = \frac{1}{c + \theta} > 0$$

is the coefficient of absolute risk aversion of type θ. The random noise which affects after-tax income allows the government to exploit heterogeneity in individual risk aversion, captured by $A(c^*, \theta)$. The specific form of preferences for consumption makes the coefficient of risk aversion decreasing with θ. The least type $\theta^{\inf}$ agents, who are also those that the government would like to favor through deterministic taxation, display the highest risk aversion. As will be shown, however, this pattern does not prevent useful randomization.

Let $y(\theta)$ denote the before-tax income of type θ in the presence of such after-tax income perturbations; it is actually $y^* + \mathrm{d}y(\theta)$ for some deterministic $\mathrm{d}y(\theta)$ close to 0. The indirect utility $\tilde{V}(\theta) = \mathbb{E}[u(c^* + \tilde{\varepsilon}(\theta), \theta)] - y(\theta)$ of type θ under thruthful reporting can be expressed as

$$\tilde{V}(\theta) \simeq u(c^*, \theta) + U(\theta), \tag{13}$$

where

$$U(\theta) = \lambda S(c^*, \theta) v(\theta) - y(\theta) \tag{14}$$

and

$$S(c, \theta) = u_c'(c, \theta)\left(1 - \frac{A(c, \theta)}{2}\right) = \frac{1}{c + \theta}\left(1 - \frac{1}{2}\frac{1}{c + \theta}\right). \tag{15}$$

The incentive constraints (3) then reduce to

$$U(\theta) = \lambda S(c^*, \theta)v(\theta) - y(\theta) \geq \lambda S(c^*, \theta)v(\tau) - y(\tau). \tag{16}$$

for all τ and θ. That is, these constraints only involve a component $U(\theta)$ of the overall utility $\tilde{V}(\theta)$ of the agents, with monotonicity properties of this sub-utility that may differ from those of $\tilde{V}(\theta)$. From a formal viewpoint, our parametrization of lotteries gives rise to incentive constraints which have the familiar quasilinear shape in $v(\theta)$ and $y(\theta)$. Usual textbook arguments following the methodology developed by Myerson (1981) can thus be used to deal with lotteries. They yield:

Lemma 2. *Consider a menu where the government uses random perturbations $\tilde{\varepsilon}(\theta)$ to the optimal deterministic after-tax income c^* with mean and variance $\lambda v(\theta)$ for some $\lambda \geq 0$ close to* 0. *The incentive constraints (3) are satisfied if and only if*

$$U'(\theta) = \lambda S'_\theta(c^*, \theta)v(\theta)$$

and $v(\theta) \geq 0$ is non-increasing for all θ.

Proof. The sum of the incentive constraints linking two different types in (16) implies a non-increasing variance $v(\theta)$. Using the envelope theorem, a necessary first-order condition for a local truthful report is $U'(\theta) = \lambda S'_\theta(c^*, \theta)v(\theta)$ for all θ. Finally,

$$\frac{\partial}{\partial \tau}\left(S(c^*, \theta)v(\tau) - y(\tau)\right) = \int_\tau^\theta S'_\theta(c^*, z)v'(\tau)\,\mathrm{d}z.$$

This has the same sign as $\theta - \tau$ since $S'_\theta(c^*, z) < 0$ for all z. It follows that (3) holds for all τ and θ. This concludes the proof. ∎

Lemma 2 highlights a reversal in the incentive pattern following the introduction of noise. Indeed, since $\theta \geq 0$,

$$S'_\theta(c, \theta) = -\frac{1}{(c+\theta)^2}\left(1 - \frac{1}{c+\theta}\right) \leq 0 \tag{17}$$

when evaluated at $c = c^* = 1$. It follows that the sub-utility $U(\theta)$ driving incentives is decreasing with θ in the presence of noise ($v(\theta) > 0$ for some θ). This is so despite the fact that from (13) and Lemma 2 we have $\tilde{V}'(\theta) = u'_\theta(c^*, \theta) + \lambda S'_\theta(c^*, \theta)v(\theta) \simeq u'_\theta(c^*, \theta) > 0$ for λ close enough to 0 and $v(\theta)$ bounded from above. The overall utility $\tilde{V}(\theta)$ remains increasing with type, but the informational rent $\lambda(S(c^*, \theta) - S(c^*, \tau))v(\tau)$ given to type θ to avoid she mimics type τ now is positive for $\tau > \theta$, i.e., type θ now is willing to mimic type τ above her, not below her (as was the case for deterministic redistribution policies). This fits an upward neighboring pattern of incentives, much in contrast with the downward pattern that prevails when deterministic redistribution is used. The introduction of randomness in the redistribution policy thus succeeds to align incentives with social preferences.

