Federal Reserve Bank of Minneapolis
Research Department

October 2004
Preliminary

Tax Riots*

Marco Bassetto

University of Minnesota
Federal Reserve Bank of Minneapolis

Christopher Phelan

Federal Reserve Bank of Minneapolis

ABSTRACT

This paper considers an optimal taxation environment where households must voluntarily report their incomes, and governments randomly audit and punish households found to be underreporting. We prove that the optimal mechanism derived using standard mechanism design techniques has a bad equilibrium (a tax riot) where households underreport their incomes, precisely because other households are expected to do so as well. We then consider three alternative approaches to designing a tax scheme when one is worried about bad equilibria.

*We are indebted to Narayana Kocherlakota for useful conversation. Marco Bassetto thanks the National Science Foundation for financial support. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis, the Federal Reserve System, or NSF.

1. Introduction

Two common characteristic of tax systems are that households must voluntarily report their incomes, and that governments randomly audit and punish households found to be underreporting. Furthermore, net taxes (taxes minus transfers) and government spending generally respond to aggregate shocks, usually for simple budgetary reasons. That is, the transfers and spending a government can afford given many households reporting low incomes is not as great as the transfers and spending a government can afford given many households reporting high incomes. This paper focuses generally on the design of tax/spending systems in such an environment and particularly on the question of existence of bad equilibria or "tax riots" where households underreport income precisely because other households are expected to underreport income.

The approach of the paper is to initially derive tax/spending policies using standard mechanism design tools. While it is well understood that mechanism design tools ensure only that truthful revelation of income is one of (possibly) multiple equilibria of the game induced by the optimal tax mechanism, it is tempting to consider the possibility of other, non-truthful, equilibria as a *theoretical* issue or a technical nuisance.[1] Our point in this paper is that non-truthful equilibria a real *economic* issue when it comes to taxation. Specifically, we argue the following point: The incentives sufficient to induce honest reporting of income if other households are expected to be honest are generally <u>insufficient</u> to induce honest

[1]The literature on *implementation* carefully considers more generally the question we consider here for our specific application, i.e., how one can ensure a good, unique equilibrium outcome. See Moore [12] and Jackson [8] for surveys of the literature under perfect information, and Jackson [8] for an analysis of Bayesian implementation. While the implementation literature only considers pure-strategy equilibria, we consider the potential for mixed-strategy equilibria. Empirically, these equilibria seem relevant: *some*, but not all households may be underreporting their income. For this reason, we cannot apply standard implementation theorems.

reporting if other households are expected to be dishonest. Other (bad) equilibria are not simply a possibility, but, in fact, certain. Given this, tax systems which are optimal using our traditional definition of optimality will generally be susceptible to tax riots. This implies that what is "optimal" may not be advisable if one is worried about bad equilibrium outcomes.

The forces at work ensuring this are likely present in a range of optimal taxation models far wider than the one we present. Mechanism design tools have been a useful tool in public finance since the pioneering work of Mirrlees [11]. More recently, the same tools have been adopted to study dynamic settings by Golosov, Kocherlakota and Tsyvinski [6], Golosov and Tryvinski [7], Albanesi and Sleet [1] and Da Costa and Werning [2]. It would be interesting to revisit the policy prescriptions in these papers when the potential for multiple equilibria is accounted for.

In section 2, we present a simple, static, tax/audit model. In it, households have random, privately observed endowments which they report, perhaps untruthfully, to the government. The government has the ability to randomly audit a given fraction of households and impose a finite punishment if a household is found to have misreported its endowment. In section 3, we characterize the optimal direct mechanism which, by construction, has as an equilibrium that all households report truthfully. Our main result is that when the number of households is large, there always exists another equilibrium of this mechanism where households with high endowments underreport their income. The following sections proceed to consider three alternative approaches to designing a tax scheme when one is worried about bad equilibria. In section 4, we derive a non-direct mechanism which implements the truthful equilibrium uniquely. In section 5, we consider dominant strategy implementation, and in section 6, a minimax direct mechanism approach. Section 7 concludes.

2. The Model

Consider a static world populated by a large finite number, $N+1$, of identical households. Each household receives an endowment $e \in \{\underline{e}, \overline{e}\}$. With probability p, a household draws a low endowment $e = \underline{e}$ and with probability $(1-p)$, it draws a high endowment $e = \overline{e}$. This probability p is an aggregate random variable with support $[\underline{p}, \overline{p}] \subset (0, 1]$ and density $f(p)$. (We will also consider the degenerate case where $\underline{p} = \overline{p}$, or p is known.[2] Given p, household realizations are independent. Households and the government do not directly observe p or the endowment realizations of other households. We do assume, however, there exists an ability for high endowment households to prove they are indeed high endowment households (say by displaying their endowment), but no corresponding ability for low endowment households to prove they are indeed low endowment households.

After endowment realizations, the government implements some tax mechanism and uses the revenues to make transfers to households or create a public good through a linear technology. After taxes, transfers, and public good production occur, the government can, at zero cost, audit a fraction $\overline{\pi}$ of households and observe the endowments of those audited. If it wishes, the government can, at this point, punish a household by destroying its ability to consume the consumption good. (This is assumed to be a pure "wasteful" punishment. The government cannot transfer the household's consumption to other households.)

Household preferences are represented by $u(c) + v(G)$, where c is the household's private consumption level and G is the common, per-capita level of the public good. Assume $u(0) = 0$, $u'(c) > 0$, $u''(c) < 0$, and $\lim_{c\to\infty} u'(c) = 0$. Further assume that $-u''(c)/u'(c)$ is weakly decreasing in c. (Decreasing absolute risk aversion, or DARA). Preferences over

[2] With this assumption, our model is close to the multiple agent model of Krasa and Villamil [9].

the public good are such that $v'(G) > 0$, $v''(G) < 0$, and $\lim_{G\to 0} v'(G) < +\infty$.[3] The last assumption implies that v is bounded below, so we normalize $v(0) = 0$.

3. Optimal Direct Mechanisms

Given the above model, a natural question to ask is how a government interested in maximizing ex-ante expected utility should tax, audit and spend. The revelation principle allows this search among mechanisms to be restricted to a simple class — direct mechanisms — but is subject to one crucial (and well known) caveat: while truth-telling will be, by construction, an equilibrium of the mechanism, it may not be the only equilibrium. Our strategy is to use standard mechanism design methods to characterize an optimal tax/audit mechanism and then ask whether these other equilibria indeed exist.

In a direct mechanism, households report their endowment realization to the government, and taxes and transfers depend on these reports. Optimal auditing and punishing is particularly simple: audit only those households which reveal the low endowment (by assumption, households which reveal $e = \overline{e}$ cannot be lying) and audit as many of them as possible given the assumption that at most a fraction $\overline{\pi}$ of households can be audited. Without loss of generality, if a household is audited and found to have lied, its consumption is set to zero.

A household's probability of being audited will depend on the number of households which announce a low endowment. Let $\overline{n}(N)$ be the largest integer such that $\frac{\overline{n}(N)}{N+1} \leq \overline{\pi}$, and n the number of households which report the low endowment. If $n \leq \overline{n}(N)$, it is optimal to audit all n households. If $n > \overline{n}(N)$, $\overline{n}(N)$ households will be audited. This auditing mechanism implies that a household which reports a low endowment has probability

[3] All of the results carry over to the case of no public good preference, i.e., $v(G) \equiv 0$, by setting G identically to 0.

$\pi(\frac{n}{N+1}) \equiv \min\{1, \frac{\overline{n}(N)}{n}\}$ of being audited. (As $N \to \infty$, this probability converges uniformly to $\min\{1, \frac{\overline{\pi}}{m}\}$ where $m = \frac{n}{N+1}$.)

A direct tax-spending system is a vector of functions $(\overline{\tau}, \underline{\tau}_0, \underline{\tau}_1, G)(m)$ which represents, respectively, the tax paid by those who reveal $e = \overline{e}$, the tax paid by those who reveal $e = \underline{e}$ and are not audited, the tax paid by those who reveal $e = \underline{e}$, are audited, and found to have been truthful, and the per-capita level of government spending G, all as functions of the fraction of households which report the low endowment, m. Government spending is required to be non-negative; taxes may be negative or positive, but must be no more than the endowment of the people they are levied upon.[4] Note that the functions $(\overline{\tau}, \underline{\tau}_1, \underline{\tau}_0, G)(m)$ implicitly depend on the number of agents. This will become important later as we take $N \to \infty$.

