Markets Versus Governments: Political Economy of Mechanisms*

Daron Acemoglu
MIT

Michael Golosov
MIT

Aleh Tsyvinski
Harvard

Still Preliminary and Incomplete.

Abstract

We study the optimal Mirrlees taxation problem in a dynamic economy with idiosyncratic (productivity or preference) shocks. In contrast to the standard approach, which implicitly assumes that the mechanism is operated by a benevolent planner with the full commitment power, we assume that any centralized mechanism can only be operated by a self-interested ruler/government without commitment power, who can therefore misuse the resources and the information it collects. An important result of our analysis is that there will be truthful revelation along the equilibrium path, which shows that truth-telling mechanisms can be used despite the commitment problems and the different interests of the government and the citizens. Using this tool, we show that if the government is as patient as the agents, the best sustainable mechanism leads to an asymptotic allocation where the highest types face zero marginal tax rate on their labor supply as in the classical Mirrlees setup and there are no aggregate capital taxes—i.e., marginal distortions arising from political economy disappear asymptotically. In contrast, when the government is less patient than the citizens, aggregate distortions remain, and there is both positive marginal labor tax on the highest types and positive aggregate capital taxes even asymptotically. Under some additional assumptions on preferences, these results generalize to the case when the government is benevolent but unable to commit to future tax policies. We conclude by providing a brief comparison of centralized mechanisms operated by self-interested rulers to anonymous markets.

Keywords: dynamic incentive problems, mechanism design, optimal taxation, political economy.
JEL Classification: H11, H21, E61, P16.

*We thank Manuel Amador, Marios Angeletos, Abhijit Banerjee, Tamer Basar, Olivier Blanchard, Ricardo Caballero, V. V. Chari, Caroline Hoxby, Larry Jones, Patrick Kehoe, Narayana Kocherlakota, Jonathan Levin, Guido Lorenzoni, Asuman Ozdaglar, Christopher Phelan, Vasiliki Skreta and seminar participants at Boston University, Carnegie Mellon, Federal Reserve Bank of Minneapolis, Harvard, MIT and Princeton for useful comments and to Oleg Itskhoki for excellent research assistance.
1 Introduction

The first-generation approach to public finance, best exemplified by models of Ramsey taxation, sought to determine the optimal policy of a benevolent government in a world with a given set of fiscal or regulatory instruments. The second-generation approach, pioneered by Mirrlees, has made major progress over this approach by explicitly modeling the choice of tax instruments that the government can use. This literature has modeled the informational problems restricting the potential tax-transfer programs, and formulated the determination of optimal tax and transfer programs as one of mechanism design.

Some of the theoretical limitations of the second-generation approach have long been apparent, however. First, in intertemporal settings, this approach assumes that the mechanism designer (government or planner) can commit to a dynamic mechanism, even though such commitment is ex post costly. Second, it is assumed that there is a body (a “benevolent government”) that can operate the optimal tax-transfer program (mechanism), even though the lessons of the political economy literature are that governments or politicians do not simply maximize welfare, but have their own selfish objectives, including reelection or personal enrichment. This paper aims to contribute to a potential third-generation approach to public finance where both the informational constraints on tax instruments and the incentive problems associated with governments, politicians and bureaucrats are taken into account. For this purpose, we investigate how mechanisms work and should be designed in the presence of self-interested and time-inconsistent governments.

1 See Mirrlees (1971) for the seminal reference and Baron and Myerson (1982), Dasgupta, Hammond and Maskin (1979), Green and Laffont (1977), Harris and Townsend (1981), Myerson (1979), and Holmstrom and Myerson (1983) for some of the important papers in the early literature. Albanesi and Sleet (2005), Golosov and Tsyvinski (2004), Golosov, Kocherlakota and Tsyvinski (2003), Kocherlakota (2005), Werning (2002), and Battaglini and Coate (2005) consider applications of the Mirrlees framework to dynamic taxation.

2 Given this formulation of the optimal tax-transfer program as a mechanism, we will use the terms “optimal tax-transfer program” and “mechanism” interchangeably.


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

4 One can think of 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 of the communication between the mechanism designer and individual agents. Although this may be a useful modeling tool, it does not circumvent the substantive issues raised here: the party entrusted with operating
Two questions motivate this analysis. First, we would like to understand whether optimal tax-transfer programs in the presence of these additional constraints, which we refer to as sustainable mechanisms, look similar to “Mirrlees” mechanisms, which assume a benevolent planner and full commitment. If they do, then we can have more confidence in the mechanism design approach as a tool to analyze the practice of policy design as well as a normative benchmark. Second, in the presence of self-interested and time-inconsistent government behavior, anonymous market allocations cannot always be replicated by centralized (sustainable) mechanisms. This opens the door to a theoretical analysis of the relative costs and benefits of centralized resource allocation methods versus those that preserve anonymity and limit government intervention, and we would like to take a first step in this direction.

To highlight the problems that arise when we depart from the benchmark of a benevolent planner with full commitment, it is useful to start with Roberts’ (1984) seminal contribution, which highlighted the importance of the commitment assumption. Roberts considered an optimal taxation environment similar to the static one analyzed by Mirrlees (1971), where risk-averse individuals are subject to unobserved shocks affecting the marginal disutility of labor supply. But differently from the benchmark Mirrlees model, he assumed that the economy was 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 subgame perfect equilibrium of this game involves the highly inefficient outcome in which all types declare to be the worst type at all dates, supply the lowest level of labor and receive the lowest level of consumption. This example shows the potential inefficiencies that can arise once we depart from the unrealistic case of full commitment, even with benevolent governments.

Our benchmark economy incorporates the lack of commitment present in Roberts’ (1984) optimal policy has neither the same interests as those of the citizens nor much commitment power. Naturally, the best mechanism we characterize can be represented as a solution to this fictional mechanism design problem. In line with this objective, throughout the paper we look for the allocation that maximizes the ex ante utility of the citizens (agents) subject to the political economy and commitment constraints introduced by the self-interested nature of the government.
paper, but also assumes that the government is self-interested and maximizes its own utility. This latter assumption brings our model closer to the political economy literature, but is particularly useful for our purposes because it simplifies the structure of the dynamic game relative to the case with a benevolent time-inconsistent planner. At the end of the paper, we generalize our results to a situation where the government has an arbitrary degree of benevolence (thus nesting the fully-benevolent time-inconsistent government).

Our main departure from Roberts’ (1984) framework is that instead of a finite-horizon economy, we study an infinite-horizon economy. This enables us to use punishment strategies against the government to construct a sustainable mechanism, defined as an equilibrium tax-transfer program in the dynamic game that is both incentive compatible for the citizens and for the government (i.e., it satisfies a sustainability constraint for the government). The (best) sustainable mechanism gives a fraction of the output to the government in every period, and if the government deviates from this implicit agreement, citizens switch to supplying zero labor, implicitly punishing the government. The infinite-horizon setup enables us to prove our first important result, that a version of the revelation principle, the truthful revelation along the equilibrium path, applies and is a useful tool of analysis for this class of dynamic incentive problems with self-interested mechanism designers and without commitment. The fact that truthful revelation principle applies only along the equilibrium path is important, since the potential actions that can be taken off the equilibrium path place restrictions on what type of mechanisms are allowed (these are encapsulated in the sustainability constraints).

The truthful revelation along the equilibrium path enables us to write the problem of finding the best sustainable mechanism as an infinite-dimensional maximization problem. We characterize the solution to this program by defining a quasi-Mirrlees problem, where expected utility of a representative agent is maximized subject to the standard incentive compatibility constraints and two additional resource constraints at every date; the first requires that the sum of total labor supply in the economy be no less than some amount $L_t$ and the second that the sum of total consumption be no greater than some amount $C_t$. When the mechanism also

\footnote{Clearly, the punishment strategies that sustain this equilibrium are not “renegotiation proof”. This is common with many other analyses of repeated games, so we do not view this as a special shortcoming of our approach. Moreover, in practice, citizens have many other recourses against governments that misbehave, including voting or throwing them out of office, and these recourses will have the same impact as supplying zero labor in the punishment phase. We do not incorporate these possibilities to simplify the analysis in this paper. In Acemoglu, Golosov and Tsyvinski (2006), we study the equilibrium of an economy without any informational restrictions on taxes and transfers, where the citizens have the option of replacing the current government.}
optimizes over the sequences of $C_t$ and $L_t$ subject to the aggregate resource constraints, the quasi-Mirrlees problem is identical to the full-commitment dynamic Mirrlees problem. We show that the best sustainable mechanism is a solution to a quasi-Mirrlees problem, and distortions resulting from the opportunistic behavior of the government only affect the parameters of this quasi-Mirrlees problem (in particular, $L_t$ and $C_t$). This formulation therefore gives us a clean way of characterizing the differences between the best sustainable mechanism and the full-commitment Mirrlees mechanism in terms of the aggregate distortions caused by the former.

The rest of our results concern a comparison of the best sustainable mechanism to the full-commitment Mirrlees mechanism. First, we show that at the initial date, there will always be further distortions in the sustainable mechanism relative to the full-commitment Mirrlees mechanism. Second, we provide tight conditions under which these further distortions disappear or persist over time. In particular, we prove that when the government is as patient as, or more patient than, the citizens, the sustainability constraint of the government eventually becomes slack. In the absence of a binding sustainability constraint, aggregate distortions disappear, and the marginal products of labor and capital asymptotically converge to the marginal rates of substitution for the agents; consequently, as in the static Mirrlees economy, marginal distortions on the labor supply of the highest types disappear, and the limiting allocation can be implemented with zero capital taxes.\footnote{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 also important to note that this limiting allocation can be decentralized in different ways, and some of those may involve positive taxes on individual capital holdings.} We also prove that when the government is less patient than the citizens, the results are very different; aggregate distortions never disappear, and even in the long run, there are positive aggregate capital taxes and a positive marginal labor income tax on the highest types. 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.

These results are derived under very general assumptions on the utility functions of citizens. We also show that when individuals have instantaneous utility functions that are separable between consumption and leisure, the same results apply for any utility function of the government, in particular, in the case where the government is fully-benevolent but time-inconsistent.

Finally, since with a sustainable mechanism part of the output has to be given to the
government, the anonymous market allocation cannot be achieved by a centralized mecha-
nism. This raises the question of when centralized sustainable mechanisms are preferable to
anonymous markets where there is limited government intervention in the economy and limited
information about individual attributes and transactions. We conclude the paper by a brief
discussion of this issue, and show that better institutional controls on government behavior
and a higher discount factor of the government makes centralized mechanisms more likely to
be ex ante preferable to anonymous markets.

This paper is related to a number of different strands of research. These include both the
original and the more recent applications of the mechanism design approach to the optimal
taxation problem already mentioned in footnote 1. The major difference between our work and
these papers is that they assume a benevolent government and full commitment. Secondly, our
paper is related to the recently burgeoning political economy literature mentioned in footnote
3. What distinguishes our paper from this literature is the explicit modeling of the incentive
problems on the side of the individuals, which underpins the power to the government.8 Finally,
our analysis is also related to work on optimal taxation with time-inconsistency, for example,

As well as Roberts (1984), most closely related to our research is the recent important paper
by Bisin and Rampini (2005), who also consider the problem of mechanism design without
commitment in a two-period setting.9 Bisin and Rampini extend Roberts’s analysis and show
how the presence of anonymous markets acts as an additional constraint on the government,
ameliorating the commitment problem. This lack of commitment is related to the lack of
commitment by the self-interested government in our model. The most important distinction
between the two approaches is that our model is infinite horizon. This allows us to construct
sustainable mechanisms with government commitment and the revelation principle along the
equilibrium path. The use of the revelation principle, which is not possible in Roberts’ or Bisin
and Rampini’s models, enables us to analyze substantially more general environments. Second,
the infinite-horizon setting enables an investigation of the limiting behavior of distortions and
taxes.

8 In this context, Hart, Shleifer and Vishny (1997), Chari (2000) and Acemoglu, Kremer and Mian (2003)
also contrast the incentive costs of governments and markets, but do not derive the costs of governments from
the centralization of power and information in the process of operating a mechanism.
9 See also the work by Freixas, Guesnerie and Tirole (1985) on the ratchet effect and recent work on general
mechanisms without commitment, for example, Bester and Strausz (2001), Skreta (2004), and Miller (2005).
The rest of the paper is organized as follows. Section 2 describes the basic environment. Section 3 starts the analysis of sustainable mechanisms. In this section, we set up the problem of constructing sustainable mechanisms and prove a version of the revelation principle. Section 4 formulates the problem of characterizing the best sustainable mechanism as a solution to a quasi-Mirrlees program. Section 5 characterizes the best sustainable mechanism using a simple but restrictive case to illustrate the main ideas. Section 6 characterizes the best sustainable mechanism without any restrictions on the set of mechanisms and preferences. Section 7 extends our analysis to cover the case of fully-benevolent, but time-inconsistent governments. Section 8 briefly compares allocations with anonymous markets to those under sustainable mechanisms. Section 9 concludes, while the Appendices contain some technical material necessary for the analysis as well the proofs not provided in the text.

2 Demographics, Preferences and Technology

The model economy is infinite horizon in discrete time. It is populated by a continuum of citizens with measure normalized to 1 and a ruler. The ruler/government can be thought of as a single agent or as a group of agents such as a bureaucracy, whose preferences can be consistently represented by a standard von Neuman-Morgenstern utility function.

Let \( \Theta = \{\theta_0, \theta_1, ..., \theta_N\} \) be a finite ordered set of potential types, with the convention that \( \theta_i \) corresponds to “higher skills” than \( \theta_{i-1} \), and in particular, \( \theta_0 \) is the worst type. Let \( \Theta^T \) be the \( T \)-fold product of \( \Theta \), representing the set of sequences of length \( T = 1, 2, ..., \infty \), with each element belonging to \( \Theta \). We think of each agent’s type is drawn from \( \Theta^\infty \) according to some measure \( \mu^\infty \). This imposes no restriction on the time-series properties of individual skills. Both iid draws from \( \Theta \) in every period and constant types, as well as arbitrary temporal dependence are allowed. We only assume that each individual’s draw from \( \Theta^\infty \) is according to the same measure \( \mu^\infty \) and is independent from the draws of all other individuals, so that there is no aggregate uncertainty. In addition, to simplify the notation, we also assume (without loss of generality) that within each period, there is an aggregate invariant distribution of types.

---

10 The continuum assumption implies that whenever we think of deviations, these should be by an individual with positive measure \( \varepsilon \), and then we should take the limit as \( \varepsilon \to 0 \). Moreover, equilibrium statements should be read as “almost-everywhere”. These technical details do not matter except in the proof of Theorem 5.

11 Finiteness of \( \Theta \) is adopted for simplicity and without loss of any economic insight. The more general case where \( \Theta \) is a compact interval of \( \mathbb{R}_+ \) introduces a number of additional technical details, not central for our analysis.
denoted by $G$.

Let $\theta^{i,\infty}$ be the draw of individual $i$ from $\Theta^\infty$. The $t$-th element of $\theta^{i,\infty}$, $\theta^i_t$, is the skill level of this individual at time $t$. We use the standard notation $\theta^{i,t}$ to denote the history of this individual’s skill levels up to and including time $t$, and make the standard measurability assumption that the individual only knows $\theta^{i,t}$ at time $t$. Since this will be a private information economy, no other agent in the economy will directly observe this history.\(^{12}\)

The instantaneous utility function of individual $i$ at time $t$ is given by

$$u(c^i_t, l^i_t \mid \theta^i_t)$$

where $c^i_t$ is the consumption of this individual and $l^i_t$ is her labor supply. We assume that labor supply of an individual with skill $\theta$ comes from a compact set, i.e., $l^i_t \in [0, \bar{l}(\theta)]$. In addition, we make a number of standard assumptions on $u$.

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

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

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

The first two assumptions are standard. Assumption 3 states that for the worst type, $\theta_0$, supplying positive labor is impossible. This suggests that we can think of the worst type as “disabled,” meaning unable to supply any labor. It also requires $\mu^\infty$ to have full support in the sense that any individual can become disabled at any point. This assumption will simplify the analysis of sustainable mechanisms by making it possible to have off-the-equilibrium path actions where all types supply zero labor. As described in Remark 1, this assumption can

\(^{12}\)Technically, this means that there exists a set of nested information sets (sub-sigma fields) representing each individual’s information sets, so that the individual only knows the information contained in $\mathcal{F}^i_t$ at time $t$. In particular, let the triple $(\Theta^\infty, \mathcal{F}, \mu^\infty)$ be a probability space and $\{\mathcal{F}^i_t : t \in \mathbb{N}\}$ be a filtration, i.e., a collection of sub-sigma fields of $\mathcal{F}$, such that $\mathcal{F}^i_t \subseteq \mathcal{F}^i_{t'}$ for all $t' > t$. Let $\Theta^t$ be the set $\Theta^\infty$ truncated at $t$. Then $\theta^{i,t} \in \Theta^t$ and all decisions taken at time $t$ by individual $i$ must be $\mathcal{F}^i_t$-measurable. See, for example, Pollard (2002).
be replaced by an alternative one, described below, which imposes “freedom of labor supply”
directly. The advantage of Assumption 3 is that it leads to the freedom of labor supply as an
equilibrium outcome.

Each individual maximizes the discounted sum of their utility with discount factor \( \beta \), so
their objective function at time \( t \) is

\[
\mathbb{E} \left[ \sum_{s=0}^{\infty} \beta^s u \left( c_{t+s}^i, \ell_{t+s}^i \mid \theta_{t+s}^i \right) \mid \mathcal{F}_t^i \right] = \mathbb{E} \left[ \sum_{s=0}^{\infty} \beta^s u \left( c_{t+s}^i, \ell_{t+s}^i \mid \theta_{t+s}^i \right) \mid \theta_{t}^{i,t} \right]
\]

where \( \mathbb{E} [\cdot | \mathcal{F}_t^i] \) or \( \mathbb{E} [\cdot | \theta_{t}^{i,t}] \) denote the expectations operator conditional on having observed
the history \( \theta_{t}^{i,t} \) (in addition to any public information).

The production side of the economy is described by a continuously differentiable constant
returns to scale aggregate production function

\[
F(K, L)
\]

where \( K \) is capital and \( L \) is labor. Each agent in the economy has access to this production
function. We assume that capital fully depreciates after use, and that \( F(K, 0) = 0 \), so that
without labor there is no production. Both of these assumptions are adopted to simplify the
notation.

In addition, the ruler has an instantaneous utility function \( v(x) \) where \( x \) denotes govern-
ment consumption and \( v : \mathbb{R}_+ \rightarrow \mathbb{R} \) is twice continuously differentiable, strictly increasing
(with positive derivative) everywhere and concave, and satisfies \( v(0) = 0 \). The ruler’s discount
factor, \( \delta \), is potentially different from that of the citizens, \( \beta \), so its objective function at time
\( t \) is

\[
\mathbb{E}_t \left[ \sum_{s=0}^{\infty} \delta^s v (x_{t+s}) \right]
\]

where \( \mathbb{E}_t \) refers to the expectations operated conditional on public information at time \( t \).

