Federal Reserve Bank of Minneapolis
Research Department

January 2006
Preliminary

Tax Riots*

Marco Bassetto

Federal Reserve Bank of Chicago
University of Minnesota
NBER

Christopher Phelan

Federal Reserve Bank of Minneapolis
University of Minnesota

ABSTRACT

This paper considers an optimal taxation environment where household income is private information, and the government randomly audits and punishes 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, and Roozbeh Hosseini and Vadym Lepetyuk for excellent research assistance. Marco Bassetto thanks the National Science Foundation and the Sloan Foundation for financial support. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Banks of Chicago, 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) generally respond to aggregate shocks, usually for simple budgetary reasons. That is, the transfers a government can afford given many households reporting low incomes is not as great as those a government can afford given many households reporting high incomes. This paper focuses generally on the design of tax 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 policies using standard mechanism design tools. As is well known, these tools ensure only that truthful revelation of income is one of possibly multiple equilibria of the game induced by the optimal tax mechanism. In this paper, we document several forces that make the optimal mechanism subject to "tax riots:" equilibria in which households find it optimal to misreport their income because they expect other households to do so. In common with the literature on crime prevention (see e.g. Bond and Hagerty [2]), tax riots emerge as a consequence of the limited resources available to the government to investigate suspicious reports. In addition, we find the optimal tax system to be increasingly redistributive in states in which many households are hit (or claim to be hit) by negative shocks. This feature provides a rationale for the many instances in which governments provide more generous benefits when many households face a common shock. As an example, the duration of unemployment insurance benefits has been routinely raised in response to increases in unemployment (see e.g. Moffitt [14]). While part of an optimal plan if the potential for multiple equilibria is neglected, the additional insurance available to poor people when there are more of them is a new engine of "tax riots:" when more of the other households are expected to claim adverse shocks, the benefits from misreporting increase and make it optimal for each household to deviate from truth telling.

The potential for an optimal direct mechanism to be subject to multiple equilibria is well known in the theoretical literature, and a significant body of research has studied the design of mechanisms that uniquely implement desirable outcomes.[1] In our analysis, we consider three possible solutions to the emergence of multiple equilibria. First, we consider a more general mechanism, that offers lotteries to induce households to indirectly report their beliefs about the strategy followed by their peers. This mechanism is successful in implementing outcomes arbitrarily close to the best, but it relies very heavily on the households' knowledge of the parameters of the economy and their sophistication in anticipating very precisely their peers' strategies. Our second solution considers the polar opposite case: we analyze the optimal ex post mechanism, which has (essentially) a unique equilibrium and is very robust to the households' (and the planner's) knowledge of parameters and beliefs about other players' strategies. This robustness comes at a significant loss. The extent of insurance is severely limited; furthermore, contrary to the optimal ex-ante incentive-compatible allocation, less redistribution is implemented in states in which more people claim low endowment. Finally, we consider a maximin mechanism, in which the government chooses the direct mechanism that achieves the highest worst-equilibrium utility. This mechanism resembles the optimal ex post mechanism, and in some cases it coincides with it.

[1]See e.g. the surveys in Moore [15] and Jackson [10], and the analysis of Bayesian implementation in Jackson [9].

The forces at work in our paper are present, in differing degrees, 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 [13]. More recently, the same tools have been adopted to study dynamic settings by Golosov, Kocherlakota and Tsyvinski [7], Golosov and Tryvinski [8], Albanesi and Sleet [1] and Da Costa and Werning [3]. Our analysis suggests that the optimal design of an optimal tax system may be significantly impacted if preventing multiple equilibria is an important consideration.

In section 2, we describe the economic environment that we consider. Section 3 characterizes 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. In section 4, we derive a non-direct mechanism which implements the truthful equilibrium uniquely. In section 5, we consider ex post implementation, and in section 6, a maximin 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

[2] With this assumption, our model is close to the multiple agent model of Krasa and Villamil [11].

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. After taxes and transfers 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 $u(c)$, where c is the household's private consumption level. 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).

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 $\pi(\frac{n}{N+1}) \equiv \min\{1, \frac{\overline{n}(N)}{n}\}$ of being audited. (As $N \to \infty$, $\pi(m)$ converges uniformly to $\min\{1, \frac{\overline{\pi}}{m}\}$ for $m \in [0, 1]$.)

A direct tax system is a vector of functions $(\overline{\tau}, \underline{\tau}_0, \underline{\tau}_1)(m)$ 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, and the tax paid by those who reveal $e = \underline{e}$, are audited, and found to have been truthful, all as functions of the fraction of households which report the low endowment, m. Taxes may be negative or positive, but must be no more than the endowment of the people they are levied upon.[3] Note that the functions $(\overline{\tau}, \underline{\tau}_1, \underline{\tau}_0)(m)$ implicitly depend on the number

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

of agents. This will become important later as we take $N \to \infty$.

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

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

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

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

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

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

where $Q(n|p,N) \equiv \binom{N}{n}p^n(1-p)^{N-n}$ 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 $(\overline{\tau}^*, \underline{\tau}_1^*, \underline{\tau}_0^*)(m)$ maximizes the utility of a given agent, or solves

(4)

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

subject to (1) and (3).

The objective function is continuous in the tax functions, and the constraint set is compact, hence a solution to this problem exists.[4] While the constraint set (3) is not convex, the appendix shows that changing variables from taxes to utility entitlements makes the objective function linear and the constraint set convex, and that the maximum is unique. 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.[5]

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 mechanism $(\overline{\tau}^*, \underline{\tau}_0^*, \underline{\tau}_1^*)$ 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 $(\overline{\tau}^*, \underline{\tau}_1^*, \underline{\tau}_0^*)$. 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

[4]Notice that the tax 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 functions are finite-dimensional vectors.

[5]The results carry easily over to the case in which the solution may be at a corner, but the notation becomes significantly more cumbersome. A technical appendix for that case is available upon request from the authors.

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

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

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

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

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

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

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

The resource constraint, which binds at an optimum, is

$$m\left[\pi\left(m\right)\underline{\tau}_1\left(m\right)+\left(1-\pi\left(m\right)\right)\underline{\tau}_0\left(m\right)\right]+\left(1-m\right)\overline{\tau}\left(m\right)=0, \qquad \text{for } m=0,\frac{1}{N+1},\ldots,1 \tag{8}$$

From now on, we use * superscripts to denote the optimal tax plan and the Lagrange multipliers that solve the first-order conditions. Further, let $(\bar{\tau}^*, \underline{\tau}_1^*, \underline{\tau}_0^*)(m; N)$ denote the optimal tax system for a given value of N. Note that for a given N, a tax system $(\bar{\tau}, \underline{\tau}_1, \underline{\tau}_0)(m; N)$ 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 $N \to \infty$.

Proposition 1. *There exists a unique value λ^∞, and unique functions $\bar{\tau}^\infty(m)$, $\underline{\tau}_1^\infty(m)$, $\mu^\infty(m)$ 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), and (8), as $N \to +\infty$, and the limiting incentive-compatibility constraint*

$$(9) \qquad \int_{\underline{p}}^{\bar{p}} (1 - m)\Delta^\infty(m) f(m) dm = 0,$$

where $\Delta^\infty(m) = u(\bar{e} - \bar{\tau}^\infty(m)) - (1 - \pi(m))u(\bar{e} - \underline{\tau}_0^\infty(m))$ 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.

The functions $\bar{\tau}^\infty$, $\underline{\tau}_1^\infty$, μ^∞, and $\underline{\tau}_0^\infty$ are continuous everywhere and continuously differentiable everywhere but points $\bar{\pi}$, $\underline{p}$, and $\bar{p}$, where they have finite left and right derivatives.

Finally, let $(m_N, N)_{N=1}^\infty$ be a sequence converging to (m, ∞). Then $\bar{\tau}^*(m_N; N) \to \bar{\tau}^\infty(m)$, $\underline{\tau}_1^*(m_N) \to \underline{\tau}_1^\infty(m)$, and $\underline{\tau}_0^*(m_N; N) \to \underline{\tau}_0^\infty(m)$.

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 mechanism; $\Delta^\infty(m)$ is the limit of $\Delta^*(m; N)$ as N goes to infinity.

Proposition 2. *There exists* $\hat{m} \in [\max\{\underline{p}, \overline{\pi}\}, \overline{p}]$ *such that for all* $m \in (\hat{m}, 1]$, $\Delta^\infty(m) < 0$.

Proof. See appendix ■

Proposition 2. 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 1. and 2. then imply our main result.

Theorem 1. *There exists* $\overline{N}$ *such that for all* $N \geq \overline{N}$, *the optimal direct mechanism denoted* *by* $(\overline{\tau}^*, \underline{\tau}_1^*, \underline{\tau}_0^*)(m)$ *admits an equilibrium where all households report the low endowment.*

Proof. From propositions 1. and 2., $\lim_{N\to\infty} \Delta^*\left(\frac{N}{N+1}; N\right) = \Delta^\infty(1) < 0$. 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) ex post implementation, and 3), a saddle-approach of maximizing the value of the worst equilibrium, which we prove to be equivalent to implementation in interim-dominant strategies.[6]

4. A Numerical Example

Calculating the optimal direct mechanism is straightforward. For a given value of λ (the Lagrange multiplier on the incentive constraint), and m (the fraction of agents with a low endowment), the system of equations defined by the limits as $N \to +\infty$ of the first-order conditions (5), (6), (7), and the resource constraint (8), define the variables $(\overline{\tau}(m), \underline{\tau}_1(m), \underline{\tau}_0(m), \mu(m))$. Given these functions, one can then calculate whether the incentive constraint is actually satisfied. One then iterates over λ to find the value of λ such that the incentive constraint holds at equality.

For a simple representative example, let $u(c) = 2\sqrt{c}$, $\overline{e} = \frac{3}{2}$, $\underline{e} = \frac{1}{2}$, $\overline{\pi} = .02$, and $f(p)$ be uniform $[0, 1]$. In this environment with full information, the functions $(\overline{\tau}(m), \underline{\tau}_1(m), \underline{\tau}_0(m))$ are each linear.

[6] Ennis and Keister [4, 5] 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.

Figure 1 displays $\overline{\tau}(m)$ under three scenarios: full information, private information under the optimal ex ante mechanism (which we have discussed so far), and private information under the optimal ex post mechanism, which we will discuss in section 5. As m increases, the amount necessary to tax those with high endowments to achieve consumption equality (the full information outcome) increases as well. With private information, taxes are, not surprisingly, lower since society is providing less insurance.

Figure 2 displays $\underline{\tau}_1(m)$. As m increases, the subsidy necessary to achieve consumption equality decreases with m, thus taxes (which here are negative) increase with m as well. The effect of private information is to provide less of a subsidy to the poor, decreasing insurance.