A tax/spending mechanism is considered *feasible* if aggregate taxes weakly exceed government spending for each realization of $n = 0, \ldots, N+1$,

$$\begin{aligned}(1)\quad & \frac{n}{N+1}\left[\pi\left(\frac{n}{N+1}\right)\ \underline{\tau}_1\left(\frac{n}{N+1}\right) + \left(1 - \pi\left(\frac{n}{N+1}\right)\right)\ \underline{\tau}_0\left(\frac{n}{N+1}\right)\right] + \\ & \left(1 - \frac{n}{N+1}\right)\ \overline{\tau}\left(\frac{n}{N+1}\right) - G\left(\frac{n}{N+1}\right) \geq 0.\end{aligned}$$

A tax/spending mechanism is considered *incentive compatible* if, for a high endowment household, the expected utility of revealing the high endowment weakly exceeds the expected utility of falsely revealing a low endowment. Let $\Delta(m)$ denote the difference in expected utility between truthfully announcing a high endowment and falsely claiming a low endowment,

[4] $\underline{\tau}_0(m)$ is only defined for values of m such that $m > \overline{\pi}$. To save on notation, we will sometimes write $(1 - \pi(m))\underline{\tau}_0(m)$ even for $m \leq \overline{\pi}$; this should be interpreted as 0. Similarly, $\overline{\tau}(m)$ is undefined when $m = 1$, and $\underline{\tau}_1(m)$ is undefined when $m = 0$, but $(1 - m)\overline{\tau}(m)$ should be interpreted as 0 when $m = 1$, and $m\underline{\tau}_1(m)$ should be interpreted as 0 when $m = 0$.

conditional on a fraction m of households realizing a low endowment. Thus,

$$(2)\quad \Delta\left(\frac{n}{N+1}\right) \equiv u\left(\overline{e} - \overline{\tau}\left(\frac{n}{N+1}\right)\right) + v\left(G\left(\frac{n}{N+1}\right)\right) - \left(1-\pi\left(\frac{n+1}{N+1}\right)\right) u\left(\overline{e} - \underline{\tau}_0\left(\frac{n+1}{N+1}\right)\right) - v\left(G\left(\frac{n+1}{N+1}\right)\right).$$

To be incentive compatible, an allocation $(\overline{\tau}, \underline{\tau}_1, \underline{\tau}_0, G)(m)$ must satisfy the incentive constraint

$$(3)\quad \int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{N} Q(n|p,N)\Delta\left(\frac{n}{N+1}\right)(1-p)f(p)dp \geq 0,$$

where $Q(n|p,N) \equiv \binom{N}{n}p^n p^{N-n}$ denotes the probability of exactly n out of N agents realizing the low endowment when each has an i.i.d. probability p of doing so. Note that $(1-p)f(p)$ is a household's updated density of p conditional on it receiving a high endowment (up to a constant).

An optimal symmetric allocation $(\overline{\tau}^*, \underline{\tau}_1^*, \underline{\tau}_0^*, G^*)(m)$ maximizes the utility of a given agent, or solves

$$(4)\quad \max_{(\overline{\tau},\underline{\tau}_1,\underline{\tau}_0,G)(m)} \int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{n} Q(n|p,N)\Big[(1-p)\ \Big[u\left(\overline{e} - \overline{\tau}\left(\frac{n}{N+1}\right)\right) + v\Big(G\Big(\frac{n}{N+1}\Big)\Big)\Big] + p\Big[\pi\left(\frac{n+1}{N+1}\right) u\left(\underline{e} - \underline{\tau}_1\left(\frac{n+1}{N+1}\right)\right) + \left(1-\pi\left(\frac{n+1}{N+1}\right)\right) u\left(\underline{e} - \underline{\tau}_0\left(\frac{n+1}{N+1}\right)\right) + v\Big(G\Big(\frac{n+1}{N+1}\Big)\Big)\Big]\Big] f(p)dp$$

subject to (1) and (3).

The objective function is continuous in the tax-spending functions, and the constraint set is compact, hence a solution to this problem exists.[5] We will assume that the solution is interior and that full risk sharing is not an incentive-compatible allocation, at least for all values of N above some threshold.

By construction, if all other households report truthfully, it is in the (weak) interest for each household to report truthfully. Thus one equilibrium of the optimal tax/spending mechanism $(\overline{\tau}^*, \underline{\tau}_0^*, \underline{\tau}_1^*, G^*)$ is for all households to tell the truth. But what if others are expected to lie? Is it optimal to lie if all other households are lying? The key to answering this question is characterizing the optimal allocation $(\overline{\tau}^*, \underline{\tau}_1^*, \underline{\tau}_0^*, G^*)$. In particular, an optimal allocation will imply a function $\Delta^*(m; N)$ — the difference in utility between truthfully announcing a high endowment and falsely claiming a low endowment, conditional on a fraction m of households reporting a low endowment. Whether other equilibria exist depends on the shape of $\Delta^*(m; N)$ which itself depends on how taxes vary in response to the aggregate realization of m. We show that when N and m are large, $\Delta^*(m; N) < 0$. This precisely implies that high endowment households wish to lie if they expect other high endowment households to lie as well.

We start characterizing $(\overline{\tau}^*, \underline{\tau}_0^*, \underline{\tau}_1^*, G^*)$ by deriving the first-order, necessary conditions for optimality. We write the Lagrange multipliers on the constraints (1) as

$$\mu\left(\frac{n}{N+1}\right)\int_{\underline{p}}^{\overline{p}} Q(n|p, N+1) f(p) dp,$$

and let λ be the multiplier on the incentive-compatibility constraint (3).

[5] Notice that the tax-spending functions have finite domain, (for a given N, there is only a finite number of possible realizations of the fraction of agents which report a low endowment). This implies the tax-spending functions are finite-dimensional vectors.

The first order conditions with respect to $\overline{\tau}$, $\underline{\tau}_1$ and $\underline{\tau}_0$ are

$$(5)\quad u'(\bar{e}-\overline{\tau}(m)) - \frac{\mu(m)}{1+\lambda} = 0, \qquad \text{for } m = 0, \frac{1}{N+1}, \ldots, \frac{N}{N+1}$$

$$(6)\quad u'(\underline{e}-\underline{\tau}_1(m)) - \mu(m) = 0, \qquad \text{for } m = \frac{1}{N+1}, \ldots, \frac{N}{N+1}, 1$$

$$(7)\quad \begin{aligned} &u'(\underline{e}-\underline{\tau}_0(m)) - \mu(m) - \lambda\, u'(\bar{e}-\underline{\tau}_0(m)) \frac{\int_{\underline{p}}^{\bar{p}}(1-p)[p^m(1-p)^{1-m}]^{N+1} f(p)dp}{\int_{\underline{p}}^{\bar{p}} p\, [p^m(1-p)^{1-m}]^{N+1} f(p)dp} = 0, \\ &\text{for } m = \frac{\bar{n}(N)+1}{N+1}, \ldots, \frac{N}{N+1}, 1. \end{aligned}$$

The first order conditions with respect to G is

$$(8)\quad \begin{aligned} &\left[1+\lambda\left[1-\frac{\int_{\underline{p}}^{\bar{p}}\left(\frac{m}{p}\right)[p^m(1-p)^{1-m}]^{N+1} f(p)dp}{\int_{\underline{p}}^{\bar{p}}[p^m(1-p)^{1-m}]^{N+1} f(p)dp}\right]\right] v'(G(m)) - \mu(m) = 0 \\ &m = 0, \frac{1}{N+1}, \ldots, 1. \end{aligned}$$

It is straightforward to show that, at an optimum, the resource constraint must be binding. We can thus rewrite it as

$$(9)\quad \begin{aligned} &m\left[\pi(m)\underline{\tau}_1(m) + (1-\pi(m))\underline{\tau}_0(m)\right] + (1-m)\overline{\tau}(m) - G(m) = 0 \\ &m = 0, \frac{1}{N+1}, \ldots, 1 \end{aligned}$$

Proposition 1. *Given λ, the resource constraint (9) and the first order conditions (5), (6), (7), and (8) define at most one set of functions $\mu^*(m)$, $\overline{\tau}^*(m)$, $\underline{\tau}_1^*(m)$, $\underline{\tau}_0^*(m)$, and $G^*(m)$.*

Proof. See appendix. ■

From now on, we use * superscripts to denote the optimal tax-spending plan and the

Lagrange multipliers that solve the first-order conditions.[6] Further, let $(\overline{\tau}^*, \underline{\tau}_1^*, \underline{\tau}_0^*, G)(m;N)$ denote the optimal tax/spending system for a given value of N. Note that for a given N, a tax/spending system $(\overline{\tau}, \underline{\tau}_1, \underline{\tau}_0, G)(m;N)$ is defined only for those fractions m which are compatible with N. (That is, $m \in \{0, \frac{1}{N+1}, \ldots, 1\}$). However, for each N, the first order conditions are nevertheless well defined for all $m \in [0,1]$ and we can consider their pointwise limits as $N \to \infty$.

Lemma 1. There exists a unique value λ^∞, and unique functions $\overline{\tau}^\infty(m)$, $\underline{\tau}_1^\infty(m)$, $G^\infty(m)$, $\mu^\infty(m)$ defined on $[0,1]$ and $\underline{\tau}_0^\infty(m)$ defined on $[\bar{\pi},1]$, that satisfy the limits of first-order conditions (5), (6), (7), (8), and (9), as $N \to +\infty$, and the limiting incentive-compatibility constraint

$$\int_{\underline{p}}^{\bar{p}} (1-m)\Big[u\left(\bar{e} - \overline{\tau}^\infty(m)\right) + v\left(G^\infty(m)\right)\Big] f(m)dm =$$
$$\int_{\underline{p}}^{\bar{p}} (1-m)\left[\max\left\{0, 1 - \frac{\bar{\pi}}{m}\right\} u\left(\underline{e} - \underline{\tau}_0^\infty(m)\right) + v\left(G^\infty(m)\right)\right] f(m)dm$$

which can further be simplified to

$$\text{(10)} \quad \int_{\underline{p}}^{\bar{p}} (1-m)\left[u\left(\bar{e} - \overline{\tau}^\infty(m)\right) - \max\left\{0, 1 - \frac{\bar{\pi}}{m}\right\} u\left(\underline{e} - \underline{\tau}_0^\infty(m)\right)\right] f(m)dm = 0.$$

Further, these functions are continuous everywhere and continuously differentiable everywhere but point $\overline{\pi}$ where they have finite left and right derivatives.

