Review of Economic Studies (2009) 0034-6527/09/00411011$02.00
© 2009 The Review of Economic Studies Limited doi: 10.1111/j.1467-937X.2009.00587.x

Dynamic Mirrlees Taxation under Political Economy Constraints

DARON ACEMOGLU
Massachusetts Institute of Technology and Canadian Institute for Advanced Research

MIKHAIL GOLOSOV
Yale University and New Economic School

and

ALEH TSYVINSKI
Yale University and New Economic School

First version received February 2008; *final version accepted August* 2009 *(Eds.)*

We study the structure of non-linear taxes in a dynamic economy subject to political economy problems. In contrast to existing literature, taxes are set by a self-interested politician, without any commitment power, who is partly controlled by the citizens. We prove that: (1) a version of the revelation principle applies; and (2) the provision of incentives to politicians can be separated from the provision of incentives to individuals. Using these results, we provide conditions under which distortions created by political economy problems persist or disappear. We then extend these results to environments with partially benevolent governments and potential *ex post* conflict among the citizens.

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. This literature characterizes the structure of optimal (non-linear) taxes assuming that policies are decided by a benevolent government with full commitment power. The optimal tax structure typically involves a significant amount of information gathered in the hands of the government as well as a range of transfers to and from the government using the available fiscal instruments. In practice, however, tax structures are designed by politicians who care about re-election, self-enrichment, or their own individual biases and cannot commit to future policies or to dynamic mechanisms. This observation gives greater weight to the famous quote by Juvenal, which is at the root of much of political economy analysis: "*who will guard the guardians?*". In such environments, "guarding the guardians" becomes even more challenging because of the amount of information and enforcement power concentrated in the government's hands.

In this paper, we study the structure of dynamic non-linear taxation under political economy constraints, that is, when there is no commitment to policies and the government (politicians) also need to be "guarded" (controlled).[1]

The challenges created by time inconsistency in dynamic non-linear taxation environments are especially severe and were 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 labour supply. But unlike 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 $T \to \infty$, the unique sequential equilibrium of this game involves the highly inefficient outcome in which all individuals declare to be the worst type at all dates, supply the lowest level of labour, 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, may also fail (in Roberts' economy there is no truthful reporting of types).

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 as in Mirrlees's baseline analysis? We show that, under reasonable assumptions, taxation and redistribution policies resembling those resulting from the normative Mirrleesean analysis with commitment and a benevolent planner can be supported as equilibria. 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*. This terminology emphasizes that we are characterizing 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 implementing 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, where politicians have no commitment power and can even deviate from their within-period commitments, but they are subject to electoral accountability. If they pursue policies not in line with the expectations (wishes) of the electorate, they can be punished by being removed from office.[2]

These two ingredients enable us to develop a tractable framework for the analysis of dynamic taxation in the presence of political economy and commitment constraints. In

[1] 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", "optimal non-linear taxation", and "mechanism" interchangeably.

[2] 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, e.g. Persson and Tabellini, 2000, Chapter 4).

particular, they lead to two results that are both important for our modelling approach and of potentially broader interest. First, a version of the revelation principle, *truthful revelation along the equilibrium path*, applies in our environment regardless of the discount factors of various parties (Theorem 1).[3] This result relies on the possibility of "harsh" punishments by the citizens (which here take the form of replacing the politician). These ensure that information revealed along the equilibrium path is of limited use when the politician deviates. Second, we show that the best sustainable mechanism enables a separation between private and public incentives (Theorem 3). This result implies that incentive compatibility for individuals can be treated separately from ensuring that the politician in power does not wish to deviate from the candidate *social plan* (proposed tax-transfer scheme).

The results on truthful revelation along the equilibrium path and on separation enable 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 labour supply. We call this a *quasi-Mirrlees problem* as it is a usual dynamic Mirrlees problem with two additional constraints on aggregate labour and consumption. Its solution leads to an *indirect utility functional* representing expected utility as a function of the aggregate levels of consumption and labour supply. (2) We then characterize the provision of 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 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, therefore, implies that incorporating lack of commitment and self-interest of politicians does not necessarily invalidate the methodology of approaching dynamic taxation problems as one of dynamic mechanism design; it simply adds additional constraints on aggregates.

Using this formulation, we provide a systematic characterization of the evolution of aggregate distortions. 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 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 must be associated with increased rents for the politician in power. This in turn increases the opportunity cost of production and leads to a reduction in labour supply and capital accumulation. Therefore, labour supply and capital are depressed in the best sustainable mechanism *as a way of relaxing the sustainability constraint* (and thus reducing the rents allocated to the politicians in power). We also show that when politicians are as patient as (more patient than) the citizens, the additional political economy distortions 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

[3] The fact that this result holds regardless of the discount factors emphasizes that it is *not* a folk theorem type result.

Note also that 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.

additional taxes on labour beyond those implied by the optimal Mirrleesean taxation and no aggregate taxes on capital.[4] 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 labour 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 and might provide a possible perspective for why capital taxation is pervasive in practice.[5]

These results are derived under a variety of assumptions on preferences and the form of political economy constraints, which ensure the separation of private and public incentives. We then show how these results can be extended to richer political and economic environments, though this necessitates an alternative approach that does not rely on a separation of private and public incentives. Using this alternative approach, we first show that similar results hold when politicians are partially benevolent. This case is particularly important since it enables us to revisit the stark negative result in Roberts' (1984) seminal paper discussed above. We demonstrate that, in contrast to Roberts' analysis, aggregate distortions created by commitment problems disappear if the politician is as patient as the citizens. Second, we show that many forms of *ex post* political conflict among the citizens (for example, between those with different histories and different "wealth levels") can also be introduced into our framework and lead to a mathematical formulation identical to that with partially benevolent governments. Finally, we highlight that our separation theorem (Theorem 3) does not hold when the best deviation by the government is a function of the distribution of resources within the population, and we also emphasize that our results do not hold when strategies are restricted to be Markovian.

Our paper is related to a number of different literatures. First, it is closely related to recent advances in the theory of mechanism design without commitment, including Bester and Strausz (2001), Skreta (2006, 2007), Sleet and Yeltekin (2006), and Bisin and Rampini (2005). Sleet and Yeltekin (2006) provide a proof of the revelation principle for sufficiently high discount factors in a dynamic economy with time inconsistency, but without political economy constraints.

Bisin and Rampini (2005) extend Roberts' analysis and show how the presence of anonymous markets acts as an additional constraint on the government, ameliorating the commitment problem. A version of the revelation principle also applies in their model but crucially does not involve truthfully reporting along the equilibrium path. In their model, it is the ability of agents to trade secretly that disciplines the behaviour of the government, whereas in our model politicians are disciplined by the threat of removal from office. The most important distinction between our work and that of Bisin and Rampini is the infinite horizon nature of our model, which enables us to construct sustainable mechanisms with the revelation principle (where agents reports the truth) holding along the equilibrium path. This enables us to analyse substantially more general environments and to characterize the limiting behaviour of distortions and taxes.

[4] 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.

[5] Naturally, the usual Mirrleesean "wedges" in labour supply and intertemporal decisions are still present and affect private incentives even though political economy distortions may disappear.

In addition to these papers, our work is related to the burgeoning literatures on dynamic political economy,[6] and on dynamic non-linear taxation. In particular, our framework incorporates the general model of dynamic non-linear taxation considered in Golosov, Kocherlakota, and Tsyvinski (2003), Kocherlakota (2005), Albanesi and Sleet (2005), and Farhi and Werning (2008). A recent interesting paper by Albanesi and Armenter (2007) studies general structure of intertemporal distortions in a variety of contexts and uses a technique similar to those in this paper in separating aggregate from idiosyncratic distortions.

The results in this paper are also closely related to our previous work, Acemoglu, Golosov, and Tsyvinski (2006, 2008*a*, *b*). Many of the results here were first presented in the working paper, Acemoglu, Golosov, and Tsyvinski (2008*a*).[7] A special case of these results has been developed in Acemoglu, Golosov and Tsyvinski (2008). That paper focuses on the problem of controlling a self-interested politician in a representative agent neoclassical growth model and thus does not feature any incomplete information, heterogeneity, or non-linear taxation, which are our present focus. Consequently, the results in the current paper are more general and more broadly applicable than those in Acemoglu, Golosov and Tsyvinski (2008*a*) and enable us to investigate questions related to the structure of taxation under political economy constraints. In particular, the key results of the present paper related to truthful revelation, separation of private and public incentives, the structure of non-linear taxes, and the interaction between political economy and time-inconsistency are not present in that paper. The provision of incentives to politicians in our model is also related to the structure of optimal contracts 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 (2008*a*) extend Ray's results to the case in which discount factors are different between the principal and the agent (or the citizens and the politician) and highlight the role of the relative discount factors, which also play a similar role in the characterization of the long-run evolution of distortions here.

The rest of the paper is organized as follows. Section 2 introduces the general economic environment and describes the design of the tax-transfer mechanisms and the political economy environment. Section 3 establishes truthful revelation along the equilibrium path and shows how the provision of incentives to individuals can be separated from the provision of incentives to politicians, enabling a relatively tractable analysis of a sustainable tax-transfer mechanism. Section 4 applies these results to characterize the behaviour of political economy distortions and derives their implications for taxes. Section 5 shows how these results can be generalized to environments with partially benevolent government and *ex post* political conflict among citizens and also discusses their potential limitations. Section 6 concludes. Several omitted proofs are presented in Appendix A, while Appendix B, which is available on the http://www.wiley.com/bw/journal.asp?ref=0034-6527&site=1, contains some additional technical results and proofs.

[6] For general discussions of the implications of self-interested behaviour 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 (2007) provide introductions to various aspects of the recent developments and the basic theory. For dynamic analysis of political economy, focusing mostly on Markovian equilibria, see, among others, Krusell and Rios-Rull (1999), Acemoglu and Robinson (2006), Hassler *et al.* (2005), and Battaglini and Coate (2008).

[7] Results related to the comparison of market-based and government-controlled allocations in that working paper have been extended in Acemoglu, Golosov and Tsyvinski (2008*a*). None of these results is present in the current paper.

2. MODEL

2.1. *Environment*

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

$$u\left(c_t^i, l_t^i \mid \theta_t^i\right), \tag{1}$$

where $c_t^i \geq 0$ is the consumption of this individual, $l_t^i \in \left[0, \bar{l}\right]$ is labour supply, and θ_t^i is the individual's "type". This formulation is general enough to nest both preference shocks and productivity shocks.[8]

Let $\Theta = \{\theta_0, \theta_1, ..., \theta_N\}$ be a finite ordered set of real numbers denoting potential types, with the convention that θ_i corresponds to "higher skills" than θ_{i-1}, and in particular, θ_0 is the worst type. Let Θ^T be the T-fold product of Θ, representing the set of sequences of length $T = 1, 2, ..., \infty$, with each element belonging to Θ. We think of each agent's lifetime type sequence θ^∞ as drawn from Θ^∞ according to some measure μ^∞. Let $\theta^{i,\infty}$ be the draw of individual i from Θ^∞. The t-th element of $\theta^{i,\infty}$, θ_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 skills 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 identically and independently from Θ^∞ according to the same measure μ^∞, so that there is no aggregate uncertainty in the type distribution.[9] We denote the distribution of the vector θ^t across agents by G^t.

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

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

where $\mathbb{E}\left[\cdot|\theta^{i,t}\right]$ denotes the expectations conditional on having observed the history $\theta^{i,t}$. We impose the following standard assumption on the utility function, which also introduces the single crossing property.

Assumption 1. ***(utility function)*** *For all $\theta \in \Theta$, $u(c, l \mid \theta) : \mathbb{R}_+ \times \left[0, \bar{l}\right] \rightarrow \mathbb{R}$ is twice continuously differentiable and jointly concave in c and l, and is non-decreasing in c and non-increasing in l. Moreover, $u_c(c, l \mid \theta)/|u_l(c, l \mid \theta)|$ is increasing in θ for all c and l and all $\theta \in \Theta$, where u_c and u_l denote the partial derivatives of u.*

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

$$Y = F(K, L),$$

where K is capital and L is labour, and the economy starts with a positive endowment of capital stock, $K_0 > 0$ at $t = 0$. In addition, we assume the following:

Assumption 2. ***(production structure)*** *F is strictly increasing and continuously differentiable in K and L with partial derivatives denoted by F_K and F_L, exhibits constant returns to*

[8] For example, productivity shocks would correspond to the case where $u\left(c_t^i, l_t^i \mid \theta_t^i\right) = u\left(c_t^i, l_t^i/\theta_t^i\right)$.

[9] This structure imposes no restriction on the time-series properties of individual skills. Both identical independent draws and arbitrary temporal dependence are allowed.

scale, and satisfies $\lim_{L\to 0} F_L(K,L) = \infty$ *for all* $K > 0$ *and* $\lim_{K\to\infty} F_K(K,L) < 1$ *for all* $L \in [0,\bar{l}]$. *Moreover, capital fully depreciates after use, and* $F(K,0) = 0$.

Both the full depreciation assumption and the assumption that labour is essential for production are adopted to simplify the notation. The condition that $\lim_{K\to\infty} F_K(K,L) < 1$ together with $L \in [0,\bar{l}]$ implies that there is a maximum steady-state level of output that is uniquely defined by $\overline{Y} = F(\overline{Y},\bar{l}) \in (0,\infty)$, where recall that $\bar{l}$ is the maximum amount of labour supply per capita (and thus the maximum total labour supply). The condition that $\lim_{L\to 0} F_L(K,L) = \infty$ implies that in the absence of distortions there will be positive production.

2.2. *Political economy*

The allocation of resources in this economy is entrusted to a politician who is in charge of operations of the government. This politician has the power to tax and redistribute resources across agents, and can also allocate some of the tax revenue to himself as rents (government consumption). We interpret the government (and thus the politician) as being necessary for the operation of the tax-redistribution mechanisms (as well as other functions, such as implementation of law and order and provision of public goods). The key dilemma facing the society is how to control the government once the powers to tax and redistribute resources have been vested with it.

We adopt the simplest and most conventional approach to this problem, and incorporate the classic electoral accountability setup of Barro (1973) and Ferejohn (1986) into our environment: there is a large number of potential (and identical) politicians. We 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), $v : \mathbb{R}_+ \to \mathbb{R}$ is the politician's instantaneous utility function. Notice also that the politician's discount factor, δ, is potentially different from that of the citizens, β. 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.[10]

Assumption 3. ***(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 expropriate 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

[10] 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 below, e.g. (7), need to be modified.

the politicians and the citizens, corresponding to the *best sustainable mechanism*—meaning the sustainable mechanism that maximizes the *ex ante* utility of citizens.[11]

The rest of this section formally defines the structure of the game. We first describe the feasible actions by citizens and the politicians, and the timing of events. We then provide a formal definition of mechanisms.

