Markets Versus Governments: Political Economy of Mechanisms*

Daron Acemoglu MIT

Michael Golosov MIT

Aleh Tsyvinski Harvard

September 2005
Very Preliminary and Incomplete Version.

Abstract

This paper investigates the political economy of (centralized) mechanisms and the comparison of these mechanisms to markets. In contrast to the standard approach to mechanism design, we assume that the mechanism is operated by a self-interested agent (ruler/government), who can misuse the resources and the information he or she collects.

The main contribution of the paper is the analysis of the form of mechanisms to insure idiosyncratic (productivity) risks as in the classical Mirrlees setup, but in the presence of a self-interested government. We construct sustainable mechanisms where the government is given incentives not to misuse resources and information. 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 provide a characterization of the best sustainable mechanism. A number of features are interesting to note. 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 an agent is maximized subject to incentive compatibility constraints, as well as two additional constraints on the total amount of consumption and labor supply in the economy. Second, we characterize the conditions under which the best sustainable mechanism will lead to an asymptotic allocation where the highest type faces a zero marginal tax rate on his or her labor supply as in the classical Mirrlees setup and there is no aggregate capital taxes as in the standard dynamic taxation literature. In particular, if the government is sufficiently patient (typically as patient as the agents), the Lagrange multiplier on the sustainability constraint of the government tends to zero, and marginal distortions arising from political economy disappear asymptotically. In contrast, when the government has a small discount factor, we show that aggregate distortions remain, and there is both positive marginal labor tax on the highest type and positive aggregate capital taxes even asymptotically. We also investigate when markets are likely to be less desirable relative to centralized mechanisms.

*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, perhaps 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 has instead attempted to explicitly model the informational problems restricting the set of instruments that the government can use, and has therefore treated the problem of optimal policy as one of mechanism design.[1] At the simplest level, we can think of each individual reporting their "type" to a centralized mechanism designer, who then implements the allocation. If the environment is one in which the first welfare theorem holds, so that a competitive market equilibrium is Pareto efficient, then the mechanism can simply replicate this allocation. Despite the numerous theoretical insights from this approach, many real-world resource allocation problems, at least superficially, do not appear to work this way, and rely on a range of *anonymous* market transactions.

An obvious reason for this is political economic.[2] Who will operate the mechanism? Who will collect information about each individual and then make centralized decisions on the basis of this (potentially sensitive and valuable) information?[3] The implicit answer seems to be the "government," but 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.[4]

[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), and Werning (2002) consider applications of this framework to dynamic taxation.

[2] 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.

[3] Naturally, one can think of an extended game in which there is a fictional disinterested mechanism designer, with the government who has the authority to tax and regulate as an additional player. Although this may be a useful modeling tool, it does not circumvent the substantive issues raised here: the party entrusted with operating optimal policy has neither the same interests as those of the citizens nor much commitment power.

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

Throughout the paper, we use the term "government" or ruler to stand for a group of individuals with the political power to operate a transfer scheme across individuals and also to divert some of those resources for different purposes, including their own consumption.

The purpose of this paper is to contribute to a potential third-generation approach to public finance where both the informational problems on tax instruments and the incentive problems associated with governments, politicians and bureaucrats are taken into account. More concretely, our purpose is to investigate how mechanisms work and should be designed in the presence of self-interested governments.

Two questions motivate this analysis. First, we would like to understand whether mechanisms, say to deal with idiosyncratic risks, would look similar to those described by Mirrlees when they are operated by a self-interested rather than a selfless government/ruler. In other words, we would like to understand the differences between *Mirrlees mechanisms* and *sustainable mechanisms.* This is important to evaluate whether some of the conclusions of the standard mechanism design approach are robust to incorporating such realistic political economy features. If they are, then we can have more confidence in the mechanism design approach as a tool to analyze the practice as well as a normative benchmark.[5] Second, the recognition of the costs of having self-interested governments designing and operating mechanisms makes a meaningful comparison of markets versus mechanisms possible, and we would like to take a first step in this direction.

To investigate these questions, we construct a dynamic Mirrlees economy in which each individual is risk-averse and is subject to productivity shocks. A mechanism corresponds to an arrangement in which each individual at every date reports her type. The mechanism designer (government) dictates how much labor will be supplied by each individual, collects the output and distributes part of it as consumption to the individuals. In this environment, there is an extreme form of informational rent for the government, who can decide to consume all (or part) of the output once it has collected it. Moreover, the government can exploit the information that it gathers about individuals (from their past reports). Environments of this sort are difficult to analyze because the powerful tools of mechanism design theory that rely on direct truth-telling mechanisms and the revelation principle cannot be applied.

Our approach to this problem is to show that a version of the revelation principle holds along the equilibrium path as long as the mechanism is "sustainable". More specifically, the society needs to design a *sustainable mechanism,* which is not only incentive compatible for

[5] 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.

the citizens (agents), but also provides dynamic incentives to the the government (i.e., it must satisfy a sustainability constraint). These incentives take the form of a fraction of the output being given to the government in every period. If the government deviates from this implicit agreement and consumes more than this amount, then individuals switch to supplying zero labor, implicitly punishing the ruler.[6] The important point is that, along the equilibrium path, when the government pursues the (implicit) rules specified by the sustainable mechanism, there is effective commitment on the side of the government. This is the idea that allows us to prove a version of the revelation principle, the *truthful revelation along the equilibrium path*, as a 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).

A major result of our analysis is that since, with a sustainable mechanism, part of the output has to be given to the government, the market allocation cannot be replicated by the mechanism. Thus, under certain conditions, markets are strictly preferred to mechanisms. Nevertheless, a centralized mechanism may be preferred to markets because of its insurance benefits.

The bulk of the paper characterizes properties of sustainable mechanisms and their comparison to market outcomes. We first define 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 by citizens in the economy be no greater than some amount C_t. When the mechanism also optimizes over the levels of C_t and L_t subject to the aggregate resource constraints, the quasi-Mirrlees problem is identical to the (dynamic) Mirrlees problem of maximizing expected utility subject to incentive compatibility and feasibility constraints.

We prove that, under some conditions, a sustainable mechanism will always solve a quasi-

[6] In practice, citizens have many other recourses against governments that misbehave, including voting or throwing them out of office. We do not incorporate these possibilities to simplify the analysis. Future and more realistic models pursuing this approach will certainly need to incorporate detailed analyses of the procedures of government replacement.

Mirrlees problem. 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), and the extent of aggregate distortions and how implicit taxes on different individuals change as a result of the opportunistic behavior of the government can be determined from the properties of the quasi-Mirrlees problem. We also clarify the conditions under which there will be further distortions and the sustainable mechanism will not even solve a quasi-Mirrlees problem.

Most importantly, although a sustainable mechanism is typically different from a Mirrlees mechanism, we obtain a number of interesting results regarding taxation under sustainable mechanisms. In particular, when the government is sufficiently patient (in some cases as patient as, or more patient than, the citizens), we show that in any sustainable mechanism the Lagrange multiplier on the sustainability constraint of the government will tend to zero (i.e., the sustainability constraint will eventually become slack). In this case, we can establish two important results: first, aggregate distortions disappear asymptotically, so that the marginal products of labor and capital are equal to some well-defined marginal rates of substitution for the agents. This implies, for example, that as in the Chamley-Judd type dynamic macro models, sustainable mechanisms are consistent with zero aggregate capital taxes in the long run (Chamley, 1986, Judd, 1985);[7] second, lack of aggregate distortions implies that, as in Mirrlees mechanisms, again in the long run, the labor supply of the highest type will not be distorted. In contrast, when the government has a small discount factor, 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 type agents. 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.

We conclude the paper by a brief comparison of sustainable mechanisms to anonymous markets. Anonymous markets are modeled as Bewley-Aiyagari economies with incomplete markets and only self-insurance (Bewley, 1977, Aiyagari, 1994). We show that a higher discount factor of the government or greater institutional controls on the government always make mechanisms more desirable. Also in an important special case, we show that greater risk aversion on the part of individuals makes mechanisms more desirable relative to autonomous

[7] The "aggregate" qualification here is important, since the allocation in a sustainable mechanism can be decentralized in different ways. Some of those may involve positive taxes on individual capital holdings.

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 literature already discussed in footnote 1. The major difference between our work and all of 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 4. What distinguishes our paper from this literature is the explicit modeling of the incentive problems on the side of the individuals, which then give additional (informational and other) power to the government.[8] Finally, our analysis is also related to work on optimal taxation with time-inconsistency, for example, Chari and Kehoe (1990, 1993), Phelan and Stacchetti (2001), and Sleet and Yeltekin (2004).[9]

Most closely related to our paper is the recent important work by Bisin and Rampini (2005), who consider the problem of mechanism design without commitment in a two-period setting. In their model, the government is benevolent, but cannot commit to not exploiting the information it has collected in the first period. Bisin and Rampini 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. Aside from the differences in modeling details, 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 Bisin and Rampini's model, enables us to analyze substantially more general environments. Second, the infinite horizon setting enables an investigation of the limiting behavior of distortions and taxes, giving us a number of important results as discussed above. Third, while Bisin and Rampini view anonymous markets as a constraint on government behavior, in our work, we compare under what conditions the society is better off with anonymous markets versus government-controlled mechanisms.

The rest of the paper is organized as follows. Section 2 describes the basic environment.

[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 recent work on general mechanisms without commitment as in Bester and Strausz (2001), Skreta (2004), and Miller (2005), as well as early classic by Roberts (1984).

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 characterizes the equilibrium distortions and the limiting behavior of sustainable mechanisms under a variety of scenarios. It also clarifies conditions under which the quasi-Mirrlees formulation does not apply. Section 5 describes the equilibrium with anonymous markets and discusses conditions under which anonymous markets will be preferred to sustainable mechanisms from an ex ante point of view. Section 6 concludes, while the Appendix contains a number of more difficult technical material, especially useful in establishing concavity of value functions using lotteries, as well as a number of other proofs.

2 Demographics, Preferences and Technology

In this section, we describe preferences and technology. The model economy is infinite horizon in discrete time. It is populated by a continuum 1 of agents, each denoted by i, 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.

We next describe the evolution and distribution of agents' skills. Let $\Theta = \{\theta_0, \theta_1, ..., \theta_N\}$ be a finite ordered set of potential types, with the convention that θ_i corresponds to "higher skills" than θ_{i-1}, and in particular, θ_0 is the worst type.[10] Let Θ^T be the T-fold product of Θ, representing the set of sequences of length T, with each element belonging to Θ. In this definition, $T = \infty$ is allowed and in fact is the case of most interest. We think of each agent's type is drawn from Θ^∞ according to some measure μ^∞. This imposes no restriction on the time-series properties of individual skills. Both iid draws from Θ in every period and constant types, as well as arbitrary temporal dependence are allowed. We only assume that each individual's draw from Θ^∞ is according to the same measure μ^∞ 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 denoted by G.

[10] Finiteness of Θ is adopted for simplicity and without loss of any economic insight. The more general case where Θ is a compact interval of $\mathbb{R}_+$ introduces a number of additional technical details, not central for our analysis.

Let $\theta^{i,\infty}$ be the draw of individual i from Θ^{∞}. The t-th element of $\theta^{i,\infty}$, θ_t^i, is the skill level of this individual at time t. We use the standard notation $\theta^{i,t}$ to denote the history of this individual's skill levels up to and including time t, and make the standard 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. 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 $\mathcal{F}_t^i$ at time t.[11] We will drop the index i when this causes no confusion. Finally, we will make a *full support* assumption on μ^{∞}, meaning that for any $\theta^{i,t-1} \in \Theta^{t-1}$, θ_t^i can take any value of the set Θ with some positive probability.

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

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

where c_t^i is the consumption of this individual and l_t^i is her labor supply. We assume that labor supply of an individual with skill θ comes from a compact set, i.e., $l_t^i \in \left[0, \bar{l}\,(\theta)\right]$. In addition, we make a number of standard assumptions on u. Let $\mathbb{R}_+$ denote the nonnegative real numbers.

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

Assumption 2 ***(worst type and full support)*** *$\bar{l}\,(\theta_0) = 0$, while $\bar{l}\,(\theta) = \bar{l} < \infty$ for all $\theta \in \Theta$ and $\theta \neq \theta_0$, and μ^{∞} has full support in the sense that $\theta_t^i = \theta_0$ has positive probability after any history.*

Assumption 3 ***(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 θ for all c and l and all $\theta \in \Theta$.*

These assumptions are standard. The only one that requires some explanation is Assumption 2. This assumption first states that for the worst type, θ_0, supplying positive labor is

[11] More formally, let $\mathcal{F}$ be the set of subsets of Θ^{∞}, so that the triple $(\Theta^{\infty}, \mathcal{F}, \mu^{\infty})$ is a probability space. In addition, let $\left\{\mathcal{F}_t^i : t \in \mathbb{N}\right\}$ be a filtration, i.e., a collection of sub-sigma fields of $\mathcal{F}$, such that $\mathcal{F}_t^i \subseteq \mathcal{F}_{t'}^i$ for all $t' > t$. Let Θ^t be the set Θ^{∞} truncated at t. Then $\theta^{i,t} \in \Theta^t$ and all decisions taken at time t by individual i must be $\mathcal{F}_t^i$-measurable. See, for example, Pollard (2002).

impossible. This suggests that we can think of the worst type as "disabled," meaning unable to supply any labor. It also states that there is full support on μ^{∞}, so 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. The single-crossing property, Assumption 3, will enable us later to reduce the number of incentive compatibility constraints.

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

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

where $\mathbb{E}\left[\cdot|\mathcal{F}_{t}^{i}\right]$ or $\mathbb{E}\left[\cdot|\theta^{i,t}\right]$ denote the expectations operator conditional on having observed the history $\theta^{i,t}$ (in addition to any public information). Throughout, we will assume that no other agent knows $\theta^{i,t}$, making this a hidden-state economy.

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

$$F\left(K,L\right) \tag{2}$$

where K is capital and L is labor. Each agent in the economy has access to this production function. Below, we will analyze both the case in which there is capital and also the economy without capital, in which case $F\left(K,L\right)$ will simply be linear in L. We assume that capital fully depreciates after use (which will turn out to simplify the notation and the discussion of savings under the mechanism). Finally, we also make the simplifying assumption that $F\left(K,0\right)=0$, so that without labor there is no production.

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

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

where $\mathbb{E}_{t}$ refers to the expectations operated conditional on public information at time t.

3 Sustainable Mechanisms

In this section, we described the game between the government and the individuals. As in the standard mechanism design problem, the government provides incentive-compatible insurance to individuals in order to maximize ex ante welfare. The main difference is that as in the analysis of Chari and Kehoe (1990, 1993), the interaction between the citizens and the government is a game. As already emphasized in the Introduction, the government both lacks commitment power and also has the ability to extract all of 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).[12] The sustainability constraints will ensure that the government both abides by a given implicitly-agreed sequence of (sub)mechanisms and also does not extract more of the output than it is supposed to.

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.[13] Nevertheless, before discussing the best sustainable mechanisms, we need to characterize properties of sustainable mechanisms in general, which is the task in this section.

3.1 The Game Form Between Government and Citizens

In this section, we will describe the game played between the government and the citizens.

For this purpose, let us first 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 $\hat{\theta}^t$ to stand for reports regarding types in the direct mechanisms. A submechanism specifies what happens at a given date. In particular, let Z_t

[12] More generally, we can allow the government to extract and consume some fraction $\eta \leq 1$ of the total output of the economy, where the level of η will be related to the institutional controls on government or politician behavior. For now, we set $\eta = 1$ to simplify notation, and return to issues of institutional limits on government action below.

[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".

be a general message space for time t, with a generic element z_t.[14] This message space may include messages about current type of the individual, $\hat{\theta}_t^i \in \Theta$, and past types $\hat{\theta}^{i,t-1} \in \Theta^{t-1}$ (even though the individual may have made some different reports about his or her types in the past), and might also include other messages. For future reference, we write $z_t = z_t\left[\theta^t\right]$ if, as part of her messages, the individual reports her true type (or something that reveals her true type).