3 Sustainable Mechanisms

As in the analysis of Chari and Kehoe (1990, 1993), the interaction between the citizens and the
government is a game. The government both lacks commitment power and also has the ability
to expropriate the output of the economy and use it for its own consumption (or for some other
activities useful for itself but not for the citizens). Our purpose throughout is to characterize
the equilibrium of this game between the government and the citizens, corresponding to the best sustainable mechanism, meaning the sustainable mechanism that maximizes the ex ante utility of citizens.  

3.1 The Game Form Between Government and Citizens

There is a number of alternative ways of specifying the game form between the government and the citizens, with identical results. Our choice here is motivated to maximize the similarity with Roberts’ (1984) model.

We define a submechanism (or t-mechanism) as a subcomponent of the overall mechanism between the government and the individuals. Recall that Θ is the type space, and Θ^t is the t-fold product of the type space, with a typical element θ^t denoting the history of types up to and including time t. We denote a generic element of Θ^t by θ^t (which corresponds to the true “type”) and use ̂θ^t to stand for reports regarding types in the direct mechanisms. 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.  

Given Assumption 3, any submechanism must allow for some messages which will lead to \( l = 0 \).

13 Since we are dealing with a dynamic game, our focus on the best sustainable mechanism is essentially a selection among the many equilibria. Alternatively, one can think of the “social plan” as being designed by the citizens to maximize their utility subject to the constraints placed by the self-interested behavior of the government (see, in particular, the last paragraph of the Concluding Remarks, and also Acemoglu, 2005). In addition, throughout the paper we ignore the issue of renegotiation, both to simplify the analysis and also because, clearly, allowing for renegotiation would put more constraints on the “best sustainable mechanism”.

14 More formally, \( \theta^t, ̂\theta^t \) and \( z_t \) have to be \( F_t \)-measurable.

15 The mechanisms we describe here allow for general message spaces, but impose two restrictions. First, they are non-stochastic. Allowing for randomization may be important to ensure convexity of the constraint set. The Appendix discusses a more general formulation with lotteries, and the definition generalizes in a natural way for this case. 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 and no aggregate uncertainty, this latter restriction is without loss of generality here.
We denote the set of submechanisms that satisfy this restriction and also the relevant resource constraints (which will be specified below) by $\mathcal{M}_t$.\(^\text{16}\)

The typical assumption in models with no commitment is that the mechanism designer can commit to a submechanism within a given date, but cannot commit to what mechanisms will be offered in the future. In our context, there is an additional type of deviation for the government whereby it can use its power to extract resources from the society even within the same period. The interaction between the government and the individuals is modeled with the following game form at each date:

1. At the beginning of period $t$, the government offers a submechanism $\tilde{M}_t \in \mathcal{M}_t$.

2. Individuals send a message $z_t \in Z_t$, which together with $z^{t-1} \in Z^{t-1}$, determines their labor supply according to the submechanism $\tilde{M}_t$.

3. Production takes place according to the labor supplies of the individuals.

4. The government 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 government chooses $\tilde{x}_t$, and next period’s capital stock is determined as $\tilde{K}_{t+1} = F(K_t, L_t) - \tilde{x}_t - \int_0^1 c'_t(z'_t) \, dt$. If $\xi_t = 1$, the government chooses $\tilde{x}_t$ and a new consumption function $\tilde{c}_0 : Z^t \to \mathbb{R}_+$, and next period’s capital stock is: $\tilde{K}_{t+1} = F(K_t, L_t) - \tilde{x}_t - \int_0^1 c'_t(z'_t) \, dt$.\(^\text{17}\)

This game form emphasizes that the only difference between the standard models with no commitment and our setup is that the government, in the last stage, can also decide to expropriate the output produced in the economy. Notice that at this stage labor 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 government is expropriating some

\(^{16}\)Alternatively, we could define a mechanism as a mapping $M_t[K_t]$ conditional on the capital stock of the economy at that date to emphasize that what can be achieved will be a function of the capital stock. We suppress this dependence to simplify notation.

\(^{17}\)More generally, we can allow the government to capture a fraction $\eta \leq 1$ of the total output of the economy when $\xi = 1$, where the level of $\eta$ could be related to the institutional controls on government or politician behavior. In this case, the constraint on the government following a deviation would be $\tilde{K}_{t+1} = \eta F(K_t, L_t) - \tilde{x}_t - \int_0^1 c'_t(z'_t) \, dt$, with the remaining $1 - \eta$ fraction of the output on getting destroyed. This generalization has no effect on our results and for now, we set $\eta = 1$ to simplify notation. We return to issues of institutional limits on government action below.
of the output for itself. Consequently, we also let the government choose a new consumption allocation function, \( \hat{c}_t^\alpha : Z^t \rightarrow \mathbb{R}_+ \) at this point.

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

We represent the action of the government at time \( t \) by \( \rho_t = (\bar{M}_t, \xi_t, \hat{x}_t, \hat{c}_t^\alpha) \in \mathbb{R}_t \equiv \mathcal{M}_t \times \{0, 1\} \times \mathbb{R}_+^2 \times \mathcal{C}_t \). The first element is what the submechanism that the government offers at stage 1 of time \( t \), and the second is the government’s expropriation decision. The third element of \( \rho_t \) is what the government consumes itself if \( \xi_t = 0 \). Since \( \bar{M}_t \) specifies both total production and total consumption by the citizens, given \( \hat{x}_t \) the capital stock for next period, \( \bar{K}_{t+1} \), is determined as a residual from the resource constraint and is not specified as part of the action profile of the government. The fourth element, \( \hat{x}_t \), is the government consumption level when \( \xi_t = 1 \). Finally, the fifth element is the function \( \hat{c}_t^\alpha \) that the government 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, \( \bar{K}_{t+1} \), is determined as a residual from the resource constraint. Government consumption levels must satisfy: \( \hat{x}_t \leq F(K_t, L_t) \) and \( \hat{x}_t \leq F(K_t, L_t) \), but to simplify notation we write \( \hat{x}_t \in \mathbb{R}_+ \). We use \( \rho^t \in \mathbb{R}^t \) to denote the history of \( \rho_t \)’s up to and including time \( t \).

Turning to the citizens, define \( \alpha_t^i \left( \theta^t | z^{t-1}, \rho^{t-1} \right) \) as the action of individual \( i \) at time \( t \) when her type history is \( \theta^t \) and her history of messages so far is \( z^{t-1} \) and the publicly observed history of government behavior up at the time \( t - 1 \) is \( \rho^{t-1} \). \( \alpha_t^i \) specifies a message \( z_t \in Z_t \):

\[
\alpha_t^i : Z^{t-1} \times \mathbb{R}^{t-1} \times \Theta^t \rightarrow Z_t.
\]

We sometimes abbreviate notation by writing \( z^t \left( \alpha_t \left( \theta^t \right) \right) \) to denote the message resulting from strategy \( \alpha_t \) with type \( \theta^t \). We call a strategy truth telling if it satisfies

\[
\alpha^\ast \left( \theta^t | z^{t-1}, \rho^{t-1} \right) = z_t \left[ \theta^t \right] \text{ for all } \theta^t \in \Theta^t, z^{t-1} \in Z^{t-1} \text{ and } \rho^{t-1} \in \mathbb{R}^{t-1}, \tag{3}
\]

where the notation \( z_t \left[ \theta^t \right] \) means that the individual is sending a message that reveals her true type. To economize on notation, we represent the truth-telling strategy by \( \alpha_t^i \left( \theta_t | z^{t-1} \left[ \theta^{t-1} \right], \rho^{t-1} \right) = \alpha^\ast \). Notice that this strategy only imposes truth-telling following truthful reports in the past.
(since instead of an arbitrary history of messages $z^{t-1}$, we have conditioned on $z^{t-1}[\theta^{t-1}]$). This is without loss of any generality. In addition, let us define the null strategy
\[
\alpha^0(\theta_t | \hat{\theta}^{t-1}, \rho^{t-1}) = z_t[\theta_0] \text{ for all } \theta' \in \Theta^t, z^{t-1} \in Z^{t-1} \text{ and } \rho^{t-1} \in \mathcal{R}^{t-1},
\]
where $z_t[\theta_0]$ stands for a message signifying that the individual is disabled. Such a message must always be allowed in any submechanism that is an element of $\mathcal{M}_t$ because of Assumption 3. Therefore, the individual can always choose to supply zero labor, or in other words, any feasible mechanism (submechanism) must allow for “freedom of labor supply”. We will use the notation $\alpha^0_i(\theta_t | \hat{\theta}^{t-1}, \rho^{t-1}) = \alpha^0$ to denote that the individual is playing the null strategy.

Finally, we denote the strategy profile of all the individuals in society by $\alpha$, with $\mathcal{A}$ denoting the set of all such strategy profiles.

Next, turning to the government’s strategies, note that at each date, the government chooses $\rho_t \in \mathcal{R}_t$. In addition, let $z_t \in Z_t$ be a profile of reports at time $t$. As usual, we define $Z^t = \prod_{s=0}^{t} Z_s$. The government’s strategy at time $t$ is therefore
\[
\Gamma_t: \mathcal{R}^{t-1} \times Z^{t-1} \rightarrow \mathcal{R},
\]
i.e., it determines $\tilde{M}_t \in \mathcal{M}_t$, $\xi_t \in \{0, 1\}$, $\tilde{x}_t \in [0, F(K_t, L_t)]$, $\hat{x}_t \in [0, F(K_t, L_t)]$ and $\tilde{c}_t \in \mathcal{C}_t$ as a function of the government’s own past actions and the entire history of reports by citizens. We denote the entire strategy profile of the government by $\Gamma$ and the set of strategy profiles by $\mathcal{G}$.

**Definition 1** A (sequential) equilibrium in the game between the government and the citizens is given by strategy profiles $\hat{\Gamma}$ and $\hat{\alpha}$ that are best responses to each other in all information sets given beliefs, and beliefs are derived from Bayesian updating given the strategy profiles. We write the requirement that these strategy profiles are best responses to each other as $\hat{\Gamma} \succeq_{\hat{\alpha}} \Gamma$ for all $\Gamma \in \mathcal{G}$ and $\hat{\alpha} \succeq_{\hat{\Gamma}} \alpha$ for all $\alpha \in \mathcal{A}$.

Let us define $\Gamma_{M,x} = \left\{ (\tilde{M}_t, \xi_t, \tilde{x}_t, \hat{x}_t, \tilde{c}_t) \right\}_{t=0}^{\infty}$ as the action profile of the government induced by strategy $\Gamma$ given a social plan $(M, x)$. Conditioning on the social plan here is simply for emphasis.

---

18 More formally, $z_t$ assigns a report to each individual, thus it is a function of the form $z_t : [0, 1] \rightarrow Z_t$, and $Z_t$ is the set of all such functions.
Definition 2  \( M \) is a sustainable mechanism if there exists \( x = \{x_t\}_{t=0}^{\infty} \), a strategy profile \( \alpha \) for the citizens and a strategy profile \( \Gamma \in \mathcal{G} \) for the government, which constitute a sequential equilibrium and induce an action profile \( \Gamma_{M,x} = \left\{ \tilde{M}_t, \xi_t, \tilde{x}_t, \tilde{\tau}_t \right\}_{t=0}^{\infty} \) for the government such that \( \tilde{M}_t = M_t, \xi_t = 0, \) and \( \tilde{x}_t = x_t, \) and satisfies \( \hat{\Gamma} \succeq_{\alpha} \Gamma \).

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

3.2 Truthful Revelation Along the Equilibrium Path

The revelation principle is a powerful tool for the analysis of mechanism design and implementation problems (see, e.g., MasCollel, Winston and Green, 1995). Since, in this environment, the government, who operates the mechanism, cannot commit and has different interests than those of the agents, the simplest version of the revelation principle does not hold; there will exist situations in which individuals will prefer not to report their true type (e.g., Roberts, 1984, Freixas, Guesnerie and Tirole, 1985, or Bisin and Rampini, 2005). The key result of this section will be that along the equilibrium path, a version of the revelation principle will still hold.

Let us first consider the problem of finding the best allocation for individuals. Without loss of any generality, we can restrict attention to sustainable mechanisms, since we can always choose the social plan \((M, x)\) to replicate the equilibrium path actions.

The best allocation is therefore a solution to the following program for determining the best sustainable mechanism:

\[
\max \mathbb{E} \left[ \sum_{t=0}^{\infty} \beta^t u \left( c_t \left( z^t \left( \alpha_t \left( \theta^t \right) \right) \right), l_t \left( z^t \left( \alpha_t \left( \theta^t \right) \right) \right) \mid \theta^t \right) \right]
\]

subject to a resource constraint of the form

\[
K_{t+1} = F \left( K_t, \int l_t \left( z^t \left( \alpha_t \left( \theta^t \right) \right) \right) dG^t \left( \theta^t \right) \right) - \int c_t \left( z^t \left( \alpha_t \left( \theta^t \right) \right) \right) dG^t \left( \theta^t \right) - x_t,
\]

\(19\) As noted in the Introduction, this statement refers to the case in which messages are sent to the government. It is possible to construct alternative environments with fictional mechanism designers with full commitment power, so that the revelation principle holds.
a set of incentive compatibility constraints for individuals, i.e.,

\[ \alpha \text{ is a best response to (incentive compatible against)} \left[ \{ \hat{M}_t, \xi_t, \hat{x}_t, \hat{c}_t^{\ell} \}_{t=0}^\infty | (M, x) \right] \] (6)

and the “sustainability” constraint of the government:

\[ \sum_{s=0}^\infty \delta^s v(x_{t+s}) \geq \max_{\hat{x}_t, \hat{K}_{t+1}} \left\{ v(\hat{x}_t) + \delta v^c_t \left( \hat{M}^t, \hat{K}_{t+1} \right) \right\}, \] (7)

for all \( t \geq 0 \). This last constraint encompasses all the possible deviations by the government at date \( t \): the left-hand side is what the government will receive from date \( t \) onwards by sticking with the implicitly-agreed consumption schedule for itself. The right-hand side is the maximum it can receive by deviating. The deviations include both the offer of a new submechanism at time \( t + 1 \) (encapsulated into \( v^c_t \)) and also a deviation at the last stage of the subgame at time \( t \) to expropriation, \( \xi_t = 1 \). In the latter case, the government chooses \( \hat{x}_t \) and \( \hat{K}_{t+1} \) to maximize its deviation value, which is given by current utility, \( v(\hat{x}_t) \), and continuation value, written as \( v^c_t \left( \hat{M}^t, \hat{K}_{t+1} \right) \), to emphasize that this continuation value depends on the entire history of submechanisms (thus information) up to time \( t \), \( \hat{M}^t \), and on the capital stock from then on, \( \hat{K}_{t+1} \).

Our first result establishes that \( v^c_t \left( \hat{M}^t, \hat{K}_{t+1} \right) = 0 \) for all \( \hat{M}^t \in \mathcal{M}^t \) and \( \hat{K}_{t+1} \in \mathbb{R}_+ \), so (7) takes the form

\[ \sum_{s=0}^\infty \delta^s v(x_{t+s}) \geq v(F(K_t, L_t)), \] (8)

where \( K_t \) is total capital stock and \( L_t \) is total labor supply.

**Proposition 1** In the best sustainable mechanism, \( v^c_t \left( \hat{M}^t, \hat{K}_{t+1} \right) = 0 \) for all \( \hat{M}^t \in \mathcal{M}^t \) and \( \hat{K}_{t+1} \in \mathbb{R}_+ \), so (7) takes the form

\[ \sum_{s=0}^\infty \delta^s v(x_{t+s}) \geq v(F(K_t, L_t)), \]

where \( K_t \) is total capital stock and \( L_t \) is total labor supply.

**Proof.** Reducing \( v^c_t \left( \hat{M}^t, \hat{K}_{t+1} \right) \) is equivalent to relaxing the constraint on problem (4), so is always preferred. Since \( v^c_t \geq 0 \) (i.e., \( x \geq 0 \) and \( v(0) = 0 \)), we only need to show that \( v^c_t \left( \hat{M}^t, \hat{K}_{t+1} \right) = 0 \) is achievable for all \( \hat{M}^t \in \mathcal{M}^t \) and \( \Gamma' \in \mathcal{G} \). The following simple combination of strategies would achieve this objective. Let \( \rho^t \) be the history of actions by the government. Also denote \( \hat{c}_t = \hat{c}_t^0 \) be the mapping that allocates zero consumption to all
individuals irrespective of past and current reports. Let \( \rho^t = \tilde{\rho}^t \) if \( \tilde{x}_{t-s} = x_{t-s} \) and \( \tilde{M}_{t-s} = M_{t-s} \) for all \( s > 0 \). Then the following strategy combination would ensure \( v_F^t \left( \tilde{M}^t, \tilde{K}_{t+1} \right) = 0 \) for all \( t \): (1) for the citizens, \( \alpha = (\tilde{\alpha} \mid \alpha^0) \), for some \( \tilde{\alpha} \), which means that for each citizen \( i \) and for all \( t \), if \( \rho^{t-1} = \tilde{\rho}^{t-1} \), then \( \alpha_i^t = \tilde{\alpha}^t \), and if \( \rho^{t-1} \neq \tilde{\rho}^{t-1} \), then \( \alpha_i^t = \alpha^0 \); (2) for the government, \( \Gamma \), such that if \( \rho^{t-1} = \tilde{\rho}^{t-1} \), then \( \Gamma \) involves \( \tilde{x}_t = x_t, \tilde{M}_t = M_t \), and \( \xi_t = 0 \); and if \( \rho^{t-1} \neq \tilde{\rho}^{t-1} \), then it involves \( \xi_t = 1, \tilde{x}_t = F(K_t, L_t) \), and \( \tilde{c}_t = \tilde{c}_t^0 \). We next need to show that these strategies are sequentially rational. Consider the citizens; it suffices to note that following a history where \( \rho^{t-1} \neq \tilde{\rho}^{t-1} \), the government is playing \( \xi_{t+s} = 1, \tilde{x}_{t+s} = F(K_{t+s}, L_{t+s}) \) and \( \tilde{c}_{t+s} = \tilde{c}_{t+s}^0 \) for all \( s \geq 0 \). Therefore, any strategy other than \( \alpha^0 \) will give some utility less than \( \mathbb{E} \left[ \sum_{s=0}^{\infty} \beta^s u(0,0 \mid \theta_{t+s}) \mid \theta^t \right] \) to an individual with type history \( \theta^t \) in the current period, which is the utility that always playing \( \alpha^0 \) delivers. This argument proves that this strategy is sequentially rational for the citizens. It is also sequentially rational for the government, since after any history of \( \rho^{t-1} \neq \tilde{\rho}^{t-1} \), there will be no future output to expropriate, thus playing \( \xi_{t+s} = 1, \tilde{x}_{t+s} = F(K_{t+s}, L_{t+s}) \) and \( \tilde{c}_{t+s} = \tilde{c}_{t+s}^0 \) is a best response for the government starting in all of its information sets for all \( s \geq 0 \). The fact that \( v_F^t \left( \tilde{M}^t, \tilde{K}_{t+1} \right) = 0 \) for all \( \tilde{M}^t \in M^t \), then implies that the best deviation for the government is also \( \xi_t = 1, \tilde{x}_t = F(K_t, L_t) \), \( \tilde{c}_t = \tilde{c}_t^0 \) and \( \tilde{K}_{t+1} = 0 \) establishing that (7) takes the form (8).