2.3. *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 $Z^t \equiv \prod_{s=0}^t Z_s$, z^t denote a generic element of Z^t and by Z the space of all such lifetime reports.

A submechanism consists of two mappings, i.e. $M_t \equiv \left(\tilde{c}_t, \tilde{l}_t\right)$ such that $\tilde{c}_t : Z^t \to \mathbb{R}_+$ assigns consumption levels for each complete history of messages and public histories, and $\tilde{l}_t : Z^t \to \left[0, \bar{l}\right]$ assigns corresponding labour supply levels.[12] However, as the timing of events below will make it clear, we also assume that there is *freedom of labour supply*, in the sense that each individual can always disobey the labour supply allocation implied by $\tilde{l}_t$ and choose $l_t^i = 0$.[13] Instead of introducing an additional action designating this choice, we introduce a message $z^{\emptyset}$ such that if $z^t = \left(z^{t-1}, z^{\emptyset}\right)$ for any $z^{t-1} \in Z^{t-1}$, then $\tilde{l}_t$ specifies $l_t^i = 0$ for the individual in question. This is clearly without loss of any generality, since if such a message did not exist, the individual could always disobey $\tilde{l}_t$ and choose $l_t^i = 0$ himself. We denote the set of feasible submechanisms, which allows for message $z^{\emptyset}$ specified above and satisfies the relevant resource constraints (specified 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. A natural assumption in the political economy context is that 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.

[11] 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 behaviour of the government. In addition, throughout the paper we focus on perfect Bayesian equilibria (see Definition 1).

[12] 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 Appendix B, 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 $M_t\left[K_t\right]$ 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.

[13] In Acemoglu, Golosov and Tsyvinski (2006), we derive the "freedom of labour supply" endogenously using an environment in which individuals could be disabled and unable to supply any labour. Directly introducing the freedom of labour supply is without loss of generality for our main focus and simplifies the analysis.

We next summarize the game between the politician in power and the citizens. 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 $z_t^i \in Z_t$. The message z_t^i together with the history of messages $z^{i,t-1} \in Z^{t-1}$ determines labour supplies $\tilde{l}_t\left(z^{i,t}\right)$ 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. At this point, individual i can also choose to supply zero labour and receive zero consumption.
3. Production takes place according to the labour 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 $L_t = \int_{i\in I} \tilde{l}_t\left(z^{i,t}\right) di$.
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 $\tilde{K}_{t+1} = F(K_t, L_t) - \tilde{x}_t - \int_{i\in I} \tilde{c}_t\left(z^{i,t}\right) di$. If $\xi_t = 1$, the politician chooses $\tilde{x}_t' \leq F(K_t, L_t)$, and a new consumption function $\tilde{c}_t' : Z^t \to \mathbb{R}_+$, and next period's capital stock is: $\tilde{K}_{t+1}' = F(K_t, L_t) - \tilde{x}_t' - \int_{i\in I} \tilde{c}_t'\left(z^{i,t}\right) di$.
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.

Note the difference between the standard models with no commitment and our setup where, in stage 4, the politician can decide to expropriate the output produced in the economy, and citizens can replace the politician at the last stage. Notice that at stage 4 labour supply decisions have already been made according to the pre-specified 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, $\tilde{c}_t' : Z^t \to \mathbb{R}_+$ 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. Here we simplify the discussion by assuming that the decision $\rho_t \in \{0,1\}$ is taken by a randomly chosen citizen.[14]

2.4. *Histories and reporting strategies*

Let $M = \{M_t\}_{t=0}^{\infty}$ with $M_t \in \mathcal{M}_t$ be a mechanism (i.e. a sequence of submechanisms defined above), with the set of mechanisms denoted by $\mathcal{M}$. Let $x = \{x_t\}_{t=0}^{\infty}$ be the sequence of consumption levels (rents) for the politician. We define a *social plan* as (M, x), which is an implicitly agreed sequence of submechanisms and politician consumption levels.

[14] Exactly the same equilibrium would be obtained if there are majoritarian elections over the replacement decision and each individual votes sincerely or uses strategies that are not weakly dominated in the election. We discuss this issue further after Lemma 1. How the introduction of *ex post* political conflict among the citizens influences the results is studied in Section 5.

We represent the action of the politician at time t by $\upsilon_t = \left(\tilde{M}_t, \xi_t, \tilde{x}_t, \tilde{x}'_t, \tilde{c}'_t\right)$. The first element of υ_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 υ_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.[15] 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: $\tilde{x}_t \le F(K_t, L_t)$ and $\tilde{x}'_t \le F(K_t, L_t)$, but to simplify notation we write $\tilde{x}_t, \tilde{x}'_t \in \mathbb{R}_+$. Let Υ_t be the set of υ_t's.[16]

Let $h^t \equiv \left(K_0, \iota_0, \upsilon_0, \rho_0, K_1, ..., K_t, \iota_t, \upsilon_t, \rho_t, K_{t+1}\right)$ denote the public history of the game up to date t, and H^t be the set of all such histories. The electoral decision at time t, ρ_t, defined as

$$\rho_t : H^{t-1} \times \Upsilon_t \to \{0, 1\},$$

designates whether the society chooses to replace a politician, given the public history at time $t-1$, h^{t-1}, and actions of politicians at time t, υ_t.

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

$$\alpha^i_t : Z^{t-1} \times H^{t-1} \times \Theta^t \to Z_t.$$

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

$$\alpha^*\left(\theta^t \mid z^{t-1}, h^{t-1}\right) = z_t\left[\theta^t\right] \text{ for all } \theta^t \in \Theta^t,\ z^{t-1} \in Z^{t-1}, \text{ and } h^{t-1} \in H^{t-1}, \tag{2}$$

where the notation $z_t\left[\theta^t\right]$ 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 $\alpha^i_t\left(\theta_t \mid z^{t-1}\left[\theta^{t-1}\right], h^{t-1}\right) = \alpha^*$. Notice that this strategy only imposes truth-telling following truthful reports in the past (because instead of an arbitrary history of messages z^{t-1}, we have conditioned on $z^{t-1}\left[\theta^{t-1}\right]$). In addition, let us define the null strategy

$$\alpha^{\emptyset}\left(\theta^t \mid z^{t-1}, h^{t-1}\right) = z^{\emptyset} \text{ for all } \theta^t \in \Theta^t,\ z^{t-1} \in Z^{t-1},\ \ h^{t-1} \in H^{t-1}, \tag{3}$$

where recall that $z^{\emptyset}$ stands for the message corresponding to zero labour supply. We will use the notation $\alpha^i_t\left(\theta_t \mid z^{t-1}, h^{t-1}\right) = \alpha^{\emptyset}$ to denote that the individual is playing the null strategy.

Let $\underline{z}_t \in \mathcal{Z}_t$ be a profile of reports at time t.[17] As usual, we define $\mathcal{Z}^t = \prod_{s=0}^{t} \mathcal{Z}_s$. We denote the *reporting strategy profile* of all the individuals in society by $\underline{\alpha}$,[18] with $\mathbf{A}$ corresponding to

[15] Since we are characterizing a (sustainable) mechanism, there is no need to specify the ownership of the capital stock $\tilde{K}_{t+1}$. Instead, this is simply the amount of resources used in production in the following period, and the government (the mechanism) decides how this production will be distributed.

[16] In fact, υ^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 υ^t should be observed publicly. This slight abuse of notation is without any consequence for the analysis.

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

[18] More formally, $\underline{\alpha}_t$ assigns a report to each individual, thus it is a function of the form $\underline{\alpha} : [0, 1] \to Z$, where $i \in [0, 1]$ denotes individual i.

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

2.5. *Definition of equilibrium*

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

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

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 by Γ 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 $\{\hat{\underline{\alpha}}, \hat{\rho}\}$ and a belief system $\mathcal{B}$, such that $\hat{\Gamma}$ and $\{\hat{\underline{\alpha}}, \hat{\rho}\}$ are sequentially rational, i.e. best responses to each other in all information sets, given $\mathcal{B}$, and whenever possible, the belief system $\mathcal{B}$ is derived from Bayesian updating given the strategy profiles $\hat{\Gamma}$ and $\{\underline{\alpha}, \rho\}$. We write the requirement that these strategy profiles are best responses to each other as $\hat{\Gamma} \succeq_{\{\hat{\underline{\alpha}}, \hat{\rho}\}} \Gamma$ for all $\Gamma \in \mathcal{G}$ and $\{\hat{\underline{\alpha}}, \hat{\rho}\} \succeq_{\hat{\Gamma}} \{\underline{\alpha}, \rho\}$ for any $\{\underline{\alpha}, \rho\} \in \mathbf{A} \times \mathcal{R}$.

In what follows, there will be no need to explicitly characterize or condition on the belief system $\mathcal{B}$ (though this is always in the background). 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 Γ given a social plan (M, x). For a given equilibrium, let us also denote the set of histories that are observed with positive probability along the equilibrium path by $\tilde{H}^t \subset H^t$.

Definition 2. M is a sustainable mechanism if there exists $x = \{x_t\}_{t=0}^{\infty}$, a strategy profile $\{\hat{\underline{\alpha}}, \hat{\rho}\}$ for the citizens and a strategy profile $\hat{\Gamma}_{M,x} \in \mathcal{G}$ 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 $\tilde{x}_t = x_t$ for all $h^t \in \tilde{H}^t$. In this case, we say that equilibrium strategy profiles $\hat{\Gamma}_{M,x}$ and $\{\hat{\underline{\alpha}}, \hat{\rho}\}$ 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, $\{\hat{\underline{\alpha}}, \hat{\rho}\}$, of the citizens. The notation $\hat{\Gamma}_{M,x} \succeq_{\{\hat{\underline{\alpha}}, \hat{\rho}\}} \Gamma$ makes this explicit, stating that given the strategy profile, $\{\hat{\underline{\alpha}}, \hat{\rho}\}$, of the citizens, the politician weakly prefers this strategy profile to any other strategy profile based on the same implicit agreement.

3. TRUTHFUL REVELATION AND SEPARATION OF INCENTIVES

The revelation principle is a powerful tool for the analysis of mechanism design and implementation problems (see, e.g. Mas-Collel, 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 may not hold. As Roberts' (1984) paper discussed in the Introduction demonstrates, there may exist situations in which no equilibrium would involve individuals reporting their true type.[19]

The key result of this section is 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 (e.g. with the ratchet effect) is the possibility here that the agents can punish the deviating politician (mechanism designer) by replacing him. 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 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.

3.1. *Truthful revelation along the equilibrium path*

We focus on (Perfect Bayesian) equilibria that maximize utility of the citizens, which we refer to as the *best sustainable mechanism*. 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:

$$\mathbf{MAX}_0: \quad \max_{\left\{\tilde{c}_t(\cdot),\tilde{l}_t(\cdot),\tilde{x}_t,K_{t+1}\right\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u\left(\tilde{c}_t\left(z^t\left[\alpha_t\left(\theta^t\right)\right]\right), \tilde{l}_t\left(z^t\left[\alpha_t\left(\theta^t\right)\right]\right) \mid \theta_t^i\right)\right] \tag{4}$$

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

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

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

$$\left\{\underline{\alpha}, \rho\right\} \text{ is a best response to } \Gamma_{M,\tilde{x}}, \tag{6}$$

and the *sustainability constraint* of the politician in power:

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

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

[19] As noted in the Introduction, this statement refers to the case in which messages are sent to the politician in charge of the mechanism. It is possible to construct alternative environments with fictional mechanism designers with full commitment power, so that the revelation principle holds. See Bester and Strausz (2001) and Skreta (2006, 2007) for general versions of the revelation principle.

$\xi_t = 1$, 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 $v^c_t\left(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t\right)$, to emphasize that this continuation value depends on the entire history of submechanisms (thus on the information about individual types that has been revealed so far) 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), is 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 nonempty, since the trivial allocation with zero production and zero consumption for all parties is in the set.

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

Lemma 1. *Suppose Assumptions 1–3 hold. In any sustainable mechanism, the condition*

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

is necessary.

Moreover, the allocation of resources in the 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* $v^c_t\left(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t\right) = 0$ *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 *any* sustainable mechanism. Recall, from the definition of $\tilde{H}^t$, that if $h^t \in \tilde{H}^t$, then $\{M_s, x_s\} = \{\tilde{M}_s, \tilde{x}_s\}$ for all $s \leq t$. Consider the strategy profile $\rho^{\varnothing}$ for the citizens such that $\rho^{\varnothing}(h^t) = 0$ if $h^t \in \tilde{H}^t$ and $\rho^{\varnothing}(h^t) = 1$ if $h^t \notin \tilde{H}^t$. 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 $h^t \in \tilde{H}^t$ only if

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

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

If (8) is violated following some history h^t, the best deviation for the politician is $\xi_t = 1$ and $\tilde{x}'_t = F(\tilde{K}_t, \tilde{L}_t)$. 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, completing the proof of the first part of the lemma.

To prove the second part, that (8) is sufficient for the best sustainable mechanism, note that reducing $v^c_t\left(\tilde{K}'_{t+1}, \tilde{c}'_t \mid \tilde{M}^t\right)$ is equivalent to relaxing the constraint on problem (4), so is always preferred. Since from Assumption 3, $v^c_t \geq 0$ (i.e. $x \geq 0$ and $v(0) = 0$), we only need to show that $v_t^c\left(\tilde{K}_{t+1}', \tilde{c}_t' \mid \tilde{M}^t\right) = 0$ 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 play up to this date. This establishes the sufficiency of (8).

Next suppose $\{\tilde{M}_t, \tilde{x}_t\}_{t=0}^{\infty}$ that is a solution to ($\mathbf{MAX}_0$) 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 ($\mathbf{MAX}_0$). This proves that there is no replacement of the initial politician along the equilibrium path.

To complete the proof of the second part, we only need to show that citizens' strategy (in particular, the replacement strategy $\rho^{\varnothing}$) is sequentially rational. This follows by considering the continuation strategy for each politician such that if $h^t \notin \tilde{H}^t$, then $x_s = F(K_s, L_s)$ and $\xi_s = 1$, $\forall s \geq t$. This ensures that $\rho^{\varnothing}$ and $\underline{\alpha}{=}\alpha^{\emptyset}$ are a best response for the citizens. ||

This lemma uses the fact that regardless 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 a politician's 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 expropriating the whole output, $\tilde{x}_t' = F(K_t, L_t)$. 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.[20] Moreover, the lemma also shows that in any sustainable mechanism (8) is necessary. We can also note that the substantive conclusions of the lemma, and the results that follow, do not depend on the specific political procedure used for replacing politicians.