Proof. See appendix. ■

Proposition 2. *Let $(m_N, N)_{N=1}^\infty$ be a sequence converging to (m, ∞). Then* $\overline{\tau}^*(m_N;N) \to$

[6] The optimal plan satisfies the first-order conditions, since we assumed it to be interior.

$\overline{\tau}^{\infty}(m)$, $\underline{\tau}_1^*(m_N) \to \underline{\tau}_1^{\infty}(m)$, $\underline{\tau}_0^*(m_N; N) \to \underline{\tau}_0^{\infty}(m)$, *and* $G^*(m_N; N) \to G^{\infty}(m)$.

Proof. See appendix. ■

Recall, for a given N, that $\Delta(m)$ denoted the difference in expected utility between truthfully announcing a high endowment and falsely claiming a low endowment, conditional on a fraction m of households realizing a low endowment. Let $\Delta^*(m; N)$ denote this difference given the optimal tax/spending mechanism and let $\Delta^{\infty}(m)$ denote the limit of $\Delta^*(m; N)$ as N goes to infinity.

Let $\Delta^{\infty}(m) = u(\overline{e} - \overline{\tau}^{\infty}(m)) - (1 - \pi(m))u(\overline{e} - \underline{\tau}_0^{\infty}(m))$. This represents, for the limiting allocation, the difference in expected utility between truthfully announcing a high endowment and falsely claiming a low endowment, conditional on a fraction m of households realizing a low endowment.

Proposition 3. *There exists* $\hat{m} \in [\max\{\underline{p}, \overline{\pi}\}, \overline{p}]$ *such that for all* $m \in (\hat{m}, \overline{p}]$, $\Delta^{\infty}(m) < 0$. *If* $v(G) \equiv 0$ *(so* $G \equiv 0$*), it is also true that* $\Delta^{\infty}(m) < 0$ *for all* $m \in (\hat{m}, 1]$.

Proof. See appendix ■

Proposition 3. relies on two economic forces in the model both of which point in the direction of providing more incentives to lie in high m states.

1. that $\pi(m)$ is decreasing implies that as more households report a low endowment, the probability of a lying household being caught decreases.
2. the optimal plan provides more insurance in poorer (high m) aggregate states.

The first effect is not crucial for our analysis. Our proof relies only on the fact that $\pi'(m) \leq 0$. The optimal plan provides more insurance in poorer (high m) aggregate states, for two reasons. First, decreasing absolute risk aversion implies this directly. Second, the efficient provision of incentives requires the optimal plan to be relatively less generous to low endowment households in low m (rich) states. This occurs because in low m states, unhappy low endowment households do not affect the objective function much (since there are not many of them) but loosen the incentive constraint greatly since the many high endowment households are all deterred from emulating them. More insurance implies higher taxes on high endowment households (relative to low endowment households) and thus greater incentives to lie. Propositions 2. and 3. then imply our main result.

Theorem 1. *If either $\overline{p} = 1$ or $v(G) \equiv 0$, there exist $\overline{N}$ such that for all $N \geq \overline{N}$, the optimal direct mechanism $(\overline{\tau}^*, \underline{\tau}_1^*, \underline{\tau}_0^*, G^*)(m)$ admits an equilibrium where all households report the low endowment.*

Proof. From propositions 2. and 3., $\lim_{N\to\infty} \Delta^*\left(\frac{N}{N+1}; N\right) = \Delta^\infty(1) < 0$. If a high endowment household believes all other households will announce low (regardless of their actual endowment), then $\Delta^*\left(\frac{N}{N+1}; N\right)$ represents its incentive to tell the truth. ■

Given this result, how should a policymaker concerned with bad equilibria proceed? In the following sections we offer three alternatives: 1) a non-direct mechanism which (almost) delivers the best equilibrium as the unique equilibrium, 2) implementation using dominant strategies, and 3), a saddle-approach of maximizing the value of the worst equilibrium.[7]

[7]Ennis and Keister [3, 4] suggest yet another possibility: selecting among the multiple equilibria using some other refinement, such as risk dominance. In line with the main message of our paper, their approach would also imply that the optimal plan is in general affected by the potential for multiple perfect Bayesian equilibria, since the plan itself would affect which equilibrium is selected.

4. General Mechanisms

In the previous mechanisms, we only considered mechanisms in which households report only their endowment. As is well known, strict implementation is much easier with more-general mechanisms. While their own endowment is the only piece of information that households know and that is not common knowledge, households also form beliefs about every other household's reporting strategy. Since these beliefs must be correct in equilibrium, the government can exploit them to design more sophisticated mechanisms that rule out undesirable equilibria. We now describe such a mechanism; this mechanism has the best outcome as a unique pure strategy equilibrium, and its mixed-strategy equilibria can be made arbitrarily close to the best outcome.

The new mechanism requires households to report two messages to the government: its own endowment and, if its endowment is low, a "flag." A household's transfer is contingent on its own report, as well as on the aggregate fractions of people that report low endowment and that flag the outcome. We will denote the taxes paid by high endowment households as $\overline{\tau}_{GM}(m_e, m_f)$, where m_e is the fraction of households with the low endowment and m_f is the fraction of households flagging. Likewise, define $\underline{\tau}_{1,GM}(m_e, m_f, f)$ and $\underline{\tau}_{0,GM}(m_e, m_f, f)$ as the taxes paid by the low endowment types (audited or not audited) as functions of m_e, m_f, and whether the household itself flagged. The mechanism we construct uses this final argument only on households which are audited. Similarly, government spending depends on the aggregate fractions: $G_{GM}(m_e, m_f)$.

Our main result of this section is then,

THEOREM 2. *Assume* $\lim_{c\to\infty} u(c) = \infty$. *Then there exists* $\overline{N}$ *such that if* $N \geq \overline{N}$, *then for* *all $\epsilon > 0$, there exists a mechanism such that all equilibria given this mechanism have value within ϵ of the optimum.*

Proof. See appendix. ■

The general idea behind our mechanism is that taxes and spending depend on the fraction of households which flag in such a way as to 1) create incentive to flag when there is more than a small probability that other households will lie, and 2) create an incentive to tell the truth when there is more than a small probability that other households will flag. This expanded mechanism eliminates the equilibrium outcome of all high endowment households claiming low endowments since this belief will cause all low endowment households to flag, and thus trigger a tax/spending policy where telling the truth is a dominant strategy. The proof also handles mixed strategy equilibria and shows that while these cannot necessarily be eliminated, they can be made arbitrarily close to the best equilibrium.

This mechanism works as follows: If at least one household flags, then the taxes and transfers to non-flagging households are independent of their endowment report. Nevertheless, such households still face a positive probability of being audited if reporting low. Thus there is no gain and a possible punishment from falsely claiming a low endowment. Thus high endowment households should tell the truth if they expect another household to flag. If no households flag, then the taxes, transfers and spending emulate the optimal direct mechanism. Finally, the mechanism offers each low endowment household a lottery with a negative expected utility payoff if other households are expected to tell the truth, and a positive expected utility payoff if other households are expected to lie with more than a small probability. The "flag" determines whether the household purchases this lottery or not. In essence, the government uses the lottery to elicit household expectations regarding the strategy of other households.

5. Dominant Strategy Implementation

Wilson [13] argues for dominant strategy implementation for robustness reasons. In particular, in our environment, the direct mechanism and the general mechanism both depend on households having precise knowledge of $f(p)$. An added benefit is such an approach eliminates the bad equilibrium in our environment and is particularly easy to characterize. In particular, one simply replaces the original incentive constraint

$$(11) \quad \int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{N} Q(n|p, N)\Delta\left(\frac{n}{N+1}\right)(1-p)f(p)dp \geq 0,$$

with a requirement $\Delta(m) \geq 0$ for all m. That is, dominant strategy implementation requires that the incentive to tell the truth must no longer be on average non-negative, but always non-negative. The ability to do this follows directly from the fact that the revelation principle applies within dominant strategy mechanisms.

Proposition 4. *For all $\epsilon > 0$, there exists a dominant strategy mechanism with a unique equilibrium whose payoff is within ϵ of the solution to maximizing (4) subject to (1) and $\Delta(m) \geq 0$ for all m.*

Proof. If $\Delta(m) \geq 0$ is replaced with $\Delta(m) \geq \delta$ for any $\delta > 0$, then every household strictly prefers to tell the truth, regardless of the strategy of other households. As $\delta \to 0$, the solution convergences in value to the solution when $\delta = 0$. ■

Recall that the optimal direct mechanism displayed multiple equilibria because it offered relatively more insurance in poorer aggregate states. In essence, the optimal dominant strategy implementation eliminates this dependence of the level of insurance on the aggregate state. This has the benefit of an unique equilibrium, but at the cost of a tighter constraint set and thus a lower social value.

