

# Dynamic Mirrlees Taxation and Political Economy\*

Daron Acemoglu  
MIT

Michael Golosov  
MIT

Aleh Tsyvinski  
Harvard

Current Version: July 2007.

## Abstract

We study the structure of nonlinear incentive-compatible taxes, as in Mirrlees's classic taxation environment, in a dynamic economy subject to political economy and commitment problems. In particular, in contrast to existing analyses of dynamic and/or nonlinear taxation problems, we relax the assumptions that taxes are set by a benevolent government (politician) and that there is commitment to policies. Instead, in our model economy taxes are set by a self-interested politician, without any commitment power. This politician is partly controlled by the citizens via elections. The resulting environment is one of a dynamic mechanism design without commitment. We focus on the best sustainable mechanism, which is the mechanism that maximizes the ex ante utility of the citizens. Even though that this class of mechanism design problems is complex, we show that our model enables a tractable analysis and relatively tight characterization results. Towards a full characterization of the best sustainable mechanism, we first prove that a version of the revelation principle applies in our environment and that attention can be restricted to direct truth-telling mechanisms. Using this result, we prove that the provision of incentives to politicians can be separated from the provision of incentives to individuals and from redistribution across individuals in the economy. This also enables us to model the political economy and commitment constraints as additional aggregate constraints. Using this formulation, we provide conditions under which distortions created by political economy and commitment problems persist or disappear in the long run. In particular, if politicians are as patient as (or more patient than) the citizens, the additional distortions introduced by political economy and commitment problems disappear asymptotically. Finally, we extend our analysis to the case where the government cares both about its own consumption and the future utility of the citizens. This extension generalizes our results to environments where the key constraint is time-inconsistency of government policy.

**Keywords:** dynamic incentive problems, mechanism design, optimal taxation, political economy, revelation principle.

**JEL Classification:** H11, H21, E61, P16.

---

\*We thank audiences at many seminars for helpful comments and Oleg Itskhoki for excellent research assistance.

# 1 Introduction

The major insight of the optimal taxation literature pioneered by Mirrlees (1971) is that the tax structure ought to provide incentives to individuals to work, exert effort and invest, while also providing insurance. This insight is also central to the recent optimal dynamic taxation literature.<sup>1</sup> This literature characterizes the structure of optimal taxes assuming that policies are decided by a benevolent government with full commitment power. In practice, however, tax structures are designed by politicians, who care about reelection, self-enrichment or their own individual biases and cannot commit to future policies or to dynamic mechanisms.<sup>2</sup>

In this paper, we study the structure of dynamic nonlinear taxation under more realistic assumptions than the optimal taxation literature.<sup>3</sup> In particular, we assume that the government (politicians) are, at least in part, self-interested and that there is no exogenous commitment to policies. Although implications of the self-interested behavior of politicians for *feasible* optimal policies has not received much attention, the difficulties introduced by lack of commitment in the design of optimal policies are generally well-recognized. Chari and Kehoe (1990), for example, provide a comprehensive discussion of the implications of incorporating time-consistency constraints.

The challenges created by time inconsistency in dynamic nonlinear taxation environments was first pointed out by Roberts (1984). Roberts considered an example economy, where, similar to Mirrlees (1971), risk-averse individuals are subject to unobserved shocks affecting the marginal disutility of labor supply. But differently from the benchmark Mirrlees model, the economy is repeated  $T$  times, with individuals having perfectly persistent types. Under full commitment, a benevolent planner would choose the same allocation at every date, which coincides with the optimal solution of the static model. However, a benevolent government without full commitment cannot refrain from exploiting the information that it has collected at previous dates to achieve better risk sharing ex post. This turns the optimal taxation problem into a dynamic game between the government and the citizens. Roberts showed that as discounting disappears and  $T \rightarrow \infty$ , the unique sequential equilibrium of this game involves

---

<sup>1</sup>See, for example, Golosov, Kocherlakota, and Tsyvinski (2003), Werning (2002), Sleet and Yeltekin (2004), Kocherlakota (2005), Albanesi and Sleet (2005), Battaglini and Coate (2005), and Golosov, Tsyvinski and Werning (2007), among others.

<sup>2</sup>For general discussions of the implications of self-interested behavior of governments, petitions and bureaucrats, see, among others, Buchanan and Tullock (1962), North and Thomas (1973), North (1981), Olson (1982), North and Weingast (1989), and Dixit (2004). Austen-Smith and Banks (1999), Persson and Tabellini (2000) and Acemoglu (2005a) provide introductions to various aspects of the recent developments and the basic theory.

Another potential difficulty with centralized system is that they may involve excessive communication relative to trading systems. See Segal (2005) for a recent model developing this insight.

<sup>3</sup>Since in the literature following Mirrlees the optimal tax-transfer program is a solution to a mechanism design problem, we use the terms “optimal tax-transfer program” and “mechanism” interchangeably. The results in this paper can be generalized to other, non-tax, dynamic mechanism design problems.

the highly inefficient outcome in which all types declare to be the worst type at all dates, supply the lowest level of labor and receive the lowest level of consumption. This example not only shows the potential inefficiencies that can arise once we depart from the unrealistic case of full commitment, even with benevolent governments, but also highlights that the main tool of analysis in dynamic taxation problems, the celebrated revelation principle, also fails (in Roberts's economy there is no truthful reporting of types).<sup>4</sup>

In light of this stark difficulty highlighted by Roberts (1984), is there any hope of constructing equilibrium taxation policies in the presence of political economy and commitment constraints that can provide incentives to and redistribution (risk sharing) among agents in the economy as in Mirrlees's baseline analysis? The main contribution of the current paper is to show that, under suitable assumptions, reasonable equilibrium taxation policies can be supported as equilibria and that these equilibrium policies retain many of the insights of the purely normative Mirrleesian analysis. To present our main results in the clearest possible way, throughout the paper we focus on the equilibrium of the dynamic game between citizens and politicians that maximizes the ex ante utility of the citizens, and refer to this as the *best sustainable mechanism*, which emphasizes that this is the the best tax-transfer scheme that is sustainable in the sense of being incentive compatible both for the citizens and for the politicians entrusted with the policies.

Two ingredients are essential for our approach. First, instead of a finite horizon economy as in Roberts, we consider an infinite horizon environment. This makes it possible for us to use standard repeated game strategies to sustain better equilibria than those emphasized in Roberts. Second, we choose a particularly tractable model of political economy and commitment problems. In particular, we assume that politicians have no commitment power and can even deviate from their within-period commitments, but they are subject to electoral accountability. Both of these assumptions are similar to those made in the baseline models of political economy based on the approach first proposed by Barro (1973) and Ferejohn (1986). In the Barro-Ferejohn model, politicians can choose any policy vector they prefer, but if their policy choice is not in line with the electorate's expectations, then they can be voted out of office (see Persson and Tabellini, 2000, Chapter 4, for a modern exposition and references).

These two ingredients enable us to develop a tractable framework for analysis of dynamic taxation in the presence of political economy and commitment constraints. Our first result is that because of the potential electoral controls on politicians, a version of the revelation principle, the *truthful revelation along the equilibrium path*, applies in our environment regardless

---

<sup>4</sup>Freixas, Guesnerie and Tirole (1985) point out similar, though less extreme, problems as Roberts in a two-period regulation problem without commitment on the part of the regulator.

of the discount factors of various parties (that is, this is *not* and folk theorem type result).<sup>5</sup> It should be noted that this result does not generalize to other dynamic mechanism design problems and does exploit the specific structure of our economy.<sup>6</sup>

Our second result is key to a tractable analysis of dynamic nonlinear taxation problems in the presence of political economy and commitment constraints: we show that the best sustainable mechanism enables a separation between private and public incentives—that is, incentive compatibility for individuals can be treated separately from ensuring that the politician in power does not wish to deviate from the candidate tax-transfer scheme. More specifically, this separation result enables us to characterize the best sustainable mechanism in two steps:

1. We first solve the problem of providing incentives to individuals given aggregate levels of consumption and labor supply. We call this problem a *quasi-Mirrlees problem* as it is a usual dynamic Mirrlees problem with two additional constraints on aggregate labor and consumption. Its solution leads to an *indirect utility functional* representing expected utility as a function of the aggregate levels of consumption and labor supply.
2. We then provide incentives to politicians by choosing aggregate variables and the level of rents paid to the politician.

This formulation not only provides us with a tractable strategy for characterizing the best sustainable mechanism, but it also enables a direct comparison between the best sustainable mechanism and the full-commitment Mirrlees mechanism in terms of the *aggregate distortions* caused by the former relative to the full-commitment Mirrlees allocation. This result also shows that incorporating lack of commitment and self-interest of politicians does not invalidate the methodology of approaching dynamic taxation problems as one of dynamic mechanism design, but now with additional constraints.

Our third and main set of results concerns the characterization of these aggregate distortions. First, we show that political economy and commitment problems always introduce further distortions in the sustainable mechanism relative to the full-commitment Mirrlees mechanism. Intuitively, if the sustainability constraint of the politician were always slack, then the

---

<sup>5</sup>One can always construct an extended game in which there is a *fictional* disinterested mechanism designer, with the government as an additional player that has the authority to tax and regulate and the ability to observe all the communication between the fictitious mechanism designer and individual agents. Although a version of the revelation principle would apply in this extended game, this does not circumvent the substantive issues raised here: the party entrusted with taxes and transfers has neither the same interests as those of the citizens nor much commitment power.

<sup>6</sup>There has recently been important progress in the characterization of dynamic mechanisms without commitment in more general settings than the one we study here. See, in particular, the important work by Bester and Strausz (2001) and Skreta (2006). See also the interesting paper by Sleet and Yeltekin (2006), who also derive a version of the revelation principle in a dynamic taxation problem with time-inconsistency (though without any political economy).

politician would receive zero consumption and would find it beneficial to deviate and expropriate some of the output. If, on the other hand, the sustainability constraint binds, then any increase in output has to be associated with increased government consumption. This in turn increases the opportunity cost of increasing output and leads to a reduction in labor supply and capital accumulation. Second, we provide precise conditions under which these further distortions disappear or persist over time. In particular, we show that when politicians are as patient as (more patient than) the citizens, additional distortions created by political economy disappear in the long run and the allocation of resources converges to that of a dynamic Mirrlees economy with an exogenous level of public good spending. In this limiting equilibrium, there are no additional taxes on labor beyond those implied by the optimal Mirrleesian taxation and no aggregate taxes on capital.<sup>7</sup> In contrast, when politicians are (strictly) less patient than the citizens, the structure of taxes never converges to that of a dynamic Mirrlees economy and features additional labor and capital taxes even asymptotically. This last set of results is important, since it provides an exception to most existing models, which predict that long-run taxes on capital should be equal to zero (cfr. footnote 7).

Our final set of results focus on politicians that are at least partly benevolent. In this case, the main constraint on the form of taxation is commitment to the sequence of policies rather than the self-interested behavior of the politicians. This case is particularly important since it enables us to revisit Roberts's (1984) stark negative result. We show that in this case also aggregate distortions created by commitment problems disappear if the politician is as patient as the citizens, though now some additional technical assumptions need to be imposed.

Our results are related to a number of different literatures. The relationship to dynamic mechanism design problems without commitments, such as Bester and Strausz (2001) and Skreta (2006), has already been mentioned. In this respect, our paper is also related to the important paper by Bisin and Rampini (2005), who extend Roberts's analysis and show how the presence of anonymous markets acts as an additional constraint on the government, ameliorating the commitment problem. This lack of commitment is related to the lack of commitment by the self-interested government in our model. The most important distinction between the two approaches is that our model is infinite horizon. This enables us to construct sustainable mechanisms with the revelation principle holding along the equilibrium path, to analyze substantially more general environments, and to characterize the limiting behavior of distortions and taxes.

---

<sup>7</sup>This result is therefore similar to that of zero limiting taxes on capital in the first-generation Ramsey-type models, e.g., Chamley (1986) or Judd (1985), but is derived here without any exogenous restriction on tax instruments (see Kocherlakota, 2005, for the zero capital tax result using the second-generation approach).

It is important to emphasize, however, that this limiting allocation can be decentralized in different ways, and some of those may involve positive taxes on individual capital holdings.

Our work is also clearly related to the burgeoning literatures on dynamic political economy (see footnote 2) and on dynamic nonlinear taxation (e.g., the papers mentioned in footnote 1). In particular, our framework incorporates the general model of dynamic nonlinear taxation considered in Golosov, Kocherlakota, and Tsyvinski (2003).

The results in this paper are also closely related to our previous work, Acemoglu, Golosov and Tsyvinski (2006, 2007a,b). Many of the results here were first presented in the working paper, Acemoglu, Golosov and Tsyvinski (2006).<sup>8</sup> A special case of these results, which apply to a representative agent neoclassical growth model (and thus without any aspect of nonlinear taxation) have been developed in Acemoglu, Golosov and Tsyvinski (2007a). The results in this paper are therefore a significant generalization of those in Acemoglu, Golosov and Tsyvinski (2007a). In particular, the results related to truthful revelation, separation of private and public incentives, and the structure of nonlinear taxes are not present in that paper. The main parallel between the current paper and Acemoglu, Golosov and Tsyvinski (2007a) is the formulation of the political economy problem as an electoral accountability problem and the emphasis on the relative discount factors of politicians and citizens. It is also noteworthy that the provision of incentives to politicians in our model is also related to the optimal provision of incentives in dynamic principal-agent analyses (see, among others, Harris and Holmstrom, 1982, Lazear, 1981, Ray, 2002). Ray (2002) provides the most general results in this context. Acemoglu, Golosov and Tsyvinski (2007a) extend Ray’s results to the case in which discount factors are different between the principal and the agent (or the politician and the citizens).

The rest of the paper is organized as follows.

## 2 Model

We consider a general dynamic Mirrlees optimal taxation setup in an infinite horizon economy. There is a continuum of individuals and we denote the set of individuals, which has measure 1, by  $I$ . The instantaneous utility function of individual  $i \in I$  at time  $t$  is given by

$$u(c_t^i, l_t^i | \theta_t^i) \tag{1}$$

where  $c_t^i \geq 0$  is the consumption of this individual,  $l_t^i \geq 0$  is labor supply, and  $\theta_t^i$  is his “type”. This formulation is general enough to nest both preference shocks and productivity shocks.<sup>9</sup>

Let  $\Theta = \{\theta_0, \theta_1, \dots, \theta_N\}$  be a finite ordered set of potential types, with the convention that  $\theta_i$  corresponds to “higher skills” than  $\theta_{i-1}$ , and in particular,  $\theta_0$  is the worst type. Let  $\Theta^T$

---

<sup>8</sup>Results related to the comparison of market-based and government-controlled allocations in that working paper have been extended in Acemoglu, Golosov and Tsyvinski (2007b). None of these results are present in the current paper.

<sup>9</sup>In particular, productivity shocks would correspond to the case where  $u(c_t^i, l_t^i | \theta_t^i) = u(c_t^i, l_t^i/\theta_t^i)$ .

be the  $T$ -fold product of  $\Theta$ , representing the set of sequences of length  $T = 1, 2, \dots, \infty$ , with each element belonging to  $\Theta$ . We think of each agent's lifetime type sequence  $\theta^\infty$  as drawn from  $\Theta^\infty$  according to some measure  $\mu^\infty$ . Let  $\theta^{i,\infty}$  be the draw of individual  $i$  from  $\Theta^\infty$ . The  $t$ -th element of  $\theta^{i,\infty}$ ,  $\theta_t^i$ , is the skill level of this individual at time  $t$ . We use the standard notation  $\theta^{i,t}$  to denote the history of this individual's skill levels up to and including time  $t$ , and make the standard measurability assumption that the individual only knows  $\theta^{i,t}$  at time  $t$ . No other agent in the economy will directly observe this history. We assume that each individual's lifetime type sequence is drawn from  $\Theta^\infty$  according to the same measure  $\mu^\infty$  and independently from the draws of all other individuals, so that there is no aggregate uncertainty in the type distribution.<sup>10</sup> In addition, to simplify the notation, we also assume that within each period, there is an aggregate invariant distribution of types denoted by  $G$ .<sup>11</sup>

We assume that labor supply of an individual with skill  $\theta$  comes from a compact set, i.e.,  $l_t^i \in [0, \bar{l}(\theta)]$ .

**Assumption 1 (utility function)** For all  $\theta \in \Theta$ ,  $u(c, l | \theta) : \mathbb{R}_+ \times [0, \bar{l}(\theta)] \rightarrow \mathbb{R}$  is twice continuously differentiable and jointly concave in  $c$  and  $l$ , and is non-decreasing in  $c$  and non-increasing in  $l$ .

**Assumption 2 (single crossing)** Let the partial derivatives of  $u$  be denoted by  $u_c$  and  $u_l$ . Then  $u_c(c, l | \theta) / |u_l(c, l | \theta)|$  is increasing in  $\theta$  for all  $c$  and  $l$  and all  $\theta \in \Theta$ .

**Assumption 3 (worst type and full support)** We have  $\bar{l}(\theta_0) = 0$  and  $\bar{l}(\theta) = \bar{l} < \infty$  for all  $\theta \in \Theta$  and  $\theta \neq \theta_0$ . Moreover,  $\mu^\infty$  has full support in the sense that  $\theta_t^i = \theta_0$  has positive probability after any history.

The first two assumptions are standard. Assumption 3 states that for the worst type,  $\theta_0$ , supplying positive labor is impossible. This suggests that we can think of the worst type as “disabled”—unable to supply any labor at that date. It also requires  $\mu^\infty$  to have full support in the sense that any individual can become disabled at any point. This assumption simplifies the analysis of sustainable mechanisms by making it possible to have off-the-equilibrium path actions where all types supply zero labor. This assumption can be replaced, instead, by an alternative assumption that there is “freedom of labor supply”, meaning that it is individuals who decide how much labor to supply and they can always choose to supply zero labor. The advantage of Assumption 3 is that it leads to the freedom of labor supply as an equilibrium

<sup>10</sup>This structure imposes no restriction on the time-series properties of individual skills. Both identical independent draws and arbitrary temporal dependence are allowed. For concreteness, one may wish to think that  $\theta_t^i$  follows a Markov process.

<sup>11</sup>This assumption plays no role in our analysis except for reducing the notation.

outcome and also enables us to impose full compliance with mechanisms, which is a typical feature of the general mechanism design problems.

