

State Verification and the Incentives to Save

Roc Armenter
F.R.B. Philadelphia

Thomas Mertens
Harvard University

February 2, 2009

INCOMPLETE DRAFT

Abstract

We consider a simple state verification technology in a dynamic private information economy. We find that the marginal benefit may be below the marginal cost of investment under the constrained efficient allocation. In particular, if the planner can verify a the type of a sufficiently large fraction of agents, then savings should be either subsidized or not taxed. In contrast, the prescriptions with respect to the labor supply are unchanged with the state verification technology.

1 Introduction

How should assets and capital be taxed? Economists have increasingly turned to dynamic economies with private information for an answer. Golosov, Kocherlakota, and Tsyvinski (2003) study a dynamic version of the framework famously introduced by Mirrlees (1971). The main finding is that the constrained efficient allocation features a wedge between the marginal benefit and the marginal cost of investing in capital. Thus efficiency is consistent with a positive tax on capital income in contrast to the results in Judd (1985) and Chamley (1986) for models with linear factor tax rates. Subsequent research has further explored the Mirrlees approach to dynamic optimal taxation and disability insurance.¹

¹A necessarily incomplete list in a rapidly increasing literature is Albanesi and Sleet (2006), Kocherlakota (2005), and Golosov and Tsyvinski (2006).

The standard Mirrlees framework makes stark assumptions on the information structure of the economy. Most models assume that the agents differ in their skill level. Neither the skill level nor the effort is directly observable, and thus incentives must be provided for an agent to truthfully report her type. However, it is reasonable to think that, at some cost, an agent's effort can be verified, or an independent observation on the skill level can be obtained. If the Mirrlees approach is to influence the design of tax systems, it is important to incorporate richer models of information. After all, tax audits are as old as taxes are.

In this paper we take a first step into a more realistic information framework. We consider a fairly simple audit technology: in each period the planner is able to verify the skill level of a fraction of agents. For the remaining agents, the skill level remains private information. The planner dictates the consumption and output of an agent as a function of the reported type and, if the skill type of the agent is verified, on whether the report was truthful or not. It is thus possible to “punish” an agent caught misreporting her type. Not surprising, the possibility of an audit improves the provision of incentives. If the skill level is verifiable for a large fraction of agents, or the planner can administer a very stern punishment, it is then possible to implement the first best allocations.

Our main finding is that the marginal benefit may be below the marginal cost of investment under the constrained efficient allocation. In other words, it may be optimal to subsidize assets rather than tax them. Whether a subsidy or a tax is optimal depends on the audit technology. If the fraction of agents whose type is verifiable is high enough, then savings should be either subsidized or not taxed. In contrast, the prescriptions of the Mirrlees approach with respect to the labor supply are unchanged with the auditing technology: we find that the labor decision of the agent with top skill level is left undistorted.²

Before we discuss our results further, let us first revisit briefly why the dynamic Mirrlees problem calls for a positive capital tax. The incentive-compatibility constraints require some spread between the consumption of high and low skill level agents. The more resources are available in the economy, the larger the spread must be. In other words, savings make the provision of incentives more difficult. As a result, consumption and welfare

²This result mirrors the findings of Cremer and Gahvari (1995) for static private information economies.

are frontloaded under the constrained efficient allocation.

Observing ex-post the agent's type relaxes incentive-compatibility constraints as it allows to prescribe a punishment for any agent caught misreporting the skill level. Since punishment is never administered on the equilibrium path, the planner equates the welfare of any misreporting agent to the lowest possible welfare level. Everything else constant, the more the welfare promised for truthful reporting, the sterner the punishment for misreporting, and thus the more effective auditing is in relaxing the incentive-compatibility constraints. Hence, additional resources, and thus savings, can improve the provision of incentives by raising the continuation welfare.