This proposition therefore establishes that irrespective of the history of submechanisms and the amount of capital stock left for future production, if the government deviates from the implicitly-agreed social plan, there is an equilibrium continuation play which gives the government zero utility from that point onwards. Intuitively, citizens can coordinate on a punishment strategy, which gives zero utility to the government after deviation. This proposition therefore establishes that the best deviation at time \( t \) for the government involves \( \xi_t = 1 \) and \( \tilde{x}_t = F(K_t, L_t) \), and thus enables us to simplify the sustainability constraints of the government to a single constraint (8), which also has the virtue of not depending on the history of submechanisms up to that point.

**Remark 1** By allowing all citizens to claim to be the worst type in the punishment phase, Assumption 3 plays a crucial role in the proof of Proposition 1. An alternative game form delivering Proposition 1 without Assumption 3 is as follows

1. At the beginning of period \( t \), the government offers a menu of labor-consumption bundles,
possibly dependent on past histories, denoted by $M^C_t$.

2. Individuals choose how much labor to supply, $l_t \in [0, \bar{l}]$, and production takes place.

3. The government decides $\xi_t \in \{0, 1\}$. If $\xi_t = 1$, all output is expropriated, and consumed or distributed as before. If $\xi_t = 0$, production is distributed among agents who have chosen labor supply level as specified in the menu $M^C_t$, and those who have chosen a labor supply level that is not in the menu ($i.e., l_t$ such that $\exists c_t \in R_+ \text{ with } (l_t, c_t) \in C_t$), receive zero consumption.

This game form directly imposes “freedom of labor supply”, meaning that it is individuals who decide how much labor to supply, whereas with our original game form, the submechanism determines how much labor individuals supply. With this game form, Proposition 1 applies exactly without Assumption 3. We chose the game form in the main text and Assumption 3 for two reasons: first, the game form in the main text is closer to the standard Mirrlees setup and the mechanism design structure; second, the game form described here imposes an additional restriction on the set of mechanisms, such that individual allocations can be history dependent only to the extent that individuals have chosen different labor-consumption bundles in the past. In contrast, our more general game form allows history-dependent allocations in which two individuals who have made different reports, but received the same consumption-labor bundle in the past may be treated differently in the future.

Next, we define a direct (sub)mechanism as $M^*_t : \Theta_t \to [0, \bar{l}] \times R$. 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 government inducing direct submechanisms along the equilibrium path by $\Gamma^*$. Let us also introduce the notation $\alpha = (\alpha | \alpha')$ to denote a strategy profile where individuals play $\alpha$ along the equilibrium path and $\alpha'$ off the equilibrium path.

**Definition 3** A strategy profile for the citizens, $\alpha^*$, is truthful if, along the equilibrium path, we have that $\alpha^*_i (\theta^i | \theta^{i-1}, \rho^{i-1}) = \alpha^*$. We write $\alpha^* = (\alpha^* | \alpha')$ to denote a truthful strategy profile.

The notation $\alpha^* = (\alpha^* | \alpha')$ emphasizes that individuals play truth-telling along the equilibrium path, but may play some different strategy profile, $\alpha'$, off the equilibrium path. Clearly,
a truthful strategy against a direct mechanism simply amounts to reporting the true type of the agent. We are now ready to define the revelation principle as it applies to our environment. Before doing this, let us define $c[\Gamma, \alpha], l[\Gamma, \alpha]$ and $x[\Gamma, \alpha]$ as equilibrium consumption and labor supply distributions across individuals (as a function of their types) and sequence of government consumption levels resulting from the strategy profiles of the government and individuals.

**Theorem 1 (Truthful Revelation along the Equilibrium Path)** For any combination of equilibrium strategy profiles $\Gamma$ and $\alpha$, there exists another pair of equilibrium strategy profiles $\Gamma^*$ and $\alpha^*$ such that $\Gamma^*$ induces direct submechanisms and $(\alpha^* | \alpha')$ induces truth telling along the equilibrium path, and moreover $c[\Gamma, \alpha] = c[\Gamma^*, \alpha^*], l[\Gamma, \alpha] = l[\Gamma^*, \alpha^*]$ and $x[\Gamma, \alpha] = x[\Gamma^*, \alpha^*]$.

**Proof.** Take any equilibrium strategy profiles $\Gamma$ and $\alpha$. Let the best response of type $\theta_t$ at time $t$ according to $\alpha$ be to announce $z_{t, \Gamma} (\theta_t)$ given a history of reports $z_{t-1} (\theta_t)$. Let $z_t (\theta_t) = (z_t^{t-1} (\theta_t), z_{t, \Gamma} (\theta_t))$. Denote the utility of this individual under this mechanism be $\hat{u} [z_t (\theta_t) | \theta_t, \Gamma]$. By definition of $z_t (\theta_t)$ being a best response, we have

$$
\hat{u} [z_t (\theta_t) | \theta_t, \Gamma] \geq \hat{u} [\hat{z}_t (\theta_t) | \theta_t, \Gamma] \text{ for all } \hat{z}_t \in Z_t.
$$

Now consider the alternative strategy profile for the government $\Gamma^*$, which induces the action profile $\Gamma^*_{M^*, x} = \left\{ (\tilde{M}_t, \xi_t, \tilde{x}_t, \tilde{c}_t)_{t=0}^\infty | (M^*, x) \right\}$ such that $\tilde{M}_t = M_t^*$ (where $M_t^*$ is a direct submechanism) and $c[\Gamma^*, \alpha^*] = c[\Gamma, \alpha], l[\Gamma^*, \alpha^*] = l[\Gamma, \alpha]$, and $x[\Gamma, \alpha] = x[\Gamma^*, \alpha^*]$. Therefore, by construction,

$$
\hat{u} [\theta_t | \theta_t, \Gamma^*] = \hat{u} [\hat{z}_t (\theta_t) | \theta_t, \Gamma^*] \geq \hat{u} [\hat{z}_t (\theta_t) | \theta_t, \Gamma] = \hat{u} [\hat{\theta}_t | \theta_t, \Gamma^*] \text{ for all } \hat{\theta}_t \in \Theta_t. \tag{9}
$$

Equation (9) implies that $\alpha^* = (\alpha^* | \alpha')$ is a best response along the equilibrium path for the agents against the mechanism $M^*$ and government strategy profile $\Gamma^*$. Moreover, by construction, the resulting allocation when individuals play $\alpha^* = (\alpha^* | \alpha')$ against $\Gamma^*$ is the same as when they play $\alpha$ against $\Gamma$. Therefore, by the definition of $\Gamma$ being sustainable, we have $\Gamma \succeq_\alpha \Gamma'$ for all $\Gamma' \in \mathcal{G}$. Now choose $\alpha'$ to be identical to $\alpha$ off-the-equilibrium path, which implies that $\Gamma^* \succeq_\alpha \Gamma'$ for all $\Gamma' \in \mathcal{G}$, establishing that $(\Gamma^*, \alpha^*)$ is an equilibrium, completing the proof. ■
The most important implication of this theorem is that for the rest of the analysis, we can restrict attention to truth-telling (direct) mechanisms on the side of the agents. The reason why, despite the lack of commitment and the self-interested preferences of the mechanism designer, a revelation principle type result holds is that the structure of the game allows individuals to use punishment strategies involving zero labor supply following a deviation by the government. These punishments strategies support a sustainable mechanism, i.e., the government prefers not to deviate from the implicitly-agreed social plan \((M, x)\). Given this sustainability, there is effective commitment on the side of the government along the equilibrium path. This notion is important to distinguish from the commitment that exists in the standard mechanism design problems where there is unconditional commitment. In contrast, in our environment, there is no commitment off the equilibrium path, when the government chooses a different sequence of mechanisms. In this case, it can exploit the information it has gathered or expropriate part of the output. Nevertheless, by definition along a sustainable mechanism, the government prefers not to deviate from the implicitly-agreed social plan and thus individuals can report their types without the fear that this information or their labor supply will be misused.

In addition to facilitating the analysis in this paper, we believe that the use of a version of the revelation principle in this class of environments is an important methodological contribution, since it demonstrates that dynamic games between agents and governments (mechanism designers) with commitment problems and self-interested objectives can be analyzed without giving up the revelation principle along the equilibrium path. Instead, we simply need to ensure that the mechanism is sustainable.

### 3.3 The Best Sustainable Mechanism

Theorem 1 enables us to focus on direct mechanisms and truth-telling strategy \(\alpha^*\) by all individuals. This implies that the best sustainable mechanism (and thus the best allocation) can be achieved by individuals simply reporting their types. Recall that at every date, there is an invariant distribution of \(\theta\) denoted by \(G(\theta)\). This implies that \(\theta^t\) has an invariant distribution, which is simply the \(t\)-fold version of \(G(\theta), G^t(\theta)\) (since there is a continuum of individuals, each history \(\theta^t\) occurs infinitely often).\(^{20}\) Given this construction, we can write total labor supply as \(L_t = \int l_t(\theta^t) \ dG^t(\theta^t)\), and total consumption as \(C_t = \int c_t(\theta^t) \ dG^t(\theta^t)\).

\(^{20}\)More formally, given the continuum of agents, we can apply a law of large numbers, and each history \(\theta^t\) will have positive measure. See, for example, Uhlig (1996).
Moreover, since Theorem 1 establishes that any sustainable mechanism is equivalent to a direct mechanism with truth-telling on the side of the agents, it also implies:

**Proposition 2** The best sustainable mechanism is a solution to the following maximization program:

\[
U_{SM} = \max_{\{c_t(\theta^t), l_t(\theta^t), x_t, K_{t+1}\}_{t=0}^\infty} \mathbb{E} \left[ \sum_{t=0}^\infty \beta^t u \left( c_t \left( \theta^{i,t} \right), l_t \left( \theta^{i,t} \right) \right) \right]
\]

subject to some initial condition \(K_0\), the resource constraint

\[
K_{t+1} = F \left( K_t, \int l_t (\theta^t) dG^t (\theta^t) \right) - \int c_t (\theta^t) dG^t (\theta^t) - x_t,
\]

a set of incentive compatibility constraints for individuals,

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

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

for all \(\theta^{i,t}\) and all possible sequences of \(\{\theta^{i,t+s}_t\}_{s=0}^\infty\), and the sustainability constraint of the government

\[
\sum_{s=0}^\infty \delta^s v (x_{t+s}) \geq v \left( F \left( K_t, \int l_t (\theta^t) dG^t (\theta^t) \right) \right),
\]

for all \(t\).

**Proof.** The proof immediately follows from Proposition 1 and Theorem 1. Suppose there exists a sequential equilibrium \((\alpha^{**}, \Gamma^{**})\), which maximizes (10). By definition, \((\alpha^{**}, \Gamma^{**})\) will not feature \(\xi_t = 1\) for any \(t\), since this would involve zero output and zero consumption after some date, which could not be a solution to maximizing ex ante citizen utility. Therefore, \((\alpha^{**}, \Gamma^{**})\) features a sequence of submechanisms \(\{M_t\}_{t=0}^\infty\), some consumption levels for the government, \(\{x_t\}_{t=0}^\infty\) and \(\xi_t = 0\) for all \(t\). Then set \((M, x) = (\{M_t\}_{t=0}^\infty ; \{x_t\}_{t=0}^\infty)\), and use Theorem 1 to find \((\alpha^*, \Gamma^*)\) corresponding to a sustainable direct mechanism. This direct mechanism has to satisfy the resource constraint, (11), the incentive compatibility constraints of individuals at all dates, which instead of (6) can be written as (12) since \(\Gamma^*\) induces direct mechanisms. Finally, from Proposition 1, the constraint (13) ensures that \(\Gamma^*\) is a best response to citizens’ strategies, \(\alpha^*\).  


Note also that this optimization problem defines $U^{SM}$ as the ex ante value of the best sustainable mechanism for an individual. The role of Theorem 1 in this formulation is obvious, since it enables us to write the program for the best sustainable mechanism as a direct mechanism with truth-telling, thus reducing the larger set of incentive compatibility constraints of individuals to (12).21

4 Sustainable Mechanisms and the Quasi-Mirrlees Program

The quasi-Mirrlees program is defined as the following maximization problem:

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

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

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

and

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

Here $c_t(\theta^t)$ and $l_t(\theta^t)$ are assumed to be $\mathcal{F}_t$-measurable. Note also that the optimal value of this program is defined as a function of the sequence $\{C_t, L_t\}_{t=0}^{\infty}$. In other words, this program takes this infinite sequence as a given and maximizes the ex ante utility of an individual subject to the usual incentive compatibility constraints as well as two additional constraints. The first, (15), requires the sum of consumption levels across agents for all report histories to be no greater than some number $C_t$, while the second, (16), requires the sum of labor supplies to be no less than some amount $L_t$.

The constraint set of this program may be empty for some sequences $\{C_t, L_t\}_{t=0}^{\infty}$. For example, if $C_t = 0$ and $L_t > 0$, there will be no way of satisfying the incentive compatibility constraints to extract positive labor supply from the individuals. We denote the set of sequences such that the constraint set is non-empty by $\Lambda^\infty$, i.e.,

$$
\Lambda^\infty = \{ \{C_t, L_t\}_{t=0}^{\infty} \text{ such that } \exists \{c_t(\theta^t), l_t(\theta^t)\}_{t=0}^{\infty} \text{ satisfying (12), (15) and (16)} \}.
$$

---

21It is also useful to note that (12) encapsulates a small subset of all potential incentive compatibility constraints because we are focusing attention on those that apply along the equilibrium path (recall (3)). This can be seen from the fact that expectations on both sides of the constraints are taken conditional on $\theta^{t,s}$; this implies that such constraints should hold after any history of truth telling. Nevertheless, there is no loss of generality in this way of writing, since (12) needs to hold for any sequence of reports $\{\tilde{\theta}_{t+s}^i\}_{s=0}^{\infty}$, thus any potential deviation from time $t = 0$ is covered by this set of constraints.
We show in Appendix B that the set of constraints on the problem (14) forms a compact set, while the objective function, $u$, is clearly continuous. Therefore, $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is well defined as a functional. Nevertheless, the incentive compatibility constraints embedded in (12) do not form a convex set. For this reason, in the Appendix B, we follow Prescott and Townsend (1984a,b) and allow lotteries to convexify the constraint set and establish concavity of $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ in $\{C_t, L_t\}_{t=0}^{\infty}$. This will change the exact form of the optimization problem, but not its economic essence. For this reason, we relegate the formalism of the lotteries to Appendix B, and in the text, we assume that $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is concave. In Appendix B, we also prove that $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ is differentiable, which we again assume to be the case in the text.

It is now useful to relate the quasi-Mirrlees program to the dynamic optimal taxation a la Mirrlees. It is evident that the maximization problem

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

subject to

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

(18)

is equivalent to the dynamic Mirrlees optimal taxation problem as analyzed, for example, in Golosov, Kocherlakota and Tsyvinski (2003) or Werning (2002). Therefore, the quasi-Mirrlees problem decomposes the dynamic Mirrlees problem into two subproblems; one of finding the best allocation for a given sequence of $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^\infty$, and a second one of choosing the sequence $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^\infty$.

For us, however, it has a different use. Returning to the program of characterizing the best sustainable mechanism, (10), this problem can be written as

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

subject to (13), $\{C_t,L_t\}_{t=0}^{\infty} \in \Lambda^\infty$, and $x_t = F(K_t, L_t) - C_t - K_{t+1}$. This formulation therefore establishes the following theorem.

**Theorem 2** The best sustainable mechanism solves a quasi-Mirrlees program for some sequence $\{C_t, L_t\}_{t=0}^{\infty} \in \Lambda^\infty$. 

21
Proof. This follows immediately from rewriting (10)-(13) as a two-step maximization program, and expressing (11) as \( x_t = F(K_t, L_t) - C_t - K_{t+1} \). ■

The significance of this result lies in the fact that by using the revelation principle along the equilibrium path (Theorem 1) and decomposing the problem of finding the best sustainable mechanism, we have shown that any allocation consistent with the best sustainable mechanism is a solution to a problem which maximizes the ex ante utility of the citizens. Therefore, despite the political economy constraints and the resources extracted by the government from the society, the mechanism will maximize the ex ante utility of the citizens given some resource constraints (which are in addition to the resource constraints imposed by feasibility).

To make more progress, we need to characterize the behavior of the sequences \( \{C_t, L_t\}_{t=0}^\infty \) and \( \{x_t\}_{t=0}^\infty \) under the best sustainable mechanism, which is what we turn to next.

5 The Economy with Private Histories

The dynamic behavior of the optimal sustainable mechanism is simultaneously determined by the need to provide dynamic incentives both to the government and to individual agents. As is well known from the dynamic mechanism design problems (e.g. Green, 1987 or Atkeson and Lucas, 1992), the behavior of aggregate variables in these environments is typically very complicated even in the absence of sustainability constraints on the government.

In order to highlight the effect of government sustainability constraints, we first consider mechanisms with private histories, i.e., where individual histories are not observed by the government. The restriction to private histories is purely a heuristic device, useful in separating different parts of the analysis. Section 6 below will drop this assumption and characterize the best sustainable mechanism in the history-dependent case without any restrictions on potential mechanisms or strategies.

5.1 Best Sustainable Mechanism with Private Histories

To further simplify the notation in this case, we also assume that there is no capital in the economy, so that the aggregate production function of the economy is

\[
L_t = F(K_t, L_t) = \int l_t(\theta_t) dG(\theta_t),
\]

(20)

with \( K_0 = 0 \).
The restriction to private histories implies that in admissible mechanisms, allocations must depend only on agents’ current report. In such an environment the incentive compatibility constraints for agents can be separated across time periods, and written as

$$u(c_t(\theta_t), l_t(\theta_t) | \theta_t) \geq u(c_t(\hat{\theta}_t), l_t(\hat{\theta}_t) | \theta_t)$$

for all $\hat{\theta}_t \in \Theta$ and $\theta_t \in \Theta$, and for all $t$. Moreover, given the single crossing property in Assumption 3, (21) can be reduced to a set of incentive compatibility constraints only for neighboring types. Since there are $N + 1$ types in $\Theta$, this implies that (21) is equivalent to $N$ incentive compatibility constraints.\footnote{More specifically, in pure strategy direct mechanisms, there will be $N (N + 1)$ incentive compatibility constraints, and Assumption 2 makes sure that only $N$ of those, i.e., those between neighboring types, where the higher type may want to misreport to be the next lower type, may be binding.}

The best sustainable mechanism with private histories maximizes (10) subject to (13), (20), and (21).