However, by Lemma 2, incentive compatibility also requires a variance non-increasing with θ. This means that type $\theta^{\inf}$ agents face both the greatest income transfers and the greatest noise. As they also display the highest coefficient of risk aversion, it is not clear at this stage whether random redistribution can be socially useful.

4.3 Welfare improving randomization

Given the simple form taken by the incentive constraints, it is routine to compute the level of social welfare, which is the utility $\tilde{V}(\theta^{\inf})$ of the lowest type in the presence of random taxes. The utility $\tilde{V}(\theta^{\inf})$ can then be compared to the utility $V(\theta^{\inf}) = -1$ of this same lowest type of agents evaluated at the optimal deterministic optimum. Details are given in Appendix A. This leads to the following result:

Proposition 1. Optimal random redistribution. *Consider a menu where the government uses random perturbations $\tilde{\varepsilon}(\theta)$ to the optimal deterministic after-tax income c^* with mean and variance $\lambda v(\theta)$ for some $\lambda \geq 0$ close to 0. Let $v(\theta)$ be non-negative, non-increasing and bounded from above. The random menu improves upon the deterministic optimum if and only if*

$$\int_{\theta^{\inf}}^{\theta^{\sup}} \phi(c^*, \theta) v(\theta) \, \mathrm{d}F(\theta) > 0,$$

where

$$\phi(c, \theta) = -\left(1 - \frac{1}{c+\theta} + \frac{1}{2}\frac{1}{(c+\theta)^2}\right) + \frac{m(\theta)}{(c+\theta)^2}\left(1 - \frac{1}{c+\theta}\right).$$

An intuition for the trade-off in the condition given in Proposition 1 obtains by applying the so-called perturbation argument to the total collected tax. Once random noise is introduced, the tax paid by a θ agent is $y(\theta) - [c^* + \lambda v(\theta)]$. Using the definition of $U(\theta) = \lambda S(c^*, \theta) v(\theta) - y(\theta)$, the total collected tax can be written

$$\int_{\theta^{\inf}}^{\theta^{\sup}} [\lambda S(c^*, \theta) v(\theta) - U(\theta) - c^* - \lambda v(\theta)] \, \mathrm{d}F(\theta). \tag{18}$$

Consider a reform that increases the (rescaled) variance $v(\theta)$ of the after-tax income by a small amount $\mathrm{d}v$ for all types between θ and $\theta + \mathrm{d}\theta$, $\mathrm{d}\theta$ positive close to 0. These types, who are directly concerned by the reform, are in total number $f(\theta)\,\mathrm{d}\theta$.

Let us first focus on the change in collected tax abstracting from the adjustments needed to meet incentives, i.e., $U(\theta)$ is temporarily maintained fixed at its initial level. In doing so, one captures the net social cost of the randomization that transits through the noise bearing on the socially favored agents. Every type θ directly concerned by the reform works more, which increases her before-tax income by $\lambda S(c^*, \theta)\,\mathrm{d}v$. The total tax resources thus increase by $\lambda S(c^*, \theta) f(\theta) \, \mathrm{d}v \, \mathrm{d}\theta$. By assumption, the government takes advantage of the noise to increase the average transfer to every such agents, which costs $\lambda \, \mathrm{d}v$ per agent. Overall the change in collected tax is $\lambda \left[S(c^*, \theta) - 1\right] f(\theta) \, \mathrm{d}v \, \mathrm{d}\theta$. Using the expression of $S(c^*, \theta)$ given in (15), we have

$$S(c^*, \theta) < 1 \Leftrightarrow \frac{1}{c^* + \theta} \left(1 - \frac{1}{2} \frac{1}{c^* + \theta}\right) < 1,$$

a condition that is always satisfied since $c^* = 1$ and $\theta \geq 0$. Hence, given $U(\theta)$, the reform yields a lower amount of collected taxes, which represents a net cost for the society.

