Fiscal Rules and Discretion under Persistent Shocks

Marina Halac and Pierre Yared

No 1014

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF WARWICK

Fiscal Rules and Discretion under Persistent Shocks*

Marina Halac[†] Pierre Yared[‡]

November 9, 2012

Abstract

This paper studies the optimal level of discretion in policymaking. We consider a fiscal policy model where the government has time-inconsistent preferences with a present-bias towards public spending. The government chooses a fiscal rule to trade off its desire to commit to not overspend against its desire to have flexibility to react to privately observed shocks to the value of spending. We analyze the optimal fiscal rule when the shocks are persistent. Unlike under i.i.d. shocks, we show that the ex-ante optimal rule is not sequentially optimal, as it provides dynamic incentives. The ex-ante optimal rule exhibits history dependence, with high shocks leading to an erosion of future fiscal discipline compared to low shocks, which lead to the reinstatement of discipline. The implied policy distortions oscillate over time given a sequence of high shocks, and can force the government to accumulate maximal debt and become immiserated in the long run.

Keywords: Institutions, Asymmetric and Private Information, Macroeconomic Policy, Structure of Government, Political Economy

JEL Classification: D02, D82, E6, H1, P16

---

*We would like to thank Manuel Amador, Sylvain Chassang, Navin Kartik, Ciamac Moallemi, Emi Nakamura, Ali Shourideh, Paolo Siconolfi, Jon Steinsson, Bruno Strulovici, Steve Tadelis, Aleh Tsyvinski, Iván Werning, Steve Zeldes, seminar participants at Brown, Columbia, Duke, Minneapolis Fed, New York Fed, Princeton, and Yeshiva, and participants at the NBER SI 2012 Political Economy Public Finance workshop and the SED conference for comments. Sergey Kolbin provided excellent research assistance.

[†]Columbia University, Graduate School of Business. email: mhalac@columbia.edu.

[‡]Columbia University, Graduate School of Business and NBER. email: pyared@columbia.edu.

1 Introduction

Governments often impose rules on themselves to constrain their behavior in the future. One of the most prevalent form of such rules are fiscal rules, typically adopted in response to rising public debts. In 2009, 80 countries had fiscal rules in place, a dramatic increase from 1990 when only seven countries had them.[1]

Despite their prevalence, little is known about the optimal structure of fiscal rules. How much discipline does an optimal fiscal rule impose? How does this level of discipline evolve over time? And does an optimal rule restrict the growth of debt in the long run? Any theory of fiscal rules must take into account a fundamental tradeoff between commitment and flexibility: on the one hand, rules provide valuable commitment as they can limit distorted incentives in policymaking which result in a spending bias and excessive deficits; on the other hand, there is a cost of reduced flexibility as fiscal constitutions cannot spell out policy prescriptions for every single shock or contingency, and some discretion may be optimal.[2]

This paper studies the tradeoff between commitment and flexibility in a dynamic self-control setting. In our model, a present-biased government privately observes a shock to the economy in each period, and a fiscal rule is a mechanism that assigns a policy as a function of the government's reported shocks. We follow a similar approach to that used in Amador, Werning, and Angeletos (2006), but we depart from their work by considering an environment in which shocks are persistent over time.[3] We are motivated by the fact that shocks underlying fiscal policy are likely to be autocorrelated, consistent with the observation that fiscal policy variables are persistent in the data.[4] As is well known, persistence introduces new difficulties into the mechanism design problem; we consider a simple framework that allows for a full characterization of the optimal mechanism.[5]

Our environment is a small open economy in which the government makes repeated spending and borrowing decisions. In each period, a shock to the social value of deficit-financed

[1]See International Monetary Fund (2009). Of those 80 countries, 60 had a deficit limit, 60 a debt limit, and 25 a spending limit. These limits vary in their specification and the extent to which they adjust to levels of GDP and to the business cycle. Moreover, these limits also vary in the degree to which they are enforced.

[2]The importance of commitment devices in policymaking dates back to Kydland and Prescott (1977). In terms of fiscal policy, there are various micro-foundations for how a deficit bias can emerge in a political economy environment. See, among others, Aguiar and Amador (2011), Alesina and Perotti (1994), Alesina and Tabellini (1990), Battaglini and Coate (2008), Caballero and Yared (2010), Lizzeri (1999), Persson and Svensson (1989), and Tornell and Lane (1999).

[3]See also Amador and Bagwell (2011), Ambrus and Egorov (2012), and Athey, Atkeson, and Kehoe (2005). All these papers consider settings with i.i.d. shocks.

[4]For example, see Barro (1990) for evidence.

[5]Persistence complicates the mechanism design problem because single-crossing conditions generally used in the analysis may fail and a recursive representation is more difficult. Recent work addresses this issue in principal-agent and optimal taxation settings (see, e.g., Battaglini and Lamba, 2012, Farhi and Werning, 2010, Golosov, Troshkin, and Tsyvinski, 2011, and Pavan, Segal, and Toikka, 2010), but this analysis does not apply to a self-control setting, which is our interest.

government spending is realized, where this shock follows a first-order Markov process. We consider two shocks in our benchmark setting and a continuum of shocks in an extension. The government in each period is benevolent ex ante, prior to the realization of the shock, but present-biased ex post, when it is time to choose fiscal policy.[6] The shocks are privately observed by the government, capturing the fact that not all contingencies in fiscal policy are contractible or observable. A fiscal rule is defined as a mechanism in which the government reports the shock in each period and is assigned a policy for every reported shock. Note that in the absence of private information, the first-best policy could be implemented with full commitment; i.e. by giving the government no discretion and committing it to the efficient path of spending. Similarly, absent a present bias, the first best could be implemented with full flexibility; i.e. by giving the government full discretion to choose spending in each period. In the presence of both private information and a present bias, however, a tradeoff between commitment and flexibility arises, and the optimal rule is then not trivial.

We study the ex-ante optimal fiscal rule and the sequentially optimal fiscal rule. The ex-ante optimal rule is the optimal dynamic mechanism that the government chooses at the beginning of time. This is a sequence of spending and borrowing levels as a function of the history of the government's reports, which maximizes ex-ante social welfare subject to a sequence of dynamic incentive compatibility constraints. The sequentially optimal rule, on the other hand, corresponds to a static mechanism that is chosen by the government in every period t (prior to the realization of the shock), which maximizes social welfare from t onward taking into account that future governments do the same.