Recall now the quasi-Mirrlees program defined above. It is straightforward to see that because of history independence, the optimal allocations of $(c_t, l_t)$ depend only on $C_t$ and $L_t$ and are independent of any $C_s, L_s$ for $s \neq t$. This implies that $U(\{C_t, L_t\}_{t=0}^\infty)$ is time separable, i.e., $U(\{C_t, L_t\}_{t=0}^\infty) = \sum_{t=0}^\infty \beta^t U(C_t, L_t)$ for some real-valued differentiable function $U : \mathbb{R}_+^2 \to \mathbb{R}$. The quasi-Mirrlees program therefore becomes:

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

subject to $C_t + x_t \leq L_t$, and

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

for all $t$.

As before, this problem is well defined only for some $(C, L)$. We denote the set of such $(C, L)$ pairs, which is a simplified version of (17), by $\Lambda$, i.e., $\Lambda \equiv \{(C, L) : \exists (c(\theta), l(\theta)) \text{ s.t. (15), (16) and (21) are satisfied}\}$. Define $\hat{v} (C, L) \equiv v(L - C) - (1 - \delta) v(L)$, which is the gap between the left and the right hand side of (22). Also define $\bar{w} \equiv \max_{(C, L) \in \Lambda} v(L - C) / (1 - \delta)$. Clearly, only values $w \leq \bar{w}$ can be promised to the government.

Assumption 4 (sustainability) There exists $\left(\hat{C}, \hat{L} \right) \in \text{arg max} v(L - C) / (1 - \delta)$, such that $\hat{v} \left(\hat{C}, \hat{L} \right) > 0$. 

22
This assumption ensures that the highest discounted utility that can be given to the government is sufficient to satisfy its sustainability constraint. Clearly this assumption is satisfied if the discount factor of the government, \( \delta \), is sufficiently large.\(^{23}\) It is also evident that, since \( U(C, L) \) is differentiable, the solution to the (full) Mirrles program (19) has to satisfy:\(^{24}\)

\[
U_C(C, L) = -U_L(C, L),
\]

where \( U_C \) is the partial derivative of \( U(C, L) \) with respect to \( C \) and \( U_L(C, L) \) is defined likewise. In Appendix B, we show that \( U(C, L) \) is differentiable, so equation (23) is meaningful. This observation immediately leads to the following definition:

**Definition 4** In the model with no capital and with private histories, we say that a (potentially stochastic) sequence \( \{C_t, L_t\}_{t=0}^{\infty} \) is undistorted at \( t \) if equation (23) holds (almost surely) for \( C_t \) and \( L_t \), and we say that it is asymptotically undistorted, if (23) holds (almost surely) as \( t \to \infty \).

This is a natural definition. Equation (23) implies that the marginal benefit from one more unit of consumption is equal to the marginal cost of one more unit of output produced by additional labor supply given the utility function \( U(C, L) \), which is the ex ante utility function of the agents in this economy once we take the incentive compatibility and feasibility constraints into account. Consequently, equation (23) implies that there are no “aggregate distortions” (beyond those implied by private information as in the typical Mirrles economy).

**Lemma 1** Consider a sequence of \( \{C_t, L_t\}_{t=0}^{\infty} \). Then:

1. the marginal labor tax rate on the highest type of agent, \( \theta_N \), at time \( t \) is given by \( \tau_{N,t} = 1 + U_L(C_t, L_t) / U_C(C_t, L_t) \).

2. if \( \{C_t, L_t\}_{t=0}^{\infty} \) is undistorted at \( t \), the labor supply decision of the highest type of agent is undistorted, i.e., \( u_c(c_t(\theta_N), l_t(\theta_N) \mid \theta_N) = -u_l(c_t(\theta_N), l_t(\theta_N) \mid \theta_N) \).

**Proof.** Assumption 2 implies that we only need to check incentive compatibility constraints for neighboring types. Appendix B establishes that Lagrange multipliers exist, and let \( u_c \) and

\(^{23}\)Notice that since \( \Lambda \) does not depend on \( \delta \), so as \( \delta \to 1 \), this assumption is surely satisfied.

\(^{24}\)It is straightforward to see that in this case \( (C, Y) \in \text{Int} \Lambda \).
Let $u_l$ be the partial derivatives of $u$ (which exist by Assumption 1). Therefore, we have

\[ u_c(c_t(\theta_N), l_t(\theta_N) \mid \theta_N)(1 + \lambda_{Nt}) = \nu_{Ct}, \]

\[ u_l(c_t(\theta_N), l_t(\theta_N) \mid \theta_N)(1 + \lambda_{Nt}) = -\nu_{Lt}, \]

where $\lambda_{Nt}$ is the multiplier on incentive compatibility constraint between types $\theta_N$ and $\theta_{N-1}$ at time $t$, $\nu_{Ct}$ is the multiplier on (15) at $t$ and $\nu_{Lt}$ is the multiplier on (16) 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

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

where the first equality defines $\tau_{N,t}$, and the second equality establishes the first result. 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 (23).

To make further progress, let us follow Thomas and Worrall (1988) and consider the recursive formulation of our problem, whereby

\[ V(w) = \max_{C, L, x, w'} \{ U(C, L) + \beta V(w') \} \tag{24} \]

subject to

\[ C + x \leq L, \]

\[ w = v(x) + \delta w', \tag{25} \]

\[ v(x) + \delta w' \geq v(L), \tag{26} \]

\[ w' \in \mathbb{W} \text{ and } (C, L) \in \Lambda. \tag{27} \]

where $w$ is a future utility promised to the government, $\mathbb{W}$ is a set of feasible values for $w$, and the requirement that $(C, Y) \in \Lambda$ make sure that we only look at feasible levels of aggregate consumption and labor supply. The program in (24) determines optimal policies for a given level of promised utility $w$. The problem of finding the best sustainable mechanism corresponds to solving (24) and then choosing the initial value $w_0$ such that $w_0 \in \arg \max_w V(w)$.\footnote{There are a number of technical details related to this program. First, we have that $\mathbb{W} = [0, \bar{w}]$ where $\bar{w}$ is the maximal feasible and sustainable promised utility to the government, defined above. Moreover, in the Appendix we show that there may be room for improving on this program by randomizing over the values...}
This program also makes the role of the sustainability constraint (26) clear. If the society wishes to produce more output (or supply more labor $L$), it can only do so by providing greater consumption to the government either today or in the future. Therefore when this constraint is binding, the social cost of increasing output will be greater than $U_L$, thus leading to further aggregate distortions.

The main result of this section is the following theorem:

**Theorem 3** Consider the economy with no capital and with private histories and suppose that Assumptions 1-4 hold.

1. At $t = 0$, there is an aggregate distortion and there is positive marginal tax on the labor supply of the highest type, $\theta_N$.

2. Suppose that $\beta \leq \delta$. Let $\Gamma$ be the best sustainable mechanism inducing a possibly stochastic sequence of values $\{w_t\}_{t=0}^{\infty}$ such that there exists a sequence of sets $\{W_t\}_{t=0}^{\infty}$ whereby $w_t \in W_t$. Then, we have that $\{w_t\}_{t=0}^{\infty}$ is a non-decreasing stochastic sequence in the sense that if $w_t = w_t' \in W_t$, then any $w_{t+1}' \in W_{t+1}$ satisfies $w_{t+1}' \geq w_t'$. Moreover, a steady state exists in that $\{w_t\}_{t=0}^{\infty}$ converges (almost surely) to some $w^*$ and $\{C_t, L_t, x_t\}_{t=0}^{\infty}$ converges (almost surely) to some $(C^*, L^*, x^*)$. Moreover, we have that $\lim_{t \to \infty} -U_C/U_L = 1$, so that asymptotically $\{C_t, L_t\}_{t=0}^{\infty}$ is (almost surely) undistorted, and the marginal tax rate on the labor supply of the highest type goes to zero.

3. If $\beta > \delta$, then aggregate distortions and the positive marginal tax on the labor supply of the highest type do not disappear even asymptotically.

**Proof.** See Appendix B. ■

The most important results are in parts 2 and 3. Part 2 states that as long as $\beta \leq \delta$, asymptotically the economy converges to an equilibrium where there are no aggregate distortions and the marginal tax rate on the highest type is equal to zero. Therefore, this theorem, in combination with Theorem 2, implies that despite the political economy constraints and
the commitment problems, many of the insights of the optimal taxation literature inspired by Mirrlees (1971) will continue to hold. Consequently, that when the government is at least as patient as the citizens, lessons from the optimal taxation literature are not only normative, but may also help us understand how tax systems are designed in practice where politicians are motivated by their own objectives, such as self-enrichment or reelection.

Part 3 of the theorem is perhaps more important, however. This part states that if the government is less patient than the agents, distortions will not disappear. Since in many realistic political economy models, the government or politicians are more short-sighted than citizens, this part of the theorem may imply that in a number of important cases, political economy considerations will lead to additional distortions that will not disappear even asymptotically.

We now give a heuristic argument to support this theorem (while the full proof is in the Appendix). Let $\gamma$ and $\psi \geq 0$ be the Lagrange multipliers on the constraints (25) and (26) respectively. Lemma 10 in the Appendix shows that $V(w)$ is differentiable. Furthermore, in the text, we simplify the discussion by assuming that $(C,Y) \in \text{Int} \Lambda$ and $w' \in \text{Int} \bar{W}$. Therefore, taking the first order condition with respect to $w'$ and using the Envelope theorem, we obtain that

$$\frac{\beta}{\delta} V'(w') = -\psi - \gamma = V'(w) - \psi$$

The other first order conditions with respect to $C$, $L$ and $x$ imply:

$$U_C + U_L = \psi v'(L), \text{ and}$$

$$v'(x)(\psi + \gamma) = U_C.$$  

Equation (29) makes it clear that aggregate distortions are related to $\psi$. It is also evident that we must have $\psi > 0$ at $t = 0$, otherwise the government must receive $w_0 = 0$ initially, which together with the sustainability constraint (26) would imply $C_t = L_t = 0$ for all $t$, which cannot be optimal. This implies that $U_C + U_L > 0$, and from Lemma 1, this yields $\tau_N > 0$ at $t = 0$.

Part 2 of Theorem 3, on the other hand, states that, as long as $\beta \leq \delta$, eventually aggregate distortions will disappear and marginal labor income taxes on the highest type will tend to zero. In many ways, this is a surprising result, but the structure of the model makes the intuition clear. To see why, let us start with the case where $\beta = \delta$, in which case equation (28)
implies
\[ V'(w') = V'(w) - \psi \leq V'(w) \]

The inequality above is strict when the sustainability constraint on the government (26) binds. This, combined with the concavity of the value function \( V(\cdot) \), which is proved in Lemma 8 in Appendix B, implies that \( w' \geq w \), with \( w' > w \) if \( \psi > 0 \) and \( w' = w \) if \( \psi = 0 \). This shows that the promised utilities for the government are nondecreasing as stated in part 2 of Theorem 3.

The intuition for why the rewards to the government are increasing (nondecreasing) is as follows. The incentives for the government in the current period are provided by both consumption in the current period, \( x \), and by consumption in future periods represented by the promised utility \( w \). Therefore, future government consumption not only relaxes the sustainability constraint in the future, but also in all prior periods. Thus, all else equal, optimal incentives for government are backloaded. The intuition for this backloaded compensation scheme is similar to the reasons why in principal-agent models backloading compensation may be useful (see, for example, Ray, 2002).

Since promised values to the government are in a compact set, this implies that they will converge to some value \( w^* \). Recall that \( \mathcal{W} = [0, \bar{w}] \). If \( w^* < \bar{w} \), (28) immediately implies that \( \psi = 0 \). In other words, the Lagrange multiplier on the sustainability constraint of the government, (26), eventually reaches zero, and at this point, aggregate distortions disappear. Lemma 1 then implies that the marginal tax rate on the labor supply of the highest type, \( \theta_N \), also vanishes as claimed in the theorem. The intuition for why the multiplier on the sustainability constraint eventually reaches zero is related to the fact that promised utilities to the government are increasing. Loosely speaking, we can remove some of the sustainability constraints in the very far future, and this will have no influence on the sequence of utilities promised to the government at time \( t = 0 \). This implies that eventually the multiplier on these sustainability constraint must tend to zero.\(^{26}\)

Finally, let us consider the case with \( \delta < \beta \). Since government is less patient than the agents, backloading incentives for government becomes costly for agents. Consider any \( w \) for

\(^{26}\)Note however that when \( w^* = \bar{w} \), (28) may no longer be valid, since it applies only on the interior of \( \mathcal{W} \). Nevertheless, in this case, a different argument spelt out in Lemma 11 in Appendix B establishes that aggregate distortions again disappear. Intuitively, \( \bar{w} \) involves the maximum utility for the government, and this is achieved without any distortions. The idea for why aggregate distortions disappear when \( \delta > \beta \) is also similar; with this configuration, the steady-state utility of the government necessarily reaches \( \bar{w} \), and the same argument establishes the desired result.
which constraint (26) does not bind. Then (28) implies that
\[ V'(w') > V'(w) \]
and \( w' < w \), so that promised utilities will be decreasing when the sustainability constraint, (26), is slack. In fact, if a steady state \((C^*, L^*, x^*)\) is ever reached, it will solve the following system of equations
\[
1 + \frac{U_L}{U_C} = \left(1 - \frac{\delta}{\beta}\right) \frac{v'(L^*)}{v'(x^*)} 
\]
(31)
\[ C^* + x^* = L^*, \text{ and} \]
\[ v(x^*) = (1 - \delta)v(L^*), \]
with the steady-state utility of the government equal to \( w^* = v(L^*) \). Equation (31) immediately shows that if a steady state is reached, there will be a positive labor distortion as long as \( \delta < \beta \), as claimed in part 3 of Theorem 3. The intuition for the presence of (asymptotic) aggregate distortions in this case is directly related to the fact that when the government is less patient than the agents, backloading does not work. Since backloading was essential for the multiplier on the sustainability constraint (26) going to zero, this multiplier remains positive, and the additional distortions created by the sustainability constraint remain even asymptotically.

5.2 An Example for an Economy with Private Histories and No Capital

We now illustrate the results from the previous subsection with a simple numerical example. Consider an economy with two types, i.e., \( \Theta = \{\theta_0, \theta_1\} \) and
\[
u(c, l | \theta) = \sqrt{c} - \frac{l^2}{m\theta}, \quad (32)\]
where \( m \) is a parameter determining the relative disutility of labor. We continue to assume that type \( \theta_0 \) is disabled and cannot supply any labor, so \( \theta_0 = 0 \), and we normalize \( \theta_1 = 1 \). Let us also assume that a fraction \( \pi = 1/2 \) of the population is of type \( \theta_1 \) and that the utility function of the government is also given by \( v(x) = \sqrt{x} \).

Since type \( \theta_0 \) cannot supply any labor, we have \( l(\theta_1) = L/\pi \). Moreover, the incentive compatibility constraint for type \( \theta_1 \) is
\[
\sqrt{c(\theta_1)} - \frac{l(\theta_1)^2}{m\theta_1} \geq \sqrt{c(\theta_0)}. \quad (33)
\]
Then $c(\theta_0)$ and $c(\theta_1)$ can be determined as solutions to (33) holding as equality and to the resource constraint, $(1 - \pi)c(\theta_0) + \pi c(\theta_1) = C$. Given this structure, $U(C, L)$ can be computed directly and used with the recursive program (24) to derive the value function $V(w)$.

We illustrate the shape of the value function using this recursive formulation. Let us assume that the government is as patient as the citizens, $\beta = \delta = 0.5$ and that $m = 2$. This case illustrates part 2 of Theorem 3. The resulting value function is plotted in Figure 1. Note that $V(w)$ is inverse U-shaped. The increasing part is due to the fact that if the government is given too low a level of utility, the sustainability constraint will force the economy to produce a very low level of output. For this reason, the relevant part of the value function $V(w)$ is the segment after the peak, which is everywhere decreasing. In fact, as noted above, with the best sustainable mechanism, the economy will start at the peak of the function $V(w)$ and move rightwards (towards higher values of $w$) over time.

![Figure 1: Theorem 3 part 2. Value function with $\beta = \delta = 0.5$.](image)

Next, we show the time path of promised values to the government and aggregate distortions for a range of different parameter values. In these computations, we use non-recursive methods, assuming that a steady state is reached after $N$ periods and then solving a sequence problem, which turns out to be faster and more accurate. Let us take the disutility of labor to be $m = 5$, the discount factor of agents $\beta = 0.9$ and consider four different values for the discount factor for the government $\delta = 0.9, 0.8, 0.7, \text{and } 0.6$. Figure 2 plots the time path of the promised value to the government, $w$, for the four different discount factors of the government. The lowest curves is for $\delta = 0.6$, and then, respectively, for $\delta = 0.7, 0.8 \text{ and } 0.9$. Consistent with
Figure 2: Time path of \( \{w_t\} \) with \( \beta = 0.9 \) and \( \delta = 0.9; 0.8; 0.7; 0.6 \). Higher curves correspond to higher values of \( \delta \).

the results in Theorem 3 part 2, when \( \delta = \beta \), the sequence \( \{w_t\} \) is an increasing sequence and converges to some level \( w^* \). Interestingly, in these examples the sequence \( \{w_t\} \) is increasing even when \( \delta < \beta \).

Figure 3, in turn, depicts the evolution of the aggregate distortion, \( 1 + U_L/U_C \) (which is also equivalent to the marginal tax on type \( \theta_1 \) in this case, recall Lemma 1). The lowest curve shows the case where \( \beta = \delta \), and consistent with part 2 of Theorem 3, the aggregate distortion converges to zero. An interesting feature of the example is that the convergence of \( \{w_t\} \) and of distortions is rather fast. This suggests that the best sustainable mechanism may converge to a mechanism without aggregate distortions and with zero marginal tax rate on the highest type agents very rapidly.