This cost has to be compared to the social gain from getting incentives aligned with redistribution tastes. The introduction of noise affects incentives through $U(\theta)$. Let us consequently allow $U(\theta)$ to respond to the reform. By Lemma 2, $U'(\theta)$ changes by $\mathrm{d}U'(\theta) = \lambda S'_\theta(c^*, \theta) \, \mathrm{d}v$ for every type directly concerned by the reform. It follows that the utility changes by $\mathrm{d}U = \mathrm{d}U'(\theta) \, \mathrm{d}\theta = \lambda S'_\theta(c^*, \theta) \, \mathrm{d}v \, \mathrm{d}\theta$ for every type above $\theta + \mathrm{d}\theta$, implying a change in total collected tax equal to $-(1 - F(\theta))\lambda S'_\theta(c^*, \theta) \, \mathrm{d}v \, \mathrm{d}\theta$. Since $S'_\theta \leq 0$ these are indeed additional tax resources. They arise due to the opportunity created by noise to increase before-tax income resources from socially disfavored high types.

The final change in taxes following the introduction of the noise is

$$\lambda \left[S(c^*, \theta) - 1\right] f(\theta) \, \mathrm{d}v \, \mathrm{d}\theta - (1 - F(\theta))\lambda S'_\theta(c^*, \theta) \, \mathrm{d}v \, \mathrm{d}\theta,$$

or equivalently,

$$\lambda \left[S(c^*, \theta) - 1 - m(\theta) S'_\theta(c^*, \theta)\right] f(\theta) \, \mathrm{d}v \, \mathrm{d}\theta.$$

Using the expressions of $S(c^*, \theta)$ and $S'_\theta(c^*, \theta)$ given in (15) and (17), the term into brackets is actually $\phi(c^*, \theta)$. It balances the loss from agents concerned by the reform (the increase in their production do not compensate the cost from the additional transfers they receive) and the greater taxes allowed by the reduced informational rents given to high types above θ.

4.4 Non-uniform partial deterministic bunching

So far we have considered the polar case of uniform bunching in the deterministic optimum where every type faces the same option (c^*, y^*). In practice, tax authorities often rely on complex tax schemes with more than two different tax brackets. If (11) is met almost everywhere, then the general form of the optimal deterministic tax schedule consists of a collection of income pairs (c_i^*, y_i^*) assigned to every agent with a type ranging from $\bar{\theta}_i$ to $\bar{\theta}_{i+1}$ ($\bar{\theta}_i < \bar{\theta}_{i+1}$). Condition (19) below provides a natural generalization of the uniform bunching alternative examined in Proposition 1 to such schedules with discontinuities.

Proposition 2. Non-uniform deterministic bunching. *Suppose that the optimal deterministic redistribution policy consists of n different tax brackets, with every type of agents in*

$\left[\bar{\theta}_i^*, \bar{\theta}_{i+1}^*\right)$ *earning* y_i^* *before tax and* c_i^* *after-tax. There exists a random policy that improves upon the deterministic optimum if*

$$\sum_{i=1}^{n} \int_{\bar{\theta}_i^*}^{\bar{\theta}_{i+1}^*} \phi(c_i^*, \theta) \mathrm{d}v(\theta)\, \mathrm{d}F(\theta) > 0, \tag{19}$$

for some non-increasing profile of after-tax income variance $(\mathrm{d}v(\theta))$ *close to* 0.

We outline the argument for the two-interval configuration $n = 2$ characterized by an interior threshold type $\bar{\theta}^*$ such that every type $\theta < \bar{\theta}^*$ earns c_1^* as after-tax income, while the remaining higher types $\theta \geq \bar{\theta}^*$ earns c_2^* ($c_1^* > c_2^*$). This deterministic schedule is dominated if the introduction of random noise on the after-tax income yields a higher amount of collected taxes while the socially favored type $\theta^{\inf}$ agents do not loose, $\mathrm{d}U(\theta^{\inf}) \geq 0$. The additional tax generated by the reform can then be redistributed to every agent through a uniform adjustment in before-tax income without violating incentive requirements.

The utility of every type $\theta < \bar{\theta}^*$ changes by

$$\mathrm{d}U(\theta) = \mathrm{d}U(\theta^{\inf}) + \int_{\theta^{\inf}}^{\theta} \lambda S'_\theta \left(c_1^*, z\right) \mathrm{d}v(z)\, \mathrm{d}z,$$

so that the total change in utility from these agents can be written, after using the integration by parts formula,

$$\mathrm{d}U(\theta^{\inf}) - \left[1 - F(\bar{\theta}^*)\right] \mathrm{d}U(\bar{\theta}^*) + \int_{\theta^{\inf}}^{\bar{\theta}^*} \lambda S'_\theta \left(c_1^*, \theta\right) m(\theta) \mathrm{d}v(\theta)\, \mathrm{d}z\, \mathrm{d}F(\theta).$$