All individuals have the same discount factor  $\beta \in (0, 1)$ , thus at time  $t$ , they maximize

$$\mathbb{E} \left[ \sum_{s=0}^{\infty} \beta^s u(c_{t+s}^i, l_{t+s}^i \mid \theta_{t+s}^i) \mid \theta^{i,t} \right]$$

where  $\mathbb{E}[\cdot \mid \theta^{i,t}]$  denotes the expectations conditional on having observed the history  $\theta^{i,t}$ .

The production side of the economy is described by the aggregate production function

$$Y = F(K, L) \tag{2}$$

where  $K$  is capital and  $L$  is labor. We assume:

**Assumption 4 (production structure)** *F is strictly increasing and continuously differentiable in both of its arguments, with derivatives denoted by  $F_K$  and  $F_L$ , exhibits constant returns to scale and satisfies the Inada condition,  $\lim_{K \rightarrow \infty} F_K(K, L) = 0$  for all  $L \in \mathbb{R}_+$ . Moreover, capital fully depreciates after use, and  $F(K, 0) = 0$ .*

Both the full depreciation assumption and the assumption that labor is essential for production are adopted to simplify the notation. The Inada condition, together with the fact that the maximum amount of labor in the economy is bounded, implies that there is a maximum steady-state output  $\bar{Y} \in (0, \infty)$  that can be produced, given by  $\bar{Y} = F(\bar{Y}, \bar{L})$ , where  $\bar{L} \equiv \int \bar{l}(\theta) dG(\theta)$  is the maximum amount of total labor.

### 3 Political Economy

The allocation of resources in this economy is entrusted to a politician. The political economy side of our model is a classic electoral accountability setup of Barro (1973) and Ferejohn (1986). We assume that there is a large number of potential (and identical) politicians, denote the set of politicians by  $\mathcal{I}$ . The utility of a politician at time  $t$  is given by

$$\sum_{s=0}^{\infty} \delta^s v(x_{t+s}),$$

where  $x$  denotes the politician's consumption (rents),  $v : \mathbb{R}_+ \rightarrow \mathbb{R}$  is the politician's instantaneous utility function. Notice also that the politician's discount factor,  $\delta$ , is potentially different from that of the citizens,  $\beta$ . To simplify the analysis, we assume that potential politicians are distinct from the citizens and never engage in production, and that once they are replaced they do not have access to capital markets.<sup>12</sup>

---

<sup>12</sup>All of the results in this paper hold if a politician has access to capital markets after deviation, and only the right hand side of the sustainability constraints need to be modified.

**Assumption 5 (politician utility)**  $v$  is twice continuously differentiable, concave, and satisfies  $v'(x) > 0$  for all  $x \in \mathbb{R}_+$  and  $v(0) = 0$ . Moreover  $\delta \in (0, 1)$ .

Since the politician in power both lacks commitment power and has the ability to appropriate output for its own consumption, we model the interaction between the citizens and the politicians as a dynamic game following the literature on sustainable plans (Chari and Kehoe, 1990, 1993). Our purpose throughout is to characterize the equilibrium of this game between the politicians and the citizens, corresponding to the *best sustainable mechanism*, meaning the sustainable mechanism that maximizes the ex ante utility of citizens.<sup>13</sup>

We first describe the feasible actions by citizens and the politicians, and the timing of events. We then provide a formal definition of mechanisms

### 3.1 Timing and Actions in Period $t$

We define a *submechanism* (or mechanism at time  $t$ ) as a subcomponent of the overall mechanism between the politician and the individuals. A submechanism specifies what happens at a given date. In particular, let  $Z_t$  be a general message space for time  $t$ , with a generic element  $z_t$ . This message space may include messages about current type of the individual,  $\hat{\theta}_t^i \in \Theta$ , and past types  $\hat{\theta}^{i,t-1} \in \Theta^{t-1}$  (even though the individual may have made some different reports about his or her types in the past), and might also include other messages. Let  $Z^t \equiv \prod_{s=0}^t Z_s$  and  $z^t$  denote a generic element of  $Z^t$ .

A submechanism consists of two mappings, i.e.,  $M_t \equiv (\tilde{c}_t, \tilde{l}_t)$  such that  $\tilde{c}_t : Z^t \rightarrow \mathbb{R}_+$  assigns consumption levels for each complete history of messages and public histories, and  $\tilde{l}_t : Z^t \rightarrow [0, \bar{l}]$  assigns corresponding labor supply levels.<sup>14</sup> Given Assumption 3, any submechanism must allow for some messages which will lead to  $l = 0$ . We denote the set of submechanisms that satisfy this restriction and also the relevant resource constraints (which will be specified

<sup>13</sup>Since we are dealing with a dynamic game, our focus on the best sustainable mechanism is essentially a selection among the many equilibria. Alternatively, one can think of the “social plan” as being designed by the citizens to maximize their utility subject to the constraints placed by the self-interested behavior of the government. In addition, throughout the paper we focus on perfect Bayesian equilibria (see Definition 1).

<sup>14</sup>The mechanisms we describe here allow for general message spaces, but impose two restrictions. First, they are non-stochastic. This is only to simplify notation in the text. In the Appendix, we consider potentially stochastic mechanisms to convexify the constraint set. Second, a more general mechanism would be a mapping from the message histories of all agents, not just the individual’s history. Since there is a continuum of agents that do not share any information, this latter restriction is without loss of generality here (except that off the equilibrium path, some submechanisms would violate the resource constraint, though this is not important for our equilibrium analysis). Notice also that while the submechanism restricts each individual’s allocations to be a function of only his own history of reports, as it will become clear below, the government’s strategies allow submechanisms to be functions of the reports of *all* agents in the past.

Finally, we could define a submechanism as a mapping  $M_t [K_t]$  conditional on the capital stock of the economy at that date to emphasize that what can be achieved will be a function of the capital stock. We suppress this dependence to simplify notation.

below) by  $\mathcal{M}_t$ .

The typical assumption in models with no commitment is that the mechanism designer can commit to a submechanism at a given date, but cannot commit to what mechanisms will be offered in the future. In our context, there is an additional type of deviation for the politician in power whereby she can use her power to extract resources from the society even within the same period.

We consider the following game. At each time  $t$ , the economy starts with a politician  $\iota_t \in \mathcal{I}$  in power and a stock of capital inherited from the previous period,  $K_t$ . Then:

1. At the beginning of period  $t$ , the politician offers a submechanism  $\tilde{M}_t \in \mathcal{M}_t$ .
2. Individuals send a message  $z_t \in Z_t$ . The message  $z_t$  together with the history of messages  $z^{t-1} \in Z^{t-1}$  determine labor supplies  $\tilde{l}_t(z^{i,t})$  according to the submechanism  $\tilde{M}_t$ , where  $i \in [0, 1]$  indexes individuals and  $z^{i,t} \in Z^t$  denotes the history of reports by individual  $i$ .
3. Production takes place according to the labor supplies of the individuals, with  $Y_t = F(K_t, L_t)$ , where  $K_t$  is the capital stock inherited from the previous period, and  $L_t = \int_0^1 \tilde{l}_t(z^{i,t}) di$ .
4. The politician decides whether to deviate from the submechanism  $\tilde{M}_t$ , denoted by  $\xi_t \in \{0, 1\}$ . If  $\xi_t = 0$ , production is distributed among agents according to the pre-specified submechanism  $\tilde{M}_t \in \mathcal{M}_t$ , the politician chooses  $\tilde{x}_t \leq F(K_t, L_t)$ , and next period's capital stock is determined as  $\tilde{K}_{t+1} = F(K_t, L_t) - \tilde{x}_t - \int_0^1 \tilde{c}_t(z^{i,t}) di$ . If  $\xi_t = 1$ , the politician chooses  $\tilde{x}'_t \leq F(K_t, L_t)$ , and a new consumption function  $\tilde{c}'_t : Z^t \rightarrow \mathbb{R}_+$ , and next period's capital stock is:  $\tilde{K}'_{t+1} = F(K_t, L_t) - \tilde{x}'_t - \int_0^1 \tilde{c}'_t(z^{i,t}) di$ .<sup>15</sup>
5. Elections are held and citizens jointly decide whether to keep the politician or replace him with a new one, denoted by  $\rho_t \in \{0, 1\}$ , where  $\rho_t = 1$  denotes replacement. Denote by  $\mathcal{R}^t \in \{0, 1\}^t$  the set of all possible histories of electoral decisions at time  $t$  and by  $\mathcal{R}$  the set of all possible electoral decisions. Replacement of politicians is without any costs.

This game emphasizes that the only difference between the standard models with no commitment and our setup is that the politician, in stage 4, can also decide to expropriate the output produced in the economy, and citizens can replace the politician at the last stage. Notice that at stage 4 labor supply decisions have already been made according to the pre-specified

---

<sup>15</sup>More generally, we can allow the government to capture a fraction  $\eta \leq 1$  of the total output of the economy when  $\xi = 1$ , where the level of  $\eta$  could be related to the institutional controls on government or politician behavior. In this case, the constraint on the government following a deviation would be  $\tilde{K}'_{t+1}(h^t) = \eta F(K_t, L_t) - \tilde{x}'_t - \int_0^1 \tilde{c}'_t(z^{i,t}) di$ , with the remaining  $1 - \eta$  fraction of the output getting destroyed. This generalization has no effect on our results, and we set  $\eta = 1$  to simplify notation.

submechanism  $\tilde{M}_t$ . However, consumption allocations cannot be made according to  $\tilde{M}_t$ , since the politician is expropriating some of the output for herself. Consequently, we also let the politician in power choose a new consumption allocation function,  $\tilde{c}'_t : Z^t \rightarrow \mathbb{R}_+$  at this point.

The important feature of stage 5 is that even though individuals make their economic decisions independently, they make their political decisions—elections to replace the politician—jointly. This is natural since there is no conflict of interest among the citizens over the replacement decision. Joint political decisions can be achieved by a variety of procedures, including various voting schemes (see, for example, Persson and Tabellini, 2000, Chapter 4). Here we simplify the discussion by assuming that the decision  $\rho_t \in \{0, 1\}$  is taken by a randomly chosen citizen.<sup>16</sup>

### 3.2 Histories and Reporting Strategies

Let  $M = \{M_t\}_{t=0}^\infty$  with  $M_t \in \mathcal{M}_t$  be a mechanism, with the set of mechanisms denoted by  $\mathcal{M}$ . Let  $x = \{x_t\}_{t=0}^\infty$  be the sequence of politician's consumption levels. We define a *social plan* as  $(M, x)$ , which is an implicitly-agreed sequence of submechanisms and consumption levels for the politician.

We represent the action of the politician at time  $t$  by  $v_t = (\tilde{M}_t, \xi_t, \tilde{x}_t, \tilde{x}'_t, \tilde{c}'_t)$ . The first element of  $v_t$  is the submechanism that the politician offers at stage 1 of time  $t$ , and the second is the politician's expropriation decision. The third element of  $v_t$  is what the politician consumes herself if  $\xi_t = 0$ . Since  $\tilde{M}_t$  specifies both total production and total consumption by the citizens, given  $\tilde{x}_t$  the capital stock for next period,  $\tilde{K}_{t+1}$ , is determined as a residual from the resource constraint and is not specified as part of the action profile of the politicians.<sup>17</sup> The fourth element,  $\tilde{x}_t$ , is the consumption level for the politician in power when  $\xi_t = 1$ . Finally, the fifth element is the function  $\tilde{c}'_t$  that the politician chooses after deviating from the original submechanism, with  $\mathcal{C}_t$  denoting the set of all such functions. Once again the capital stock for the following period,  $\tilde{K}'_{t+1}$ , is determined as a residual from the resource constraint. Government (politician) consumption levels must satisfy:  $\tilde{x}_t \leq F(K_t, L_t)$  and  $\tilde{x}'_t \leq F(K_t, L_t)$ , but to simplify notation we write  $\tilde{x}_t, \tilde{x}'_t \in \mathbb{R}_+$ . Let  $\Upsilon_t$  be the set of  $v_t$ 's and  $v^t \in \Upsilon^t$  denote the history of  $v_t$ 's up to and including time  $t$ , and assume that this is publicly observable.<sup>18</sup>

<sup>16</sup>Exactly the same equilibrium is obtained if there are majoritarian elections over the replacement decision and each individual votes sincerely (which is the weakly dominant strategy for each citizen in the election).

<sup>17</sup>Since we are characterizing a (sustainable) mechanism, the ownership of the capital stock  $\tilde{K}_{t+1}$  is not specified. Instead, this is simply the amount of resources used in production in the following period, and the government decides how this production will be distributed.

<sup>18</sup>In fact,  $v^t$  includes the action  $\tilde{x}'_t$  and the function  $\tilde{c}'_t$ , which are not observed when  $\xi_t = 0$ . Thus, more appropriately, only a subset of  $v^t$  should be observed publicly. This slight abuse of notation is without any consequence for the analysis.

Let

$$\tilde{h}^t \equiv (K_0, \iota_0, v_0, x_0, \rho_0, K_1, \dots, K_t, \iota_t, v_t, x_t, \rho_t, K_{t+1})$$

denote the public history of the game up to date  $t$ , and  $\tilde{H}^t$  be the set of all such histories.

The electoral decision at time  $t$ ,  $\rho_t$ , is formally defined as

$$\rho_t : \tilde{H}^{t-1} \times v_t \rightarrow \{0, 1\},$$

given the public history at time  $t - 1$  and actions of politicians at time  $t$ , the society chooses whether to replace a politician.

For the citizens, define  $\alpha_t^i(\theta^t | z^{t-1}, \tilde{h}^{t-1})$  as the reporting action of an individual  $i$  at time  $t$  when her type history is  $\theta^t$ , her history of messages so far is  $z^{t-1}$  and the publicly observed history up to time  $t - 1$  are  $\tilde{h}^{t-1}$ . The action  $\alpha_t^i$  specifies a message  $z_t \in Z_t$ , so:

$$\alpha_t^i : Z^{t-1} \times \tilde{H}^{t-1} \times \Theta^t \rightarrow Z_t.$$

We write  $z^t(\alpha_t(\theta^t))$  to denote the message resulting from strategy  $\alpha_t$  for an agent of type  $\theta^t$ . A strategy is *truth telling* if it satisfies

$$\alpha^* \left( \theta^t | z^{t-1}, \tilde{h}^{t-1} \right) = z_t[\theta^t] \text{ for all } \theta^t \in \Theta^t, z^{t-1} \in Z^{t-1}, \text{ and } \tilde{h}^{t-1} \in \tilde{H}^{t-1}, \quad (3)$$

where the notation  $z_t[\theta^t]$  means that the individual is sending a message that fully reveals her true type. To economize on notation, we represent the truth-telling strategy by  $\alpha_t^i(\theta_t | z^{t-1}[\theta^{t-1}], \tilde{h}^{t-1}) = \alpha^*$ . Notice that this strategy only imposes truth-telling following truthful reports in the past (since instead of an arbitrary history of messages  $z^{t-1}$ , we have conditioned on  $z^{t-1}[\theta^{t-1}]$ ). In addition, let us define the null strategy

$$\alpha^0 \left( \theta_t | z^{t-1}, \tilde{h}^{t-1} \right) = z_t^0 \text{ for all } \theta^t \in \Theta^t, z^{t-1} \in Z^{t-1}, \tilde{h}^{t-1} \in \tilde{H}^{t-1},$$

where  $z_t^0$  stands for a message signifying that the individual is disabled (i.e.,  $\theta_t^i = \theta_0$ ). Such a message must always be allowed in any submechanism that is an element of  $\mathcal{M}_t$  because of Assumption 3.<sup>19</sup> Therefore, the individual can always choose to supply zero labor, or in other words, any feasible mechanism (submechanism) must allow for “freedom of labor supply”. We will use the notation  $\alpha_t^i(\theta_t | z^{t-1}, \tilde{h}^{t-1}) = \alpha^0$  to denote that the individual is playing the null strategy.

We denote the *reporting strategy profile* of all the individuals in society by  $\underline{\alpha}$ , with  $\mathbf{A}$  corresponding to the set of all such reporting strategy profiles. We denote the *strategy profile* of all the individuals in society by  $\{\underline{\alpha}, \rho\}$ .

<sup>19</sup>Since an individual with  $\theta_t^i = \theta_0$  cannot supply any labor, he must always send the message  $z_t^0$  (or an equivalent message).

### 3.3 Definition of Equilibrium

Let  $\underline{z}_t \in \mathcal{Z}_t$  be a profile of reports at time  $t$ .<sup>20</sup> As usual, we define

$$\mathcal{Z}^t = \prod_{s=0}^t \mathcal{Z}_s.$$

The strategy of the politician in power at time  $t$  is therefore

$$\Gamma_t : \tilde{H}^{t-1} \times \mathcal{Z}^{t-1} \rightarrow \Upsilon,$$

that is, it determines  $\tilde{M}_t \in \mathcal{M}_t$ ,  $\xi_t \in \{0, 1\}$ ,  $\tilde{x}_t \in [0, F(K_t, L_t)]$ ,  $\tilde{x}'_t \in [0, F(K_t, L_t)]$  and  $\tilde{c}_t \in \mathcal{C}_t$  as a function of the public history and the entire history of reports by citizens. We denote the strategy profile of the politician's by  $\Gamma$  and the set of these strategies by  $\mathcal{G}$ .

**Definition 1** A (Perfect Bayesian) equilibrium in the game between the politicians and the citizens is given by strategy profiles  $\hat{\Gamma}$  and  $\{\underline{\alpha}, \rho\}$  that are sequentially rational, i.e., best responses to each other in all information sets given beliefs, and whenever possible, beliefs are derived from Bayesian updating given the strategy profiles.<sup>21</sup> We write the requirement that these strategy profiles are best responses to each other as  $\hat{\Gamma} \succeq_{\{\underline{\alpha}, \rho\}} \Gamma$  for all  $\Gamma \in \mathcal{G}$  and  $\{\underline{\alpha}, \rho\} \succeq_{\hat{\Gamma}} \{\underline{\alpha}', \rho'\}$  for all  $\underline{\alpha}, \underline{\alpha}' \in \mathbf{A}$  and  $\rho, \rho' \in \mathcal{R}$ .

Let us define  $\Gamma_{M,x} = \left[ \left\{ \tilde{M}_t, \xi_t, \tilde{x}_t, \tilde{x}'_t, \tilde{c}_t \right\}_{t=0}^{\infty} \right]$  as the action profile of the politician induced by strategy  $\Gamma$  given a social plan  $(M, x)$ .