Figure 3 displays $\underline{\tau}_0(m)$. Here, private information plays a much more prominent role, and taxes are nonmonotonic. For low m (states where society is rich), results are driven by the discrepancy between the objective function and the constraint due to Bayes' law. While the objective function is the unconditional expected utility of a household, the incentive constraint is conditional on the household having high endowment. In computing expected values, Bayes' rule implies that a high-income household will regard as more likely states of nature in which the probability of low income (m) is low. Ceteris paribus, depressing consumption of low-income households has therefore a disproportionate benefit on incentives, compared to its welfare cost. For intermediate and high values of m, the primary role of taxes as an insurance mechanism ensure that transfers decrease as the number of high-income households decreases.

Finally, Figure 4 displays $\Delta(m)$, the gain to truth-telling as a function of m. That $\Delta(1)$ is negative confirms all households announcing low is an equilibrium.

5. 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.

Our main result of this section is then,

Theorem 2. *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 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 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 and transfers 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.

6. Ex Post Implementation

Wilson [16] argues that we should strive for more robust notions of implementation. In particular, in our environment, the direct mechanism and the general mechanism both depend on households having precise knowledge of $f(p)$.

We now consider the much stricter requirement of ex post implementation; this requires designing mechanisms such that the prescribed equilibrium actions to the households turn out to be optimal ex post, i.e., independently of the state of nature and of the actions taken by other households. This stricter notion no longer requires households to have any information on $f(p)$, nor on the strategy that other households will follow. An added benefit is such an approach eliminates the bad equilibrium in our environment and is particularly easy to characterize. Since the revelation principle applies in this context as well, there exists a direct mechanism that attains the best equilibrium within this restricted class. To find it, one simply replaces the original incentive constraint (3) with a requirement $\Delta(m) \geq 0$ for all m.

Proposition 3. *For all $\epsilon > 0$, there exists a mechanism with a unique ex post 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 $\delta \rightarrow 0$, the solution converges 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, ex post 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.

Figures 1-3 display taxes and transfers in the optimal ex post mechanism. Compared to the best incentive compatible allocation (the best equilibrium of the ex ante mechanism), there are two striking differences:

- The overall level of social insurance is greatly reduced. The impossibility of trading off incentives across states is very detrimental to the degree of insurance that is incentive compatible over a wide range of aggregate outcomes.
- The pattern of redistribution as a function of the aggregate performance of the economy is the opposite. The best incentive-compatible allocation exploits the cheap incentives available when few people report low income to provide generous redistribution when many people report a negative outcome. In the ex post mechanism, cheap incentives when few people report low income can only be used to provide close to full insurance in those states, while the limited auditing capacity restricts the government to trivial amounts of redistribution if many households report the low realization.

7. Maximin Direct Mechanisms and Interim-Dominant Strategy Implementation

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 3) or the best ex post equilibrium (our focus in section 5). 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 ex post implementation described in the previous section uses a direct mechanism, ex post mechanisms (direct or not) impose a very tight constraint set on a policymaker concerned with bad equilibria. An interesting question is the direct mechanism which delivers the best worst case scenario, or the best worst perfect Bayesian equilibrium. We will call this the "maximin" problem for brevity.

We restrict our attention to symmetric mechanisms and equilibria for simplicity of notation, though none of our results seems to hinge on this assumption.

A (symmetric) mechanism is now given by $(\bar{\tau}_0, \bar{\tau}_1, \underline{\tau}_0, \underline{\tau}_1)$ as functions of $n/(N+1)$, the fraction of people reporting low endowment (and of $N+1$, the total number of people in the economy).[7]

Compared to previous notation, we have the following differences:

- $\bar{\tau}_0$ was simply referred to as $\bar{\tau}$ previously, and denotes taxes of high-income households that declare high income.
- $\bar{\tau}_1$ represents taxes on high-income households that report low and are audited. When solving for the best equilibrium, we set $\bar{\tau}_1 \equiv \bar{e}$ without loss of generality. However, if households misreport with positive probability in equilibrium, we cannot take for granted that households found cheating will be assessed the maximal punishment.

The following proposition is very useful to characterize the solution to the maximin

[7]While in principle the government could offer lotteries (introduce randomness even conditional on the outcome $n/(N+1)$), DARA implies that this will not happen in the maximin, so we can keep our notation that prescribes single numbers. This can be proven along the same lines of the proof of theorem 4. below. Similarly, we assume without loss of generality that the government audits exactly $\pi\left(\frac{n}{N+1}\right)$ households. A mechanism where the government audits fewer households than the maximum it is able is equivalent to one where it audits the maximum and it then offers a lottery to audited households, some of them being treated as if they were not audited.

problem.

Proposition 4. *Let* x^B *be the maximum equilibrium probability of misreporting for a mechanism* $(\bar{\tau}_0^B, \bar{\tau}_1^B, \underline{\tau}_0^B, \underline{\tau}_1^B)$ *(such a probability exists by continuity of the incentive-compatibility constraint). Then there exists a mechanism* $(\bar{\tau}^F, \bar{e}, \underline{\tau}_0^F, \underline{\tau}_1^F)$ *that has a unique equilibrium in which households report truthfully with probability 1 and that attains an equilibrium utility level that is at least as high as the equilibrium of the original mechanism at* x^B*. The equilibrium of* $(\bar{\tau}^F, \bar{e}, \underline{\tau}_0^F, \underline{\tau}_1^F)$ *is also an interim dominant strategy equilibrium (i.e., given any probability* $x > 0$ *of misreports by other households, each household strictly prefers to report truthfully).*

Proof. See appendix. ■

Proposition (4.) implies that we can restrict our search to mechanisms where truth-telling is the (interim) dominant strategy; this is a great simplification, which offers a way of characterizing and/or computing solutions. It is worth noting that the maximin will not in general be attained, but will rather be a "supmin." This is due to the open-set nature of the problem: to rule out multiple equilibria, we require that households *strictly* prefer reporting truthfully when other households are not doing so with probability 1. We will thus be interested in approximately optimal mechanisms.

In section 3.we proved that the best direct mechanism necessarily has a second, very undesirable equilibrium. However, when N is large, there may be other direct mechanisms that have a unique dominant strategy equilibrium close to the best incentive-compatible allocation. In the next theorem, we establish a convenient necessary and sufficient condition to check whether this is the case.

THEOREM 3. *Consider the following tax plan:*

$$\bar{\tau}^{test}(m) = \begin{cases} \bar{\tau}^{\infty}(m) & \text{if } m \leq \bar{p} \\ -\frac{m}{1-m}\underline{e} & \text{if } m > \bar{p} \end{cases}$$

$$\underline{\tau}_1^{test}(m) = \begin{cases} \underline{\tau}_1^{\infty}(m) & \text{if } m \leq \bar{p} \\ \underline{e} & \text{if } m > \bar{p} \end{cases}$$

$$\underline{\tau}_0^{test}(m) = \begin{cases} \underline{\tau}_0^{\infty}(m) & \text{if } m \leq \bar{p} \\ \underline{e} & \text{if } m > \bar{p} \end{cases}$$

A necessary and sufficient condition for the solution to the maximin problem to converge to the best incentive-compatible allocation as $N \to \infty$ is that the following inequality holds:

(10)

$$\int_{\underline{p}}^{\bar{p}} \left[u\left(\bar{e} - \bar{\tau}^{test}(m + (1-m)x)\right) - (1 - \pi(m + (1-m)x))u\left(\bar{e} - \underline{\tau}_0^{test}(m + (1-m)x)\right) \right] \cdot$$
$$(1-m)f(m)dm \geq 0, \quad x \in [0,1)$$

Proof. See appendix. ■

The intuition of theorem 3. is simple. As N grows large, if households report truthfully, the probability of observing outcomes where more than a fraction $\bar{p}$ of households report low endowment goes to 0. Hence, the expected utility cost from perturbing the plan on those outcomes becomes negligible.[8] The "test" plan takes advantage of this flexibility to redistribute

[8]For this, it is important that utility is bounded below.

as much as possible in favor of the rich over these outcomes, which retain positive probability when high-income households are expected to misreport their type with probability $x > 0$. Equation (10) is the limiting condition for truth-telling to be a dominant strategy. If it holds, then for large N there will be plans close to the test plan for which truth-telling is dominant and that have a vanishing utility difference with respect to the best incentive-compatible allocation. If (10) does not hold, then imposing that truth-telling be a dominant strategy will force further changes in the tax plan, and those changes will have a discrete cost, even in the limit.

While (10) is an intuitive and computationally convenient way to check, it still involves endogenous variables. The following corollaries display two special cases where the result can be derived in terms of the exogenous parameters directly.

Corollary 1. If $\underline{p} = \bar{p}$, then the solution to the maximin problem converges to the best incentive-compatible allocation.

Proof. In this case, 10 becomes

$$\bar{\tau}^{\text{test}}(m) - (1 - \pi(m))\underline{\tau}_0^{\text{test}}(m) \geq 0, \quad m \in [\bar{p}, 1)$$

For $m = \bar{p}$, this equation coincides with (9); for greater values, it is satisfied given the definition of the test plan. ∎

When there is no aggregate uncertainty, as N becomes large, the law of large numbers implies that with very high probability a fraction of households very close to $\bar{p}$ will receive a low endowment. The utility of a tax plan depends only on the allocation it promises in a neighborhood of that point, leaving the government free to use arbitrary tax functions on all the values of m that would occur if households misreported their income.

As a numerical example, consider the preferences and technology of section 4.except for the probability distribution over p: assume that the probability of low income for each person is known and equal to 0.5. In an economy with 1,000,000 people, the probability that more than 50.2% of the people obtain the low endowment is less than 1/1,000,000. We consider changing the optimal plan as follows: if $m < .502$, the new plan coincides with the optimal plan; if $m > .502$, the new plan offers the highest possible incentives to report high, by switching to the "test" tax plan. In the truth-telling equilibrium, the utility cost of this change is trivial. Yet, the resulting plan ensures uniqueness: as the fraction of people filing false reports rises, the probability of surpassing the 50.2% threshold increases dramatically, and makes it very undesirable to falsely claim low endowment.

Corollary 2. If $\underline{p} < \bar{p} = 1$, then the solution to the maximin problem does not converge to the best incentive-compatible allocation.

Proof. In this case, the test plan coincides with $(\bar{\tau}^\infty, \underline{\tau}_1^\infty, \underline{\tau}_0^\infty)$. We proved earlier that Δ^∞ is continuous, and that $\Delta^\infty(1) < 0$; this implies that (10) will fail for sufficiently high values of x. ■

This case is the polar opposite of the preceding one. Here, the support of the distribution f extends all the way to 1, and the probability of outcomes where most households receive low endowment does not vanish even in the limit. In this case, perturbing the optimal tax plan to get rid of unwanted equilibria will necessarily involve a cost, no matter how large N is.