Let $Z^t = \prod_{s=0}^{t} Z_s$ and z^t denote a generic element of Z^t. Then a direct submechanism would specify a mapping from Z^t to $\left[0, \bar{l}\right] \times \mathbb{R}_+$, that is, a level of consumption and labor supply for any history of messages up to this point. With direct mechanisms, the individuals make their reports and the mechanism designer allocates labor supply and consumption levels according to these reports. We now define a submechanism:

Definition 1 *A submechanism is a mapping*

$$M_t : Z^t \to \left[0, \bar{l}\right] \times \mathbb{R}_+,$$

which assigns labor supply and consumption levels for each complete history of messages.

In other words, for every complete history of messages, the submechanism promises a level of labor supply $l \in \left[0, \bar{l}\right]$ and a level of consumption $c \in \mathbb{R}_+$.[15] Given Assumption 2, any submechanism must allow for some messages which will lead to $l = 0$. We denote the set of submechanisms that satisfy this restriction and also the relevant resource constraints (which will be specified below) by $\mathcal{M}_t$.[16]

The typical assumption in models with no commitment (e.g., Roberts, 1984) 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, the possibility of the government using its power to extract resources from the society necessitates a modification where there is the possibility of deviation even within the same period. The interaction between the government and the individuals is modeled with the following game form at each date:

[14] More formally, θ^t, $\hat{\theta}^t$ and z_t have to be $\mathcal{F}_t$-measurable.

[15] What we have described here are "pure strategy" submechanisms that do not allow for lotteries. Allowing for lotteries 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.

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

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.

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

4. The government decides whether to distribute consumption among agents according to $\tilde{M}_t \in \mathcal{M}_t$ or to deviate to some other mechanism $\tilde{M}_t' \in \mathcal{M}_t\left(\tilde{M}_t\right)$ to determine the distribution of consumption. It also decides how much of this production to consume itself, x_t, and how much to invest, K_{t+1}.

In this game form, the notation $\tilde{M}_t' \in \mathcal{M}_t\left(\tilde{M}_t\right)$ emphasizes the fact that the submechanism $\tilde{M}_t'$ can only be different from the submechanism $\tilde{M}_t$ in its allocation of consumption, because by the time the government chooses $\tilde{M}_t'$, labor decisions have already been made. Given this, we write $\tilde{M}_t' = \tilde{M}_t$ (only) when these two mechanisms have identical distributions of consumption across agents.

This game form emphasizes that the only difference between the standard models with no commitment and our setup is that, because of the ability of the government to take some of the resources for its own consumption, we have to allow a deviation at the last stage to a different mechanism. In particular, if the government chooses to extract all of the resources of the society for itself, clearly it cannot abide by the consumption levels promised in $\tilde{M}_t$, hence we have to allow for a deviation to a different mechanism at this point.

Definition 2 *$M = \{M_t\}_{t=0}^{\infty}$ with $M_t \in \mathcal{M}_t$ is a mechanism, and we denote the set of mechanisms 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. The social plan also implicitly determines how much will be invested in next period's capital stock from the resource constraint.

Denote the action of the government at time t is $\rho_t = \left(\tilde{M}_t, \tilde{x}_t, \tilde{M}_t'\right) \in \Re_t \equiv \mathcal{M}_t \times \mathbb{R}_+ \times \mathcal{M}_t$. The first element is what the government offers at stage 1 of time t, the second is what the government consumes itself, and the third is what the government uses to distribute consumption at the end. The largest amount that the government may choose to take is equal to the output in the economy, i.e., $\tilde{x}_t \leq F(K_t, L_t)$, but to simplify notation we write $\tilde{x}_t \in \mathbb{R}_+$. With standard notation, we use $\rho^t \in \Re^t$ to denote the history of ρ_t's up to and including time t.

Let us further define $\Gamma_{M,x} = \left[\left\{\tilde{M}_t, \tilde{x}_t, \tilde{M}_t'\right\}_{t=0}^{\infty} \mid (M, x)\right]$ as a government strategy profile. The notation emphasizes that what matters is the sequence of actions by the government given the implicitly agreed social plan (M, x).

Turning to the citizens, define $\alpha_t^i\left(\theta^t \mid z^{t-1}, \rho^{t-1}\right)$ as the action of individual i at time t when her type history is θ^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 ρ^{t-1}. α_t^i specifies a message $z_t \in Z_t$:

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

We will sometimes abbreviate notation by writing $z^t\left(\alpha_t\left(\theta^t\right)\right)$ to denote the message resulting from strategy α_t with type θ^t.

We call a strategy *truth telling* if it satisfies

$$\alpha^*\left(\theta^t \mid z^{t-1}, \rho^{t-1}\right) = z_t\left[\theta^t\right] \text{ for all } \theta^t \in \Theta^t,\ z^{t-1} \in Z^{t-1} \text{ and } \rho^{t-1} \in \Re^{t-1}. \tag{3}$$

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 will shorten the truth-telling strategy to $\alpha_t^i\left(\theta_t \mid \theta^{t-1}, \rho^{t-1}\right) = \alpha^*$. Notice that this strategy only imposes truth-telling following truthful reports in the past. As we will see below, this is without loss of any generality. In addition, let us define the null strategy

$$\alpha^{\emptyset}\left(\theta_t \mid \hat{\theta}^{t-1}, \rho^{t-1}\right) = z_t\left[\theta_0\right] \text{ for all } \theta^t \in \Theta^t,\ z^{t-1} \in Z^{t-1} \text{ and } \rho^{t-1} \in \Re^{t-1},$$

where $z_t\left[\theta_0\right]$ 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 2. Therefore, the individual can always choose to supply zero labor, or in other words, any feasible mechanism (submechanism) must allow for "freedom of labor supply". We will use the notation $\alpha_t^i\left(\theta_t \mid \theta^{t-1}, \rho^{t-1}\right) = \alpha^{\emptyset}$ to denote that the individual is playing the null strategy. Finally, we denote the *strategy profile* of all the individuals in society by $\boldsymbol{\alpha}$. Moreover, we use the notation $\boldsymbol{\alpha} = (\alpha \mid \alpha')$ to denote a strategy profile where individuals play α along the equilibrium path and α' off the equilibrium path, with $\mathbf{A}$ denoting the set of all such strategy profiles.

Next, turning to the government's strategies, note that at each date, the government chooses three objects, $\tilde{M}_t \in \mathcal{M}_t$, $\tilde{x}_t \in [0, F(K_t, L_t)]$ and $\tilde{M}'_t \in \mathcal{M}_t\left(\tilde{M}_t\right)$. We have already denoted the vector of these actions by ρ_t and the set of such vectors by $\Re_t$. In addition, let $\mathbf{z}_t \in \mathcal{Z}_t$ be a profile of reports at time t.[17] As usual, we define $\mathcal{Z}^t = \prod_{s=0}^{t} \mathcal{Z}_s$. The government's strategy at time t is therefore

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

i.e., it determines $\tilde{M}_t \in \mathcal{M}_t$, $\tilde{x}_t \in [0, F(K_t, L_t)]$ and $\tilde{M}'_t \in \mathcal{M}_t\left(\tilde{M}_t\right)$ 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 $\Gamma \in \mathcal{G}$. A (*sequential) equilibrium* in the game between the government and the citizens is given by strategy profiles $\hat{\Gamma}$ and $\hat{\boldsymbol{\alpha}}$ that are best responses to each other in all information sets given beliefs, 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 $\hat{\Gamma} \succeq_{\hat{\boldsymbol{\alpha}}} \Gamma$ for all $\Gamma \in \mathcal{G}$ and $\hat{\boldsymbol{\alpha}} \succeq_{\hat{\Gamma}} \boldsymbol{\alpha}$ for all $\boldsymbol{\alpha} \in \mathbf{A}$.

For our analysis the more important concept is that of a *sustainable mechanism*. For this purpose, let us define $\Gamma_{M,x} = \left[\left\{\tilde{M}_t, \tilde{x}_t, \tilde{M}'_t\right\}_{t=0}^{\infty} \mid (M, x)\right]$ to the along-the-equilibrium-path action profile (or action profile for short) induced by strategy profile Γ given a social plan (M, x). Conditioning on the social plan here is simply for emphasis.

Definition 3 *M is a sustainable mechanism if there exists $x = \{x_t\}_{t=0}^{\infty}$, a strategy profile $\boldsymbol{\alpha}$ for the citizens and a strategy profile $\Gamma \in \mathcal{G}$ for the government, which induces an action profile $\Gamma_{M,x} = \left[\left\{\tilde{M}_t, \tilde{x}_t, \tilde{M}'_t\right\}_{t=0}^{\infty} \mid (M, x)\right]$ for the government such that $\tilde{M}_t = \tilde{M}'_t = M_t$ and $\tilde{x}_t = x_t$, and satisfies $\hat{\Gamma} \succeq_{\boldsymbol{\alpha}} \Gamma$*

In essence, a sustainable mechanism is part of a social plan (M, x) from which the government does not wish to deviate given the strategy profile, $\boldsymbol{\alpha}$, of the citizens. In this context, not deviating means offering the implicitly-agreed submechanism at every date, consuming exactly as much as agreed, and distributing consumption across agents according to the implicitly-agreed submechanism. The notation $\hat{\Gamma} \succeq_{\boldsymbol{\alpha}} \Gamma$ makes this explicit, stating that given the strat-

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

egy profile, $\boldsymbol{\alpha}$, 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). It enables mechanism design problems to be formulated in terms of a truth-telling mechanism, where each type prefers to report his or her true type. 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, Guesneries and Tirole, 1985, or Bisin and Rampini, 2005).[18] 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:

$$\max_{(M,x)} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t\left(z^t\left[\alpha_t\left(\theta^t\right)\right]\right), l_t\left(z^t\left[\alpha_t\left(\theta^t\right)\right]\right) \mid \theta_t^i\right)\right] \tag{4}$$

subject to a resource constraint of the form

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

a set of incentive compatibility constraints for individuals, i.e.,

$$\boldsymbol{\alpha} \text{ is a best response to (incentive compatible against) } \left[\left\{\tilde{M}_t, \tilde{x}_t, \tilde{M}_t'\right\}_{t=0}^{\infty} \mid (M, x)\right] \tag{6}$$

[18] As already noted in the Introduction, it is possible to construct a "grand" mechanism with a fictitious disinterested mechanism designer, and treat the government as a player. However, to capture the substantive issues we are dealing with here, one would need to impose the additional restriction on this mechanism that any communication between citizens and the mechanism designer are also observed by the government. Without this additional restriction, the revelation principle will naturally apply to this grand mechanism, but would be of little relevance to the environment we are studying here. An alternative way of viewing our theorem in this section is therefore as showing that the revelation principle applies along the equilibrium path of the ground mechanism even when we impose the additional restriction that the government observes all communications between citizens and the fictitious mechanism designer.

and the "sustainability" constraint of the government:

$$\sum_{s=0}^{\infty} \delta^s v\left(x_{t+s}\right) \geq v\left(F\left(K_t, \int l_t\left(z^t\left[\alpha_t\left(\theta^t\right)\right]\right) dG^t\left(\theta^t\right)\right)\right) + \delta v_t^c\left(\tilde{M}^t\right), \tag{7}$$

for all $t \geq 0$. This last constraint can be explained as follows: the left-hand side is what the government will receive by sticking with the implicitly-agreed consumption schedule for itself. The right-hand side is what it can receive with the best deviation this period, which involves extracting as much as possible (i.e., confiscating all of the output), plus some continuation value, which may depend on the sequence (history) of submechanisms used so far, $\tilde{M}^t$.

Our first result establishes that $v_t^c\left(\tilde{M}^t\right) = 0$. This is essentially equivalent to the results in repeated games where the most severe punishments against deviations are optimal (e.g., Abreu, 1988).

Proposition 1 *In the best sustainable mechanism, $v_t^c\left(\tilde{M}^t\right) = 0$ for all $\tilde{M}^t \in \mathcal{M}^t$, so (7) takes the form*

$$\sum_{s=0}^{\infty} \delta^s v\left(x_{t+s}\right) \geq v\left(F\left(K_t, L_t\right)\right), \tag{8}$$

where K_t is total capital stock and L_t is total labor supply.

Proof. Reducing $v_t^c\left(\tilde{M}^t\right)$ is equivalent to relaxing the constraint on problem (4), so is always preferred. Since $v_t^c \geq 0$ (i.e., $x \geq 0$ and $v(0) = 0$), we only need to show that $v_t^c\left(\tilde{M}^t\right) = 0$ is achievable for all $\tilde{M}^t \in \mathcal{M}^t$. The following simple combination of strategies would achieve this objective. Let ρ^t be the history of actions by the government. Also denote $M_t' = M^{\emptyset}$ as the strategy that allocates zero consumption to all individuals. Let $\rho^t = \hat{\rho}^t$ if $\tilde{x}_{t-s} = x_{t-s}$ and $\tilde{M}_{t-s} = \tilde{M}_{t-s}' = M_{t-s}$ for all $s > 0$. Then the following strategy combination would ensure $v_t^c\left(\tilde{M}^t\right) = 0$ for all t: (1) for the citizens, $\boldsymbol{\alpha} = \left(\tilde{\alpha} \mid \alpha^{\emptyset}\right)$, for some $\tilde{\alpha}$, which means that for each citizen i and for all t, if $\rho^{t-1} = \hat{\rho}^{t-1}$, then $\alpha_t^i = \tilde{\alpha}$, and if $\rho^{t-1} \neq \hat{\rho}^{t-1}$, then $\alpha_t^i = \alpha^{\emptyset}$; (2) for the government, Γ, such that if $\rho^{t-1} = \hat{\rho}^{t-1}$, then Γ implies $\tilde{x}_t = x_t$ and $\tilde{M}_t = \tilde{M}_t' = M_t$, and if $\rho^{t-1} \neq \hat{\rho}^{t-1}$, then it implies $\tilde{x}_t = F\left(K_t, L_t\right)$ and any $\hat{M}_t \in \mathcal{M}_t$ and $\hat{M}_t' = M^{\emptyset}$. To complete the proof, we need to show that these strategies are sequentially rational. Consider the citizens; it suffices to note that following a history where $\rho^{t-1} \neq \hat{\rho}^{t-1}$, the government is playing $\tilde{x}_{t+s} = F\left(K_{t+s}, L_{t+s}\right)$ and $\hat{M}_{t+s}' = M^{\emptyset}$ for all $s \geq 0$. Therefore, any strategy other than $\alpha^{\emptyset}$ will give some utility less than $u\left(0, 0 \mid \theta\right) / \left(1 - \beta\right)$, which is the utility that always playing $\alpha^{\emptyset}$ will deliver. 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 $\rho^{t-1} \neq \hat{\rho}^{t-1}$, there will be no future output to expropriate, thus playing $\tilde{x}_{t+s} = F\left(K_{t+s}, L_{t+s}\right)$ (i.e., expropriating any output at that date), as well as $\tilde{M}_{t+s} = M^{\emptyset}_{t+s}$, is a best response for the government starting in all of its information sets for all $s \geq 0$. ■

This proposition therefore establishes that irrespective of the history of submechanisms, 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. As a result, the sustainability constraint of the government can be simplified to (8), which has the virtue of not depending on the history of submechanisms up to that point. This will be crucial in proving truthful revelation along the equilibrium path.

Next, we define a *direct (sub)mechanism* as $M^*_t : \Theta^t \to \left[0, \bar{l}\right] \times \mathbb{R}$. In other words, direct mechanisms involve restricted message spaces, $Z_t = \Theta_t$, where individuals only report their current type. We denote a strategy profiled by the government which induces direct submechanisms along the equilibrium path by Γ^*.

Definition 4 *A strategy profile for the citizens, $\boldsymbol{\alpha}^*$, is truthful if along the equilibrium path we have that $\alpha^i_t\left(\theta^t \mid \theta^{t-1}, \rho^{t-1}\right) = \alpha^*$. We write $\boldsymbol{\alpha}^* = (\alpha^* \mid \alpha')$ to denote a truthful strategy profile.*

The notation $\boldsymbol{\alpha}^* = (\alpha^* \mid \alpha')$ emphasizes that individuals play truth-telling along the equilibrium path, but may play some different strategy profile, α', off the equilibrium path. Clearly, a truthful strategy against a direct mechanism simply amounts to reporting the true type of the agent. We are now ready to define the revelation principle as it applies to our environment. Before doing this, let us define $\mathbf{c}\left[\Gamma, \boldsymbol{\alpha}\right]$, $l[\Gamma, \boldsymbol{\alpha}]$ and $x\left[\Gamma, \boldsymbol{\alpha}\right]$ 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, and recall that $(\alpha^* \mid \alpha')$ stands for truthful revelation along the equilibrium path.