**Definition 2**  $M$  is a sustainable mechanism if there exists  $x = \{x_t\}_{t=0}^{\infty}$ , a strategy profile  $\{\underline{\alpha}, \rho\}$  for the citizens and a strategy profile  $\Gamma_{M,x} \in \mathcal{G}$  for the government, which constitute an equilibrium and induce an action profile  $\left[ \left\{ \tilde{M}_t, \xi_t, \tilde{x}_t, \tilde{x}'_t, \tilde{c}_t \right\}_{t=0}^{\infty} \right]$  for the politicians such that  $\tilde{M}_t = M_t$ ,  $\xi_t = 0$ , and  $\tilde{x}_t(h^t) = x_t(h^t)$  for all  $h^t \in H^t$ , and satisfies  $\Gamma_{M,x} \succeq_{\underline{\alpha}, \rho} \Gamma$  for all  $\Gamma \in \mathcal{G}$ . In this case, we say that equilibrium strategy profiles  $\Gamma_{M,x}$  and  $\{\underline{\alpha}, \rho\}$  support the sustainable mechanism  $M$ .

In essence, this implies that the politician in power does not wish to deviate from the social plan  $(M, x)$  given the strategy profile,  $\{\underline{\alpha}, \rho\}$ , of the citizens. The notation  $\hat{\Gamma} \succeq_{\{\underline{\alpha}, \rho\}} \Gamma$  makes this explicit, stating that given the strategy profile,  $\{\underline{\alpha}, \rho\}$ , of the citizens, the politician weakly prefers its strategy profile to any other strategy profile based on the same implicit agreement.

<sup>20</sup> More formally,  $\underline{z}_t$  assigns a report to each individual, thus it is a function of the form  $\underline{z} : [0, 1] \rightarrow \mathcal{Z}_t$ , where  $i \in [0, 1]$  denotes individual  $i$ , and  $\mathcal{Z}_t$  is the set of all such functions.

<sup>21</sup> We do not introduce explicit notation to describe beliefs, since these do not play any role in any of the analysis or the proofs.

## 4 Truthful Revelation Along the Equilibrium Path

The revelation principle is a powerful tool for the analysis of mechanism design and implementation problems (see, e.g., MasCollé, Winston and Green, 1995). Since, in our environment, the politician in power, who operates the mechanism, cannot commit and has different interests than those of the agents, the simplest version of the revelation principle does not hold; there will exist situations in which individuals will prefer not to report their true type (e.g., Roberts, 1984, Freixas, Guesnerie and Tirole, 1985, or Bisin and Rampini, 2005).<sup>22</sup>

The key result of this section will be that along the equilibrium path, a version of the revelation principle will hold (without introducing a fictional mechanism designer and for all positive discount factors). The main difference between our approach and the literature on dynamic mechanism design without commitment (with the ratchet effect) is that the possibility that the agents can punish the deviating politician (mechanism designer) by replacing her. Such punishments are natural in the context of political economy models, though they are typically not present in other mechanism design problems without commitment. Another important difference is that, as it will become clear below, the punishments that can be imposed on deviating politicians will be independent of the history of mechanisms to date. These differences are responsible for truthful revelation along the equilibrium path in our model

### 4.1 Truthful Revelation

We focus on SPE that maximize utility of the citizens, which we refer to as the *best SPE*. As we will see below, as long as the set of sustainable mechanisms (i.e., the constraint set, (5)-(7)) is nonempty, this is equivalent to choosing the best sustainable mechanism, given by the following program:

$$\mathbf{MAX}_0: \max \mathbb{E} \left[ \sum_{t=0}^{\infty} \beta^t u \left( \tilde{c}_t(z^t[\alpha_t(\theta^t)]), \tilde{l}_t(z^t[\alpha_t(\theta^t)]) \mid \theta_t^i \right) \right] \quad (4)$$

subject to an initial capital stock  $K_0$ , the resource constraint,

$$K_{t+1} = F \left( K_t, \int \tilde{l}_t(z^t[\alpha_t(\theta^t)]) dG^t(\theta^t) \right) - \int \tilde{c}_t(z^t[\alpha_t(\theta^t)]) dG^t(\theta^t) - \tilde{x}_t, \quad (5)$$

a set of incentive compatibility constraints and electoral decisions for individuals,

$$\{\underline{\alpha}, \rho\} \text{ is a best response to } \Gamma_{M,x}, \quad (6)$$

---

<sup>22</sup>As noted in the Introduction, this statement refers to the case in which messages are sent to the government. It is possible to construct alternative environments with fictional mechanism designers with full commitment power, so that the revelation principle holds.

and the “sustainability” constraint of the politician in power:

$$\mathbb{E} \left[ \sum_{s=0}^{\infty} \delta^s v(\tilde{x}_{t+s}) \right] \geq \max_{\tilde{x}'_t, \tilde{K}'_{t+1}, \tilde{c}'_t} \mathbb{E} \left[ \left\{ v(\tilde{x}'_t) + \delta v_t^c(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t) \right\} \right], \quad (7)$$

for all  $t \geq 0$ .

The last constraint, (7), encompasses all the possible deviations by the politician at date  $t$ : the left-hand side is what the politician will receive from date  $t$  onwards by sticking with the implicitly-agreed consumption schedule for herself. The right-hand side is the maximum she can receive by deviating. The potential deviations include a deviation at the last stage of the subgame at time  $t$  to expropriation,  $\xi_t = 1$ , together with a new consumption schedule for individuals,  $\tilde{c}'_t$ ; or  $\xi_t = 0$  and a choice of  $\tilde{x}'_t$  different from  $x_t$ ; or the offer of a new submechanism at time  $t + 1$  (encapsulated into the continuation value  $v_t^c$ ). In the case where  $\xi_t = 1$ , the politician chooses  $\tilde{x}'_t$ ,  $\tilde{K}'_{t+1}$  and  $\tilde{c}'_t$  to maximize her deviation value, which is given by current utility,  $v(\tilde{x}'_t)$ , and continuation value, written as  $v_t^c(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t)$ , to emphasize that this continuation value depends on the entire history of submechanisms (thus information) up to time  $t$ ,  $\tilde{M}^t$ , and on the capital stock from then on,  $\tilde{K}'_{t+1}$ , as well as potentially on  $\tilde{c}'_t$ . If this constraint, (7), were not satisfied, it is either because the politician prefers  $\xi_t = 0$  and some sequence of submechanisms or consumption levels different from  $(M, x)$ , or because the politician prefers  $\xi_t = 1$ . In the former case, we can always change  $(M, x)$  to ensure that (7) is satisfied. The latter, i.e.,  $\xi_t = 1$ , cannot be part of the best equilibrium allocation from the viewpoint of the citizens, since it involves government expropriation. Consequently, as long as the constraint set given by (5)-(7) is nonempty, the best allocation must satisfy (7) and is thus a solution to the program of maximizing (4) subject to (5)-(7). Finally, this constraint set is indeed nonempty, since the trivial allocation with zero production and zero consumption for all parties is in the set.

Let us also introduce the notation  $\underline{\alpha} = (\alpha \mid \alpha')$  to denote a strategy profile where all individuals play  $\alpha$  along the equilibrium path and  $\alpha'$  off the equilibrium path. We then have:

**Lemma 1** *Suppose Assumptions 1-5 hold. In any sustainable mechanism,*

$$\mathbb{E} \left[ \sum_{s=0}^{\infty} \delta^s v(x_{t+s}) \right] \geq v \left( F \left( K_t(\tilde{h}^{t-1}), L_t(\tilde{h}^t) \right) \right) \text{ for all } t, \quad (8)$$

*is necessary. The allocation of resources in the best SPE (best sustainable mechanism) involves no replacement of the initial politician along the equilibrium path, is identical to the solution of the maximization problem in  $(\mathbf{MAX}_0)$  with  $v_t^c(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t) = 0$  for all  $\tilde{M}^t \in \mathcal{M}^t$ ,  $\tilde{K}'_{t+1} \in \mathbb{R}_+$  and  $\tilde{c}'_t \in \mathcal{C}_t$ , and the sustainability constraint (7) is equivalent to (8).*

**Proof.** Let  $\{\tilde{M}, \tilde{x}_t\}_{t=0}^\infty$  be a solution to **(MAX<sub>0</sub>)**. Introduce the following notation:  $h^t = \hat{h}^t$  if  $\{M_s, x_s\} = \{\tilde{M}_s, \tilde{x}_s\}$  for all  $s \leq t$ . Consider the strategy profile  $\rho^\varnothing$  for the citizens such that  $\rho^\varnothing(\tilde{h}^t) = 0$  if  $h^t = \hat{h}^t$  and  $\rho^\varnothing(h^t) = 1$  if  $h^t \neq \hat{h}^t$ . That is, citizens replace the politician unless the politician has always chosen a strategy inducing the allocation  $\{\tilde{M}_t, \tilde{x}_t\}_{t=0}^\infty$  in all previous periods. It is a best response for the politician to choose  $\{\tilde{M}_t, \tilde{x}_t\}_{t=0}^\infty$  after history  $\tilde{h}^t$  only if

$$\mathbb{E} \left[ \sum_{s=0}^{\infty} \delta^s v(\tilde{x}_{t+s}) \right] \geq \max_{\tilde{x}'_t, \tilde{K}'_{t+1}, \tilde{c}'_t} \mathbb{E} \left[ \left\{ v(\tilde{x}'_t) + \delta v_t^c(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t) \right\} \right]$$

where  $v_t^c(x'_t, \tilde{c}'_t, K'_{t+1})$  is the politician's continuation value following a deviation to a feasible  $(x'_t, \tilde{c}'_t, K'_{t+1})$ .

If (8) is violated following some public history  $h^t$ , the best deviation for the politician is  $\xi_t = 1$  and  $x'_t = F(\tilde{K}_t, \tilde{L}_t)$ . This deviation payoff is greater than its equilibrium payoff following  $h^t$ , given by the left-hand side of (8). This contradicts sustainability and establishes that (8) is necessary in any sustainable mechanism.

To see that (8) is sufficient for the best sustainable mechanism, note that reducing  $v_t^c(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t)$  is equivalent to relaxing the constraint on problem (4), so is always preferred. Since from Assumption 5,  $v_t^c \geq 0$  (i.e.,  $x \geq 0$  and  $v(0) = 0$ ), we only need to show that  $v_t^c(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t) = 0$  is achievable for all  $\tilde{M}^t \in \mathcal{M}^t$ ,  $\Gamma' \in \mathcal{G}$ ,  $\tilde{K}'_{t+1} \in \mathbb{R}_+$  and  $\tilde{c}'_t \in \mathcal{C}_t$ . Under the candidate equilibrium strategy  $\rho^\varnothing$  which involves replacing the politician when she deviates, the continuation value of the politician is clearly  $v^c = 0$  regardless of the history of pass play. This establishes the sufficiency of (8).

Next suppose  $\{\tilde{M}_t, \tilde{x}_t\}_{t=0}^\infty$  that is a solution to **(MAX<sub>0</sub>)** can be supported as a perfect Bayesian equilibrium with replacement of the initial politician. Now consider an alternative allocation  $\{\tilde{M}'_t, \tilde{x}'_t\}_{t=0}^\infty$  such that the initial politician is kept in power along the equilibrium path and receives exactly the same consumption sequence as the new politicians would have received after replacement. Since  $\{\{\tilde{M}_t, \tilde{x}_t\}_{t=0}^\infty$  satisfies (8) for the new politicians at all  $t$ ,  $\{\tilde{M}'_t, \tilde{x}'_t\}_{t=0}^\infty$  satisfies (8) for all  $t$  for the initial politician. Moreover, since  $\{\tilde{M}_t, \tilde{x}_t\}_{t=0}^\infty$  must involve at least some positive consumption for the new politicians,  $\{\tilde{M}'_t, \tilde{x}'_t\}_{t=0}^\infty$  yields a higher  $t = 0$  utility to the initial politician. Thus,  $x_0$  can be reduced and consumption of agents at  $t = 0$  can be increased without violating (8), so  $\{\tilde{M}'_t, \tilde{x}'_t\}_{t=0}^\infty$  cannot be a solution to **(MAX<sub>0</sub>)**. This proves that there is no replacement of the initial politician along the equilibrium path.

To complete the proof, we only need to show that citizens' strategy (in particular, the replacement strategy  $\rho^\varnothing$ ) is sequentially rational. This follows by considering the following continuation strategy for a politician: if  $h^t \neq \hat{h}^t$ , then  $x_t = F(K_t, L_t)$  and  $\xi_t = 1$ . This ensures that  $\rho^\varnothing$  and  $\underline{\alpha} = \alpha^0$  are a best response for the citizens. ■

This lemma uses the fact that irrespective of the history of submechanisms and the amount

of capital stock left for future production, there is an equilibrium continuation play that replacing a politician gives the deviator zero utility from that point onwards (which is analogous to the results in repeated games where the most severe punishments against deviations are optimal, e.g., Abreu, 1988). This continuation play is used as the threat against politician deviation from the implicitly-agreed social plan. The implication is that, along the best sustainable mechanism, the best deviation for the politician involves  $\xi_t = 1$  and  $\tilde{x}'_t = F(K_t, L_t)$ . This enables us to simplify the sustainability constraints of the politician to (8), which also has the virtue of not depending on the history of submechanisms up to that point.<sup>23</sup> Moreover, the lemma also shows that in any sustainable mechanism (8) is necessary.

Next, we define a *direct (sub)mechanism* as  $M_t^* : \Theta^t \rightarrow [0, \bar{l}] \times \mathbb{R}_+$ . In other words, direct mechanisms involve a restricted message space,  $Z_t = \Theta_t$ , where individuals only report their current type. We denote a strategy profile by the politician's inducing direct submechanisms along the equilibrium path by  $\Gamma^*$ .

**Definition 3** *A strategy profile for the citizens,  $\underline{\alpha}^*$ , is truthful if, along the equilibrium path, we have that  $\alpha_t^i(\theta^t | \theta^{t-1}, \tilde{h}^{t-1}) = \alpha^*$ . We write  $\underline{\alpha}^* = (\alpha^* | \alpha')$  to denote a truthful strategy profile.*

The notation  $\underline{\alpha}^* = (\alpha^* | \alpha')$  emphasizes that individuals play truth-telling along the equilibrium path, but may play some different strategy profile,  $\alpha'$ , off the equilibrium path. Clearly, a truthful strategy against a direct mechanism simply amounts to reporting the true type of the agent. Let us next define  $\underline{c}[\Gamma, \underline{\alpha}]$ ,  $\underline{l}[\Gamma, \underline{\alpha}]$  and  $x[\Gamma, \underline{\alpha}]$  as, respectively, the equilibrium consumption and labor supply distributions across individuals (as a function of the history of their reports), and the sequence of government consumption levels resulting from the strategy profiles of the politicians and citizens, such that all of these functions only condition on information available up to time  $t$  for allocations of time  $t$ .

**Theorem 1 (Truthful Revelation Along the Equilibrium Path)** *Suppose Assumptions 1-5 hold and that  $\Gamma$  and  $\{\underline{\alpha}, \rho\}$  are a combination of strategy profiles and electoral decisions that support a sustainable mechanism. Then, there exists another pair of equilibrium strategy profiles  $\Gamma^*$  and  $\underline{\alpha}^* = (\alpha^* | \alpha')$  for some  $\alpha'$  such that  $\Gamma^*$  induces direct submechanisms and  $\underline{\alpha}^*$  induces truth telling along the equilibrium path, and moreover  $\underline{c}[\Gamma, \underline{\alpha}] = \underline{c}[\Gamma^*, \underline{\alpha}^*]$ ,  $\underline{l}[\Gamma, \underline{\alpha}] = \underline{l}[\Gamma^*, \underline{\alpha}^*]$ , and  $x[\Gamma, \underline{\alpha}] = x[\Gamma^*, \underline{\alpha}^*]$ .*

**Proof.** Take equilibrium strategy profiles  $\Gamma$  and  $\underline{\alpha}$  that support a sustainable mechanism. Then by definition  $\xi_t = 0$  for all  $t$ , and from Lemma 1, (8) is satisfied. Let the best response

<sup>23</sup>This statement refers to the sustainability constraint, (8). The optimal mechanism will clearly make allocations depend on the history of individual messages.

of type  $\theta^t$  at time  $t$  according to  $\underline{\alpha}$  be to announce  $z_{t,\Gamma}(\theta^t, \tilde{h}^{t-1})$  given a history of reports  $z_{\Gamma}^{t-1}(\theta^{t-1}, \tilde{h}^{t-1})$  and public history  $\tilde{h}^{t-1}$ . Let  $z_{\Gamma}^t(\theta^t, \tilde{h}^{t-1}) = (z_{\Gamma}^{t-1}(\theta^{t-1}, \tilde{h}^{t-1}), z_{t,\Gamma}(\theta^t, \tilde{h}^{t-1}))$ .