Our motivation for studying sequentially optimal rules is twofold. First, in practice, fiscal rules often have a bite in the short term for the current fiscal year, but can be renegotiated and changed by the government in advance for the following year. Second, a central result from the previous literature is that the ex-ante optimal and sequentially optimal rules coincide when the shocks are i.i.d.[7] That is, under the ex-ante optimal mechanism, at any given date (prior to the realization of the shock) the government would not want to change the continuation mechanism. The reason is that no dynamic incentives are provided, and the mechanism at any date depends only on the payoff-relevant states. Notably, this contrasts with the results of the principal-agent literature, where, even under i.i.d. shocks, dynamic incentives are provided to an agent with private information along the equilibrium path. The difference is due to the fact that there is a single player in a self-control setting, and dynamic incentives cannot be provided by increasing one player's welfare at the expense of another. Thus, under i.i.d. shocks, dynamic incentives would affect the government's welfare on the equilibrium path (when reporting the shock truthfully) and off the equilibrium path (when misreporting the

[6] The preference structure corresponds to the quasi-hyperbolic consumption model; see Laibson (1997).

[7] See Amador, Werning, and Angeletos (2006) and Athey, Atkeson, and Kehoe (2005).

shock) equally, and therefore result in no welfare gains.

We first show that, as in the case of i.i.d. shocks, the sequentially optimal fiscal rule under persistent shocks is history independent and simple, as by definition dynamic incentives are not provided to the government. Both under two shocks as well as under a continuum of shocks, the sequentially optimal mechanism can be implemented with a debt limit that is a function of only the payoff-relevant states, namely the current level of debt and the previous period's shock which forecasts the current shock.

Our main result, however, shows that unlike under i.i.d. shocks, the ex-ante optimal rule does not coincide with the sequentially optimal rule when shocks are persistent. The ex-ante optimal rule takes into account that the government in every period learns about its current and future spending needs, and now provides dynamic incentives for the government not to overspend and overborrow. Specifically, consider a government that is tempted to overspend today when its needs are low. Dynamic incentives can be provided by introducing excessively lax and ex-post suboptimal rules tomorrow if the government chooses high levels of spending today. The expected cost to the government of such lax rules tomorrow is greater if spending needs are actually low today, as spending needs are then more likely to be low tomorrow. Thus, lax rules tomorrow affect welfare on and off the equilibrium path differently, and the threat of no discipline in the future ironically imposes discipline today.

We characterize the ex-ante optimal rule and the dynamics that it induces in fiscal policy. We show that this rule is history dependent: the mechanism at a given date is a function of not only the payoff-relevant states, but also the entire history of shocks. History dependence follows from the fact that the mechanism provides dynamic incentives. Intuitively, because the shock at date $t-2$ predicts the realization of the shock at $t-1$ (since shocks follow a first-order Markov process), the shock at $t-2$ affects the relative tightness of incentive compatibility constraints at $t-1$, which in turn affect the policies that are chosen at t in providing dynamic incentives at $t-1$. In particular, the shock at date $t-2$ tells us how likely it is that the shock to the value of spending will be low at $t-1$, and thus how beneficial it is to make rules excessively lax at t following high spending by the government at $t-1$.

We explicitly characterize the dynamics implied by the ex-ante optimal fiscal rule in an infinite horizon economy using a recursive technique similar to that developed by Fernandes and Phelan (2000) for a principal-agent setting. We show that high shocks to the value of spending lead to an erosion of future fiscal discipline compared to low shocks, which lead to the reinstatement of fiscal discipline. This is related to the "no distortion at the top, distortion at the bottom" result in standard adverse selection problems, although here the result obtains from the combination of persistent private information and a present bias, and the induced distortions are reflected in changes to spending over time. For a sequence of high shocks, we further show that fiscal policy exhibits oscillatory dynamics, with large (small) distortions being followed by relatively smaller (larger) distortions. The logic stems from the self-control nature of our problem: a reduction in fiscal discipline at date t to provide dynamic incentives at $t-1$ relaxes incentive constraints at t, and hence implies that a smaller reduction in fiscal discipline at date $t+1$ is sufficient to provide dynamic incentives at t. Finally, we study the implied long-run debt dynamics when the market and social discount rates coincide. We show that while the self-insurance motive leads to the infinite accumulation of assets in the first best and the sequentially optimal rule, periods of non-discipline used to provide dynamic incentives in the ex-ante optimal rule can force the government accumulate maximal debt and become immiserated in the long run.

The paper is related to several literatures. First, as mentioned, the paper fits into the mechanism design literature on the tradeoff between commitment and flexibility.[8] In contrast to this literature, we study the optimal dynamic mechanism in a setting with persistent shocks. Second, the paper relates to the literature on the political economy of fiscal policy.[9] Most closely related is Azzimonti, Battaglini, and Coate (2010), which considers the quantitative welfare implications of a balanced budget rule in an i.i.d. setting where the government is present-biased towards pork-barrel spending. Our main departure is that we study optimal fiscal rules in a private information economy, and we use mechanism design tools to derive the optimal rule without restricting its structure. Third, our work is related to various papers studying principal-agent contracts under persistent private information, although, because our application is a self-control environment, their methods do not directly apply here.[10] Finally, more broadly, our paper contributes to the literature on hyperbolic discounting and the benefits of commitment devices.[11]

Section 2 describes our benchmark environment. Section 3 defines the ex-ante optimal and sequentially optimal fiscal rules. Section 4 illustrates the main insights from our model using a simple three-period example. Section 5 characterizes optimal fiscal rules in the infinite horizon economy and their implications for debt in the long run. Section 6 extends the analysis to an economy with a continuum of shocks. Section 7 concludes. Formal proofs are contained in the Appendix.

[8]In addition to the work previously cited, see Bernheim, Ray, and Yeltekin (1999), Bond and Sigurdsson (2011), and Sleet (2004) for related studies of self-control problems. More generally, the paper relates to the literature on delegation in principal-agent settings, including Alonso and Matouschek (2008), Ambrus and Egorov (2009), and Holmström (1977, 1984).