In the case of our numerical example, the maximin mechanism coincides with the optimal ex post mechanism. Given the assumption of a uniform distribution, this is true independently of preferences and endowment (a proof of this result is in the works). Contrary to the ex post mechanism, the maximin solution does allow trading off incentives across states. However, in order to rule out all mixed-strategy equilibria, as well as the undesirable pure-strategy equilibrium, requires either of the following:

- The government could concentrate its incentive provision in a region close to $m = 1$; given enough rewards for reporting truthfully in that region, households would then always strictly prefer reporting truthfully, independently of their expectations about other households' (possibly mixed) strategies. However, this is very costly: the best incentive-compatible allocation offers greatest redistribution precisely in this region, and the limited capacity to audit when m is close to 1 implies that incentive provision requires large changes in tax and transfer payments, at a very big insurance loss.
- Alternatively, the government could spread additional incentives throughout the range of possible realizations of m. One particular way of doing so is following the optimal ex post mechanism, which turns out to be optimal for a uniform distribution.[9]

8. Conclusion

In this paper we studied the design of optimal tax mechanisms when the government is concerned about the presence of multiple equilibria. To prevent households from coordinating on an undesirable equilibrium, the government faces a trade-off. At one end of the

[9] In experimenting with a truncated normal distribution, we found that, given our preferences, the maximin coincides with the optimal ex post mechanism when the variance is not too low, and oscillates very close to the ex post mechanism otherwise.

spectrum, it can rely on complicated mechanisms that hinge on the households' sophistication in understanding their environment and the strategic interaction with other households. At the opposite extreme, there are very blunt mechanisms that are simple and robust, but impose severe limits on the ability to achieve the governments' objectives (social insurance, in our case). The economic forces we emphasized are common to many other problems in public finance, and accounting for them has important implications for policy advice.

Appendix

A1. Sufficiency of the First-Order Conditions and Uniqueness of the Maximum of (4) subject to (1) and (3)

In the problem of maximizing (4) subject to (1) and (3), define $\bar{U}\left(\frac{n}{N+1}\right)$ the utility that the planner delivers to households reporting high-income. Similarly, let $\underline{U}_1\left(\frac{n}{N+1}\right)$ and $\underline{U}_0\left(\frac{n}{N+1}\right)$ be the utility promises to households that report low income and are or are not audited, respectively. It is straightforward to check that $\bar{U}$, $\underline{U}_1$ and $\underline{U}_0$ are strictly decreasing transformations of $\bar{\tau}$, $\underline{\tau}_0$ and $\underline{\tau}_1$; hence, the first-order conditions of the two maximization problems coincide. In terms of utility promises, the problem becomes

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

subject to

$$\frac{n}{N+1}\Big[\pi\left(\frac{n}{N+1}\right)\ u^{-1}\left(\underline{U}_1\left(\frac{n}{N+1}\right)\right)+\left(1-\pi\left(\frac{n}{N+1}\right)\right)\ u^{-1}\left(\underline{U}_0\left(\frac{n}{N+1}\right)\right)\Big]+ \left(1-\frac{n}{N+1}\right)u^{-1}\left(\overline{U}\left(\frac{n}{N+1}\right)\right)\leq\frac{n}{N+1}\underline{e}+\left(1-\frac{n}{N+1}\right)\bar{e}.$$

and (3), where Δ is now equal to

$$\Delta\left(\frac{n}{N+1}\right)=\bar{U}\left(\frac{n}{N+1}\right)-\left(1-\pi\left(\frac{n+1}{N+1}\right)\right)u\left[\bar{e}-\underline{e}+u^{-1}\left(\underline{U}_0\left(\frac{n}{N+1}\right)\right)\right].$$

The objective function is linear in the utility promise. Since u^{-1} is strictly convex, the resource constraint yields a strictly convex set. For the (3), we notice that the second derivative of Δ

with respect to $\underline{U}_0$ is

$$\frac{1}{\left[u'\left[u^{-1}\left(\underline{U}_0\left(\frac{n}{N+1}\right)\right)\right]\right]^2 u'\left[\bar{e}-\underline{e}+u^{-1}\left(\underline{U}_0\left(\frac{n}{N+1}\right)\right)\right]}\cdot$$
$$\left[\frac{u''\left[\bar{e}-\underline{e}+u^{-1}\left(\underline{U}_0\left(\frac{n}{N+1}\right)\right)\right]}{u'\left[\bar{e}-\underline{e}+u^{-1}\left(\underline{U}_0\left(\frac{n}{N+1}\right)\right)\right]}-\frac{u''\left[u^{-1}\left(\underline{U}_0\left(\frac{n}{N+1}\right)\right)\right]}{u'\left[u^{-1}\left(\underline{U}_0\left(\frac{n}{N+1}\right)\right)\right]}\right]\geq 0,$$

where the inequality follows from the assumption of decreasing absolute risk aversion. Hence, the set of utility promises that satisfies (3) is (weakly) convex.

Finally, notice that the resource constraint must be binding at an optimum: otherwise, it would be possible to improve upon the objective function without tightening the incentive constraint by increasing $\bar{\tau}$ and/or $\underline{\tau}_1$. Since the resource constraint set is strictly convex, this implies a unique maximum.

A2. Proof of Proposition 1.

This proof of this proposition requires several lemmas.

LEMMA 1. $\limsup_{N\to\infty}\lambda^*(N)<\infty$.

Proof. Suppose by contradiction that there exists a subsequence $\{N_t\}_{t=1}^{\infty}$ such that $\lambda^*(N_t)\to_{t\to\infty}\infty$. The first-order conditions imply $\bar{e}-\bar{\tau}^*(m;N_t)>\underline{e}-\underline{\tau}_1^*(m;N_t)>\underline{e}-\underline{\tau}_0^*(m;N_t)$. This fact, along with the fact that the resource constraint must be binding, implies that $\bar{e}-\bar{\tau}^*(m;N_t)$ is uniformly bounded away from 0. We call this bound $\underline{c}$.

Consider now $\epsilon>0$. We analyze the solution of the system the system (8), (5), (6), and (7) on $m\in[0,1-\epsilon]$. Equation (8) implies that $u'(\bar{e}-\bar{\tau}^*(m;N_t))$ must remain uniformly bounded away from 0 on this interval. Equation (5) then implies that $\mu^*(m;N_t)$ must diverge uniformly. If $\lim_{c\to 0}u'(c)<+\infty$, (7) would have no interior solution in the limit, contrary to our assumption. If $\lim_{c\to 0} u'(c) = \infty$, then (7) implies that $\underline{e} - \underline{\tau}_0^*(m; N_t)$ converges uniformly to 0. Hence, $\Delta^*(m; N_t)$ is asymptotically uniformly greater or equal than $u(\underline{c})$ on $[0, 1-\epsilon]$.

Consider now the interval $m \in [1-\epsilon, 1]$. From (6) and (7) and feasibility, we obtain

$$\underline{\tau}_0^*(m; N) \geq -\frac{\epsilon \bar{e}}{1-\epsilon},$$

which implies

$$\Delta^*(m; N_t) \geq -u\left(\bar{e} + \frac{\epsilon \bar{e}}{1-\epsilon}\right), m \in [1-\epsilon, 1]$$

As $t \to +\infty$, the left-hand side of (3) is asymptotically greater or equal to

$$u(\underline{c}) \int_{\underline{p}}^{\min\{\bar{p}, 1-\epsilon\}} f(p)dp - I_{1-\epsilon<\bar{p}} u\left(\bar{e} + \frac{\epsilon \bar{e}}{1-\epsilon}\right) \int_{1-\epsilon}^{\bar{p}} f(p)dp,$$

where $I_{1-\epsilon<\bar{p}} = 1$ if $1-\epsilon < \bar{p}$, and is 0 otherwise. For ϵ sufficiently small,[10] the quantity above is strictly positive. This would imply that the incentive constraint is asymptotically slack, which contradicts $\lim_{t\to+\infty} \lambda^*(N_t) = +\infty$. ■

Lemma 2. $\mu^*(m; N)$ is uniformly bounded as $N \to +\infty$.

Proof. By contradiction, suppose $\exists (m_t, N_t)_{t=0}^{\infty} : \lim_{t\to+\infty} \mu^*(m_t, N_t) \to \infty$. We previously established that $\lambda^*(N_t)$ is a bounded sequence. If $\lim_{c\to 0} u'(c) < +\infty$, this would imply that (5), (6), (7) have no interior solution, contradicting our assumption. Otherwise, these conditions imply that $\bar{\tau}^*(m_t, N_t) \to_{t\to+\infty} \bar{e}$, $\underline{\tau}_0^*(m_t, N_t) \to_{t\to+\infty} \underline{e}$, and $\underline{\tau}_1^*(m_t, N_t) \to_{t\to+\infty} \underline{e}$.

[10]Notice that the lower bound $\underline{c}$ is independent of ϵ.

In this case, the resource constraint (8) is asymptotically slack, which yields a contradiction.

■

Lemma 3. 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 4. There exists a value E_c such that consumption of all households in the optimal plan is below E_c, independently of N.

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

$$\text{(A1)}\quad 0 = \begin{cases} u'\left(\underline{e} - \underline{\tau}_0\right) - \lambda u'\left(\bar{e} - \underline{\tau}_0\right)\left(\frac{1}{m} - 1\right) - \mu & \text{if } m \in [\underline{p}, \bar{p}] \\ u'\left(\underline{e} - \underline{\tau}_0\right) - \lambda u'\left(\bar{e} - \underline{\tau}_0\right)\left(\frac{1}{\underline{p}} - 1\right) - \mu & \text{if } m \in [0, \underline{p}] \\ u'\left(\underline{e} - \underline{\tau}_0\right) - \lambda u'\left(\bar{e} - \underline{\tau}_0\right)\left(\frac{1}{\bar{p}} - 1\right) - \mu & \text{if } m \in [\bar{p}, 1] \end{cases}$$

$$\text{(A2)}\quad 0 = \min\left\{m, \bar{\pi}\right\}\underline{\tau}_1 + \max\left\{0, m - \bar{\pi}\right\}\underline{\tau}_0 + (1 - m)\,\bar{\tau}$$

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)-(7) and (8), with the extensions (A1) and (A2), are continuous functions of $\bar{\tau}, \underline{\tau}_0, \underline{\tau}_1, \mu, \lambda, m, N$ over the domain and with the metric defined above.[11] By compactness, they also are uniformly continuous.

We are now ready to prove proposition 1..

Proof.