Denote the expected utility of this individual under this mechanism given history  $\tilde{h}^{t-1}$  be  $\tilde{u} [z_{\Gamma}^t(\theta^t, \tilde{h}^{t-1}) \mid \theta^t, \Gamma, \tilde{h}^{t-1}]$ . By definition of  $z_{\Gamma}^t(\theta^t, \tilde{h}^{t-1})$  being a best response, we have

$$\tilde{u} [z_{\Gamma}^t(\theta^t, \tilde{h}^{t-1}) \mid \theta^t, \Gamma, \tilde{h}^{t-1}] \geq \tilde{u} [\tilde{z}_{\Gamma}^t(\theta^t, \tilde{h}^{t-1}) \mid \theta^t, \Gamma, \tilde{h}^{t-1}] \text{ for all } \tilde{z}_{\Gamma}^t(\theta^t, \tilde{h}^{t-1}) \in Z^t \text{ and } \tilde{h}^{t-1} \in \tilde{H}^{t-1}.$$

Now consider the alternative strategy profile for the politician  $\Gamma^*$ , which induces the action profile  $\left[ \left\{ \tilde{M}_t, \xi_t, \tilde{x}_t, \tilde{x}'_t, \tilde{c}'_t \right\}_{t=0}^{\infty} \right]$  such that  $\xi_t = 0$  for all  $t$ ,  $\tilde{M}_t = M_t^*$  (where  $M_t^*$  is a direct submechanism) and  $\underline{c}[\Gamma^*, \underline{\alpha}^*, \underline{h}] = \underline{c}[\Gamma, \underline{\alpha}, \underline{h}]$ ,  $\underline{l}[\Gamma^*, \underline{\alpha}^*, \underline{h}] = \underline{l}[\Gamma, \underline{\alpha}, \underline{h}]$ , and  $x[\Gamma, \underline{\alpha}, \underline{h}] = x[\Gamma^*, \underline{\alpha}^*, \underline{h}]$ . Therefore, by construction,

$$\begin{aligned} \tilde{u} [\theta^t, \tilde{h}^{t-1} \mid \theta^t, \Gamma^*, \tilde{h}^{t-1}] &= \tilde{u} [z_{\Gamma}^t(\theta^t, \tilde{h}^{t-1}) \mid \theta^t, \Gamma, \tilde{h}^{t-1}] \\ &\geq \tilde{u} [\tilde{z}_{\Gamma}^t(\theta^t, \tilde{h}^{t-1}) \mid \theta^t, \Gamma, \tilde{h}^{t-1}] = \tilde{u} [\hat{\theta}^t, h^{t-1} \mid \theta^t, \Gamma^*, \tilde{h}^{t-1}] \end{aligned} \quad (9)$$

for all  $\hat{\theta}^t \in \Theta^t$  and all  $\tilde{h}^{t-1} \in \tilde{H}^{t-1}$ . Equation (9) implies that  $\underline{\alpha}^* = (\alpha^* \mid \alpha')$  is a best response along the equilibrium path for the agents against the mechanism  $M^*$  and politician strategy profile  $\Gamma^*$ . Moreover, by construction, the resulting allocation when individuals play  $\underline{\alpha}^* = (\alpha^* \mid \alpha')$  against  $\Gamma^*$  is the same as when they play  $\underline{\alpha}$  against  $\Gamma$ . Therefore, by the definition of  $\Gamma$  being sustainable, we have  $\Gamma \succeq_{\{\underline{\alpha}, \rho\}} \Gamma'$  for all  $\Gamma' \in \mathcal{G}$ . Now choose  $\alpha'$  to be identical to  $\underline{\alpha}$  off-the-equilibrium path, which implies that  $\Gamma^* \succeq_{\{\underline{\alpha}^*, \rho\}} \Gamma'$  for all  $\Gamma' \in \mathcal{G}$  or that (8) is satisfied, thus establishing that  $(\Gamma^*, \underline{\alpha}^*)$  is an equilibrium. ■

The most important implication of this theorem is that for the rest of the analysis, we can restrict attention to truth-telling (direct) mechanisms on the side of the agents. The reason why, despite the lack of commitment and the self-interested preferences of the mechanism designer, a revelation principle type result holds is twofold: first, the politician has a deviation within the same period; and second, individuals can use punishment strategies involving replacing the politician. The punishment strategies of citizens support a sustainable mechanism, making it the best response for the politician in power to pursue the implicitly-agreed social plan  $(M, x)$ . Given this sustainability, there is effective commitment on the side of the politician *along the equilibrium path*.

This notion is important to distinguish from the commitment that exists in the standard mechanism design problems where there is unconditional commitment (i.e., along all paths). In contrast, in our environment, there is no commitment *off the equilibrium path*, where the politician can exploit the information it has gathered or expropriate part of the output. In fact, off the equilibrium path, non-truthful reporting by the individuals is important to ensure

sustainability. Nevertheless, by definition, along the equilibrium path induced by a sustainable mechanism, the politician prefers not to deviate from the implicitly-agreed social plan and thus individuals can report their types without the fear that this information or their labor supply will be misused.

## 4.2 The Best Sustainable Mechanism

Theorem 1 enables us to focus on direct mechanisms and truth-telling strategy  $\alpha^*$  by all individuals. This implies that the best sustainable mechanism can be achieved by individuals simply reporting their types. Recall that at every date, there is an invariant distribution of  $\theta$  denoted by  $G(\theta)$ . This implies that  $\theta^t$  has an invariant distribution, which is simply the  $t$ -fold version of  $G(\theta)$ ,  $G^t(\theta)$  (since there is a continuum of individuals, each history  $\theta^t$  occurs infinitely often).<sup>24</sup> Given this construction, we can write total labor supply as  $L_t = \int_{\Theta^t} l_t(\theta^t) dG^t(\theta^t)$ , and total consumption as  $C_t = \int_{\Theta^t} c_t(\theta^t) dG^t(\theta^t)$ .<sup>25</sup> Moreover, since Theorem 1 establishes that any sustainable mechanism is equivalent to a direct mechanism with truth-telling on the side of the agents, we obtain the main result from the section, which will be used in the rest of the paper:

**Proposition 1** *Suppose Assumptions 1-5 hold. Then, the best sustainable mechanism is a solution to the following maximization program:*

$$\mathbf{MAX}_1 : \mathbf{U}^{SM} = \max_{\{c_t(\theta^t), l_t(\theta^t), x_t, K_{t+1}\}_{t=0}^{\infty}} \mathbb{E} \left[ \sum_{t=0}^{\infty} \beta^t u(c_t(\theta^{i,t}), l_t(\theta^{i,t}) \mid \theta_t^i) \right] \quad (10)$$

subject to some initial condition  $K_0$ , the resource constraint

$$K_{t+1} = F(K_t, L_t) - C_t - x_t, \quad (11)$$

a set of incentive compatibility constraints for individuals,

$$\begin{aligned} & \mathbb{E} \left[ \sum_{s=0}^{\infty} \beta^s u(c_{t+s}(\theta^{i,t+s}), l_{t+s}(\theta^{i,t+s}) \mid \theta_{t+s}^i) \mid \theta^{i,t} \right] \\ & \geq \mathbb{E} \left[ \sum_{s=0}^{\infty} \beta^s u(c_{t+s}(\hat{\theta}^{i,t+s}), l_{t+s}(\hat{\theta}^{i,t+s}) \mid \theta_{t+s}^i) \mid \theta^{i,t} \right] \end{aligned} \quad (12)$$

for all  $t$ , all  $\theta^{i,t} \in \Theta^t$  and all possible sequences of  $\{\hat{\theta}_{t+s}^i\}_{s=0}^{\infty}$ , and the sustainability constraint of the politician

$$\mathbb{E} \left[ \sum_{s=0}^{\infty} \delta^s v(x_{t+s}) \right] \geq v(F(K_t, L_t)), \quad (13)$$

<sup>24</sup>More formally, given the continuum of agents, we can apply a law of large numbers type argument, and each history  $\theta^t$  will have positive measure. See, for example, Uhlig (1996).

<sup>25</sup>From now on, we suppress the  $\sim$ 's to simplify notation and simply use  $c_t$ ,  $l_t$  and  $x_t$ . Note also that  $\int_{\Theta^t}$  here denotes Lebesgue integrals, and in what follows, we will suppress the range of integration,  $\Theta^t$ .

for all  $t$ .

**Proof.** The proof follows from Lemma 1 and Theorem 1. Suppose there exists an equilibrium  $(\underline{\alpha}^{**}, \Gamma^{**})$ , that maximizes (10). By the argument in the text,  $(\underline{\alpha}^{**}, \Gamma^{**})$  will not feature  $\xi_t = 1$  for any  $t$ . Therefore,  $(\underline{\alpha}^{**}, \Gamma^{**})$  features a sequence of submechanisms  $\{\hat{M}_t\}_{t=0}^{\infty}$ , consumption levels for the politician,  $\{\hat{x}_t\}_{t=0}^{\infty}$  and  $\xi_t = 0$  for all  $t$ . Then setting  $(M, x) = \left(\{\hat{M}_t\}_{t=0}^{\infty}, \{\hat{x}_t\}_{t=0}^{\infty}\right)$  implies that  $(\underline{\alpha}^{**}, \Gamma^{**})$  support a sustainable mechanism. Then, use Theorem 1 to find  $(\underline{\alpha}^*, \Gamma^*)$  corresponding to a sustainable direct mechanism. This direct mechanism has to satisfy the resource constraint, (11), the incentive compatibility constraints of individuals at all dates, which instead of (6) can be written as (12) since  $\Gamma^*$  induces direct mechanisms. Finally, from Lemma 1, the constraint (13) ensures that  $\Gamma^*$  is a best response to citizens' strategies,  $\underline{\alpha}^*, \rho$ . ■

The role of Theorem 1 in this formulation is that it enables us to write the program for the best sustainable mechanism as a direct mechanism with truth-telling, thus reducing the larger set of incentive compatibility constraints of individuals to (12).<sup>26</sup>

It should also be noted that Proposition 1 does not make use of “self-harming” punishments by citizens in order to ensure the sustainability constraint (13). In particular, citizens simply replace a politician who has deviated from the implicitly-agreed social plan. Using this idea, it can also be shown that a version of Proposition 1 holds if we focus on *renegotiation-proof* equilibria, with the appropriate definition of renegotiation-proofness. Acemoglu, Golosov and Tsyvinski (2007) establish a similar result in the context of a related political economy model embedded in a neoclassical growth framework. We do not present this additional result here to economize on space.

## 5 Separation of Private and Public Incentives

We now proceed to our next result. We show that our analysis of the dynamic Mirrlees economy with self-interested politicians is simplified by separating the provision of incentives to individuals from the provision of incentives to politicians.

Let us first define the *dynamic Mirrlees program* (with full-commitment, benevolent government, and exogenous government expenditures). Imagine the economy needs to finance an exogenous government expenditure  $X_t \geq 0$  at time  $t$ . Then the dynamic Mirrlees program of maximizing the time  $t = 0$  (*ex ante*) utility of a representative agent, can be written as (e.g.,

---

<sup>26</sup>The equations in (12) focus on the incentive compatibility constraints that apply along the equilibrium path (expectations on both sides of the constraints are taken conditional on  $\theta^{i,t}$ ). This is without any loss of generality, since (12) needs to hold for any sequence of reports  $\{\hat{\theta}_{t+s}^i\}_{s=0}^{\infty}$ , thus any potential deviation from time  $t = 0$  is covered by this set of constraints.

Golosov, Kocherlakota and Tsyvinski, 2003, Kocherlakota, 2005):

$$\max_{\{c_t(\theta^t), l_t(\theta^t)\}_{t=0}^{\infty}} \mathbb{E} \left[ \sum_{t=0}^{\infty} \beta^t u(c_t(\theta^{i,t}), l_t(\theta^{i,t}) \mid \theta_t^i) \right] \quad (14)$$

subject to the incentive compatibility constraints, (12), and  $C_t + X_t + K_{t+1} \leq F(K_t, L_t)$ . Moreover, we add the feasibility constraint that  $\{X_t\}_{t=0}^{\infty}$  should be such that

$$\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty},$$

where

$$\Lambda^{\infty} = \{ \{C_t, L_t\}_{t=0}^{\infty} \text{ such that } \exists \{c_t(\theta^t), l_t(\theta^t)\}_{t=0}^{\infty} \text{ satisfying (12)} \} \quad (15)$$

In other words,  $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}$  implies that there exist incentive feasible and compatible  $\{c_t(\theta^t), l_t(\theta^t)\}_{t=0}^{\infty}$ . This set is important to define as for certain government expenditure sequences,  $\{X_t\}_{t=0}^{\infty}$ 's, the constraint set of this Mirrlees maximization problem can be empty (e.g., if  $C_t = 0$  and  $L_t > 0$ , the incentive compatibility constraints of individuals cannot be satisfied).

For a sequence  $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}$ , we can define the *quasi-Mirrlees program* as

$$\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) \equiv \max_{\{c_t(\theta^t), l_t(\theta^t)\}_{t=0}^{\infty}} \mathbb{E} \left[ \sum_{t=0}^{\infty} \beta^t u(c_t(\theta^{i,t}), l_t(\theta^{i,t}) \mid \theta_t^i) \right] \quad (16)$$

subject to the incentive compatibility constraints, (12), and two additional constraints

$$\int c_t(\theta^t) dG(\theta^t) \leq C_t, \quad (17)$$

and

$$\int l_t(\theta^t) dG(\theta^t) \geq L_t. \quad (18)$$

This program takes the sequence  $\{C_t, L_t\}_{t=0}^{\infty}$  as given and maximizes ex ante utility of an agent subject to incentive constraints and to two additional constraints. The first, (17), requires the sum of consumption levels across agents for all report histories to be no greater than some number  $C_t$ , while the second, (18), requires the sum of labor supplies to be no less than some amount  $L_t$ . The functional  $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$  defines the maximum *ex ante* ( $t = 0$ ) utility of an agent in this economy for a given sequence  $\{C_t, L_t\}_{t=0}^{\infty}$ . In Appendix A, we show that the functional  $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$  is well-defined, nondecreasing in  $C_t$ , nonincreasing in  $L_t$ , concave and differentiable (as long as we allow for randomizations). In the text, we will make use of these properties of  $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$  to characterize the best sustainable mechanism.

Returning to the dynamic Mirrlees program, for a given sequence of government expenditures  $\{X_t\}_{t=0}^{\infty}$ , this can be written as:

$$\max_{\{C_t, L_t, K_{t+1}\}_{t=0}^{\infty}} \mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) \quad (19)$$

subject to a given level of  $K_0$  and to

$$C_t + X_t + K_{t+1} \leq F(K_t, L_t), \text{ and } \{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}. \quad (20)$$

Therefore, we can represent the dynamic Mirrlees program as a solution to a two-step maximization problem, in which the first step is the quasi-Mirrlees formulation, yielding the functional  $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ , and the second step is the maximization of  $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$  over sequences  $\{C_t, L_t, K_{t+1}\}$  subject to a resource constrained and feasibility.

Now again using the quasi-Mirrlees formulation, the characterization of the best sustainable mechanism, (10), can be written as

$$\max_{\{C_t, L_t, x_t, K_t\}_{t=0}^{\infty}} \mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) \quad (21)$$

subject to

$$C_t + x_t + K_{t+1} \leq F(K_t, L_t) \text{ and } \{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}, \quad (22)$$

and also subject to (13). The only difference between the dynamic Mirrlees program in (19)-(20) and the best sustainable mechanism in (21)-(22)-(13) is the presence of the sustainability constraint for the politician in power, (13), which also makes  $\{x_t\}_{t=0}^{\infty}$  an endogenously chosen sequence instead of the exogenously given  $\{X_t\}_{t=0}^{\infty}$ . This formulation establishes the following theorem.

**Theorem 2 (*Separation of Private and Public Incentives*)** *Suppose Assumptions 1-5 hold. Then, the best sustainable mechanism solves a quasi-Mirrlees program for some sequence  $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}$ .*

**Proof.** This follows immediately from rewriting (10)-(13) from Proposition 1 as a two-step maximization program, and expressing (11) as  $x_t = F(K_t, L_t) - C_t - K_{t+1}$ . ■

Consequently, the allocation induced by the best sustainable mechanism is a solution to a problem that maximizes the ex ante utility of the citizens as given in (14), but must also choose levels of aggregate consumption and labor supply consistent with the sustainability constraint of the politician in power.<sup>27</sup> An important implication of this result is that political economy

<sup>27</sup>The key feature necessary for Theorem 2 is that politicians' deviation payoffs depend only on aggregates. If, instead of  $x_t = F(K_t, L_t)$ , the maximum consumption for the politician were a nonlinear function of the entire distribution of labor supplies,  $[l_{i,t}]_{i \in I}$ , Theorem 2 would not necessarily hold.

considerations do not fundamentally alter the optimal taxation problem; instead, they modify the aggregate constraints in this dynamic maximization problem. From a technical point of view, this theorem implies that we can separate the analysis of the political economy of dynamic taxation into two parts:

1. We first solve the problem of providing incentives to individuals given aggregate levels of consumption and labor supply.
2. We then provide incentives to politicians by choosing aggregate variables and the level of rents.

Accordingly, the best sustainable mechanism will be *undistorted* when it can achieve the same allocation as that of a full dynamic Mirrlees economy with the same sequence of  $\{x_t\}_{t=0}^{\infty}$  (which naturally involves no marginal distortions in addition to those implied by Mirrleesian optimal taxation).

This theorem also enables us to represent the differences between the dynamic Mirrlees program and the best sustainable mechanism purely in terms of *aggregate distortions*, corresponding to what the sequences  $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}$  are (and how they differ from the solution to the dynamic Mirrlees program in (19)-(20)).

The quasi-Mirrlees program was defined above in (16), and Theorem 2 established that the best sustainable mechanism solves a quasi-Mirrlees program. In addition, Appendix C shows that  $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$  is differentiable in the sequences  $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}$ . This implies that we can think of variations in sequences  $\{C_t, L_t\}_{t=0}^{\infty}$  where only one element,  $C_s$  or  $L_s$  for some specific  $s$  is varied. We denote the derivative of  $\mathcal{U}$  with respect to such variations by  $\mathcal{U}_{C_s}(\{C_t, L_t\}_{t=0}^{\infty})$  and  $\mathcal{U}_{L_s}(\{C_t, L_t\}_{t=0}^{\infty})$  or simply by  $\mathcal{U}_{C_s}$  and  $\mathcal{U}_{L_s}$ . We also denote the partial derivatives of the production function with respect to labor and capital at time  $s$  by  $F_{L_s}$  and  $F_{K_s}$ .

**Definition 4** *We say that the sequence  $\{C_t, L_t, K_{t+1}, x_t\}_{t=0}^{\infty}$  induced by the best sustainable mechanism  $\Gamma^*$  is undistorted at  $t'$  if  $\{\hat{C}_t, \hat{L}_t, \hat{K}_{t+1}\}_{t=0}^{\infty}$  is a solution to (19) subject to (20) with  $\{X_t\}_{t=0}^{\infty} = \{x_t\}_{t=0}^{\infty}$  and  $C_{t'} = \hat{C}_{t'}$ ,  $\hat{L}_{t'} = L_{t'}$ ,  $\hat{K}_{t'+1} = K_{t'+1}$ . We say that  $\{C_t, L_t, K_{t+1}, x_t\}_{t=0}^{\infty}$  is asymptotically undistorted if it is undistorted as  $t \rightarrow \infty$ .*

This definition is a natural specification of what it means for an allocation to be undistorted. When  $\{C_t, L_t\}_{t=0}^{\infty} \in \text{Int}\Lambda^{\infty}$ , we say that an allocation is *undistorted* if

$$\mathcal{U}_{C_t} \cdot F_{L_t} = -\mathcal{U}_{L_t} , \tag{23}$$

$$F_{K_{t+1}} \cdot \mathcal{U}_{C_{t+1}} = \mathcal{U}_{C_t}. \quad (24)$$

at time  $t$  (or as  $t \rightarrow \infty$ ). Here, the first condition requires the marginal cost of effort at time  $t$  given the utility function  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$  to be equal to the increase in output from the additional effort times the marginal utility of additional consumption. The second one requires the cost of a decline in the utility by saving one more unit to be equal to the increase in output in the next period times the marginal utility of consumption then. Once again, these are aggregate conditions since they are defined in terms of the utility functional  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$ , which represents the ex ante maximal utility of an individual subject to incentive constraints. Moreover, if a steady state exists and the conditions in (23) and (24) hold as  $t \rightarrow \infty$ , then it is also clear that  $\{C_t, L_t, K_{t+1}, x_t\}_{t=0}^\infty$  must be undistorted. We therefore say that there are no asymptotic aggregate distortions on capital accumulation (or no aggregate capital taxation) if  $F_{K_{t+1}} \cdot \mathcal{U}_{C_{t+1}} = \mathcal{U}_{C_t}$  and no aggregate distortions on labor supply if  $\mathcal{U}_{C_t} \cdot F_{L_t} = -\mathcal{U}_{L_t}$  as  $t \rightarrow \infty$ .