The figure also shows that, as predicted by part 3 of Theorem 3 and in contrast to the case with \( \beta = \delta \), with \( \delta < \beta \) aggregate distortions do not disappear asymptotically, and in fact, they could be quite sizable. For example, when \( \delta = 0.6 \), the aggregate distortion converge to an asymptotic value of 0.15 (the highest curve in the graph).
6 Optimal History-Dependent Sustainable Mechanisms

6.1 Characterization of Best Sustainable Mechanism

We now return to the analysis of the general problem in Sections 2-4 without the restriction to private histories, and we also incorporate capital. The analysis parallels the discussion of the best sustainable mechanism with private histories in Section 5. The quasi-Mirrlees program was defined above in (14), and Theorem 2 established that the best sustainable mechanism solves a quasi-Mirrlees program. In addition, note that Appendix C shows that $U(\{C_t, L_t\}_{t=0}^\infty)$ is differentiable in the sequences $\{C_t, L_t\}_{t=0}^\infty \in \Lambda^\infty$. This implies that we can think of changes in sequences $\{C_t, L_t\}_{t=0}^\infty$ where only one element, $C_s$ or $L_s$ for some specific $s$ is varied. We denote the derivative of $U$ with respect to such variations by $U_{C_s}(\{C_t, L_t\}_{t=0}^\infty)$ and $U_{L_s}(\{C_t, L_t\}_{t=0}^\infty)$ or simply $U_{C_s}$ and $U_{L_s}$. We also denote the partial derivatives of the production function by $F_{L_t}$ and $F_{K_t}$. Then we have:

**Definition 5** In the general environment, we say that (potentially stochastic) sequence $\{C_t, L_t\}_{t=0}^\infty$
is undistorted at time \( t \) (or as \( t \to \infty \)), if we have

\[
\mathcal{U}_{C_t} \cdot F_{L_t} = -\mathcal{U}_{L_t} \quad \text{and} \quad F_{K_{t+1}} \cdot \mathcal{U}_{C_{t+1}} = \mathcal{U}_{C_t}
\]

at time \( t \) (or as \( t \to \infty \)).

This definition is the natural generalization of Definition 4. In particular, the first condition requires the marginal cost of effort at time \( t \) given the utility function \( \mathcal{U}(\{C_t, L_t\}_{t=0}^\infty) \) to be equal to the increase in output from the additional effort times the marginal utility of additional consumption. The second one requires the cost of a decline in the utility by saving one more unit to be equal to the increase in output in the next period times the marginal utility of consumption then. Once again, these are aggregate conditions since they are defined in terms of the utility function \( \mathcal{U}(\{C_t, L_t\}_{t=0}^\infty) \), which represents the ex ante utility of an individual subject to incentive and feasibility constraints. Moreover, we say that there is no aggregate capital taxation at time \( t \) if \( F_{K_{t+1}} \cdot \mathcal{U}_{C_{t+1}} = \mathcal{U}_{C_t} \). It is important that this condition refers to no aggregate capital taxation, and does not rule out capital taxes on individuals in some possible decentralizations of these mechanisms.

The main result of this section parallels Theorem 3, but is weaker in some respects. To state this theorem, we need an assumption analogous to Assumption 4 in Section 5. We write \( \{C, L\} \in \tilde{\Lambda}^\infty \) if \( \{C_t, L_t\}_{t=0}^\infty \in \Lambda^\infty \) and \( C_t \to C, L_t \to L \). In other words, this notation implies that there exists a sequence \( \{C_t, L_t\}_{t=0}^\infty \) in \( \Lambda^\infty \) converging to a steady state with \( \{C, L\} \).\(^{27}\)

Next, with an analogy to our previous analysis, we define \( \hat{v}(C, L, K) \equiv v(F(K, L) - C - K) - (1 - \delta) v(F(K, L)) \) and \( \bar{w} \equiv \max_{\{C, L\} \in \Lambda, K \geq 0} v(F(K, L) - C - K) / (1 - \delta) \). The key sustainability assumption is a generalization of Assumption 4:

**Assumption 5 (general sustainability)** There exists

\[
\hat{v}(\hat{C}, \hat{L}, \hat{K}) \in \arg\max_{\{C, L\} \in \Lambda, K \geq 0} v(F(K, L) - C - K) / (1 - \delta) \quad \text{such that} \quad \hat{v}(\hat{C}, \hat{L}, \hat{K}) > 0.
\]

Moreover, when \( \{C_t, L_t, K_t\} \to (C^*, L^*, K^*) \) almost surely, let \( \mathcal{U}^*_{C_t} = \mathcal{U}_{C_t}(\{C_t, L_t\}_{t=0}^\infty) \).

Then, we have (proof in the Appendix):

**Theorem 4** Consider the model with the general environment and suppose that Assumptions 1-3 and 5 hold.

\(^{27}\)Note that despite the similarity of the symbols, \( \tilde{\Lambda}^\infty \) and \( \Lambda^\infty \) are very different sets. \( \Lambda^\infty \) is a subset of the vector space \( \mathcal{L}^\infty \), while \( \tilde{\Lambda}^\infty \subset \mathbb{R}^2 \).
1. At \( t = 0 \), there are aggregate distortions, so that there is positive aggregate capital and labor taxation.

Let \( \Gamma \) be the best sustainable mechanism, inducing a sequence of consumption, labor supply and capital levels \( \{C_t, L_t, K_t\} \). Suppose a steady state exists such that as \( t \to \infty \), \( \{C_t, L_t, K_t\} \to (C^*, L^*, K^*) \) almost surely. Moreover, let \( \varphi = \sup \{ \varepsilon \in [0,1] : \text{plim}_{t \to \infty} \varepsilon^{-t}U^*_{C_t} = 0 \} \), where \( \varphi < 1 \).

2. If \( \varphi \leq \delta \), then (almost surely) there are no asymptotic aggregate distortions on capital accumulation and labor supply.

3. If \( \varphi > \delta \), then aggregate distortions, the positive aggregate capital and labor taxation do not disappear even asymptotically.

**Proof.** See Appendix C. ■

The major results from Theorem 3 continue to hold here.\(^{28}\) The most important difference is that instead of comparing the \( \beta \), to the discount factor of the government, \( \delta \), we now compare the rate at which the ex ante marginal utility of consumption \( U^*_{C_t} \) is declining in the steady state, denoted by \( \varphi \), to the discount factor of the government, \( \delta \). Clearly, in the case where \( U(\{C_t, L_t\}_{t=0}^\infty) \) is time separable as in Theorem 3, the rate at which \( U^*_{C_t} \) declines in steady state is exactly equal to \( \beta \), so that the results in this theorem are closely related to those in Theorem 3. Moreover, in reality, \( \varphi \) is the fundamental discount factor of the citizens, since it measures how one unit of resources at time \( t \) compares with one unit of resources at time \( t + 1 \). Only in special cases (e.g., without any dynamic incentive linkages) does this fundamental discount factor coincide with \( \beta \). Therefore, the case of \( \varphi \leq \delta \) indeed corresponds to the situation in which the government is as patient as or more patient than the citizens. It is also noteworthy that, as long as a steady state exists, the rate at which \( U^*_{C_t} \) declines is independent of \( \delta \), so it is always possible to choose \( \delta \) to make sure that an economy is in part 2 or part 3 of this theorem.

The most important results here are again contained in parts 2 and 3. Part 2 states that as long as a steady state exists and \( U^*_{C_t} \) declines sufficiently rapidly, the multiplier of the sustainability constraint goes to zero. This establishes that the sequence \( \{C_t, L_t, K_t\}_{t=0}^\infty \)

---

\(^{28}\)The parts that are missing from this theorem relative to Theorem 3 are that the sequence of promised values to the government is increasing and a statement that a steady state exists.
is asymptotically undistorted, which no aggregate labor supply distortion and no aggregate capital tax. This generalizes the results from the economy with no capital and private histories to the much more general environment here. Part 3, on the other hand, states that if the discount factor of the government $\delta$ is sufficiently low, then aggregate distortions will not disappear, even asymptotically. The significance of this result is even greater than in Theorem 3, since it implies not only a marginal labor tax on the highest type but also aggregate capital taxes contrary to the literature on dynamic fiscal policy.

Once again, we provide a heuristic argument here, leaving the proof to the Appendix. Since the objective function is no longer time separable, to characterize the best sustainable mechanism in this case, we follow Marcet and Marimon (1998) and form a Lagrangian of the form

$$\max_{\{C_t, L_t, K_t, x_t\}_{t=0}^{\infty}} \mathcal{L} = \mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) + \sum_{t=0}^{\infty} \delta^t \left\{ \mu_t v(x_t) - (\mu_t - \mu_{t-1})v(F(K_t, L_t)) \right\} \tag{34}$$

subject to

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

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

for all $t$, where $\mu_t = \mu_{t-1} + \psi_t$ with $\mu_{-1} = 0$ and $\delta^t \psi_t \geq 0$ is the Lagrange multiplier on the constraint (13).$^{29}$

The differentiability of $\mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty})$ implies that for $\{C_t, L_t\}_{t=0}^{\infty} \in \text{Int} \Lambda^\infty$, we have:$^{30}$

$$\mathcal{U}_{Lt} - \delta^t (\mu_t - \mu_{t-1}) v'(F(K_t, L_t)) F_{Lt} = -\mathcal{U} C_t F_{Lt} \tag{36}$$

$^{29}$To derive (34), form the Lagrangian

$$\mathcal{L}' = \mathcal{U}(\{C_t, L_t\}_{t=0}^{\infty}) + \sum_{t=0}^{\infty} \delta^t \psi_t \left[ \sum_{s=0}^{\infty} \delta^s v(x_{t+s}) - v(F(K_t, L_t)) \right],$$

then note that

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

where $\mu_t = \mu_{t-1} + \psi_t$ with $\mu_{-1} = 0$. Substituting this in $\mathcal{L}'$ above gives (34).

$^{30}$To obtain these equations, let the multiplier on constraint (35) at time $t$ be $\kappa_t$. Then the first-order condition with respect to $C_t$ gives $\mathcal{U}_C = \kappa_t$, while the first-order condition with respect to $L_t$ gives

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

Substituting for $\kappa_t$ gives (36). The first-order condition with respect to $K_{t+1}$, on the other hand, gives

$$-\delta^t (\mu_{t+1} - \mu_t) v'(F(K_{t+1}, L_{t+1})) F_{K_{t+1}} + \kappa_t - \kappa_{t+1} F_{K_{t+1}} = 0$$

Substituting for $\kappa_t$ and $\kappa_{t+1}$ and rearranging gives (37).
\[ UC_t = [UC_{t+1} + \delta^t(\mu_{t+1} - \mu_t)v'(F(K_{t+1}, L_{t+1}))] F_{K_{t+1}} \]  

(37)

Since \( \mu_t \geq \mu_{t-1} \), this implies that

\[ U_{C_t} \geq UC_{t+1} F_{K_{t+1}}, \quad \text{and} \]

\[ UC_t \geq UC_{t+1} F_{K_{t+1}}. \]  

(38)

(39)

With the same argument as in the previous section, both of these inequalities have to be strict at \( t = 0 \), since the sustainability constraint, now (13), has to be binding at \( t = 0 \). This explains part 1 of the theorem.

Next, again for \( \{C_t, L_t\}_{t=0}^{\infty} \in \text{Int} \Lambda^\infty \), the first-order condition with respect to \( x_t \) yields:

\[ \frac{UC_t}{\delta^t v'(x_t)} = \mu_t \leq \mu_{t+1} = \frac{UC_{t+1}}{\delta^{t+1} v'(x_{t+1})}. \]  

(40)

We know that, by construction, \( \mu_t \) is an increasing sequence, so it must either converge to some value \( \mu^* \) or go to infinity. Suppose that \( (C_t, L_t, K_t) \) converges to some \( (C^*, L^*, K^*) \)—and \( x_t \) converges to \( x^* = L^* - C^* - K^* \). If \( U_{C_t}^* \) is proportional to some \( \varphi \leq \delta \), then we can show that \( \mu_t \) (almost surely) converges to some value \( \mu^* < \infty \), and that both (38) and (39) must hold as equality (see the proof of Theorem 4), establishing the result stated in part 2 of the Theorem. In contrast, if \( U_{C_t}^* \) is proportional to some \( \varphi > \delta \), then \( \mu_t \) tends to infinity and aggregate distortions do not disappear.

6.2 Example for History-Dependent Mechanisms

We now briefly illustrate the results of Theorem 4 and show how in some simple cases, \( \varphi \) defined as \( \text{sup}\{\varepsilon \in [0, 1] : \text{plim}_{t \to \infty} \varepsilon^{-t}U_{C_t}^* = 0\} \) is again equivalent to the discount factor of the agents, \( \beta \). In particular, let us consider the following economy with “almost constant types” as explained below. There are two types \( \Theta = \{\theta_0, \theta_1\} \) and the utility function is

\[ u(c, l | \theta) = u(c) - g(l/\theta), \]

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

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

subject to

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

and

$$\pi c_t(\theta_1) + (1 - \pi) c_t(\theta_0) \leq \pi l_t(\theta_1) - x_t,$$

where $L_t = \pi l_t(\theta_1)$ and $C_t = L_t - x_t$. The first constraint is the incentive compatibility constraint sufficient for the high type to reveal its identity given the presence of effective commitment along the equilibrium path. The second constraint is the resource constraint for each $t$. Note that because $\varepsilon$ is assumed to converge to 0, we do not specify other incentive compatibility constraints. Assigning Lagrange multipliers $\lambda$ and $\beta^t \mu_t$ to these constraints, the first-order necessary conditions of this problem can be written as:

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

Equations (42) imply that $u'(c_t(\theta_1)) = (1 - \pi - \lambda) u'(c_t(\theta_0)) = (1 - \pi) \mu_t$, and

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

Consequently, there is constant risk-sharing between the two types in all periods. Moreover, if a steady state exists, so that $x_t \to x^*$, (42) and (43) combined imply that $c_t(\theta_1) \to e^{1x^*}$, $c_t(\theta_0) \to e^{0x}$, and $l_t(\theta_1) \to l^*$, and hence $\mu_t \to \mu^*$. Consequently, in this case $\varphi = \beta$, so Theorem 4 applies in exactly the same form as Theorem 3. Therefore, in this particular case, the rate at which the derivative $U^*_C_t$ declines is very easy to determine, and it does so at the same rate as the discount factor of the citizens. It is also straightforward to see that the same argument generalizes to the case where there are more than two types.31

---

31 In fact, we conjecture that whenever there exists a stationary redistribution of consumption among individuals, $\varphi = \beta$, though we have not been able to prove this conjecture yet.
For a numerical illustration, we again consider the utility function (32) with two types and the same parameter values from subsection 5.2. For brevity, we simply show the aggregate distortion, $1 + U_{L_t}/U_{C_t}$, in Figure 4. Consistent with part 2 of Theorem 4, when $\beta = \delta = 0.9$, the lowest curve shows that the aggregate distortion converges to zero and the convergence is again rather fast. Instead, when $\delta < \beta$, the aggregate distortion converges to a positive, and potentially large, asymptotic value.32

6.3 Limits to Quasi-Mirrlees Programs

The analysis so far has exploited the structure of the problem in which the government can expropriate all of the output, and using this structure, established Theorem 2, showing that the best sustainable mechanism solves a quasi-Mirrlees program. This indicates that despite the political economy considerations, there is a close connection between the optimal dynamic taxation results a la Mirrlees and the equilibrium results presented here. This connection was further emphasized by the results showing that the marginal tax on the highest type and the aggregate tax on capital may go to zero asymptotically. In this subsection, we briefly discuss a more general class of economies which do not satisfy Theorem 2. This is useful to highlight what is involved in this quasi-Mirrlees result.

32 Note that the distortions are exactly the same here as for the case with the private histories studied in subsection 5.2. The reason is that without the sustainability constraints, the problems would have the same solution. When sustainability constraints are present but satisfied, there are no other reasons for the solutions to differ.
In particular, imagine an environment without capital and suppose that labor supply is equal to output. Recall that to simplify the exposition, we have so far assumed that the government can extract the full output within a given period. A more general assumption, as mentioned in footnote 17 is to allow the government to extract a fraction $\eta$ of the total output. It can be verified that this assumption does not change anything substantial in our analysis.

Here, let us make an alternative assumption and suppose that the government can tax individuals according to how much they produce. In particular, let the maximum amount that can be extracted from an individual supplying labor $l$ be a function, where $\tilde{\eta} : [0, \bar{l}] \to [0, \bar{l}]$. In this case, using an analog of Proposition 1, we have the sustainability constraint as:

$$\sum_{s=0}^{\infty} \delta^s v(x_{t+s}) \geq v \left( \int \tilde{\eta}(l(\theta^t)) \, dG^t(\theta^t) \right),$$

where $l(\theta^t)$ is the labor supply of an individual with type history $\theta^t$, and the term $\int \tilde{\eta}(l(\theta^t)) \, dG^t(\theta^t)$ captures the maximum amount that the government can expropriate given the technological restriction embedded in the function $\tilde{\eta}(\cdot)$ and the distribution of types given by $G^t(\theta^t)$. It is clear that unless $\tilde{\eta}(\cdot)$ is a linear function Theorem 2 does not apply.

The lesson from this brief analysis is that when the government, for informational or other reasons, can expropriate different amounts from individuals supplying different levels of labor, there will be further distortions relative to the baseline analysis presented above. In particular, the best sustainable mechanism will no longer solve a quasi-Mirrlees program. Because of space restrictions, we do not pursue a general analysis of this class of problems here and leave this for future work.

7 Best Sustainable Mechanism With Benevolent Time-Inconsistent Governments

The analysis so far was simplified by the fact that the government was purely self-interested. Although this case is of relevance for many political economy applications, it is also important to understand how the results generalize to the case considered by Roberts (1984), Freixas, Guesneries and Tirole (1985), or Bisin and Rampini (2005), where the government is still benevolent, but “time inconsistent”, i.e., unable to commit to a full dynamic mechanism. To do this, in this section we consider a more general utility function for the government of the
form:
\[ \mathbb{E}_t \sum_{s=0}^{\infty} \delta^s \left[ (1 - a) v(x_{t+s}) + a \left( \mathbb{E} \int (c_{t+s}, l_{t+s} \mid \theta^{t+s}) \, dG^{t+s} (\theta^{t+s}) \right) \right], \quad (44) \]
where the second term is the average (expected) utility of the citizens at time \( t + s \). Therefore, this utility function is identical to that of a purely-self-interested government when \( a = 0 \), and identical to the fully-benevolent case when \( \delta = \beta \) and \( a = 1 \).

In this case, we need to strengthened Assumption 1 to:

**Assumption 1’ (separability)** \( u(c, l \mid \theta) = u(c) - h(l \mid \theta) \), with \( u(\cdot) \) continuously differentiable, strictly increasing and concave, and \( h(\cdot \mid \theta) \) continuously differentiable, strictly increasing and convex for all \( \theta \in \Theta \).