[9]In addition to the work cited in fn. 2, see also Acemoglu, Golosov, and Tsyvinski (2008), Azzimonti (2011), Krusell and Rios-Rull (1999), Song, Storesletten, and Zilibotti (2012), and Yared (2010).

[10]In addition to the work cited in fn. 5, see Golosov and Tsyvinski (2006), Halac (2012), Kapicka (2010), Strulovici (2011), and Williams (2011).

[11]See for example Barro (1999), Krusell, Kruscu, and Smith, Jr. (2010), Krusell and Smith, Jr. (2003), Laibson (1997), and Phelps and Pollak (1968).

2 The Model

We consider a simple model of fiscal policy in which a government makes repeated spending and borrowing decisions. Our environment is the same as that analyzed in Amador, Werning, and Angeletos (2006), with the exception that we allow for persistent private information.[12]

At the beginning of each period, $t \in \{0, 1, \ldots\}$, the government observes a shock to the economy, which is the government's private information or *type*. The government's type can be low or high, $\theta_t \in \{\theta^L, \theta^H\} \equiv \Theta$, where $\theta^H > \theta^L > 0$.[13] This type follows a first-order Markov process, with $p(\theta_{t+1}|\theta_t)$ corresponding to the probability of type θ_{t+1} at date $t+1$ conditional on type θ_t at date t. We consider $p\left(\theta^L|\theta^L\right) = p\left(\theta^H|\theta^H\right) \in [0.5, 1)$, and we compare the case where types are i.i.d., i.e., $p\left(\theta^i|\theta^i\right) = 0.5$ for $\theta^i \in \{\theta^L, \theta^H\}$, to the case where types are persistent over time, i.e., $p\left(\theta^i|\theta^i\right) > 0.5$ for $\theta^i \in \{\theta^L, \theta^H\}$.[14]

In each period t, following the realization of θ_t, the government chooses public spending $g_t \geq 0$ and debt b_{t+1} subject to a budget constraint:

$$g_t = \tau + b_{t+1}/(1+r) - b_t, \tag{1}$$

where $\tau > 0$ is the exogenous fixed tax revenue collected by the government in each period, b_t is the level of debt with which the government enters the period, and r is the exogenous interest rate. b_0 is exogenous and $\lim_{t\to\infty} b_{t+1}/(1+r)^t = 0$, so that all debts must be repaid and all assets must be consumed. Constraint (1) can be rewritten as a weak inequality constraint to allow for money burning without affecting any of our results. Such a weak inequality constraint would take into account the possibility of introducing fines in this setting; we ignore this possibility here to simplify the exposition.[15]

The government's welfare at date t, *prior* to the realization of its type θ_t, is

$$\sum_{k=0}^{\infty} \delta^k \mathbb{E}\left[\theta_{t+k} U(g_{t+k}) | \theta_{t-1}\right], \tag{2}$$

where $\theta_t U(g_t)$ is the social utility from public spending at date t and δ is the discount factor. The government's welfare *after* the realization of its type θ_t at date t, when choosing spending

[12] Amador, Werning, and Angeletos (2006) consider a two-period setting, but as shown in Amador, Werning, and Angeletos (2003), under i.i.d. shocks the results apply directly to a multiple-period environment.

[13] Section 6 extends our analysis to an economy with a continuum of types.

[14] Our main results are robust to considering $p\left(\theta^i|\theta^i\right) < 0.5$, although the intuition is different. Given that fiscal policy variables are positively autocorrelated in the data, we focus our attention on $p\left(\theta^i|\theta^i\right) \geq 0.5$. Our results are also robust to considering $p(\theta^L|\theta^L) \neq p(\theta^H|\theta^H)$; we focus on the symmetric case to simplify the exposition.

[15] In the proofs of our results in the Appendix, we allow (1) to be a weak inequality and show that this constraint must bind in equilibrium.

g_t, is

$$\theta_t U(g_t) + \beta \sum_{k=1}^{\infty} \delta^k \mathbb{E}\left[\theta_{t+k} U(g_{t+k}) | \theta_t\right], \tag{3}$$

where $\beta \in (0,1)$.

There are two important features of this environment. First, the government's objective (3) following the realization of its type does not coincide with its objective (2) prior to this realization. In particular, the government's welfare after θ_t is realized overweighs the importance of current public spending compared to its welfare before θ_t is realized. This formulation captures a friction that is common in various models of political economy interactions. For instance, preferences such as these may emerge naturally in settings with political uncertainty where policymakers place a higher value on public spending when they hold power and can make spending decisions. In such settings, policymakers are biased towards present public spending relative to future public spending and incur excessively high debts.

The second feature of this environment is that the realization of θ_t—which affects the marginal social utility of public spending—is privately observed by the government. One possible interpretation is that θ_t is not verifiable ex-post by a rule-making body; therefore, even if θ_t is observable, fiscal rules cannot explicitly depend on the value of θ_t.[16] An alternative interpretation is that the exact cost of public goods is only observable to the policymaker, who may be inclined to overspend on these goods. A third possibility is that citizens have heterogeneous preferences or heterogeneous information regarding the optimal level of public spending, and the government sees an aggregate that the citizens do not see (see Sleet, 2004).

To facilitate an explicit characterization of optimal fiscal rules, we make the following assumption.[17]

Assumption 1. $U(g_t) = \log(g_t)$.

Assumption 1 implies that welfare is separable with respect to the level of debt. To see this, define

$$\widetilde{\theta}^i = \sum_{k=1}^{\infty} \delta^k \mathbb{E}\left[\theta_{t+k} | \theta_t = \theta^i\right], \tag{4}$$

for $i = \{L, H\}$, so that at any date t, $\widetilde{\theta}_t = \widetilde{\theta}^i$ if $\theta_t = \theta^i$. Let the savings rate at t be $s_t \in [0,1]$,

[16] Because our focus is a fiscal constitution, we implicitly rule out other punishments that citizens can inflict on policymakers such as replacement. See Ales, Maziero, and Yared (2012) for a related model with private government information that allows for this possibility.