We say that an allocation  $\{C_t, L_t, K_{t+1}, x_t\}_{t=0}^\infty$  features *labor distortions* at time  $t$  if (23) is not satisfied at  $t$ . We refer to these as *downward labor distortions* if the left-hand side of (23) is strictly greater than the right-hand side. If (24) is not satisfied, then there are *intertemporal distortions* at time  $t$ , and if the left-hand side of (24) is strictly less than the right-hand side, then there are *downward intertemporal distortions*. Downward distortions imply that there is less labor supply and less capital accumulation than in an undistorted allocation. We will interpret these distortions as corresponding to “aggregate tax distortions,” since allocations that involve downward labor and intertemporal distortions can be decentralized by using marginal (e.g., linear) labor and capital taxes.

## 6 Best Sustainable Mechanisms

We next to characterize the behavior of the sequences  $\{C_t, L_t, K_t\}_{t=0}^\infty$  (and  $\{x_t\}_{t=0}^\infty$ ) under the best sustainable mechanism, which is what we turn to next. We now provide the theorem that characterizes the behavior of allocations and distortions for this general environment.

**Theorem 3 (*Characterization of the Best Sustainable Mechanism*)** *Consider the optimal dynamic Mirrlees economy with self-interested politicians and suppose that individual allocations can be a function of the entire individual history. Suppose that there exists  $\{C_t, L_t\}_{t=0}^\infty \in \text{Int}\Lambda^\infty$  (with  $L_t > 0$ ) for some  $t$ . Then, in the best sustainable mechanism:*

1. *there are downward labor distortions at some  $t < \infty$  and downward intertemporal distortions at  $t - 1$  (provided that  $t \geq 1$ ).*

Let the best sustainable mechanism induce a sequence of consumption, labor supply and capital levels  $\{C_t, L_t, K_{t+1}\}_{t=0}^\infty$ . Suppose a steady state exists such that as  $t \rightarrow \infty$ ,  $\{C_t, L_t, K_{t+1}\}_{t=0}^\infty \rightarrow (C^*, L^*, K^*)$ , where  $(C^*, L^*)$  is interior. Moreover, let  $\varphi = \inf\{\varrho \in (0, 1] : \text{plim}_{t \rightarrow \infty} \varrho^{-t} \mathcal{U}_{C_t}^* = 0\}$ , where  $\varphi < 1$ . Then:

2. if  $\varphi = \delta$ , then there are no asymptotic aggregate distortions on capital accumulation and labor supply;
3. if  $\varphi > \delta$ , then aggregate distortions on capital accumulation and labor supply do not disappear even asymptotically.

**Proof.** We show in the Appendix that, when randomizations are introduced,  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$  is a well-defined functional and is continuous, concave, and differentiable. In this proof, we suppress randomization to simplify notation.

We write the problem of characterizing the best sustainable mechanism non-recursively following Marcet and Marimon (1998) as

$$\max_{\{C_t, L_t, K_t, x_t\}_{t=0}^\infty} \mathcal{L} = \mathcal{U}(\{C_t, L_t\}_{t=0}^\infty) + \sum_{t=0}^{\infty} \delta^t \{\mu_t v(x_t) - (\mu_t - \mu_{t-1})v(F(K_t, L_t))\} \quad (25)$$

subject to

$$C_t + x_t + K_{t+1} \leq F(K_t, L_t), \text{ and} \quad (26)$$

$$\{C_t, L_t\}_{t=0}^\infty \in \Lambda^\infty,$$

for all  $t$ , where  $\mu_t = \mu_{t-1} + \psi_t$  with  $\mu_{-1} = 0$  and  $\delta^t \psi_t \geq 0$  is the Lagrange multiplier on the constraint (13). The differentiability of  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$  implies that for  $\{C_t, L_t\}_{t=0}^\infty \in \text{Int}\Lambda^\infty$ , we have:

$$\mathcal{U}_{L_t} - \delta^t (\mu_t - \mu_{t-1}) v'(F(K_t, L_t)) F_{L_t} = -\mathcal{U}_{C_t} \cdot F_{L_t} \quad (27)$$

$$\mathcal{U}_{C_t} = [\mathcal{U}_{C_{t+1}} - \delta^t (\mu_{t+1} - \mu_t) v'(F(K_{t+1}, L_{t+1}))] F_{K_{t+1}} \quad (28)$$

Since  $\mu_t \geq \mu_{t-1}$ , there will be downward labor and intertemporal distortions whenever  $\mu_t > \mu_{t-1}$  and  $\mu_{t+1} > \mu_t$ , i.e., whenever  $\psi_t > 0$  and  $\psi_{t+1} > 0$ .

**Part 1:** Suppose to obtain a contradiction that  $\mu_t = 0$  for all  $t \geq 0$ . Then,  $x_t = 0$  for all  $t$ . But in this case, if  $L_t > 0$  for any  $t$ , then the politician can improve by expropriating the entire output at  $t$ . Thus we must have  $L_t = 0$  for all  $t$ . Since, by hypothesis,  $\{C_t, L_t\}_{t=0}^\infty \in \text{Int}\Lambda^\infty$  with  $L_t > 0$  is feasible and the associated  $\{C_t, L_t\}_{t=0}^\infty \in \text{Int}\Lambda^\infty$  necessarily gives higher ex ante utility to citizens than  $L_t = C_t = 0$ , the plan with  $L_t = 0$  for all  $t$  cannot be optimal. Therefore, the sustainability constraint of politician (13) must bind at some  $t$  with  $\psi_t > 0$ . Then (27) implies

that there will be downward labor distortions at that  $t$ , and (28) implies that there will be downward intertemporal distortions at  $t - 1$ .

**Part 2:** We start by proving that  $\varphi = \sup\{\varrho \in [0, 1] : \text{plim}_{t \rightarrow \infty} \varrho^{-t} \mathcal{U}_{C_t}^* = 0\}$  defined in the theorem is well-defined and strictly less than 1. To see this, recall that by hypothesis, a steady state exists, so that  $\{C_t, L_t, K_{t+1}\}_{t=0}^\infty \rightarrow (C^*, L^*, K^*)$ , thus  $\{C_t\}_{t=0}^\infty$  is in the space  $c$  of convergent infinite sequences (rather than simply in the space of all bounded infinite sequences,  $\ell_\infty$ ). The dual of  $c$  is  $\ell_1$ , that is, the space of sequences  $\{y_t\}_{t=0}^\infty$  such that  $\sum_{t=0}^\infty |y_t| < \infty$ . Since  $\mathcal{U}_{C_t}$  is equal to the Lagrange multiplier for the constraint (17), it lies in the dual space of  $\{C_t\}_{t=0}^\infty$  (see, e.g., Luenberger, 1969, Chapter 9), thus in  $\ell_1$ , which implies that  $\lim_{t \rightarrow \infty} \mathcal{U}_{C_t} = 0$ , hence  $\varphi < 1$ .

Rearranging equations (27) and (28) and substituting for  $\mathcal{U}_{C_t}^*$ , we have

$$-\frac{\mathcal{U}_{L_t}^*}{\mathcal{U}_{C_t}^* F_{L_t}(K^*, L^*)} = 1 - \frac{(\mu_t - \mu_{t-1})v'(F(K^*, L^*))}{\mu_t v'(x^*)} \quad (29)$$

and

$$\frac{F_{K_{t+1}}(K^*, L^*) \mathcal{U}_{C_{t+1}}^*}{\mathcal{U}_{C_t}^*} = 1 + \frac{(\mu_{t+1} - \mu_t)v'(F(K^*, L^*))F_{K_{t+1}}(K^*, L^*)}{\mu_t v'(x^*)}, \quad (30)$$

where all derivatives are evaluated at the limit  $(C^*, L^*, K^*)$ .

The first-order condition with respect to  $x_t$  then implies:

$$\frac{\mathcal{U}_{C_t}}{\delta^t v'(x_t)} = \mu_t \leq \mu_{t+1} = \frac{\mathcal{U}_{C_{t+1}}}{\delta^{t+1} v'(x_{t+1})}. \quad (31)$$

By construction,  $\mu_t$  is an increasing sequence, so it must either converge to some value  $\mu^*$  or go to infinity. Since as  $t \rightarrow \infty$ , an interior steady state  $(C^*, L^*, K^*, x^*)$  exists by hypothesis and  $\mathcal{U}_{C_t}^*$  is proportional to  $\varphi^t$ , (31) can be written as

$$\frac{\varphi^t \mathcal{U}_{C_t}^*}{\delta^t v'(x_t^*)} = \mu_t \leq \mu_{t+1} = \frac{\varphi^{t+1} \mathcal{U}_{C_{t+1}}^*}{\delta^{t+1} v'(x_{t+1}^*)} \text{ as } t \rightarrow \infty. \quad (32)$$

Since  $\varphi = \delta$ , we have that (32) implies that as  $t \rightarrow \infty$ ,  $|\mu_{t+1} - \mu_t| \rightarrow 0$  and  $\mu_t \rightarrow \mu^* \in (0, \infty]$  (where the fact that  $\mu^* > 0$  follows from Part 1, since  $\mu_{t+1} \geq \mu_t$  and  $\mu_t > 0$  for some  $t$ ). Therefore,  $(\mu_t - \mu_{t-1})/\mu_t \rightarrow 0$ , and distortions disappear asymptotically.

**Part 3:** Suppose that  $\varphi > \delta$ . In this case, (31) implies that  $\mathcal{U}_{C_t}^*$  is proportional to  $\varphi > \delta$  as  $t \rightarrow \infty$ . This implies that  $(\mu_t - \mu_{t-1})/\mu_t > 0$  as  $t \rightarrow \infty$ , so from (27) and (28), aggregate distortions cannot disappear, completing the proof. ■

The first part of the theorem states that the sustainability constraint of the politician, (13), necessarily introduces a distortion.<sup>28</sup> Intuitively, this additional aggregate distortion results

<sup>28</sup>Because of the ex ante utility function of the citizens,  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$ , is nonseparable, this distortion is not necessarily present at  $t = 0$ .

because, as output increases, the sustainability constraint (13) requires that more be given to the politicians in power and this increases the effective cost of production. The best SPE creates distortions so as to reduce the level of output and thus the rents that have to be paid to the politician.

The most important results are contained in parts 2 and 3 of the theorem.

Part 2 states that as long as an interior steady state exists and  $\mathcal{U}_{C_t}^*$  declines sufficiently rapidly (which is related to the rate of discounting by the citizens, see below), the multiplier of the sustainability constraint goes to zero. This result is important as it implies that in the long run there will be “efficient” provision of rents to politicians, with the necessary tax revenues raised without distortions. Intuitively, current incentives to the politician are provided by both consumption in the current period,  $x_t$ , and by consumption in the future. Future consumption by the politician not only relaxes the sustainability constraint in the future but does so in all prior periods as well. Thus, all else equal, optimal incentives for the politician should be backloaded. Backloading leads to sustainability constraint not binding in the long run and distortions disappearing.

Notice that the results in this theorem compare  $\delta$  to  $\varphi$ , which is the rate at which the ex ante marginal utility of consumption  $\mathcal{U}_{C_t}^*$  is declining in the steady state. Clearly, in the case where  $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$  is time separable, the rate at which  $\mathcal{U}_{C_t}^*$  declines is exactly equal to  $\beta$ . We will show that in an important special case that we consider in the next section this is indeed the case. In the more general case of present section,  $\varphi$  is the “fundamental discount factor” of the citizens, since it measures how one unit of resources at time  $t$  compares with one unit of resources at time  $t + 1$  (from the viewpoint of  $t = 0$ ). Only in special cases (e.g., without any dynamic incentive linkages) does this fundamental discount factor coincide with  $\beta$ . Therefore, the case of  $\varphi = \delta$  indeed corresponds to a situation in which the politician is as patient as the citizens.<sup>29</sup>

Part 3, on the other hand, states that if the discount factor of the politician  $\delta$  is sufficiently low compared to the fundamental discount factor  $\varphi$ , then aggregate distortions will not disappear, even asymptotically. The significance of this result is that it also implies *positive aggregate capital taxes* in contrast to the existing literature on dynamic fiscal policy. Since in many realistic political economy models politicians are—or act as—more short-sighted than the citizens, this part of the theorem implies that in a number of important cases, political economy considerations will lead to additional distortions that will not disappear even asymptotically.

---

<sup>29</sup>In part 2 of this theorem, we limit attention to the case in which  $\varphi = \delta$ , since when  $\varphi > \delta$  we will not converge to an interior steady state. Theorem 4 in the next section explicitly deals with non-interior steady states.

## 7 Private Histories

In this section we consider an important special case that allows us both to strengthen the results of the previous section and also to present them in a way that makes interpretation slightly easier. This special case involves restriction to *private histories*, that is, we assume that individual histories are not observed by the politicians and therefore current publications can only be conditioned on current reports. To further simplify the notation in this case, let us also assume that there is no capital, so that the aggregate production function of the economy is

$$Y_t = L_t, \quad (33)$$

where  $K_0 = 0$  and  $L_t$  denotes the aggregate labor supply at time  $t$ .

The restriction to private histories implies that in admissible mechanisms, allocations must depend only on agents' current report. In such an environment the incentive compatibility constraints for agents can be separated across time periods, and written as

$$u(c_t(\theta_t), l_t(\theta_t) \mid \theta_t) \geq u(c_t(\hat{\theta}_t), l_t(\hat{\theta}_t) \mid \theta_t) \quad (34)$$

for all  $\hat{\theta}_t \in \Theta$  and  $\theta_t \in \Theta$ , and for all  $t$ . Moreover, given the single crossing property in Assumption 3, (34) can be reduced to a set of incentive compatibility constraints only for neighboring types. Since there are  $N + 1$  types in  $\Theta$ , this implies that (34) is equivalent to  $N$  incentive compatibility constraints. The best sustainable mechanism with private histories maximizes (10) subject to (13), (34) and the resource constraint

$$C_t + x_t \leq L_t. \quad (35)$$

Recall now the quasi-Mirrlees program defined above. It is straightforward to see that because of “private histories”, the optimal allocations of  $(c_t, l_t)$  depend only on the aggregate variables in the same period,  $C_t$  and  $L_t$ , and are independent of any  $C_s, L_s$  with  $s \neq t$ . This implies that  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$  is time separable, i.e.,  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty) = \mathbb{E} \sum_{t=0}^\infty \beta^t U(C_t, L_t)$  for some real-valued differentiable function  $U : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ . The program for the best sustainable mechanism, (21)-(22), therefore becomes:

$$\max_{\{C_t, L_t, x_t\}_{t=0}^\infty} \mathbb{E} \sum_{t=0}^\infty \beta^t U(C_t, L_t) \quad (36)$$

subject to the resource constraint, (35), and the sustainability constraint,

$$w_t \equiv \mathbb{E} \left[ \sum_{s=0}^\infty \delta^s v(x_{t+s}) \right] \geq v(L_t), \quad (37)$$

for all  $t$ , where  $w_t$  denotes the present value of utilities delivered to the politician at time  $t$ .

The incentive compatibility constraints for individuals in (34) play a similar role to (15) in our formulation above. In particular, we can define

$$\Lambda = \left\{ (C, L) \text{ such that } \exists \left\{ \{c_t(\theta), l_t(\theta)\}_{\theta \in \Theta} \right\}_{t=0}^{\infty} \text{ satisfying (34), and} \right. \quad (38)$$

$$\left. C = \int c(\theta) dG(\theta), \text{ and } L = \int l(\theta) dG(\theta) \right\}.$$

We next adopt the following sustainability assumption. This assumption is used only in part 2 of the next theorem to characterize the equilibrium when the utility provided to a politician reaches the boundary of the set of feasible values and enables us to obtain sharper results than in Theorem 3.<sup>30</sup> Let  $\bar{w} \equiv \max_{(C,L) \in \Lambda} v(L - C) / (1 - \delta)$ .

**Assumption 6 (sustainability)** *There exists  $(\bar{C}, \bar{L}) \in \arg \max_{(C,L) \in \Lambda} v(L - C) / (1 - \delta)$ , such that  $v(\bar{L} - \bar{C}) / (1 - \delta) > v(\bar{L})$ .*

This assumption ensures that the highest discounted utility that can be given to the politician is sufficient to satisfy its sustainability constraint (37). Clearly this assumption is satisfied if the discount factor of the politician,  $\delta$ , is sufficiently large.

Following the general case, it is easy to extend the analysis and introduce randomization to show that  $U(C, L)$  is well defined, concave, and differentiable.

The concept of the aggregate distortion is also simpler in this setup. When  $(C, L) \in \text{Int}\Lambda$  the solution to the dynamic (full-commitment) Mirrlees program (19)-(20) satisfies:

$$U_C(C, L) = -U_L(C, L), \quad (39)$$

where  $U_C$  and  $U_L$  are the partial derivatives of  $U(C, L)$  with respect to  $C$  and  $L$ . We refer to a *downward labor distortions* if the left-hand side of (39) is strictly greater than the right-hand side.

The next proposition illustrates the relationship between distortions, taxes, and this condition more explicitly:

**Proposition 2** *Suppose Assumptions 1-2 hold. Consider a sequence of  $\{C_t, L_t\}_{t=0}^{\infty}$ . Then:*

1. *the marginal labor tax rate on the highest type of agent,  $\theta_N$ , at time  $t$  is given by  $\tau_{N,t} = 1 + U_L(C_t, L_t) / U_C(C_t, L_t)$ .*
2. *if  $\{C_t, L_t\}_{t=0}^{\infty}$  is undistorted at  $t$ , the labor supply decision of the highest type of agent is undistorted, i.e.,  $u_c(c_t(\theta_N), l_t(\theta_N) \mid \theta_N) = -u_l(c_t(\theta_N), l_t(\theta_N) \mid \theta_N)$ .*

---

<sup>30</sup>This set of feasible values is described in greater detail in Appendix A.