1. Existence:

   By lemma 1., 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 λ^{∞} (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 $m_{N_t} \to_{t\to+\infty} m$. $\bar{\tau}^*(m_{N_t};N_t)$, $\underline{\tau}_1^*(m_{N_t};N_t)$, $\underline{\tau}_0^*(m_{N_t};N_t)$, $\mu^*(m_{N_t};N_t)$ satisfy the first-order conditions (5)-(7), and the resource constraint (8). Since the tax policy and the multipliers are uniformly bounded, there exists a convergent subsequence of these values. Let $\bar{\tau}^{\infty}(m)$, $\underline{\tau}_1^{\infty}(m)$, $\underline{\tau}_0^{\infty}(m)$, $\mu^{\infty}(m)$ be the limit. By uniform continuity of the first-order conditions and the resource constraint, these limits satisfy equations (5), (6), (A1), and (A2). By the implicit function theorem, the resulting values are continuously differentiable, except at $m=\bar{\pi}$, $m=\underline{p}$, and $m=\bar{p}$, where they are only continuous and have finite left and right derivative.

   Let F_{N_t} be the c.d.f. of the fraction of N_t households receiving low income, conditional on the N_t+1st household having high income. As $N_t\to\infty$, this distribution converges weakly to one with p.d.f. proportional to $(1-m)f(m)$, where m is the fraction of people

[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 $\frac{[p^{m_N}+(1-p)^{1-m_N}]^{N+1}f(p)}{\int_{\underline{p}}^{\bar{p}}[p^{m_N}+(1-p)^{1-m_N}]^{N+1}f(p)dp}$ 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.

receiving low income. Next, extend $\bar{\tau}^*(m, N_t)$ to $[0,1]$ by defining

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

where $i(m, N_t)$ is the smallest integer such that $\frac{i(m,N_t)}{N_t+1} \geq m$ if $m < 1$, and is equal to N_t if $m = 1$. We can extend $\underline{\tau}_0^*$ to $[\bar{\pi}, 1]$. The functions $\bar{\tau}^{**}(m; N_t)$, and $\underline{\tau}_0^{**}(m; N_t)$ coincide with $\bar{\tau}^*(m; N_t)$ and $\underline{\tau}_0^*(m; N_t)$ on all points of positive probability mass in equation (3), and they converge uniformly to the continuous functions $\bar{\tau}^\infty(m)$ and $\underline{\tau}_0^\infty(m)$. Since (3) holds for all values of N, it then follows that λ^∞, $\bar{\tau}^\infty(m)$, and $\underline{\tau}_0^\infty(m)$ satisfy (9) as well.

2. Uniqueness:

If there were multiple subsequences converging to different values, it would imply that the limiting first-order conditions and constraints (5), (6), (A1), (A2) and (9) have more than one solution. We prove that this is impossible. First, we prove that, given λ and m, (5), (6), (A1) and (A2) have a unique solution. Then, we show that, at all solutions, (9) is strictly increasing in λ when $\bar{\tau}$, $\underline{\tau}_0$, $\underline{\tau}_1$ and μ are adjusted to ensure that (5), (6), (A1) and (A2) are met.

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

For $m \in [\underline{p}, \bar{p}]$, equation (A1) and $\mu(m) > 0$ implies

$$\text{(A3)} \quad \frac{u'(\underline{e} - \underline{\tau}_0)}{u'(\bar{e} - \underline{\tau}_0)} > \lambda \frac{1-m}{m}.$$

That u displays nonincreasing absolute risk aversion (DARA) implies

$$\text{(A4)} \quad \frac{u''(\underline{e}-\underline{\tau}_0)}{u''(\overline{e}-\underline{\tau}_0)} \geq \frac{u'(\underline{e}-\underline{\tau}_0)}{u'(\overline{e}-\underline{\tau}_0)}.$$

Equations (A3) and (A4) imply

$$\frac{u''(\underline{e}-\underline{\tau}_0)}{u''(\overline{e}-\underline{\tau}_0)} > \lambda\frac{1-m}{m}$$

or

$$\text{(A5)} \quad -u''(\underline{e}-\underline{\tau}_0)+\lambda\frac{1-m}{m}u''(\overline{e}-\underline{\tau}_0) > 0.$$

This last equation is the derivative of (A1) with respect to $\underline{\tau}_0$, thus the left hand side of (A1) 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. The same steps can be repeated for $m \in [0, \underline{p})$ and $m \in (\bar{p}, 1]$. Thus $\overline{\tau}$, $\underline{\tau}_1$, and $\underline{\tau}_0$ are all increasing functions 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.

We now need to prove that there is a unique value of λ which, in combination with the functions above, satisfies (9) as well. We do so by proving that the left-hand side of (9) is locally strictly increasing in λ at all solutions, when the functions $\bar{\tau}^{\infty}(m;\lambda)$, $\underline{\tau}_1^{\infty}(m;\lambda)$, $\underline{\tau}_0^{\infty}(m;\lambda)$, and $\mu^{\infty}(m;\lambda)$ are adjusted to satisfy (5), (6), (A1), and (A2). 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 strictly decrease to maintain (5), and that (if $m > \bar{\pi}$) $\underline{\tau}_0$ will have to strictly increase to maintain (A1). In general, the resulting change will no longer satisfy the resource constraint (A2); therefore, the change in λ will require a change in μ, which will affect $\underline{\tau}_1$ as well. However, since (locally) all taxes are increasing in μ, to satisfy (A2), the final outcome will necessarily involve a lower value for $\bar{\tau}$ and (if $m > \bar{\pi}$) a higher value for $\underline{\tau}_0$. Since this is true for all values of m, this proves that the left-hand side of (9) is strictly increasing in λ.

■

A3. Proof of proposition 2.

The proof requires some lemmas.

LEMMA 5. If $\Delta^\infty(m) \le 0$, then $\overline{\tau}^\infty(m) > \underline{\tau}_0^\infty(m)$.

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

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

$$u(\overline{e} - \overline{\tau}^\infty(m)) - (1 - \pi(m))u(\overline{e} - \overline{\tau}^\infty(m)) = \pi(m)u(\overline{e} - \overline{\tau}^\infty(m)) \ge \pi(m)u(\overline{e} - \underline{\tau}_0^\infty(m)) > 0,$$

a contradiction. ■

LEMMA 6. If $\Delta^\infty(m) \le 0$ and $m \in [\max\{\underline{p}, \bar{\pi}\}, 1]$, then $\mu^{\infty\prime}(m) > 0$.

Proof. Fix $\lambda = \lambda^\infty$. For all $m \le \bar{p}$ and μ, let $x(\mu, m)$ denote the value of $\underline{\tau}_0$ which solves (A1), and let $c(\mu, m) = \underline{e} - x(\mu, m)$. The partial derivative of (A1) with respect to $\underline{\tau}_0$ is strictly positive from equation (A5), while its partial derivative with respect to μ is strictly negative. Thus $x_\mu(\mu, m) > 0$. Likewise, the partial derivative of (A1) with respect to m is strictly positive, thus $x_m(\mu, m) < 0$. These imply $c_\mu(\mu, m) < 0$ and $c_m(\mu, m) > 0$.

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

$$\text{(A6)}\ \ m\underline{e} + (1-m)\overline{e} - (m - \overline{\pi})c(\mu, m) - \overline{\pi}u'^{-1}(\mu) - (1-m)u'^{-1}(\mu/(1-\lambda^\infty)) = 0.$$

Let $z(\mu, m)$ denote the left hand side (A6). The partials of z with respect to m and μ are

$$\text{(A7)}\ \ z_m(\mu, m) = \underline{e} - \overline{e} + [u'^{-1}(\mu/(1+\lambda^\infty)) - c(\mu, m)] - (m - \overline{\pi})c_m(\mu, m)$$

and

$$\text{(A8)}\ \ z_\mu(\mu, m) = -(m - \overline{\pi})c_\mu(\mu, m) - \frac{\overline{\pi}}{u''(u'^{-1}(\mu))} - \frac{1-m}{u''(u'^{-1}(\mu/(1+\lambda)))}.$$

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

$$\text{(A9)}\ \ z_m(\mu^\infty(m), m) = \underline{\tau}_0^\infty(m) - \overline{\tau}^\infty(m) - (m - \overline{\pi})c_m(\mu^\infty(m), m) \leq \underline{\tau}_0^\infty(m) - \overline{\tau}^\infty(m),$$

from $c_m(\mu, m) > 0$. By lemma 5., $\Delta^\infty(m) \leq 0$ implies $\overline{\tau}^\infty(m) > \underline{\tau}_0^\infty(m)$, hence $\mu^{\infty\prime}(m) > 0$. The proof for $m \in [\bar{p}, 1]$ is entirely analogous, except that $x_m(\mu, m) = 0$. ■

Lemma 7. For all $m \in [\max\{\underline{p}, \overline{\pi}\}, 1]$, if $\Delta^{\infty}(m) \leq 0$, then $\Delta^{\infty\prime}(m) < 0$.

Proof. We provide here the proof of the result on the range $m < \bar{p}$. The proof for $m > \bar{p}$ follows from repeating identical steps, adjusting the equations to take into account the appropriate case in equation (A1).[12]

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

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

Differentiating (5) with respect to m delivers

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

Likewise differentiating (A1) with respect to m delivers

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

[12] At the point $\bar{p}$, $\Delta^{\infty}(m)$ is not differentiable, but both its right and left derivatives are negative, as implied by the separate proofs on the two intervals. Similarly, at the lower bound the derivative below should be interpreted as the right derivative.

These imply

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

since the denominator of the second term is negative from equation (A5). Next,

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

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

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

$$\text{(A10)}\quad (1+\lambda^{\infty})u''(\overline{e}-\overline{\tau}^{\infty}(m))>u''(\underline{e}-\underline{\tau}_0^{\infty}(m))-\lambda^{\infty}\frac{1-m}{m}u''(\overline{e}-\underline{\tau}_0^{\infty}(m)).$$

Equations (5) and (A1) imply

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

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

$$\text{(A11)}\quad \theta u'(\underline{e}-\underline{\tau}_0^{\infty}(m))=u'(\overline{e}-\overline{\tau}^{\infty}(m))(1+\lambda^{\infty}),$$

and

$$\text{(A12)}\quad (1-\theta)u'(\underline{e}-\underline{\tau}_0^\infty(m)) = \lambda^\infty \frac{1-m}{m} u'(\overline{e}-\underline{\tau}_0^\infty(m)).$$

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

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

which combined with (A11) implies

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

or

$$\text{(A13)}\quad \theta u''(\underline{e}-\underline{\tau}_0^\infty(m)) < (1+\lambda^\infty)u''(\overline{e}-\overline{\tau}^\infty(m)).$$

Decreasing absolute risk aversion also implies

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

which combined with (A12) and rearranging implies

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

Adding (A13) and (A14) delivers equation (A10), proving $\Delta^{\infty\prime}(m) < 0$. ■

We are now ready to prove proposition 2..