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

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

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

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

$$\tilde{u}\left[\theta^t \mid \theta^t, \Gamma^*\right] = \tilde{u}\left[z_\Gamma^t(\theta^t) \mid \theta^t, \Gamma\right] \geq \tilde{u}\left[\tilde{z}_\Gamma^t(\theta^t) \mid \theta^t, \Gamma\right] = \tilde{u}\left[\hat{\theta}^t \mid \theta^t, \Gamma^*\right] \text{ for all } \hat{\theta}^t \in \Theta^t. \tag{9}$$

Equation (9) implies that $\boldsymbol{\alpha}^* = (\alpha^* \mid \alpha')$ is a best response along the equilibrium path for the agents against the mechanism M^* and government strategy profile Γ^*. Moreover, by construction, the resulting allocation when individuals play $\boldsymbol{\alpha}^* = (\alpha^* \mid \alpha')$ against Γ^* is the same as when they play $\boldsymbol{\alpha}$ against Γ. Therefore, by the definition of Γ being sustainable, we have $\Gamma \succeq_{\boldsymbol{\alpha}} \Gamma'$ for all $\Gamma' \in \mathcal{G}$. Now choose α' to be identical to $\boldsymbol{\alpha}$ off-the-equilibrium path, which implies that $\Gamma^* \succeq_{\boldsymbol{\alpha}^*} \Gamma'$ for all $\Gamma' \in \mathcal{G}$, establishing that $(\Gamma^*, \boldsymbol{\alpha}^*)$ is an equilibrium, completing the proof. ■

It is useful to note what this theorem entails. The most important implication 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-interest of the mechanism designer, is twofold: the first is our focus on sustainable mechanisms, and the second is the structure of the game which allows individuals to use punishment strategies that supply zero labor following the deviation by the government.[19] Consequently, it is in the

[19] Other papers, for example, Ausubel and Deneckere (1989), Fudenberg, Levine and Maskin (1994), or Miller (2005), make use of punishment strategies in infinitely-repeated incomplete information games, and obtain results with full information revelation (or extraction) as the discount factor approaches 1. The distinctive feature of the result here is that we obtain truthful revelation along the equilibrium path irrespective of the level of the discount factor. This has two reasons: first, the form of the political economy interactions make the sustainability constraints particularly simple; second (and more importantly), Proposition 1 implies that the equilibrium involves zero continuation payoff for the government after deviation irrespective of the past mechanisms that have been used.

interest of the government to stick with the implicit mechanism promised at $t = 0$. 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. However, a sustainable mechanism will be such that along the equilibrium path the government will have no interest in doing so, ensuring effective commitment along the equilibrium path. This in turn implies that individuals can report their types without the fear that this information or their labor supply will be misused.[20]

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 governments (mechanism designers) and agents 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 α^* 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 θ denoted by $G(\theta)$. This implies that θ^t has an invariant distribution, which is simply the t-fold version of $G(\theta)$, $G^t(\theta)$ (since there is a continuum of individuals, each history θ^t occurs infinitely often).[21] Given this construction, we can write total labor supply as $L_t = \int l_t\left(\theta^t\right) dG^t\left(\theta^t\right)$, and total consumption as $C_t = \int c_t\left(\theta^t\right) dG^t\left(\theta^t\right)$. Moreover, since Theorem 1 establishes that any sustainable mechanism is equivalent to a direct

[20] It is also useful to note at this point that this theorem could have been derived under somewhat different assumptions. Instead of choosing the game form with the government moving first and choosing the sub-mechanisms, an alternative game form would involve individuals first deciding their labor supply and then the government potentially rewarding them with consumption. In this case, the game form de facto introduces "freedom of labor supply", and we could have obtained exactly the same result as Theorem 1 without Assumption 2 above. With the game form used in the text, the disabled type and the full support assumption (Assumption 2) is important for the punishments off-the-equilibrium-path, thus for Theorem 1. Our choice of the structure in the text was motivated by our desire to maximize the parallel between our game form and those in the literature on mechanism design without commitment.

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

mechanism with truth-telling on the side of the agents, it immediately establishes the following proposition characterizing the best sustainable mechanism.

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

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

subject to some initial condition K_0, the resource constraint

$$K_{t+1} = F\left(K_t, \int l_t\left(\theta^t\right) dG^t\left(\theta^t\right)\right) - \int c_t\left(\theta^t\right) dG^t\left(\theta^t\right) - x_t, \tag{11}$$

a set of incentive compatibility constraints for individuals,

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

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

$$\sum_{s=0}^{\infty} \delta^s v\left(x_{t+s}\right) \geq v\left(F\left(K_t, \int l_t\left(\theta^t\right) dG^t\left(\theta^t\right)\right)\right), \tag{13}$$

for all t.

Note also that this optimization problem defines $\mathbf{U}^{SM}$ 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).[22]

In the next section, we characterize the solution to (10) and investigate conditions under which sustainable mechanisms are preferred to markets.

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

4 Sustainable Mechanisms and the Quasi-Mirrlees Program

In this section, we show that the best sustainable mechanism solves a quasi-Mirrlees program and describe some of the properties of the optimal allocations. We first establish a general result on the role of quasi-Mirrlees programs. We then analyze a particular example where there is no capital accumulation and the allocations to the agents are restricted to depend only on their current reports (and not on the history of their past reports). This economy can be justified by assuming that the government has no access to the past history of reports, thus we refer to it as *an economy with private histories*, though for us its main role is to clarify the basic trade-offs. Finally, we present the analysis for the general case, where agents' skills follow an arbitrary stochastic process, there is accumulation of physical capital, and the optimal allocations are history dependent.

4.1 Quasi-Mirrlees Formulation

Let us define the *quasi-Mirrlees program* as the following maximization problem:

$$U(\{C_t, L_t\}_{t=0}^{\infty}) \equiv \max_{\{c_t(\theta^t), l_t(\theta^t)\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t\left(\theta^t\right), l_t\left(\theta^t\right)\right)\right] \tag{14}$$

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

$$\int c_t\left(\theta^t\right) dG\left(\theta^t\right) \leq C_t, \tag{15}$$

and

$$\int l_t\left(\theta^t\right) dG\left(\theta^t\right) \geq L_t. \tag{16}$$

Here $c_t\left(\theta^t\right)$ and $l_t\left(\theta^t\right)$ are assumed to be $\mathcal{F}_t$-measurable bounded functions with the $\|\cdot\|_\infty$ norm. 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 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.

An important point which will play a crucial role in the analysis below is that 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 Λ^∞, i.e.,

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

We show in the Appendix that the set of constraints on the problem (14) form a compact set, while the objective function, u, is clearly continuous. Therefore, by the Weierstrass maximum theorem for the general (possibly infinite-dimensional) normed linear spaces, a solution exists (see Luenberger, 1969, Theorem 1, p. 40), so $U(\{C_t, L_t\}_{t=0}^\infty)$ is well defined as a functional. Moreover, given the concavity of the objective function u in $\left\{ c_t\left(\theta^t\right), l_t\left(\theta^t\right) \right\}_{t=0}^\infty$, convexity of the constraint set would be sufficient to ensure the concavity of the functional $U(\{C_t, L_t\}_{t=0}^\infty)$ in $\{C_t, L_t\}_{t=0}^\infty$. Nevertheless, the incentive compatibility constraints embedded in (12) do not form a convex set. For this reason, in the Appendix, we follow Prescott and Townsend (1984a,b) and allow lotteries to convexify the constraint set. 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 the Appendix, and in the text, we assume that $U(\{C_t, L_t\}_{t=0}^\infty)$ is concave. Another technical detail relates to the existence of Lagrange multipliers and to the differentiability of $U(\{C_t, L_t\}_{t=0}^\infty)$. These results are also derived in the Appendix and in the text, we assume that Lagrange multipliers for problem (14) exist and are unique (given a solution), and hence $U(\{C_t, L_t\}_{t=0}^\infty)$ is differentiable.

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

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

subject to

$$K_{t+1} \leq F\left(K_t, L_t\right) - C_t, \tag{17}$$

as well as

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

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

$$\max_{\{C_t, L_t, x_t, K_t\}_{t=0}^{\infty}} U(\{C_t, L_t\}_{t=0}^{\infty}) \tag{18}$$

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 (proof in the text).

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

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.

4.2 Best Sustainable Mechanism 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 such constraints, in this subsection we consider mechanisms with *private histories*, i.e., where individual histories are not observed by the government, and leave the analysis of the general environment to subsection 4.4 below. We also assume that there is no capital in the economy, so that the aggregate production function of the economy is

$$L_t = F(K_t, L_t) = \int l_t(\theta_t) dG(\theta_t), \tag{19}$$

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

$$u\left(c_t\left(\theta_t\right), l_t\left(\theta_t\right) \mid \theta_t\right) \geq u\left(c_t\left(\hat{\theta}_t\right), l_t\left(\hat{\theta}_t\right) \mid \theta_t\right) \tag{20}$$

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, (20) can be reduced to a set of incentive compatibility constraints only for neighboring types. Since there are $N+1$ types in Θ, this implies that (20) is equivalent to N incentive compatibility constraints.[23]

The best sustainable mechanism with private histories maximizes (10) subject to (13), (19), and (20).

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 $s \neq t$. Therefore we can represent the objective function in this problem in a time-separable form, so that the quasi-Mirrlees program becomes:[24]

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

subject to

$$C_t + x_t \leq L_t,$$

and

$$\sum_{s=0}^{\infty} \delta^s v(x_{t+s}) \geq v(L_t) \tag{21}$$

for all t.

[23] More specifically, in pure strategy direct mechanisms, there will be $N(N+1)$ incentive compatibility constraints, and Assumption 3 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.

[24] With a slight abuse of notation, we use $U(C, L)$ as a function of the level of C and L, whereas previously $U(\{C_t, L_t\}_{t=0}^{\infty})$ was defined as a function of entire sequences of $\{C_t, L_t\}_{t=0}^{\infty}$. This avoids further proliferation of notation.

As before, this problem is well defined only for some (C, L). We denote the set of such (C, L) pairs by Λ, i.e.,

$$\Lambda \equiv \{(C, L) : \exists (c(\theta), l(\theta)) \text{ s.t. } (15), (16) \text{ and } (20) \text{ are satisfied}\}.$$

Moreover, consider the maximization problem

$$\left(\hat{C}, \hat{L}\right) \in \arg \max_{(C,L)\in\Lambda} \hat{v}(C, L) \equiv v(L - C) - (1-\delta) v(L), \tag{22}$$

which essentially looks for the largest gap between the left and the right hand side of (21) in steady state. This problem always has a solution, since $C = 0$ and $L = 0$ are in Λ, in which case we have $\hat{v}(0, 0) = 0$. Also define

$$\bar{w} \equiv \arg \max_{(C,L)\in\Lambda} \frac{v(L - C)}{1 - \delta}$$

such that $\hat{v}(C, L) \geq 0$. Therefore, $\bar{w}$ is the highest steady state utility that can be given to the government without violating feasibility or its sustainability constraint (21). Clearly, only values $w \leq \bar{w}$ can be promised to the government. Let us now assume

Assumption 4 ***(sustainability)*** $\hat{v}\left(\hat{C}, \hat{L}\right) > 0$.

Inspection of (22) shows that this assumption is satisfied if the discount factor of the government, δ, is sufficiently large.[25]

It is also evident that, since $U(C, L)$ is differentiable, the solution to the (full) Mirrlees program (18) has to satisfy:[26]

$$U_C(C, L) = -U_L(C, L), \tag{23}$$

where U_C is the partial derivative of $U(C, L)$ with respect to C and $U_L(C, L)$ is defined likewise. In the Appendix (cfr. Lemma 7), we show that $U(C, L)$ is differentiable, so equation (23) is meaningful. This observation immediately leads to the following definition:

Definition 5 *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* $t \to \infty$.

[25] Notice that since Λ does not depend on δ, so as $\delta \to 1$, this assumption is surely satisfied.

[26] It is straightforward to see that in this case $(C, Y) \in \text{Int}\Lambda$.

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".

While the classical Mirrlees problem generates a number of important insights, a central one is that the marginal (labor) tax rate on the highest type should be equal to zero. This will play a role in our analysis below. For this reason, we also introduce the following terminology

Definition 6 *We say that the first Mirrlees principle holds at time t if the labor supply decision of the highest type of agent, θ_N, is undistorted at time t, i.e., if we have*

$$u_c\left(c_t\left(\theta_N\right), l_t\left(\theta_N\right) \mid \theta_N\right) = -u_l\left(c_t\left(\theta_N\right), l_t\left(\theta_N\right) \mid \theta_N\right).$$

Clearly, this is equivalent to the marginal tax rate on the labor supply of the highest type individual, which we denote by τ_N, being equal to zero.[27] Motivated by this discussion, we now have the following useful lemma:

Lemma 1 *Consider a sequence of $\{C_t, L_t\}_{t=0}^{\infty}$, then the marginal labor tax rate on the highest type of agent, θ_N, at time t is given by $\tau_{N,t} = 1 + U_L\left(C_t, L_t\right)/U_C\left(C_t, L_t\right)$.*

Proof. Assumption 3 (single crossing) implies that we only need to check incentive compatibility constraints for neighboring types. Lemma 6 establishes that the solution to the best sustainable mechanism is at a regular point (see the Appendix), so that Lagrange multipliers exist. Therefore, we have

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

where λ_{Nt} is the multiplier on incentive compatibility constraint between types θ_N and θ_{N-1} at time t, ν_{Ct} is the multiplier on (15) at t and ν_{Lt} is the multiplier on (16) at t. By the

[27] Recall that u_c denotes the partial derivative of the function u with respect to its first argument and u_y is its partial derivative with respect to the second argument. These partial derivatives exist by Assumption 1.

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 result. ■

An immediate corollary of this lemma links aggregate distortions (or the lack thereof) to the taxes on individual labor supply:

Corollary 1 *If $\{C_t, L_t\}_{t=0}^{\infty}$ is undistorted at time t (or as $t \to \infty$), then the first Mirrlees principle result holds at time t and $\tau_{N,t} = 0$ (or as $t \to \infty$, $\tau_{N,t} \to 0$).*

Proof. The hypothesis that $\{C_t, L_t\}_{t=0}^{\infty}$ is undistorted at time t implies that we have $\nu_{Ct} = \nu_{Lt}$, so that

$$u_c(c_t((\theta_N)), l_t(\theta_N) \mid \theta_N) = -u_l(c_t((\theta_N)), l_t(\theta_N) \mid \theta_N),$$

which immediately yields $\tau_{N,t} = 0$. When $\{C_t, L_t\}_{t=0}^{\infty}$ is asymptotically undistorted, then $\tau_{N,t} \to 0$, establishing the result. ■

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

$$V(w) = \max_{C,L,x,w'} \left\{U(C,L) + \beta V(w')\right\} \tag{24}$$

subject to

$$C + x \leq L,$$

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

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

$$w \in \mathbb{W} \text{ and } (C,Y) \in \Lambda. \tag{27}$$

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 choosing the initial value w_0 such that $w_0 \in \arg\max_w V(w)$.

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 of w', and thus over C and L (i.e., considering lotteries for the government in the same way as we do for individuals). We relegate the discussion of this issue to the Appendix, but here we take the sequence of values given (promised) to the government $\{w_t\}_{t=0}^{\infty}$ as a stochastic process, with each element taking values from the set $\mathbb{W}$, and consequently, the sequence $\{C_t, L_t\}_{t=0}^{\infty}$ that results from the best sustainable mechanism is also a stochastic sequence. Finally, in the text, we assume that V is concave and differentiable, which are both proved in the Appendix as well.