**Proof.** Assumption 2 implies that we only need to check incentive compatibility constraints for neighboring types. Let  $u_c$  and  $u_l$  be the partial derivatives of  $u$  (which exist by Assumption 1). Therefore, we have

$$\begin{aligned} u_c(c_t(\theta_N), l_t(\theta_N) | \theta_N)(1 + \lambda_{Nt}) &= \nu_{Ct}, \\ u_l(c_t(\theta_N), l_t(\theta_N) | \theta_N)(1 + \lambda_{Nt}) &= -\nu_{Lt}, \end{aligned}$$

where  $\lambda_{Nt}$  is the multiplier on incentive compatibility constraint between types  $\theta_N$  and  $\theta_{N-1}$  at time  $t$ ,  $\nu_{Ct}$  is the multiplier on (17) at  $t$  and  $\nu_{Lt}$  is the multiplier on (18) at  $t$ . By the differentiability of  $U(C, L)$  and the definition of Lagrange multipliers,  $\nu_{Ct} = U_C(C_t, L_t)$  and  $\nu_{Lt} = -U_L(C_t, L_t)$ . Combining these equations, we have

$$-\frac{u_l(c_t(\theta_N), l_t(\theta_N) | \theta_N)}{u_c(c_t(\theta_N), l_t(\theta_N) | \theta_N)} = (1 - \tau_{N,t}) = -\frac{U_L(C_t, L_t)}{U_C(C_t, L_t)},$$

where the first equality defines  $\tau_{N,t}$ , and the second equality establishes the first part of the lemma. The second result follows immediately from setting  $U_L(C_t, L_t) = -U_C(C_t, L_t)$  from the definition of an undistorted sequence, in particular, equation (39). ■

The main result of this section is the following theorem:

**Theorem 4 (*Best Sustainable Mechanisms with Private Histories*)** *Consider the economy with no capital and with private histories and suppose that Assumptions 1- 6 hold.*

1. *At  $t = 0$ , there is an aggregate distortion.*
2. *Suppose that  $\beta \leq \delta$ . Let  $\Gamma^*$  be the best sustainable mechanism inducing a sequence of values  $\{w_t\}_{t=0}^\infty$ . Then  $\{w_t\}_{t=0}^\infty$  is a non-decreasing sequence in the sense that  $w_{t+1} \geq w_t$  for all  $t$ . Moreover, a steady state exists in that  $\{w_t\}_{t=0}^\infty$  converges (almost surely) to some  $w^* \in [0, \bar{w}]$  and  $\{C_t, L_t, x_t\}_{t=0}^\infty$  converges (almost surely) to some  $(C^*, L^*, x^*)$ , which is asymptotically undistorted.*
3. *If  $\beta > \delta$ , then aggregate distortions do not disappear even asymptotically.*

**Proof.** Most of the results in this theorem follow as corollaries of the corresponding results in Theorem 3. The three additional results are that there are distortions at the initial date,  $t = 0$ , rather than at some possible future date, that  $\{w_t\}_{t=0}^\infty$  is a non-decreasing, and that when  $\delta \leq \beta$  a steady state necessarily exists. All three of these results follow from Theorem 1 and 2 in Acemoglu, Golosov, and Tsyvinski (2007), and we do not repeat these proofs to economize on space. ■

Theorem 4 provides a tighter characterization of the best sustainable equilibrium for this special case than the general results in Theorem 3. It also enables us to see the role of the relative discount factors of the politicians and the citizens more clearly.

More specifically, Part 1 of Theorem 4 establishes that there is distortion in period 0, rather at some period  $t \geq 0$ . It is possible to compare the discount factor of the politician  $\delta$  to the discount factor of the agent as function  $U(C, L)$  is separable across time. We show that a sequence of values delivered to politicians,  $\{w_t\}_{t=0}^{\infty}$  is non-decreasing providing an easily interpretable notion of backloading of incentives for politicians. The theorem also does not require existence of the interior steady state. Assumption 6 guarantees that if the boundary  $\bar{w}$  is reached the allocation will be undistorted. Finally, Part 2 of the Theorem extends results for the case of politicians being more patient than agents.

## 8 Example for History-Dependent Mechanisms

We now briefly illustrate the results of Theorem 3 and show how in some simple cases,  $\varphi$  defined as  $\inf\{\varrho \in (0, 1] : \text{plim}_{t \rightarrow \infty} \varrho^{-t} U_{C_t}^* = 0\}$  is again equivalent to the discount factor of the agents,  $\beta$ . In particular, let us consider the following economy without capital and with “almost constant types” as explained below. There are two types  $\Theta = \{\theta_0, \theta_1\}$  and the utility function is

$$u(c, l | \theta) = u(c) - \chi(l/\theta),$$

where  $u$  is continuously differentiable, increasing and strictly concave and  $\chi$  is continuously differentiable, increasing and strictly convex. Furthermore, suppose that  $u$  satisfies Inada-type conditions, so that first-order conditions are always satisfied as equality. We take  $\theta_0 = 0$ , so that the low type is again disabled and cannot supply any labor. Suppose that with probability  $\pi$  an individual is born as high type, and remains so with (iid) probability  $1 - \varepsilon$  in every period. With probability  $1 - \pi$ , individual is born as low type, and remains low type forever. By almost constant types, we mean the limit of this economy as  $\varepsilon \rightarrow 0$ . Then the quasi-Mirrlees formulation can be written as

$$\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) \equiv \max_{\{c_t(\theta_0), c_t(\theta_1), l_t(\theta_1)\}_{t=0}^{\infty}} \pi \sum_{t=0}^{\infty} \beta^t [u(c_t(\theta_1)) - \chi(l_t(\theta_1)/\theta_1)] + (1 - \pi) \sum_{t=0}^{\infty} \beta^t [u(c_t(\theta_0))], \quad (40)$$

subject to  $\pi c_t(\theta_1) + (1 - \pi) c_t(\theta_0) \leq \pi l_t(\theta_1) - x_t$  for all  $t$ , and

$$\sum_{t=0}^{\infty} \beta^t [u(c_t(\theta_1)) - \chi(l_t(\theta_1)/\theta_1)] \geq \sum_{t=0}^{\infty} \beta^t [u(c_t(\theta_0))],$$

where  $L_t = \pi l_t(\theta_1)$  and  $C_t = L_t - x_t$ . The first constraint is the resource constraint for each  $t$ , while the second constraint is the incentive compatibility constraint sufficient for the high

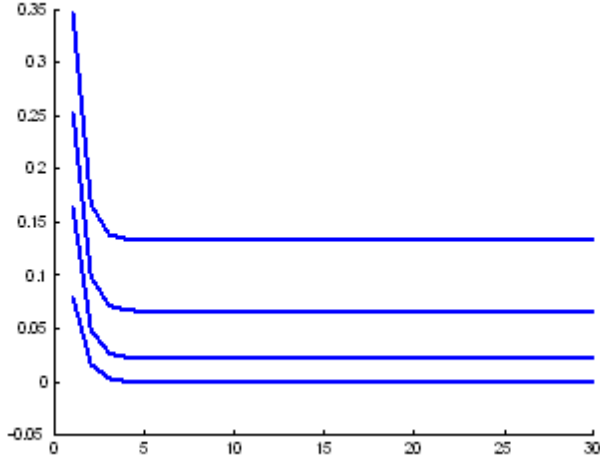


Figure 1: The time path of distortions for almost constant types with  $\beta = 0.9$  and  $\delta = 0.6, 0.7, 0.8, 0.9$ . Higher curves correspond to lower values of  $\delta$ .

type to reveal his identity given the presence of effective commitment along the equilibrium path. Because  $\varepsilon \rightarrow 0$ , we do not specify other incentive compatibility constraints. Assigning Lagrange multipliers  $\lambda$  and  $\beta^t \mu_t$  to these constraints, the first-order necessary conditions of this problem can be written as:

$$(\pi + \lambda) u'(c_t(\theta_1)) = \pi \mu_t \quad (41)$$

$$(1 - \pi - \lambda) u'(c_t(\theta_0)) = (1 - \pi) \mu_t, \quad (42)$$

and

$$\frac{(\pi + \lambda)}{\theta_1} \chi'(l_t(\theta_1)/\theta_1) = \mu_t. \quad (43)$$

Equations (41)-(42) imply that

$$\frac{u'(c_t(\theta_1))}{u'(c_t(\theta_0))} = \frac{(1 - \pi - \lambda)}{(\pi + \lambda)}.$$

Consequently, there is constant risk-sharing between the two types in all periods. Moreover, if a steady state exists, so that  $x_t \rightarrow x^*$ , (41)-(43) combined imply that  $c_t(\theta_1) \rightarrow c^{1*}$ ,  $c_t(\theta_0) \rightarrow c^{0*}$ , and  $l_t(\theta_1) \rightarrow l^*$ , and hence  $\mu_t \rightarrow \mu^*$ . Consequently, in this case  $\varphi = \beta$ , so Theorem 3 applies in exactly the same form as Theorem 4. Therefore, in this particular case, the rate at which the derivative  $\mathcal{U}_{C_t}^*$  declines is easy to determine, and it does so at the same rate as the discount factor of the citizens, i.e.,  $\varphi = \beta$ . It is also straightforward to see that the same argument generalizes to the case where there are more than two types.

We now numerically illustrate the above obtained theoretical results. We consider an economy with two “almost constant” types, i.e.,  $\Theta = \{\theta_0, \theta_1\}$ , and individual utility functions given by

$$u(c, l | \theta) = \sqrt{c} - \frac{l^2}{5\theta}. \quad (44)$$

Suppose that type  $\theta_0$  is disabled and cannot supply any labor, so  $\theta_0 = 0$ , and we normalize  $\theta_1 = 1$ . Let us also assume that a fraction  $\pi = 1/2$  of the population is of type  $\theta_1$  and that the utility function of the politician is given by  $v(x) = \sqrt{x}$ . We consider the case without capital, so that the production function is  $F(K, L) = L$ .

We show the aggregate distortion,  $1 + \mathcal{U}_{L_t}/\mathcal{U}_{C_t}$ , in Figure 1 for politician for the baseline case, with  $\delta = 0.9$ , and also for a range of lower discount factors for the politician,  $\delta = 0.8, 0.7$ , and  $0.6$ .

Consistent with part 2 of Theorem 3, when  $\beta = \delta = 0.9$ , the lowest curve shows that the aggregate distortion converges to zero and the convergence is again rather fast. Instead, when  $\delta < \beta$ , the aggregate distortion converges to a positive, and potentially large, asymptotic value. For example, when  $\delta = 0.6$ , the aggregate distortion converges to an asymptotic value of 0.15 (the highest curve in the graph).

Another question concerns how much of the economy’s output has to be allocated to the politician (as rents or government consumption). Figure 2 answers this question, again for  $\beta = 0.9$  and  $\delta = 0.9, 0.8, 0.7$ , and  $0.6$ . When the politician’s discount factor is equal to that of the citizens, he receives a very small fraction of the output even in the asymptotic equilibrium. As we consider lower discount factors for the politician, his temptation to deviate increases and consequently, he receives a higher fraction of the output. But even with  $\delta = 0.6$ , this is only 16% of total output.

## 9 Benevolent Time-Inconsistent Governments

The analysis so far considered only the case when politicians were purely self-interested. Although this case is of relevance for many political economy applications, it is also important to understand how the results generalize to the case considered by Roberts (1984), Freixas, Guesnerie and Tirole (1985), or Bisin and Rampini (2005), where the government is still benevolent, but “time inconsistent”, i.e., unable to commit to a full dynamic mechanism. To do this, we now consider a more general utility function for the government of the form:

$$\sum_{s=0}^{\infty} \delta^s \left[ (1-a)v(x_{t+s}) + a \left( \mathbb{E}_{t+s} \int u(c_{t+s}, l_{t+s} | \theta^{t+s}) dG^{t+s}(\theta^{t+s}) \right) \right], \quad (45)$$

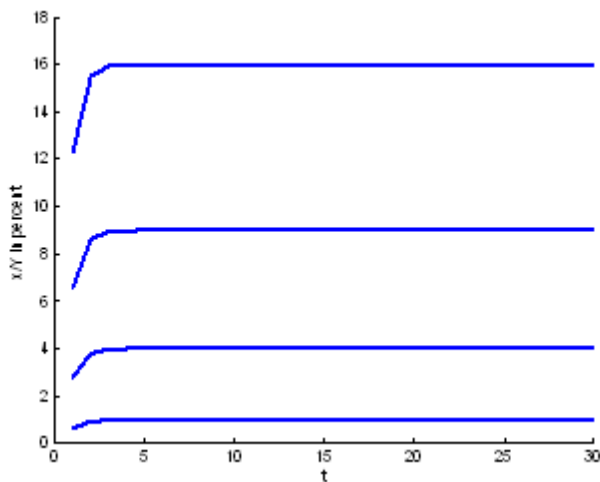


Figure 2: Time path of  $x_t/Y_t$  with  $\beta = 0.9$  and  $\delta = 0.6, 0.7, 0.8, 0.9$ . Higher curves correspond to lower values of  $\delta$ .

where the second term is the average (expected) utility of the citizens at time  $t + s$ , and  $0 < a < 1$ .<sup>31</sup> Therefore, this utility function is identical to that of a purely-self-interested government when  $a = 0$ .

To incorporate the case of a benevolent time-inconsistent government, we also need to change the political game. In particular, we no longer allow citizens to vote the current government out of office. Instead, the same government is always in power. Despite this, all of the main results so far continued to hold, since citizens have another effective punishment against the government, to produce zero. In particular, if the government deviates from the prescribed policy (or from the implicitly-agreed social plan), individuals can claim to be the worst type from then on and produce zero output. It can be verified that all of the results so far hold under this alternative game form (see Acemoglu, Golosov and Tsyvinski, 2006). The advantage of this alternative game form is that it naturally adapts to the case of a partly benevolent government. In addition, we also need we need to strengthen Assumption 1 and assume separable utility, which is a standard assumption in most analyses of dynamic taxation (e.g., Golosov, Kocherlakota and Tsyvinski, 2003, Kocherlakota, 2005).

**Assumption 1' (separable utility)**  $u(c, l | \theta) = u(c) - \chi(l | \theta)$ , where  $u : \mathbb{R}_+ \rightarrow \mathbb{R}$  is continuously differentiable, strictly increasing and concave, and  $\chi(\cdot | \theta)$  is continuously differentiable, strictly increasing and convex for all  $\theta \in \Theta$ , and satisfies  $\chi(0 | \theta) = 0$  for all  $\theta \in \Theta$ .

<sup>31</sup>We do not allow the case of  $a = 1$  (purely benevolent government) as it requires a different proof strategy. Our environment obviously allows the government to be arbitrarily benevolent ( $a \rightarrow 1$  and  $\delta = \beta$ )

The next theorem shows that Theorem 1 and Proposition 1 continue to hold in this more general environment.

**Theorem 5** *Suppose that government utility is given by (45) and that Assumptions 1', 2, 3, 4 and 5 hold. Then for any combination of strategy profiles  $\Gamma$  and  $\underline{\alpha}$  that support a sustainable mechanism, there exists another pair of equilibrium strategy profiles  $\Gamma^*$  and  $\underline{\alpha}^* = (\alpha^* | \alpha')$  for some  $\alpha'$  such that  $\Gamma^*$  induces direct submechanisms,  $\underline{\alpha}^*$  induces truth telling along the equilibrium path, and  $\underline{c}[\Gamma, \underline{\alpha}] = \underline{c}[\Gamma^*, \underline{\alpha}^*]$ ,  $l[\Gamma, \underline{\alpha}] = l[\Gamma^*, \underline{\alpha}^*]$  and  $x[\Gamma, \underline{\alpha}] = x[\Gamma^*, \underline{\alpha}^*]$ . Moreover, the best sustainable mechanism is a solution to maximizing (10) subject to (11), (12) and the government sustainability constraint:*

$$\begin{aligned} & \sum_{s=0}^{\infty} \delta^s [(1-a)v(x_{t+s}) + \\ & a \left( \mathbb{E}_{t+s} \int [u(c(\theta^{t+s})) - \chi(l(\theta^{t+s}) | \theta_{t+s})] dG^{t+s}(\theta^{t+s}) \right)] \geq \\ & \max_{\tilde{x}'_t + \int \tilde{c}'_t(\theta^t) dG^t(\theta^t) \leq F(K_t, L_t)} (1-a)v(\tilde{x}_t) + a \int u(\hat{c}_t(\theta^t)) dG^t(\theta^t), \end{aligned} \quad (46)$$

for all  $t$ .

**Proof.** See Appendix B. ■

The difference in the proof with the previous environment is that instead of replacing politicians, now agents use the null strategy following the deviation by a politician. In particular imagine that the government has undertaken a deviation in which it has used some of its past information in order to improve the ex post allocation of resources. This could clearly be desirable given the utility function of the government in (45), but as illustrated with the Roberts' (1984) example, it may have very negative consequences ex ante. Therefore, the best sustainable mechanism will have to discourage such deviations. To do this, imagine a punishment strategy, in which following any type of deviation, all individuals supply zero labor. To establish Theorem 5, all we need to show is that such punishment strategies are sequentially rational. When all other agents choose zero labor supply, following any deviation to positive labor supply, the government would consume some of the increase in output itself, and would redistribute the rest equally among all agents given the separable utility function assumed in Assumption 1'. Since there is a very large number of citizens, this implies the deviating individual will receive no additional consumption from supplying positive labor, and thus it is sequentially rational for all citizens to supply zero labor following a deviation by the government.