Corollary 1. *The results of Lemma 1 remain valid if majoritarian elections are used to replace politicians.*

Proof. This corollary follows immediately by noting that when all other individuals use a voting strategy that is equivalent to $\rho^{\varnothing}$ in the proof of Lemma 1, it is a best response for each to do so and this gives the best sustainable mechanism. ||

Next, we define a *direct (sub)mechanism* as $M_t^* : \Theta^t \to \left[0, \bar{l}\right] \times \mathbb{R}_+$. 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 Γ^*.

[20] This statement refers to the sustainability constraint, (8). The optimal mechanism will clearly make allocations depend on the history of individual messages.

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

The notation $\underline{\alpha}^* = \left(\alpha^* \mid \alpha'\right)$ emphasizes that individuals play truth-telling along the equilibrium path, but may play some different strategy profile, α', 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 $\underline{c}\left[\Gamma, \underline{\alpha}\right]$, $\underline{l}\left[\Gamma, \underline{\alpha}\right]$ and $x\left[\Gamma, \underline{\alpha}\right]$ as, respectively, the *equilibrium* consumption and labour 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 1–3 hold. Then there is truthful revelation* along the equilibrium path. *In particular, if* Γ *and* $\left\{\underline{\alpha}, \rho\right\}$ *form a combination of strategy profiles and electoral decisions that support a sustainable mechanism, then there exists another pair of equilibrium strategy profiles* Γ^* *and* $\left\{\underline{\alpha}^*, \rho^{\varnothing}\right\}$*, where* Γ^* *induces direct submechanisms and* $\underline{\alpha}^* = \left(\alpha^* \mid \alpha'\right)$*, for some* α'*, induces truth telling along the equilibrium path, and moreover* $\underline{c}\left[\Gamma, \underline{\alpha}\right] = \underline{c}\left[\Gamma^*, \underline{\alpha}^*\right]$*,* $\left[\Gamma, \underline{\alpha}\right] = \underline{l}\left[\Gamma^*, \underline{\alpha}^*\right]$*, and* $x\left[\Gamma, \underline{\alpha}\right] = x\left[\Gamma^*, \underline{\alpha}^*\right]$.

Proof. Take equilibrium strategy profiles Γ and $\left\{\underline{\alpha}, \rho\right\}$ that support a sustainable mechanism. Then by definition $\xi_t = 0$ for all t (cf. Definition 2), and from Lemma 1, (8) is satisfied. Let the best response of type θ^t at time t according to $\underline{\alpha}$ be to announce $z_{t,\Gamma}\left(\theta^t, h^{t-1}\right)$ given a history of reports $z_{\Gamma}^{t-1}\left(\theta^{t-1}, h^{t-2}\right)$ and public history h^{t-1}. Let $z_{\Gamma}^t\left(\theta^t, h^{t-1}\right) = \left(z_{\Gamma}^{t-1}\left(\theta^{t-1}, h^{t-2}\right), z_{t,\Gamma}\left(\theta^t, h^{t-1}\right)\right)$.

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

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

Now consider the alternative strategy profile for the politician Γ^*, 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 $\underline{c}\left[\Gamma^*, \underline{\alpha}^*\right] = \underline{c}\left[\Gamma, \underline{\alpha}\right]$, $\underline{l}\left[\Gamma^*, \underline{\alpha}^*\right] = \underline{l}\left[\Gamma, \underline{\alpha}\right]$, and $x\left[\Gamma^*, \underline{\alpha}^*\right] = x\left[\Gamma, \underline{\alpha}\right]$ (for any equilibrium path history $h^{t-1} \in \tilde{H}^{t-1}$). Therefore, by construction,

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

for all $\hat{\theta}^t \in \Theta^t$ and all $h^{t-1} \in \tilde{H}^{t-1}$. Now consider the strategy $\left\{\underline{\alpha}^*, \rho^{\varnothing}\right\}$ for the citizens, where $\underline{\alpha}^* = \left(\alpha^* \mid \alpha^{\varnothing}\right)$, $\alpha^{\varnothing}$ is defined in (3) and $\rho^{\varnothing}$ is the replacement strategy defined in the proof of Lemma 1 (which involves replacing the politician following any deviation). Equation (9) then implies that $\left\{\underline{\alpha}^*, \rho^{\varnothing}\right\}$ is a best response along the equilibrium path (for $h^{t-1} \in \tilde{H}^{t-1}$) for the agents against the mechanism M^* and politician strategy profile Γ^*. Moreover, by construction, the resulting equilibrium path allocation when individuals play $\underline{\alpha}^* = \left(\alpha^* \mid \alpha^{\varnothing}\right)$ against Γ^* is the same as when they play $\underline{\alpha}$ against Γ, and the replacement strategy $\rho^{\varnothing}$ ensures that the deviation of the politician under $\left(\Gamma^*, \left\{\underline{\alpha}^*, \rho^{\varnothing}\right\}\right)$ is no higher than under $\left(\Gamma, \left\{\underline{\alpha}, \rho\right\}\right)$. Then since Γ is sustainable, $\Gamma \succeq_{\{\underline{\alpha},\rho\}} \Gamma'$ for all $\Gamma' \in \mathcal{G}$. Therefore, $\Gamma^* \succeq_{\{\underline{\alpha}^*,\rho^{\varnothing}\}} \Gamma'$ for all $\Gamma' \in \mathcal{G}$ or that (8) is satisfied, thus establishing that $\left(\Gamma^*, \left\{\underline{\alpha}^*, \rho^{\varnothing}\right\}\right)$ 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 intuition for Theorem 1 is related to the fact that along the equilibrium path there is effective commitment by the ruler. The proof also exploits this observation and constructs truthful reports that give the same equilibrium utility to the ruler. Off the equilibrium path, one can use the same punishment strategies as those used in the game without direct revelation. Therefore, the sustainability constraint is satisfied.

Two observations are worth making at this point. First, truthful revelation along the equilibrium path and the politician's decision to pursue the implicitly agreed social plan are important to distinguish from truth telling and commitment that are present in the standard mechanism design problems. In these problems, there exist mechanisms that induce truth telling along all paths and there is unconditional commitment (i.e. again along all paths). In contrast, in our environment, there is no commitment *off the equilibrium path*, where the politician can exploit the information he has gathered or expropriate part of the output. Relatedly, off the equilibrium path, non-truthful reporting by the individuals is both present and also important to ensure sustainability. However, along the equilibrium path induced by a sustainable mechanism, the politician prefers not to deviate from the implicitly agreed social plan, and given this equilibrium behaviour, individuals can report their types without the fear that this information or their labour supply will be misused. Second, the results in Theorem 1 are not related to "folk theorem" type results. In particular, Theorem 1 is not a limiting result and applies for all discount factors.

The reason why, despite the lack of commitment, a version of the revelation principle applies is twofold. *First*, the setup where the politician has a deviation within the same period ensures that its continuation payoff after deviation is independent of the information revealed along the equilibrium path. If the continuation value of the politician depended on such information, ensuring truthful reporting would become more difficult. Note that deviation within the period, following production, is a natural assumption in our setup, since the politician has the power to make transfers after production is realized, and thus there is no reason for him not to deviate and take a greater fraction of these resources for himself than specified in the mechanism if such a deviation is beneficial. *Second*, in our model 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*.

Theorem 1, like the rest of our analysis, focuses on perfect Bayesian equilibria. One could also impose additional refinements, such as *renegotiation-proofness*. It can be shown that the main results in this paper (except those in Section 5) hold without any modification if we focus on renegotiation-proof equilibria. In other words, truthful revelation in Theorem 1 and the best sustainable mechanisms characterized in Theorems 2, 3, and 4, and Proposition 2 can be supported as renegotiation-proof equilibria. The argument is similar to that developed in Acemoglu, Golosov and Tsyvinski (2008*a*) for an economy without incomplete information. The main idea is that citizens can punish a politician who deviates from the social plan by replacing him and following an equilibrium along the relevant Pareto frontier thereafter (thus there is no reason for punishments that are harmful for the citizens themselves). We do not provide the details here to economize on space.

3.2. *The best sustainable mechanism*

Theorem 1 enables us to focus on direct mechanisms and truth-telling strategy α^* 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 θ denoted by $G(\theta)$. This implies that θ^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 θ^t occurs infinitely often).[21] Given this construction, we can write total labour supply as $L_t = \int_{\Theta^t} l_t\left(\theta^t\right) dG^t\left(\theta^t\right)$, and total consumption as $C_t = \int_{\Theta^t} c_t\left(\theta^t\right) dG^t\left(\theta^t\right)$.[22] 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 of this section, which will be used throughout the rest of the paper:

Theorem 2. *Suppose Assumptions 1–3 hold. Then, the best sustainable mechanism is a solution to the following maximization program:*

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

subject to some initial condition $K_0 > 0$, the resource constraint

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

a set of incentive compatibility constraints for individuals,

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

and

$$\mathbb{E}\left[\sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}\left(\theta^{i,t+s}\right), l_{t+s}\left(\theta^{i,t+s}\right) \mid \theta_{t+s}^i\right) \middle| \theta^{i,t}\right] \geq \mathbb{E}\left[\sum_{s=0}^{\infty} \beta^s u\left(0, 0 \mid \theta_{t+s}^i\right) \middle| \theta^{i,t}\right] \tag{13}$$

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

$$\mathbb{E}\left[\sum_{s=0}^{\infty} \delta^s v\left(x_{t+s}\right)\right] \geq v\left(F\left(K_t, L_t\right)\right), \tag{14}$$

for all t.

[21] More formally, given the continuum of agents, we can apply a law of large numbers type argument, and each history θ^t will have positive measure. See, for example, Uhlig (1996).

[22] From now on, we suppress the ˜'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, Θ^t.

Proof. The proof of this theorem follows from Lemma 1 and Theorem 1. Suppose there exists an equilibrium $\left(\Gamma^{**},\left\{\underline{\alpha}^{**},\rho\right\}\right)$, that maximizes (10). By the argument in the text, $\left(\Gamma^{**},\left\{\underline{\alpha}^{**},\rho\right\}\right)$ will not feature $\xi_t = 1$ for any t. Therefore, $\left(\Gamma^{**},\left\{\underline{\alpha}^{**},\rho\right\}\right)$ features a sequence of submechanisms $\left\{\hat{M}_t\right\}_{t=0}^{\infty}$, consumption levels for the politician, $\left\{\hat{x}_t\right\}_{t=0}^{\infty}$ and $\xi_t = 0$ for all t. Setting $(M, x) = \left(\left\{\hat{M}_t\right\}_{t=0}^{\infty}, \left\{\hat{x}_t\right\}_{t=0}^{\infty}\right)$ implies that $\left(\Gamma^{**},\left\{\underline{\alpha}^{**},\rho\right\}\right)$ supports a sustainable mechanism. Then, use Theorem 1 to find $\left(\Gamma^{*},\left\{\underline{\alpha}^{*},\rho^{\varnothing}\right\}\right)$ 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 replaced by (12) and (13) since Γ^* induces direct mechanisms and at any date individuals can decide not to supply any labour, after which the mechanism would optimally allocate zero consumption to such individuals in perpertuity. Finally, from Lemma 1, the constraint (14) ensures that Γ^* is a best response to citizens' strategies, $\left\{\underline{\alpha}^{*},\rho^{\varnothing}\right\}$. ||

The role of Theorem 1 in this formulation is clear: it enables us to write the program for the best sustainable mechanism as a direct mechanism with truth-telling reports along the equilibrium path, thus reducing the larger set of incentive compatibility constraints of individuals to (12), which require truth telling, and to (13), which ensures that no individual "disobeys" the mechanism, thus supplying and consuming zero from then on.[23]

3.3. *Separation of private and public incentives*

We next show the second main result of the paper, 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. Golosov, Kocherlakota and Tsyvinski, 2003; Kocherlakota, 2005):

$$\max_{\left\{c_t(\cdot), l_t(\cdot), K_{t+1}\right\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t\left(\theta^{i,t}\right), l_t\left(\theta^{i,t}\right) \mid \theta_t^i\right)\right] \tag{15}$$

subject to the incentive compatibility constraints, (12), and $C_t + X_t + K_{t+1} \leq F\left(K_t, L_t\right)$.

Next, 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

$$\Lambda^{\infty} = \{\{C_t, L_t\}_{t=0}^{\infty} \text{ such that } \exists \left\{c_t\left(\theta^t\right), l_t\left(\theta^t\right)\right\}_{t=0}^{\infty} \text{ satisfying (12) and (13)}\}. \tag{16}$$

In other words, $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}$ implies that there exists incentive compatible and feasible $\left\{c_t\left(\theta^t\right), l_t\left(\theta^t\right)\right\}_{t=0}^{\infty}$. This set is important to define, since, given certain government expenditure

[23] 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 $\left\{\hat{\theta}_{t+s}^{i}\right\}_{s=0}^{\infty}$, thus any potential deviation from time $t = 0$ is covered by this set of constraints.

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

$$\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) \equiv \max_{\{c_t(\cdot), l_t(\cdot)\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t\left(\theta^{i,t}\right), l_t\left(\theta^{i,t}\right) \mid \theta_t^i\right)\right] \tag{17}$$

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

$$\int c_t\left(\theta^t\right) dG^t\left(\theta^t\right) \leq C_t, \tag{18}$$

and

$$\int l_t\left(\theta^t\right) dG^t\left(\theta^t\right) \geq L_t. \tag{19}$$

This program takes the sequence $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}$ as given and maximizes *ex ante* utility of an agent subject to incentive constraints and two additional constraints. The first, (18), requires the sum of consumption levels across agents for all report histories to be no greater than some number C_t, while the second, (19), requires the sum of labour supplies to be no less than some amount L_t. The functional $\mathcal{U}\left(\{C_t, L_t\}_{t=0}^{\infty}\right)$ defines the maximum *ex ante* ($t = 0$) utility of an agent in this economy for a given sequence $\{C_t, L_t\}_{t=0}^{\infty}$ and can be interpreted as the *indirect utility function* of the individuals (from the viewpoint of time $t = 0$). In Appendix B, we show that the functional $\mathcal{U}\left(\{C_t, L_t\}_{t=0}^{\infty}\right)$ 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 $\mathcal{U}\left(\{C_t, L_t\}_{t=0}^{\infty}\right)$ 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:

$$\max_{\{C_t, L_t, K_{t+1}\}_{t=0}^{\infty}} \mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) \tag{20}$$

subject to an initial level of capital stock $K_0 > 0$ and to

$$C_t + X_t + K_{t+1} \leq F\left(K_t, L_t\right), \text{ and } \{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}. \tag{21}$$

This derivation implies that we can represent the standard 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 constraint and to feasibility.