In short, savings can help the provision of incentives in the presence of audits as the additional resources increase the welfare loss in case of misreporting. Whether savings are taxed or subsidized depends, ultimately, in which effect dominates. It is thus not surprising that the effect associated with auditing dominates when it is easier to verify the agent's type.³

By taking the audit technology as given, we may be missing some key features of a properly designed audit mechanism.⁴ For example, we have assumed that the probability of verifying an agent type is independent of the message. The planner thus cannot target a particular set of agents for audits. It turns out that our assumption is without loss of generality. There is no need and no gain in verifying the type of an agent reporting to be of high skill: the low-skill agents have no incentives to misreport.

In the next section we illustrate the key results of the paper in a simple two period economy.

2 A Simple Two-Period Economy

The economy lasts for two periods, $t = 1, 2$ and is populated by a continuum of agents. All agents are identical at date $t = 1$. The only decision at $t = 1$ is whether to save or consume the endowment y_1 ,

$$c_1 \leq y_1 - s.$$

In the second period agents learn their skill level. Half of the agents are

³Note that the audit technology allows to frontload distortions in a private information economy. Albanesi and Armenter (2008) have argued that such ability is the key determinant in the absence of intertemporal distortions in the constrained efficient allocation.

⁴See Border and Sobel (1987) for optimal audit mechanism design.

able to produce one unit of output with $1/\theta$ units of effort, with $\theta > 1$. The other half produces only one unit of output for each unit of effort. The skill level of an agent is her type.

The resource constraint at date $t = 2$ is

$$\frac{1}{2}(c_2(\theta) - y_2(\theta)) + \frac{1}{2}(c_2(1) - y_2(1)) \leq s.$$

where $\{c_2(\theta), y_2(\theta)\}$ and $\{c_2(1), y_2(1)\}$ denote the consumption and output produced at date $t = 2$ for agents with skill level θ and 1 respectively. For simplicity, we assume that the return to savings is 1. A feasible allocation x is a vector $\{c_1, c_2(\theta), y_2(\theta), c_2(1), y_2(1), s\}$ that satisfies the resource constraint at both dates.

Agents value an allocation x according to standard separable preferences,

$$U(x) = u(c_1) + \frac{1}{2}\{u(c_2) - v(y_2/\theta)\} + \frac{1}{2}\{u(c_2) - v(y_2)\},$$

where u and v satisfy the usual properties. Note that the discount rate is set to unity. We also assume there is a lower bound on welfare, v , associated with feasible allocations.

As in the classic Mirrlees setting, the skill level is private information. For an allocation to be incentive compatible, the agents must thus be compelled to reveal their type. The key innovation of the paper is to introduce the possibility that the skill level is verifiable. We have to modify the standard incentive-compatibility constraints to reflect the possibility that an agent may be caught misreporting their type.

Let us be more precise about how the information is revealed and communicated between agents and the “planner.” At the beginning of the period $t = 2$, agents learn their type, θ or 1. They then report a type $m \in \{1, \theta\}$ to the planner. After the message has been received, the planner gets to verify the skill level of a fraction p of agents. For the remaining agents the skill level remains private information. We assume that the probability an agent’s type is observed is independent of the type or message. Finally the planner assigns a level of consumption and output to each agent depending on the message and, if the skill type of the agent is verified, the true type.

It is important that information is revealed sequentially. Agents have to report their type after learning their skill level but before they know whether they will be audited or not. Clearly this timing improves the incentives

compared to the case without any auditing technology as it becomes possible to punish an agent that misreports. However, private information still constraints the set of feasible allocations that can be implemented.

2.1 Constrained Efficient Allocation

It is easy to see that the revelation principle still applies in our economy, and agents report their true type. In the off-equilibrium event that an agent is caught misreporting, the maximum punishment is applied, which is given by the lower bound on welfare v . The incentive compatibility constraints are then

$$u(c_2(\theta)) - v(y_2(\theta)/\theta) \geq pv + (1-p)(u(c_2(1)) - v(y_2(1)/\theta)), \quad (1)$$

$$u(c_2(1)) - v(y_2(1)) \geq pv + (1-p)(u(c_2(\theta)) - v(y_2(\theta))). \quad (2)$$

Clearly the probability of observing the true type, p , and the severity of the punishment, v , relax the incentive compatibility constraints. If $p = 0$, we recover the standard formulation. Also for v low enough or p sufficiently close to 1, the first best allocation becomes incentive compatible.