[17] This assumption is made in previous work studying economies with hyperbolic discounting, such as Barro (1999). Our main result that dynamic incentives are provided under persistent shocks while not under i.i.d. shocks does not depend on this assumption.

corresponding to the fraction of lifetime resources which is not spent at t:

$$g_t = (1 - s_t)[(1+r)\tau/r - b_t]. \tag{5}$$

Using this notation and Assumption 1, welfare in (2), at date t prior to the realization of the type θ_t, can be rewritten as

$$\sum_{k=0}^{\infty} \delta^k \mathbb{E}\big[\theta_{t+k} U(1 - s_{t+k}) + \widetilde{\theta}_{t+k} U(s_{t+k}) | \theta_{t-1}\big] + \chi(b_t), \tag{6}$$

for a constant $\chi(b_t)$ which depends on b_t.[18] Analogously, welfare in (3), at date t following the realization of θ_t, can be rewritten as

$$\theta_t U(1 - s_t) + \beta \widetilde{\theta}_t U(s_t) + \beta \sum_{k=1}^{\infty} \delta^k \mathbb{E}\big[\theta_{t+k} U(1 - s_{t+k}) + \widetilde{\theta}_{t+k} U(s_{t+k}) | \theta_t\big] + \varphi(b_t), \tag{7}$$

for a constant $\varphi(b_t)$ which depends on b_t.[19] Given the representation in (6) and (7), hereafter we consider the problem of a government which chooses a savings rate s_t in every period t. In this environment, the first-best policy is defined by a stochastic sequence of savings rates $\mathbf{s}^{fb} \equiv (s_0^{fb}, s_1^{fb}, \ldots)$ that satisfy $s_t^{fb} = s^{fb}(\theta^i)$ if $\theta_t = \theta^i$ for $\theta^i \in \{\theta^L, \theta^H\}$, where

$$\theta^i U'(1 - s^{fb}(\theta^i)) = \widetilde{\theta}^i U'(s^{fb}(\theta^i)). \tag{8}$$

3 Equilibrium Definition

We define a fiscal rule as a mechanism where the government reports the shock in every period and is assigned a policy as a function of the reports. We distinguish between the ex-ante optimal rule and the sequentially optimal rule. The ex-ante optimal rule is a dynamic mechanism chosen by the government at the beginning of time. In contrast, the sequentially optimal rule is a static mechanism chosen in every period by the current government, taking into account the future static mechanisms chosen by future governments.

We let $\theta^t = (\theta_0, \theta_1, \ldots, \theta_t) \in \Theta^t$ denote the history of shocks through time t.

3.1 Ex-ante Optimal Rule

Let $h_{t-1} = (\widehat{\theta}_0, \widehat{\theta}_1, \ldots, \widehat{\theta}_{t-1}) \in \Theta^{t-1}$ be the history of reported types through time $t-1$. A mechanism is a sequence of savings rates $s_t(h_{t-1}, \widehat{\theta}_t)$ for all $\big\{(h_{t-1}, \widehat{\theta}_t)\big\}_{t=0}^{\infty}$, which effectively

[18] This constant is equal to $\sum_{k=0}^{\infty} \delta^k \mathbb{E}\left[\theta_{t+k} U((1+r)^k[\tau(1+r)/r - b_t]) | \theta_{t-1}\right]$.

[19] This constant is equal to $\theta_t U(\tau(1+r)/r - b_t) + \beta \sum_{k=1}^{\infty} \delta^k \mathbb{E}\left[\theta_{t+k} U((1+r)^k[\tau(1+r)/r - b_t]) | \theta_t\right]$.

specify levels of public spending $g_t(h_{t-1}, \widehat{\theta}_t)$ and debt $b_{t+1}(h_{t-1}, \widehat{\theta}_t)$, as a function of the history of past reports and the current report.

Given the mechanism, the government chooses a reporting strategy $m_t(h_{t-1}, \theta_t)$ for all $\{(h_{t-1}, \theta_t)\}_{t=0}^{\infty}$, where θ_t is the government's type at date t and $m_t(h_{t-1}, \theta_t) \in \{\theta^L, \theta^H\}$ is the government's report of its type at t. We restrict attention to public strategies, that is, strategies that depend only on the public history—reports and policies—and on the government's current private information, but not on privately observed history. It follows by standard arguments that if all future governments choose public strategies, and if the mechanism is a function of the public history, then the current government's best response is also a public strategy. From the Revelation Principle, we can restrict attention to truthtelling equilibria in which $m_t(h_{t-1}, \theta_t) = \theta_t$ for all h_{t-1} and θ_t.

A perfect Bayesian equilibrium of this revelation game is a mechanism and a reporting strategy such that the budget constraint (1) is satisfied in every period following every history, and the policy under the mechanism is incentive compatible, meaning that following every history and type realization, the government prefers to report $m_t(h_{t-1}, \theta_t) = \theta_t$ rather than $m_t(h_{t-1}, \theta_t) = \widehat{\theta}_t \neq \theta_t$. An ex-ante optimal rule in this framework is one that selects a mechanism and a reporting strategy that maximize the ex-ante welfare (6) in period 0.

We formulate the ex-ante optimal rule as a solution to a sequence program. Given history θ^{t-1}, let $W_{t+1}(\theta^{t-1}, \theta_t)$ be the expected continuation value from $t+1$ on, normalized by $b_t(\theta^{t-1})$, for a type θ_t who truthfully reports $\widehat{\theta}_t = \theta_t$:

$$W_{t+1}(\theta^{t-1}, \theta_t) = \sum_{\theta_{t+1} \in \{\theta^L, \theta^H\}} p(\theta_{t+1}|\theta_t) \left[\begin{array}{c} \theta_{t+1} U(1 - s_{t+1}(\theta^{t-1}, \theta_t, \theta_{t+1})) \\ +\widetilde{\theta}_{t+1} U(s_{t+1}(\theta^{t-1}, \theta_t, \theta_{t+1})) + \delta W_{t+2}(\theta^{t-1}, \theta_t, \theta_{t+1}) \end{array} \right]. \tag{9}$$

In contrast, given history θ^{t-1}, let $V_{t+1}(\theta^{t-1}, \widehat{\theta}_t)$ be the expected continuation value from $t+1$ on, normalized by $b_t\left(\theta^{t-1}\right)$, for a type θ_t who lies and reports $\widehat{\theta}_t \neq \theta_t$:[20]

$$V_{t+1}(\theta^{t-1}, \widehat{\theta}_t) = \sum_{\theta_{t+1} \in \{\theta^L, \theta^H\}} p(\theta_{t+1}|\theta_t) \left[\begin{array}{c} \theta_{t+1} U(1 - s_{t+1}(\theta^{t-1}, \widehat{\theta}_t, \theta_{t+1})) \\ +\widetilde{\theta}_{t+1} (U(s_{t+1}(\theta^{t-1}, \widehat{\theta}_t, \theta_{t+1})) + \delta W_{t+2}(\theta^{t-1}, \widehat{\theta}_t, \theta_{t+1}) \end{array} \right]. \tag{10}$$

Note that in the special case of i.i.d. shocks, expectations over future shocks do not depend on the current shock θ_t, implying that $W_{t+1}(\theta^{t-1}, \widehat{\theta}_t) = V_{t+1}(\theta^{t-1}, \widehat{\theta}_t)$.