This theorem therefore shows that revelation principle applies to the case of benevolent, but time-inconsistent governments as well, though under the additional assumption of Assumption 1'. The next example shows why this assumption is necessary:

**Example 1** To avoid issues of deviation among continuum of agents, let us consider a finite economy with  $n$  agents for this example, where  $n$  is large (exactly the same example can be constructed in an economy with a continuum of agents). There are two types of agents,  $\theta \in \{0, 1\}$ , with  $\theta = 0$  corresponding to the disabled type, who can only supply  $l = 0$ , and has utility  $u(c, \cdot | \theta = 0) = u(c)$ , while the utility of type  $\theta = 1$  is  $u(c, l | \theta = 1) = u(c - \chi_1(l))$ , where with  $\chi_1(\cdot)$  strictly increasing in  $l$ . Furthermore, suppose that aggregate output is linear in labor and that the government is fully benevolent, i.e.,  $a = 1$  in terms of the utility function in (45). Now imagine the economy has entered the punishment phase where each citizen is supposed to supply  $l = 0$  and consume  $c = 0$ . Consider a deviation by an agent,  $i'$ , of type  $\theta = 1$  to  $l' > 0$  such that  $\chi_1(l') < 1$ . Following this deviation, the benevolent planner will distribute consumption (output  $l' > 0$ ) to maximize its own utility, which involves maximizing average utility of the citizens, thus equating the marginal utility of consumption across agents, i.e.,

$$u'(c_i) = u'(c_{i'} - \chi_1(l')) \text{ for all } i \neq i'$$

thus,  $c_{i'} = c_i + \chi_1(l')$  for all  $i \neq i'$ . The resource constraint is  $(n - 1)c_i + c_{i'} = l'$ , or  $c_i = (l' - \chi_1(l'))/n$  and  $c_{i'} = (l' - \chi_1(l'))/n + \chi_1(l')$ . The resulting utility of individual  $i'$  is

$$u((l' - \chi_1(l'))/n) > u(0),$$

for any  $n$ , thus giving him greater utility than supplying zero labor. This proves that the punishment phase where each citizen is supposed to supply zero labor is not sequentially rational and thus cannot be part of a (Perfect Bayesian) equilibrium with this utility function.

The next theorem provides a generalization of Theorem 3 for the most-commonly studied case where types are constant (in our context, types are “almost constant” as in subsection 8) and  $\beta = \delta$ .<sup>32</sup>

**Theorem 6** *Suppose that government utility is given by (45) with  $a \in (0, 1)$  and that Assumptions 1', 2, 3, 4 and 5 hold. Furthermore, assume that there are (almost) constant types,  $\beta = \delta$  and  $au'(0) \neq (1 - a)v'(0)$ . Then, asymptotically there are no aggregate distortions on labor supply and capital accumulation.*

---

<sup>32</sup>Once again with the game form in Remark 1, this theorem can be stated for constant types.

**Proof.** See Appendix B. ■

This theorem implies that in an economy with (almost) constant types, aggregate distortions disappear irrespective of the degree of benevolence of the government. Consequently, there will be no aggregate capital taxes and no further taxes on labor beyond those implied by the full-commitment Mirrlees economy.<sup>33</sup> In the case where  $a \rightarrow 1$ , the government is arbitrarily close to the fully-benevolent case, and the theorem contrasts with the results in Roberts (1984), where in a very similar environment, the equilibrium always involved extreme distortions. Once again, the main source of the difference is the infinite-horizon nature of our economy, which allows us to construct equilibria in which the government will be punished if it exploits the information it gathers via the earlier submechanisms.

## 10 Conclusions

In this paper, we take a first step towards a political-economic analysis of dynamic and non-linear taxation. Importantly, we provide a tractable framework to analyze issues of non-commitment, self interest of policymakers, and provision of incentives.

We focus on the best sustainable equilibrium, i.e., the best equilibrium that satisfies the incentive compatibility constraints of politicians. Given the infinite horizon nature of the environment in question, we can construct *sustainable mechanisms* where the politician in power is given incentives not to misuse resources and information. An important result of our analysis is the *revelation principle along the equilibrium path*, which shows that truth-telling mechanisms can be used despite the commitment problems and the different interests of the government (politicians) and the citizens. Using this tool, we provide a characterization of the best sustainable mechanism. Political economy considerations introduce additional constraints on the optimal taxation problem, but these constraints are relatively simple. In particular, we show that the provision of incentives to politicians can be separated from the provision of incentives and insurance to agents. Political economy constraints, instead, take the form of additional constraints on aggregate consumption and labor supply in the economy. These constraints then lead to new (political economy) distortions and thus change the structure of taxation. Our analysis provides a characterization of these distortions and their evolution over time. We show that when politicians are as patient as, or more patient than, citizens, aggregate capital and labor distortions disappear in the long run. The politician in power still receives rents, but these rents are provided without additional distortions. This result therefore implies that the insights from Mirrlees' classical analysis and from the more recent

---

<sup>33</sup>The assumption that  $au'(0) \neq (1-a)v'(0)$  rules out a special case in which our method of proof does not work (though other more complicated approaches may work even without this assumption).

dynamic taxation literature may generalize to certain environments featuring political economy constraints and commitment problems. However, we also show that when politicians are less patient than the citizens, aggregate distortions remain positive even asymptotically. In this case, in contrast to the classical results in optimal taxation, there will be positive distortions and positive aggregate capital taxes even in the long run.

Our analysis relies on the infinite horizon nature of the economy and especially on the infinite planning horizon of the politicians. Nevertheless, we believe that similar insights apply even when politicians have finite horizons, and a detailed investigation of this issue would be an interesting area for future research. For example, we conjecture that in a model with either finitely-lived politicians or with term limits, a society consisting of infinitely-lived citizens or overlapping generations of citizens will be able to commit to providing a continuation value (e.g., “pension”) to politicians that have not deviated from the social plan. In this case, even though distortions will not disappear in the long run, they will decline during the tenure of the politician. Such a model would also enable an analysis of the effects of term limits and other realistic institutional constraints on politicians.

The results in this paper can also be extended to an environment with competing parties or interest groups. For example, in Acemoglu, Golosov and Tsyvinski (2007), we consider a model in which political power fluctuates between different parties. We show that distortions decline as a particular party remains in power for longer and they increase when power switches to a new party (see also Dixit, Grossman and Gul, 2000). Another interesting area for future work may be to extend the analysis to different types of government intervention in the economy. For example, an important role that governments play in practice is contract enforcement. However, the power delegated to governments to enforce contracts can be misused in the same way as their taxation powers are potentially misused in this paper. A similar analysis might reveal what types of constraints political economy considerations will place on equilibrium contracting institutions.

## 11 Appendix A: Properties of $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$

### 11.1 Some Technical Results

We first present some technical results that will be useful in establishing the properties of the functional  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$ .

**Definition 5** *Let  $X$  and  $Z$  be Banach spaces and  $G : X \rightarrow Z$  be a vector-valued mapping. Suppose that  $G$  is continuously (Fréchet) differentiable in the neighborhood of  $x_0$  with the derivative denoted by  $G'(x_0)$ . Then  $x_0$  is said to be a regular point of  $G$  if  $G'(x_0)$  maps  $X$  onto  $Z$ .*

**Lemma 2** *Let  $X$  and  $Z$  be Banach spaces. Consider the maximization problem of*

$$P(u) = \max_{x \in X} f(x) \quad (47)$$

*subject to*

$$g_0(x) \leq u \quad (48)$$

*and*

$$G(x) \leq \mathbf{0} \quad (49)$$

*where  $f : X \rightarrow \mathbb{R}$  and  $g_0 : X \rightarrow \mathbb{R}$  are real-valued functions and  $G : X \rightarrow Z$  is a vector-valued mapping and  $\mathbf{0}$  is the zero of the Banach space  $Z$ . Suppose that  $f$  is concave and  $g_0$  is convex, and moreover that the solution at  $u = 0$ ,  $x_0$ , is a regular point. Let  $\mu$  be any multiplier of (48). Then  $\mu$  is a subgradient of  $P(0)$ .*

**Proof.** This lemma is a direct generalization of Proposition 6.5.8 of Bertsekas, Nedic and Ozdaglar (2003, p. 382) to an infinite dimensional maximization problem. ■

**Theorem 7** *Let  $X$  and  $Z$  be Banach spaces. Consider the maximization problem of*

$$P(\mathbf{u}) = \max_{x \in X} f(x)$$

*subject to*

$$G(x) \leq \mathbf{0} + \mathbf{u}$$

*where  $f : X \rightarrow \mathbb{R}$  is a real-valued concave function and  $G : X \rightarrow Z$  is a convex vector-valued mapping and  $\mathbf{0}$  is the zero of the vector space  $Z$  and  $\mathbf{u}$  is a perturbation. Suppose that  $x_0$  is a solution to this program. Suppose also that  $x_0$  is a regular point of  $G$  and that  $f$  and  $G$  are continuously (Fréchet) differentiable in the neighborhood of  $x_0$ . Then  $P(\mathbf{0})$  is differentiable.*

**Proof.** From Lemma 2, it follows that if there is a unique multiplier,  $P$  has a unique subgradient and is thus differentiable. Proposition 4.47 in Bonnans and Shapiro (2000) establishes that under a weaker constraint qualification condition than regularity, this problem has a unique multiplier. ■

**Theorem 8** *Let  $X$  be a compact metric space, then  $\mathbf{M}(X)$  is a compact metric space with the weak topology.*

**Proof.** See Parthasarathy (1967, p. 45). ■

## 11.2 Randomizations

We next introduce randomizations to show concavity and differentiability of  $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ . To simplify notation, in this appendix, we suppress dependence on public histories  $h^t$ . The original maximization problem without randomization is to maximize (10) subject to (11), (12), and (13) as stated in Proposition 1. Recall also that  $\theta_t \in \Theta$ , where  $\Theta$  is a finite set (with  $N + 1$  elements). Therefore  $\Theta^t$  for any  $t < \infty$  is also a finite set. Consider next the functions  $c_t : \Theta^t \rightarrow \mathbb{R}_+$  and  $l_t : \Theta^t \rightarrow [0, \bar{l}]$ . By definition, these functions assign values to a finite number of points in the set  $\Theta^t$  for any  $t < \infty$ , thus can simply be thought of as vectors of  $(N(N + 1))^t$  dimension. Moreover

$$\int c_t(\theta^t) dG(\theta^t) \leq \bar{Y}, K_{t+1} \leq \bar{Y} \text{ and } x_t \leq \bar{Y}, \quad (50)$$

where  $\bar{Y} = F(\bar{Y}, \bar{l}) < \infty$ . Therefore,  $X_t = \{c_t(\theta^t), l(\theta^t), K_{t+1}, x_t\}$  is a vector (of dimension  $(N(N + 1))^{2t} + 2$ ). Let  $\mathbf{X}_t$  be the set of all such vectors that satisfy the inequalities in (50), and for  $X_t \in \mathbf{X}_t$ , let  $X_t(i)$  denote the  $i$ th component of this vector, and  $T_t$  be the dimension of vectors in the set  $\mathbf{X}_t$  (i.e.,  $T_t = (N(N + 1))^{2t} + 2$ ).  $\mathbf{X}_t$  is a compact metric space with the usual Euclidean distance metric,  $d_t(X_t, X') = \left(\sum_{i=1}^{T_t} (X'_t(i) - X_t(i))^2\right)^{1/2}$

Let us now construct the product space of the  $\mathbf{X}_t$ 's

$$\mathbf{X} = \prod_{t=1}^{\infty} \mathbf{X}_t$$

Clearly the sequence  $\{c_t(\theta^t), l_t(\theta^t), x_t, K_{t+1}\}_{t=0}^{\infty}$  must belong to  $\mathbf{X}$ . In fact, it must belong to the subset of  $\mathbf{X}$ , which satisfy (11), (12), and (13), denoted by  $\bar{\mathbf{X}}$ .

Now by Tychonoff's theorem (e.g., Dudley, 2002, Theorem 2.2.8),  $\mathbf{X}$  is compact in the product topology. Since (11), (12), and (13) are (weak) inequalities,  $\bar{\mathbf{X}}$  is a closed subset of  $\mathbf{X}$ , and therefore it is also compact in the product topology. Moreover,  $\mathbf{X}$  with the product topology is metrizable, with the metric

$$d(X, X') = \sum_{t=1}^{\infty} \phi^t d_t(X_t, X'_t) \quad (51)$$

for some  $\omega \in (0, 1)$  and  $X \equiv \{X_t\}_{t=0}^\infty \in \mathbf{X}$ . This shows that  $\mathbf{X}$  endowed with the product topology is a metric space, and so is  $\bar{\mathbf{X}}$ .

From Theorem 8, the set of probability measures defined over a compact metric space is compact in the weak topology. This establishes that the set of probability measures  $\mathcal{P}^\infty$  defined over  $\bar{\mathbf{X}}$  is compact in the weak topology.

We are concerned not with all probability measures, but those that condition at  $t$  on information revealed up to  $t$ . Let  $\mathcal{C} = \{(c, l) \in \mathbb{R}^2 : 0 \leq c \leq \bar{c}, 0 \leq l \leq \bar{l}\}$  be the set of possible consumption-labor allocations for agents, so that  $\mathcal{P}^\infty$  defined above is the set of all probability measures over  $\mathcal{C}^\infty$ . Now, for each  $t \in \mathbb{N}$  and  $\theta^{t-1} \in \Theta^{t-1}$ , let  $\mathcal{P}[\theta^{t-1}]$  be the space of  $N + 1$ -tuples of probability measures on Borel subsets of  $\mathcal{C}$  for an individual with history of reports  $\theta^{t-1}$ . Thus each element  $\zeta(\cdot | \theta^{t-1}) = [\zeta(\theta_0 | \theta^{t-1}), \dots, \zeta(\theta_N | \theta^{t-1})]$  in a  $\mathcal{P}^t[\theta^{t-1}]$  consists of  $N + 1$  probability measures for each type  $\theta_i$  given their past reports,  $\theta^{t-1}$ , and is thus closed. Consider  $\mathcal{P} \equiv \bigcup_{t \in \mathbb{N}} \bigcup_{\theta^t \in \Theta^t} \mathcal{P}^t[\theta^{t-1}]$ , which is a closed subset of  $\mathcal{P}^\infty$ . Since a closed subset of a compact space is compact (e.g., Dudley, 2002, Theorem 2.2.2),  $\mathcal{P}$  is compact in the weak topology.

Finally, choosing  $\phi \leq \beta$  in (51) shows that the objective function is continuous in the weak topology. This establishes that including randomizations, we have a maximization problem over probability measures in which the objective function is continuous in the weak topology, and the constraint set is compact in the weak topology, and thus there exists a probability measure that reaches the maximum.

### 11.3 Properties of $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$

We now established the main properties of  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$ . The only additional restriction is that in all the proofs we assume that the solution to the maximization problem (10) is at a regular point. This needs to be imposed an assumption, since it is impossible to check that the solution is indeed at a regular point. Nevertheless, this assumption is not a strong one, since if the solution is not at their regular point, a perturbation of the utility functions or the production function will ensure that the solution shifts to regular point (i.e., solutions that are not at regular points in this context are “non-generic,” though we do not present a precise mathematical statement of this property to economize on further notation and space).

**Lemma 3**  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$  is continuous and concave on  $\Lambda^\infty$ , nondecreasing in  $C_s$  and nonincreasing in  $L_s$  for any  $s$  and differentiable in  $\{C_t, L_t\}_{t=0}^\infty$ .

**Proof.** The above argument established that in the problem of maximizing (10) subject to (11), (12), and (13) over probability measures, a maximum exists and  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$  is

therefore well defined.

To show concavity, consider  $(C^0, L^0)$  and  $(C^1, L^1)$  and corresponding  $\zeta^0, \zeta^1$ . We have

$$\begin{aligned} & \int (u(c, l; \theta) - u(c, l; \hat{\theta})) \zeta^\alpha(d(c, l), \theta) \\ &= \alpha \int (u(c, l; \theta) - u(c, l; \hat{\theta})) \zeta^0(d(c, l), \theta) + (1 - \alpha) \int (u(c, l; \theta) - u(c, l; \hat{\theta})) \zeta^1(d(c, l), \theta) \\ &\geq 0 \end{aligned}$$

In a similar way we can show that  $\zeta^\alpha$  satisfies (11), (12), and (13), this convex combination is feasible and it gives the same utility as  $\alpha \zeta^0 \cdot u(\theta) + (1 - \alpha) \zeta^1 \cdot u(\theta)$ .

Next, note that the constraint set expands if  $C_s$  increases or  $L_s$  decreases for any  $s$ , therefore  $U$  must be weakly increasing in  $C_s$  and weakly decreasing in  $L_s$ .

Finally, returning to the original topology,  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$  is defined over a Banach space. Given the assumption that the solution to (10) is at their regular point, we can use Theorem 7 to conclude that  $\mathcal{U}(\{C_t, L_t\}_{t=0}^\infty)$  is differentiable in  $\{C_t, L_t\}_{t=0}^\infty$ , completing the proof. ■

**Lemma 4**  $\Lambda^\infty$  is compact and convex.

**Proof. Convexity:** Consider  $\{C_t, L_t\}_{t=0}^\infty$  and  $\{C'_t, L'_t\}_{t=0}^\infty \in \Lambda^\infty$  and some  $\zeta^0, \zeta^1$  feasible for  $\{C_t, L_t\}_{t=0}^\infty$  and  $\{C'_t, L'_t\}_{t=0}^\infty$  respectively. Now for any  $\alpha \in (0, 1)$   $\zeta^\alpha \equiv \alpha \zeta^0 + (1 - \alpha) \zeta^1$  is feasible for  $(\alpha \{C_t, L_t\}_{t=0}^\infty + (1 - \alpha) \{C'_t, L'_t\}_{t=0}^\infty)$ , so that this set is non-empty. Moreover, since  $\zeta^0, \zeta^1$  satisfy the incentive compatibility constraints,  $\zeta^\alpha$  satisfies it as well. Similarly,  $\zeta^\alpha$  satisfies the constraints on aggregate  $\{C_t, L_t\}_{t=0}^\infty$ .

**Compactness:** For any sequence  $\{C_t^n, L_t^n\}_{t=0}^\infty \in \Lambda^\infty, \{C_t^n, L_t^n\}_{t=0}^\infty \rightarrow \{C_t^\infty, L_t^\infty\}_{t=0}^\infty$ , there exists a sequence  $\{\zeta_t^n\}_{t=0}^\infty$  corresponding to  $\{C_t^n, L_t^n\}_{t=0}^\infty$ , such that  $\zeta^n \rightarrow \zeta^\infty$ , satisfying the incentive compatibility, aggregate constraints and feasibility, therefore  $\{C_t^\infty, L_t^\infty\}_{t=0}^\infty \in \Lambda^\infty$  is closed. Boundedness follows from boundedness of  $C$  and  $L$ . ■

## 12 Appendix B: Proofs for Section 9

### 12.1 Proof of Theorem 5:

The proof of this theorem follows the structure of the proofs of Lemma 1, Theorem 1 and Proposition 1. The main difference here is that instead of replacing the politician, citizens play a null strategy of supplying zero labor.