We are interested in the constrained efficient allocation, that is, the incentive-compatible allocation attaining the highest welfare. Our main finding concerns the allocation of resources across periods: an audit technology can lead to savings being *subsidized* rather than taxed. The latter is always the case in dynamic Mirrlees economies without auditing.⁵ We find that if a fraction of agents larger than $\bar{p}$ are audited, then savings are subsidized under the constrained efficient allocation. The threshold level $\bar{p}$ can be quite small. In particular, it must be emphasized that the first-best allocation is not incentive compatible when $p = \bar{p}$.

The result on savings contrasts with the findings on the allocation between consumption and labor. In there the prescriptions of the standard Mirrlees problem are robust to the audit technology: the labor decision of the agent with top skill level is left undistorted, and the labor supply of the low skill level agent is taxed. The only exception is when the fraction of agents audited is high enough that the first best allocation is incentive compatible.

⁵See Golosov, Kocherlakota and Tsyvinski (2003).

2.2 Main Findings

We illustrate our main findings by deriving the necessary first order conditions associated with the constrained efficient allocations. As in the standard Mirrlees problem, the incentive constraint for the agent with a low skill level (2) is not binding. In other words, only agents of type θ have incentives to misreport.

The necessary first order conditions associated with $c_2(1)$ and $c_2(\theta)$ are

$$\begin{aligned} u_2^c(\theta)(1 + \mu) &= \lambda_2, \\ u_2^c(1)(1 - (1 - p)\mu) &= \lambda_2 \end{aligned}$$

where μ and λ_2 are the Lagrangian multipliers associated with the incentive compatibility constraint (1) and the resource constraint at date $t = 2$ respectively.⁶ By combining both we have that

$$u_2^c(\theta) = \frac{1 - (1 - p)\mu}{1 + \mu} u_1^c(\theta).$$

As long as the first best allocations are not incentive compatible, that is, $\mu > 0$, we have that the marginal utilities of consumption are not equated. More precisely, the agent type θ has a higher level of consumption, as in the standard Mirrlees problem.

Let us now turn to the question of whether savings will be taxed or subsidized from the point of view of the agents. By simple manipulation of the f.o.n.c. derived above we have that

$$.5 \frac{1}{u_2^c(\theta)} + .5 \frac{1}{u_2^c(1)} = \frac{1}{\lambda_2} \left(1 + \frac{p\mu}{2}\right).$$

From the f.o.n.c. for savings and consumption at date $t = 1$ is immediate that

$$u_1^c = \lambda_1 = \lambda_2.$$

We have thus that

$$.5 \frac{1}{u_2^c(\theta)} + .5 \frac{1}{u_2^c(1)} = \frac{1}{u_1^c} \left(1 + \frac{p\mu}{2}\right). \quad (3)$$

⁶We have normalized the Lagrangian multiplier μ by the measure of agents of type θ . As a result we have that $0 \geq \mu < 1$.

When $p = 0$, equation (3) is the well-known *inverse Euler equation*,

$$u_1^c = \left[E \left\{ \frac{1}{u_2^c} \right\} \right]^{-1}.$$

As shown above, as long as the first best allocations are not incentive compatible, $u_2^c(\theta) \neq u_2^c(1)$ and by Jensen's inequality we have that

$$u_1^c < E \{u_2^c\}.$$

Thus agents' savings must be taxed in order to implement the constrained efficient allocations. This result carries on for infinite horizon economies with arbitrary stochastic processes for the agents' type.⁷

Now consider the possibility of an audit, $p > 0$. We have then that

$$u_1^c = \left(1 + \frac{p\mu}{2} \right) \left[E \left\{ \frac{1}{u_2^c} \right\} \right]^{-1}$$

and thus

$$u_1^c < \left(1 + \frac{p\mu}{2} \right) E \{u_2^c\}$$

again by use of Jensen's inequality. As long as $p\mu > 0$, we cannot establish that $u_1^c < E \{u_2^c\}$.