Similarly, the utility of every type $\theta \geq \bar{\theta}^*$ changes by

$$\mathrm{d}U(\theta) = \mathrm{d}U(\theta^{\inf}) + \int_{\theta^{\inf}}^{\bar{\theta}^*} \lambda S'_\theta \left(c_1^*, z\right) \mathrm{d}v(z)\, \mathrm{d}z + \int_{\bar{\theta}^*}^{\theta} \lambda S'_\theta \left(c_2^*, z\right) \mathrm{d}v(z)\, \mathrm{d}z,$$

which now yields a total utility change for these agents equal to

$$\left[1 - F(\bar{\theta}^*)\right] \mathrm{d}U(\bar{\theta}^*) + \int_{\bar{\theta}^*}^{\theta^{\sup}} \lambda S'_\theta \left(c_2^*, \theta\right) m(\theta) \mathrm{d}v(\theta)\, \mathrm{d}F(\theta).$$

After reintroducing these utility changes into (18), the additional amount of collected tax implied by the introduction of random noise writes

$$\int_{\theta^{\inf}}^{\bar{\theta}^*} \lambda \phi(c_1^*, \theta) \mathrm{d}v(\theta)\, \mathrm{d}F(\theta) + \int_{\bar{\theta}^*}^{\theta^{\sup}} \lambda \phi(c_2^*, \theta) \mathrm{d}v(\theta)\, \mathrm{d}F(\theta) - \mathrm{d}U(\theta^{\inf}).$$

Given a profile $(\mathrm{d}v(\theta))$, which Lemma 2 shows must have a decreasing variance to meet incentive requirements, the highest amount of collected tax that does not hurt type $\theta^{\inf}$ agents obtains by setting $\mathrm{d}U(\theta^{\inf}) = 0$. A social welfare improvement therefore obtains if () is met for $n = 3$, with $\bar{\theta}_1^* = \theta^{\inf}$, $\bar{\theta}_2^* = \bar{\theta}^*$ and $\bar{\theta}_3^* = \theta^{\sup}$.

Remark. Partial bunching. In the numerical example in Section 4.5, the inequality (11) is not satisfied for θ high enough, and so the quantity $c^*(\theta)$ is decreasing at the top of the distribution. Condition (19) actually applies for $\bar{\theta}_{n+1}^* < \theta^{\sup}$, i.e., in the absence of bunching at the top of the distribution. Then, one can indeed set $\mathrm{d}v(\theta) = 0$ for types that are not concerned by bunching, $\theta \geq \bar{\theta}_{n+1}^*$. The change in collected tax is

$$\sum_{i=1}^{n} \int_{\bar{\theta}_i^*}^{\bar{\theta}_{i+1}^*} \lambda\phi(c_i^*, \theta)\mathrm{d}v(\theta)\,\mathrm{d}F(\theta) - \mathrm{d}U(\theta^{\inf}) + \left[1 - F(\bar{\theta}_{n+1}^*)\right] \mathrm{d}U(\bar{\theta}_{n+1}^*).$$

As above, one should set $\mathrm{d}U(\theta^{\inf}) = 0$ to maximize the additional tax revenue. In addition, types $\theta \geq \bar{\theta}_{n+1}^*$ now should be given $\left[1 - F(\bar{\theta}_{n+1}^*)\right] \mathrm{d}U(\bar{\theta}_{n+1}^*)$ to meet incentive compatibility.

4.5 Generalized Weibull specification

In this section we exhibit a specific parametrization with types distributed according to a generalized Weibull distribution where tax randomization is socially useful (see Dimitrakopoulou, Adamidis, and Loukas (2007) for properties of this distribution). Let

$$F(\theta) = 1 - \exp\left[1 - \left(1 + \lambda\theta^b\right)^a\right]$$

for $\theta \geq \theta^{\inf} = 0$, with a and b positive shape parameters and λ a positive scale parameter. The Mills ratio $m(\theta)$ is increasing for $a < 1$ and $b \leq 1$.