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 (9) holds as an equality. Thus,

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

proving their exists $\hat{m} \in [\max\{\underline{p}, \bar{\pi}\}, \overline{p}]$ such that $\Delta(\hat{m}) \leq 0$.[13] Lemma 7. then recursively implies $\Delta(m) < 0$ for all $m \in (\hat{m}, 1]$. ■

A4. Proof of Theorem 2.

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

$$\bar{\tau}_{GM}(m_e, 0; N) = \bar{\tau}^*(m; N)$$

$$\underline{\tau}_{0,GM}(m_e, 0, 0; N) = \underline{\tau}_0^*(m_e; N)$$

$$\underline{\tau}_{1,GM}(m_e, 0, 0; N) = \underline{\tau}_1^*(m_e; N)$$

Consider next the case in which one household flags.[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 a component that incorporates a bet on the

[13]Note that $\bar{\pi} < \bar{p}$, for otherwise the incentive constraint would not be binding in the limit.

[14]The mechanism could be made more robust, at the expense of complicating the proof, by not 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 [12] and Fudenberg, Levine and Pesendorfer [6].

aggregate distribution of reports. This bet is advantageous when high-endowment households misreport their income with a probability greater than δ. Formally,

$$\underline{\tau}_{0,GM}(m_e, 1/(N+1), 1; N) = \underline{\tau}_0^*(m_e; N)$$

$$\underline{\tau}_{1,GM}(m_e, 1/(N+1), 1; N) = \underline{\tau}_1^*(m_e; N) + \tau^{(1)}(m_e; N)$$

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

$$\tau^{(i)}(m_e; N) = (m_\delta^{(i)}(N) - m_e)\tau_\delta$$

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

$$\int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{N} Q(n|p + (1-p)\delta)\pi\left(\frac{n+1}{N+1}\right) \cdot$$
$$\left[u\left(\underline{e} - \underline{\tau}_1^*\left(\frac{n+1}{N+1}\right) - \tau^{(1)}\left(\frac{n+1}{N+1}; N\right)\right) - u\left(\underline{e} - \underline{\tau}_1^*\left(\frac{n+1}{N+1}; N\right)\right)\right] pf(p)dp = 0$$

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

For households that do not flag, we set

$$\bar{\tau}_{GM}(m_e, 1/(N+1); N) = \underline{\tau}_{0,GM}(m_e, 1/(N+1), 0; N) =$$
$$\underline{\tau}_{1,GM}(m_e, 1/(N+1), 0; N) = \max_m \max\left\{0, -\frac{\underline{\tau}_1^*(m; N)}{N}\right\} + \tau_\delta$$

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

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

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

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

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

For households that do not flag, we set

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

As a last remark, notice that this mechanism is feasible even when all households flag, since their tax will remain nonnegative in this case.

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. 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

[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.

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 converge to 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. ■

A5. Proof of Proposition 4.

Proof. The utility attained by $(\bar{\tau}_0^B, \bar{\tau}_1^B, \underline{\tau}_0^B, \underline{\tau}_1^B$ and a given probability of misreporting x is

$$\int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{N} Q(n|p+(1-p)x, N) \Bigg\{ (1-p)(1-x) u\left(\bar{e} - \bar{\tau}_0^B\left(\frac{n}{N+1}\right)\right) +$$
$$(1-p)x\left[\pi\left(\frac{n+1}{N+1}\right) u\left(\bar{e} - \bar{\tau}_1^B\left(\frac{n+1}{N+1}\right)\right) + \left(1-\pi\left(\frac{n+1}{N+1}\right)\right) u\left(\bar{e} - \underline{\tau}_0^B\left(\frac{n+1}{N+1}\right)\right)\right] +$$
$$p\left[\pi\left(\frac{n+1}{N+1}\right) u\left(\underline{e} - \underline{\tau}_1^B\left(\frac{n+1}{N+1}\right)\right) + \left(1-\pi\left(\frac{n+1}{N+1}\right)\right) u\left(\underline{e} - \underline{\tau}_0^B\left(\frac{n+1}{N+1}\right)\right)\right]\Bigg\} f(p)dp$$

The resource constraint requires

$$\frac{N+1-n}{N+1}\bar{\tau}_0^B + \frac{n}{N+1}\pi\left(\frac{n}{N+1}\right) \min\left\{\bar{\tau}_1^B\left(\frac{n}{N+1}\right), \underline{\tau}_1^B\left(\frac{n}{N+1}\right)\right\} +$$
$$\frac{n}{N+1}\left[1-\pi\left(\frac{n}{N+1}\right)\right]\underline{\tau}_0^B\left(\frac{n}{N+1}\right) \geq 0$$

Note that this constraint embeds the assumption that households that are audited can be treated differently based on their income only ex post, by sending some of their endowment to waste.

For interior misreporting probabilities $x \in (0,1)$ to be an equilibrium, we need

$$\int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{N} Q(n|p+(1-p)x, N) u\left(\bar{e} - \bar{\tau}_0^B\left(\frac{n}{N+1}\right)\right)(1-p)f(p)dp =$$
$$\int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{N} Q(n|p+(1-p)x, N)\Bigg[\pi\left(\frac{n+1}{N+1}\right) u\left(\bar{e} - \bar{\tau}_1^B\left(\frac{n+1}{N+1}\right)\right) +$$
$$\left(1-\pi\left(\frac{n+1}{N+1}\right)\right) u\left(\bar{e} - \underline{\tau}_0^B\left(\frac{n+1}{N+1}\right)\right)\Bigg](1-p)f(p)dp$$

For equilibria with $x = 0$ or $x = 1$, the appropriate inequality has to hold.

First, if the highest equilibrium probability of misreporting for the mechanism at hand is $x^B = 0$, then the proof is immediate.

If $x^B = 1$, then strict concavity of u and the resource constraint imply that the no-tax mechanism improves utility:

$$\bar{\tau}^F \equiv 0, \quad \underline{\tau}_0^F \equiv 0, \quad \underline{\tau}_1^F \equiv 0$$

This mechanism has truth-telling as the unique equilibrium.

Consider now the case in which $x^B \in (0,1)$. We set the new mechanism so that, when exactly n households report low, a household gets the same expected utility as it would get under the original mechanism, conditional on n households having (not just reporting) the low endowment. The only exception is for high-income households that misreport and are audited, for whom we set the penalty at the maximum rate (though we could have preserved their utility as well). For now, define $\phi^F\left(\frac{n}{N+1}\right)$ as the auditing probability in the new mechanism when n people report low endowment. In the construction below, ϕ^F will sometimes be strictly below (but never above) the maximal auditing rate π. We will later argue that the mechanism can be changed to increase the auditing probability to the maximum without loss.

$$u\left(\bar{e}-\bar{\tau}^F\left(\frac{n}{N+1}\right)\right)=\sum_{k=0}^{N-n}Q(k|x^B,N-n)u\left(\bar{e}-\bar{\tau}_0^B\left(\frac{n+k}{N+1}\right)\right),\quad n=0,\dots,N$$

$$\phi^F\left(\frac{n}{N+1}\right)u\left(\underline{e}-\underline{\tau}_1^F\left(\frac{n}{N+1}\right)\right)=$$
$$\sum_{k=0}^{N+1-n}Q(k|x^B,N+1-n)\pi\left(\frac{n+k}{N+1}\right)u\left(\underline{e}-\underline{\tau}_0^B\left(\frac{n+k}{N+1}\right)\right),\quad n=1,\dots,N+1$$

$$\phi^F\left(\frac{n}{N+1}\right)=\sum_{k=0}^{N+1-n}Q(k|x^B,N+1-n)\pi\left(\frac{n+k}{N+1}\right),\quad n=1,\dots,N+1$$

$$\left(1-\phi^F\left(\frac{n}{N+1}\right)\right)u\left(\underline{e}-\underline{\tau}_0^F\left(\frac{n}{N+1}\right)\right)=$$
$$\sum_{k=0}^{N+1-n}Q(k|x^B,N+1-n)\left(1-\pi\left(\frac{n+k}{N+1}\right)\right)u\left(\underline{e}-\underline{\tau}_0^B\left(\frac{n+k}{N+1}\right)\right)\quad n=1,\dots,N+1$$

Strict concavity of u and weak convexity of the resource constraint set imply that the solution to the equations above (which is unique on all positive-probability events) satisfies the resource constraint. The expected utility of the households is unaffected by this change, provided every high-income household was misreporting with probability x^B under the old mechanism and is telling the truth under the new one. To see that this is the case, the only point worth noting is that we are using the fact that (because of the incentive-compatibility constraint) high-income households are indifferent between reporting truthfully or not under the old mechanism.

Furthermore, DARA implies

$$\left(1-\pi\left(\frac{n}{N+1}\right)\right)u\left(\bar{e}-\underline{\tau}_0^F\left(\frac{n}{N+1}\right)\right)\leq$$
$$\sum_{k=0}^{N-n}Q(k|x^B,N-n)\left(1-\phi^F\left(\frac{n+k}{N+1}\right)\right)u\left(\bar{e}-\underline{\tau}_0^B\left(\frac{n+k}{N+1}\right)\right)$$

Next, notice

$$\sum_{n=0}^{N}Q(n|p+(1-p)x,N)u\left(\bar{e}-\bar{\tau}^F\left(\frac{n}{N+1}\right)\right)=$$
$$\text{(A15)}\quad \sum_{n=0}^{N}Q(n|p+(1-p)x,N)\sum_{k=0}^{N-n}Q(k|x^B,N-n)u\left(\bar{e}-\bar{\tau}_0^B\left(\frac{n+k}{N+1}\right)\right)=$$
$$\sum_{n=0}^{N}Q(n|(1-x^B)(1-p(1-x))+x^B,N)u\left(\bar{e}-\bar{\tau}_0^B\left(\frac{n}{N+1}\right)\right)$$

The last equality comes from the following observation. The random count that is the argument of taxes in the first two lines is the sum of the number of reported failures (n), which occur with probability $p+(1-p)x$, and of occurrences of a further event that occurs with probability x^B conditional on a success. Hence the overall distribution of a single observation in the count is Bernoulli with probability $(1-x^B)(1-p(1-x))+x^B$.