This representation is particularly useful because it highlights the parallel between the dynamic Mirrlees program and the best sustainable mechanism. In particular, the maximization problem characterizing the best sustainable mechanism, (10), can be written as one of maximizing the indicted utility function subject to constraints:

$$\max_{\{C_t, L_t, x_t, K_t\}_{t=0}^{\infty}} \mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) \tag{22}$$

subject to

$$C_t + x_t + K_{t+1} \leq F\left(K_t, L_t\right) \text{ and } \{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}, \tag{23}$$

and also subject to (14). The only difference between the dynamic Mirrlees program in (20)–(21) and the best sustainable mechanism in (22)–(23)–(14) is the presence of the sustainability constraint for the politician in power, (14), 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 3. ***(Separation of Private and Public Incentives)*** *Suppose Assumptions 1–3 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)–(14) from Theorem 2 as a two-step maximization program, and expressing (11) as $x_t = F(K_t, L_t) - C_t - K_{t+1}$. ||

The allocation induced by the best sustainable mechanism is therefore a solution to a problem that maximizes the *ex ante* utility of the citizens as given in (15), but must also choose levels of aggregate consumption and labour supply consistent with the sustainability constraint of the politician in power. It is important to point out the limitations of this theorem. First, it relies on our version of the revelation principle, Theorem 1. Second, it exploits the complete separation between the consumptions of the ruler and the citizens. For this reason, Theorem 3 will no longer hold in the more general environments discussed in Section 5, where the ruler cares about the utility of the citizens. In that section, we develop an alternative approach to provide extensions of our main results to these more general environments.

An important implication of Theorem 3 is that when it holds, political economy 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 labour 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 Mirrleesean optimal taxation).

4. BEST SUSTAINABLE MECHANISMS

We now use Theorems 1–3 to characterize the behaviour of the sequences $\{C_t, L_t, K_t\}_{t=0}^{\infty}$ (and $\{x_t\}_{t=0}^{\infty}$) and aggregate distortions under the best sustainable mechanism.

4.1. *Aggregate distortions*

Theorem 3 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 (20)–(21)). Appendix B shows that $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is differentiable in the sequences $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^{\infty}$. We can therefore consider 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 (with all C_t, L_t for $t \neq s$ held constant). 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 labour 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 Γ^* is **undistorted** at t' if $\left\{\hat{C}_t, \hat{L}_t, \hat{K}_{t+1}\right\}_{t=0}^{\infty}$ is a solution to (20) subject to (21) 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 $t \to \infty$.

This definition states that an undistorted allocation is exactly the allocation that would result in the standard dynamic Mirrlees problem where, in addition to restriction across agents, the government also has to finance an exogenously given sequence of public good expenditures $\{X_t\}_{t=0}^{\infty}$. If an allocation is *undistorted* and $\{C_t, L_t\}_{t=0}^{\infty} \in \text{Int}\Lambda^{\infty}$, then

$$\mathcal{U}_{C_t} \cdot F_{L_t} = -\mathcal{U}_{L_t} \ , \tag{24}$$

$$F_{K_{t+1}} \cdot \mathcal{U}_{C_{t+1}} = \mathcal{U}_{C_t}. \tag{25}$$

at time t (or as $t \to \infty$). Here, the first condition states that the marginal cost of effort at time t given the utility function $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is 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 (24) and (25) hold as $t \to \infty$, then it is also clear that $\{C_t, L_t, K_{t+1}, x_t\}_{t\to\infty}$ must be undistorted. We then 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 labour supply if $\mathcal{U}_{C_t} \cdot F_{L_t} = -\mathcal{U}_{L_t}$ as $t \to \infty$. By implication, an allocation $\{C_t, L_t, K_{t+1}, x_t\}_{t=0}^{\infty}$ features *labour distortions* at time t if (24) is not satisfied at t. We refer to these as *downward labour distortions* if the left-hand side of (24) is strictly greater than the right-hand side. If (25) is not satisfied, then there are *intertemporal distortions* at time t, and if the left-hand side of (25) is strictly less than the right-hand side, then there are *downward intertemporal distortions.* Downward distortions imply that there is less labour supply and less capital accumulation than in an undistorted allocation.

Proposition 3 below clarifies the connection between aggregate labour distortions and non-linear labour income taxation. It should be noted, however, that there is no such connection between aggregate intertemporal distortions and capital taxes in general. In particular, lack of intertemporal distortions does not necessarily imply that the intertemporal decisions of individual agents will be undistorted. In particular, since these distortions are expressed in terms of the indirect utility function $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ and there are additional individual-level incentive compatibility constraints, there is no general guarantee that the intertemporal rate of substitution of each agent will coincide with those implied by the indirect utility function. In Section 4.4, we provide two canonical economies where the behaviour of individual-level distortions can be determined explicitly.

4.2. *The best sustainable mechanism*

Theorem 4. ***(Best Sustainable Mechanism)*** *Consider the optimal dynamic Mirrlees economy with self-interested politicians described above. Suppose that Assumptions 1–3 hold and there exists $\{C_t, L_t\}_{t=0}^{\infty} \in Int\Lambda^{\infty}$ (with $L_t > 0$) for some t. Then, in the best sustainable mechanism:*

1. *there are downward labour 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, labour supply and capital levels $\{C_t, L_t, K_{t+1}\}_{t=0}^{\infty}$. Suppose a steady state exists such that as $t \to \infty$, $\{C_t, L_t, K_{t+1}\}_{t=0}^{\infty} \to (C^*, L^*, K^*)$, where (C^*, L^*) is interior. Moreover, let $\varphi \equiv \inf\{\varrho \in (0,1] : \text{plim}_{t\to\infty} \varrho^{-t}\mathcal{U}^*_{C_t} = 0\}$, where $\varphi < 1$. Then:
2. *if $\varphi = \delta$, then there are no asymptotic aggregate distortions on capital accumulation and labour supply.*
3. *if $\varphi > \delta$, then aggregate distortions on capital accumulation and labour supply do not disappear even asymptotically.*

Proof. The proof of this theorem is related to that of Theorem 2 in Acemoglu, Golosov and Tsyvinski (2008*a*). But it requires a different mathematical argument. The details are presented in Appendix B, where we first show that, when randomizations are introduced, $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is a well-defined, continuous, concave, and differentiable functional. We then provide a non-recursive optimization problem that characterizes the best sustainable mechanism and study the properties of the limit of the solution to this problem as $t \to \infty$. ||

The first part of the theorem states that the sustainability constraint of the politician, (14), necessarily introduces a distortion. Intuitively, this additional (aggregate) distortion arises because, as output increases, the sustainability constraint (14) requires that more rents be given to the politicians in power. These additional rents increase the effective cost of production. The best sustainable mechanism creates distortions so as to reduce the level of output and thus the rents that have to be paid to the politician. Intuitively, starting from an undistorted allocation, a small reduction in labour supply and capital causes a second-order loss in output, but a first-order decline in the amount of rents that need to be paid to the politician. Consequently, some amount of distortion reducing labour supply and capital is optimal from the viewpoint of the citizens.

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 ensures that the sustainability constraint will not bind in the long run and thus distortions will vanish.

Notice that the results in this theorem compare δ to φ. Here φ is the rate at which the *ex ante* marginal utility of consumption $\mathcal{U}^*_{C_t}$ is declining in the steady state and cannot be excessively derived in terms of primitives without further assumptions.[24] Clearly, in the case

[24] This is a common problem in dynamic incentive problems and is due to the intertemporal nature of the incentive constraints. Nevertheless, the numerical computation of φ in most applications is straightforward; see, for example, Golosov, Tsyvinski and Werning (2006) and Albanesi and Sleet (2005).

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 β. We will show next that in two important cases this will indeed be the case. In the more general case of the present section, φ 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 $t+1$. Additionally, we will show in the next section that without any dynamic incentive linkages, e.g. with constant types, this fundamental discount factor coincides with β, though in general it may be different from β. Therefore, the case of $\varphi = \delta$ indeed corresponds to a situation in which the politician is as patient as the citizens.[25]

Part 3, on the other hand, states that if the discount factor of the politician δ is sufficiently low compared to the fundamental discount factor φ, 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.

4.3. *Application 1: constant types*

Theorem 4 characterizes the behaviour of the distortions introduced by political economy and commitment problems and their asymptotic behaviour. But the results are in terms of the fundamental discount factor φ obtained from the indirect utility function $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$. In this and the next subsection, we strengthen the results of Theorem 4 and derive results in terms of the discount factor of the citizens β. In this subsection, we focus on economies with *constant types*, where $\theta_t^i = \theta_{t+1}^i$ for all i and t, that is, economies in which individual types are realized in the first date and remain constant thereafter. This is the assumption that is used in much of the literature on dynamic mechanisms without commitment (e.g. Roberts, 1984; Freixas, Guesnerie and Tirole, 1985; Bisin and Rampini, 2005).

With constant types, truthful reporting along the equilibrium path (cf. Theorem 1) implies that individual incentive compatibility constraints, (12) and (13), can now be written as

$$\sum_{t=0}^{\infty} \beta^t u\left(c_t(\theta), l_t(\theta) \mid \theta\right) \geq \sum_{t=0}^{\infty} \beta^t u\left(c_t(\hat{\theta}), l_t(\hat{\theta}) \mid \theta\right), \text{ and}$$

$$\sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}(\theta_{t+s}), l_{t+s}(\theta_{t+s}) \mid \theta_t\right) \geq \frac{1}{1-\beta} u\left(0, 0 \mid \theta_t\right) \text{ (for each } t) \tag{26}$$

for all $\hat{\theta} \in \Theta$ and $\theta \in \Theta$ (with a slight abuse of notation). Since types are known at all dates, the only reason why aggregates L_t and C_t will vary in this case is because of changes in the rents paid to the politician in power, x_t. In this case, we have the following result.

Proposition 1. ***(Best Sustainable Mechanisms with Constant Types)*** *Consider the optimal dynamic Mirrlees economy with self-interested politicians described above. Assume that types are constant, that is, $\theta_t^i = \theta_{t+1}^i$ for all $i \in I$ and $t = 0, 1, \ldots$. Suppose moreover that*

[25] 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. Proposition 2 in the next section explicitly deals with non-interior steady states. Proposition 2 also provides primitive conditions under which we will have $\varphi = \delta$, thus illustrating that this condition is not "knife-edge".

Assumptions 1–3 hold and there exists $\{C_t, L_t\}_{t=0}^{\infty} \in Int\Lambda^{\infty}$ *(with* $L_t > 0$*) for some* t*. Then, in the best sustainable mechanism,* $\varphi = \beta$*.*

Proof. See Appendix A. ||

This result means that we can directly apply Theorem 4 to establish the characterization of the best sustainable mechanism for this case. There are downward labour distortions at some $t < \infty$ and downward intertemporal distortions at $t - 1$ (provided that $t \geq 1$). Suppose an interior steady state. If $\beta = \delta$, then there are no asymptotic aggregate distortions on capital accumulation and labour supply. If $\beta > \delta$, then aggregate distortions on capital accumulation and labour supply do not disappear even asymptotically.

4.4. *Application 2: private histories*

In this subsection, we focus on economies with *private histories*, where individual histories will not be observed by the politicians (so in this case allocations can only be conditioned on current reports). This restriction enables us to focus on the main interplay between private and public incentives, without introducing the substantial complications that arise when individuals are given complex dynamic incentives. We show that in this case we again have $\varphi = \beta$ and in addition, we provide a tighter characterization of the behaviour of distortions.

Let us simplify the exposition and the notation by assuming that within each period, there is an aggregate invariant distribution of types, denoted by G, and also by removing capital, so that the aggregate production function of the economy is

$$Y_t = L_t, \tag{27}$$

where $K_0 = 0$ and L_t denotes the aggregate labour supply at time t. These simplifications are without any significant consequences for our analysis.

Private histories imply 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

$$u\left(c_t\left(\theta_t\right), l_t\left(\theta_t\right) \mid \theta_t\right) \geq u\left(c_t\left(\hat{\theta}_t\right), l_t\left(\hat{\theta}_t\right) \mid \theta_t\right) \text{ and } \geq u\left(0, 0 \mid \theta_t\right) \tag{28}$$

for all $\hat{\theta}_t \in \Theta$ and $\theta_t \in \Theta$, and for all t.

The best sustainable mechanism with private histories therefore maximizes (10) subject to (14), (28) and the resource constraint

$$C_t + x_t \leq L_t. \tag{29}$$

Returning to the quasi-Mirrlees program defined above, it is straightforward to see that with 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 $U : \mathbb{R}_+^2 \to \mathbb{R}$. The results for $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ in Appendix B immediately imply that $U\,(C, L)$ is also well defined, concave, and differentiable.

The incentive compatibility constraints for individuals in (28) play a similar role to (16) in our formulation above. In particular, we can define a static set of feasible aggregate consumption and labour supply levels,

$$\Lambda = \{ \quad (C, L) \text{ such that } \exists \{c(\theta), l(\theta)\} \text{ satisfying (28) and} \\ C = \int c(\theta) \, dG(\theta), \text{ and } L = \int l(\theta) \, dG(\theta)\}. \tag{30}$$

The program for the best sustainable mechanism, (22)–(23), can now be written as:

$$\max_{\{C_t, L_t, x_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t U(C_t, L_t) \tag{31}$$

subject to the resource constraint, (29), (30), and the sustainability constraint,

$$w_t \equiv \mathbb{E}\left[\sum_{s=0}^{\infty} \delta^s v(x_{t+s})\right] \geq v(L_t), \tag{32}$$

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

Finally, we also adopt the following sustainability assumption, which will be used in establishing convergence to a steady state and in Part 2 of the next proposition (in particular, when the utility provided to a politician reaches the boundary of the set of feasible values). Let $\overline{w} \equiv \max_{(C,L)\in\Lambda} v(L - C) / (1 - \delta)$.

Assumption 4. ***(sustainability)*** *There exists $\left(\overline{C}, \overline{L}\right) \in \arg\max_{(C,L)\in\Lambda} v(L - C) / (1 - \delta)$, such that $v\left(\overline{L} - \overline{C}\right) / (1 - \delta) > v\left(\overline{L}\right)$.*

Intuitively, this assumption ensures that the highest discounted utility that can be given to the politician is sufficient to satisfy the sustainability constraint (32). Clearly this assumption is satisfied when the politician's discount factor, δ, is sufficiently large.