For $a = 0.5$, $b = 0.05$ and $s = 0.5$, (11) is satisfied for all $\theta \leq 19.91$, which corresponds to the 22.68 percent of the population with the lowest types. The optimal deterministic income tax schedule consists of a given income pair $(\bar{c}^*, \bar{y}^*)$ offered to every type $\theta \leq \bar{\theta}^*$ while all the other types are designed the optimal relaxed income pair $(c^*(\theta), y^*(\theta))$. The social objective $V(\theta^{\inf})$ thus is

$$\int_0^{\bar{\theta}^*} \left[\ln(\bar{c}^* + \theta) - \bar{c}^* - \frac{m(\theta)}{\bar{c}^* + \theta}\right] \mathrm{d}F(\theta) \tag{20}$$

$$+ \int_{\bar{\theta}^*}^{\theta^{\sup}} \left[\ln(c^*(\theta) + \theta) - c^*(\theta) - \frac{m(\theta)}{c^*(\theta) + \theta}\right] \mathrm{d}F(\theta).$$

Using the expression of $c^*(\theta)$ given in (10), the optimal threshold is such that

$$\bar{c}^* = c^*(\bar{\theta}^*) = \frac{1 + \left(1 + 4m(\bar{\theta}^*)\right)^{1/2}}{2} - \bar{\theta}^* \tag{21}$$

To characterize the amount of after-tax income $\bar{c}^*$, we apply the integration by parts formula and rewrite the contribution (20) as $-\left[1-F(\bar{\theta}^*)\right]\ln\left(\bar{c}^*+\bar{\theta}^*\right)+\ln\bar{c}^*-\bar{c}^*F(\bar{\theta}^*)$. Solving for the first-order condition for $\bar{c}^*$ to maximize this contribution, which is a quadratic equation in $\bar{c}^*$, the only positive root is

$$\bar{c}^*=\frac{1-\bar{\theta}^*}{2}+\frac{1}{2}\sqrt{\left(1-\bar{\theta}^*\right)^2+\frac{4\bar{\theta}^*}{F(\bar{\theta}^*)}}.$$

Replacing this expression of $\bar{c}^*$ into (21) defines the optimal threshold $\bar{\theta}^*$. Numerical computations (see the `R` code in Appendix B) yield $\bar{c}^*=4.02$ and $\bar{\theta}^*=74.87$, with $F(\bar{\theta}^*)=23.88$ percent. This completes the characterization of the optimal deterministic income tax schedule.

We now expose to random noise a subset of low types among those facing the after-tax income $\bar{c}^*$ in the deterministic income tax schedule. Namely we set $\mathrm{d}v(\theta)=\mathrm{d}v>0$ for all $\theta\leq\bar{\theta}$, with $\bar{\theta}$ some threshold type below $\bar{\theta}^*$. With this variance step-profile, condition (19) for socially useful randomness in redistribution can be re-expressed as

$$-F(\bar{\theta})-\frac{1-F\left(\bar{\theta}\right)}{\bar{c}^*+\bar{\theta}}\left(1-\frac{1}{2\left(\bar{c}^*+\bar{\theta}\right)}\right)+\frac{1}{\bar{c}^*}\left(1-\frac{1}{2\bar{c}^*}\right)>0.$$

The shape of the left-hand side is depicted in Figure 1. It is 0 when evaluated at $\bar{\theta}=0$ and decreasing for $\bar{\theta}$ close enough to 0, so that one should not set random taxes on a narrow subset of types close to $\theta^{\inf}$. For higher values of $\bar{\theta}$, it is single-peaked, reaching its global maximum for $\bar{\theta}=73.83$, hence close to $\bar{\theta}^*=74.87$. It takes positive values for all $\bar{\theta}\in\left[9.76,\bar{\theta}^*\right]$, with $F(9.76)=22.05$ percent.

We conclude that introducing random taxes designed for approximately the bottom quartile of the distribution improves upon the deterministic optimum. Relying on the interpretation of (19) as a change in collected tax following the introduction of small random noise on the after-tax income, the highest amount of additional taxes would be generated by subjecting almost all agents affected by bunching to random perturbations.

5 Conclusion

Our paper reexamines optimal income taxation in a Mirrlees setup with a continuum of types of taxpayers. Our focus is on the choice between deterministic versus random taxation. We have shown that the random alternative is preferred only if the best deterministic policy implies a uniform treatment of different types of taxpayers. The introduction of randomness then allows the government to exploit taxpayers' risk aversion and implement discriminatory tax treatment.