Why this assumption is necessary is illustrated below in Example 1. The main results from our analysis above continue to hold under this additional assumption. In particular we have:

**Theorem 5** Suppose government utility is given by (44) and individual utility functions satisfy Assumptions 1’, 2, and 3. Then any combination of equilibrium strategy profiles \( \Gamma \) and \( \alpha \), there exists another pair of equilibrium strategy profiles \( \Gamma^* \) and \( \alpha^* = (\alpha^* \mid \alpha') \) for some \( \alpha' \) such that \( \Gamma^* \) induces direct submechanisms and \( (\alpha^* \mid \alpha') \) induces truth telling along the equilibrium path, and moreover \( c[\Gamma, \alpha] = c[\Gamma^*, \alpha^*] \), \( l[\Gamma, \alpha] = l[\Gamma^*, \alpha^*] \) and \( x[\Gamma, \alpha] = x[\Gamma^*, \alpha^*] \). Moreover, the best sustainable mechanism is a solution to maximizing (10) subject to (11), (12) and the government sustainability constraint:

\[
\mathbb{E}_t \sum_{s=0}^{\infty} \delta^s \left[ (1 - a) v(x_{t+s}) + a \left( \mathbb{E} \int (c(\theta^{t+s})) - h(l(\theta^{t+s}) \mid \theta_{t+s}) \right) dG^{t+s} (\theta^{t+s}) \right] 
\geq \max_{\hat{x}_t + \hat{c}_t \leq F(K_t, L_t)} (1 - a) v(\hat{x}_t) + au(\hat{c}_t). \quad (45)
\]

**Proof.** The proof of this theorem follows the structure of the proofs of Propositions 1 and 2, and Theorem 1. Similar to the proof of Proposition 1, we first need to show that there exists a sequentially rational continuation play in which all agents supply zero labor. Suppose that the government has announced a submechanism \( \tilde{M}_t \) at time \( t \) and has capital stock \( K_t \), and \( \alpha^i_{t+s} = \alpha^i_0 \) for all \( i \in [0, 1] \) and for all \( s \geq 0 \). We first show that a deviation by an individual, \( i' \) with type \( \theta_{i'} \neq \theta_0 \) to some other strategy that involves supplying positive labor is not profitable (as noted in footnote 10, we think of an individual with positive measure \( \varepsilon \) deviating, and take the limit \( \varepsilon \to 0 \), since there is a continuum of agents). Without the deviation, \( i' \) obtains utility \( (u(0) - h(0 \mid \theta_{i'})) / (1 - \beta) \). Now imagine a deviation to a message that corresponds
to positive labor supply, say \( l' \), with \( h(l' \mid \theta_i) > h(0 \mid \theta_i) \) by definition. This will generate output \( F(K_t, \varepsilon l') \), since all other agents are supplying zero labor. Now imagine the behavior of the government at the last stage of the game, conditional on \( \alpha_{t+s}^i = \alpha^\varnothing \) for all \( i \in [0,1] \) and for all \( s \geq 1 \). Then the sequentially rational strategy of the government is to maximize (44) with \( K_{t+1} = 0 \), since there will be no production in future periods. Consequently, the utility-maximizing program of the government in the information set following the deviation is

\[
\max_{x, c_t(\theta^t)} (1 - a) v(x_t) + a \left( \int [u(c^t_i(z^t(\alpha_t(\theta^t)))) + h(l_t(z^t(\alpha_t(\theta^t))) \mid \theta_t)] dG^t(\theta^t) \right)
\]

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

Now consider two different types of deviations by the government. First, imagine the government offers \( \tilde{M}_t \neq M_t \), i.e., a different mechanism at the beginning of time \( t \) than the one implicitly agreed in the social plan \((M, x)\). Given the above-constructed continuation equilibrium, \( \alpha_{t+s}^i = \alpha^\varnothing \) 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, as before, 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 play 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:
\[
\mathbb{E} \sum_{s=0}^{\infty} \delta^s \left[ (1 - a) v(x_{t+s}) + a \left( \int [u(c_t) + h(l_t \{ z^t(\alpha^t(\theta^t)) \} | \theta_t)] dG^t(\theta^t) \right) \right] \geq \max_{\hat{x}_t + \hat{c}_t \leq F(K_t, L_t)} (1 - a) v(\hat{x}_t) + au(\hat{c}_t).
\]

Now, given an equilibrium pair of strategy profiles \( \Gamma \) and \( \alpha \), exactly the same argument as in the proof of Theorem 1 implies that exists another pair of equilibrium strategy profiles \( \Gamma^* \) and \( \alpha^* = (\alpha^* | \alpha') \) for some \( \alpha' \) such that \( \Gamma^* \) induces direct submechanisms. Consequently, we can write (46), in terms of a direct mechanism, which gives (45).

Finally, exactly 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 as captured by (45).

The idea of this theorem is that exactly the same type of punishment strategies that were used in the case of the purely self-interested government also work when the government is benevolent. 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 (44), 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 the same punishment strategies as above, in which following any type of deviation, all individuals supply zero labor. To establish Theorem 5, all we need to show is that such punishment strategies are sequentially rational. This is straightforward to see given the separability assumption, Assumption 1’. When all other agents choose zero labor supply, following any deviation to positive labor supply, the government would consume some of the increase in output itself, and would redistribute the rest equally among all agents given the separable utility function assumed in Assumption 1’. Since there is a very large number of citizens, this implies the deviating individual will receive no additional consumption from supplying positive labor, and thus it is sequentially rational for all citizens to supply zero labor following a deviation by the government.

This theorem therefore shows that revelation principle applies to the case of benevolent, but time-inconsistent governments as well, though under the additional assumption of Assumption 1’. The next example shows why this assumption is necessary:
Example 1 To avoid issues of deviation among continuum of agents, let us consider a finite economy with \( N \) agents for this example, where \( N \) is large (exactly the same example can be constructed in an economy with a continuum of agents). There are two types of agents, \( \theta \in \{0, 1\} \), with \( \theta = 0 \) corresponding to the disabled type, who can only supply \( l = 0 \), and has utility \( u(c, \cdot | \theta = 0) = u(c) \), while the utility of type \( \theta = 1 \) is \( u(c, l | \theta = 0) = u(c - h(l)) \), where with \( h(\cdot) \) strictly increasing in \( l \). Furthermore, suppose that aggregate output is linear in labor. Also assume that the government is fully benevolent, i.e., \( a = 1 \) in terms of the utility function in (44). 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 \( h(l') < 1 \). The level of output generated by aggregate labor supply \( l' \) is \( l' > 0 \) by assumption. Following this deviation, the benevolent planner will distribute consumption to maximize its own utility, which involves maximizing average utility of the citizens, thus equating the marginal utility of consumption across agents, i.e.,

\[
u'(c_i) = u'(c_{i'} - h(l')) \text{ for all } i \neq i'
\]

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

\[
u\left(\frac{(l' - h(l'))}{N}\right) > u(0),
\]

thus giving him greater utility than supplying zero labor. This proves that the punishment phase where each citizen is supposed to supply zero labor cannot be part of a sequential equilibrium in this case.

8 Anonymous Markets versus Mechanisms

We have so far characterized the behavior of the best sustainable mechanism under political economy constraints. Although this was largely motivated by our objective of understanding the form of optimal policy in an environment with both informational problems on the side of agents and selfish behavior on the side of the government (or bureaucrats), an additional motivation is to investigate when certain activities should be regulated by (sustainable) mechanisms and when they should be organized in anonymous markets. In this section, we begin this analysis. Space restrictions preclude a detailed discussion of how anonymous markets
should be modeled, so we take the simplest conception of anonymous markets as one in which there is no intervention by the government, and consequently more limited insurance. For the purposes of the exercise in this section, we do not need to assume anything specific about how the anonymous markets work, except that there exists a well-defined anonymous market equilibrium, which yields utility ex ante $U^{AM}$ to individuals before they know anything about their types. The important point is that $U^{AM}$ is independent of both the discount factor of the government and any other institutional controls imposed on the government (since there is no government involvement in the anonymous markets).

Given this, we can provide some simple comparisons between anonymous markets versus sustainable mechanisms. Our first comparative static result states that an increase in the discount factor of the government, $\delta$, makes mechanisms more attractive relative to markets.

**Proposition 3** Suppose $U^{SM}(\delta) \geq U^{AM}$, then $U^{SM}(\delta') \geq U^{AM}$ for all $\delta' \geq \delta$. Moreover, as $\delta \to 0$, $U^{AM} > U^{SM}(\delta)$.

**Proof.** Let $S(\delta)$ be the feasible set of allocation rules when the government discount factor is equal to $\delta$ (meaning that they are feasible and also satisfy the sustainability constraint (13)). Let $\{c_{i,t}(\delta), l_{i,t}(\delta), x(\delta)\} \in S(\delta)$ represent the best sustainable mechanism. Since $\delta' \geq \delta$, we immediately have $\{c_{i,t}(\delta), l_{i,t}(\delta), x(\delta)\} \in S(\delta')$, since, when the government’s discount factor is $\delta'$, the left-hand side of (13) is higher, while the right-hand side is unchanged, so $\{c_{i,t}(\delta), l_{i,t}(\delta), x(\delta)\}$ satisfies (13). Therefore, $\{c_{i,t}(\delta), l_{i,t}(\delta), x(\delta)\}$ is feasible and yields expected utility $U^{SM}(\delta)$ when the government’s discount factor is $\delta'$. This implies that $U^{SM}(\delta')$ is at least as large as $U^{AM}$, therefore $U^{SM}(\delta') \geq U^{SM}(\delta) \geq U^{AM}$.

The second part follows from the observation that with anonymous markets, individuals can always achieve the autarchy allocation, thus $U^{AM} \geq \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c^a(\theta), l^a(\theta) \mid \theta)$, where $c^a$ and $l^a$ denote the optimal autarchy choices of an agent with type $\theta$. In contrast, with $\delta \to 0$, the centralized mechanism leads to utility of $\mathbb{E} \sum_{t=0}^{\infty} \beta^t u(0, 0 \mid \theta) < \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c^a(\theta), l^a(\theta) \mid \theta)$.

\[\blacksquare\]

Let us next consider a modification of our main setup along the lines mentioned in footnote 44.
Proposition 4 Suppose \( U^{SM}(\eta) \geq U^{AM} \), then \( U^{SM}(\eta') \geq U^{AM} \) for all \( \eta' \leq \eta \). Moreover, as \( \eta \to 0 \), \( U^{SM}(\eta) > U^{AM} \).

The proof of this proposition is similar to that of Proposition 3 and is omitted. It states the intuitive result that better institutional controls on government make mechanisms more desirable relative to markets. It also implies that with sufficiently good controls on centralized mechanisms (government behavior), sustainable mechanisms are preferred to anonymous markets.

9 Concluding Remarks

The optimal taxation literature pioneered by Mirrlees (1971) has generated a number of important insights about the optimal tax policy in the presence of insurance-incentive trade-offs. The recent optimal dynamic taxation literature has extended these insights to a macroeconomic setting where issues of dynamic behavior of taxes is of central importance. A potential criticism against all of this literature is that they consider the optimal tax scheme from the viewpoint of a benevolent government with full commitment power. A relevant and important question in this context is whether the insights of this literature apply to real world situations where politicians care about reelection, self-enrichment or their own individual biases, and cannot commit to sequences of future policies or to mechanisms.

This paper investigated this question and characterized the conditions under which these insights hold even when mechanisms are operated by self-interested politicians, who can misuse the resources and the information they collect. The potential misuse of resources and information by the government (politicians or bureaucrats) makes mechanisms less desirable relative to markets than in the standard mechanism design approach, and implies that certain allocations resulting from anonymous market transactions will not be achievable via centralized mechanisms. Nevertheless, centralized mechanisms may be preferable to anonymous markets because of the additional insurance they provide to risk-averse agents.
The main contribution of the paper is an analysis of the form of mechanisms to insure idiosyncratic (productivity) risks as in the classical Mirrlees setup in the presence of the self-interested government. Given the infinite horizon nature of the environment in question, we can construct sustainable mechanisms where the government is given incentives not to misuse resources and information. An important result of our analysis is the revelation principle along the equipment path, which shows that truth-telling mechanisms can be used despite the commitment problems and the different interests of the government and the citizens. Using this tool, we provide a characterization of the best sustainable mechanism.

The other major results of our analysis are as follows. First, under fairly general conditions, the best sustainable mechanism is a solution to a quasi-Mirrlees problem, defined as a problem in which the ex ante utility of (ex ante identical) agents is maximized subject to incentive compatibility, feasibility constraints as well as two additional constraints on the total amount of consumption and labor supply in the economy. Second, under additional conditions, we can characterize the initial and asymptotic distortions created by the best sustainable mechanism. In particular, when the government is sufficiently patient (in many cases as patient as, or more patients than, the citizens), we can show that the Lagrange multiplier on the sustainability constraint of the government goes to zero and aggregate distortions disappear asymptotically. Consequently, in the long run the highest type individuals will face zero marginal tax rate on their labor supply as in classical Mirrlees setup and there will be no aggregate capital taxes as in the classical dynamic taxation literature. These latter results therefore imply that some of the insights from Mirrlees’ classical analysis and from the dynamic taxation literature may follow despite the presence of political economy constraints and commitment problems. However, we also show that when the government is not sufficiently patient, aggregate distortions remain, even asymptotically. In this case, in contrast to many existing studies of optimal taxation, there will be positive distortions and positive aggregate capital taxes even in the long run.

In addition, under the further assumption that individual preferences are separable between consumption and leisure, we also generalized the main results to an environment with more general preferences for the government, in particular, including the fully-benevolent time-inconsistent government.

We view this paper as a first step in investigating political economy of mechanisms. There are both technical and substantial issues left unanswered. First, we would like to generalize
the results on time-inconsistent fully-benevolent government to non-separable utility functions. Secondly, it is important to undertake a more detailed comparison of centralized mechanisms subject to commitment problems and government misbehavior to more realistic models of anonymous markets. Finally and perhaps most importantly, our investigation introduces an interesting question: how should the society be structured so that the government (the mechanism designer) is easier to control. In other words, the recognition that governments need to be given the right incentives in designing mechanisms opens the way for the analysis of “mech-anism design squared”, where the structure of incentives and institutions for governments and individuals are simultaneously determined (for example, in the form of “constitutional design”). This becomes relevant in particular when we want to think about the interaction of different types of institutions in society, for example between contracting institutions that regulate the relationship between individual citizens versus “property rights institutions” that regulate the relationship between the state and individuals (Acemoglu and Johnson, 2005). While the existence of these distinct types of institutions have been recognized, how they should be simultaneously designed has not been investigated. We believe that the approach and tools in this paper will be useful to address this class of questions.


10 References


49


ilies, 51, 177-195.


Appendix (Still Incomplete)

11 Appendix A: Some Technical Results and Definitions

In this appendix, we provide a number of technical results and definitions that will be used in the rest of the appendix.

We take the definition of regular point from Luenberger (1969, p. 240).

**Definition 6** Let $X$ and $Z$ be Banach spaces and $G : X \to Z$ be a continuously (Fréchet) differentiable vector-valued mapping, with the derivative denoted by $G'$. Then $x_0$ is said to be a regular point of $G$ if $G'(x_0)$ maps $X$ onto $Z$.

**Example 2** If $G : \mathbb{R}^n \to \mathbb{R}^m$, $x_0$ is a regular point of $G$ if the Jacobian matrix of $G$ at $x_0$ has rank $m$.

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

$$P(u) = \max_{x \in X} f(x)$$  \hspace{1cm} (47)

subject to

$$g_0(x) \leq u$$  \hspace{1cm} (48)

and

$$G(x) \leq 0$$  \hspace{1cm} (49)

where $f : X \to \mathbb{R}$ and $g_0 : X \to \mathbb{R}$ are real-valued functions and $G : X \to Z$ is a vector-valued mapping and 0 is the zero of the vector space $Z$. $P(u)$ is the primal function denoting the optimal value of the problem. Suppose that $P$ is concave and let $\mu$ be any multiplier of (48). Then $\mu$ is a subgradient of $P(0)$.

This lemma follows, for example, from Proposition 6.5.8 of Bertsekas, Nedic and Ozdaglar (2003, p. 382). It immediately implies that if there is a unique multiplier, $P$ has a unique subgradient and is thus differentiable.

**Lemma 3** Let $X$ and $Z$ be Banach spaces. Consider the maximization problem of

$$\max_{x \in X} f(x)$$

53
subject to

\[ G(x) \leq 0 \]

where \( f : X \rightarrow \mathbb{R} \) is a real-valued function and and \( G : X \rightarrow Z \) is a vector-valued mapping and \( 0 \) is the zero of the vector space \( Z \). Suppose that \( x_0 \) is a solution to this program. Suppose also that \( f \) and \( G \) are continuously (Fréchet) differentiable in the neighborhood of \( x_0 \), and that \( x_0 \) is a regular point of \( G \). Then there exists a unique vector \( b^* \) such that

\[ f(x) - b^* G(x) \]

has a stationary point at \( x_0 \).

**Proof.** The existence of \( b^* \) follows from Theorem 1 of Luenberger (1969, p. 249), since all of the conditions of that theorem are satisfied. The stationarity of \( f(x) + b^* G(x) \) implies that

\[ \nabla f(x_0) = b^* \nabla G(x_0) . \]

The regularity of \( x_0 \) implies that \( b^* \) is uniquely defined. ■

**Example 3** In this special case where \( G : \mathbb{R}^n \rightarrow \mathbb{R}^m \) and \( f : \mathbb{R}^n \rightarrow \mathbb{R} \), this is particularly easy to see since the regularity condition implies that \( \nabla G(x_0) \) has rank equal to \( m \). In this case, there cannot exist \( b^* \neq \tilde{b} \) such that

\[ \nabla f(x_0) = b^* \nabla G(x_0) \]

\[ \nabla f(x_0) = \tilde{b} \nabla G(x_0) . \]

Now combining these lemmas, we obtain:

**Theorem 6** Let \( X \) and \( Z \) be Banach spaces. Consider the maximization problem of

\[ P(u) = \max_{x \in X} f(x) \]

subject to

\[ G(x) \leq 0 + u \]

where \( f : X \rightarrow \mathbb{R} \) is a real-valued function and and \( G : X \rightarrow Z \) is a vector-valued mapping and \( 0 \) is the zero of the vector space \( Z \) and \( u \) is a perturbation. Suppose that \( x_0 \) is a solution to this program. Suppose also that \( x_0 \) is a regular point of \( G \) and that \( f \) and \( G \) are continuously (Fréchet) differentiable in the neighborhood of \( x_0 \). Then \( P(0) \) is differentiable.
Next, we define consider a metric space $X$ and $M(X)$ to be the space of all measures defined on Borel sets of $X$. An element $\xi \in M(X)$ is no negative, comfortably additive, and has the property that $\xi(X) = 1$. Let $C(X)$ be the space of all bound the real-valued continuous functions on $X$. Then, following Parthasarathy (1967, p. 40) we have that:

**Definition 7** The weak topology on $M(X)$ is the topology with the following sets as basis:

$$S_\xi(f_1, \ldots, f_k; \varepsilon_1, \ldots, \varepsilon_k) = \left\{ \nu : \nu \in M(X), \left| \int f_i d\xi - f_i d\nu \right| < \varepsilon_i \text{ for } i = 1, \ldots, k \right\}.$$

The weak topology is also referred to as weak-* topology when $\xi$ and $\nu$ are not probability measures. The key result for us is Theorem 6.5 from Parthasarathy (1967, p. 45), which we state here is:

**Theorem 7** Let $X$ be a compact metric space, then $M(X)$ is a compact metric space with the weak topology.

Next, recall the Caratheodory’s Theorem, with $\text{conv}(X)$ denoting the convex hull of $X$ for some $X \in \mathbb{R}^N$:

**Theorem 8** (Caratheodory’s Theorem) Let $X \in \mathbb{R}^N$, then any $x \in \text{conv}(X)$ can be represented as a convex combination of vectors $x_1, \ldots, x_m$ from $X$ such that $x_2 - x_1, \ldots, x_m - x_1$ are linearly independent.