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 (20)–(21) satisfies:

$$U_C(C, L) = -U_L(C, L), \tag{33}$$

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 labour distortion* if the left-hand side of (33) is strictly greater than the right-hand side.

The main result of this section is the following proposition:

Proposition 2. ***(Best Sustainable Mechanisms with Private Histories)*** *Consider the economy with no capital and with private histories. Suppose also that Assumptions 1, 3, and 4 hold.*

1. *At $t = 0$, there is an aggregate distortion.*
2. *Suppose that $\beta \leq \delta$. Let Γ^* 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 $w^* \in [0, \overline{w}]$ 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 proposition follow as corollaries of the corresponding results in Theorem 4. 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 (2008*a*), and we do not repeat these proofs to economize on space. ||

Proposition 2 provides a tighter characterization of the best sustainable equilibrium when histories are private. 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 Proposition 2 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 δ 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 4 guarantees that if the boundary $\overline{w}$ is reached the allocation will be undistorted. Finally, Part 2 of the Proposition extends results for the case of politicians being more patient than agents.

4.5. *Interpretation of distortions*

We can also use the economies with private histories or with constant types to clarify the meaning of aggregate distortions. To do this in the cleanest possible fashion, let us focus on the economy with private histories (the results are identical with constant types).

Given the single crossing property (from Assumption 1), the set of incentive compatibility constraints with private histories, (28), can be reduced to a set of incentive compatibility constraints only for neighbouring types. Since there are $N + 1$ types in Θ, this implies that (28) is equivalent to N incentive compatibility constraints. This then enables us to establish the following proposition, which illustrates the relationship between aggregate distortions and individual income taxes:

Proposition 3. *Suppose Assumption 1 holds and suppose that the best sustainable mechanism does not involve randomization. Consider a sequence of* $\{C_t, L_t\}_{t=0}^{\infty}$. *Then:*
1. the marginal labour tax rate on the highest type of agent, θ_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 labour supply decision of the highest type of agent is undistorted, i.e.* $u_c(c_t(\theta_N), l_t(\theta_N) \mid \theta_N) = -u_l(c_t(\theta_N), l_t(\theta_N) \mid \theta_N)$.

Proof. The single crossing property in Assumption 1 implies that we only need to check incentive compatibility constraints for neighbouring types. Let u_c and u_l be the partial derivatives of u (which exist by Assumption 1). Since there is no randomization, we have

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

where λ_{Nt} is the multiplier on incentive compatibility constraint between types θ_N and θ_{N-1} at time t, ν_{Ct} is the multiplier on (18) at t, and ν_{Lt} is the multiplier on (19) 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

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

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 (33). ||

This proposition therefore further clarifies the meaning of the aggregate distortions, which have been our focus so far, and shows that they are naturally linked to the marginal labour taxes in the standard Mirrlees problem. In particular, if in the standard Mirrlees problem the marginal income tax on the highest type should be equal to zero, then the added distortion is exactly equal to the tax that will be imposed on the highest type under the best sustainable mechanism.

5. BENEVOLENT GOVERNMENTS AND *EX POST* POLITICAL CONFLICT

Our analysis so far has provided a general approach and various characterization results for the analysis of the structure of taxation in a dynamic economy subject to political economy constraints. This analysis was predicated on a number of assumptions. In particular, the politician in power was assumed to be entirely self-interested, there was no political economy conflict among the citizens (the only political economy interaction being between the citizens and the politician), various assumptions were imposed on possible deviations of the politician, and we focused on perfect Bayesian equilibria. These assumptions enabled us to obtain a tractable characterization of the dynamic non-linear taxation problem. In particular, Theorem 1 established truthful revelation, and Theorem 3 showed a strong separation between the provision of private and public incentives. In this section, we consider environments with partially benevolent governments and *ex post* political conflict among the citizens. In these extended environments, Theorem 3 no longer applies. We therefore develop an alternative (though related) mathematical approach, which enables us to generalize the main insights. We end by emphasizing various theoretical limits to our results.

5.1. *Benevolent government*

An important question is whether the results presented so far are informative about (generalize to) environments where politicians or the government are partially benevolent. These environments include those considered by Roberts (1984), Freixas, Guesnerie and Tirole (1985), or Bisin and Rampini (2005), where the government is benevolent, but “time inconsistent”, i.e. unable to commit to a full dynamic mechanism.

Benevolent government can be modelled by considering a more general utility function for the government of the form:

$$\sum_{s=0}^{\infty} \delta^s \left[(1-a)\, v(x_{t+s}) + a \left(\mathbb{E}_{t+s} \int u\left(c_{t+s}, l_{t+s} \mid \theta^{t+s}\right) dG^{t+s}\left(\theta^{t+s}\right) \right) \right], \tag{34}$$

where the second term is the average (expected) utility of the citizens at time $t+s$, and $0 < a < 1$. Therefore, this utility function is arbitrarily close to that of a purely self-interested government when $a \to 0$. Below we provide a generalization of our results to an environment, for which the benevolent government with utility function (34) is a special case. These results show that, under an additional technical assumption, essentially all of the results derived so far apply in this case.

5.2. *Ex post political conflict*

The results provided so far only allowed for conflict between the politician and the citizens as a whole. In many societies, political economy conflicts are more multifaceted, involving both a conflict of interest between the politician and the citizens and redistributive conflict among the citizens. We now outline how such *ex post* conflict can be incorporated into our model, and in the next subsection we provide general results that apply in these cases.

The main insight that enables us to model *ex post* conflict is that many models of redistributive conflict lead to equilibria that are isomorphic to the maximization of a weighted social welfare function. These include: (1) benchmark models of lobbying, where different lobbies compete by offering payment schedules to politicians to obtain redistributive policies in their favour (e.g. Grossman and Helpman, 1994); (2) models of probabilistic voting, in which individuals vote over redistributive policies and also receive idiosyncratic and common taste shocks affecting their voting decisions (e.g. Lindbeck and Weibull, 1987); (3) models with party capture.[26] For example, in the benchmark lobbying model of Grossman and Helpman (1994), which builds on Becker's (1983) insights, equilibria maximize a weighted average of the utilities of different individuals in the society, with those who are part of organized lobbies that can offer contributions to politicians receiving greater weights.

Alternatively, the following result on probabilistic voting models is presented in Acemoglu (2007). Consider a society consisting of N individuals, each with a utility function $u^i(p)$ defined over some policy vector p. Suppose that there is electoral competition between two parties that care about their share of votes, and individuals vote according to the utility from the policies offered by the two parties and also a stochastic variable ε_i that determines their relative preference for one of the parties. If each ε_i has a smooth distribution F_ε^i, then any pure-strategy equilibrium of the probabilistic voting model will give the same allocation as the maximization of $\sum_{i=1}^N \phi^i u^i(p)$ for some sequence of strictly positive numbers ϕ^i.[27]

Intuitively, *ex post* conflict creates incentives for politicians to shift resources across individuals or groups until the electoral or monetary return to the politician from providing resources to different groups is equalized. This leads to the maximization of a weighted average of their utilities. In the context of a dynamic model such as ours, this would mean maximizing a weighted average of the continuation utilities of all individuals (which are themselves functions of their realized histories up to that date). Therefore, we can consider the following general maximization problem for the government

$$\sum_{s=0}^{\infty} \delta^s \left[(1-a)\, v(x_{t+s}) + a \left(\mathbb{E}_{t+s} \int u\left(c_{t+s}, l_{t+s} \mid \theta^{t+s}\right) d\tilde{G}^{t+s}\left(\theta^{t+s}\right) \right) \right], \tag{35}$$

which is similar to (34) except that the integration in the second term is with respect to some arbitrary distribution $\tilde{G}$ rather than G. We assume throughout that $\tilde{G}^t$ (for each t) is a probability measure (which is just a normalization, since v can be changed accordingly) and that $\tilde{G}^t$ and G^t (for each t) are absolutely continuous with respect to each other (so that all realized histories of types receive positive weight under the $\tilde{G}$'s as well as the G's). This utility function incorporates both the weighted average of the utilities of the citizens (representing

[26] See the discussion of these three sets of models in Acemoglu and Robinson (2007, Appendix) and in Acemoglu (2007). The lobbying and probabilistic voting models are also extensively discussed in Persson and Tabellini (2000).

[27] However, guaranteeing that a pure-strategy equilibrium exists is harder. Furthermore, if each F_ε^i is symmetric (treating the two parties symmetrically), then $\phi^i = F_\varepsilon^i(0)$ for each i.

ex post conflict) and the utility of the politician in power, given by the term with $v(x)$, corresponding to the self-interested motives of those controlling the government.[28]

In the next subsection, we show that when the government in power maximizes (35) and the appropriate punishments are instituted, analogues of the main theorems presented so far continue to hold. Since the utility function of the partially benevolent government is a special case of (35), this analysis also generalizes our results to societies with partially benevolent politicians and governments.

5.3. *General results*

To derive the general results with government utility given by (35), we change the political game. In particular, we no longer allow citizens to vote the current government out of office (thus their strategy simply corresponds to α rather than $\{\alpha, \rho\}$ as before). Instead, the same government is always in power. Despite this, all of the main results so far continue to hold, because 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 exercise their freedom of labour supply and produce zero output thereafter.[29] 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 partially benevolent government. In addition, we also 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; Farhi and Werning, 2008).

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

The next theorem shows that Theorem 1 and Proposition 2 continue to hold in this more general environment.

Theorem 5. ***(Truthful Revelation)*** *Suppose that government utility is given by (35) and that Assumptions 1′, 2 and 3 hold. Then for any combination of strategy profiles Γ and $\underline{\alpha}$ that support a sustainable mechanism, there exists another pair of equilibrium strategy profiles Γ^* and $\underline{\alpha}^* = (\alpha^* \mid \alpha')$ for some α' such that Γ^* induces direct submechanisms, $\underline{\alpha}^*$ induces truth telling along the equilibrium path, and $\underline{c}[\Gamma, \underline{\alpha}] = \underline{c}[\Gamma^*, \underline{\alpha}^*]$, $\underline{l}[\Gamma, \underline{\alpha}] = \underline{l}[\Gamma^*, \underline{\alpha}^*]$ and $x[\Gamma, \underline{\alpha}] = x[\Gamma^*, \underline{\alpha}^*]$. Moreover, the best sustainable mechanism is a solution to maximizing*

[28] The recent paper by Farhi and Werning (2008) also uses a social welfare function representation in order to study the political economy of non-linear taxation. They focus on an unweighted average of the utilities of all agents and motivate this with probabilistic voting. Their results rely on the specific structure of the social welfare function and do not feature self-interest of politicians. On the other hand, Farhi and Werning's results correspond to the case with $a = 1$, which is not covered by our theorems, and more importantly, they provide a tight and elegant characterization of the degree of progressivity of taxes, which is not possible given the level of generality here.

[29] With these punishments strategies, equilibria are no longer renegotiation-proof. Nevertheless, it is possible to extend the game considered here, so that even though politicians are partially benevolent, there is still replacement of politicians (and a politician who is replaced still cares about the average utility of the citizens), and obtain similar results. We do not introduce this somewhat more involved game form to economize of space.

(10) subject to (11), (12) and the government sustainability constraint:

$$\sum_{s=0}^{\infty}\delta^{s}\left[(1-a)\,v\left(x_{t+s}\right)+a\left(\mathbb{E}_{t+s}\int\left[u\left(c\left(\theta^{t+s}\right)\right)-\chi\left(l\left(\theta^{t+s}\right)\mid\theta_{t+s}\right)\right]d\tilde{G}^{t+s}\left(\theta^{t+s}\right)\right)\right] \tag{36}$$

$$\geq \max_{\tilde{x}_t'+\int\tilde{c}_t'(\theta^t)dG^t(\theta^t)\leq F(K_t,L_t)}(1-a)\,v\left(\tilde{x}_t'\right)+a\int\left[u\left(\tilde{c}'\left(\theta^t\right)\right)-\chi\left(l\left(\theta^t\right)\mid\theta_t\right)\right]d\tilde{G}^t\left(\theta^t\right)$$

for all t.

Proof. See Appendix A. ||

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 (35), 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 labour. To establish Theorem 5, all we need to show is that such punishment strategies are sequentially rational. When all other agents choose zero labour supply, following any deviation to positive labour 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 labour, and thus it is sequentially rational for all citizens to supply zero labour following a deviation by the government.

This theorem therefore shows that revelation principle applies to the case where the government maximizes (35), 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 a disabled type, who can only supply $l = 0$, and has utility $u\left(c, \cdot \mid \theta = 0\right) = u\left(c\right)$, while the utility of type $\theta = 1$ is $u\left(c, l \mid \theta = 1\right) = u\left(c - \chi_1\left(l\right)\right)$, with $\chi_1\left(\cdot\right)$ strictly increasing in l. Furthermore, suppose that aggregate output is linear in labour and that the government is fully benevolent, i.e. $a = 1$ in terms of the utility function in (35). 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 $\chi_1\left(l'\right) < 1$. 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.

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

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

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

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

The next theorem provides a characterization of the structure of distortions and their asymptotic behaviour for the case of constant types and under the assumption that the politician and the citizens have the same discount factor, i.e., $\beta = \delta$. We start with this environment, since constant types constitute the most commonly-studied case and enable us to obtain the main results in a succinct fashion. Theorem 7 below generalizes the results of this theorem to non-constant types.