6. Minimax Direct Mechanisms

The revelation principle allows the restriction to direct mechanisms either when looking for a mechanism which delivers the best perfect Bayesian equilibrium (our focus in section three) or the best dominant strategy equilibrium (our focus in section five). Direct mechanisms may also be interesting in their own right for the simplicity of the reports they require. Given this, one may wish to arbitrarily restrict attention to direct mechanisms.

While the dominant strategy implementation described in the previous section uses a direct mechanism, dominant strategy mechanisms (direct or not) impose a very tight constraint set on a policymaker concerned with bad equilibria. Even if a mechanism has multiple equilibria, it is possible that all of them have a value better than the best dominant strategy outcome. An interesting question is the direct mechanism which delivers the best worst case scenario, or the best worst perfect Bayesian equilibrium. Such a mechanism solves

$$\begin{aligned}
\max_{(\overline{\tau},\underline{\tau}_1,\underline{\tau}_0,G)(m)} \min_{x\in[0,1]} \int_{\underline{p}}^{\overline{p}} \sum_{n=0}^{n} Q(n|p+(1-p)x,N)\Big[(1-p)(1-x)\ \Big[u\left(\overline{e}-\overline{\tau}\left(\frac{n}{N+1}\right)\right)+ \\
v\Big(G\Big(\frac{n}{N+1}\Big)\Big)\Big]+(1-p)x\Big[\left(1-\pi\left(\frac{n+1}{N+1}\right)\right)u\left(\overline{e}-\underline{\tau}_0\left(\frac{n+1}{N+1}\right)\right)+ \\
v\Big(G\Big(\frac{n+1}{N+1}\Big)\Big)\Big]+ \\
p\Big[\pi\left(\frac{n+1}{N+1}\right)u\left(\underline{e}-\underline{\tau}_1\left(\frac{n+1}{N+1}\right)\right)+\left(1-\pi\left(\frac{n+1}{N+1}\right)\right)u\left(\underline{e}-\underline{\tau}_0\left(\frac{n+1}{N+1}\right)\right)+ \\
v\Big(G\Big(\frac{n+1}{N+1}\Big)\Big)\Big]\Big]f(p)dp
\end{aligned} \tag{12}$$

subject to the resource constraint (1) and the incentive constraints

$$\begin{cases} U_{truth}(x) \leq U_{lie}(x), & \text{if } x > 0 \\ U_{truth}(x) \geq U_{lie}(x), & \text{if } x < 1, \end{cases} \tag{13}$$

where

$$U_{truth}(x) = \int_{\underline{p}}^{\overline{p}} \sum_{n=0}^{n} Q(n|p+(1-p)x,N)\Big[u\left(\overline{e}-\overline{\tau}\left(\frac{n}{N+1}\right)\right)+v\Big(G\Big(\frac{n}{N+1}\Big)\Big)\Big], \tag{14}$$

and

$$\begin{aligned}
U_{lie}(x) = \int_{\underline{p}}^{\overline{p}} \sum_{n=0}^{n} Q(n|p+(1-p)x,N)\Big[\left(1-\pi\left(\frac{n+1}{N+1}\right)\right)u\left(\overline{e}-\underline{\tau}_0\left(\frac{n+1}{N+1}\right)\right)+ \\
v\Big(G\Big(\frac{n+1}{N+1}\Big)\Big)\Big].
\end{aligned} \tag{15}$$

We argue two characteristics regarding solutions to this problem. First, if $f(p)$ is degenerate (or p is common knowledge) that as $N \to \infty$, the solution to this minmax problem approaches (in value) the best equilibrium outcome. Second, we conjecture if $\overline{p} = 1$, that as $N \to \infty$, the solution to this minmax problem does not approach the best equilibrium outcome.

Proposition 5. *Assume $\underline{p} = \overline{p}$. Then for all $\epsilon > 0$, there exists an $\overline{N}$ such that if $N \geq \overline{N}$, there exists a direct mechanism with a worst equilibrium within ϵ (in value) of the best equilibrium outcome.*

Proof.

As before, let m denote the fraction of households claiming a low endowment. Consider the tax/spending policy

$$(16)\quad (\overline{\tau}, \underline{\tau}_1, \underline{\tau}_0, G)(m) = \begin{cases} (0,0,0,0) & \text{if } m > \overline{p} + \delta \\ (\overline{\tau}^*, \underline{\tau}_1^*, \underline{\tau}_0^*, G^*)(m) & \text{if } m \leq \overline{p} + \delta, \end{cases}$$

for arbitrary $\delta > 0$. Let x be the probability a high endowment household lies. As $N \to \infty$, the probability that $m > \overline{p} + \delta$ goes to unity if $x > \delta/(1-\overline{p})$ and goes to zero if $x < \delta/(1-\overline{p})$. If $x > \delta/(1-\overline{p})$, there exists an $\overline{N}$ such that for $N \geq \overline{N}$, the best reply to this policy is truthtelling, thus ruling out any equilibrium with $x > \delta/(1-\overline{p})$. By choosing a sequence $\delta_i \to 0$, one can choose a sequence $\overline{N}_i$ to rule out equilibria with $x > \delta_i/(1-\overline{p})$. Boundedness in payoffs then implies the result. ■

Proposition 6. (Conjecture) *Assume $\overline{p} = 1$. There exists an $\epsilon > 0$ and an $\overline{N}$ such that if $N \geq \overline{N}$ the worst equilibrium of all direct mechanisms is not within ϵ (in value) of the best equilibrium.*

Proof. To be completed. (We hope). ■.

7. Conclusion

In this paper we show that the optimal direct mechanism provides a relatively large amount of insurance in poor aggregate states, and very little insurance in rich aggregate states. We show that this implies a second equilibrium where all households claim to be poor. An obvious remedy to the uncertainty presented by the existence of this bad equilibrium is to shift social insurance away from poor aggregate states, as does our dominant strategy implementation. Overall, the forces present in our environment delivering these results appear to be general and thus the consideration of bad equilibria in public finance environments warrants further attention.

Appendix

A1. Proof of Proposition 1.

Proof. Take m and λ as given. For a given μ, equations (5), (6), and (8) have at most one solution for $\overline{\tau}$, $\underline{\tau}_1$, and G respectively. Further, as μ increases, the $\overline{\tau}$ and $\underline{\tau}_1$ which solve (5) and (6) must increase, and G which solves (8) must decrease.[8]

Equation (7) and $\mu(m) > 0$ implies

$$\text{(A1)}\quad \frac{u'(\underline{e}-\underline{\tau}_0)}{u'(\overline{e}-\underline{\tau}_0)} > \lambda \frac{\int_{\underline{p}}^{\bar{p}}[p^m(1-p)^{1-m}]^{N+1}f(p)dp}{\int_{\underline{p}}^{\bar{p}}[p^m(1-p)^{1-m}]^{N+1}f(p)dp}.$$

That u displays nonincreasing absolute risk aversion (DARA) implies

$$\text{(A2)}\quad \frac{u''(\underline{e}-\underline{\tau}_0)}{u''(\overline{e}-\underline{\tau}_0)} \geq \frac{u'(\underline{e}-\underline{\tau}_0)}{u'(\overline{e}-\underline{\tau}_0)}.$$

Equations (A1) and (A2) imply

$$\frac{u''(\underline{e}-\underline{\tau}_0)}{u''(\overline{e}-\underline{\tau}_0)} > \lambda \frac{\int_{\underline{p}}^{\bar{p}}[p^m(1-p)^{1-m}]^{N+1}f(p)dp}{\int_{\underline{p}}^{\bar{p}}[p^m(1-p)^{1-m}]^{N+1}f(p)dp}$$

or

$$\text{(A3)}\quad -u''(\underline{e}-\underline{\tau}_0) + \lambda \frac{\int_{\underline{p}}^{\bar{p}}[p^m(1-p)^{1-m}]^{N+1}f(p)dp}{\int_{\underline{p}}^{\bar{p}}[p^m(1-p)^{1-m}]^{N+1}f(p)dp} u''(\overline{e}-\underline{\tau}_0) > 0.$$

[8] In (8), the coefficient multiplying v' must be positive for the system to have a solution.

This last equation is the derivative of (7) with respect to $\underline{\tau}_0$, thus the left hand side of (7) is an increasing function of $\underline{\tau}_0$ around the solution, and thus has at most one solution. Note further that this implies as μ increases, $\underline{\tau}_0$ must increase as well for equation (7) to continue to hold. Thus $\overline{\tau}$, $\underline{\tau}_1$, and $\underline{\tau}_0$ are all increasing functions of μ, and G is a decreasing function of μ. It follows that there exists at most one value of μ, for each m and λ, such that the resource constraint holds as well. ■

A2. Proof of lemma 1.

This proof of this lemma itself requires several lemmas.

LEMMA 2. $\limsup_{N\to\infty} \lambda^*(N) < \infty$.