Fortunately, it is still possible to sign $u_1^c - E \{u_2^c\}$. Using the f.o.n.c. for $c_2(\theta)$ and $c_2(1)$ is possible to solve for $E \{u_2^c\}$ in terms of μ and p ,

$$E \{u_2^c\} = \frac{1}{2} \lambda_2 \left[\frac{1}{1 + \mu} + \frac{1}{1 - (1 - p)\mu} \right]$$

and since $\lambda_2 = u_1^c$, we have that $u_1^c \geq E \{u_2^c\}$ whenever

$$\frac{1}{1 + \mu} + \frac{1}{1 - (1 - p)\mu} \leq 2$$

or

$$\mu \leq \frac{1}{2} \frac{p}{1 - p}. \quad (4)$$

Moreover, $u_1^c = E \{u_2^c\}$ if and only if $\mu = 0$ or (4) holds with strict equality. Under specific choices for preferences, we can write condition (4) in terms of

⁷See Kocherlakota (2005).

primitives. Of the parameters in our economy we are interested in the audit technology in particular. It is possible to express the condition (4) in terms of a threshold $\bar{p}$ for the audit technology. Heuristically, the Lagrangian multiplier μ is decreasing in p as the incentive-compatibility constraints become relaxed; and the right hand side of (4) is increasing in p .

We find thus that savings are subsidized if the audit technology is good enough, that is, the fraction of agents whose skill level becomes observable is larger than some threshold, $p \geq \bar{p}$. The exact value of the threshold depends on the primitives of the economy—the curvature of the preferences, the dispersion in skills, as well as the lower bound on welfare. It must be emphasized, though, that the first best allocations are not incentive compatible at $p = \bar{p}$. We show below that the intratemporal distortions remain for some $p \geq \bar{p}$. Indeed the main results of the classic Mirrlees problem on the allocation of consumption and labor across agents remain intact—a remarkable contrast with the prescriptions on the intertemporal margin.

Before we discuss our results further, let us first revisit briefly why the dynamic Mirrlees problem calls for savings to be taxed. The incentive-compatibility constraints require some spread between the consumption of high and low skill level agents. The more resources are available in the economy, the larger the spread must be. In other words, savings make the provision of incentives more difficult. As a result, consumption and welfare are frontloaded under the constrained efficient allocation.

The possibility of observing ex-post the agent's type relaxes incentive compatibility constraints as it allows to prescribe a punishment for any agent caught misreporting the skill level. Since punishment is never administered on the equilibrium path, the planner equates the welfare of any misreporting agent to the lower bound on welfare. Everything else constant, the more the welfare promised for truthful reporting, the sterner the punishment for misreporting, and thus the more effective auditing is in relaxing the incentive-compatibility constraints. Hence, additional resources, and thus savings, can improve the provision of incentives by raising overall welfare.

In short, savings can help the provision of incentives in the presence of audits as the saved resources increase the welfare loss in case of misreporting. Whether savings are taxed or subsidized depends, ultimately, in which effect dominates. It is thus not surprising that the effect associated with auditing dominates when it is easier to observe the agent's type.

While auditing can overturn the implications of the standard dynamic Mirrlees model for savings, the prescriptions for the taxation of labor are remarkably robust. For any auditing technology, the marginal rate of substitution is equated to the marginal rate of transformation for the agent with the top skill level. As long as the first-best allocation is not incentive compatible, the agent with the low skill level has her labor supply taxed positively.