Suppose that there are $N+1$ types, i.e., $\Theta = \{\theta_0, \theta_1, ..., \theta_N\}$, ranked in ascending order of skills, and with respective probabilities $\{\pi_0, \pi_1, ..., \pi_N\}$. To prove the next theorem, we will impose the following technical assumption, which is the analogue of Assumption 4 for this environment. In particular, this assumption ensures that a limiting allocation without distortions is feasible. More specifically, let us define:

$$W\,(K_0) \equiv \max_{\left\{K_{t+1}, x_t, \{c_t(\theta_i), l_t(\theta_i)\}_{i=0}^{N}\right\}_{t=0}^{\infty}} \times \sum_{t=0}^{\infty} \delta^t \left\{(1-a)\,v\,(x_t) + a\left(\sum_{i=0}^{N} \pi_i' \left[u\,(c_t\,(\theta_i)) - \chi\,(l_t\,(\theta_i) \mid \theta_i)\right]\right)\right\}$$

subject to the resource constraint $x_t + K_{t+1} + \sum_{i=0}^{N} \pi_i c_t\,(\theta_i) \leq F\left(K_t, \sum_{i=1}^{N} \pi_i l_t\,(\theta_i)\right)$ (for each t) and the freedom of labour supply constraint $\sum_{s=0}^{\infty} \beta^{t+s} [u\,(c_{t+s}\,(\theta_i)) - \chi\,(l_{t+s}\,(\theta_i) \mid \theta_i)] \geq [u\,(0) - \chi\,(0 \mid \theta_i)]\,/\,(1-\beta)$ (for each t), and starting with initial capital stock K_0. Therefore, $W\,(K_0)$ is the maximum utility that the government can reach starting with capital stock K_0 and subject only to the freedom of labour supply of the agents (and of course the resource constraint). Let the implied labour supply of type θ_i in the current period in the solution of this maximization be denoted by $l^*\,(\theta_i, K_0)$. Finally, let us also define

$$V\left(K, \{l\,(\theta_i)\}_{i=0}^{N}\right) \equiv \max_{\left\{\tilde{x}', \{\tilde{c}'(\theta_i)\}_{i=0}^{N}\right\}} (1-a)\,v\left(\tilde{x}'\right) + a\left(\int \left[u\left(\tilde{c}'\,(\theta_i)\right) - \chi\,(l\,(\theta_i) \mid \theta_i)\right] d\tilde{G}\,(\theta_i)\right) \tag{37}$$

subject to $\tilde{x}' + \sum_{i=0}^{N} \pi_i \tilde{c}'\,(\theta_i) \leq F\left(K, \sum_{i=0}^{N} l\,(\theta_i)\right)$ as the deviation utility of the government starting with capital stock K and after labour supply decisions $\{l\,(\theta_i)\}$ have been made. This expression uses the fact that the government gives weights $\tilde{G}\,(\theta_i)$ to the different types and incorporates the result of Theorem 4 that after a deviation by the government there will be zero labour supply in all future dates. Then the analogue of Assumption 4 is:

Assumption 5. ***(generalized sustainability)*** *For any* $K > 0$,

$$W\,(K) \geq V\left(K, \left\{l^*\,(\theta_i, K)\right\}_{i=0}^{N}\right).$$

Intuitively, Assumption A states that if, starting with some positive capital stock, all future allocations are chosen to maximize the utility of the government, then the sustainability constraint of the government is slack. Under this assumption, we prove the following result establishing that when types are constant and $\beta = \delta$, distortions disappear in the long run.

Theorem 6. ***(Best Sustainable Mechanism with Constant Types)*** *Suppose that government utility is given by (35) with $a \in (0, 1)$ and that Assumptions 1′, 2, 3 and 5 hold. Furthermore, assume that there are constant types and $\beta = \delta$. Then, asymptotically there are no aggregate distortions on labour supply and capital accumulation.*

Proof. See Appendix A. ||

This theorem implies that in an economy with constant types, aggregate distortions disappear regardless of the degree of benevolence of the government and also regardless of the exact weights that the government puts on the utilities of different citizens. Consequently, this result applies to both situations in which the government is benevolent and also to situations in which there is *ex post* political economy conflict among the citizens. Theorem 6 therefore shows that in an important benchmark economy with fully persistent types and either *ex post* conflict or partially benevolent governments, there will be no aggregate capital taxes and no further taxes on labour beyond those implied by a full-commitment Mirrlees economy.[30] In the case where $a \to 1$, 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.

We next present a generalization of Theorem 6, which parallels our general result, Theorem 4. This theorem illustrates the importance of the discount factor of the politicians in the environment with that partially benevolent government, though, now, the fundamental discount factor to which the discount factor of the politicians is being compared to is a more complicated object than φ defined in Theorem 4, making the result somewhat weaker.

For this theorem, let $\{\mu_t^*\}_{t=0}^{\infty}$ denote the multipliers associated with the sustainability constraints of the government and let $\mathcal{U}\left(\{C_t, L_t\}_{t=0}^{\infty}; \{\mu_t^*\}_{t=0}^{\infty}\right)$ denote the indirect utility functional, which now has to be conditioned on this sequence of multipliers as well. As in Theorem 4, let $\varphi \equiv \inf\{\varrho \in [0, 1] : \operatorname{plim}_{t\to\infty} \varrho^{-t}\mathcal{U}_{C_t}^* \left(\{C_t, L_t\}_{t=0}^{\infty}; \{\mu_t^*\}_{t=0}^{\infty}\right) = 0\}$ (which is now a function of the sequence $\{\mu_t^*\}$ as well as $\{C_t^*, L_t^*\}$). Let us also define Λ^∞ analogously to (16).

Theorem 7. ***(Best Sustainable Mechanism: General Case)*** *Suppose that government utility is given by (35) with $a \in (0, 1)$, that Assumptions 1′, 2 and 3 hold, and that there exists $\{C_t, L_t\}_{t=0}^{\infty} \in Int\Lambda^\infty$ (with $L_t > 0$) for some t. Then, in the best sustainable mechanism:*

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

[30] 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).

Let the best sustainable mechanism induce a sequence of consumption, labour supply and capital levels $\{C_t, L_t, K_{t+1}\}_{t=0}^{\infty}$. Suppose a steady state exists such that as $t \to \infty$, $\{C_t, L_t, K_{t+1}\}_{t=0}^{\infty} \to (C^*, L^*, K^*)$, where (C^*, L^*) is interior. Then:

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

Proof. See Appendix A. ||

This theorem therefore shows that the fundamental results from Theorem 4 generalize to the case with a partially benevolent government and with potential *ex post* political economy conflict. Since the fundamental discount factor, φ, is now an even more complicated object and certainly not easy to compute, the results in this theorem are weaker than those in Theorem 4. Nevertheless, Theorem 6 showed that even with this more general formulation, sharper results can be obtained when types are constant.

5.4. *Theoretical limits to separation of private and public incentives*

In this subsection, we discuss various theoretical limits to our main characterization results, in particular, to Theorems 1 and 3.

5.4.1. Markovian strategies. Several recent studies focus on Markov equilibria of dynamic political economy games (Hassler *et al.*, 2005; Battaglini and Coate, 2008). In contrast to these approaches, the results in this paper use trigger strategies (for example, in deriving Theorems 1 and 3). These results do not generalize when attention is restricted to Markovian strategies. A simple example of how the results change radically even when we restrict attention to stationary strategies (which is a superset of Markovian strategies) is illustrated in the "representative agent" version used in Acemoglu, Golosov and Tsyvinski (2008*a*, Proposition 2). We do not repeat these results here to save space. Nevertheless, it can also be noted that all of the results in this paper can be supported as "quasi-Markovian equilibria", where the set of state variables is augmented by a single non-payoff-relevant variable, which denotes whether the government has deviated from its promises. Citizens vote the government out or use other punishments if there has been such a deviation. Note however that this type of quasi-Markovian equilibrium is considerably more complex than Markov perfect equilibria, and in particular would be nonstationary.

5.4.2. Different deviations. Our results have also assumed a specific form for the (best) deviation by the government. In particular, the politician in power can deviate to capture all of the output produced within the period. This feature is not important *per se*. For example, we can generalize the results to an environment where the government can only freely use a fraction $\eta \leq 1$ of the total output of the economy. In this case, the level of η could be related to the institutional controls on government or politician behaviour. In this case, the constraint on the government following a deviation would be $\tilde{K}'_{t+1}\left(h^t\right) = \eta F\left(K_t, L_t\right) - \tilde{x}'_t - \int_{i \in I} \tilde{c}'_t\left(z^{i,t}\right) di$, with the remaining $1 - \eta$ fraction of the output getting destroyed. Nevertheless, other types of plausible deviations by the politician would not be sufficient to establish the key separation result, Theorem 3. The essential feature of this theorem is that politicians' deviation payoffs depend only on aggregates. For example, if, instead of $x_t = F\left(K_t, L_t\right)$, the maximum consumption for the politician were a non-linear function of the entire distribution of labour supplies, $[l_{i,t}]_{i \in I}$, Theorem 3 would not necessarily hold. In essence, the separation of private and public incentives in Theorem 3 requires that the best deviation utility of the politician (or the government) should be independent of the distribution of resources (including labour supply) among the citizens. While this is true in our environment and would be true in a number of other situations, there are also various important dynamic resource allocation problems in which it may not hold.

6. CONCLUSIONS

In this paper, we took a first step towards a political-economic analysis of dynamic and non-linear taxation. Political economy considerations become particularly important in the context of non-linear taxation, which involves a significant amount of information and enforcement power being concentrated in the hands of the government and policymakers. Unless these policymakers can commit to the entire future path of taxes, the structure of taxation must not only provide incentives to individuals, but also respect political economy constraints—that is, it should provide the appropriate incentives to policymakers. Despite the complex set of issues that arise in balancing private and public incentives, it is possible to develop a relatively tractable framework for the analysis of how political economy constraints affect the structure of taxation.

To achieve this objective, we focused on the best sustainable equilibrium, i.e. the best equilibrium that satisfies the incentive compatibility constraints of politicians. We showed how *sustainable mechanisms*, where the politician in power is given incentives not to misuse resources and information, can be constructed in the infinite-horizon economy we study. 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 provided a characterization of the best sustainable mechanism. Political economy considerations introduce additional constraints on the optimal taxation problem, but these constraints are intuitive and relatively simple to characterize. In particular, we showed 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 labour supply in the economy. These constraints then lead to new (political economy) distortions and change the structure of taxation.

Using this approach, we provided a systematic characterization of these distortions and their evolution over time. We showed that when politicians are as patient as, or more patient than, citizens, aggregate capital and labour 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 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. To the extent that smaller discount factors for politicians than for the citizens are a reasonable approximation to reality, our results also suggest a possible explanation for understanding distortionary long run taxes on labour and capital.

We also showed how our results generalize to certain other environments, in particular, to situations in which there may be *ex post* political economy conflict among the citizens and also to environments with partially benevolent governments. A dynamic political economy analysis of other policies, including regulation, contract enforcement, and other forms of redistribution, is an important and fruitful area for future research.

APPENDIX A. PROOFS

Proof of Proposition 1

Proof. By Theorem 3, the best sustainable mechanism also solves in the problem of maximizing (17) subject to incentive compatibility constraints (26) as well as (18) and (19) (for a given sequence $\{C_t, L_t\}_{t=0}^{\infty}$). We now write this quasi-Mirrlees program explicitly taking into account randomizations. To simplify exposition, we focus on randomizations across individuals and ignore randomizations across different levels of x_t's. Applying Caratheodory's Theorem (e.g. Proposition 1.3.1 in Bertsekas, Nedic and Ozdaglar, 2003, pp. 37–38) to the problem of providing a certain level of utility to each type at a given date, it is sufficient to focus on a finite number of consumption levels for each type $\theta \in \Theta$ at each date. Let the set of these finite consumption levels for type θ at time t be $M_t(\theta)$. Denote these consumption levels for each $m(\theta) \in M_t(\theta)$ by $c_t^{m(\theta)}(\theta)$ for $\theta \in \Theta$ and time t, and denote the probability that this consumption level will be given to an individual who has announced his type as θ by $q_t^{m(\theta)}(\theta)$. Finally, let us denote the probability that an individual will be of type θ by $\pi(\theta)$. Then the quasi-Mirrlees program can be written as

$$\mathcal{U}\left(\{C_t, L_t\}_{t=0}^{\infty}\right) \equiv \max_{\left\{c_t^{m(\theta)}, l_t^{m(\theta)}\right\}_{t, m(\theta)\in M_t(\theta), \theta\in\Theta}} \sum_{\theta\in\Theta} \pi(\theta) \sum_{t=0}^{\infty} \beta^t \sum_{m(\theta)\in M_t(\theta)} \left[q_t^{m(\theta)}(\theta)\, u\left(c_t^{m(\theta)}(\theta), l_t^{m(\theta)}(\theta) \mid \theta\right)\right]$$

subject to the incentive compatibility constraints (equivalent to (26)), which take the form

$$\sum_{t=0}^{\infty} \beta^t \sum_{m(\theta)\in M_t(\theta)} \left[q_t^{m(\theta)}(\theta)\, u\left(c_t^{m(\theta)}(\theta), l_t^{m(\theta)}(\theta) \mid \theta\right)\right] \geq \sum_{t=0}^{\infty} \beta^t \sum_{m(\widehat{\theta})\in M_t(\theta)} \left[q_t^{m(\widehat{\theta})}(\theta)\, u\left(c_t^{m(\widehat{\theta})}(\hat{\theta}), l_t^{m(\widehat{\theta})}(\hat{\theta}) \mid \theta\right)\right]$$

$$\sum_{s=0}^{\infty} \beta^s \pi(\theta) \left[q_{t+s}^{m(\theta)}(\theta)\, u\left(c_{t+s}^{m(\theta)}(\theta), l_{t+s}^{m(\theta)}(\theta) \mid \theta\right)\right] \geq \frac{1}{1-\beta} u(0, 0 \mid \theta) \text{ (for each } t) \tag{37}$$

for all $\theta \in \Theta$ and $\widehat{\theta} \in \Theta$, and versions of (18) and (19), which take the form

$$\begin{aligned} \sum_{\theta\in\Theta} \sum_{m(\theta)\in M_t(\theta)} \pi(\theta)\, q_t^{m(\theta)}(\theta)\, c_t^{m(\theta)}(\theta) &\leq C_t \\ \sum_{\theta\in\Theta} \sum_{m(\theta)\in M_t(\theta)} \pi(\theta)\, q_t^{m(\theta)}(\theta)\, l_t^{m(\theta)}(\theta) &\geq L_t \end{aligned} \tag{38}$$

for all t. Denote the multiplier of (38) by λ_t. Then, the differentiability of $\mathcal{U}\left(\{C_t, L_t\}_{t=0}^{\infty}\right)$ implies that $\mathcal{U}_{C_t}\left(\{C_t, L_t\}_{t=0}^{\infty}\right) = \lambda_t$. Let the multipliers for the incentive compatibility constraints be denoted by $\eta\left(\theta, \widehat{\theta}\right)$. There will exist a type θ^*, typically the highest type θ_N, such that $\eta(\theta, \theta^*) = 0$ for all $\theta \in \Theta$. Single crossing property (Assumption 1) implies that (37) does not bind for such θ^*. Then the first-order conditions with respect to consumption allocations in the quasi-Mirrlees program imply

$$\beta^t u_c\left(c_t^{m(\theta^*)}(\theta^*), l_t^{m(\theta^*)}(\theta^*) \mid \theta^*\right)\left[1 + \frac{\sum_{\theta\neq\theta^*} \eta(\theta^*, \theta)}{\pi(\theta^*)}\right] = \lambda_t. \tag{39}$$

Note that $1 + \sum_{\theta\neq\theta^*} \eta(\theta^*, \theta)/\pi(\theta^*)$ here is a constant independent of time and we denote it by $\overline{\eta}$. Since the (stochastic) sequences $\{c_t(\theta)\}_{t=0}^{\infty}$ and $\{l_t(\theta)\}_{t=0}^{\infty}$ are bounded by feasibility and u is continuously differentiable, $\left\{u_c\left(c_t^{m(\theta)}(\theta), l_t^{m(\theta)}(\theta) \mid \theta\right)\right\}_{t=0}^{\infty}$ is a bounded stochastic sequence. Fix a sequence $\{m_t(\theta^*)\}$ such that $m_t(\theta^*) \in M_t(\theta^*)$ for all t and $\limsup_{t\to\infty} q_t^{m_t(\theta^*)} > 0$ (such a sequence clearly exists). Then, every subsequence of $\left\{u_c\left(c_t^{m_t(\theta^*)}(\theta^*), l_t^{m_t(\theta^*)}(\theta^*) \mid \theta^*\right)\right\}_{t=0}^{\infty}$ is bounded, and in particular, $\limsup_{t\to\infty} u_c\left(c_t^{m_t(\theta^*)}(\theta^*), l_t^{m_t(\theta^*)}(\theta^*) \mid \theta^*\right) = \overline{u}_c^*$. This implies that $\limsup_{t\to\infty} \beta^{-t}\lambda_t = \overline{u}_c^*\overline{\eta}$. Therefore, $\varphi \equiv \inf\{\varrho \in (0, 1] : \operatorname{plim}_{t\to\infty} \varrho^{-t}\mathcal{U}_{C_t}^* = 0\} = \beta$, establishing that $\varphi = \beta$. This completes the proof of the proposition. ‖

Proof of Theorem 5

The proof of this theorem follows the structure of the proofs of Lemma 1, Theorem 1, and Proposition 2. The main difference here is that instead of replacing the politician, citizens play a null strategy of supplying zero labour.