Proof. Suppose by contradiction that there exists a subsequence $\{N_t\}_{t=1}^{\infty}$ such that $\lambda^*(N_t) \to_{t\to\infty} \infty$. Let $\epsilon > 0$. We consider the solution of the system the system (9), (5), (6), (7), and (8) on $m \in [0, 1-\epsilon]$. The resource constraint implies that $u'(\bar{e} - \bar{\tau}^*(m; N))$ must remain uniformly bounded away from 0 on this interval. (5) then implies that $\mu^*(m; N)$ must diverge uniformly. Substituting (5) into (8), we obtain

$$\frac{u'(\bar{e} - \bar{\tau}^*(m; N))}{v'(G^*(m; N))} = 1 - \frac{\lambda^*(N) \left(\frac{\int_{\underline{p}}^{\bar{p}} \left(\frac{m}{p}\right) [p^m (1-p)^{1-m}]^{N+1} f(p) dp}{\int_{\underline{p}}^{\bar{p}} [p^m (1-p)^{1-m}]^{N+1} f(p) dp} \right)}{1 + \lambda^*(N)}$$

As $N \to \infty$ and $\lambda^*(N) \to \infty$, the right-hand-side of this expression is a continuous function converging pointwise to 0 on $[\underline{p}, \bar{p}]$, to $1 - \frac{m}{\underline{p}}$ on $[0, \underline{p})$, and to $1 - \frac{m}{\bar{p}} < 0$ on $(\bar{p}, 1]$. This implies a contradiction, even if $\bar{p} = 1$, since $\lim_{G\to 0} v'(G) < +\infty$.[9] ■

[9] This could just mean that in the limit the solution might be at a corner, which we ruled out by assumption. However, a more complete proof by contradiction assumption can be developed, in which the incentive constraint is shown to become slack if $\lambda^*(N)$ diverged.

Lemma 3. $\mu^*(m; N)$ is uniformly bounded as $N \to +\infty$.

Proof. By contradiction, suppose $\exists (m_t, N_t)_{t=0}^{\infty} : \lim_{t \to +\infty} \mu^*(m_t, N_t) \to \infty$. Since $\lambda^*(N_t)$ is a bounded sequence and $\lim_{G \to 0} v'(G) < +\infty$, (8) will not have a solution.[10] ■

Lemma 4. If $\lim_{c \to 0} u'(c) = +\infty$, then there exists a value ϵ_c such that consumption of all households (that report truthfully) in the optimal plan is above ϵ_c independently of N.

Proof. Since we know from the preceding lemmas that the Lagrange multipliers are uniformly bounded, the first-order conditions imply that the marginal utilities must also be uniformly bounded. The conditions in the lemma are necessary for this to be the case. ■

We will use the bound ϵ_c below. If $\lim_{c \to 0} u'(c) < +\infty$, then $\epsilon_c = 0$.

Lemma 5. There exist values E_c $[E_G]$ such that consumption of all households [government spending] in the optimal plan is below E_c $[E_G]$, independently of N.

Proof. The resource constraint implies $G \leq \bar{e}$, so we can set $E_G = \bar{e}$. For private consumption, suppose by contradiction there is a sequence $(m_t, N_t)_{t=0}^{\infty}$ such that, along the sequence, a selection from the taxes $(\bar{\tau}^*, \underline{\tau}_1^*, \underline{\tau}_0^*)$ diverges to $-\infty$. Notice $\underline{\tau}_0^*(m; N) > \underline{\tau}_1^*(m; N)$ for all values of N and all values of m for which $\underline{\tau}_0^*(m; N)$ is defined. Furthermore, $\underline{\tau}_1^*(m; N) > \bar{\tau}^*(m; N) + \underline{e} - \bar{e}$, except for $m = 1$, when $\bar{\tau}^*(1; N)$ is not defined.

From the sequence (m_t, N_t), we select a subsequence in which $m_t = 1$. We distinguish two cases, at least one of which must be true.

1. We obtain a new subsequence $(1, \hat{N}_s)_{s=0}^{\infty}$ in which the selection is still divergent. In this

[10] Again, a proof by contradiction with the resource constraint becoming slack could be constructed when corners are an issue.

case, it must be the case that $\underline{\tau}_1^*(1, \hat{N}_s) \to -\infty$, but this eventually violates the resource constraint.

2. After taking all occurrences of $m_t = 1$ out, the remaining subsequence $(\tilde{m}_t, \tilde{N}_t)_{t=0}^{\infty}$ admits a divergent selection. Given the relationship among taxes stated above, it must be the case that $\lim_{t\to\infty} \bar{\tau}^*(\tilde{m}_t, \tilde{N}_t) = -\infty$. Since $\lambda^*(N)$ is bounded, it follows that $\lim_{t\to\infty} \mu^*(\tilde{m}_t, \tilde{N}_t) = 0$, which implies $\lim_{t\to\infty} \underline{\tau}_1^*(\tilde{m}_t, \tilde{N}_t) = -\infty$. We must also have $\lim_{t\to\infty} N_t = \infty$, since in any problem with finite N taxes are a finitely dimensional vector and the resource constraint imposes bounds at all values for which they are defined. It follows that, in the limit, there is at least a fraction $\bar{\pi}$ of people that "pay" either $\bar{\tau}^*(\tilde{m}_t, \tilde{N}_t)$ or $\underline{\tau}_1^*(\tilde{m}_t, \tilde{N}_t) = -\infty$. This would again violate the resource constraint eventually, leading to a contradiction.

■

We now study the left-hand sides of the system of first-order conditions (5)-(8) and of the resource constraint (9) as functions of $\bar{\tau}, \underline{\tau}_0, \underline{\tau}_1, G, \mu, \lambda, m, N$. We will consider the following domain: $\bar{\tau} \in [\bar{e} - E_c, \bar{e} - \epsilon_c]$, $\underline{\tau}_i \in [\underline{e} - E_c, \underline{e} - \epsilon_c]$, $i = 0, 1$, $G \in [0, \bar{e}]$, $\mu \in [0, \bar{\mu}]$, $\lambda \in [0, \bar{\lambda}]$, $m \in [0, 1]$, $N = 1, 2, \ldots, +\infty$. When $N = +\infty$, we take pointwise limits, so the expressions for the left-hand side of (7), (8) and (9) converge to the right-hand side of

$$\text{(A4)} \quad 0 = \begin{cases} u'\left(\underline{e} - \underline{\tau}_0\right) - \lambda u'\left(\bar{e} - \underline{\tau}_0\right)\left(\frac{1}{m} - 1\right) - \mu & \text{if } m \in [\underline{p}, \bar{p}] \\ u'\left(\underline{e} - \underline{\tau}_0\right) - \lambda u'\left(\bar{e} - \underline{\tau}_0\right)\left(\frac{1}{\underline{p}} - 1\right) - \mu & \text{if } m \in [0, \underline{p}] \\ u'\left(\underline{e} - \underline{\tau}_0\right) - \lambda u'\left(\bar{e} - \underline{\tau}_0\right)\left(\frac{1}{\bar{p}} - 1\right) - \mu & \text{if } m \in [\bar{p}, 1] \end{cases}$$

$$
\text{(A5)}\quad 0 = \begin{cases} v'(G) - \mu & \text{if } m \in [\underline{p}, \bar{p}] \\ \left[1 + \lambda\left(1 - \frac{m}{\underline{p}}\right)\right] v'(G) - \mu & \text{if } m \in [0, \underline{p}] \\ \left[1 + \lambda\left(1 - \frac{m}{\bar{p}}\right)\right] v'(G) - \mu & \text{if } m \in [\bar{p}, 1] \end{cases}
$$

$$
\text{(A6)}\quad 0 = \min\{m, \bar{\pi}\}\,\underline{\tau}_1 + \max\{0, m - \bar{\pi}\}\,\underline{\tau}_0 + (1-m)\,\overline{\tau} - G
$$

The domain specified above is compact if the set $\{1, \ldots, +\infty\}$ is endowed with the metric $d(N_1, N_2) \equiv |1/N_1 - 1/N_2|$. The left-hand sides of (5)-(8) and (9), with the extensions (A4), (A5) and (A6), are continuous functions of $\bar{\tau}, \underline{\tau}_0, \underline{\tau}_1, G, \mu, \lambda, m, N$ over the domain and with the metric defined above.[11] By compactness, they also are uniformly continuous.

We are now ready to prove lemma 1..

Proof.