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 (marginal) 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 Assumption 4 holds.*

1. *At $t = 0$, there is an aggregate distortion and the first Mirrlees principle fails to hold.*

2. *Suppose that $\beta \leq \delta$. Let Γ 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 $\{\mathcal{W}_t\}_{t=0}^{\infty}$ whereby $w_t \in \mathcal{W}_t$. Then, we have that $\{w_t\}_{t=0}^{\infty}$ is a non-decreasing stochastic sequence in the sense that if $w_t = w_t' \in \mathcal{W}_t$, then any $w_{t+1}' \in \mathcal{W}_{t+1}$ 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 $\text{plim}_{t\to\infty} -U_C/U_L = 1$, so that asymptotically $\{C_t, L_t\}_{t=0}^{\infty}$ is (almost surely) undistorted, and the first Mirrlees principle holds.*

3. *If $\beta > \delta$, then there are aggregate distortions and the first Mirrlees principle fails to hold even asymptotically.*

This theorem is proved in the Appendix. Here we give a heuristic argument justifying the results. Before doing this, however, it is useful to interpret the consequences of this theorem. 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 first Mirrlees principle holds (i.e., 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. This implies 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 equally important, and may have even more empirical relevance. 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 γ and $\psi \geq 0$ be the Lagrange multipliers on the constraints (25) and (26) respectively. Lemma 12 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 $w' \in \text{Int}\mathbb{W}$. Therefore, taking the first order condition with respect to w' and using the Envelope theorem, we obtain that

$$\frac{\beta}{\delta} V'(w') = -\psi - \gamma = V'(w) - \psi \tag{28}$$

The other first order conditions are

$$U_C + U_L = \psi v'(L), \tag{29}$$

and

$$v'(x)(\psi + \gamma) = U_C. \tag{30}$$

Equation (29) makes it clear that aggregate distortions are related to ψ. 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 the proof of Lemma 1, this yields $\tau_N > 0$. Consequently, there is an aggregate distortion and the first Mirrlees principle fails to hold in the initial period.

Part 2 of Theorem 3, on the other hand, states that, as long as $\beta \leq \delta$, eventually aggregate distortions will disappear and the first Mirrlees principle will hold. 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 as the sustainability constraint on the government (26) binds. This, combined with the concavity of the value function $V(\cdot)$, which is proved in Lemma 10 in the Appendix, 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 (and they are increasing if and only if the sustainability constraints binds). In the text, this is stated as if $\{w_t\}_{t=0}^{\infty}$ is a non-stochastic sequence, but the Appendix establishes the more general result claimed in Theorem 3.

The intuition for why the rewards to the government are increasing 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.[28] 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 $\mathbb{W} = [0, \bar{w}]$. If $w^* < \bar{w}$, (28) immideately implies that $\psi = 0$. In other words, the Lagrange multiplier on the sustainability constraint of the

[28] This backloading effect disappears if the sustainability constraint does not bind and $\psi = 0$.

government, (26), eventually reaches zero, and at this point, aggregate distortions disappear. Corollary 1 then implies that the marginal tax rate on the labor supply of the highest type, θ_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. Essentially, high levels of future utilities are being promised to the government in order to relax constraints now. Loosely speaking, we can remove some of the sustainability constraints in the very far future, and this will effectively have no influence on the sequence of utilities promised to the government. This implies that eventually the multiplier on these sustainability constraint must tend to zero.

Next, suppose that $w^* = \bar{w}$. In this case, (28) may no longer be valid, since it applies only on the interior of $\mathbb{W}$. Nevertheless, in this case, it can be again proved (but now by a different argument, see Lemma 13) that aggregate distortions disappear. Essentially, $\bar{w}$ involves the maximum (steady-state) utility for the government, and it can be proved that this is achieved without any distortions (even though we reach the boundary of the set Λ). Therefore, the aggregate distortion again goes to zero, and the first Mirrlees principle applies.

Consider next the situation when government's and agents' discount factors differ. It can be shown that when $\delta > \beta$, the steady-state utility of the government will again reach $\bar{w}$, and an argument similar to that in the previous paragraph establishes that aggregate distortions disappear and the first Mirrlees principle applies.

It can also be noted that the same conclusions do not necessarily apply when Assumption 4 is not satisfied. For example when $\delta = 0$, which would make sure that Assumption 4 is violated, the only sustainable mechanism involves all agents providing zero labor, and the levels of consumption of the government and of the agents will be zero in all periods. Thus the steady state is reached in the initial period, but distortions are always present.

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 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

$$1+\frac{U_L}{U_C}=\left(1-\frac{\delta}{\beta}\right)\frac{v'(L^*)}{v'(x^*)} \tag{31}$$

$$C^*+x^*=L^*$$

$$v(x^*)=(1-\delta)v(L^*)$$

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 always remains positive, and the additional distortions created by the sustainability constraint remain even asymptotically.

In summary, the analysis in this section showed how the best sustainable mechanism can be characterized in the simplest economy with no capital and with private histories. The most interesting results concern the asymptotic behavior of distortions. When the government is as patient as (or more patient than) the citizens, despite the presence of initial aggregate distortions resulting from no commitment and the self-interested behavior of the government, these distortions disappear asymptotically. Consequently, in the limit, the first Mirrlees principle holds and the marginal tax rate on the highest type is equal to zero. In contrast, when the government is less patient than the citizens, aggregate distortions remain even asymptotically. In subsection 4.4, we will see that similar results hold in the more general environment described in Section 3.

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

It is useful to illustrate the results of subsection 4.2 with a simple example. Consider an economy with two types, i.e., $\Theta=\{\theta_0,\theta_1\}$ and

$$u\left(c,l\mid\theta\right)=\sqrt{c}-\frac{l^2}{2\theta}. \tag{32}$$

We continue to assume that type θ_0 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 θ_1, and suppose that the utility function of the government is also given by $v(x) = \sqrt{x}$.

Since type θ_0 cannot supply any labor, the derivation of $U(C, L)$ is particularly simple, and we must have $l(\theta_1) = L/\pi$. Moreover, the incentive compatibility constraint for type θ_1 is

$$\sqrt{c(\theta_1)} - \frac{l(\theta_1)^2}{2\theta_1} \geq \sqrt{c(\theta_0)}, \tag{33}$$

which will naturally hold as equality, pinning down $c(\theta_0)$ and $c(\theta_1)$ as a solution to (33) holding as equality, and $(1-\pi)c(\theta_0) + \pi c(\theta_1) = C$. Given this structure, we can easily determine $U(C, L)$, and then use the recursive program (24) to derive the value function $V(w)$.

We first start with the case where the government is as patient as the citizens, in particular with $\beta = \delta = 0.5$. This case illustrates part 2 of Theorem 3. The resulting value function in this case is plotted in Figure 1. Note that $V(w)$ is inverse U-shaped, first increasing and then decreasing. 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.

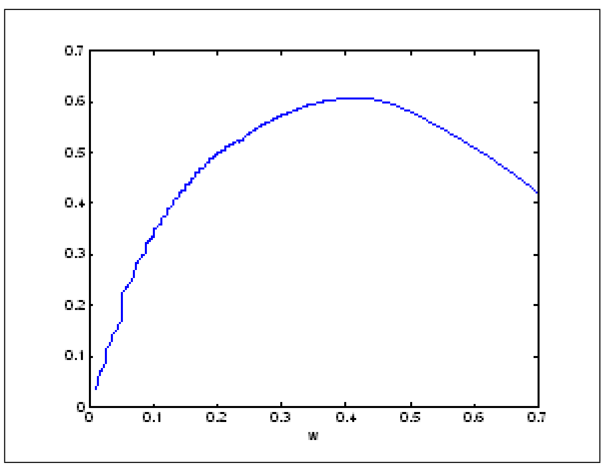

Figure 1: Theorem 3 part 2. Value function with $\beta = \delta = 0.5$.

More interesting for our purposes is the behavior of the promised value to the government, w. Its time path can be computed from program (24) once we have $V(w)$. To compute this time path, we start from the peak of $V(w)$, and follow the policy function implied by this value function. The resulting time path of $\{w_t\}$ is shown in Figure 2. Consistent with the results in Theorem 3 part 2, $\{w_t\}$ is an increasing sequence and convergenes to some level w^*.

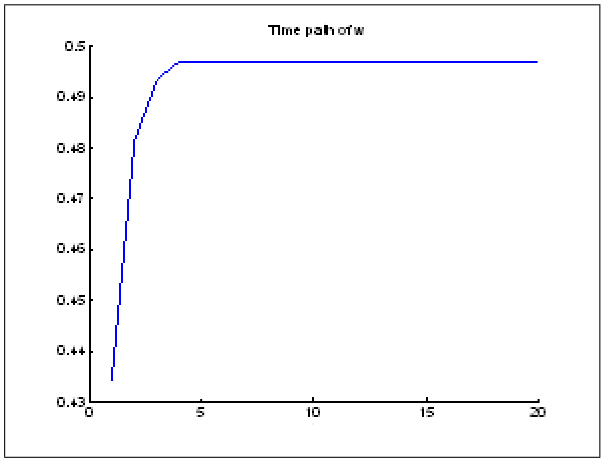

Figure 2: Theorem 3 part 2. Time path of w with $\beta = \delta = 0.5$.

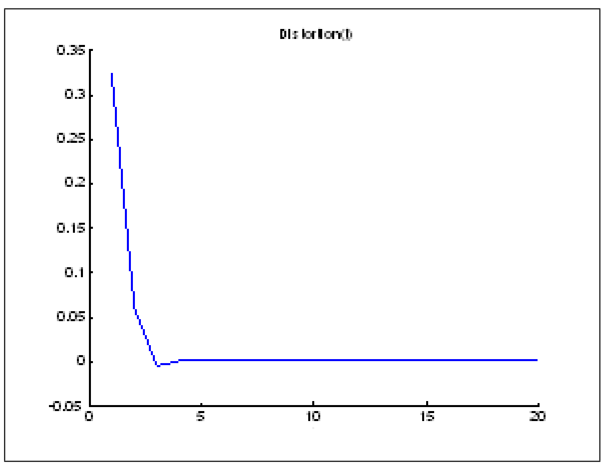

Figure 3: Theorem 3 part 2. Time path of distortions with $\beta = \delta = 0.5$.

Figure 3, in turn, depicts the evolution of the aggregate distortion (which is also equivalent to the marginal tax on type θ_1 in this case). As the sequence $\{w_t\}$ converges to w^*, as claimed in part 2 of Theorem 3, these distortions converge 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.

We next consider a modification of this environment to illustrate part 3 of Theorem 3. In particular, we change the above example so that the government's discount factor is $\delta = 0.4$, i.e., less than that of the agents, $\beta = 0.5$.

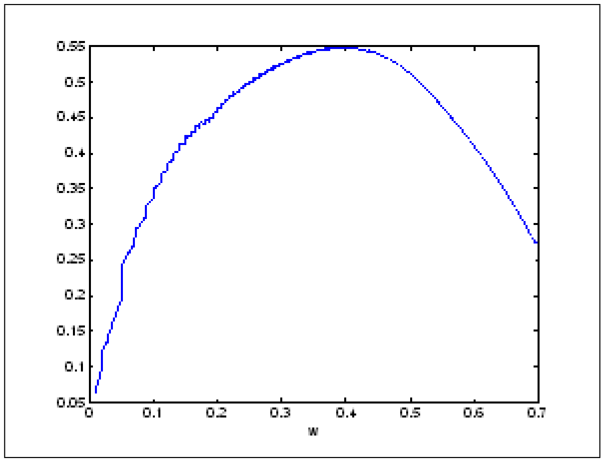

Figure 4: Theorem 3 part 3. Value function with $\beta = 0.5$ and $\delta = 0.4$.

The resulting value function is plotted in Figure 4. It is evident the level of utility achieved in this case is lower for each w than for the case when government has discount factor $\delta = 0.5$, since in this case the sustainability constraint (26) is more tight.

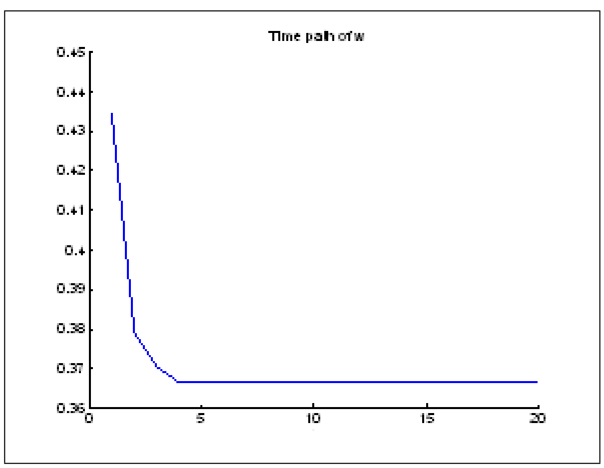

Figure 5: Theorem 3 part 3. Time path of w with $\beta = 0.5$ and $\delta = 0.4$.

To compute the time path of promised utilities for the government, we again start from the peak of $V(w)$, and follow the policy function implied by this value function. The resulting time path of $\{w_t\}$ is shown in Figure 5. The contrast to the economy with $\beta = \delta$ is quite stark, in particular, the promised utility sequence is now decreasing.

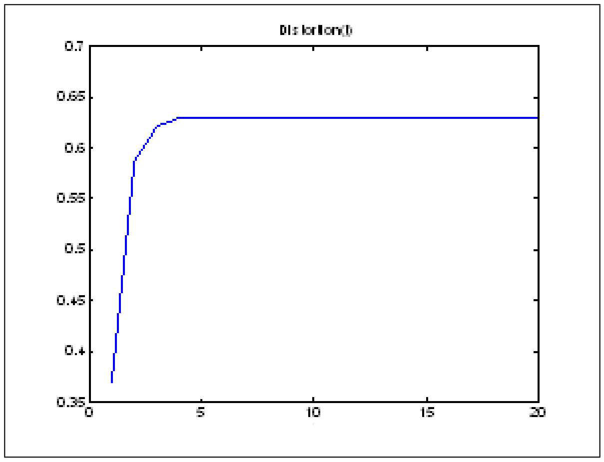

Figure 6: Theorem 3 part 3. Time path of distortions with $\beta = 0.5$ and $\delta = 0.4$.

Finally, Figure 6 depicts the evolution of the aggregate distortion (which is again equivalent to the marginal tax on type θ_1 in this case). As the sequence $\{w_t\}$ converges to w^* distortions converge. Again in contrast to the case with equal discount factors, this distorion increases (rather than decrease) and reaches a significant size of a 65 percent tax rate (on the highest type agent). This example shows that even relatively small differences in discount factors of the agent and the government can lead to very significant distortions in the long run.

4.4 Optimal History-Dependent Sustainable Mechanisms

We now return to the analysis of the general problem in Section 3 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 subsection 4.2. The quasi-Mirrlees program was defined above in (14), and the analysis there established that the best sustainable mechanism solves a quasi-Mirrlees program.

Recall first that Lemma 7 in the Appendix shows $U(\{C_t, L_t\}_{t=0}^{\infty})$ 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 $F_{L_t}(K_t, L_t)$ and $F_{K_t}(K_t, L_t)$, or simply by F_{L_t} and F_{K_t}. Then we have:

Definition 7 *In the general environment, we say that $\{C_t, L_t, K_t\}_{t=0}^{\infty}$ is undistorted at time t (or as $t \to \infty$), if we have*

$$U_{C_t} \cdot F_{L_t} = -U_{L_t} \text{ and } F_{K_{t+1}} \cdot U_{C_{t+1}} = U_{C_t}$$

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

This definition is the natural generalization of Definition 5. In particular, the first condition requires the marginal cost of effort at time t given the utility function $U(\{C_t, L_t\}_{t=0}^{\infty})$ 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 $U(\{C_t, L_t\}_{t=0}^{\infty})$, which represents the ex ante utility of an individual subject to incentive and feasibility constraints.

Moreover, we say that

Definition 8 *There is no aggregate capital taxation at time t if $F_{K_{t+1}} \cdot U_{C_{t+1}} = U_{C_t}$.*

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 following lemma is a straightforward generalization of Lemma 1 and Corollary 1, and its proof is omitted.