Using (7), the incentive compatibility constraint for a government of type θ_t is

$$\begin{aligned} &\theta_t U(1 - s_t(\theta^{t-1}, \theta_t)) + \beta \widetilde{\theta}_t U(s_t(\theta^{t-1}, \theta_t)) + \beta\delta W_{t+1}(\theta^{t-1}, \theta_t) \\ \geq\ &\theta_t U(1 - s_t(\theta^{t-1}, \widehat{\theta}_t)) + \beta \widetilde{\theta}_t U(s_t(\theta^{t-1}, \widehat{\theta}_t)) + \beta\delta V_{t+1}(\theta^{t-1}, \widehat{\theta}_t) \text{ for } \widehat{\theta}_t \neq \theta_t \text{ and all } \theta^{t-1}. \end{aligned} \tag{11}$$

[20]Because there are only two types, we only need to define one such expected continuation value.

Condition (11) says that the government prefers to report its true type θ_t rather than to lie and report $\widehat{\theta}_t \neq \theta_t$. To understand this constraint, note that the government's true type θ_t not only directly affects the government's immediate payoff by determining the marginal cost and benefit of current savings, but, if shocks are persistent, it can also affect the government's continuation payoff by changing the expectations over the realizations of future types.[21]

Let $\rho = \left\{ \left\{ s_t(\theta^t), W_{t+1}(\theta^t), V_{t+1}(\theta^t) \right\}_{\theta^t \in \Theta^t} \right\}_{t=0}^{\infty}$ be a stochastic sequence of savings rates and continuation values. The ex-ante optimal rule solves the following sequence problem:

$$\max_{\rho} \sum_{\theta_0 \in \{\theta^L, \theta^H\}} p\left(\theta_0 | \theta_{-1}\right) \left[\theta_0 U(1 - s_0(\theta^0)) + \widetilde{\theta}_0 U(s_0(\theta^0)) + \delta W_1(\theta^0)\right] \tag{12}$$

$$\text{s.t. } (9), (10), \text{ and } (11).$$

It is clear that the solution to this program is invariant to the initial level of debt b_0.

3.2 Sequentially Optimal Rule

The sequentially optimal fiscal rule is the one that results if, at every history, the government chooses a static mechanism that maximizes social welfare given that future governments will do the same. That is, given θ_{t-1} and b_t, the government chooses a mechanism $\{g_t(\widehat{\theta}_t), b_{t+1}(\widehat{\theta}_t)\}$, assigning a level of spending and debt conditional on the report $\widehat{\theta}_t$, which maximizes social welfare taking the actions of future governments as given. The future government knows the true value of θ_t and, given θ_t and $b_{t+1}(\widehat{\theta}_t)$, it analogously chooses an optimal static mechanism taking the actions of future governments as given.

The sequentially optimal rule thus solves the following problem:

$$J(\theta_{t-1}, b_t) = \max_{\{g_t(\theta_t), b_{t+1}(\theta_t)\}_{\theta_t \in \{\theta^L, \theta^H\}}} \sum_{\theta_t \in \{\theta^L, \theta^H\}} p(\theta_t | \theta_{t-1}) \big(\theta_t U(g_t(\theta_t)) + \delta J(\theta_t, b_{t+1}(\theta_t))\big) \tag{13}$$

$$\text{s.t.}$$

$$g_t(\theta_t) = \tau + b_{t+1}(\theta_t)/(1 + r) - b_t \text{ and} \tag{14}$$

$$\theta_t U\left(g_t(\theta_t)\right) + \beta\delta J(\theta_t, b_{t+1}(\theta_t)) \geq \theta_t U(g_t(\widehat{\theta}_t)) + \beta\delta J(\theta_t, b_{t+1}(\widehat{\theta}_t)) \text{ for } \widehat{\theta}_t \neq \theta_t. \tag{15}$$

$J(\theta_{t-1}, b_t)$ is the value at t under the payoff-relevant states θ_{t-1} and b_t if the current government chooses an optimal static mechanism given that future governments do the same. In a sense, $J(\theta_{t-1}, b_t)$ thus corresponds to the solution of a two-period mechanism design problem. In choosing its report $\widehat{\theta}_t$, the government's flow welfare is $\theta_t U(g_t(\widehat{\theta}_t))$ and its continuation welfare is $J(\theta_t, b_{t+1}(\widehat{\theta}_t))$. Condition (15) is the incentive compatibility constraint, where the

[21] Because θ_t follows a first-order Markov process, the single period deviation principle holds.

government knows that if it lies and reports $\widehat{\theta}_t \neq \theta_t$, this affects its payoff from tomorrow onward only through the implied level of debt $b_{t+1}(\widehat{\theta}_t)$, since, given $b_{t+1}(\widehat{\theta}_t)$, the government which knows the true value of θ_t will choose an optimal static mechanism going forward. We interpret the sequentially optimal fiscal rule that emerges from this recursion as a rule that has a bite in the short term, but can be renegotiated for the future.[22]