Proof. Let $\tilde{c}_t\,[c]$ be the mapping that allocates a consumption level of $c \in [0, F(K_t, L_t)]$ to each individuals regardless of past and current reports (and the remainder to government consumption). As before, let $h^{t-1} \in \tilde{H}^{t-1}$ if $x_{t-s}\left(h^{t-s}\right) = \tilde{x}_{t-s}\left(h^{t-s}\right)$ and $M_{t-s} = \tilde{M}_{t-s}$ for all $s > 0$. Then the following strategy combination would ensure $v_t^c\left(\tilde{K}'_{t+1}, \tilde{c}'_t \mid M^t\right) = 0$ for all t: (1) for the citizens, $\underline{\alpha}= \left(\tilde{\alpha} \mid \alpha^{\emptyset}\right)$ for some $\tilde{\alpha}$, which means that for each citizen i and for all t, we have that if $h^{t-1} \in \tilde{H}^{t-1}$, then $\alpha_t^i = \tilde{\alpha}$, and if $h^{t-1} \notin \tilde{H}^{t-1}$, then $\alpha_t^i = \alpha^{\emptyset}$; (2) for the politician, Γ, such that if $h^{t-1} \in \tilde{H}^{t-1}$, then Γ involves $\tilde{x}_t = x_t$, $\tilde{M}_t = M_t$, and $\xi_t = 0$; and if $h^{t-1} \notin \tilde{H}^{t-1}$, then it involves $\xi_t = 1$, $\tilde{x}'_t = F(K_t, L_t) - c^*$, and $\tilde{c}'_t = \tilde{c}_t\left[c^*\right]$, where

$$c^* \in \arg\max_{c}\,(1-a)\,v\,(F(K_t, L_t) - c) + au\,(c)\,.$$

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 labour. Suppose that the government has announced a submechanism $\tilde{M}_t$ at time t and has capital stock K_t, and $\alpha_{t+s}^i = \alpha^{\varnothing}$ 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 labour is not profitable (we think of an individual with positive measure ε deviating, and take the limit $\varepsilon \to 0$, since there is a continuum of agents). Without the deviation, i' obtains utility $u\,(0)\,/\,(1-\beta)$ (since from Assumption $1'$, $\chi\,(0 \mid \theta) = 0$ for all $\theta \in \Theta$ and there will be no labour supply for any type in the continuation game). Now imagine a deviation to a message that corresponds to positive labour supply, say l', with $\chi\left(l' \mid \theta_t^{i'}\right) > \chi\left(0 \mid \theta_t^{i'}\right) = 0$ by definition. This will generate output $F\left(K_t, \varepsilon l'\right)$, since all other agents are supplying zero labour. Now imagine the behaviour of the government at the last stage of the game, conditional on $\alpha_{t+s}^i = \alpha^{\varnothing}$ for all $i \in [0, 1]$ and for all $s \geq 1$. Then the sequentially rational strategy of the government is to maximize (35) with $K_{t+1} = 0$, since there will be no production in future periods. Consequently, the utility-maximizing program of the government following the deviation is:

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

subject to $\tilde{x}'_t + \int \tilde{c}'_t\left(z^t\left(\alpha_t\left(\theta^t\right)\right)\right) dG^t\left(\theta^t\right) \leq F\left(K_t, \varepsilon l'\right)$, where recall that $z^t\left(\alpha_t\left(\theta^t\right)\right)$ is the history of reports up to time t by an individual of type θ^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 is to choose $\tilde{c}_t\left[c^*\right]$ as specified above, which involves redistributing what it does not consume itself equally across agents, i.e. $\tilde{c}'_t\left(z^t\left(\alpha_t\left(\theta^t\right)\right)\right) = c^*$ for all $z^t\left(\alpha_t\left(\theta^t\right)\right) \in Z^t$. However, as $\varepsilon \to 0$, $c^* \to 0$, and thus the deviation payoff of i' is $u\,(0) - \chi\left(l' \mid \theta^{i'}\right) + \beta\left(u\,(0) - \chi\left(0 \mid \theta^{i'}\right)\right) / (1-\beta) < \left(u\,(0) - \chi\left(0 \mid \theta^{i'}\right)\right) / (1-\beta)$, showing that a continuation strategy profile where all agents supply zero labour 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, $\alpha_{t+s}^i = \alpha^{\varnothing}$ for all $i \in [0, 1]$ and for all $s \geq 0$ is a best response against this deviation. Since maximal punishments are optimal, $\alpha_{t+s}^i = \alpha^{\varnothing}$ 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, $\alpha_{t+s}^i = \alpha^{\varnothing}$ 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:

$$\begin{aligned}
&\mathbb{E}_t \sum_{s=0}^{\infty} \delta^s \Bigg[(1-a)\,v\,(x_{t+s}) \\
&\quad + a\left(\int \left[u\left(c_{t+s}\left(z^{t+s}\left(\alpha_{t+s}\left(\theta^{t+s}\right)\right)\right)\right) - \chi\left(l_{t+s}\left(z^{t+s}\left(\alpha_{t+s}\left(\theta^{t+s}\right)\right)\right) \mid \theta_t\right)\right] d\tilde{G}^t\left(\theta^t\right)\right)\Bigg] \qquad (40)\\
&\geq \max_{\tilde{x}'_t + \int \tilde{c}'_t(\theta^t) dG(\theta^t) \leq F(K_t, L_t)} (1-a)\,v\,(\tilde{x}_t) + a\int \left[u\left(\tilde{c}'_t\left(\theta^t\right)\right) - \chi\left(l_t\left(\left(z^t\left(\alpha_t\left(\theta^t\right)\right)\right)\right) \mid \theta_t\right)\right] d\tilde{G}^t\left(\theta^t\right) \text{ for all } t.
\end{aligned}$$

Now, given an equilibrium pair of strategy profiles Γ and $\underline{\alpha}$, exactly the same argument as in the proof of Theorem 1 implies that there exists another pair of equilibrium strategy profiles Γ^* and $\underline{\alpha}^* = \left(\alpha^* \mid \alpha'\right)$ for some α' such that Γ^* induces direct submechanisms. Consequently, we can write (40), in terms of a direct mechanism, which gives (36).

Finally, the same argument as in the proof of Proposition 2 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 (36). ‖

Proof of Theorem 6

Recall that the $N+1$ types, i.e., $\Theta = \{\theta_0, \theta_1, ..., \theta_N\}$ have respective probabilities $\{\pi_0, \pi_1, ..., \pi_N\}$. Suppose also that the weights given to these skill groups in the utility function (35) are such that the measure $\tilde{G}$ corresponds to $\left\{\pi'_0, \pi'_1, ..., \pi'_N\right\}$. Since there are constant types, again suppressing h^t-dependence to simplify notation, we can write the program for the best sustainable mechanism as:

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

subject to the constraints

$$\sum_{t=0}^\infty \beta^t \left[u\left(c_t\left(\theta_i\right)\right) - \chi\left(l_t\left(\theta_i\right) \mid \theta_i\right)\right] \geq \sum_{t=0}^\infty \beta^t \left[u\left(c_t\left(\theta_{i-1}\right)\right) - \chi\left(l_t\left(\theta_{i-1}\right) \mid \theta_i\right)\right] \tag{41}$$

$$\sum_{s=0}^\infty \beta^{t+s} \left[u\left(c_{t+s}\left(\theta_i\right)\right) - \chi\left(l_{t+s}\left(\theta_i\right) \mid \theta_i\right)\right] \geq \frac{1}{1-\beta}\left[u(0) - \chi\left(0 \mid \theta_i\right)\right] \text{ (for each } t\text{)} \tag{42}$$

for all $i = 0, ..., N$,

$$\sum_{s=0}^\infty \beta^{t+s} \left\{(1-a)\, v\left(x_{t+s}\right) + a\left(\sum_{i=0}^N \pi'_i \left[u\left(c_{t+s}\left(\theta_i\right)\right) - \chi\left(l_{t+s}\left(\theta_i\right) \mid \theta_i\right)\right]\right)\right\} \geq V\left(K_t, \{l_t(\theta_i)\}_{i=0}^N\right) \tag{43}$$

for all t, and

$$x_t + K_{t+1} + \sum_{i=0}^N \pi_i c_t\left(\theta_i\right) \leq F\left(K_t, \sum_{i=1}^N \pi_i l_t\left(\theta_i\right)\right) \tag{44}$$

for all t (also naturally $c_t(\theta_i) \geq 0$ for all i and t and $x_t \geq 0$ for all t).

The first set of constraints, (41), ensures incentive compatibility for the citizens. Given Theorem 5, there is truthful revelation along the equilibrium path. This, together with the single crossing property in Assumption 1, implies that we only need one constraint for each type, where type i could deviate to claim to be type $i-1$. The second set of constraints, (42), follows from the freedom of labour supply (one for each date), and the third set, (43), again one for each date, imposes sustainability, with the definition of $V\left(K_t, \{l_t(\theta_i)\}_{i=0}^N\right) \equiv \max_{\left\{\tilde{x}', \{\tilde{c}'(\theta_i)\}_{i=0}^N\right\}} (1-a)\, v\left(\tilde{x}'\right) + a\left(\sum_{i=0}^N \pi'_i \left[u\left(\tilde{c}'\left(\theta_i\right)\right) - \chi\left(l^*\left(\theta_i, K\right) \mid \theta_i\right)\right]\right)$ as in (37) in the text. Finally, the last set of constraints, (44) one for each date, imposes the aggregate resource constraint.

Let λ_i be the multiplier on the incentive-compatibility constraint of type i, let ψ_t be the multiplier on (43), and also define

$$\mu_t = \mu_{t-1} + \psi_t, \tag{45}$$

with $\mu_{-1} = 0$. Following Marcet and Marimon (1998), we can incorporate the incentive compatibility constraints (41) and the sustainability constraint (43) (until date $\min\{t, T-1\}$) into the objective function, and for any $T \geq 0$, represent this maximization problem with the Lagrangian:

$$
\begin{aligned}
\max_{\left\{\{c_t(\theta_i),l_t(\theta_i)\}_{i=0}^N,x_t,K_{t+1}\right\}_{t=0}^\infty} & \sum_{t=0}^{\infty}\beta^t\sum_{i=0}^{N}\pi_i\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right] \\
&+\sum_{i=1}^{N}\lambda_i\left\{\sum_{t=0}^{\infty}\beta^t\left\{\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]-\left[u\left(c_t\left(\theta_{i-1}\right)\right)-\chi\left(l_t\left(\theta_{i-1}\right)\mid\theta_i\right)\right]\right\}\right\} \\
&+\sum_{t=0}^{\infty}\beta^t\sum_{s=0}^{\min\{t,T-1\}}\psi_s\left\{(1-a)\,v(x_t)+a\sum_{i=1}^{N}\pi_i'\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]\right\} \\
&-\sum_{t=0}^{T-1}\beta^t\psi_t V\left(K_t,\{l_t\left(\theta_i\right)\}_{i=1}^N\right)
\end{aligned}
$$

subject to (42), (44) and also (43) for dates $t \geq T$.