A simple look at the necessary first order conditions for $y_2(\theta)$ and $y_2(1)$ suffices. They are

$$\begin{aligned} v'(y_2(\theta)/\theta)(1+\mu) &= \theta\lambda_2, \\ v'(y_2(1)) \left(1 - (1-p)\mu \frac{v'(y_2(1)/\theta)}{\theta v'(y_2(1))}\right) &= \lambda_2. \end{aligned}$$

Recalling the f.o.n.c. associated with consumption, we have thus that,

$$u_2^c(\theta)\theta = v'(y_2(\theta)/\theta),$$

so the allocation between consumption and output for the type θ agent is left undistorted. In contrast, we have that

$$u_2^c(1) > v'(y_2(1))$$

for the agent of type 1.⁸ Hence the allocation between consumption and output for the type 1 agents is distorted as long as $\mu > 0$ and $p < 1$.

Finally, we discuss briefly our modeling of the audit technology. The reader may be worried that a properly designed audit mechanism would be very different from our setting. We argue that our setting captures the key features of an optimal audit mechanism, at least in a simple economy as the one above.

First, we assumed that the probability of observing an agent type is independent of the message. Thus the planner cannot target a particular set of agents for audits, which tax agencies around the world spend considerable resources doing. It turns out that our assumption is without loss of generality. Only agents with high skill level have an incentive to misreport.⁹ There is no need thus to audit agents that declare themselves to have a high skill

⁸By the convexity of v , $\frac{v'(y_2(1)/\theta)}{\theta v'(y_2(1))}$ is less than one.

⁹In other words, only the incentive-compatibility constraint (1) is binding.

level. So we could have allow for a message-specific audit probability, and our results would stay with $p(\theta) = p$ and for any value of $p(1)$.

Second, we have abstracted from the costs associated with an audit. We can include some resource cost associated with auditing, and then derive the fraction of audited agents p as an optimal choice. For our purposes, though, it is more natural to ask what resource cost would sustain the choice of a given fraction p . The comparative statics with respect to p imply that the marginal cost of increasing the fraction of agents whose type is verified is

$$\xi = \frac{\mu}{p\lambda} (u(c_2(1)) - v(y_2(1)) - v).$$

As long as the punishment is effective, namely, there is some welfare forgone if an agent is caught misreporting, and the incentive compatibility is binding, $\mu > 0$, there is some positive marginal cost that sustains the choice of p . Behind this exercise, there is, though, one important assumption on transferable utility. If the costs of an audit are borne individually and the planner cannot compensate an agent for it, then the audit mechanism becomes much more complex.

References

- [1] Albanesi, Stefania and Roc Armenter. 2008. Intertemporal Distortions in the Second Best. *Working Paper*, Columbia University.
- [2] Albanesi, Stefania and Christopher Sleet. 2006. Dynamic Optimal Taxation with Private Information. *The Review of Economic Studies*, 73: 1-30.
- [3] Border, Kim C. and Joel Sobel. 1987. Samurai Accountant: A Theory of Auditing and Plunder. *The Review of Economic Studies*, 54(4): 525-540.
- [4] Chamley, Christophe, 1986. Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica* 54 (3): 607-622.
- [5] Cremer, Helmuth and Firouz Gahvari. 1995. Tax evasion and the optimum general income tax. *Journal of Public Economics*, 60:235-249.
- [6] Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski. 2003. Optimal Indirect and Capital Taxation. *The Review of Economic Studies*, Volume 70, Issue 3, Page 569.

- [7] Golosov, Mikhail, and Aleh Tsyvinski. 2006. Designing Optimal Disability Insurance: A Case for Asset Testing. *Journal of Political Economy*.
- [8] Judd, Kenneth. 1985. Redistributive Taxation in a Perfect Foresight Model. *Journal of Public Economics* 28, 59-84.
- [9] Kocherlakota, Narayana. 2005b. Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation. *Econometrica*.
- [10] Mirrlees, James. 1971. An exploration in the theory of optimum income taxation. *The Review of Economic Studies* 38: 175-208.
- [11] Rogerson, William. 1985. Repeated moral hazard. *Econometrica* 53:69-76.