Lemma 2 *In the general environment, if $\{C_t, L_t, K_t\}_{t=0}^{\infty}$ is undistorted at time t (or as $t \to \infty$), then the first Mirrlees principle holds and there is no aggregate capital taxation at time t (or as $t \to \infty$).*

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 subsection 4.2. For this purpose, first define the following notation: we write $\{C, L\} \in \bar{\Lambda}^{\infty}$ 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 Λ^{∞} converging to a steady state with $\{C, L\}$. Note that despite the similarity of the symbols, $\bar{\Lambda}^{\infty}$ and Λ^{∞} are very different sets. Λ^{∞} is a subset of $\mathcal{L}^{\infty}$ (the set of infinite sequences bounded with the $\|\cdot\|_{\infty}$ norm), while $\bar{\Lambda}^{\infty} \subset \mathbb{R}^2$. Then we define:

$$\left(\hat{C}, \hat{L}, \hat{K}\right) \in \arg \max_{\{C,L\} \in \Lambda^{\infty}, K} \hat{v}\left(C, L, K\right) \equiv v\left(F\left(K, L\right) - C - K\right) - \left(1 - \delta\right) v\left(F\left(K, L\right)\right) \quad (34)$$

subject to (17), where the constraint $\{C, L\} \in \Lambda^{\infty}$ means that a sequence $\{C_t, L_t\}_{t=0}^{\infty}$, with each element equal to (C, L) belongs to Λ^{∞}. This program again looks for the largest gap between the left and the right hand side of the sustainability constraint for the general case, (13), in steady state. Once again $\hat{v}(0, 0, 0) = 0$ is a possible solution. Suppose instead that the maximization problem (34) has a solution with $\hat{v}(C, L, K) > 0$. Let us then define

$$\bar{w} \equiv \arg \max_{\{C,L\} \in \Lambda, K} \frac{v\left(F\left(K, L\right) - C - K\right)}{1 - \delta}$$

subject to (17) and $\hat{v}(C, L, K) \geq 0$. Clearly, $\bar{w}$ is again the highest steady state utility that can be given to the government without violating feasibility or its sustainability constraint. The key sustainability assumption is a generalization of Assumption 4:

Assumption 5 ***(general sustainability)*** $\hat{v}\left(\hat{C}, \hat{L}, \hat{K}\right) > 0$.

Moreover, when $\{C_t, L_t, K_t\} \to (C^*, L^*, K^*)$ almost surely, let $U_{C_t}^* = U_{C_t}(\{C_t, L_t\}_{t=0}^{\infty})$. Then, we have (proof in the Appendix):

Theorem 4 *Consider the model with the general environment and suppose that Assumption 5 holds.*

1. *At $t = 0$, there are aggregate distortions, so that the first Mirrlees principle fails to hold and there is positive aggregate capital taxation.*

   Let Γ 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 φ =sup$\{\ \varepsilon \in [0, 1] : p\lim_{t\to\infty} \varepsilon^{-t} U_{C_t}^* = 0\}$.

2. *If $\varphi \leq \delta$, then (almost surely) there are no asymptotic aggregate distortions, so that the first Mirrlees principle holds and there is no aggregate capital taxation.*

3. *If $\varphi > \delta$, then there are aggregate distortions and the first Mirrlees principle fails to hold even asymptotically.*

The major results from Theorem 3 continue to hold here.[29] The most important difference is that instead of comparing the discount factor of agents, β, to that of the government, δ, we now compare the rate at which the ex ante marginal utility of consumption in steady state, $U^*_{C_t}$, is declining to the discount factor of the government, δ. It can be verified that 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 is exactly equal to β, so that the results in this theorem are closely related to those in Theorem 3. Moreover, as long as a steady state exists, the rate at which $U^*_{C_t}$ declines is independent of δ, so it is always possible to choose δ to make sure that an economy is in part 2 or part 3 of this theorem.

The most important results here are 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 $\{C_t, L_t, K_t\}_{t=0}^{\infty}$ is asymptotically undistorted, which implies zero marginal tax on the labor supply of the highest type 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 δ 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. This is important, since the finding of zero capital taxes in the long run is a fairly robust finding in a large class of models (e.g., Chamley, 1986, Judd, 1985), and the analysis here shows that the presence of political economy constraints may easily reverse this conclusion.

Once again, we leave the proof of this theorem to the Appendix, and provide a heuristic argument here. The difficulty relative to the subsection 4.2 is that the objective function is no longer time separable and cannot be written in a simple recursive form. Instead, to characterize the best sustainable mechanism in this case, we follow Marcet and Marimon (1998) and form

[29] 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.

a Lagrangian of the form

$$\max_{\{C_t,L_t,K_t,x_t\}_{t=0}^{\infty}} \mathcal{L} = U(\{C_t, L_t\}_{t=0}^{\infty}) + \sum_{t=0}^{\infty} \delta^t \left\{\mu_t v(x_t) - (\mu_t - \mu_{t-1}) v(F(K_t, L_t))\right\} \tag{35}$$

subject to

$$C_t + x_t + K_{t+1} = F(K_t, L_t) \tag{36}$$

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).[30]

The differentiability of $U(\{C_t, L_t\}_{t=0}^{\infty})$ implies that for $\{C_t, L_t\}_{t=0}^{\infty} \in \mathrm{Int}\Lambda^{\infty}$, we have[31]

$$U_{L_t} - \delta^t (\mu_t - \mu_{t-1}) v'(F(K_t, L_t)) F_{L_t} = -U_{C_t} F_{L_t} \tag{37}$$

$$U_{C_t} = \left[U_{C_{t+1}} + \delta^t (\mu_{t+1} - \mu_t) v'(F(K_{t+1}, L_{t+1}))\right] F_{K_{t+1}} \tag{38}$$

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

$$U_{L_t} \geq -U_{C_t} F_{L_t}, \tag{39}$$

and

$$U_{C_t} \geq U_{C_{t+1}} F_{K_{t+1}}. \tag{40}$$

[30] To derive (35), form the Lagrangian

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

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 (35).

[31] To obtain these equations, let the multiplier on constraint (36) at time t be κ_t. Then the first-order condition with respect to C_t gives $U_{C_t} = \kappa_t$, while the first-order condition with respect to L_t gives

$$U_{Y_t} - \delta^t \left(\mu_t - \mu_{t-1}\right) v'(F(K_t, L_t)) F_{L_t} = -\kappa_t F_{L_t}.$$

Substituting for κ_t gives (37). The first-order condition with respect to K_{t+1}, on the other hand, gives

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

Substituting for κ_t and κ_{t+1} and rearranging gives (38).

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:

$$\frac{U_{C_t}}{\delta^t v'(x_t)} = \mu_t \leq \mu_{t+1} = \frac{U_{C_{t+1}}}{\delta^{t+1} v'(x_{t+1})}. \tag{41}$$

We know that, by construction, μ_t is an increasing sequence, so it must either converge to some value μ^* or go to infinity. Suppose that (C_t, L_t, K_t) converge 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 μ_t (almost surely) converges to some value $\mu^* < \infty$, and that both (39) and (40) 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 μ_t tends to infinity and aggregate distortions do not disappear.

4.5 Example for History-Dependent Mechanisms

In this subsection, we briefly illustrate the results of Theorem 4 and show how in some simple cases, φ defined as $\sup\{\ \varepsilon \in [0,1] : \text{plim}_{t\to\infty} \varepsilon^{-t} U^*_{C_t} = 0\}$ is again equivalent to the discount factor of the agents, β. 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

$$u(c, l \mid \theta) = u(c) - g(l/\theta),$$

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 π 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

$$U(\{C_t, L_t\}_{t=0}^{\infty}) \equiv \max_{\{c_t(\theta_0), c_t(\theta_1), l_t(\theta_1)\}_{t=0}^{\infty}} \pi \sum_{t=0}^{\infty} \beta^t \left[u\left(c_t\left(\theta_1\right)\right) - g\left(l_t\left(\theta_1\right)/\theta_1\right)\right] + (1-\pi) \sum_{t=0}^{\infty} \beta^t \left[u\left(c_t\left(\theta_0\right)\right)\right], \tag{42}$$

subject to

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

and

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

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 ε is assumed to converge to 0, we do not specify other incentive compatibility constraints.

Assigning Lagrange multipliers λ and $\beta^t \mu_t$ to these constraints, we can form the Lagrangian:

$$\begin{aligned}
\max_{\{c_t(\theta_0), c_t(\theta_1), l_t(\theta_1), x_t\}_{t=0}^{\infty}} \mathcal{L} \quad = \quad & \pi \sum_{t=0}^{\infty} \beta^t \left[u\left(c_t\left(\theta_1\right)\right) - g\left(l_t\left(\theta_1\right)/\theta_1\right)\right] + (1-\pi) \sum_{t=0}^{\infty} \beta^t \left[u\left(c_t\left(\theta_0\right)\right)\right] \\
& + \lambda \left\{ \sum_{t=0}^{\infty} \beta^t \left[u\left(c_t\left(\theta_1\right)\right) - g\left(l_t\left(\theta_1\right)/\theta_1\right)\right] - \sum_{t=0}^{\infty} \beta^t \left[u\left(c_t\left(\theta_0\right)\right)\right] \right\} \\
& + \beta^t \mu_t \left\{ \pi l_t\left(\theta_1\right) - x_t - \pi c_t\left(\theta_1\right) - (1-\pi)\, c_t\left(\theta_0\right) \right\}.
\end{aligned}$$

The first-order necessary conditions of this problem are

$$\begin{aligned}
(\pi + \lambda)\, u'\left(c_t\left(\theta_1\right)\right) &= \pi \mu_t && (43) \\
(1 - \pi - \lambda)\, u'\left(c_t\left(\theta_0\right)\right) &= (1-\pi)\, \mu_t,
\end{aligned}$$

and

$$\frac{(\pi + \lambda)}{\theta_1} g'\left(l_t\left(\theta_1\right)/\theta_1\right) = \mu_t. \tag{44}$$

Equations (43) imply that

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

Consequently, there is constant risk-sharing between the two types in all periods. Moreover, if a steady state exists, so that $x_t \to x^*$, (43) and (44) combined imply that $c_t(\theta_1) \to c^{1*}$, $c_t(\theta_0) \to c^{0*}$, and $l_t(\theta_1) \to l^*$, and hence $\mu_t \to \mu^*$. Finally, substituting these back into (42), we see immediately that $\varphi = \beta$, 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.[32]

Finally, to see the behavior of the economy more explicitly in this case, we further specialize the utility function to[33]

$$u(c, l \mid \theta) = \sqrt{c} - \frac{l^2}{10\theta}. \tag{45}$$

Let us also assume that $\pi = 1/2$, and suppose that the utility function of the government is also given by $v(x) = \sqrt{x}$. Figure 7 shows that the aggregate distortion, $1 - U_{Y_t}/U_{C_t}$, for the case , $\beta = \delta = 0.8$. Consistent with part 2 of Theorem 4, the aggregate distortion converges to zero, and the convergence is again rather fast.

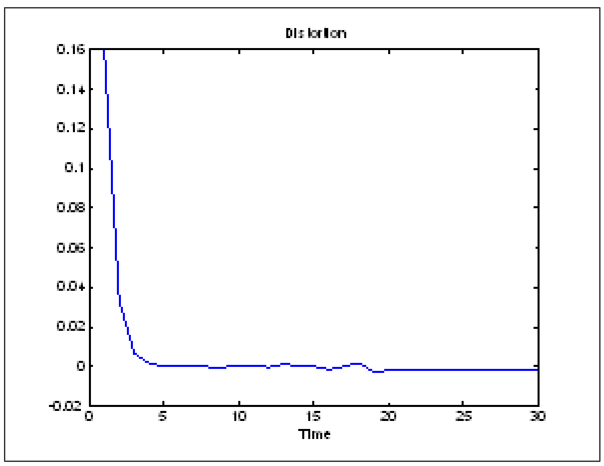

Figure 7: Theorem 4 part 2, almost constant types. The time path of distortions with $\beta = \delta = 0.8$.

Instead, when we consider the case where the government is less patient than agents, in particular, $\beta = 0.8$ and $\delta = 0.6$, once again consistent with Theorem 4, Figure 8 shows that the aggregate distortion, $1 - U_{Y_t}/U_{C_t}$, converges to a positive number, of about 10 percent.

[32] In fact, we conjecture that whenever there exists a station redistribution of consumption among individuals, $\varphi = \beta$, though we have not been able to prove this conjecture yet.

[33] The reason for choosing a utility function different from (4.3) is to illustrate the results more clearly in this case.

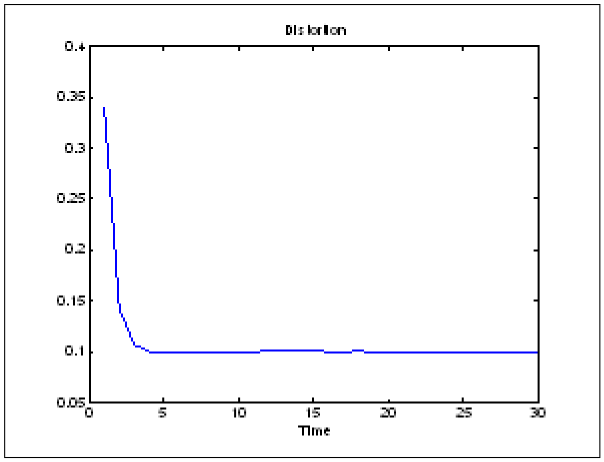

Figure 8: Theorem 4, part 3, almost constant types. The time path of distortions with $\beta = 0.8$ and $\delta = 0.6$.

4.6 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 ouput, 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.

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 12 is to allow the government to extract a fraction η of the total output. It can be verified that this assumption does not change anything substantial in our analysis.

Here, let us go one step further and suppose that the government can expropriate different amounts from different individuals, depending on 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:

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

where $l\left(\theta^t\right)$ is the labor supply of an individual with type history θ^t, and the term $\int \tilde{\eta}\left(l\left(\theta^t\right)\right) dG^t\left(\theta^t\right)$ 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 $G^t\left(\theta^t\right)$. 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.

5 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 will take the simplest conception of anonymous markets as one in which there is no insurance, and anonymity prevents any type of intervention by the government. Essentially, in anonymous markets, the benefits of insurance are forgone for the benefits of taking away the ability of the government to expropriates resources or misuse information.

5.1 Anonymous Market Equilibrium

As noted above, the anonymity of markets rules out any type of insurance mechanism, leaving self-insurance as the only option for individuals (when there is capital). Therefore, our model of anonymous markets is analogous to the Bewley-Aiyagari incomplete markets economy in which individuals can only insure against idiosyncratic risk by means of self insurance.

Let W_t be the equilibrium wage rate, R_t the rate of return on capital and denote individual i's capital holdings by k_t^i. Then the income of individual i at time t is

$$k_{t+1}^i = W_t l_t^i + R_t k_t^i - c_t^i. \tag{46}$$

Therefore, the market offers no insurance opportunities and individuals can only save in the form of a riskless bond, whose return is equal to the marginal product of capital, and we also assume that individual capital holdings cannot be negative, so $k_t^i \geq 0$.

Consequently, in anonymous markets, individuals take the sequence of prices $\{R_t, W_t\}$, as well as the stochastic process for θ_t^i as given and maximize their utility. The state variable for each individual is the history of his type $\theta^{i,t}$ and his or her capital holdings, k_t^i. Therefore, the individual maximization problem can be written as

$$\mathbf{u}\left(\theta^{i,t}, k_t^i\right) = \max_{\left\{c_{t+s}^i, l_{t+s}^i, k_{t+s+1}^i\right\}_{s=0}^{\infty}} \mathbb{E}\left[\sum_{s=0}^{\infty} \beta^s u\left(c_{t+s}^i, l_{t+s}^i \mid \theta_{t+s}^i\right) | \theta^{i,t}\right]$$

subject to (46) and $k_t^i \geq 0$. In this section, to simplify the discussion, let us assume that μ^∞ is such that θ_t^i is first-order Markov, so that we can write $\mathbf{u}\left(\theta^{it,}, k_t^i\right) = \mathbf{u}\left(\theta_t^i, k_t^i\right)$ as the value of an individual starting with type θ_t^i and capital holdings k_t^i.[34] Consequently, we can also express the maximizing arguments of this problem as $l_t^i\left(\theta_t^i, k_t^i\right)$ and $k_{t+1}^i\left(\theta_t^i, k_t^i\right)$.