**Proof.** Define the following combination. Denote  $\tilde{c}'_t = \tilde{c}_t^{\prime 0}$  be the mapping that allocates zero consumption to all individuals irrespective of past and current reports. Let  $\tilde{h}^t = \hat{h}^t$  if  $\tilde{x}_{t-s}(\tilde{h}^{t-s}) = x_{t-s}(\hat{h}^{t-s})$  and  $\tilde{M}_{t-s} = M_{t-s}$  for all  $s > 0$ . Then the following strategy

combination would ensure  $v_t^c(\tilde{K}'_{t+1}, \tilde{c}'_t | \tilde{M}^t) = 0$  for all  $t$ : (1) for the citizens,  $\underline{\alpha} = (\tilde{\alpha} | \alpha^\emptyset)$ , for some  $\tilde{\alpha}$ , which means that for each citizen  $i$  and for all  $t$ , we have that if  $\tilde{h}^{t-1} = \hat{h}^{t-1}$ , then  $\alpha_t^i = \tilde{\alpha}$ , and if  $\tilde{h}^{t-1} \neq \hat{h}^{t-1}$ , then  $\alpha_t^i = \alpha^\emptyset$ ; (2) for the politician,  $\Gamma$ , such that if  $\tilde{h}^{t-1} = \hat{h}^{t-1}$ , then  $\Gamma$  involves  $\tilde{x}_t = x_t$ ,  $\tilde{M}_t = M_t$ , and  $\xi_t = 0$  for all  $\tilde{h}^t \in \tilde{H}^t$ ; and if  $\tilde{h}^{t-1} \neq \hat{h}^{t-1}$ , then it involves  $\xi_t = 1$ ,  $\tilde{x}'_t = F(K_t, L_t)$  for all  $\tilde{h}^t \in \tilde{H}^t$ , and  $\tilde{c}'_t = \tilde{c}_t^\emptyset$ .

A difference from the proof with Lemma 1 is that we need to show that there exists a sequentially rational continuation play in which all agents supply zero labor. Suppose that the government has announced a submechanism  $\tilde{M}_t$  at time  $t$  and has capital stock  $K_t$ , and  $\alpha_{t+s}^i = \alpha^\emptyset$  for all  $i \in [0, 1]$  and for all  $s \geq 0$ . We first show that a deviation by an individual,  $i'$  with type  $\theta_t^{i'} \neq \theta_0$  to some other strategy that involves supplying positive labor is not profitable (we think of an individual with positive measure  $\varepsilon$  deviating, and take the limit  $\varepsilon \rightarrow 0$ , since there is a continuum of agents). Without the deviation,  $i'$  obtains utility  $u(0)/(1-\beta)$  (since from Assumption 1',  $\chi(0|\theta) = 0$  for all  $\theta \in \Theta$  and there will be no labor supply for any type in the continuation game). Now imagine a deviation to a message that corresponds to positive labor supply, say  $l'$ , with  $\chi(l'|\theta_t^{i'}) > \chi(0|\theta_t^{i'}) = 0$  by definition. This will generate output  $F(K_t, \varepsilon l')$ , since all other agents are supplying zero labor. Now imagine the behavior of the government at the last stage of the game, conditional on  $\alpha_{t+s}^i = \alpha^\emptyset$  for all  $i \in [0, 1]$  and for all  $s \geq 1$ . Then the sequentially rational strategy of the government is to maximize (45) with  $K_{t+1} = 0$ , since there will be no production in future periods. Consequently, the utility-maximizing program of the government in the information set following the deviation is:

$$\max_{\tilde{x}'_t, \tilde{c}'_t} (1-a)v(x_t) + a \left( \int [u(\tilde{c}'_t(z^t(\alpha_t(\theta^t)))) - \chi(l_t(z^t(\alpha_t(\theta^t))) | \theta_t)] dG^t(\theta^t) \right),$$

subject to  $x_t + \int \tilde{c}'_t(z^t(\alpha_t(\theta^t))) dG^t(\theta^t) \leq F(K_t, \varepsilon l')$ , where recall that  $z^t(\alpha_t(\theta^t))$  is the history of reports up to time  $t$  by an individual of type  $\theta^t$  given strategy profile  $\underline{\alpha}$ . In view of Assumption 1', this expression is concave in  $c$  for any strategy profile  $\underline{\alpha}$ , so the optimal policy for the government in this information set is to redistribute consumption (what it does not consume itself) equally across agents, i.e.,  $\tilde{c}'_t(z^t(\alpha_t(\theta^t))) = c_t$  for all  $z^t(\alpha_t(\theta^t)) \in Z^t$ . This implies that as  $\varepsilon \rightarrow 0$ ,  $c_t \rightarrow 0$ , and thus the deviation payoff of  $i'$  is  $u(0) - \chi(l'|\theta_t^{i'}) + \beta(u(0) - \chi(0|\theta_t^{i'})) / (1-\beta) < (u(0) - \chi(0|\theta_t^{i'})) / (1-\beta)$ , showing that a continuation strategy profile where all agents supply zero labor is sequentially rational.

Now consider two different types of deviations by the government. First, imagine the government offers  $\tilde{M}_t \neq M_t$ , i.e., a different mechanism at the beginning of time  $t$  than the one implicitly agreed in the social plan  $(M, x)$ . Given the above-constructed continuation equilibrium,  $\alpha_{t+s}^i = \alpha^\emptyset$  for all  $i \in [0, 1]$  and for all  $s \geq 0$  is a best response against this

deviation. Since maximal punishments are optimal,  $\alpha_{t+s}^i = \alpha^\emptyset$  for all  $i \in [0, 1]$  and for all  $s \geq 0$  is optimal against this deviation, implying that such a deviation would never be profitable for the government.

Second, the government can deviate at the last stage of time  $t$ . Again  $\alpha_{t+s}^i = \alpha^\emptyset$  for all  $i \in [0, 1]$  and for all  $s \geq 1$  is the maximal sequentially rational punishment against such a deviation. Consequently, after any deviation by the government, there will not be any further production. Thus the optimal deviation for the government involves  $\tilde{K}'_{t+1} = 0$ , and again exploiting the concavity of the government's continuation payoff in  $c$ , the sustainability constraint is equivalent to:

$$\begin{aligned} & \mathbb{E}_t \sum_{s=0}^{\infty} \delta^s \left[ (1-a)v(x_{t+s}) + a\mathbb{E}_{t+s} \left( \int [u(c_t) - \chi(l_t(z^t(\alpha_t(\theta^t))) | \theta_t)] dG^t(\theta^t) \right) \right] \\ & \geq \max_{\tilde{x}'_t + \tilde{c}'_t \leq F(K_t, L_t)} (1-a)v(\tilde{x}_t) + a \int u(\tilde{c}'_t(\theta^t)) dG^t(\theta^t) \text{ for all } t. \end{aligned} \quad (52)$$

Now, given an equilibrium pair of strategy profiles  $\Gamma$  and  $\underline{\alpha}$ , exactly the same argument as in the proof of Theorem 1 implies that there exists another pair of equilibrium strategy profiles  $\Gamma^*$  and  $\underline{\alpha}^* = (\alpha^* | \alpha')$  for some  $\alpha'$  such that  $\Gamma^*$  induces direct submechanisms. Consequently, we can write (52), in terms of a direct mechanism, which gives (46).

Finally, the same argument as in the proof of Proposition 1 implies that the best sustainable mechanism is a solution to maximizing (10) subject to (11), (12), and the sustainability constraints of the government given by (46). ■

## 12.2 Proof of Theorem 6

**Proof.** Suppose again that there are  $N+1$  types, i.e.,  $\Theta = \{\theta_0, \theta_1, \dots, \theta_N\}$ , ranked in ascending order of skills, and with respective probabilities  $\{\pi_0, \pi_1, \dots, \pi_N\}$ . Given the assumptions of the theorem (and again suppressing  $h^t$ -dependence to simplify notation), we can write the program for the best sustainable mechanism as:

$$\max_{\{c_t(\theta_i), l_t(\theta_i)\}_{i=0, \dots, N}^N, x_t, K_{t+1}} \sum_{t=0}^{\infty} \beta^t \sum_{i=0}^N \pi_i [u(c_t(\theta_i)) - \chi(l_t(\theta_i) | \theta_i)]$$

subject to the constraints

$$\sum_{t=0}^{\infty} \beta^t [u(c_t(\theta_i)) - \chi(l_t(\theta_i) | \theta_i)] \geq \sum_{t=0}^{\infty} \beta^t [u(c_t(\theta_{i-1})) - \chi(l_t(\theta_{i-1}) | \theta_{i-1})] \quad (53)$$

for all  $i = 1, \dots, N$ ,

$$\sum_{s=0}^{\infty} \beta^{t+s} \left\{ (1-a)v(x_{t+s}) + a \left( \sum_{i=0}^N \pi_i [u(c_{t+s}(\theta_i)) - \chi(l_{t+s}(\theta_i) | \theta_i)] \right) \right\} \geq V(K_t, L_t) \quad (54)$$

for all  $t$ , and

$$x_t + K_{t+1} + \sum_{i=0}^N \pi_i u(c_t(\theta_i)) \leq F\left(K_t, \sum_{i=1}^N \pi_i l_t(\theta_i | \theta_i)\right) \quad (55)$$

for all  $t$ , and that  $c_t(\theta_i) \geq 0$  for all  $i$  and  $t$  and  $x_t \geq 0$  for all  $t$ .

The first set of constraints, (53), ensure incentive compatibility for the citizens. Given Theorem 5, there is truthful revelation along the equilibrium path. This, together with Assumption 2, implies that we only need one constraint for each type other than the disabled type,  $\theta_0$ , where type  $i$  could deviate to claim to be type  $i - 1$ . The second set of constraints, (54), one for each date, impose sustainability, with the definition  $V(K_t, L_t) \equiv \max_{\tilde{x}'_t + \tilde{c}'_t \leq F(K_t, L_t)} (1 - a)v(\tilde{x}'_t) + a \sum_{i=0}^N \pi_i u(\tilde{c}'_t(\theta_i))$ , and finally, the last set of constraints, one for each date, impose the aggregate resource constraint.

We again follow Marcet and Marimon (1998) and form the Lagrangian:

$$\begin{aligned} \mathcal{L} = & \sum_{t=0}^{\infty} \beta^t \sum_{i=1}^N \pi_i [u(c_t(\theta_i)) - \chi(l_t(\theta_i) | \theta_i)] \\ & + \sum_{t=0}^{\infty} \beta^t \mu_t \left\{ (1 - a)v(x_t) + a \sum_{i=1}^N \pi_i [u(c_t(\theta_i)) - \chi(l_t(\theta_i) | \theta_i)] \right\} \\ & - \sum_{t=0}^{\infty} \beta^t (\mu_t - \mu_{t-1}) V\left(K_t, \sum_{i=1}^N \pi_i l_t(\theta_i | \theta_i)\right) \\ & + \sum_{i=1}^N \lambda_i \left\{ \sum_{t=0}^{\infty} \beta^t \{ [u(c_t(\theta_i)) - \chi(l_t(\theta_i) | \theta_i)] - [u(c_t(\theta_{i-1})) - \chi(l_t(\theta_{i-1}) | \theta_{i-1})] \} \right\} \\ & - \sum_{t=0}^{\infty} \beta^t \eta_t \left\{ x_t + K_{t+1} + \sum_{i=1}^N \pi_i c_t(\theta_i) - F\left(K_t, \sum_{i=1}^N \pi_i l_t(\theta_i | \theta_i)\right) \right\} \end{aligned}$$

where  $\lambda_i$  is the multiplier on the incentive-compatibility constraint of type  $i$ ,  $\beta^t \eta_t$  is the multiplier on the resource constraint at time  $t$ , and we have left the constraints that  $c_t(\theta_i) \geq 0$  for all  $i$  and  $t$  and  $x_t \geq 0$  for all  $t$  implicit.

For  $c_t(\theta_i) > 0$  and  $x_t > 0$ , we can take first-order conditions, which, after canceling out the  $\beta^t$  terms and defining  $\lambda_0 = 0$  and  $\lambda_{N+1} = 0$ , yield:

$$(1 + a\mu_t) \pi_i u'(c_t(\theta_i)) + (\lambda_i - \lambda_{i+1}) u'(c_t(\theta_i)) - \eta_t \pi_i = 0 \text{ for all } i = 0, \dots, N \text{ and all } t, \quad (56)$$

$$(1 + a\mu_t) \pi_i \chi'(l_t(\theta_i) | \theta_i) + (\lambda_i \chi'(l_t(\theta_i) | \theta_i) - \lambda_{i+1} \chi'(l_t(\theta_i) | \theta_{i+1})) \quad (57)$$

$$+ (\mu_t - \mu_{t-1}) \pi_i V_L(K_t, L_t) - \eta_t \pi_i F_L(K_t, L_t) = 0 \text{ for all } i = 0, \dots, N \text{ and all } t,$$

$$- (\mu_t - \mu_{t-1}) V_K(K_t, L_t) + \eta_t F_K(K_t, L_t) - \beta^{-1} \eta_{t-1} = 0 \text{ for all } t, \quad (58)$$

$$\mu_t (1 - a) v'(x_t) = \eta_t \text{ for all } t. \quad (59)$$

Recall that  $\{\mu_t\}_{t=0}^\infty$  is a nondecreasing sequence, thus it possesses a unique limit point on the extended real line. First suppose that  $\mu_t \rightarrow \mu^* < \infty$ , then (56)-(59) establishes that there exists an allocation with  $\eta_t \rightarrow \eta^*$ ,  $c_t(\theta_i) \rightarrow c_i^* > 0$  for all  $i$ , and  $x_t \rightarrow x^*$ , in which distortions disappear as claimed in the theorem.

To complete the proof, we need to show that there does not exist any solution to the above maximization problem with  $\mu_t \rightarrow \infty$ . To obtain a contradiction suppose that  $\mu_t \rightarrow \infty$ . Combine (56) for type  $N$  with (59) and using the fact that  $\mu_{t+1} \geq \mu_t > 0$  and that  $\lambda_{N+1} = 0$  (by definition), we have that for all  $c_t(\theta_N) > 0$  and  $x_t > 0$ ,

$$\frac{\left(1 + \frac{\lambda_N}{\pi_N}\right)}{(1-a) \frac{v'(x_t)}{u'(c_t(\theta_N))} - a} = \mu_t \leq \mu_{t+1} = \frac{\left(1 + \frac{\lambda_N}{\pi_N}\right)}{(1-a) \frac{v'(x_{t+1})}{u'(c_{t+1}(\theta_N))} - a}. \quad (60)$$

Both sides of this equation are strictly positive by the fact that  $\mu_{t+1} \geq \mu_t > 0$ . The hypothesis that  $\mu_t \rightarrow \infty$  implies that as  $t \rightarrow \infty$ ,  $\mu_t < \mu_{t+1}$ .

Next combine (56) some  $i$  and  $i' \neq i$  to obtain:

$$u'(c_t(\theta_i)) + \frac{(\lambda_i - \lambda_{i+1})}{\pi_i(1 + a\mu_t)} u'(c_t(\theta_i)) = u'(c_t(\theta_{i'})) + \frac{(\lambda_{i'} - \lambda_{i'+1})}{\pi_{i'}(1 + a\mu_t)} u'(c_t(\theta_{i'})). \quad (61)$$

The fact that  $\mu_t \rightarrow \infty$  implies that as  $t \rightarrow \infty$ ,  $|c_t(\theta_i) - c_t(\theta_{i'})| \rightarrow 0$ . This argument then establishes that  $c_t(\theta_i) \downarrow c^*$  for all  $i = 0, \dots, N$ . From the freedom of labor supply, this also implies that we must have  $l_t(\theta_i) \downarrow l^* = 0$  for all  $i = 1, \dots, N$ , since otherwise at some point all  $\theta_i \neq \theta_0$  would claim to have become disabled). From Assumption 4 and the resource constraint, this also implies that as  $t \rightarrow \infty$ ,  $c_t(\theta_i) \downarrow 0$  for all  $i = 0, \dots, N$  and  $x_t \downarrow 0$ .

Now, suppose first that  $l^* = 0$  and  $c^* = 0$  are reached in finite time, i.e.,  $c_{t'}(\theta_i) = 0$  for all  $i$  and all  $t \geq t'$  for some  $t' < \infty$ . We will show that this cannot be part of a sustainable mechanism. We have that for all  $t \geq t'$ ,  $c_t(\theta_i) = l_t(\theta_i) = 0$  for all  $i$  and  $x_t = 0$ , so the continuation utility of the government is  $(1-a)u(0)/(1-\delta)$  (since  $v(0) = 0$  and  $\chi(0|\cdot) = 0$ ). However, by hypothesis,  $c_{t'-1}(\theta_i) > 0$  for at least some  $i$ , so there is positive output at  $t' - 1$ . Moreover, from (61),  $c_{t'-1}(\theta_i) \neq c_{t'-1}(\theta_{i'})$  for some  $i$  and  $i'$ . This implies that the government would prefer to deviate at  $t' - 1$  to  $\xi_{t'-1} = 1$ , and redistribute the output between  $x_t$  and equal consumption across all individuals (i.e.,  $c_{t'-1}(\theta_i) = c_{t'-1}^*$  for all  $i$  and some  $c_{t'-1}^*$ ). This deviation will necessarily increase government utility at  $t' - 1$  (since  $c_{t'-1}(\theta_i) \neq c_{t'-1}(\theta_{i'})$  for all  $i$  and  $i'$  in the original allocation), and its continuation utility from  $t'$  onwards would still remain at  $(1-a)u(0)/(1-\delta)$ . Since this argument applies for any  $t' > 0$ , it proves that there cannot be a sustainable mechanism that reaches  $l^* = 0$  and  $c^* = 0$  in finite time. Hence, it must be the case that  $c_t(\theta_i) \downarrow 0$  and  $l_t(\theta_i) \downarrow 0$ , but  $c_t(\theta_i) > 0$  for all  $t$ . Then, combining (60) and (61) implies that, as long as  $au'(0) \neq (1-a)v'(0)$ ,  $\mu_{t+1} - \mu_t \downarrow 0$ , contradicting  $\mu_t \rightarrow \infty$ , and thus establishing the theorem. ■