1. Existence:

   By lemma 2., we can find a sequence $\{N_t\}_{t=1}^{\infty}$, with $N_t \to_{t\to\infty} \infty$, such that $\lambda^*(N_t)$ converges, to a value that we define $\lambda^*(\infty)$ (we will later prove uniqueness of this value, justifying labelling it as such). Given any $m \in [0,1]$, we can find a sequence $\{m_{N_t}\}_{t=1}^{\infty}$ such that $m_{N_t} = i_{N_t}/(N_t + 1)$ for some sequence of integers $\{i_{N_t}\}_{t=1}^{\infty}$, and $m_{N_t} \to_{t\to+\infty} m$. Since we assumed the solution to be interior for all values of N, $\bar{\tau}^*(m_{N_t}; N_t)$, $\underline{\tau}_1^*(m_{N_t}; N_t)$, $\underline{\tau}_0^*(m_{N_t}; N_t)$, $G^*(m_{N_t}; N_t)$, $\mu^*(m_{N_t}; N_t)$ satisfy the first-order conditions (5)-(8), and the resource constraint (9). Since the tax-spending policy and the multipliers are uniformly bounded, there exist a convergent subsequence of these

[11] The only step worth mentioning is that, if $(m_N, N)_{N=1}^{\infty}$ is a sequence converging to (m, ∞), a distribution with density $\frac{[p_N^m + (1-p)^{1-m_N}]^{N+1} f(p)}{\int_{\underline{p}}^{\bar{p}} [p_N^m + (1-p)^{1-m_N}]^{N+1} f(p) dp}$ on $[\underline{p}, \bar{p}]$ converges weakly to the degenerate distribution that puts mass 1 on m if $m \in [\underline{p}, \bar{p}]$, and on the closer bound otherwise.

values. Let $\bar{\tau}^*(m;\infty)$, $\underline{\tau}_1^*(m;\infty)$, $\underline{\tau}_0^*(m;\infty)$, $G^*(m;\infty)$, $\mu^*(m;\infty)$ be the limit. By uniform continuity of the first-order conditions and the resource constraint, these limits satisfy equations (5), (6), (A4), (A5) and (A6). Repeating the steps of proposition 1., we can prove that, given $\lambda^*(\infty)$ and m, there exists at most one value for $\bar{\tau}$, $\underline{\tau}_1$, $\underline{\tau}_0$, G and μ that satisfies these equations; hence, all subsequences will have to converge to the same value, defined above. Since the system is uniformly continuous including m as a variable as well, the limiting functions $\bar{\tau}^*(m;\infty)$, $\underline{\tau}_1^*(m;\infty)$, $\underline{\tau}_0^*(m;\infty)$, $G^*(m;\infty)$, $\mu^*(m;\infty)$ must be satisfy (5), (6), (A4), (A5) and (A6) for all values of m; by the implicit function theorem, this implies that they must be continuously differentiable, except at $m = \bar{\pi}$, where they are only continuous and have finite left and right derivative.

Let F_N be the c.d.f. of the fraction of N households receiving low income, conditional on the $N+1$st household having high income. As $N \to \infty$, this distribution converges weakly to one with p.d.f. proportional to $(1-m)f(m)$, where m is the fraction of people receiving low income. Next, extend $\bar{\tau}^*(m,N)$ to $[0,1]$ by defining

$$\bar{\tau}^{**}(m;N) \equiv \bar{\tau}^*\left(\frac{i(m,N)}{N+1};N\right),$$

where $i(m,N)$ is the smallest integer such that $\frac{i(m,N)}{N+1} \geq m$ if $m < 1$, and is equal to N if $m = 1$. We can extend $\underline{\tau}_0^*$ to $[\bar{\pi},1]$ and G^* to $[0,1]$. The functions $\bar{\tau}^{**}(m;N)$, $\underline{\tau}_0^{**}(m;N)$ and $G^{**}(m;N)$ coincide with $\bar{\tau}^*(m;N)$, $\underline{\tau}_0^*(m;N)$ and $G^*(m;N)$ on all points of positive probability mass in equation (3), and they converge uniformly to the continuous functions $\bar{\tau}^*(m;\infty)$, $\underline{\tau}_0^*(m;\infty)$ and $G^*(m;\infty)$. Since (3) holds for all values of N, it then follows that $\lambda^*(\infty)$, $\bar{\tau}^*(m;\infty)$, $\underline{\tau}_0^*(m;\infty)$ satisfy (10) as well.

2. Uniqueness:

   As already mentioned above, the same steps of proposition 1. imply that, given λ, there exist at most one set of continuous functions $\bar{\tau}^*(m;\infty)$, $\underline{\tau}_1^*(m;\infty)$, $\underline{\tau}_0^*(m;\infty)$, $G^*(m;\infty)$, $\mu^*(m;\infty)$ that satisfy (5), (6), (A4), (A5), and (A6). We now need to prove that there is a unique value of λ which, in combination with these functions, satisfies (10) as well. We do so by proving that the left-hand side of (10) is locally increasing in λ at all solutions, when the functions $\bar{\tau}(m;\infty)$, $\underline{\tau}_1(m;\infty)$, $\underline{\tau}_0(m;\infty)$, $G(m;\infty)$, $\mu(m;\infty)$ are adjusted to satisfy (5), (6), (A4), (A5), and (A6). Starting from a solution, suppose we increase λ by a sufficiently small increment $\Delta\lambda > 0$. Consider each value of $m \in [\underline{p}, \bar{p}]$ independently. Holding μ fixed, a change in λ implies that $\bar{\tau}$ will have to increase to maintain (5), and that (if $m > \bar{\pi}$) $\underline{\tau}_0$ will have to decrease to maintain (A4).[12] In general, the resulting change will no longer satisfy the resource constraint (A6); therefore, the change in λ will require a change in μ, which will affect $\underline{\tau}_1$ and G as well. However, since (locally) all taxes are increasing in μ and G is decreasing in μ, to satisfy (A6), the final outcome will necessarily involve a higher value for $\bar{\tau}$ and (if $m > \bar{\pi}$) a lower value for $\underline{\tau}_0$. Since this is true for all values of m, this proves that the left-hand side of (A6) is increasing in λ.

■

A3. Proof of proposition 2.

Proof. Suppose the stated convergence fails. Then, repeating the steps of lemma 1., we can find a subsequence $\{N_t\}_{t=1}^{\infty}$ such that the functions $\bar{\tau}^*(\cdot;N_t)$, $\underline{\tau}_1^*(\cdot;N_t)$, $\underline{\tau}_0^*(\cdot;N_t)$, $G^*(\cdot;N_t)$,

[12] The proof assumes that the limiting taxes and spending are interior, but it can be easily adapted to the case in which the limit hits some corners.

$\mu^*(\cdot; N_t)$, and $\lambda^*(N_t)$ converge to values that are different from $\bar{\tau}^*(\cdot;\infty)$, $\underline{\tau}_1^*(\cdot;\infty)$, $\underline{\tau}_0^*(\cdot;\infty)$, $G^*(\cdot;\infty)$, $\mu^*(\cdot;\infty)$, and $\lambda^*(\infty)$. By the same proof as before, the new limiting point will satisfy (5), (6), (A4), (A5), (A6), and (10). But this contradicts the uniqueness of the solution to this system of equations, which we established above. ■

A4. Proof of proposition 3.

The proof requires some lemmas.

LEMMA 6. If $\Delta^\infty(m) < 0$, then $\overline{\tau}^\infty(m) > \underline{\tau}_0^\infty(m)$.

Proof. Suppose $\overline{\tau}^\infty(m) \leq \underline{\tau}_0^\infty(m)$. Then

$$\Delta^\infty(m) = u(\overline{e} - \overline{\tau}^\infty(m)) - (1 - \pi(m))u(\overline{e} - \underline{\tau}_0^\infty(m)) \geq$$

$$u(\overline{e} - \overline{\tau}^\infty(m)) - (1 - \pi(m))u(\overline{e} - \overline{\tau}^\infty(m)) = \pi(m)u(\overline{e} - \overline{\tau}^\infty(m)) \geq 0,$$

a contradiction. ■

LEMMA 7. If $\Delta^\infty(m) < 0$ and $m \in [\max\{\underline{p}, \bar{\pi}\}, \bar{p}]$, then $\mu'^\infty(m) > 0$.[13] Furthermore, if $v(G) \equiv 0$ (so $G \equiv 0$), then $\mu'^\infty(m) > 0$ also on $[\bar{p}, 1]$.

Proof. For all m and μ, let $x(\mu, m)$ denote the value of $\underline{\tau}_0$ which solves (A4), and let $c(\mu, m) = \underline{e} - x(\mu, m)$. The partial derivative of (A4) with respect to $\underline{\tau}_0$ is strictly positive from equation (A3), while its partial derivative with respect to μ is strictly negative. Thus $x_\mu(\mu, m) > 0$. Likewise, the partial derivative of (A4) with respect to m is strictly positive, thus $x_m(\mu, m) < 0$. These imply $c_\mu(\mu, m) > 0$ and $c_m(\mu, m) > 0$.

[13] That $\bar{\pi} < \bar{p}$ follows from the fact that otherwise the incentive constraint could not be binding for sufficiently large N.

Using equations (5), (6), and (A5), the resource constraint for $m \geq \overline{\pi}$ can be written

(A7) $m\underline{e} + (1-m)\overline{e} - (m-\overline{\pi})c(\mu, m) - \overline{\pi}u'^{-1}(\mu) - (1-m)u'^{-1}(\mu/(1-\lambda)) - v'^{-1}(\mu) = 0.$

Let $z(\mu, m)$ denote the left hand side (A7). The partials of z with respect to m and μ are

(A8) $z_m(\mu, m) = \underline{e} - \overline{e} + [u'^{-1}(\mu/(1+\lambda)) - c(\mu, m)] - (m-\overline{\pi})c_m(\mu, m)$

and

(A9) $z_\mu(\mu, m) = -(m-\overline{\pi})c_\mu(\mu, m) - \dfrac{\overline{\pi}}{u''(u'^{-1}(\mu))} - \dfrac{1-m}{u''(u'^{-1}(\mu/(1+\lambda)))} - \dfrac{1}{v''(v'^{-1}(\mu))}.$

The partial $z_\mu(\mu^*(m), m) > 0$. Thus $\mu'^*(m) > 0$ if $z_m(\mu^*(m), m) \leq 0$. From (5), $\overline{e} - u'^{-1}(\mu^*(m)/(1+\lambda)) = \overline{\tau}^*(m)$, and from the definition of $c(\mu, m)$, $\underline{e} - c(\mu^*(m), m) = \underline{\tau}_0^*(m)$. Thus

(A10) $z_m(\mu^*(m), m) = \underline{\tau}_0^*(m) - \overline{\tau}^*(m) - (m-\overline{\pi})c_m(\mu^*(m), m) \leq \underline{\tau}_0^*(m) - \overline{\tau}^*(m),$

from $c_m(\mu, m) > 0$. Since $u(\overline{e} - \overline{\tau}^*(m)) - (1-\pi(m))u(\overline{e} - \underline{\tau}_0^*(m)) < 0$ implies $\overline{\tau}^*(m) > \underline{\tau}_0^*(m)$, $\mu'^*(m) > 0$. When $v(G) \equiv 0$, the proof for $m \in [\bar{p}, 1]$ is entirely analogous, except that $x_m(\mu, m) = 0$. ■

LEMMA 8. For all $m \in [\max\{\underline{p}, \overline{\pi}\}, \bar{p}]$, if $\Delta(m) \leq 0$, then $\Delta'(m) < 0$. If $v(G) \equiv 0$, the same is true on $[\bar{p}, 1]$.