Similarly, we have

(A16)
$$\sum_{n=0}^{N}Q(n|p+(1-p)x,N)\left(1-\pi\left(\frac{n}{N+1}\right)\right)u\left(\bar{e}-\underline{\tau}^F\left(\frac{n}{N+1}\right)\right)\leq$$
$$\sum_{n=0}^{N}Q(n|p+(1-p)x,N)\sum_{k=0}^{N-n}Q(k|x^B,N-n)\left(1-\pi\left(\frac{n+k}{N+1}\right)\right)u\left(\bar{e}-\bar{\tau}_0^B\left(\frac{n+k}{N+1}\right)\right)=$$
$$\sum_{n=0}^{N}Q(n|(1-x^B)(1-p(1-x))+x^B,N)\left(1-\pi\left(\frac{n}{N+1}\right)\right)u\left(\bar{e}-\bar{\tau}_0^B\left(\frac{n}{N+1}\right)\right)$$

Since we know that the old mechanism does not have any equilibria for $x > x^B$, (A15), (A16) and the incentive-compatibility constraint imply that the new mechanism does not have any equilibria for $x > 0$.

In the mechanism we just constructed, households that report low income have a probability of being audited that is less than $\pi(m)$. It is immediate to see that the mechanism coincides with one where the probability of being audited is exactly $\pi(m)$, but, after an audit, the government draws a lottery and ignores the audit with probability $\frac{1-\phi^F(m)}{\pi(m)}$. Because of decreasing (and positive) risk aversion, it is possible to repeat the steps of the proof above to construct a new mechanism that gets rid of this unnecessary uncertainty, while preserving expected utility and enhancing incentives to report the truth. ■

A6. Proof of theorem 3.

Proof. First, we prove sufficiency. Let $\{\delta_N\}_0^\infty$ be a strictly positive sequence converging to 0, and consider the following sequence of plans:

$$\bar{\tau}^N(m) = \begin{cases} \bar{\tau}^\infty(m) + (1-\pi(m))m\xi_N(m) & \text{if } m \leq \bar{p} \\ -\frac{m}{1-m}\underline{e} & \text{if } m > \bar{p} \end{cases}$$

$$\underline{\tau}_1^N(m) = \begin{cases} \underline{\tau}_1^\infty(m) & \text{if } m \leq \bar{p} \\ \underline{e} & \text{if } m > \bar{p} \end{cases}$$

$$\underline{\tau}_0^N(m) = \begin{cases} \underline{\tau}_0^\infty(m) - (1-m)\xi_N(M) & \text{if } m \leq \bar{p} \\ \underline{e} & \text{if } m > \bar{p}, \end{cases}$$

where $\xi_N(m)$ is such that

$$u\left(\bar{e}-\bar{\tau}^N(m)\right)-(1-\pi(m))u\left(\bar{e}-\underline{\tau}_0^N(m)\right)-u\left(\bar{e}-\bar{\tau}^\infty(m)\right)+(1-\pi(m))u\left(\bar{e}-\underline{\tau}_0^\infty(m)\right)=\delta_N$$

Such plans will always exist, provided $\{\delta_N\}_0^\infty$ remains sufficiently close to 0. The plans will uniformly converge to $(\bar{\tau}^{\text{test}}, \underline{\tau}_0^{\text{test}}, \underline{\tau}_1^{\text{test}})$. If all households report truthfully, the expected utility from this sequence of plans converges to the expected utility attained in the limit by $(\bar{\tau}^\infty, \underline{\tau}_0^\infty, \underline{\tau}_1^\infty)$. Furthermore, when the probability of high-income households misreporting their type is x, the payoff for a high-income household to report truthfully vs. misreporting converges to the value in (10). If (10) is nonnegative independently of x, we can set δ_N such that, for N large enough, high-income households strictly prefer reporting truthfully, independently of x. Hence, $(\bar{\tau}^N, \underline{\tau}_0^N, \underline{\tau}_1^N)$ is a sequence of mechanisms that achieves the maximin (for N large enough) and converges to the expected value of the best incentive-compatible allocation, which proves the claim.

We prove necessity by contradiction. The proof we include here assumes that $\bar{p} < 1$ and that f is bounded away form 0 on $[\underline{p}, \bar{p}]$. However, the proof could be easily adjusted to deal with the alternative cases, by excluding arbitrarily small intervals around the extrema and around points at which $f = 0$. The theorem would require adaptation only in the case in which the support contains a "hole," i.e., f is 0 on an interval inside $[\underline{p}, \bar{p}]$; in that case, the "test" plan should account for those additional degrees of freedom.

Suppose that (10) fails, and yet it is possible to find a sequence of mechanisms $\{(\bar{\tau}^N, \underline{\tau}_0^N, \underline{\tau}_1^N)\}_{N=1}^\infty$ for which truthtelling is a dominant strategy and whose equilibrium payoff converges to the payoff of the best equilibrium. Without loss of generality, we assume that these plans coincide with the "test" plan on $[0, \underline{p})$ and $(\bar{p}, 1]$. We also assume that they satisfy the budget constraint with equality; if they did not, it would be easy to construct a mechanism with higher utility and better incentives by decreasing the taxes levied on people reporting high income. The values the plans take on $[0, \underline{p})$ have no asymptotic effect on either the equilibrium payoff or any of the incentive constraints. The values the plans take on $(\bar{p}, 1]$ have no asymptotic effect on the equilibrium payoff, and the "test" plan relaxes as much as possible the incentive constraints when high-income households misreport their type with positive probability.

Let $\hat{x}$ be a value of x for which (10) fails. Consider the following functions:

$$\Psi(\varphi; N) := \int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{N} Q(n|p + (1-p)\hat{x}, N)(1-p) \cdot$$
$$\Bigg[u\left(\bar{e} - \varphi\bar{\tau}^N\left(\frac{n}{N+1}\right) - (1-\varphi)\bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right) -$$
$$\left(1 - \pi\left(\frac{n+1}{N+1}\right)\right)\left(\varphi y^N\left(\frac{n+1}{N+1}\right) + (1-\varphi)y^{\text{test}}\left(\frac{n+1}{N+1}\right)\right)\Bigg]$$

and

$$\Omega(\varphi; N) := \int_{\underline{p}}^{\bar{p}} \sum_{n=0}^{N} Q(n|p, N) \cdot$$
$$\Bigg[(1-p)u\left(\bar{e} - \varphi\bar{\tau}^N\left(\frac{n}{N+1}\right) - (1-\varphi)\bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right) +$$
$$p\Bigg[\pi\left(\frac{n+1}{N+1}\right)u\left(\underline{e} - \varphi\underline{\tau}_1^N\left(\frac{n+1}{N+1}\right) - (1-\varphi)\underline{\tau}_1^{\text{test}}\left(\frac{n+1}{N+1}\right)\right) +$$
$$\left(1 - \pi\left(\frac{n+1}{N+1}\right)\right)u\left(\underline{e} - \bar{e} + u^{-1}\left(\varphi y^N\left(\frac{n+1}{N+1}\right) + (1-\varphi)y^{\text{test}}\left(\frac{n+1}{N+1}\right)\right)\right)\Bigg]\Bigg] f(p)dp$$

where $y^N(m) := u(\bar{e} - \underline{e} + \underline{\tau}_0(m))$, and similarly for y^{test}.

It is straightforward to check that Ψ is weakly concave in φ. Furthermore, $\Psi(1; N) \geq 0$, while $\Psi(0;N)$ converges to a negative number. Hence, $\Psi'(0;N)$ as asymptotically larger than a strictly positive value, which we define $\bar{\Psi}$. Hence, we have

(A17)

$$0 < \bar{\Psi} < \int_{\underline{p}}^{\bar{p}} \sum_{n=\lceil \underline{p}N \rceil}^{\lfloor \bar{p}N \rfloor} Q(n|p+(1-p)\hat{x}, N)(1-p) \cdot \left[\left(\bar{\tau}^N\left(\frac{n}{N+1}\right) - \bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right) u'\left(\bar{e} - \bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right) - \left(1-\pi\left(\frac{n+1}{N+1}\right)\right)\left(y^N\left(\frac{n+1}{N+1}\right) - y^{\text{test}}\left(\frac{n+1}{N+1}\right)\right)\right] f(p)dp \leq \frac{\sup_{m\in[\underline{p}+(1-\underline{p})\hat{x},\bar{p}]} f(m)}{\inf_{m\in[\underline{p},\bar{p}]} f(m)} \cdot \left[\sup_{m\in[\underline{p},\bar{p}]} u'\left(\bar{e} - \bar{\tau}^{\text{test}}(m)\right) \sum_{n=\lceil \underline{p}N \rceil}^{\lfloor \bar{p}N \rfloor} \int_{\underline{p}}^{\bar{p}} Q(n|p,N)(1-p)dp \cdot \left|\bar{\tau}^N\left(\frac{n}{N+1}\right) - \bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right| + \frac{1-\underline{p}}{\underline{p}} \sum_{n=\lceil \underline{p}N \rceil}^{\lfloor \bar{p}N \rfloor} \int_{\underline{p}}^{\bar{p}} Q(n|p,N)pf(p)dp \left(1-\pi\left(\frac{n+1}{N+1}\right)\right)\left|y^N\left(\frac{n+1}{N+1}\right) - y^{\text{test}}\left(\frac{n+1}{N+1}\right)\right|\right]$$

Since $\bar{e} - \bar{\tau}^{\text{test}}(m) > \underline{e}$, u' is bounded.