**Corollary 1** Let $X \in \mathbb{R}^N$, then any $x \in \text{conv}(X)$ may be represented by no more than $N + 1$ vectors of $X$.

**Proof.** Suppose not, this would violate linear independence of $x_2 - x_1, \ldots, x_m - x_1$ as stated in Theorem 8.

This theorem and its corollary imply that the convex hull of any subset of the $N$-dimensional Euclidean space can be achieved by $N + 1$ vectors, and will be useful in reducing the dimension of randomization below.
12 Appendix B: Proofs for Section 5

In this appendix, we provide and prove a number of results used in the analysis of Section 5, ultimate the building up to the proof of Theorem 3. Throughout this section we assume that Assumption 4 is satisfied.

12.1 Properties of the Function $U(C,L)$

Our first task is to establish a number of properties of $U(C,L)$. As mentioned in the text, to establish the concavity of $U(C,L)$, we need to introduce lotteries to convexify the constraint set, which we will do following Prescott and Townsend (1984a, 1984b). Recall that $U(C,L)$ is a solution to a finite-dimensional maximization problem. Moreover, using the single-crossing property (Assumption 2), we can reduce the static incentive compatibility constraints to only the constraints for the neighboring types, thus to $N$ constraints (there is no incentive compatibility constraint for the lowest type, $\theta_0$). In addition, there are the resource constraints on the sum of consumption and labor supply levels. Recall also that only $(C,L) \in \Lambda$ will enable this maximization program to be well defined by making the constraint set non-empty.

Let $C = \{(c,l) \in \mathbb{R}^2 : 0 \leq c \leq \bar{c}, 0 \leq l \leq \bar{l}\}$ be the set of possible consumption-labor allocations for agents. Let $P$ be the space of $N+1$-tuples of probability measures on Borel subsets of $C$. Thus each element $\zeta = [\zeta(\theta_0), ..., \zeta(\theta_N)]$ in $P$ consists of $N+1$ probability measures for each type $\theta_i \in \Theta$.

Then the quasi-Mirrlees problem can be defined in the following way

$$ U(C,L) \equiv \max_{\zeta \in P} \left\{ \sum_{\theta \in \Theta} \pi(\theta) \int u(c,l;\theta)\zeta(d(c,l),\theta) \right\} \tag{50} $$

subject to

$$ \int u(c,l \mid \theta_i)\zeta(d(c,l),\theta_i) \geq \int u(c,l \mid \theta)\zeta(d(c,l),\theta_{i-1}) \text{ for all } i = 1, ..., N \tag{51} $$

$$ \sum_{\theta \in \Theta} \pi(\theta) \int c\zeta(d(c,l),\theta) \leq C \tag{52} $$

$$ \sum_{\theta \in \Theta} \pi(\theta) \int l\zeta(d(c,l),\theta) \geq L \tag{53} $$

for some $(C,L) \in \Lambda$.

Before deriving properties of the function $U(C,L)$, we need to ensure regularity. Let (51), (52) and (53) define the constraint mapping.
Lemma 4 The solution to (50) is a regular point of the constraint mapping.

Proof. The proof follows immediately from the fact that from single-crossing property (Assumption 2), all incentive compatibility constraints in (51) are linearly independent from each other, and also linearly independent from (52) and (53), thus the constraint mapping has full rank, \( N + 2 \), and is thus onto. 

Our main result on the function \( U(C, L) \) is:

Lemma 5 \( U(C, L) \) is well-defined, continuous and concave on \( \Lambda \), nondecreasing in \( C \) and nonincreasing in \( L \) and differentiable in \((C, L)\).

Proof. First, we show that \( U(C, L) \) is well-defined, i.e., a solution exists. For this, endow the set of probability measures \( \mathcal{P} \) with the weak topology. Since \( \mathcal{C} \) is a compact subset of \( \mathbb{R}^2 \), Theorem 7 above establishes that \( \mathcal{P} \) is compact in the weak topology, and the constraint set is compact in the weak topology as well. Moreover, the objective function is continuous in any \( \zeta \in \mathcal{P} \), thus establishing existence.

Next, to show that \( U(C, L) \) is continuous, note that with the lotteries the constraint set is convex. Then from Berge’s Maximum Theorem, \( U(C, L) \) is continuous in \((C, L)\).

Concavity then follows from the convexity of the constraint set and the fact that the objective function is concave in \( \zeta \in \mathcal{P} \).

\( U(C, L) \) is also clearly nondecreasing in \( C \), since a higher \( C \) relaxes constraint (52) and nonincreasing in \( L \), since a higher \( L \) tightens constraint (53).

Finally, to prove differentiability, note that the regularity condition is satisfied from Lemma 4 and moreover, the objective function in (50) is continuously differentiable at all points of the constraint set. We can therefore apply Theorem 6, establishing that \( U(C, L) \) is differentiable in \((C, L)\). This completes the proof of the lemma.

The necessary properties of the set \( \Lambda \) are derived in the next lemma.

Lemma 6 \( \Lambda \) is compact and convex.

Proof. Convexity: Consider \((C^0, L^0), (C^1, L^1) \in \Lambda \) and some \( \zeta^0, \zeta^1 \) feasible for \((C^0, L^0)\) and \((C^1, L^1)\) respectively. For any \( \alpha \in (0, 1) \) \( \zeta^\alpha \equiv \alpha \zeta^0 + (1 - \alpha) \zeta^1 \) is feasible for \((\alpha C^0 + (1 - \alpha) C^1, \alpha L^0 + (1 - \alpha) L^1)\), so that this set is non-empty. Moreover, since \( \zeta^0, \zeta^1 \) satisfy (51), (52) and (53), \( \zeta^\alpha \) satisfies all three of these constraints, establishing convexity.
Compactness: Λ is clearly bounded, so we only have to show that it is closed. Take a sequence \((C^n, L^n) \in \Lambda\). Since this sequence is in a bounded set, it has a convergent subsequence, \((C^n, L^n) \to (C^\infty, L^\infty)\). We just need to show that \((C^\infty, L^\infty) \in \Lambda\). Let \(\zeta^n\) be a feasible element for \((C^n, L^n)\), and since \(P\) is compact under the weak topology, \(\zeta^n \to \zeta^\infty \in P\), which implies that \(\zeta^\infty\) satisfies (51)-(53) and so \(\zeta^\infty\) is feasible for \((C^\infty, L^\infty)\), therefore \(\Lambda\) is closed.

Now define a promised utility for the government for some sequence \(x\) as

\[
w = \sum_{t=0}^{\infty} \delta^t v(x_t)
\]

Then the set of feasible promised utilities \(\mathbb{W}\) is defined as

\[
\mathbb{W} = \{ w : \exists x \in \mathbb{R}^\infty \text{ s.t. for any } t \text{ there is some } L \text{ s.t. } (L - x_t, L) \in \Lambda, \ w = \sum_{t=0}^{\infty} \delta^t v(x_t) \}
\]

**Lemma 7** \(\mathbb{W} = [0, \bar{w}]\).

**Proof.** Since \(v(0) = 0\), it is clear that 0 is the minimal element. By definition \(\bar{w}\) is the maximal element. Moreover, clearly any \(w \leq \bar{w}\) is also achievable, so \(\mathbb{W}\) must take the form \([0, \bar{w}]\). ■

To further analyze the best sustainable mechanism, it is useful to re-write the maximization problem in the following recursive way as in equations (24)-(27) in the text. The constraint set in this problem is not convex, and randomization may further improve the value of the program. So analogously to the quasi-Mirrlees problem, we now consider randomizations. Now let \(z = (C, L, x, w') \in \mathbb{R}^4, C(w) = \{ z : (24)-(27) \text{ are satisfied for given } w \}\), and let \(Z\) be the set of Borel subsets of \(C(w)\) Then let the triple \((C(w), Z, \mu)\) be a probability space. Let \(P(w)\) be the space of probability measures on \(C(w)\) endowed with the weak topology. Incorporating randomization, we can write the recursive formulation as:

**Problem A1**

\[
V(w) = \max_{\xi \in P(w)} \int [U(C, L) + \beta V(w'(\omega))] \xi(\text{d}z) \tag{54}
\]

subject to

\[
C + x \leq L \xi\text{-almost-surely} \tag{55}
\]

\[
v(x) + \delta w' \geq v(L) \xi\text{-almost-surely} \tag{56}
\]

\[
w = \int [v(x) + \delta w'] \xi(\text{d}z) \tag{57}
\]
and
\[(C, L) \in \Lambda \text{ and } w' \in \mathbb{W} \xi\text{-almost-surely.}\]  
\[(58)\]

We can now establish:

**Lemma 8** \(V(w)\) is concave.

**Proof.** Consider any \(w_0\) and \(w_1\) and \(\xi_0\) and \(\xi_1\) that are the solution to the maximization problem. Consider \(w = \alpha w_0 + (1 - \alpha)w_1\) for some \(\alpha \in (0, 1)\). Let \(\xi_\alpha = \alpha \xi_0 + (1 - \alpha)\xi_1\). Constraints (55) and (56) hold state by state, and satisfied for both \(\xi_0\) and \(\xi_1\), and therefore must be satisfied for \(\xi_\alpha\). Constraint (57) is linear in \(\xi\), therefore \(\xi_\alpha\) also satisfies this constraint. Since the objective function is linear in \(\xi\), \(V(\alpha w_0 + (1 - \alpha)w_1) \geq \alpha V(w_0) + (1 - \alpha)V(w_1)\), establishing the concavity of \(V\).

The above lemma establishes the concavity of \(V\) using arbitrary randomizations in the maximization problem (54). The next lemma shows that a particularly simple form of randomization is sufficient to achieve the maximum of (54).

**Lemma 9** There exists \(\xi \in \mathcal{P}(w)\) achieving the value \(V(w)\) with randomization between at most two points, \((C_0, L_0, x_0, w_0')\) and \((C_1, L_1, x_1, w_1')\) with probabilities \(\xi_0\) and \(1 - \xi_0\).

**Proof.** To achieve convexity, we only need to constraint set to be convex. The constraint set here is \(C(w) \in \mathbb{R}^4\). Recall Theorem 8 and its corollary, which imply that the convex hull of \(C(w)\) can be achieved with 5 points.

Suppose, to obtain a contradiction, that there are more than two points with positive probability. We consider a case of three points, since the same argument applies to any finite number of points. Suppose that randomization occurs between \((C_0, L_0, x_0, w_0')\), \((C_1, L_1, x_1, w_1')\) and \((C_2, L_2, x_2, w_2')\) with probabilities \(\xi_0, \xi_1, \xi_2 > 0\). Suppose without loss of generality that \(v(x_0) + \delta w_0' \leq v(x_2) + \delta w_2' \leq v(x_1) + \delta w_1'\) and let \(\alpha \in [0, 1]\) be such that \(v(x_2) + \delta w_2' = \alpha(v(x_0) + \delta w_0') + (1 - \alpha)(v(x_1) + \delta w_1')\). Suppose first
\[U(C_2, L_2) + \beta V(w_2') > \alpha[U(C_0, L_0) + \beta V(w_0')] + (1 - \alpha)[U(C_1, L_1) + \beta V(w_1')].\]
Then an alternative element \(\hat{\xi} \in \mathcal{P}(w)\) assigning probability \(\hat{\xi}_2 = 1\) to \((C_2, L_2, x_2, w_2')\) is feasible and yields higher utility than the original randomization, yielding a contradiction.
Next suppose that

\[ U(C_2, L_2) + \beta V(w'_2) < \alpha[U(C_0, L_0) + \beta V(w'_0)] + (1 - \alpha)[U(C_1, L_1) + \beta V(w'_1)]. \]

Now consider an alternative \( \hat{\xi} \in \mathcal{P}(w) \) assigning probability \( \xi_0 + \alpha \xi_2 \) to \((C_0, L_0, x_0, w'_0)\) and probability \( \xi_1 + (1 - \alpha)\xi_2 \) to \((C_1, L_1, x_1, w'_1)\), which is again feasible and gives a higher utility than original randomization, once again yielding a contradiction. Therefore, \( \xi \) must satisfy

\[ U(C_2, L_2) + \beta V(w'_2) = [U(C_0, L_0) + \beta V(w'_0)] + (1 - \alpha)[U(C_1, L_1) + \beta V(w'_1)]. \]

But then the optimum can be achieved by simply randomizing between \((C_0, L_0, x_0, w'_0)\) and \((C_1, L_1, x_1, w'_1)\) with probabilities \( \xi_0 + \alpha \xi_2 \) to \((C_0, L_0, x_0, w'_0)\) and probability \( \xi_1 + (1 - \alpha)\xi_2 \).

The same argument applies whenever all different points receive positive probability. ■

Lemma 9 implies that for the rest of this section, we can focus on randomizations between two points. We denote solution for any \( w \) by \( C_i(w), L_i(w), x_i(w), w'_i(w), \xi_i(w) \) for \( i \in \{0, 1\} \).

Now from Lemma 9, we can focus on the problem equivalent to Problem A1:

**Problem A2:**

\[
V(w) = \max_{\{\xi_i, C_i, L_i, x_i, w'_i\}} \sum_{i=0,1} \xi_i [U(C_i, L_i) + \beta V(w'_i)]
\]  

subject to

\[
C_i + x_i \leq L_i \quad \text{for} \ i = 0, 1
\]

\[
v(x_i) + \delta w'_i \geq v(L_i) \quad \text{for} \ i = 0, 1
\]

\[
w = \sum_{i=0,1} \xi_i [v(x_i) + \delta w'_i].
\]

\[
(C_i, L_i) \in \Lambda \quad \text{for} \ i = 0, 1 \quad \text{and} \ w' \in \mathbb{W}
\]

Next we would like to establish that \( V(w) \) is differentiable. This does not follow from Theorem 6, since \( V(w) \) includes the term \( V(w'_i) \), which may not be differentiable. Instead, we can apply an argument similar to that of Benveniste and Scheinkman (1979) to prove differentiability (see also Stokey, Lucas and Prescott, 1989).

**Lemma 10** \( V(w) \) is differentiable.
Proof. Let the maximizer for \( w' (\omega \mid w_0) \) in Problem A1 when \( w = w_0 \) be \( w_0 (\omega \mid w_0) \). Then consider the alternative maximization problem

\[
W(w) = \max_{\xi \in \mathcal{P}(w)} \int [U(C, L) + \beta V(w (\omega \mid w_0))] d\xi (\omega)
\]
subject to (55), (56) and

\[
w = \int [v(x (\omega)) + \delta w' (\omega \mid w_0)] d\xi (\omega).
\]

By the same argument as in Lemma 8, \( W(w) \) is concave (recall that the proof of Lemma 8 is for a given \( w' \), so the fact that we have \( w' (\omega \mid w_0) \) fixed, does not affect the proof). Next the same arguments as in Lemma 9 establishes that \( W(w) \) can be equivalently characterized by the following maximization problem:

**Problem A3:**

\[
W(w) = \max_{\{\xi, C_i, L_i, x_i\}_{i=0,1}} \sum_{i=0,1} \xi_i [U(C_i, L_i) + \beta V(w'_i (w_0))]
\]

\[
C_i + x_i \leq L_i \text{ for } i = 0, 1
\]

\[
v(x_i) + \delta w'_i (w_0) \geq v(L_i) \text{ for } i = 0, 1
\]

\[
w = \sum_{i=0,1} \xi_i [v(x_i) + \delta w'_i (w_0)].
\]

\[
(C_i, L_i) \in \Lambda \text{ for } i = 0, 1 \text{ and } w' \in \mathcal{W}
\]

where \( w'_i (w_0) \) for \( i = 0, 1 \) are the optimal choices for \( w_0 \).

Since \( W(w) \) is concave, Theorem 6 implies that it is also differentiable (\( \beta V(w'_i (w_0)) \) is just a constant here, so does not affect anything, and everything else is differentiable). Moreover, we have

\[
W(w) \leq V(w)
\]

and

\[
W(w_0) = V(w_0)
\]

by construction.

From Lemma 8 \( V(w_0) \) is concave, and therefore \(-V\) is convex. Convex functions have well-defined subdifferentials (see Rockafellar, 1970, Chapter 23 or Bertsekas, Nedic and Ozdaglar,
In particular, if \( f \) is convex, there exists a closed, convex and nonempty set \( \partial f \) such that for all \( v \in \partial f \) and any \( x \) and \( x' \), we have \( f(x') - f(x) \geq v(x' - x) \). Let \(-\partial V(w)\) be the set of subdifferentials of \(-V\), i.e., all \(-\nu\) such that \(-V(\hat{w}) + V(w) \geq -\nu(\hat{w} - w)\). By definition, \(-\partial V(w)\) is a closed, convex and nonempty set. Let \( \partial V(w) = \{v : -v \in -\partial V(w)\} \).

Clearly \( \partial V(w) \) is also closed, convex and nonempty. Consequently, for any subgradient \( v \) of \( \partial V(w_0) \), we have

\[
v(w - w_0) \geq V(w) - V(w_0) \geq W(w) - W(w_0),
\]

where the first inequality is by the definition of a subgradient, and the second follows from (69) and (70). This implies that \( v \) is also a subgradient of \( W(w_0) \). But since \( W(w_0) \) is differentiable, \( v \) must be unique, therefore \( V(w_0) \) is also differentiable.

Finally, before providing a proof of Theorem 3, we need a characterization of how the society would provide the maximum (steady-state) utility \( \bar{w} \) to the government. This is stated and proved in the next lemma.