Equilibrium prices are given by the marginal products of the production function (2). For this, we need to calculate aggregate levels of capital and labor. A dynamic equilibrium here is a complicated object, so let us focus on the stationary equilibrium (given Markov structure). For this reason, let us denote the stationary distribution of $\left(\theta_t^i, k_t^i\right)$ by H. Then we have:

$$K_t = \int k_t^i\left(\theta_{t-1}^i, k_{t-1}^i\right) dH\left(\theta_{t-1}^i, k_{t-1}^i\right) \tag{47}$$

and

$$L_t = \int l_t^i\left(\theta_t^i, k_t^i\right) dH\left(\theta_t^i, k_t^i\right), \tag{48}$$

for all t, where we have used the same distribution H in calculating capital stock and labor supply, since this stationary distribution applies both to $\left(\theta_{t-1}^i, k_{t-1}^i\right)$ and $\left(\theta_t^i, k_t^i\right)$. Given (47)

[34] In the absence of this assumption, the structure of the anonymous market equilibrium is more difficult to express, though none of the substantive conclusions of our analysis are affected.

and (48), factor prices are given by

$$R_t = \frac{\partial F(K_t, L_t)}{\partial K_t} \text{ and } W_t = \frac{\partial F(K_t, L_t)}{\partial L_t}. \tag{49}$$

A stationary anonymous market equilibrium is then given by a sequence of factor prices $\{R_t, W_t\}_{t=0}^{\infty}$ that satisfy (49), a stationary distribution of capital holdings and types given by $H\left(\theta_t^i, k_t^i\right)$, and a sequence of consumption, labor supply and saving decisions $\left\{c_t^i\left(\theta_t^i, k_t^i\right), l_t^i\left(\theta_t^i, k_t^i\right), k_{t+1}^i\left(\theta_t^i, k_t^i\right)\right\}_{t=0}^{\infty}$.

Theorem 5 *A stationary anonymous market equilibrium exists.*

The proof of this theorem is standard, and follows the structure in Aiyagari (1994) and is omitted here to save space.

Furthermore, to simplify the analysis, we assume that at $t = 0$, the economy starts with the stationary distribution of types and capital holdings, $H\left(\theta_t^i, k_t^i\right)$. Then, we can define the ex ante utility of anonymous market equilibrium as

$$\mathbf{U}^{AM} = \int \mathbf{u}\left(\theta_t^i, k_t^i\right) dH\left(\theta_t^i, k_t^i\right),$$

which can be interpreted as the utility of an individual in the stationary equilibrium before knowing his type and capital holdings (i.e., behind a veil of ignorance). Although behind the veil of ignorance comparisons are unattractive to make political economy claims, as a first step, it is useful to compare the behind a veil of ignorance utility of anonymous markets and sustainable mechanisms. $\mathbf{U}^{AM}$ is useful for this reason.

5.2 Anonymous Markets Versus Mechanisms

Our first comparative static result states that an increase in the discount factor of the government, δ, makes mechanisms more attractive relative to markets. This is stated and proved in the following proposition. Let $\mathbf{U}^{SM}(\delta)$ be the ex ante expected value of the best sustainable mechanism when the government discount factor is δ and $\mathbf{U}^{AM}$ be as defined above.

Proposition 3 *Suppose $\mathbf{U}^{SM}(\delta) \geq \mathbf{U}^{AM}$, then $\mathbf{U}^{SM}\left(\delta'\right) \geq \mathbf{U}^{AM}$ for all $\delta' \geq \delta$.*

Proof. Let $\mathcal{S}(\delta)$ be the feasible set of allocation rules for the problem when the government discount factor is equal to δ (meaning that they are feasible and also satisfy the sustainability constraint (13). Let $\{c_{i,t}(\delta), l_{i,t}(\delta), x(\delta)\} \in \mathcal{S}(\delta)$ represent the best sustainable mechanism. Since $\delta' \geq \delta$, we immediately have $\{c_{i,t}(\delta), l_{i,t}(\delta), x(\delta)\} \in \mathcal{S}(\delta')$, since, when the government's discount factor is δ', 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 $\mathbf{U}^{SM}(\delta)$ when the government's discount factor is δ'. This implies that $\mathbf{U}^{SM}(\delta')$ is at least as large as $\mathbf{U}^{AM}$, therefore $\mathbf{U}^{SM}(\delta') \geq \mathbf{U}^{SM}(\delta) \geq \mathbf{U}^{AM}$, completing the proof. ■

Let us next consider a modification of our main setup along the lines mentioned in footnote 12, whereby the government can consume only a portion of the output η. Let us consider the implications of better insututional constraints, which here correspond to a lower level of η (i.e., a lower fraction of output that the government can consume). It then immediately follows that better institutional controls on government make mechanisms more desirable relative to markets. For this proposition, define the value of the mechanism as $\mathbf{U}^{SM}(\eta)$, i.e., now as a function of the institutional restriction on the government. We then have:

Proposition 4 *Suppose* $\mathbf{U}^{SM}(\eta) \geq \mathbf{U}^{AM}$, *then* $\mathbf{U}^{SM}(\eta') \geq \mathbf{U}^{AM}$ *for all* $\eta' \leq \eta$.

The proof of this proposition is similar to the previous in this section and is omitted.

5.3 Risks and Mechanisms

We next want to show that when individuals become more risk averse, mechanisms become more desirable relative to markets. Unfortunately, we are unable to prove this result without further restrictions. In particular, to investigate the implications of greater risk aversion on the comparison of sustainable mechanisms to anonymous markets we will make three additional assumptions. First, we will limit ourselves to the economy with private histories and no capital. Second, we will assume that preferences take the quasi-linear form so that individuals maximize

$$\mathbb{E}\sum_{t=0}^{\infty}\beta^t u\left(c_t^i - g\left(\frac{l_t^i}{\theta_t^i}\right)\right). \tag{50}$$

Finally, it is important for the proof that $x_t = x$ for all t. For this reason, we now define and restrict attention to "second-best sustainable mechanisms", where the government is restricted to always receive the same amount, x.

It is evident from (50) that in the anonymous market equilibrium, optimal labor supply choices will satisfy $\bar{c}_i = \bar{l}_i$ such that

$$\frac{1}{\theta_i} g'\left(\frac{\bar{l}_i}{\theta_i}\right) = 1.$$

This expression clarifies the role of the quasi-linear preferences which make consumption and labor supply independent of the utility function. We denote the market equilibrium by $\{\bar{c}_i, \bar{l}_i\}$, and write the utility of an individual explicitly conditional on the utility function of agents, u as $\mathbf{U}^{AM}(u)$:

$$\mathbf{U}^{AM}(u) = \frac{1}{1-\beta} \sum_{i=0}^{N} \pi_i \left[u\left(\bar{l}_i - g\left(\frac{\bar{l}_i}{\theta_i}\right)\right)\right],$$

where we have used the fact that, ex ante each individual has a probability π_i of being type i in every period.

The restriction to second-best mechanisms implies that instead of the general sustainability constraint (13), we now have:

$$\frac{v(x)}{1-\delta} \geq v(L_t) \tag{51}$$

as the sustainability constraint.[35] Moreover, since individual allocations will not be time-varying either, the incentive compatibility constraints can be written as

$$c_i - g\left(\frac{l_i}{\theta_i}\right) \geq c_{i-1} - g\left(\frac{l_{i-1}}{\theta_i}\right) \tag{52}$$

for all $i = 1, 2, ..., N$.

Recall also that if a utility function $\tilde{u}$ is an increasing concave transformation of another one, u, then $\tilde{u}$ represents more risk-averse preferences than u. Let $\{c_i(u), l_i(u), x(u)\}$ be a solution when the utility function is u, and define let

$$b_i \equiv c_i(u) - g\left(\frac{l_i(u)}{\theta_i}\right)$$

and

$$\bar{b}_i \equiv \bar{c}_i - g\left(\frac{\bar{l}_i}{\theta_i}\right).$$

Before proving this proposition, we need the following straightforward lemma:

Lemma 3 $l_{it} \leq \bar{l}_{it}$.

[35] It is straightforward to verify that with this constraint, the limiting marginal taxes on the highest type will not go to zero (i.e., part 3 of Theorem 3 will not hold).

This lemma simply states that labor supply of all types will always be (weakly) lower under a sustainable mechanism than in the market equilibrium. The proof of this lemma is standard and is omitted.

Proposition 5 *Let us restrict attention to second-best mechanisms where $x_t = x$ for all t. Let $\tilde{u} = h(u)$ where $h(\cdot)$ is increasing and concave. Suppose $\mathbf{U}^{SM}(u) \geq \mathbf{U}^{AM}(u)$, then $\mathbf{U}^{SM}(\tilde{u}) \geq \mathbf{U}^{AM}(\tilde{u})$.*

Proof. Denote the solution to the program with utility functions u by $\{c_{it}(u), l_{it}(u), x_t(u)\}$, and recall that the market equilibrium is $\{\bar{c}_i, \bar{l}_i\}$. Also denote the set of allocation rules that are feasible for the second-best mechanism design problem when utility function is u by $\mathcal{S}(u)$. Note that (52) implies that if $\{c_i(u), l_i(u), x(u)\} \in \mathcal{S}(u)$ is a solution to this program, then $\{c_i(u), l_i(u), x(u)\} \in \mathcal{S}(\tilde{u})$, since both individual allocations remain feasible and (51) is still satisfied. We say that an allocation is u-preferred to another, if when preferences are given by u, the first allocation is feasible (sustainable) and gives greater ex ante utility. As in the previous proof, we will show that if $\{c_i(u), l_i(u), x(u)\}$ is u-preferred to $\{\bar{c}_i, \bar{l}_i\}$, then $\{c_i(u), l_i(u), x(u)\}$ is $\tilde{u}$-preferred to $\{\bar{c}_i, \bar{l}_i\}$. This would imply that since the solution to the second-best program with you to to function $\tilde{u}$, $\{c_i(\tilde{u}), l_i(\tilde{u}), x(\tilde{u})\}$, is by definition $\tilde{u}$-preferred to $\{c_i(u), l_i(u), x(u)\}$, it must also be $\tilde{u}$-preferred to $\{\bar{c}_i, \bar{l}_i\}$.

In fact, it is sufficient to prove that when $\{c_i(u), l_i(u), x(u)\}$ is u-indifferent to $\{\bar{c}_i, \bar{l}_i\}$, then $\{c_i(u), l_i(u), x(u)\}$ is $\tilde{u}$-preferred to $\{\bar{c}_i, \bar{l}_i\}$. So let us focus on this case where by hypothesis, we have

$$\sum_{i=0}^{N} \pi_i u(b_i(u)) = \sum_{i=0}^{N} \pi_i u(\bar{b}_i). \tag{53}$$

Now, we would like to prove that for any concave $h(\cdot)$, we have

$$\sum_{i=0}^{N} \pi_i h(u(b_i(u))) \geq \sum_{i=0}^{N} \pi_i h(u(\bar{b}_i)). \tag{54}$$

To accomplish this, let us define two new random variables $B_i = u(b_i(u))$ and $\bar{B}_i = u(\bar{b}_i)$. From (53), these two variables have the same mean. If, in addition, B_i is a mean-preserving spread of (i.e., second-order stochastically dominates) $\bar{B}_i$, then (54) follows for any concave h, and would prove the desired result.

Therefore, we only have to prove that B_i second-order stochastically dominates $\bar{B}_i$. To do this, recall the following equivalent characterization of second order stochastic dominance: if B_i and $\bar{B}_i$ have the same mean and their distribution functions intersect only once (with that of B_i cutting from below), then B_i second-order stochastically dominates $\bar{B}_i$.

Now consider the incentive compatibility constraint between types i and $i-1$. This constraint can either be binding or slack. First, suppose that it is binding. In that case, by adding and subtracting $g(l_{i-1}/\theta_{i-1})$ to the right-hand side, it can be written as:

$$c_i - g\left(\frac{l_i}{\theta_i}\right) = c_{i-1} - g\left(\frac{l_{i-1}}{\theta_{i-1}}\right) + g\left(\frac{l_{i-1}}{\theta_{i-1}}\right) - g\left(\frac{l_{i-1}}{\theta_i}\right).$$

Rearranging and using the definition of b, we obtain

$$b_i - b_{i-1} = g\left(\frac{l_{i-1}}{\theta_{i-1}}\right) - g\left(\frac{l_{i-1}}{\theta_i}\right) > 0$$

where the fact that this is greater than zero follows from $\theta_i > \theta_{i-1}$. Expanding $g\left(\frac{l_{i-1}}{\theta_i}\right)$ around $g\left(\frac{l_{i-1}}{\theta_{i-1}}\right)$:

$$b_i - b_{i-1} \simeq \frac{l_{i-1}}{\theta_{i-1}^2} g'\left(\frac{l_{i-1}}{\theta_{i-1}}\right)(\theta_i - \theta_{i-1}) > 0$$

Next, write the net consumption levels in the market equilibrium as

$$\bar{b}(\theta_i) = \bar{l}(\theta_i) - g\left(\frac{\bar{l}(\theta_i)}{\theta_{i-1}}\right)$$

and expand this around θ_{i-1} and use the envelope theorem to obtain:

$$\bar{b}(\theta_i) - \bar{b}(\theta_{i-1}) \simeq \frac{\bar{l}(\theta_{i-1})}{\theta_{i-1}^2} g'\left(\frac{\bar{l}(\theta_{i-1})}{\theta_{i-1}}\right)$$

Since from Lemma 3, $\bar{l}(\theta_{i-1}) \geq l_{i-1}$, we have

$$\bar{b}_i - \bar{b}_{i-1} \geq b_{it} - b_{i-1,t}. \tag{55}$$

Next, suppose that the incentive compatibility constraint between i and $i-1$ is slack at time t, this immediately implies $b_i = b_{i-1}$, so (55) is again satisfied.

Now, this observation combined with (53) implies that there exists some k such that

$$\begin{aligned} b_i &\geq \bar{b}_i \text{ for all } i \leq k \text{ and} \\ b_i &\leq \bar{b}_i \text{ for all } i > k, \end{aligned}$$

Since $u(\cdot)$ is strictly monotonic, the same applies to the ranking of B_i and $\bar{B}_i$. This implies that for all $\{\pi_i\}_{i=0}^{N}$, B_i second-order stochastically dominates $\bar{B}_i$, and completes the proof. ■

This proposition yields an important and intuitive result; it shows that when individuals become more risk averse, then they also become more willing to tolerate the costs of centralized mechanisms. It is important, however, to emphasize that a similar proof does not work with the best sustainable mechanism because payments to the government increase over time, so ranking of utilities in terms of second-order stochastic dominance becomes impossible. Based on this, we conjecture that there exist economies in which the best sustainable mechanism may become less preferable to the anonymous market allocation as individuals become more risk averse, but we have not confirmed this conjecture yet.

6 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 characterize 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.

A number of results of our analysis are important to note. 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 sustainable mechanisms. 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 insights from Mirrlees's 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.

The last part of the paper investigates when markets are likely to be less desirable relative to centralized mechanisms. Under some further simplifying assumptions, we show that greater risk aversion, more severe worst-case risks, a greater discount factor for the government and more severe limits on government expropriation make centralized mechanisms more desirable relative to markets despite the costs that they entail.

We view this paper as a first step in investigating political economy of mechanisms. There are both a number of technical and substantial issues left unanswered. First of all, the analysis in this paper has been confined to the most severe form of political economy risk, whereby the government can expropriate part of the output of the economy for its own consumption. It is important to investigate whether similar conclusions hold when there are different types of political economy considerations, for example electoral competition between politicians. Secondly, it would be interesting to compare centralized mechanisms not only to anonymous markets without any insurance, but also to decentralized insurance mechanisms, for example among sub-coalitions of individuals.

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 “mechanism design squared”, where the structure of incentives and institutions for governments and individuals are simultaneously determined. 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.

7 References

Abreu, Dilip (1988) “On the Theory of Repeated Games with Discounting” *Econometrica*, 56, 383-396.

Acemoglu, Daron (2005) “Modeling Inefficient Institutions” Prepared for the World Congress London 2005.