Next, we compute $\Omega''(\varphi;N)$ for $\varphi \in [0,1]$:

$$-\Omega''(\varphi;N) = -\int_{\underline{p}}^{\bar{p}} \sum_{n=\lceil \underline{p}N \rceil}^{\lfloor \bar{p}N \rfloor} Q(n|p,N)\left[(1-p)\left(\bar{\tau}^N\left(\frac{n}{N+1}\right) - \bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right)^2 \cdot u''\left(\bar{e} - \varphi\bar{\tau}^N\left(\frac{n}{N+1}\right) - (1-\varphi)\bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right) + \left[\pi\left(\frac{n+1}{N+1}\right)\left(\underline{\tau}_1^N\left(\frac{n+1}{N+1}\right) - \underline{\tau}_1^{\text{test}}\left(\frac{n+1}{N+1}\right)\right)^2 \cdot u''\left(\underline{e} - \varphi\underline{\tau}_1^N\left(\frac{n+1}{N+1}\right) - (1-\varphi)\underline{\tau}_1^{\text{test}}\left(\frac{n}{N+1}\right)\right) + \left(1-\pi\left(\frac{n+1}{N+1}\right)\right)\left(y^N\left(\frac{n+1}{N+1}\right) - y^{\text{test}}\left(\frac{n+1}{N+1}\right)\right)^2 \cdot z\left(\varphi y^N\left(\frac{n+1}{N+1}\right) + (1-\varphi)y^{\text{test}}\left(\frac{n+1}{N+1}\right)\right)\right]\right], \tag{A18}$$

where

$$z(y) := \frac{u'(\underline{e} - \bar{e} + u^{-1}(y))}{[u'(u^{-1}(y))]^2} \left[\frac{u''(\underline{e} - \bar{e} + u^{-1}(y))}{u'(\underline{e} - \bar{e} + u^{-1}(y))} - \frac{u''(u^{-1}(y))}{u'(u^{-1}(y))} \right] \leq 0$$

First, consider the case in which

$$\text{(A19)}\quad \liminf_{N\to\infty} \sum_{n=\lceil \underline{p}N \rceil}^{\lfloor \bar{p}N \rfloor} \int_{\underline{p}}^{\bar{p}} Q(n|p,N)(1-p)f(p)dp \left| \bar{\tau}^N\left(\frac{n}{N+1}\right) - \bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right) \right| > 0$$

Then we have

$$\begin{aligned}
-\Omega''(\varphi;N) \geq & \inf_{\underline{p}\leq n/N \leq \bar{p}, \phi\in[0,1]} \left(-u''\left(\bar{e} - \varphi\bar{\tau}^N\left(\frac{n}{N+1}\right) - (1-\varphi)\bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right)\right) \cdot \\
& \sum_{n=\lceil \underline{p}N \rceil}^{\lfloor \bar{p}N \rfloor} \int_{\underline{p}}^{\bar{p}} Q(n|p,N)(1-p)dp \left(\bar{\tau}^N\left(\frac{n}{N+1}\right) - \bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right)^2 \geq \\
& \inf_{\underline{p}\leq n/N \leq \bar{p}, \phi\in[0,1]} \left(-u''\left(\bar{e} - \varphi\bar{\tau}^N\left(\frac{n}{N+1}\right) - (1-\varphi)\bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right)\right) \cdot \\
& \frac{\left[\sum_{n=\lceil \underline{p}N \rceil}^{\lfloor \bar{p}N \rfloor} \int_{\underline{p}}^{\bar{p}} Q(n|p,N)(1-p)dp \cdot \left| \bar{\tau}^N\left(\frac{n}{N+1}\right) - \bar{\tau}^{\text{test}}\left(\frac{n}{N+1}\right)\right| \right]^2}{\sum_{n=\lceil \underline{p}N \rceil}^{\lfloor \bar{p}N \rfloor} \int_{\underline{p}}^{\bar{p}} Q(n|p,N)(1-p)dp}.
\end{aligned}$$

The feasibility constraints imply that $\{(\bar{\tau}^N, \underline{\tau}_0^N, \underline{\tau}_1^N)\}_{N=1}^{\infty}$ is uniformly bounded on $[\underline{p}, \bar{p}]$, which in turn implies $\liminf_{N\to\infty} -\Omega''(\varphi;N) > 0$. However, we know $\lim_{N\to\infty} \Omega(0;N) = \lim_{N\to+\infty} \Omega(1;N)$; this implies that, for N large enough, both $\Omega(0;N)$ and $\Omega(1;N)$ are bounded away from $\max_{\varphi\in[0,1]} \Omega(\varphi;N)$. Since the constraint set is convex, this would imply that, for an infinite sequence of values of N, some intermediate value of φ would represent a feasible, incentive-compatible tax plan with higher utility than the best incentive-compatible allocation: a contradiction.

Suppose instead that the limit in (A19) is 0. Let $\{N_s\}_{s=0}^{\infty}$ be a sequence that makes (A19) converge to 0. Then the feasibility constraints and(A17) imply

$$0 < \liminf_{s\to\infty} \sum_{n=\lceil \underline{p} N_s \rceil}^{\lfloor \bar{p} N_s \rfloor} \int_{\underline{p}}^{\bar{p}} Q(n|p, N_s) p f(p) dp \left(1 - \pi\left(\frac{n+1}{N_s+1}\right)\right) \left| y^{N_s}\left(\frac{n+1}{N_s+1}\right) - y^{\text{test}}\left(\frac{n+1}{N_s+1}\right)\right| \leq$$

$$\liminf_{s\to\infty} \left\{ \sup_{\underline{p} \leq n/N_s \leq \bar{p}, \phi\in[0,1]} u'\left(\bar{e} - u\left(\varphi y^{N_s}\left(\frac{n+1}{N_s+1}\right) + (1-\varphi) y^{\text{test}}\left(\frac{n+1}{N_s+1}\right)\right)\right) \cdot \right.$$

$$\left. \sum_{n=\lceil \underline{p} N_s \rceil}^{\lfloor \bar{p} N_s \rfloor} \int_{\underline{p}}^{\bar{p}} Q(n|p, N_s) p f(p) dp \pi\left(\frac{n+1}{N_s+1}\right) \left| \underline{\tau}_1^{N_s}\left(\frac{n+1}{N_s+1}\right) - \underline{\tau}_1^{\text{test}}\left(\frac{n+1}{N_s+1}\right)\right| \right\}$$

Substituting this into (A18), we again obtain $\liminf_{N\to\infty} -\Omega''(\varphi; N) > 0$, which, as before, leads to a contradiction. ■

References

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

[2] Philip Bond and Kathleen Hagerty. Preventing Crime Waves. Mimeo, University of Pennsylvania, 2005.

[3] 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.

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

[5] Huberto M. Ennis and Todd Keister. Optimal Fiscal Policy under Multiple Equilibria. *Journal of Monetary Economics*. forthcoming.

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

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

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

[9] Matthew O. Jackson. Bayesian Implementation. *Econometrica*, 59(2):461–477, 1991.

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

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

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

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

[14] Robert Moffitt. Unemployment Insurance and the Distribution of Unemployment Spells. *Journal of Econometrics*, 28(1):85–101, 1985.

[15] 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.

[16] 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.