Proof. For $m \geq \overline{\pi}$,

$$\begin{aligned}\Delta'(m) &= -u'(\overline{e}-\overline{\tau}^*(m))\overline{\tau}'^*(m) + (1-\frac{\overline{\pi}}{m})u'(\overline{e}-\underline{\tau}_0^*(m))\underline{\tau}_0'^*(m) - \frac{\overline{\pi}}{m^2}u(\overline{e}-\underline{\tau}_0^*(m)) \\ &\leq -u'(\overline{e}-\overline{\tau}^*(m))\overline{\tau}'^*(m) + (1-\frac{\overline{\pi}}{m})u'(\overline{e}-\underline{\tau}_0^*(m))\underline{\tau}_0'^*(m).\end{aligned}$$

Differentiating (5) with respect to $\overline{\tau}$ delivers

$$\overline{\tau}'^*(m) = \frac{-\mu'^*(m)}{(1+\lambda)u''(\overline{e}-\overline{\tau}^*(m))}.$$

Likewise differentiating (A4) with respect to $\underline{\tau}_0$ delivers

$$\underline{\tau}_0'^*(m) = \frac{-\mu'^*(m) + \frac{1}{m^2}\lambda u'(\overline{e}-\underline{\tau}_0^*(m))}{u''(\underline{e}-\underline{\tau}_0^*) - \lambda\frac{1-m}{m}u''(\overline{e}-\underline{\tau}_0^*(m))}.$$

These imply

$$\begin{aligned}\Delta'(m) &\leq u'(\overline{e}-\overline{\tau}^*(m))\frac{\mu'^*(m)}{(1+\lambda)u''(\overline{e}-\overline{\tau}^*(m))} + \\ &\quad (1-\frac{\overline{\pi}}{m})u'(\overline{e}-\underline{\tau}_0^*(m))\frac{-\mu'^*(m) + \frac{1}{m^2}\lambda u'(\overline{e}-\underline{\tau}_0^*(m))}{u''(\underline{e}-\underline{\tau}_0^*(m)) - \lambda\frac{1-m}{m}u''(\overline{e}-\underline{\tau}_0^*(m))} \\ &< u'(\overline{e}-\overline{\tau}^*(m))\frac{\mu'^*(m)}{(1+\lambda)u''(\overline{e}-\overline{\tau}^*(m))} - \\ &\quad (1-\frac{\overline{\pi}}{m})u'(\overline{e}-\underline{\tau}_0^*(m))\frac{\mu'^*(m)}{u''(\underline{e}-\underline{\tau}_0^*(m)) - \lambda\frac{1-m}{m}u''(\overline{e}-\underline{\tau}_0^*(m))},\end{aligned}$$

since the denominator of the second term is negative from equation (A3). Next,

$$\Delta'(m) < \mu'^*(m)\Big[\frac{u'(\overline{e}-\overline{\tau}^*(m))}{(1+\lambda)u''(\overline{e}-\overline{\tau}^*(m))} - \frac{u'(\overline{e}-\underline{\tau}_0^*(m))}{u''(\underline{e}-\underline{\tau}_0^*) - \lambda\frac{1-m}{m}u''(\overline{e}-\underline{\tau}_0^*(m))}\Big]$$

from $(1-\overline{\pi}/m) < 1$.

Since $\mu'^{*}(m) > 0$, $\Delta'(m) < 0$ if the expression within the square brackets is negative. Since $u'(\overline{e}-\overline{\tau}^{*}(m)) > u'(\overline{e}-\underline{\tau}_0^{*}(m))$ from $\overline{\tau}^{*}(m) > \underline{\tau}_0^{*}(m)$, it is sufficient to show that

$$\text{(A11)}\quad (1+\lambda)u''(\overline{e}-\overline{\tau}^{*}(m)) > u''(\underline{e}-\underline{\tau}_0^{*}(m)) - \lambda\frac{1-m}{m}u''(\overline{e}-\underline{\tau}_0^{*}(m)).$$

Equations (5) and (7) imply

$$u'(\underline{e}-\underline{\tau}_0^{*}(m)) = u'(\overline{e}-\overline{\tau}^{*}(m))(1+\lambda) + \lambda\frac{1-m}{m}u'(\overline{e}-\underline{\tau}_0^{*}(m)).$$

This implies there exists an $\theta \in [0,1]$ such that

$$\text{(A12)}\quad \theta u'(\underline{e}-\underline{\tau}_0^{*}(m)) = u'(\overline{e}-\overline{\tau}^{*}(m))(1+\lambda),$$

and

$$\text{(A13)}\quad (1-\theta)u'(\underline{e}-\underline{\tau}_0^{*}(m)) = \lambda\frac{1-m}{m}u'(\overline{e}-\underline{\tau}_0^{*}(m)).$$

From (5), $u'(\overline{e}-\overline{\tau}^{*}(m)) < \mu^{*}(m)$, and from (7), $u'(\underline{e}-\underline{\tau}_0^{*}(m)) > \mu^{*}(m)$. Thus $\overline{e}-\overline{\tau}^{*}(m) > \underline{e}-\underline{\tau}_0^{*}(m)$. Decreasing absolute risk aversion then implies

$$\frac{u''(\underline{e}-\underline{\tau}_0^{*}(m))}{u''(\overline{e}-\overline{\tau}^{*}(m))} > \frac{u'(\underline{e}-\underline{\tau}_0^{*}(m))}{u'(\overline{e}-\overline{\tau}^{*}(m))},$$

which combined with (A12) implies

$$\frac{u''(\underline{e}-\underline{\tau}_0^*(m))}{u''(\overline{e}-\overline{\tau}^*(m))} > \frac{1+\lambda}{\theta},$$

or

$$\text{(A14)}\quad \theta u''(\underline{e}-\underline{\tau}_0^*(m)) < (1+\lambda)u''(\overline{e}-\overline{\tau}^*(m)).$$

Decreasing absolute risk aversion also implies

$$\frac{u''(\underline{e}-\underline{\tau}_0^*(m))}{u''(\overline{e}-\underline{\tau}_0^*(m))} > \frac{u'(\underline{e}-\underline{\tau}_0^*(m))}{u'(\overline{e}-\underline{\tau}_0^*(m))},$$

which combined with (A13) and rearranging implies

$$\text{(A15)}\quad (1-\theta)u''(\underline{e}-\underline{\tau}_0^*(m)) < \lambda\frac{1-m}{m}u''(\overline{e}-\underline{\tau}_0^*(m)).$$

Adding (A14) and (A15) delivers equation (A11), proving $\Delta'(m) < 0$. ■

We are now ready to prove proposition 3..

Proof. Since we assumed that the incentive constraint is binding for all values of N (at least above a threshold $\bar{N}$), the limiting incentive constraint (10) holds as an equality. Thus,

$$\int_{m\geq\overline{\pi}} \Delta^\infty(m)(1-m)f(m)dm \leq \int_{\underline{p}}^{\overline{p}} \Delta^\infty(m)(1-m)f(m)dm = 0,$$

proving their exists $\hat{m} \in [\max\{\underline{p}, \bar{\pi}\}, \overline{\pi}]$ such that $\Delta(\hat{m}) \leq 0$. Lemma 8. then recursively implies $\Delta(m) < 0$ for all $m \in (\hat{m}, \bar{p}]$. ■

A5. Proof of Theorem 2.