The existing literature following Hellwig (2007) suggests that rationing, viewed as a form of random allocation of goods designed for the poor, may be justified as far as these agents display lower risk aversions. However we expect a greater, not lower aversion to

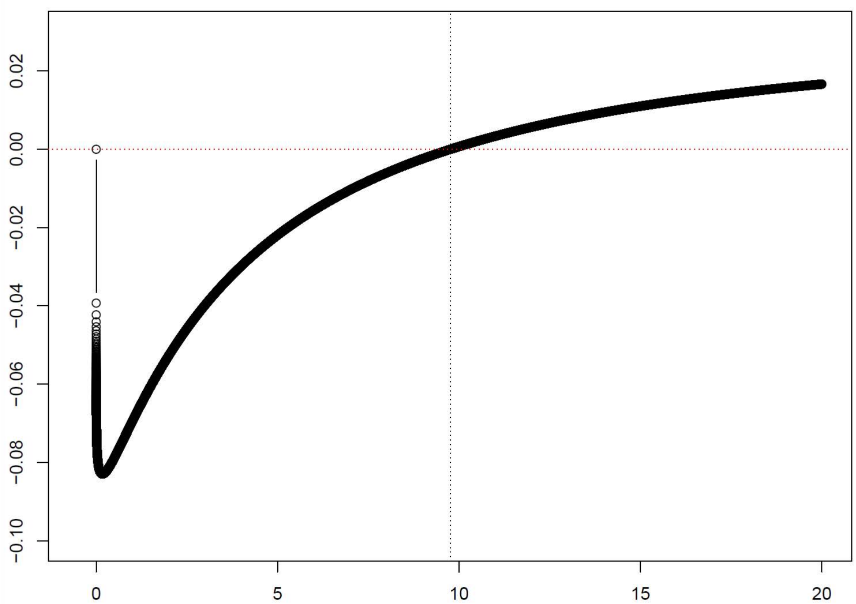

Figure 1: Random redistribution with a generalized Weibull distribution

consumption risk among the poor, who have the lowest consumption. Our paper shows that stochastic redistribution can be socially useful even though random noise bears on the most risk averse agents. In this respect, it can be used to justify policies relying on rationing the less well-off part of the population to improve its welfare.

The social acceptability of more explicit forms of randomization of taxes may be disputable. However, administrative errors, where, for instance, the assessment of before-tax income may be inaccurate, make a deterministic tax rule consistent with small random income perturbations. It is likely that they also bear on the less well-off part of the population. In practice, such errors plausibly affect the most vulnerable, as they usually fall into several social benefit regimes. In the redistributive case, where these agents have low risk aversions and high social importance, our results suggest that it may be wasteful to correct these errors.

Two features in our parametric example would be worth addressing in further work. First, we considered the case of a Rawlsian planner, which magnifies tensions from redistribution. A continuity argument suggests that the results should remain unaffected for redistributive Utilitarian preferences that place greater importance on agents who value consumption more. On the other hand, the occurrence of bunching in the deterministic optimum may be less plausible for weak redistribution motives, e.g., the unweighted (Benthamite) utilitarian social welfare objective, which suggests that low redistribution should be made deterministically.

A second feature relates to the interplay between the extent of bunching and optimal randomization. In our numerical example, bunching concerns a significant portion of the population, and most of agents affected by failures of monotonicity requirements should face random taxation, while redistribution should be deterministic at the top of the distribution. This suggests that deterministic redistribution instead is more suitable in economic circumstances where a smaller part of the population is affected by bunching.

References

Brito, D., J. Hamilton, S. Slutsky, and J. Stiglitz (1995): “Randomization in optimal income tax schedules,” *Journal of Public Economics*, 56, 189–223.

Dimitrakopoulou, T., K. Adamidis, and S. Loukas (2007): “A Lifetime Distribution With an Upside-Down Bathtub-Shaped Hazard Function,” *IEEE Transactions on Reliability*, 56(2), 308–311.

Dworczak, P., S. D. Kominers, and M. Akbarpour (2021): “Redistribution through markets,” *Econometrica*, 89(4), 1665–1698.

Ederer, F., R. Holden, and M. Meyer (2018): “Gaming and strategic opacity in incentive provision,” *The RAND Journal of Economics*, 49(4), 819–854.

Gauthier, S., and G. Laroque (2014): “On the value of randomization,” *Journal of Economic Theory*, 151, 493–507.

——— (2017): “Redistribution by means of lotteries,” *Journal of Economic Theory*, 169, 707–167.

Hellwig, M. (2007): “The undesirability of randomized income taxation under decreasing risk aversion,” *Journal of Public Economics*, 91, 791–816.

Laffont, J.-J., and D. Martimort (2002): *The theory of incentives: The principal-agent model*. Princeton University Press.

Lang, M. (2017): “Legal Uncertainty as a welfare enhancing screen,” *European Economic Review*, 91, 274–289.