In the recursive program defined above, $J(\cdot)$ may have multiple solutions. We select a unique solution by considering the limit of a finite horizon economy with end date T as T approaches infinity. $J(\theta_{t-1}, b_t)$ is thus characterized recursively via backward induction. In Appendix A, we show that under log utility the solution to (13)–(15) admits a savings rate $s_t(\theta_t)$ for each θ_t that is invariant to the level of debt b_t and only depends on the previous shock θ_{t-1}, and, hence, using the welfare representation in (6) and (7) and given θ_{t-1}, (13)–(15) can be rewritten as:

$$\max_{\{s_t(\theta_t)\}_{\theta_t \in \{\theta^L, \theta^H\}}} \sum_{\theta_t \in \{\theta^L, \theta^H\}} p(\theta_t|\theta_{t-1}) \big(\theta_t U(1 - s_t(\theta_t)) + \widetilde{\theta}_t U(s_t(\theta_t)) \big) \tag{16}$$

s.t.

$$\theta_t U(1 - s_t(\theta_t)) + \beta \widetilde{\theta}_t U(s_t(\theta_t)) \geq \theta_t U(1 - s_t(\widehat{\theta}_t)) + \beta \widetilde{\theta}_t U(s_t(\widehat{\theta}_t)) \text{ for } \widehat{\theta}_t \neq \theta_t. \tag{17}$$

Clearly, the sequentially optimal rule must satisfy the constraints of the problem defined by the ex-ante optimal rule in (12). To see this, note that the incentive compatibility constraint (17) is more strict than the constraint (11), as by definition the solution to (16)–(17) must admit a sequence of savings rates satisfying $W_{t+1}\left(\theta^{t-1}, \theta_t\right) \geq V_{t+1}\left(\theta^{t-1}, \theta_t\right)$. Thus, naturally, the sequentially optimal rule provides weakly lower welfare than the ex-ante optimal rule.

4 Three-Period Example

To provide intuition for our main results, we start by considering a three-period economy with $t \in \{0, 1, 2\}$. The purpose is threefold. First, we show that in contrast with the case of i.i.d. shocks, when shocks are persistent, the ex-ante optimal rule does not coincide with the sequentially optimal rule. We discuss how the difference between the two rules depends on parameters such as the level of persistence and the degree of time inconsistency. Second, we show that, also unlike under i.i.d shocks, the ex-ante optimal rule under persistent shocks exhibits history dependence, that is, the mechanism at $t = 1$ depends on more than the payoff-relevant variables at $t = 1$. Finally, we describe properties of the solution that provide some insight into the dynamics of the infinite horizon economy discussed in Section 5.

[22]This notion of sequential optimality is related to the notion of reconsideration-proofness in Kocherlakota (1996) and renegotiation-proofness in Farrell and Maskin (1989) and Strulovici (2011).

Given log utility and the arguments of Section 3, the problem can be stated as that of choosing a savings rate $s_0(\theta_{-1}, \theta_0)$ at date 0 and a savings rate $s_1(\theta_{-1}, \theta_0, \theta_1)$ at date 1.[23] We contrast the date 1 policies in the ex-ante optimal rule and the sequentially optimal rule.

We begin by considering the sequentially optimal rule, which can be solved for by backward induction. At date 1, the government solves program (16)–(17), where $\widetilde{\theta}_1$ is given by $\delta\mathbb{E}[\theta_2|\theta_1]$. To simplify the analysis, we assume that θ^L and θ^H are relatively close to each other:

$$\frac{\theta^H}{\theta^L} < \frac{1}{\beta}. \tag{18}$$

Condition (18) implies that the first-best savings rate at date 1, defined by (8), is not incentive compatible for the low type, who would want to pretend to be a high type so as to spend and borrow more. Moreover, this condition implies that the optimal static mechanism at date 1 features pooling: following the realization of θ_0, the sequentially optimal rule assigns a fixed savings rate, independent of the type at date 1.[24] This pooled savings rate is chosen optimally given the probabilities of a high and low type at date 1, so it is given by

$$\mathbb{E}[\theta_1|\theta_0 = \theta^i]U'(1 - s_1(\theta_{-1}, \theta^i)) - \delta\mathbb{E}[\theta_2|\theta_0 = \theta^i]U'(s_1(\theta_{-1}, \theta^i)) = 0 \tag{19}$$

for $\theta^i \in \{\theta^L, \theta^H\}$, where, with some abuse of notation, we have written $s_1(\cdot)$ as a function of θ_{-1} and θ_0 only given that savings at date 1 do not depend on θ_1 under pooling.

Consider next the ex-ante optimal rule. Suppose by contradiction that the ex-ante optimum and the sequential optimum coincide at date 1. We show that it is then possible to perturb the ex-ante optimal mechanism in a way that reduces welfare from the perspective of date 1 but increases welfare from the perspective of date 0, thus increasing ex-ante welfare and contradicting the assumption that the two rules coincide at date 1.

Assume that at date 0 the incentive compatibility constraint binds for the low type and is slack for the high type. This means that the low type is under-saving relative to first best, and he cannot be induced to save more as he would then want to pretend to be a high type to save less. Formally, the low type is indifferent between reporting the truth and receiving a payoff equal to

$$\begin{aligned}&\theta^L U(1 - s_0(\theta_{-1}, \theta^L)) + \beta\delta\mathbb{E}[\theta_1 + \delta\theta_2|\theta_0 = \theta^L]U(s_0(\theta_{-1}, \theta^L))\\ &+\beta\delta\big(\mathbb{E}[\theta_1|\theta_0 = \theta^L]U(1 - s_1(\theta_{-1}, \theta^L)) + \delta\mathbb{E}[\theta_2|\theta_0 = \theta^L]U(s_1(\theta_{-1}, \theta^L))\big),\end{aligned}$$

[23]Note that savings at date 2 are always equal to zero, since this is the final period.

[24]Intuitively, separation when types are sufficiently close is suboptimal because it would require excessively low savings by the high type so as to satisfy the incentive compatibility constraint of the low type.

and lying and receiving a payoff equal to

$$\theta^L U(1-s_0(\theta_{-1},\theta^H))+\beta\delta\mathbb{E}[\theta_1+\delta\theta_2|\theta_0=\theta^L]U(s_0(\theta_{-1},\theta^H))$$
$$+\beta\delta\big(\mathbb{E}[\theta_1|\theta_0=\theta^L]U(1-s_1(\theta_{-1},\theta^H))+\delta\mathbb{E}[\theta_2|\theta_0=\theta^L]U(s_1(\theta_{-1},\theta^H))\big),$$

where, by the contradiction assumption that the mechanism is sequentially optimal at date 1, $s_1\left(\theta_{-1},\theta^i\right)$ is the pooled savings rate given by (19). Clearly, the low type could be induced to save more at date 0 if his payoff from pretending to be a high type could be reduced. Note that this payoff can be reduced by changing either the savings rate that is assigned at date 0 given a high reported type at date 0, or the savings rate that is assigned at date 1 given a high reported type at date 0. Moreover, the effect on welfare of changing the savings rate at date 1 depends on the probabilities of a low type and a high type at date 1, which under persistent shocks depend on the realized type at date 0.