Suppose $\left\{\left\{c_t^*\left(\theta_i\right),l_t^*\left(\theta_i\right)\right\}_{i=0}^N,x_t^*,K_{t+1}^*\right\}_{t=0}^\infty$ is a solution this program. Then taking the multipliers $\{\lambda_i\}_{i=0}^N$ as given, $\left\{\left\{c_t^*\left(\theta_i\right),l_t^*\left(\theta_i\right)\right\}_{i=0}^N,x_t^*,K_{t+1}^*\right\}_{t=T}^\infty$ must be a solution to

$$
\begin{aligned}
\max_{\left\{\{c_t(\theta_i),l_t(\theta_i)\}_{i=0}^N,x_t,K_{t+1}\right\}_{t=T}^\infty} & \sum_{t=T}^{\infty}\beta^t\sum_{i=0}^{N}\pi_i\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right] \\
&+\sum_{i=1}^{N}\lambda_i\left\{\sum_{t=T}^{\infty}\beta^t\left\{\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]-\left[u\left(c_t\left(\theta_{i-1}\right)\right)-\chi\left(l_t\left(\theta_{i-1}\right)\mid\theta_i\right)\right]\right\}\right\} \\
&+\sum_{t=T}^{\infty}\beta^t\mu_{T-1}\left\{(1-a)\,v(x_t)+a\sum_{i=1}^{N}\pi_i'\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]\right\}
\end{aligned}
$$

subject to, for all $t, s \geq T$, (42), (43) and (44). For any T, μ_{T-1} is finite. Thus dividing wall terms in the previous expression by $\beta^T\mu_{T-1}$, it can be equivalently written as

$$
\begin{aligned}
\max_{\left\{\{c_t(\theta_i),l_t(\theta_i)\}_{i=0}^N,x_t,K_{t+1}\right\}_{t=T}^\infty} & \frac{1}{\mu_{T-1}}\sum_{t=T}^{\infty}\beta^{t-T}\sum_{i=0}^{N}\pi_i\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right] \qquad (46)\\
&+\frac{1}{\mu_{T-1}}\sum_{i=1}^{N}\lambda_i\left\{\sum_{t=T}^{\infty}\beta^{t-T}\left\{\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]-\left[u\left(c_t\left(\theta_{i-1}\right)\right)-\chi\left(l_t\left(\theta_{i-1}\right)\mid\theta_i\right)\right]\right\}\right\} \\
&+\sum_{t=T}^{\infty}\beta^{t-T}\left\{(1-a)\,v(x_t)+a\sum_{i=1}^{N}\pi_i'\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]\right\}.
\end{aligned}
$$

Equation (45) implies that $\{\mu_T\}$ is a nondecreasing sequence ($\psi_t \geq 0$ for all t), and thus it either converges to some $\mu^* < \infty$ or diverges to infinity. Suppose first that $\{\mu_T\}$ converges to $\mu^* < \infty$. Then (45) implies that $\psi_T \to 0$ and therefore distortions disappears as claimed in the theorem. To complete the proof, we must show that $\mu_T \to \infty$ is not possible. Suppose, to obtain a contradiction, that $\mu_T \to \infty$. Then the maximization problem (46) converges to

$$
\max_{\left\{\{c_t(\theta_i),l_t(\theta_i)\}_{i=0}^N,x_t,K_{t+1}\right\}_{t=T}^\infty}\sum_{t=T}^{\infty}\beta^{t-1}\left\{(1-a)\,v(x_t)+a\sum_{i=1}^{N}\pi_i'\left[u\left(c_t\left(\theta_i\right)\right)-\chi\left(l_t\left(\theta_i\right)\mid\theta_i\right)\right]\right\}
$$

subject to, for all $t, s \geq T$, (42), (43) and (44). However, Assumption 5 implies that in this problem, the government sustainability constraint, (43), ceases to bind, and thus $\psi_T \to 0$, which from (45) contradicts $\mu_T \to \infty$. This contradiction implies that $\mu_T \to \mu^* < \infty$ and establishes the desired result.

Proof of Theorem 7

Proof. In this case, the best sustainable mechanism can be represented as the solution to the following maximization problem:

$$\mathbf{MAX}_2 : \mathbf{U}^{SM} = \max_{\left\{c_t(\cdot), l_t(\cdot), x_t, K_{t+1}\right\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t\left(\left(\theta^{i,t}\right)\right), l_t\left(\theta^{i,t}\right) \mid \theta_t^i\right)\right]$$

subject to the initial capital stock $K_0 > 0$, (11), (12) and the modified sustainability constraint,

$$\mathbb{E}_{\tilde{G}}\left[\sum_{s=0}^{\infty} \delta^s \left\{(1-a)\, v\left(x_{t+s}\right) + au\left(c_t\left(\left(\theta^{i,t}\right)\right), l_t\left(\theta^{i,t}\right) \mid \theta_t^i\right)\right\}\right] \geq V\left(K_t, \left\{l_t\left(\theta^{i,t}\right)\right\}\right), \tag{47}$$

which again takes into account that after deviation there will be zero labour supply and thus the highest continuation value of the government after deviation is $V\left(K_t, \left\{l_t\left(\theta^{i,t}\right)\right\}\right)$, defined analogously to (37) in the text. Here $\mathbb{E}_{\tilde{G}}$ denotes the integral evaluated according to $\tilde{G}$ as in (35). As before, denote the Lagrange multipliers on the sustainability contraint by $\delta^t \psi_t$, with $\mu_t = \mu_{t-1} + \psi_t$, and rewrite $\mathbf{MAX}_2$ recursively as

$$\mathbf{MAX}_2 : \mathbf{U}^{SM} = \max_{\left\{c_t(\cdot), l_t(\cdot), x_t, K_{t+1}\right\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t\left(\theta^{i,t}\right), l_t\left(\theta^{i,t}\right) \mid \theta_t^i\right) + \mu_t \delta^t au\left(c_t\left(\left(\theta^{i,t}\right)\right), l_t\left(\theta^{i,t}\right) \mid \theta_t^i\right)\right]$$
$$+ \sum_{t=0}^{\infty} \delta^t \left\{\mu_t (1-a)\, v(x_t) - (\mu_t - \mu_{t-1}) V\left(K_t, \left\{l_t\left(\theta^{i,t}\right)\right\}\right)\right\}$$

subject to (11) and (12). Denote the Lagrange multipliers in the solution to this problem by $\left\{\mu_t^*\right\}$. Then as in the text, consider the quasi-Mirrlees program corresponding to $\mathbf{MAX}_2$, which is given by

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

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

$$\int c_t\left(\theta^t\right) dG\left(\theta^t\right) \leq C_t,$$

and

$$\int l_t\left(\theta^t\right) dG\left(\theta^t\right) \geq L_t.$$

Here $\mathcal{U}\left(\{C_t, L_t\}_{t=0}^{\infty}; \left\{\mu_t^*\right\}_{t=0}^{\infty}\right)$ has exactly the same interpretation as $\mathcal{U}\left(\{C_t, L_t\}_{t=0}^{\infty}\right)$ in the text, except that it is also a function of the sequence of Lagrange multipliers. Then, with the same arguments as in the text, the maximization problem $\mathbf{MAX}_2$ becomes

$$\max_{\{C_t, L_t, x_t, K_t\}_{t=0}^{\infty}} \mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}; \left\{\mu_t^*\right\}_{t=0}^{\infty}) + \sum_{t=0}^{\infty} \delta^t \left\{\mu_t^* (1-a)\, v(x_t) - (\mu_t^* - \mu_{t-1}^*) V(K_t, \{l_t(\theta^{i,t})\})\right\}$$

subject to (11) and (47).

Proceeding as in the proof of Theorem 4, we obtain

$$\frac{\mathcal{U}_{C_t}(\{C_t, L_t\}_{t=0}^{\infty}; \left\{\mu_t^*\right\}_{t=0}^{\infty})}{\alpha \delta^t v'(x_t)} = \mu_t^* \leq \mu_{t+1}^* = \frac{\mathcal{U}_{C_{t+1}}(\{C_t, L_t\}_{t=0}^{\infty}; \left\{\mu_t^*\right\}_{t=0}^{\infty})}{\alpha \delta^{t+1} v'(x_{t+1})}.$$

Using the definition of φ, this can be rewritten as

$$\frac{\varphi^t \mathcal{U}_{C^*}^*}{(1-a)\, \delta^t v'(x^*)} = \mu_t^* \leq \mu_{t+1}^* = \frac{\varphi^{t+1} \mathcal{U}_{C^*}^*}{(1-a)\, \delta^{t+1} v'(x^*)} \text{ as } t \to \infty.$$

Then we can proceed as in Theorem 4 to yield the conclusions in the theorem. In particular, when $\varphi = \delta$, then we must have $\mu_t^* \to \overline{\mu}^*$ and distortions disappear. If, on the other hand, $\varphi > \delta$, $\mu_{t+1}^* > \mu_t^*$ and those $\psi_t > 0$ as $t \to \infty$ and asymptotic distortions remain. ‖

Acknowledgements. We thank two anonymous referees, the editor, Kjetil Storesletten, and audiences at many seminars for helpful comments and suggestions, and Georgy Egorov and Oleg Itskhoki for excellent research assistance. All three authors gratefully acknowledge financial support from the National Science Foundation.

REFERENCES

ABREU, D. (1988), "On the Theory of Repeated Games with Discounting", *Econometrica,* **56**, 383–396.

ACEMOGLU, D. (2007), "Modeling Inefficient Institutions", in Blundell, R., Newey, W. K. and Persson, T. (eds.) *Advances in Economic Theory, Proceedings of World Congress 2005* (Cambridge University Press).

ACEMOGLU, D., GOLOSOV, M. and TSYVINSKI, A. (2006), "Markets Versus Governments: Political Economy of Mechanisms" (NBER Working Paper 12224).

ACEMOGLU, D., GOLOSOV, M. and TSYVINSKI, A. (2008*a*), "Political Economy of Mechanisms", *Econometrica* **76**, 619–641.

ACEMOGLU, D., GOLOSOV, M. and TSYVINSKI, A. (2008*b*), "Markets Versus Governments", *Journal of Monetary Economics,* **55**, 159–189.

ACEMOGLU, D. and ROBINSON, J. (2006), *Economic Origins of Dictatorship and Democracy* (New York: Cambridge University Press).

ALBANESI, S. and ARMENTER, R. (2007), "Intertemporal Distortions in the Second Best" (NBER Working Paper 13629).

ALBANESI, S. and SLEET, C. (2005), "Dynamic Optimal Taxation with Private Information", *Review of Economic Studies,* **72**, 1–29.

AUSTEN-SMITH, D. and BANKS, J. S. (1999), *Positive Political Theory I: Collective Preference* (Ann Arbor: University of Michigan Press).

BARRO, R. (1973), "The Control of Politicians: An Economic Model", *Public Choice,* **14**, 19–42.

BATTAGLINI, M. and COATE, S. (2008), "A Dynamic Theory of Spending, Taxation, and Debt", *American Economic Review,* **98**, 201–236.

BECKER, G. (1983), "A Theory of Competition Amoung Pressure Groups for Political Influence", *Quarterly Journal of Economics,* **98**, 371–400.

BERTSEKAS, D., NEDIC, A. and OZDAGLAR, A. (2003), *Convex Analysis and Optimization* (Boston: Athena Scientific).

BESTER, H. and STRAUSZ, R. (2001), "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case", *Econometrica,* **69**, 1077–1098.

BISIN, A. and RAMPINI, A. (2006), "Markets as Beneficial Constraints on the Government", *Journal of Public Economics,* **90**, 601–629.

BUCHANAN, J. M. and TULLOCK, G. (1962), *The Calculus of Consent* (Ann Arbor: University of Michigan Press).

CHAMLEY, C. (1986), "Optimal Taxation of Capital Income in General Equilibrium and Infinite Lives", *Econometrica,* **54**, 607–622.

CHARI, V. V. and KEHOE, P. (1990), "Sustainable Plans", *Journal of Political Economy,* **94**, 783–802.

CHARI, V. V. and KEHOE, P. (1990), "Sustainable Plans and Mutual Default", *Review of Economic Studies,* **60**, 175–195.

DIXIT, A. K. (2004), *Lawlessness and Economics: Alternative Modes of Governance* (Princeton: Princeton University Press).

FARHI, E. and WERNING, I. (2008), "The Political Economy of Nonlinear Capital Taxation" (Mimeo, Harvard).

FEREJOHN, J. (1986), "Incumbent Performance and Electoral Control", *Public Choice,* **50**, 5–25.

FREIXAS, X., GUESNERIE, R. and TIROLE, J. (1985), "Planning under Incomplete Information and the Ratchet Effect", *Review of Economic Studies,* **52**, 173–192.

GOLOSOV, M., KOCHERLAKOTA, N. and TSYVINSKI, A. (2003), "Optimal Indirect and Capital Taxation", *Review of Economic Studies,* **70**, 569–587.

GOLOSOV, M., TSYVINSKI, A. and WERNING, I. (2006), "New Dynamic Public Finance: A User's Guide", in Acemoglu, D., Rogoff, K. and Woodford, M. (eds.) *NBER Macroeconomics Annual 2006* (Cambridge, MA: MIT Press).

GROSSMAN, G. and HELPMAN, E. (1994), "Protection for Sale", *American Economics Review,* **84**, 833–850.

HARRIS, M. and HOLMSTROM, B. (1982), "A Theory of Wage Dynamics", *Review of Economic Studies,* **49**, 315–333.

HASSLER, J., KRUSELL, P. and ZILIBOTTI, F. (2005), "The Dynamics of Government: A Positive Analysis", *Journal of Monetary Economics,* **52**, 1331–1358.

JUDD, K. (1985), "Redistributive Taxation in a Simple Perfect Foresight Model", *Journal of Public Economics,* **28**, 59–83.

KOCHERLAKOTA, N. (2005), "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation", *Econometrica,* **73**, 1587–1621.

KRUSELL, P. and RIOS-RULL, J.-V. (1999), "On the Size of Government: Political Economy in the Neoclassical Growth Model", *American Economic Review,* **89**, 1156–1181.

LINDBACK, A. and WEIBULL, J. (1987), "Balanced-budget Redistribution as the Outcome of Political Competition", *Public Choice,* **52**, 273–297.

MARCET, A. and MARIMON, R. (1998), "Recursive Contracts" (Mimeo, University of Pompeu Fabra).

MAS-COLELL, A., WHINSTON, M. D. and GREEN, J. (1995), *Microeconomic Theory* (New York, Oxford: Oxford University Press).

MIRRLEES, J. A. (1971), "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic Studies,* **38**, 175–208.

NORTH, D. C. (1981), *Structure and Change in Economic History* (New York: W.W. Norton & Co.).

NORTH, D. C. and THOMAS, R. P. (1973), *The Rise of the Western World: A New Economic History* (Cambridge, UK: Cambridge University Press).

NORTH, D. C. and WEINGAST, B. R. (1989), “Constitutions and Commitment: Evolution of Institutions Governing Public Choice in Seventeenth Century England”, *Journal of Economic History,* **49**, 803–832.

OLSON, M. (1982), *The Rise and Decline of Nations: Economic Growth, Stagflation, and Economic Rigidities* (New Haven and London: Yale University Press).

PERSSON, T. and TABELLINI, G. (2000), *Political Economics: Explaining Economic Policy* (Cambridge, MA: MIT Press).

RAY, D. (2002), “Time Structure of Self-Enforcing Agreements”, *Econometrica,* **70**, 547–582.

ROBERTS, K. (1984), “Theoretical Limits to Redistribution”, *Review of Economic Studies,* **51**, 177–195.

SKRETA, V. (2006), “Optimal Auction Design under Non-Commitment” (Mimeo).

SKRETA, V. (2007), “Sequentially Optimal Mechanisms”, *Review of Economic Studies,* **73**, 1085–1111.

SLEET, C. and YELTEKIN, S. (2004), “Credibility and Endogenous Social Discounting,” *Review of Economic Dynamics,* **9**, 410–437.

UHLIG, H. (1996), “A Law of Large Numbers for Large Economies”, *Economic Theory,* **8**, 41–50.