——— (2023): “Stochastic Contracts and Subjective Evaluations,” *Rand Journal of Economics*, 54, 104–134.

Lollivier, S., and J.-C. Rochet (1983): “Bunching and Second-Order Conditions: A Note on Optimal Tax Theory,” *Journal of Economic Theory*, 31, 392–400.

Mirrlees, J. (1971): “An exploration in the theory of optimum income taxation,” *Review of Economic Studies*, 38, 175–208.

Myerson, R. (1981): “Optimal Auction Design,” *Mathematics of Operations Research*, 6, 58–73.

Nichols, A. L., and R. J. Zeckhauser (1982): “Targeting Transfers through Restrictions on Recipients,” *American Economic Review*, 72, 372–377.

Pavlov, G. (2011): “Optimal mechanism for selling two goods,” *The B.E. Journal of Theoretical Economics*, 11, 1–35.

Pycia, M., and U. Unver (2015): “Decomposing random mechanisms,” *Journal of Mathematical Economics*, 61, 21–33.

Stiglitz, J. (1981): “The Allocation Role of the Stock Market,” *The Journal of Finance*, 36, 235–251.

——— (1982): “Utilitarianism and horizontal equity: the case for random taxation,” *Journal of Public Economics*, 18, 1–33.

Strausz, R. (2006): “Deterministic versus stochastic mechanisms in principal-agent models,” *Journal of Economic Theory*, 128, 306–314.

Tobin, J. (1971): “On limiting the domain of inequality,” *Journal of Law and Economics*, 13, 263–277.

Weiss, L. (1976): “The desirability of cheating incentives and randomness in the optimal income tax,” *Journal of Political Economy*, 84, 1343–1352.

Weitzman, M. (1977): “Is the price system or rationing more effective in getting a commodity to those who need it most?,” *Bell Journal of Economics*, 8, 517–524.

A Proof of Proposition 1

We first express the before-tax income of type θ as a function of her indirect utility $\tilde{V}(\theta)$. From (13) and (14), we have

$$y(\theta) = u(c^*, \theta) + \lambda S(c^*, \theta)v(\theta) - \tilde{V}(\theta).$$

The expression of the indirect utility $\tilde{V}(\theta)$ obtains from the first-order necessary condition in Lemma 2 for incentive compatibility $U'(\theta) = \lambda S'_\theta(c^*, \theta)v(\theta)$ for all θ, which yields

$$U(\theta) = U(\theta^{\text{inf}}) + \int_{\theta^{\text{inf}}}^{\theta} \lambda S'_\theta(c^*, z)v(z)\,\mathrm{d}z.$$

Hence, using (13),

$$\tilde{V}(\theta) = u(c^*, \theta) + U(\theta^{\text{inf}}) + \int_{\theta^{\text{inf}}}^{\theta} \lambda S'_\theta(c^*, z)v(z)\,\mathrm{d}z.$$

The feasibility constraint (2) reads

$$\int_{\theta^{\text{inf}}}^{\theta^{\text{sup}}} [c^* + \lambda v(\theta) - y(\theta)]\ \mathrm{d}F(\theta) = 0$$

$$\Leftrightarrow \int_{\theta^{\text{inf}}}^{\theta^{\text{sup}}} \left[c^* + \lambda v(\theta) - \lambda S(c^*, \theta)v(\theta) + U(\theta^{\text{inf}}) + \int_{\theta^{\text{inf}}}^{\theta} \lambda S'_\theta(c^*, z)v(z)\,\mathrm{d}z \right]\ \mathrm{d}F(\theta) = 0.$$

Using the integration by parts formula,

$$\int_{\theta^{\text{inf}}}^{\theta^{\text{sup}}} \int_{\theta^{\text{inf}}}^{\theta} S'_\theta(c^*, z)v(z)\,\mathrm{d}z\,\mathrm{d}F(\theta) = \int_{\theta^{\text{inf}}}^{\theta^{\text{sup}}} \frac{1 - F(\theta)}{f(\theta)} S'_\theta(c^*, \theta)v(\theta)\,\mathrm{d}F(\theta),$$

the feasibility constraint allows us to get

$$U(\theta^{\text{inf}}) = -\int_{\theta^{\text{inf}}}^{\theta^{\text{sup}}} \left[c^* + \lambda v(\theta) - \lambda S(c^*, \theta)v(\theta) + \frac{1 - F(\theta)}{f(\theta)} \lambda S'_\theta(c^*, \theta)v(\theta) \right]\ \mathrm{d}F(\theta).$$