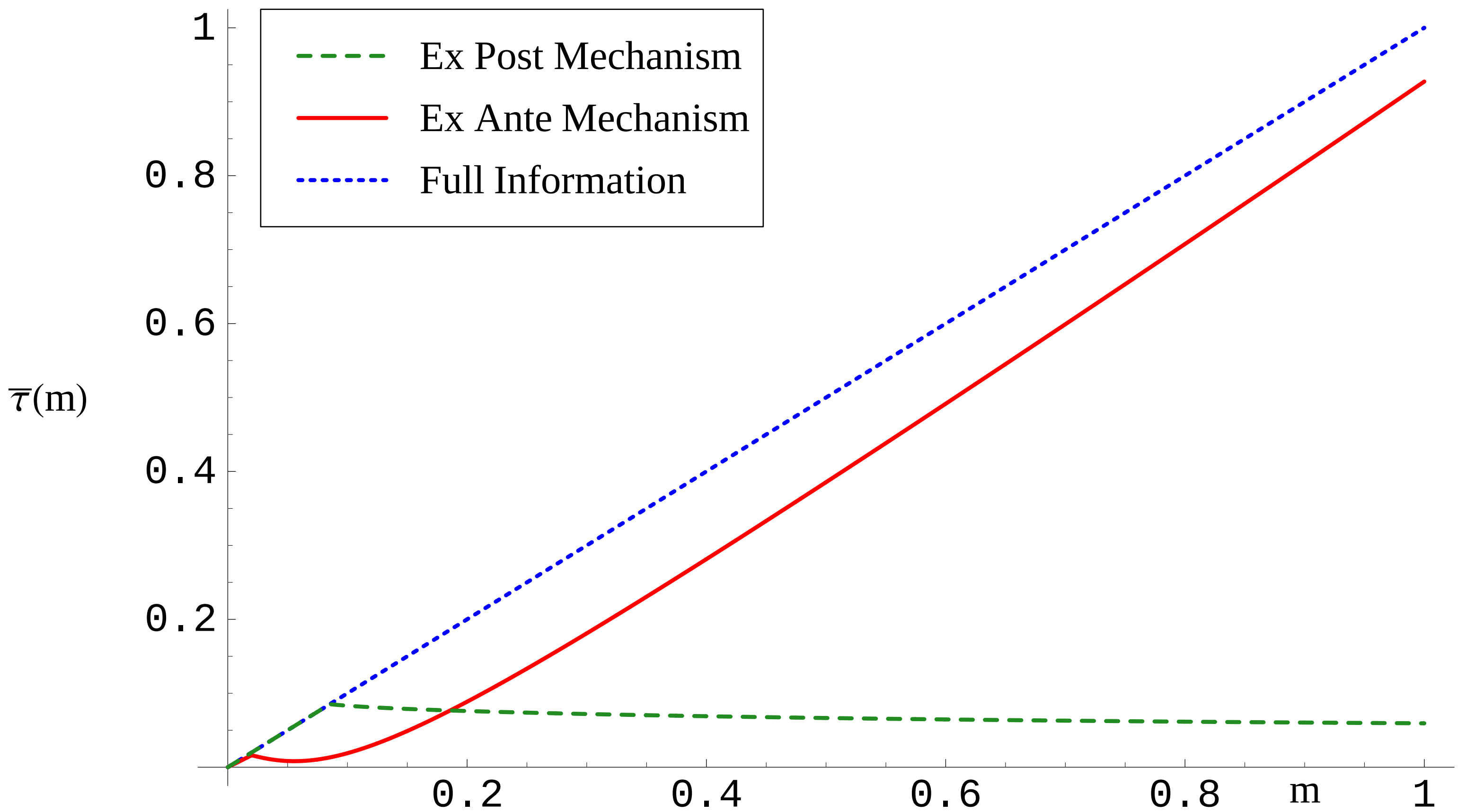

Figure 1: Taxes on rich, $\overline{\tau}(m)$

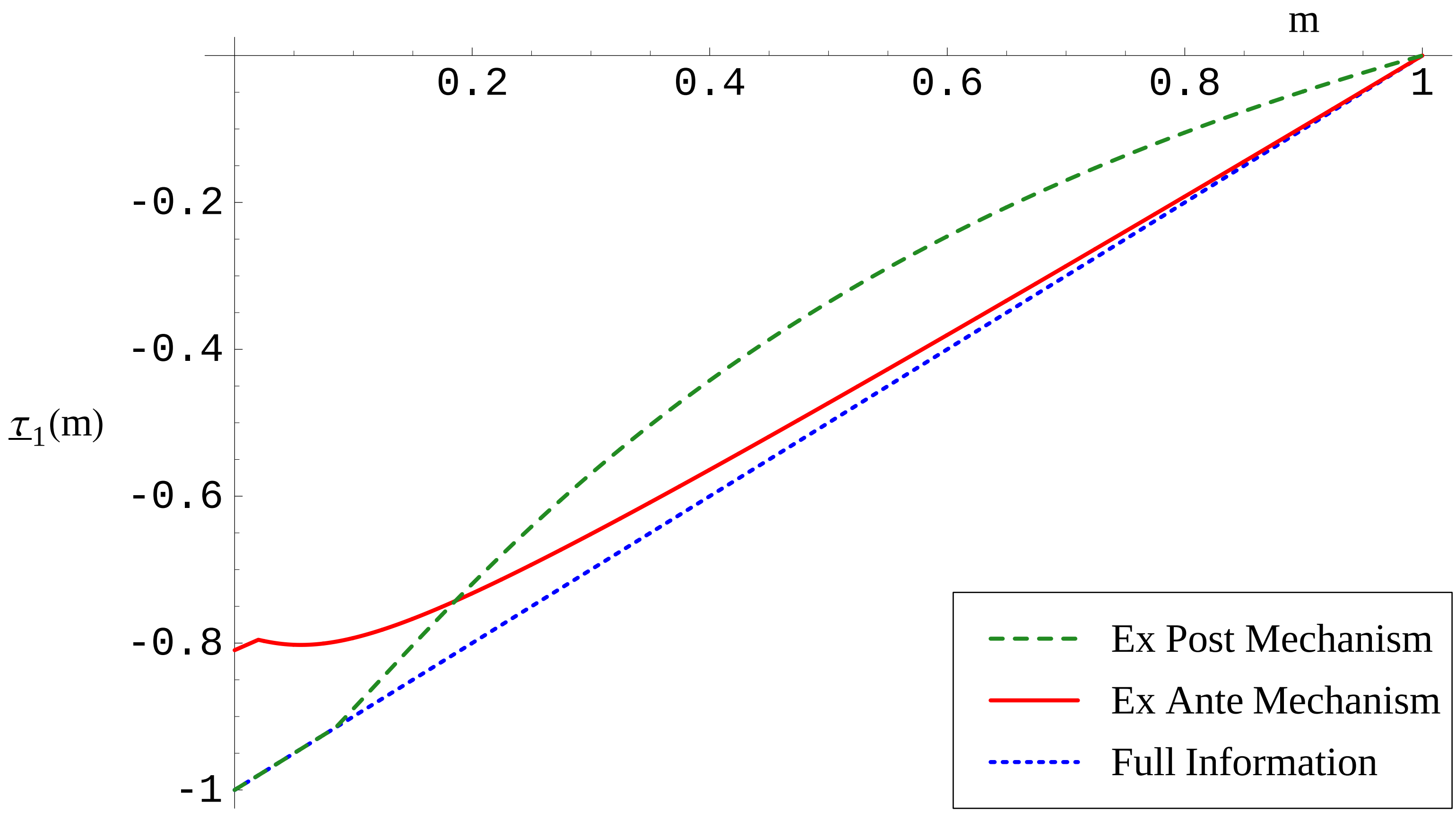

Figure 2: Taxes on audited poor, $\underline{\tau}_1(m)$

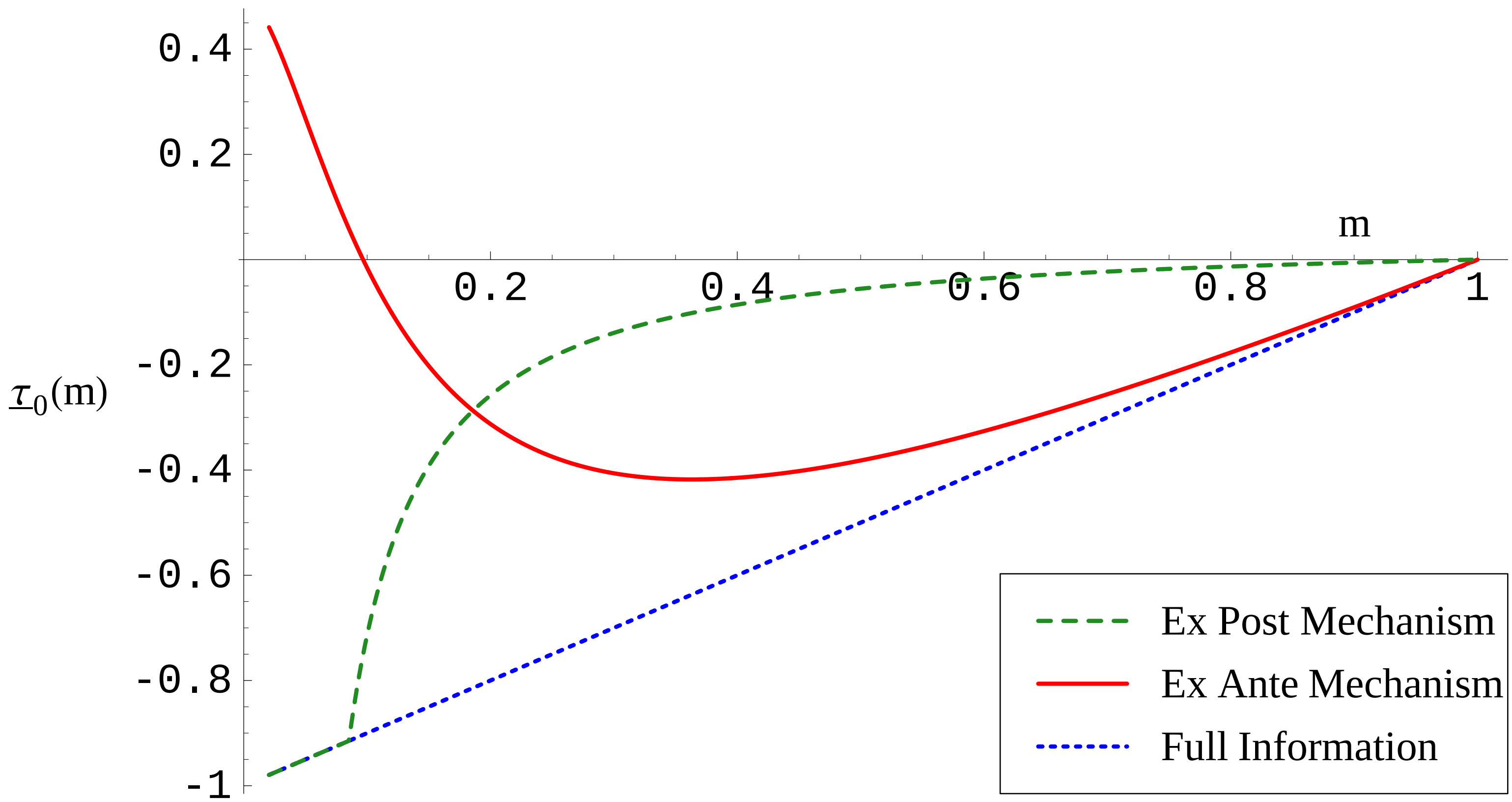

Figure 3: Taxes on non-audited poor, $\underline{\tau}_0(m)$

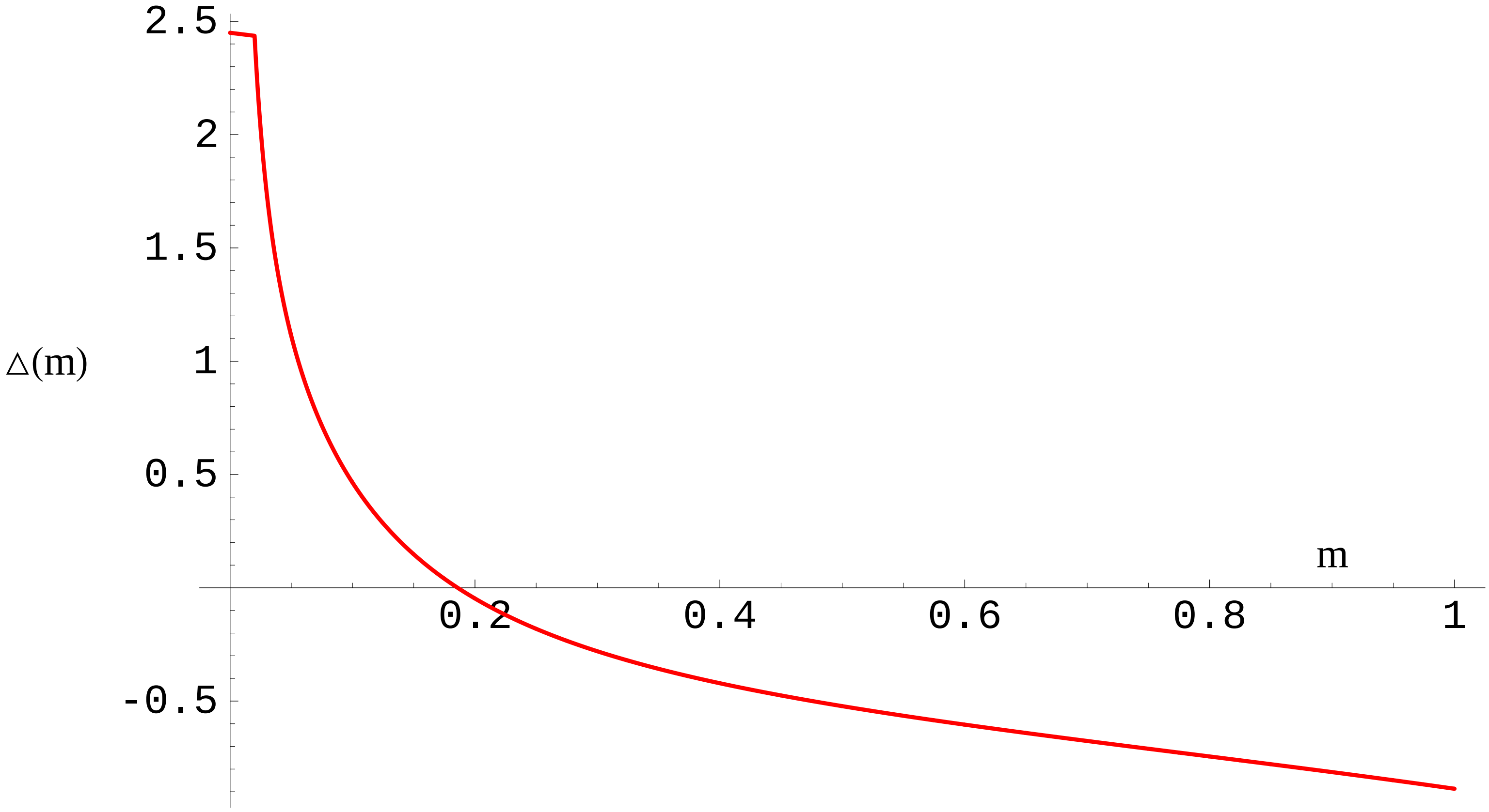

Figure 4: Ex post gain to truth-telling as function of aggregate reports, $\Delta(m)$