With this observation in mind, consider a perturbation that reduces the savings rate at date 1 given a high type at date 0, $s_1\left(\theta_{-1},\theta^H\right)$, by $\varepsilon>0$ arbitrarily small. This perturbation clearly reduces *equilibrium* welfare from date 1 onward, that is, welfare at date 1 given a truthful report of $\widehat{\theta}_0=\theta^H$ at date 0, as it induces overspending at date 1. However, from the envelope condition in (19), this is a second-order loss, so this effect approaches 0 as ε approaches 0. In contrast, consider the effect of the perturbation on *off-equilibrium* welfare from date 1 onward, that is, welfare at date 1 given a non-truthful report of $\widehat{\theta}_0=\theta^H$ at date 0. As ε approaches 0, this effect takes the same sign as

$$\mathbb{E}[\theta_1|\theta_0=\theta^L]U'(1-s_1(\theta_{-1},\theta^H))-\delta\mathbb{E}[\theta_2|\theta_0=\theta^L]U'(s_1(\theta_{-1},\theta^H)). \tag{20}$$

If shocks are i.i.d., $\mathbb{E}[\theta_1|\theta_0=\theta^L]/\mathbb{E}[\theta_2|\theta_0=\theta^L]=\mathbb{E}[\theta_1|\theta_0=\theta^H]/\mathbb{E}[\theta_2|\theta_0=\theta^H]$, so, given (19), (20) must equal zero. This means that the perturbation affects continuation welfare on and off the equilibrium path equally and hence cannot improve ex-ante welfare. In contrast, when shocks are persistent, $\mathbb{E}[\theta_1|\theta_0=\theta^L]/\mathbb{E}[\theta_2|\theta_0=\theta^L]<\mathbb{E}[\theta_1|\theta_0=\theta^H]/\mathbb{E}[\theta_2|\theta_0=\theta^H]$, so (20) must be negative. Given (19), this means that the perturbation reduces continuation welfare off the equilibrium path without affecting continuation welfare on path. Therefore, the low type's incentive compatibility constraint at date 0 can be relaxed at no social cost, and ex-ante welfare can be increased.

The intuition behind the perturbation is simple. In every period, the government learns about its current spending needs. If shocks are persistent, the current shock also informs the government about its future spending needs. The government has a temptation to overspend today even if its needs are low, but it maximizes social welfare from tomorrow on. The ex-ante optimal rule then has the feature that it becomes excessively lax in the future if spending is

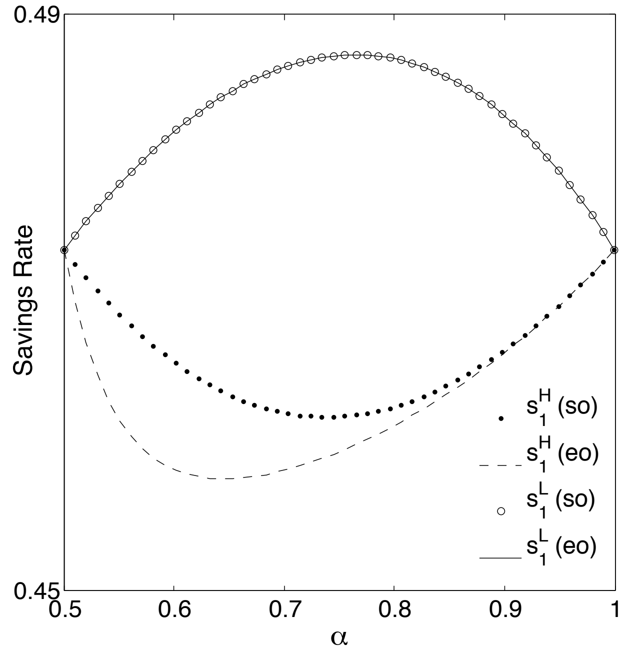

Figure 1: Savings rate at $t = 1$ in the ex-ante and sequentially optimal rules given $\theta_{-1} = \theta^H$.

high today. The reason is that the expected cost of lax rules tomorrow is greater if spending needs are low today, as spending needs are then likely to be low tomorrow. As a result, the threat of no discipline in the future ironically imposes discipline today.

We illustrate the properties of the ex-ante optimum and the sequential optimum with a numerical example.[25] It can be shown that under (18), the ex-ante optimal rule also features pooling at date 1; thus, under the two rules, the savings rate at date 1 is independent of the realized type at date 1. Figure 1 shows this savings rate, $s_1(\theta_{-1}, \theta^i)$, in the ex-ante optimal ("eo") and sequentially optimal ("so") rules for $\theta_0 = \theta^L$ ("s_1^L") and $\theta_0 = \theta^H$ ("s_1^H"), given $\theta_{-1} = \theta^H$, as a function of the persistence of types, denoted $\alpha \equiv p(\theta^i|\theta^i)$ in the figure.

Three points are evident in Figure 1. First, consistent with the perturbation just described, the figure shows that the ex-ante optimal savings rate at date 1 following $\theta_0 = \theta^H$ is below the sequentially optimal rate. As explained, inducing overspending ex post is efficient ex ante because it allows to relax the low type's incentive compatibility constraint and curb his spending at date 0. Second, the figure shows that following $\theta_0 = \theta^L$, the ex-ante optimal savings rate at date 1 coincides with the sequentially optimal rate. The logic for this "no distortion at the top" also stems from the low type's incentive compatibility constraint at date 0: to relax this constraint, the ex-ante optimal rule maximizes the continuation welfare for the low type given a truthful report at date 0, which corresponds to assigning the optimal static rate at date 1. Finally, the figure shows how the savings rate at date 1 depends on the

[25] We consider $\delta = 0.9$, $\beta = 0.6$, $\theta^L = 2$, and $\theta^H = 3$.