Finally (13) gives us the indirect utility $\tilde{V}(\theta^{\text{inf}}) = u(c^*, \theta^{\text{inf}}) + U(\theta^{\text{inf}})$, which is

$$u(c^*, \theta^{\text{inf}}) - \int_{\theta^{\text{inf}}}^{\theta^{\text{sup}}} \left[c^* + \lambda v(\theta) - \lambda S(c^*, \theta)v(\theta) + \frac{1 - F(\theta)}{f(\theta)} \lambda S'_\theta(c^*, \theta)v(\theta) \right]\ \mathrm{d}F(\theta).$$

In the absence of income noise, $v(\theta) = 0$ for all θ, we have

$$\tilde{V}(\theta^{\inf}) = u(c^*, \theta^{\inf}) - \int_{\theta^{\inf}}^{\theta^{\sup}} c^* \, \mathrm{d}F(\theta) = u(c^*, \theta^{\inf}) - c^* = V(\theta^{\inf}).$$

Therefore tax randomizations improve upon the deterministic optimum, $\tilde{V}(\theta^{\inf}) > V(\theta^{\inf})$, if and only if

$$\lambda \int_{\theta^{\inf}}^{\theta^{\sup}} \left[v(\theta) - S(c^*, \theta) v(\theta) + \frac{1 - F(\theta)}{f(\theta)} S'_\theta(c^*, \theta) v(\theta) \right] \mathrm{d}F(\theta) < 0.$$

After replacing $S(c^*, \theta)$ and $S'_\theta(c^*, \theta)$ with their expressions in (15) and (17), this inequality rewrites as in Proposition 1.

B R code for Weibull distribution

```
a <- 0.5; b <- 0.05; s <- 0.5 #

FF <- function(x) 1 - exp( 1 - (1+s*x^b)^a )
ff <- function(x)  ( a*(1+s*x^b)^(a-1)* s*b*x^(b-1)) * exp( 1 - (1+s*x^b)^a )
mm <- function(x) (1-FF(x)) / ff(x)
mmprime <- function(x) -(a*b*(b-1)*s*x^(b-2)*(1+s*x^b)^(a-1)
        + a*b*s*x^(b-1)*(a-1)*s*b*x^(b-1)*(1+s*x^b)^(a-2)) / (a*b*s*x^(b-1)*(1+s*x^b)^(a-1))^2
bunch <- function(x) mmprime(x) - (1+4*mm(x))^(1/2)

xx <- seq(1e-10,1e3,1e-2)
plot(xx, bunch(xx)) # bunching occurs for xx such that bunch(xx) is positive
max(xx[bunch(xx)>=0]); FF(max(xx[bunch(xx)>=0]))

thetabar <- function(x) x + ( (1-x)^2 +4*x/FF(x) )^(1/2) - ( 1+4*mm(x) )^(1/2)
plot(xx, thetabar(xx)) # threshold below which bunching occurs is such that thetabar = 0
min(xx[thetabar(xx)>=0]); FF(min(xx[thetabar(xx)>=0]))
theta <- min(xx[thetabar(xx)>=0])
theta

cbar <- (1-theta)/2 + ( (1-theta)^2 + 4*theta/FF(theta)  )^(1/2) / 2
cbar

sumphi <- function(x) -FF(x) - (1-FF(x))*(1-1/(2*(cbar+x)))/(cbar+x) + ( 1-1/(2*cbar) )/cbar
norm.sumphi <- function(x) sumphi(x) - sumphi(1e-10) # the normalization ensures that sumphi is 0 at 1e-10
xx <- seq(1e-10,theta,1e-3)
max(norm.sumphi(xx)); xx[norm.sumphi(xx)==max(norm.sumphi(xx))]
norm.sumphi(theta)
plot(xx, norm.sumphi(xx), type="b")
min(xx[norm.sumphi(xx)>0]); FF(min(xx[norm.sumphi(xx)>0]))

# Figure to be exported -- finer grid
xx <- seq(1e-10,2,1e-5)
plot(xx, norm.sumphi(xx), type="b", xlim=c(-0.5,20), ylim=c(-0.1,0.03))
  xx <- seq(2,20,1e-3)
  points(xx, norm.sumphi(xx), type="b")
  abline(h=0, col="red", lty="dotted")
  abline(v=min(xx[norm.sumphi(xx)>0]), lty="dotted")
```