Acemoglu, Daron and Simon Johnson (2005) “Unbundling Institutions” forthcoming *Journal of Political Economy.*

Acemoglu, Daron, Simon Johnson and James A. Robinson (2004) “Institutions As the Fundamental Cause of Long-Run Economic Growth,” forthcoming *Handbook of Economic Growth* edited by Philippe Aghion and Steve Durlauf.

Acemoglu, Daron, Michael Kremer and Atif Mian (2003) “Incentives in Markets, Firms and Governments” NBER working paper No. 9802.

Aiyagari, S. Rao (1994) “Uninsured Idiosyncratic Risk and Aggregate Saving”, *Quarterly Journal of Economics* 109, 659-684

Albanesi, Stefania and Christopher Sleet (2005) “Dynamic Optimal Taxation with Private Information”, *Review of Economic Studies*, 72, 1-29

Atkeson, Andrew and Robert E. Lucas, Jr. (1992) “On efficient distribution with private information”, *Review of Economic Studies,* 59, 427-53.

Ausubel Lawrence and Raymond Deneckere (1989) “Reputation in Bargaining and Durable Goods Monopoly” *Econometrica*, 57, 511-531.

Austen-Smith, David, and Jeffrey S. Banks (1999) *Positive Political Theory I: Collective Preference.* Ann Arbor MI: University of Michigan Press.

Baron, David and Roger Myerson (1982) “Regulating a Monopolist with Unknown Costs “ *Econometrica*, 50, 911-30.

Benveniste, Lawrence and José Scheinkman (1979) “On the Differentiability of the Value Function in Dynamic Models of Economics” *Econometrica*, 47, 727-732.

Bertsekas, Dimitri, Angelia Nedic and Asuman Ozdaglar (2003) *Convex Analysis and Optimization*, Athena Scientific Boston.

Bester, Helmut and Richard Strausz (2001) “Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case.” *Econometrica*, 69, 1077-1098.

Bewley, Truman, F. (1977) "The Permanent Income Hypothesis: A Theoretical Formulation", *Journal of Economic Theory*, 16(2), 252-292.

Bisin, Alberto and Adriano Rampini (2005) "Markets as Beneficial Constraints on the Government", forthcoming *Journal of Public Economics*

Buchanan, James M. and Gordon Tullock (1962) *The Calculus of Consent.* Ann Arbor, MI: University of Michigan Press.

Chamley, Christophe (1986) "Optimal Taxation of Capital Income in General Equilibrium and Infinite Lives" *Econometrica*, 54, 607-622.

Chari, V.V. (2000) "Limits of Markets and Limits of Governments: An Introduction to a Symposium on Political Economy" *Journal of Economic Theory*, 94, 1-6.

Chari, V.V. and Patrick Kehoe (1990) "Sustainable Plans" *Journal of Political Economy*, 94, 783-802.

Chari, V.V. and Patrick Kehoe (1993) "Sustainable Plans and Mutual Default" *Review of Economic Studies*, 60, 175-195.

Dasgupta, Partha, Peter Hammond and Eric Maskin (1979) "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility", *Review of Economic Studies*, 46(2),185-216

Dixit, Avinash K. (2004) *Lawlessness and Economics: Alternative Modes of Governance*, Princeton; Princeton University Press.

Eggertsson, Thrainn (2005) *Imperfect Institutions: Possibilities and Limits of Reform*, University of Michigan Press, Ann Arbor.

Freixas, Xavier, Roger Guesnerie and Jean Tirole (1985) "Planning under Incomplete Information and the Wretched Effect" *Review of Economic Studies*, 52, 173-192.

Fudenberg, Drew, David Levine and Eric Maskin (1994) "The Folk Theorem with Imperfect Public Information" *Econometrica*, 62, 997-1039.

Golosov, Michael, Narayana Kocherlakota and Aleh Tsyvinski (2003) "Optimal Indirect and Capital Taxation", *Review of Economic Studies* 70, 569-587

Golosov, Michael and Aleh Tsyvinski (2004) "Designing Optimal Disability Insurance: A Case for Asset Testing", NBER Working Paper #10792

Green, Edward J. (1987) Lending and the smoothing of uninsurable income, in *Contractual Arrangements for Intertemporal Trade,* ed. E. Prescott and N. Wallace, Minneapolis:

University of Minnesota Press, 3-25.

Green Jerry and Jean Jacques Laffont (1977) "Characterization of Satisfaction Mechanisms for to Revelation of Preferences for Public Goods" *Econometrica*, 45, 427-38.

Harris, Milton and Robert M. Townsend (1981) "Resource Allocation Under Asymmetric Information", *Econometrica*, 49(1), 33-64.

Hart, Oliver, Andrei Shleifer and Robert Vishny (1997) "The Proper Scope of Government: Theory and an Application to Prisons," *Quarterly Journal of Economics*, November, 112(4), 1127-1161.

Hayek, Friedrich (1945) "The Use of Knowledge in Society" *American Economic Review*, 35, 519-30.

Holmstrom Bengt and Roger Myerson (1983) "Efficient and Durable Decision Rules with Incomplete Information" *Econometrica*, 51, 1799-1819.

Jones, Eric L. (1981) *The European Miracle: Environments, Economies, and Geopolitics in the History of Europe and Asia*, Cambridge University Press, New York.

Judd, Kenneth (1985) "Redistributive Taxation in a Simple Perfect Foresight Model" *Journal of Public Economics*, 28, 59-83.

Kocherlakota, Narayana (2005) "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation", *Econometrica*, 73(5), 1587-1621.

Luenberger, David G. (1969) *Optimization by Vector Space Methods*, John Wiley & Sons New York.

Marcet, Albert and Ramon Marimon (1998) "Recursive Contracts". Mimeo. University of Pompeu Fabra

Mas-Colell, Andreu, Michael D. Whinston and Jerry R. Green (1995) *Microeconomic Theory*, Oxford University Press, New York, Oxford

Miller, David (2005) "Attainable Payoffs in Repeated Games with Interdependent Private Information" University of California San Diego mimeo.

Mirrlees, James A. (1971) "An exploration in the theory of optimum income taxation", *Review of Economic Studies* 38, 175-208.

Myerson, Roger B. (1979) "Incentive Compatibility and the Bargaining Problem", *Econometrica*, 47(1), 61-73

North, Douglass C. (1981) *Structure and Change in Economic History*, W.W. Norton

& Co., New York.

North, Douglass C. and Robert P. Thomas (1973) *The Rise of the Western World: A New Economic History*, Cambridge University Press, Cambridge UK.

North, Douglass C. and Barry R. Weingast (1989) "Constitutions and Commitment: Evolution of Institutions Governing Public Choice in Seventeenth Century England, *Journal of Economic History*, 49, 803-832.

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

Parthasarathy, K. R. (1967) *Probity Measures on Metric Spaces*, Academic Press, New York.

Persson, Torsten and Guido Tabellini (2000) *Political Economics: Explaining Economic Policy*, The MIT Press, Cambridge MA

Phelan, Christopher (1994) "Incentives and Aggregate Shocks" *Review of Economic Studies*, 61, 681-700.

Phelan, Christopher and Ennio Stacchetti (2001) "Sequential Equilibria in a Ramsey Tax Model", *Econometrica*, 69(6), 1491-1518.

Pollard, David (2002) *A User's Guide to Measure Theoretic Probability*, Cambridge University press, New York.

Prescott, Edward C. and Robert Townsend (1984a) "General Competitive Analysis in an Economy with Private Information", *International Economic Review*, 25, 1-20

Prescott, Edward C. and Robert Townsend (1984b) "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard", *Econometrica* 52, 21-45.

Ray, Debraj (2002) "Time Structure of Self-Enforcing Agreements" *Econometrica*,70, 547-82.

Roberts, Kevin (1977) "Voting Over Income Tax Schedules" *Journal of Public Economics*, 8,329-40.

Roberts, Kevin (1984) "Theoretical Limits to Redistribution" *Review of Economic Studies*, 51, 177-195.

Rockafellar, R. Tyrrell (1970) *Convex Analysis*, Princeton University Press, Princeton

Segal, Ilya (2004) "The Communication Requirements of Social Choice Rules and Supporting Budget Sets" Stanford University mimeo.

Skreta, Vasiliki (2004) "Sequentially Optimal Mechanisms," mimeo University of Minnesota & Federal Reserve Bank of Minneapolis, and NAJ Economics, Vol. 4 - March 05, 2002 , http://www.najecon.org/v4.html.

Sleet, Christopher and Sevin Yeltekin (2004) "Credible Social Insurance" Carnegie Mellon mimeo.

Stokey, Nancy, Robert E. Lucas and Edward Prescott (1989) *Recursive Methods in Economic Dynamics*, Harvard University Press, Cambridge.

Thomas, Jonathan and Tim Worrall (1990) "Income Fluctuations and Asymmetric Information: An Example of Repeated Principle-Agent Problem", *Journal of Economic Theory* 51, 367-90.

Uhlig, Harald (1996) "A Law of Large Numbers for Large Economies" *Economic Theory*, 8, 41-50.

Werning, Ivan (2002) "Optimal Dynamic Taxation and Social Insurance", University of Chicago Ph.D. dissertation.

Appendix (Still Incomplete)

8 Technical Results

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

Definition 9 *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 1 *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 4 *Let X and Z be Banach spaces. Consider the maximization problem of*

$$P(u) = \max_{x \in X} f(x) \tag{56}$$

subject to

$$g_0(x) \leq u \tag{57}$$

and

$$G(x) \leq \mathbf{0} \tag{58}$$

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 $\mathbf{0}$ 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 μ be any multiplier of (57). Then μ 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 subgredient and is thus differentiable.

Lemma 5 *Let X and Z be Banach spaces. Consider the maximization problem of*

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

subject to

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

where $f : X \to \mathbb{R}$ *is a real-valued function and and* $G : X \to Z$ *is a vector-valued mapping and* $\mathbf{0}$ *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 2 *In this special case where* $G : \mathbb{R}^n \to \mathbb{R}^m$ *and* $f : \mathbb{R}^n \to \mathbb{R}$*, 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*

$$\begin{aligned} \nabla f(x_0) &= b^* \nabla G(x_0) \\ \nabla f(x_0) &= \tilde{b} \nabla G(x_0). \end{aligned}$$

Now combining these lemmas, we obtain:

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

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

subject to

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

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

9 Proofs for Section 4.2

In this section, we provide and prove a number of results used in the analysis of Section 4.2, ultimate the building up to the proof of Theorem 3. Throughout this section we assume that Assumption 4 is satisfied.

9.1 Properties of $U(C, L)$

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

$$U(C, L) \equiv \max_{\zeta \in \mathcal{P}} \sum_{\theta \in \Theta} \pi(\theta) \int u(c, l; \theta) \zeta(d(c, l), \theta) \tag{59}$$

subject to

$$\int u(c, l \mid \theta_i) \zeta(d(c, l), \theta_i) \geq \int u(c, l \mid \theta) \zeta(d(c, l), \theta_{i-1}) \text{ for all } i = 1, ..., N \tag{60}$$

$$\sum_{\theta \in \Theta} \pi(\theta) \int c \zeta(d(c, l), \theta) \leq C \tag{61}$$

$$\sum_{\theta \in \Theta} \pi(\theta) \int l \zeta(d(c, l), \theta) \geq L \tag{62}$$

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

Before deriving properties of the function $U(C, L)$, we need to ensure regularity. Let (60), (61) and (62) define the constraint mapping.

Lemma 6 *The solution to (59) is a regular point of the constraint mapping.*

Proof. The proof follows immediately from the fact that from single-crossing property (Assumption 3), all incentive compatibility constraints in (60) are linearly independent from each other, and also linearly independent from (61) and (62), thus the constraint mapping has full rank, $N + 2$, and is thus onto. ■

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

Lemma 7 *$U(C, L)$ is well-defined, continuous and concave on Λ, 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. Parthasarathy (1967) Theorem 6.4 p. 45 establishes that the set of probability measures defined over a compact metric space is compact in the weak topology. Since $\mathcal{C}$ is a compact subset of $\mathbb{R}^2$, $\mathcal{P}$ 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 (61) and nonincreasing in L, since a higher L tightens constraint (62).

Finally, to prove differentiability, note that the regularity condition is satisfied from Lemma 6 and moreover, the objective function in (59) 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 Λ are derived in the next lemma.

Lemma 8 *Λ is compact and convex.*

Proof. (convexity) Consider $(C^0, L^0), (C^1, L^1) \in \Lambda$ and some ζ^0, ζ^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 ζ^0, ζ^1 satisfy (60), (61) and (62), ζ^α 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 ζ^n be a feasible element for (C^n, L^n), and since $\mathcal{P}$ is compact under the weak topology, $\zeta^n \to \zeta^\infty \in \mathcal{P}$, which implies that ζ^∞ satisfies (60)-(62) and so ζ^∞ is feasible for (C^∞, L^∞), therefore Λ 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 9 $\mathbb{W} = [0, \bar{w}]$

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 way

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

subject to

$$C_t + x_t \leq L_t$$

$$v(x_t) + \delta w_{t+1} \geq v(L_t)$$

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

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. Let now $\mathcal{C} = (C, L, x, w) \in \mathbb{R}^4$ and let $\mathcal{Z}$ be the set of Borel subsets of $\mathcal{C}^\infty$. Then let the triple $(\mathcal{C}^\infty, \mathcal{Z}, \bar{\mu})$ be a probability space, and let $\{\mathcal{Z}_t : t \in \mathbb{N}\}$ be a filtration defined over $\mathcal{C}^t$, whereby naturally $\mathcal{Z}_t \subseteq \mathcal{Z}_{t+1}$. Let $\mathcal{P}^\infty$ be the space of probability measures on $\mathcal{C}^\infty$ endowed with the weak topology. We will think of randomization as the realization of a random variable ω^t measurable with respect to $\mathcal{Z}_t$ at time t. Therefore, for each realizations of the random variable ω^t measurable with respect to $\mathcal{Z}_t$ the constraints take the form (with all the constraints holding $\mathcal{Z}_t$-almost-surely):

$$C_t(\omega^t) + x_t(\omega^t) \leq L_t(\omega^t)$$

$$v(x_t(\omega^t)) + \delta E\{w_{t+1}|\omega^t\} \geq v(L_t(\omega^t))$$
$$E\left\{\sum_{s=0}^{\infty} \delta^s v(x_{t+s}\left(\omega^t\right))|\mathcal{Z}_t\right\} = w_t$$

Analogously to the quasi-Mirrlees problem, we define probability measures $\xi^\infty \in \mathcal{P}^\infty$ on the set of all (C, L, x, w) tuples. However, we would like to exploit the recursive formulation of this problem. For this, let $\mathcal{P}(w)$ be the space of probability measures after a level of utility w has been promised to the government. In that case, the problem can be written recursively as follows

Problem A1

$$V(w) = \max_{\xi \in \mathcal{P}(w)} \int [U(C, L) + \beta V(w'(\omega))] d\xi(\omega) \tag{63}$$

subject to

$$C(\omega) + x(\omega) \leq L(\omega) \text{ } \xi\text{-almost-surely} \tag{64}$$

$$v(x(\omega)) + \delta w'(\omega) \geq v(L(\omega)) \text{ } \xi\text{-almost-surely} \tag{65}$$

$$w = \int [v(x(\omega)) + \delta w'(\omega)] d\xi(\omega) \tag{66}$$

and

$$(C, L) \in \Lambda \text{ and } w' \in \mathbb{W} \text{ } \xi\text{-almost-surely.} \tag{67}$$

We can now establish:

Lemma 10 *$V(w)$ is concave.*

Proof. Consider any w_0 and w_1 and ξ_0 and ξ_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 (64) and (65) hold state by state, and satisfied for both ξ_0 and ξ_1, and therefore must be satisfied for ξ_α. Constraint (66) is linear in ξ, therefore ξ_α also satisfies this constraint. Since the objective function is linear in ξ_α, $V(\alpha w_0 + (1-\alpha)w_1) \geq \alpha V(w_0) + (1-\alpha)V(w_1)$, establishing the concavity of V. ■

The above lemma establishes the concavity of V using arbitrary randomizations in the maximization problem (63). The next lemma shows that a particularly simple form of randomization is sufficient to achieve the maximum of (63).