Proof. First, if nobody flags, the mechanism coincides with optimal direct revelation:

$$
\begin{aligned}
\bar{\tau}_{GM}(m_e, 0) &= \bar{\tau}^*(m) \\
\underline{\tau}_{0,GM}(m_e, 0, 0) &= \underline{\tau}_0^*(m_e) \\
\underline{\tau}_{1,GM}(m_e, 0, 0) &= \underline{\tau}_1^*(m_e) \\
G_{GM}(m_e, 0) &= G^*(m_e)
\end{aligned}
$$

If any low-income household flags, the government sets $G_{GM}(m_e, m_f) = 0$.[14] The household that flags receives the same tax/transfer as in the optimal direct mechanism if it is not audited. If the household is audited, its tax/transfer is changed by adding two components: the first one exactly compensates for the loss in utility due to the failure by the government to provide for the public good, whereas the second one incorporates a bet on the aggregate distribution of reports that is advantageous when high-endowment households misreport their income with a probability greater than δ. Formally,

$$
\begin{aligned}
\underline{\tau}_{0,GM}(m_e, 1/(N+1), 1) &= \underline{\tau}_0^*(m_e) \\
\underline{\tau}_{1,GM}(m_e, 1/(N+1), 1) &= \tau_A(m_e) + \tau_B^{(1)}(m_e)
\end{aligned}
$$

[14]The mechanism could be made more robust, at the expense of complicating the proof, by not switching government spending to 0 and triggering big aggregate changes when only 1 household flags; we could instead only offer the "small bet" component of the change until a given proportion of people flag. This might also require to use flags from high-endowment households, and we do not pursue it further here, but it would be required if household actions were observed with noise, as in the microfoundations of anonymous games of Levine and Pesendorfer [10] and Fudenberg, Levine and Pesendorfer [5].

$\tau_A(m_e)$ is such that

$$\pi(m_e)u(e - \tau_A(m_e)) = \pi(m_e)u[e - \underline{\tau}_1^*(m_e)] + v(G^*(m_e))$$

$\tau_A(m_e)$ will necessarily exist if u is unbounded, and will not violate feasibility for the taxes of other households (to be defined below), provided N is sufficiently large.

$\tau_B^{(i)}$ is defined, for $i = 1, \ldots, N+1$, by

$$\tau_B^{(i)}(m_e) = \begin{cases} +\tau_\delta & \text{if } m_e \leq m_\delta^{(i)} \\ \\ -\tau_\delta & \text{if } m_e > m_\delta^{(i)} \end{cases}$$

with $\tau_\delta > 0$. $m_\delta^{(1)}$ is determined by

$$\begin{aligned} &\int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{N} Q(n|p + (1-p)\delta) \Bigg\{ \pi\left(\frac{n+1}{N+1}\right) \cdot \\ &\left[u\left(\underline{e} - \tau_A\left(\frac{n+1}{N+1}\right) - \tau_B^{(1)}\left(\frac{n+1}{N+1}\right)\right) - u\left(\underline{e} - \underline{\tau}_1^*\left(\frac{n+1}{N+1}\right)\right)\right] - \\ &v\left(G^*\left(\frac{n+1}{N+1}\right)\right)\Bigg\} f(p)dp = 0 \end{aligned}$$

Notice that the left-hand-side is continuous and strictly decreasing in $m_\delta^{(1)}$, strictly negative if $m_\delta^{(1)} = 1$ and strictly positive if $m_\delta^{(1)} = 0$, which ensures that $m_\delta^{(1)}$ that exactly solves the equation is unique and well defined. We will define $m_\delta^{(i)}$ for $i > 1$ below.

For households that do not flag, we set

$$\bar{\tau}_{GM}(m_e, 1/(N+1)) = \underline{\tau}_{0,GM}(m_e, 1/(N+1), 0) =$$
$$\underline{\tau}_{1,GM}(m_e, 1/(N+1), 0) = \max\left\{0, -\frac{\tau_A(m_e)}{N}, -\frac{\underline{\tau}_0^*(m_e)}{N}\right\} + \tau_\delta$$

This tax will be below $\underline{e}$, provided N is sufficiently large and τ_δ sufficiently small. Notice that we constructed the mechanism so that the government budget constraint is slack in many states; this makes the construction simpler and tidier, and will not affect the equilibrium payoffs by more than an infinitesimal amount, since all equilibria will feature flagging with arbitrarily low probability.

When more than 1 household flags, we construct the payoff of flagging households recursively, for $i = 2, \ldots, N+1$, as

$$\underline{\tau}_{0,GM}\left(m_e, \frac{i}{N+1}, 1\right) = \underline{\tau}_{0,GM}\left(m_e, \frac{i-1}{N+1}, 0\right)$$
$$\underline{\tau}_{1,GM}\left(m_e, \frac{i}{N+1}, 1\right) = \underline{\tau}_{1,GM}\left(m_e, \frac{i-1}{N+1}, 1\right) + \tau_B^{(i)}$$

where $m_\delta^{(i)}$ is now defined as

$$\int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{N} Q(n|p + (1-p)\delta)\pi\left(\frac{n+1}{N+1}\right) \cdot$$
$$\left[u\left(\underline{e} - \underline{\tau}_{1,GM}\left(m_e, \frac{i-1}{N+1}, 0\right) + \tau_B^{(i)}\right) - \right.$$
$$\left. u\left(\underline{e} - \underline{\tau}_{1,GM}\left(m_e, \frac{i-1}{N+1}, 0\right)\right)\right] f(p)dp = 0$$

As before, $m_\delta^{(i)}$ exists and is unique.

For households that do not flag, we set

$$\bar{\tau}_{GM}(m_e, i/(N+1)) = \underline{\tau}_{0,GM}(m_e, i/(N+1), 0) =$$
$$\underline{\tau}_{1,GM}(m_e, i/(N+1), 1) = \tau_\delta$$

As a last remark, notice that this mechanism is feasible even when all households flag; by the recursion, their tax in that case is exactly 0.

We next prove that, as $\delta \to 0$, all the equilibrium outcomes of this mechanism converge to the best equilibrium of the optimal direct mechanism (simply "best equilibrium" from now on). First, consider the best response of a household with low income. This household can only choose whether to flag. By the construction above, flagging is optimal if each high-income household misreports with probability greater than δ, not flagging is optimal if the probability of each high-income household misreporting is less than δ, and any flagging decision is optimal when the misreporting probability is exactly δ.[15] Consider next the best response of a household that has high-income. Its choice is whether to report truthfully or not. By construction, truthtelling is strictly optimal if any other household flags: in this case, transfers are independent of the report, but lying implies a positive probability of being audited and losing the entire consumption allocation. Truthtelling is also strictly optimal if all households have high income (in which case nobody can flag), as $\Delta^*(0; N) > 0$, at least when N is sufficiently large. It immediately follows that there can be no equilibria in which high-income households misreport their income with a probability higher than δ. The only pure-strategy equilibrium is thus one in which households report truthfully and no flags are raised.

[15] Our reasoning only considers symmetric strategies, but it can be proven that even the asymmetric equilibria of the game above converge to the best equilibrium.

There could be mixed-strategy equilibria in which high-income households misreport with probability exactly equal to δ. In order for this to be the case, the probability of a flag being raised must be such that high-income households are indifferent on their report. As $\delta \to 0$, conditional on no flags being raised, the outcomes converge to those of the best equilibrium, in which high-income households are exactly indifferent between lying and reporting truthfully. By contrast, conditional on a flag being raised, the payoff loss to a high-income household from misreporting does not converge to 0. Hence, the probability of a flag being raised must necessarily converge to 0 as $\delta \to 0$. This proves that, as $\delta \to 0$, all equilibrium outcomes of the game coincide with the best (equilibrium) outcome with probability converging to 1. Since utility is bounded below, the expected payoff loss from events in which the mixed-strategy equilibrium outcomes do not coincide with the best outcome is also converging to 0, proving the theorem. ■

References

[1] Stefania Albanesi and Christopher Sleet. Dynamic Optimal Taxation with Private Information. *CEPR Discussion Paper*, 4006, 2003.

[2] Carlos Da Costa and Ivan Werning. On the Optimality of the Friedman Rule with Heterogeneous Agents and Non-Linear Income Taxation. Mimeo, University of Chicago, 2002.

[3] Huberto M. Ennis and Todd Keister. Government Policy and the Probability of Coordination Failures. *European Economic Review*. Forthcoming.

[4] Huberto M. Ennis and Todd Keister. Optimal Fiscal Policy under Multiple Equilibria.

Journal of Monetary Economics. forthcoming.

[5] Drew Fudenberg, David K. Levine, and Wolfgang Pesendorfer. When Are Nonanonymous Players Negligible? *Journal of Economic Theory*, 79(1):46–71, 1998.

[6] Mikhail Golosov, Narayana Kocherlakota, and Aleh Tsyvinski. Optimal Indirect and Capital Taxation. *Review of Economic Studies*, 70(3):569–87, 2003.

[7] Mikhail Golosov and Aleh Tsyvinski. Designing Optimal Disability Insurance: A Case for Asset Testing. *NBER Working Paper*, 10792, 2004.

[8] Matthew O. Jackson. A Crash Course in Implementation Theory. *Social Choice and Welfare*, 18(4):655–708, 2001.

[9] Stefan Krasa and Anne P. Villamil. Optimal Multilateral Contracts. *Economic Theory*, 4(2):167–187, 1994.

[10] David K. Levine and Wolfgang Pesendorfer. When Are Agents Negligible? *American Economic Review*, 85(5):1160–1170, 1995.

[11] James A. Mirrlees. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies*, 38(2):175–208, 1971.

[12] John Moore. Implementation, Contracts, and Renegotiation in Environments with Complete Information. In Jean-Jacques Laffont, editor, *Advances in economic theory: Sixth World Congress*, volume 1, pages 182–202. Econometric Society Monographs, no. 20, 1992.

[13] Robert B. Wilson. Game-Theoretic Analyses of Trading Processes. In Truman F. Bewley, editor, *Advances in economic theory: Fifth World Congress*, pages 33–70. Econometric Society Monographs, no. 12, 1987.