**Lemma 11** Suppose \((\hat{C}, \hat{L}) \in \Lambda\) maximizes \( \bar{w} = v(L - C) / (1 - \delta) \) and Assumption 4 holds, then

\[
U_C(\hat{C}, \hat{L}) + U_L(\hat{C}, \hat{L}) = 0.
\]

**Proof.** Recall from the text that \( \bar{w} \) is a solution to the maximization problem

\[
\bar{w} \equiv \max_{(C,L) \in \Lambda} v(L - C) / (1 - \delta),
\]

and Assumption 4 ensures that at the solution \((\hat{C}, \hat{L})\), we have \( \hat{v}(\hat{C}, \hat{L}) > 0 \) (as defined in Assumption 4) which implies that \( \psi = 0 \). Moreover, the constraint that \( (C,L) \in \Lambda \) implies that we have to satisfy the \( N \) incentive compatibility constraints in addition to the resource constraints (15) and (16). Consider the first-order condition for the consumption and the labor supply of the highest type \( \theta_N \), \( c(\theta_N) \) and \( l(\theta_N) \), which are

\[
\begin{align*}
u_C(c(\theta_N), l(\theta_N) | \theta_N)(1 + \lambda_N) & = \nu_C \\
u_L(c(\theta_N), l(\theta_N) | \theta_N)(1 + \lambda_N) & = -\nu_L
\end{align*}
\]

where \( \lambda_N \) is the multiplier on incentive compatibility constraint between types \( \theta_N \) and \( \theta_{N-1} \), \( \nu_C \) is the multiplier on (15) and \( \nu_L \) is the multiplier on (16). This equation implies that these
multipliers, $\nu_C$ and $\nu_L$, and thus the derivatives $U_C(\dot{C}, \dot{L})$ and $U_L(\dot{C}, \dot{L})$ are well-defined at the solution $(\dot{C}, \dot{L})$. Next, since $\psi = 0$, the first-order conditions with respect to $C$ and $L$ yield $\nu_C = \nu_L$, therefore

$$U_C(\dot{C}, \dot{L}) + U_L(\dot{C}, \dot{L}) = 0,$$

as desired.

### 12.2 Proof of Theorem 3

**Proof.** Since $V$ is differentiable from Lemma 5 and concave from Lemma 8, the first-order conditions are necessary and sufficient for the maximization (59). Assign the multipliers $\xi_i \kappa_i$ to the constraints in (60), $\xi_i \psi_i$ to those in (61) and $\gamma$ to constraint (62), and let $V'(w)$ be the derivative of $V(w)$ at $w$, we have

$$\beta \xi_0 V'(w_0') + \delta \psi_0 \xi_0 + \delta \gamma \xi_0 \leq 0$$

$$\beta \xi_1 V'(w_1') + \delta \psi_1 \xi_1 + \delta \gamma \xi_1 \leq 0$$

with both equations holding as equality for $w'_i \in \text{Int} W$. Therefore,

$$\frac{\beta}{\delta} V'(w_i') \leq -\psi_i - \gamma, \quad (71)$$

again with equality for $w'_i \in \text{Int} W$. Moreover, since $V$ is differentiable, we have

$$V'(w) \geq -\gamma \quad (72)$$

again with equality for $w \in \text{Int} W$.

In addition, combining first-order conditions we have that for $(C, L) \in \text{Int} \Lambda$,

$$U_C(C_i, L_i) + U_L(C_i, L_i) = \psi_i v'(L_i) \quad \text{for } i = 0, 1. \quad (73)$$

**Part 1:** To establish this part of the theorem, it suffices to show that $\psi_i > 0$ at $t = 0$ for $i = 0$ or 1. First note that the initial value $w_0$ maximizes $V(w)$, and since $V(\cdot)$ is differentiable, this implies $V'(w_0) = \gamma = 0$ at $t = 0$. Suppose, to obtain a contradiction, that $\psi_i = 0$ at $t = 0$ for both $i = 0$ and 1. This implies from (71) that $\beta V'(w'_i) / \delta = 0$, so that $w'_i = w_0$. Repeating this argument yields $w_t = w_0$ for all $t$, and (61) never binds. This is in turn only consistent with $x_t = 0$ for all $t$, which then implies $C_t = L_t = 0$ for all $t$, which cannot be optimal (given Assumption 4), yielding a contradiction. Therefore $\psi_i > 0$ for $i = 0$ or 1 at
time \( t = 0 \), so initial \((C, L)\) cannot be undistorted and there is a positive marginal tax rate on even the highest type.

**Part 2:** Let \( W_t = \{w_{0,t}, w_{1,t}\} \). Since \( \beta \leq \delta \) and \( V'(w) \leq 0 \), (71) implies

\[
V'(w'_t) \leq -\psi_i - \gamma.
\]

Combining this with (72) and \( \psi_i \geq 0 \) yields:

\[
V'(w) \geq V'(w').
\]

Concavity of \( V \) then implies that \( w'_t \geq w \) for \( i = 0, 1 \), establishing the claim that the sequence \( \{w_t\}_{t=0}^{\infty} \) is nondecreasing. Since each \( w_t \) is in the compact set \([0, \bar{w}]\), the stochastic sequence \( \{w_t\}_{t=0}^{\infty} \) must converge almost surely to some point \( w^* \), meaning \( \operatorname{plim} w_t = w^* \). This immediately implies \( \operatorname{plim} x_t = x^* \). Therefore \( \operatorname{plim} C_t = C^* \) and \( \operatorname{plim} L_t = L^* \) are feasible (given \( \operatorname{plim} x_t = x^* \)) and are optimal from the concavity of \( U(C, L) \), this is a solution to the maximization in (59), establishing the existence of a steady state as claimed in the theorem.

Recall from the above argument that \( \{w_t\} \uparrow w^* \) almost surely. First suppose that \( w^* = \bar{w} \). In this case, Lemma 11 applies and implies that the sustainability constraint, (61), is still slack and (74) applies as desired.

Next consider the case in which \( \beta < \delta \). Now we have

\[
\frac{\beta}{\delta} V'(w_{i,t+1}) + \psi_i \leq V'(w_t) \quad \text{for} \quad i = 0, 1.
\]

Recall that, as established above, \( \{w_{i,t}\} \uparrow w^* \). To derive a contradiction, suppose that \( w^* < \bar{w} \). This implies that \( \frac{\beta}{\delta} V'(w^*) + \psi^* \leq V'(w^*) \) for some \( \psi^* \geq 0 \), which is impossible in view of the
fact that $\beta < \delta$ and $V'(w^*) \leq 0$ (unless $w_t = w_0$ for all $t$ so that $V'(w^*) = 0$, which is ruled out by the argument in part 1). Therefore, we must have $\{w_{i,t}\} \uparrow \bar{w}$. Therefore Lemma 11 applies and yields (74) as desired.

The rest of part 2 follows from Lemma 1.

**Part 3:** Suppose that $\beta > \delta$. Then, $\{w_t\}$ is no longer nondecreasing. If $\{w_t\}$ converges to some $w^*$, then equation (31) in the text must hold and $\lim_{t \to \infty} -U_C/U_L$ exists and is strictly greater than 1 as claimed in the theorem. Next, suppose that $\{w_t\}$ does not converge. Since it lies in a compact set, it has a convergent subsequence. Suppose that for all such convergent subsequences $\psi_i = 0$ for $i = 0, 1$, this would imply convergence to a steady state since we would have $\psi_{i,t} = 0$ for $i = 0, 1$ and for all $t$, yielding a contradiction. Therefore, there must exist a convergent subsequence with $\psi_i > 0$, so that $\limsup -U_C/U_L > 1$. Consequently, distortions do not disappear asymptotically, completing the proof. ■
13 Appendix C: Proofs of Section 6

In this appendix, we provide the proofs for the more general environment.

13.1 Properties of $\mathcal{U} \left( \{C_t, L_t\}_{t=0}^\infty \right)$

As in the above proof, to show concavity and differentiability of $\mathcal{U} \left( \{C_t, L_t\}_{t=0}^\infty \right)$, we introduce randomizations. The original maximization problem without randomization is to maximize

$$\text{(10)}$$ subject to (11), (12), and (13) as stated in Proposition 2. Recall also that $\theta_t \in \Theta$, where $\Theta$ is a finite set (with $N+1$ elements). Therefore $\Theta_t$ for any $t < \infty$ is also a finite set.

Consider next the functions $c_t : \Theta_t \to \mathbb{R}_+$ and $l_t : \Theta_t \to [0, \bar{l}]$. By definition, these functions assign values to a finite number of points in the set $\Theta_t$ for any $t < \infty$, thus can simply be thought of as vectors of $(N+1)^t$ dimension. Moreover

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

(75)

where $\bar{Y} = F (\bar{Y}, \bar{l}) < \infty$. Therefore, $X_t = \{c_t (\theta^t), l_t (\theta^t), K_{t+1}, x_t\}$ is a vector (of dimension $(N+1)^t + 2$). Let $\mathbf{X}_t$ be the set of all such vectors that satisfy the inequalities in (75), and for $X_t \in \mathbf{X}_t$, let $X_t (i)$ denote the $i$th component of this vector, and $T_t$ be the dimension of vectors in the set $\mathbf{X}_t$ (i.e., $T_t = (N+1)^t + 2$). $\mathbf{X}_t$ is a compact metric space space with the usual Euclidean distance metric, $d_t (X_t, X_t^\prime) = \left( \sum_{i=1}^{T_t} (X_t^\prime (i) - X_t (i))^2 \right)^{1/2}$

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

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

Clearly the sequence $\{c_t (\theta^t), l_t (\theta^t), x_t, K_{t+1}\}_{t=0}^\infty$ must belong to $\mathbf{X}$. In fact, it must belong to the subset of $\mathbf{X}$, which satisfy (11), (12), and (13), $\mathbf{X}$. $\mathbf{X}$ is compact in the product topology. Since (11), (12), and (13) are (weak) inequalities, $\mathbf{X}$ is a closed subset of $\mathbf{X}$, and therefore it is also compact in the product topology. Moreover, $\mathbf{X}$ with the product topology is metrizable, with the metric

$$d (X, X^\prime) = \sum_{t=1}^\infty \omega^t d_t (X_t, X_t^\prime)$$

(76)

for some $\omega \in (0, 1)$ and $X = \{X_t\}_{t=0}^\infty \in \mathbf{X}$. This shows that $\mathbf{X}$ endowed with the product topology is a metric space, and so is $\mathbf{X}$. From Theorem 7, the set of probability measures
defined over a compact metric space is compact in the weak topology. This establishes that the set of probability measures $\mathcal{P}^\infty$ defined over $\bar{X}$ is compact in the weak topology.

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

Finally, choosing $\omega \leq \beta$ in (76) shows that the objective function is continuous in the weak topology. This establishes that including randomizations, we have a maximization problem over probability measures in which the objective function is continuous in the weak topology, and the constraint set is compact in the weak topology. From the Weierstrass maximum theorem for the general (possibly infinite-dimensional) normed linear spaces (e.g., Luenberger, 1969, Theorem 1, p. 40), this establishes the existence of a probability measure that reaches the maximum.

Given this result, the rest of the analysis parallels that of Theorem 3. The key lemma is a generalization of Lemma 5, which is stated here.

**Lemma 12** $U(\{C_t, L_t\}_{t=0}^\infty)$ is continuous and concave on $\Lambda^\infty$, nondecreasing in $C_s$ and non-increasing in $L_s$ for any $s$ and differentiable in $\{C_t, L_t\}_{t=0}^\infty$.

This lemma can be proved along the lines of Lemma 5, except that in this infinite-dimensional space we are no longer able to prove Lemma 4. Thus, all the proofs assume that the solution is at a regular point.

**Proof.** The above argument established that in the problem of maximizing (10) subject to (11), (12), and (13) over probability measures, a maximum exists and $U(\{C_t, L_t\}_{t=0}^\infty)$ is therefore well defined.
To show concavity, consider \((C^0, L^0)\) and \((C^1, L^1)\) and corresponding \(\zeta^0, \zeta^1\). We have

\[
\int (u(c, l; \theta) - u(c, l; \hat{\theta})) \zeta^0(d(c, l), \theta) \alpha \int (u(c, l; \theta) - u(c, l; \hat{\theta})) \zeta^0(d(c, l), \theta) + (1 - \alpha) \int (u(c, l; \theta) - u(c, l; \hat{\theta})) \zeta^1(d(c, l), \theta) \geq 0
\]

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

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

Finally, Theorem 6 applies to this problem and implies that \(U(\{C_t, L_t\}_{t=0}^\infty)\) is differentiable in \(\{C_t, L_t\}_{t=0}^\infty\), completing the proof. ■

**Lemma 13** \(\Lambda\) is compact and convex.

**Proof.** **Convexity:** Consider \(\{C_t, L_t\}_{t=0}^\infty\) and \(\{C'_t, L'_t\}_{t=0}^\infty\) in \(\Lambda^\infty\) and some \(\zeta^0, \zeta^1\) feasible for \(\{C_t, L_t\}_{t=0}^\infty\) and \(\{C'_t, L'_t\}_{t=0}^\infty\) respectively. Now for any \(\alpha \in (0, 1)\) \(\zeta^\alpha \equiv \alpha \zeta^0 + (1 - \alpha)\zeta^1\) is feasible for \(\alpha \{C_t, L_t\}_{t=0}^\infty + (1 - \alpha) \{C'_t, L'_t\}_{t=0}^\infty\), so that this set is non-empty. Moreover, since \(\zeta^0, \zeta^1\) satisfy (51), \(\zeta^\alpha\) satisfies it as well. Similarly, \(\zeta^\alpha\) satisfies (52) and (53).

**Compactness:** For any sequence \(\{C^n_t, L^n_t\}_{t=0}^\infty\), \(\{C^n_t, L^n_t\}_{t=0}^\infty \rightarrow \{C^\infty_t, L^\infty_t\}_{t=0}^\infty\), there will be \(\zeta^n\) of elements in \(\{C^n_t, L^n_t\}_{t=0}^\infty\), such that \(\zeta^n \rightarrow \zeta^\infty\), satisfying (51)-(53) and feasibility, therefore \(\{C^\infty_t, L^\infty_t\}_{t=0}^\infty \in \Lambda^\infty\) is closed. Boundedness follows from boundedness of \(C\) and \(L\). ■

Finally, we have a generalization of Lemma 11:

**Lemma 14** Suppose \(\hat{C}, \hat{L}\) in \(\Lambda^\infty\) and \(\hat{K}\) maximize \(\tilde{w} = v \left(\hat{F}(\hat{K}, \hat{L}) - \hat{C}\right) / (1 - \delta)\) and Assumption 5 holds, then

\[
\mathcal{U}_{C_t} \left(\hat{C}, \hat{L}, \hat{K}\right) \cdot F_{L_t} \left(\hat{K}, \hat{L}\right) = -\mathcal{U}_{L_t} \left(\hat{C}, \hat{L}, \hat{K}\right) \text{ and } F_{K_{t+1}} \left(\hat{K}, \hat{L}\right) \mathcal{U}_{C_{t+1}} \left(\hat{C}, \hat{L}, \hat{K}\right) = \mathcal{U}_{C_t} \left(\hat{C}, \hat{L}, \hat{K}\right).
\]

**Proof.** To be written. ■

**13.2 Proof of Theorem 4**

The proof of Theorem 4 is similar to that of Theorem 3, except that we do not use the recursive formulation and randomizations. Instead, we work directly with the sequence problem and the necessary conditions of the sequence problem.
Proof. Part 1: An argument analogous to that of part 1 of Theorem 3 establishes this result.

Part 2: Take \( \{C_t, L_t\}_{t=0}^\infty \) to be part of the optimal mechanism. Suppose that \( \{C_t, L_t, K_t\}_{t=0}^\infty \) (almost surely) converges to a limit, \((C^*, L^*, K^*)\), and let \( x^* = L^* - C^* - K^* \).

First, suppose that \( \{C_t, L_t\}_{t=0}^\infty \in \text{Int} \Lambda^\infty \), then (36) and (37) are necessary conditions for optimality. Rearranging these equations and substituting for \( U_{C_t}(C^*, L^*) \), we have

\[
-\frac{U_{C_t}^*}{U_{C_t}^*F_{L_t}(K^*, L^*)} = 1 - \frac{(\mu_t - \mu_{t-1})v'(F(K^*, L^*)))}{\mu_tv'(x^*)} \tag{77}
\]

and

\[
\frac{F_{K_{t+1}}(K^*, L^*)U_{C_{t+1}}^*}{U_{C_t}^*} = 1 + \frac{(\mu_{t+1} - \mu_t)v'(F(K^*, L^*))F_{K_{t+1}}(K^*, L^*)}{\mu_tv'(x^*)}, \tag{78}
\]

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

Since \( \{C_t, L_t\}_{t=0}^\infty \in \text{Int} \Lambda^\infty \), equation (40) also holds and implies that as \( t \to \infty \),

\[
\frac{U_{C_t}^*}{\delta^tv'(x^*)} = \mu_t \leq \mu_{t+1} = \frac{U_{C_{t+1}}^*}{\delta^{t+1}v'(x^*)}. \tag{79}
\]

First suppose that \( \varphi = \delta < 1 \) where \( \varphi = \sup \{ \varepsilon \in [0, 1] : \lim_{t \to \infty} \varepsilon^{-t}U_{C_t}^* = 0 \} \) as defined in the theorem. This implies that as \( t \to \infty \), \( U_{C_t}^* \) is proportional to \( \varphi^t \). Therefore, \( \mu_t \) must converge (almost surely) to \( \mu^* < \infty \), thus \( (\mu_t - \mu_{t-1})/\mu_t \to 0 \) almost surely, and from (77) and (78), we have that \( -U_{L_t}/U_{C_t}F_{L_t} \) and \( F_{K_{t+1}}U_{C_{t+1}}/U_{C_t} \) almost surely converge to 1, thus \( \{C_t, L_t, K_t\}_{t=0}^\infty \) must be asymptotically undistorted. The rest of the argument parallels the rest of the proof of part 2 of Theorem 3.

Next, suppose that \( \{C_t, L_t\}_{t=0}^\infty \in \text{Bd} \Lambda^\infty \) or \( \varphi < \delta \). First consider the case, \( \varphi < \delta \). But now equation (79) cannot apply. Since this equation must hold for all \( \{C_t, L_t\}_{t=0}^\infty \in \text{Int} \Lambda^\infty \), we must have that \( \{C_t, L_t\}_{t=0}^\infty \in \text{Bd} \Lambda^\infty \). This implies that in both cases, \( \{C_t, L_t, K_t\}_{t=0}^\infty \) must converge to some \((C^*, L^*, K^*)\) (since by hypothesis a steady state exists). The same argument as in the proof of Theorem 3 shows convergence to \( \bar{w} \) (and to \( \hat{C}, \hat{L}, \hat{K} \) which satisfies Assumption 5). This combined with Lemma 14 establishes the desired result.

Part 3: Suppose that \( \varphi > \delta \). In this case, (79) implies that \( U_{C_t}^* \) is proportional to \( \varphi > \delta \) as \( t \to \infty \). This implies that \( (\mu_t - \mu_{t-1})/\mu_t > 0 \) as \( t \to \infty \), so from (77) and (78), aggregate distortions cannot disappear, completing the proof.