Lemma 11 *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 ξ_0 and $1-\xi_0$.*

Proof. First suppose 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, ξ 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.

Next, suppose that ξ assigns positive probability to a subset $\bar{\mathcal{C}}$ of $\mathcal{C}$, but zero probability to any $(C, L, x, w') \in \bar{\mathcal{C}}$ (i.e., is non-atomic over this set). Now the same argument implies that all $(C, L, x, w') \in \bar{\mathcal{C}}$ must yield the same utility, and again randomizing between two points in this set is sufficient to achieve the optimum. ■

Lemma 11 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 11, we can focus on the problem equivalent to Problem A1:

Problem A2:

$$V(w) = \max_{\{\xi_i, C_i, L_i, x_i, w_i'\}_{i=0,1}} \sum_{i=0,1} \xi_i \left[U(C_i, L_i) + \beta V(w_i')\right] \tag{68}$$

subject to

$$C_i + x_i \leq L_i \text{ for } i = 0, 1 \tag{69}$$

$$v(x_i) + \delta w_i' \geq v(L_i) \text{ for } i = 0, 1 \tag{70}$$

$$w = \sum_{i=0,1} \xi_i \left[v(x_i) + \delta w_i'\right]. \tag{71}$$

$$(C_i, L_i) \in \Lambda \text{ for } i = 0, 1 \text{ and } w' \in \mathbb{W} \tag{72}$$

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 12 *$V(w)$ is differentiable.*

Proof. Let the maximizer for $w'(\omega \mid w_0)$ in Problem A1 when $w = w_0$ be $w'(\omega \mid w_0)$. Then consider the alternative maximization problem

$$W(w) = \max_{\xi \in \mathcal{P}(w)} \int [U(C, L) + \beta V(w'(\omega \mid w_0))] d\xi(\omega) \tag{73}$$

subject to (64), (65) and

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

By the same argument as in Lemma 10, $W(w)$ is concave (recall that the proof of Lemma 10 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 11 establishes that $W(w)$ can be equivalently characterized by the following maximization problem:

Problem A3:

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

$$C_i + x_i \leq L_i \text{ for } i = 0, 1 \tag{74}$$

$$v(x_i) + \delta w_i'(w_0) \geq v(L_i) \text{ for } i = 0, 1 \tag{75}$$

$$w = \sum_{i=0,1} \xi_i \left[v(x_i) + \delta w_i'(w_0) \right]. \tag{76}$$

$$(C_i, L_i) \in \Lambda \text{ for } i = 0, 1 \text{ and } w' \in \mathbb{W} \tag{77}$$

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

$$W(w) \leq V(w) \tag{78}$$

and

$$W(w_0) = V(w_0) \tag{79}$$

by construction.

>From Lemma 10 $V(w_0)$ is concave, and therefore $-V$ is convex. Convex functions have well-defined subdifferentials (see Rockafellar, 1970, Chapter 23 or Bertsekas, Nedic and Ozdaglar, 2003, Chapter 4). In particular, if f is convex, there exists a closed, convex and nonempty set ∂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 (78) and (79). This implies that v is also a subgradients 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 13 *Suppose $(\bar{C}, \bar{L}) \in \Lambda$ is such that the utility of the government is $\bar{w} = v(\bar{L} - \bar{C})/(1-\delta)$ and the sustainability constraint (21) is satisfied with a Lagrange multiplier $\psi = 0$, then*

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

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

$$\bar{w} \equiv \arg \max_{(C,L)\in\Lambda} v(L - C)$$

such that $\hat{v}(C, L) \geq 0$. 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 θ_N, $c(\theta_N)$ and $l(\theta_N)$, which are

$$\begin{aligned} u_c(c(\theta_N), l(\theta_N) \mid \theta_N)(1 + \lambda_N) &= \nu_C \\ u_l(c(\theta_N), l(\theta_N) \mid \theta_N)(1 + \lambda_N) &= -\nu_L \end{aligned}$$

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

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

as desired. ■

9.2 Proof of Theorem 3

Proof. Since V is differentiable from Lemma 7 and concave from Lemma 10, the first-order conditions are necessary and sufficient for the maximization (68). Assign the multipliers $\xi_i \kappa_i$

to the constraints in (69), $\xi_i \psi_i$ to those in (70) and γ to constraint (71), and let $V'(w)$ be the derivative of $V(w)$ at w, we have

$$\begin{aligned} \beta \xi_0 V'\left(w_0'\right) + \delta \psi_0 \xi_0 + \delta \gamma \xi_0 &\leq 0 \\ \beta \xi_1 V'\left(w_1'\right) + \delta \psi_1 \xi_1 + \delta \gamma \xi_1 &\leq 0 \end{aligned}$$

with both equations holding as equality for $w_i' \in$Int$\mathbb{W}$. Therefore,

$$\frac{\beta}{\delta} V'\left(w_i'\right) \leq -\psi_i - \gamma, \tag{80}$$

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

$$V'(w) \geq -\gamma \tag{81}$$

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

In addition, combining first-order conditions we have that for $(C, L) \in$IntΛ,

$$U_C(C_i, L_i) + U_L(C_i, L_i) = \psi_i v'(L_i) \text{ for } i = 0, 1. \tag{82}$$

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 (80) 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 (70) 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 $\mathcal{W}_t = \{w_{0,t}, w_{1,t}\}$. Since $\beta \leq \delta$ and $V'(w) \leq 0$, (80) implies

$$V'\left(w_i'\right) \leq -\psi_i - \gamma.$$

Combining this with (81) yields

$$V'(w) \geq V'\left(w'\right).$$

Concavity of V then implies that $w_i' \geq w$ 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 $\text{plim}\, w_t = w^*$. This immediately implies $\text{plim}\, x_t = x^*$. Therefore $\text{plim}\, C_t = C^*$ and $\text{plim}\, L_t = L^*$ are feasible (given $\text{plim}\, x_t = x^*$) and are optimal from the concavity of $U(C, L)$, this is a solution to the maximization in (68), 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 $\beta = \delta$ and $w^* < \bar{w}$. Then we must have $V'(w^*) \leq -\psi_i - \gamma$ and from (81), $V'(w^*) \geq -\gamma$, which is only possible if $\psi_i = 0$ for $i = 0, 1$ (recall that $\psi_i \geq 0$), establishing the claim that the sustainability constraints, (70), become slack asymptotically. Moreover, $w^* < \bar{w}$ implies that $(C^*, L^*) \in \text{Int}\Lambda$ (since $\bar{w} = v\left(\bar{L} - \bar{C}\right)/(1-\delta)$ and $\left(\bar{C}, \bar{L}\right) \in \text{Bd}\Lambda$, $w^* = v\left(L^* - C^*\right)/(1-\delta)$ for some (C^*, L^*) with $C^* \geq \bar{C}$ and $L^* \leq \bar{L}$, with at least one of them as a strict inequality, and thus $(C^*, L^*) \in \text{Int}\Lambda$), equation (82) combined with the fact that $\psi_i \to 0$ for $i = 0, 1$ implies

$$U_C(C^*, L^*) + U_L(C^*, L^*) = 0, \tag{83}$$

as desired. Next suppose that $w^* = \bar{w}$. In this case, the sustainability constraint, (70), is still slack, so that $\psi_i \to 0$ for $i = 0, 1$. Then (83) follows from Lemma 13.

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

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

Recall that, as established above, $\{w_{i,t}\} \uparrow w^*$. First, 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 $\{w_{i,t}\} \uparrow \bar{w}$. Now consider two cases. First suppose that $\{w_{i,t}\}$ converges to $\bar{w}$ in finite time, say by time T. This implies that for all $t \geq T$, the sustainability constraints after time T are identical to that for time T, and thus have zero Lagrange multipliers as claimed. Then applying Lemma 13 yields (83) as desired.

Second, suppose that $\{w_{i,t}\}$ does not converge to $\bar{w}$ in finite time. But then since $\{w_{i,t}\}$ is a convergent sequence, it is also a Cauchy sequence, and $\lim_{t\to\infty} |w_{i,t+1} - w_{i,t}| = 0$ for $i = 0, 1$. Since $\beta < \delta$, this implies that there must exist some T, such that for all $t \geq T$, $\psi_{i,t} = 0$ for $i = 0, 1$, establishing that the sustainability constraints, (70), are slack after T as claimed. Finally, in this case we must still have that $\{w_{i,t}\} \uparrow \bar{w}$ by the above argument, so $\psi = 0$ combined with Lemma 13 yields (83) as desired. The rest of part 2 follows from Corollary 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. ■

10 Proofs of Section 4.4

In this section of the Appendix, we provide the proofs for the more general environment.

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

As in the above proof, to show concavity and differentiability of $U\left(\{C_t, L_t\}_{t=0}^{\infty}\right)$, we introduce lotteries following Prescott and Townsend (1984a, 1984b). Now the lotteries are more complicated objects, since they are conditional on the entire history of reports. Let $\mathcal{C} = \{(c,l) \in \mathbb{R}^2 : 0 \le c \le \bar{c},\ 0 \le l \le \bar{l}\}$ be the set of possible consumption-labor allocations for agents. For each $t \in \mathbb{N}$ and $\theta^{t-1} \in \Theta^{t-1}$, let $\mathcal{P}\left[\theta^{t-1}\right]$ be the space of $N+1$-tuples of probability measures on Borel subsets of $\mathcal{C}$ for an individual with history of reports θ^{t-1}. Thus each element $\zeta\left(\cdot \mid \theta^{t-1}\right) = [\zeta(\theta_0 \mid \theta^{t-1}), ..., \zeta(\theta_N \mid \theta^{t-1})]$ in a $\mathcal{P}^t\left[\theta^{t-1}\right]$ consists of $N+1$ probabilities measures for each type θ_i given their past reports, θ^{t-1}. We again endow each space $\mathcal{P}^t\left[\theta^{t-1}\right]$ with the weak topology for each θ^{t-1}. Let $\mathcal{P} \equiv \bigcup_{t\in\mathbb{N}} \bigcup_{\theta^t\in\Theta^t} \mathcal{P}^t\left[\theta^{t-1}\right]$.

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

$$U\left(\{C_t, L_t\}_{t=0}^{\infty}\right) \equiv \max_{\left\{\zeta^t\left(\cdot|\theta^{t-1}\right)\in\mathcal{P}^t\left[\theta^{t-1}\right]\right\}_{t=0}^{\infty}} \mathbb{E}\sum_{t=0}^{\infty}\beta^t \int u(c,l;\theta_t)\zeta^t\left(d(c,l),\theta_t \mid \theta^{t-1}\right) \tag{84}$$

subject to

$$\mathbb{E}\left[\sum_{s=0}^{\infty}\beta^s \int u(c,l;\theta)\zeta^{t+s}(d(c,l),\theta \mid \theta^{t+s-1})|\theta^{t-1}\right] \ \ge\ \mathbb{E}\left[\sum_{s=0}^{\infty}\beta^s \int u(c,l;\theta)\zeta^{t+s}(d(c,l),\hat{\theta} \mid \theta^{t+s-1})|\theta^{t-1}\right]$$
$$\text{for all } \theta, \hat{\theta} \text{ and any } \theta^{t-1} \tag{85}$$

$$\int \sum_{\theta\in\Theta}\pi(\theta)\int c\zeta^t(d(c,l),\theta \mid \theta^{t-1})dG^{t-1}\left(\theta^{t-1}\right) \le C_t \tag{86}$$

$$\int \sum_{\theta \in \Theta} \pi(\theta) \int l\zeta^t(d(c,l), \theta \mid \theta^{t-1}) dG^{t-1}\left(\theta^{t-1}\right) \geq L_t \tag{87}$$

for all t.

The key lemma is a generalization of Lemma 7, which is stated here.

Lemma 14 *$U(\{C_t, L_t\}_{t=0}^{\infty})$ is continuous and concave on Λ^{∞}, 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 7, except that in this infinite-dimensional space we are no longer able to prove Lemma 6. Thus, all the proofs assume that the solution is at a regular point.

Proof. The constraint set given by (85)-(87) is compact in the weak topology and the objective function (84) is continuous. Therefore, by a generalized version of Weierstrass' theorem applies to normed linear spaces, a maximum exists and $U(\{C_t, L_t\}_{t=0}^{\infty})$ is a well defined.

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

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

In a similar way we can show that ζ^{α} satisfies (86) and (87). So the 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 15 *Λ 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 ζ^0, ζ^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 ζ^0, ζ^1 satisfy (60), ζ^{α} satisfies it as well. Similarly, ζ^{α} satisfies (61) and (62).

(compactness) For any sequence $\{C_t^n, L_t^n\}_{t=0}^{\infty} \in \Lambda^{\infty}, \{C_t^n, L_t^n\}_{t=0}^{\infty} \to \{C_t^{\infty}, L_t^{\infty}\}_{t=0}^{\infty}$, there will be ζ^n of elements in $\{C_t^n, L_t^n\}_{t=0}^{\infty}$, such that $\zeta^n \to \zeta^{\infty}$, satisfying (60)-(62) and feasibility, therefore $\{C_t^{\infty}, L_t^{\infty}\}_{t=0}^{\infty} \in \Lambda^{\infty}$ is closed. Boundedness follows from boundedness of C and L. ■

The lemmas regarding concavity of V and randomization over two points also generalize, and we omit them to save space. Finally, we have a generalization of Lemma 13, which we state without proof:

Lemma 16 *Suppose* $\left\{\bar{C}, \bar{L}\right\} \in \Lambda^{\infty}$ *and* $\bar{K}$ *are such that the utility of the government is* $\bar{w} = v\left(F(\bar{K}, \bar{L}) - \bar{C}\right) / (1-\delta)$ *and the sustainability constraint (13) is satisfied with a Lagrange multiplier* $\psi = 0$*, then*

$$U_{C_t}\left(\bar{C}, \bar{L}, \bar{K}\right) \cdot F_{L_t}\left(\bar{K}, \bar{L}\right) = -U_{L_t}\left(\bar{C}, \bar{L}, \bar{K}\right) \text{ and } F_{K_{t+1}}\left(\bar{K}, \bar{L}\right) \cdot U_{C_{t+1}}\left(\bar{C}, \bar{L}, \bar{K}\right) = U_{C_t}\left(\bar{C}, \bar{L}, \bar{K}\right).$$

10.2 Proof of Theorem 4

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 (37) and (38) hold. Rearranging these equations and substituting for $U_{C_t}(C^*, L^*)$, we have

$$-\frac{U_{L_t}^*}{U_{C_t}^* F_{L_t}(K^*, L^*)} = 1 - \frac{(\mu_t - \mu_{t-1}) v'(F(K^*, L^*)))}{\mu_t v'(x^*)} \tag{88}$$

and

$$\frac{F_{K_{t+1}}(K^*, L^*)\, U_{C_{t+1}}^*}{U_{C_t}^*} = 1 + \frac{(\mu_{t+1} - \mu_t) v'(F(K^*, L^*)) F_{K_{t+1}}(K^*, L^*)}{\mu_t v'(x^*)}, \tag{89}$$

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 (41) also holds and implies that as $t \to \infty$,

$$\frac{U_{C_t}^*}{\delta^t v'(x^*)} = \mu_t \leq \mu_{t+1} = \frac{U_{C_{t+1}}^*}{\delta^{t+1} v'(x^*)}. \tag{90}$$

First suppose that $\varphi = \delta < 1$ where $\varphi = \sup\{\varepsilon \in [0,1] : \text{plim}_{t\to\infty} \varepsilon^{-t} U_{C_t}^* = 0\}$ as defined in the theorem. This implies that as $t \to \infty$, $U_{C_t}^*$ is proportional to φ^t. Therefore, μ_t must converge (almost surely) to $\mu^* < \infty$, thus $(\mu_t - \mu_{t-1})/\mu_t \to 0$ almost surely, and from (88) and (89), 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 (90) 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 $(\bar{C}, \bar{L}, \bar{K})$) or to $\psi = 0$ for the constraints (13) in finite time. This combined with Lemma 16 and the fact that the sustainability constraint (13) holds for all $t \geq T$ establishes the desired result.

Part 3: Suppose that $\varphi > \delta$. In this case, (90) 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 (88) and (89), aggregate distortions cannot disappear, completing the proof. ■