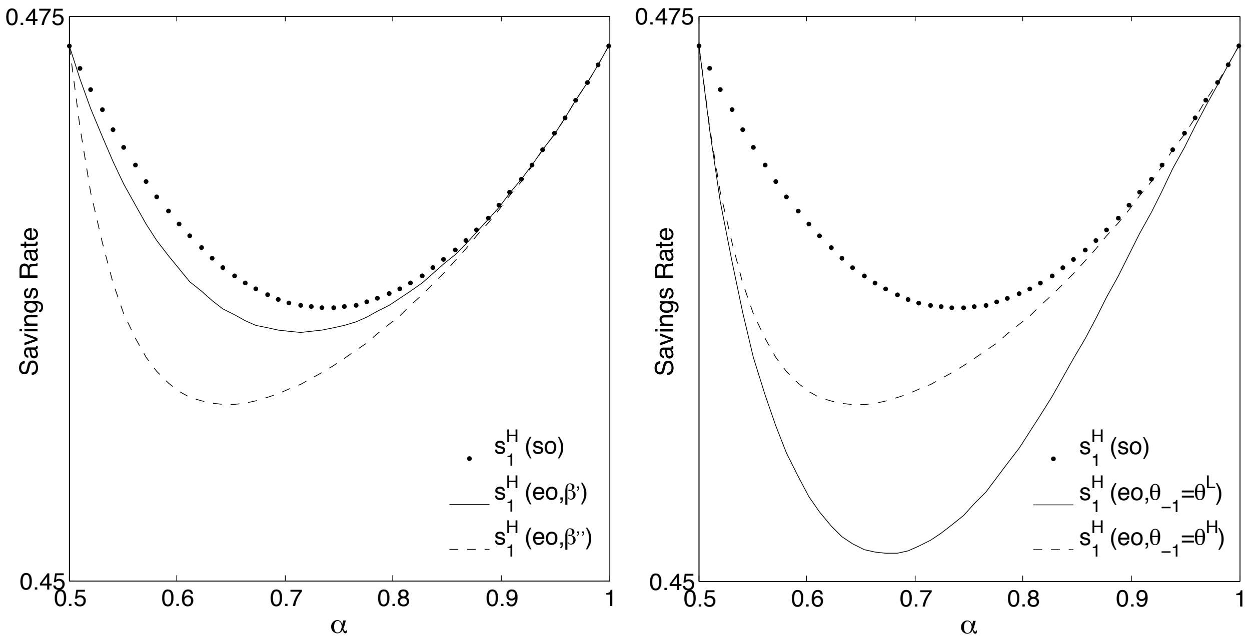

Figure 2: Savings rate at $t = 1$ under different degrees of time inconsistency (left graph, with $\beta'' > \beta'$) and under different histories (right graph).

persistence of types, and in particular that the ex-ante optimal and sequentially optimal rates coincide if types are i.i.d. ($\alpha = 0.5$) or fully persistent ($\alpha = 1$). Intuitively, in these cases, $\mathbb{E}[\theta_1|\theta_0 = \theta^i] = \mathbb{E}[\theta_2|\theta_0 = \theta^i]$, so the marginal benefit of spending at dates 1 and 2 is the same. Given (19), this means that the sequentially optimal rate at date 1 is independent of θ_0, and thus any perturbation would affect welfare on and off the equilibrium path equally.

To gain further intuition for the ex-ante optimal rule, the left graph in Figure 2 explores how the savings rate at date 1 following $\theta_0 = \theta^H$ depends on the time inconsistency problem. The graph shows that the difference between the sequentially optimal and ex-ante optimal rates is smaller the larger is the government's present bias, i.e. the lower is β.[26] The logic is simple: the less the government at date 0 cares about its welfare at date 1, the less effective is a distortion in the sequentially optimal rate at date 1 to relax incentive compatibility constraints at date 0. Since the cost of the distortion at date 1 is independent of the degree of present bias, it follows that the distortion must be smaller when this bias is larger.

Finally, the right graph in Figure 2 explores whether the ex-ante optimal mechanism exhibits history dependence. The graph shows that, indeed, the ex-ante optimal savings rate at date 1 following $\theta_0 = \theta^H$ depends on the history, namely on θ_{-1}.[27] Specifically, this rate is lower and further away from the sequentially optimal rate if $\theta_{-1} = \theta^L$ than if $\theta_{-1} = \theta^H$,

[26] We take $\beta' = 0.4$ and $\beta'' = 0.6$. Note that given other parameters fixed, β must be low enough for condition (18) to be satisfied. In our numerical example, (18) requires $\beta \leq 2/3$. Note also that the sequentially optimal savings rate is independent of β.

[27] The savings rate at date 1 following $\theta_0 = \theta^L$ is history independent, as it is sequentially optimal.

so fiscal rules become more excessively lax after a high shock at date 0 if $\theta_{-1} = \theta^L$. The intuition is that θ_{-1} affects the distribution of types at date 0, and thus the benefits and costs of perturbing the savings rate at date 1. Recall that the benefit of the perturbation is that it relaxes the low type's incentive compatibility constraint at date 0 and allows to increase savings given $\theta_0 = \theta^L$; the cost on the other hand is that it induces an ex-post suboptimal savings rate at date 1 given $\theta_0 = \theta^H$. If $\theta_{-1} = \theta^L$, the probability of $\theta_0 = \theta^L$ is higher, and therefore the relative benefits of the perturbation are larger.

5 Optimal Fiscal Rules

This section characterizes the optimal fiscal rules in an infinite horizon economy. Because the sequentially optimal rule is simple, we first study this rule in Subsection 5.1 and then use it as a benchmark in describing the ex-ante optimal rule in Subsection 5.2. In Subsection 5.3, we consider the implications of these rules for the level of debt in the long run.

To simplify the analysis and consistent with the extension to a continuum of types in Section 6, we assume that types are relatively close to each other:

Assumption 2.

$$\frac{\theta^H}{\theta^L} - \frac{\widetilde{\theta}^H}{\widetilde{\theta}^L} < \frac{1 - p(\theta^i|\theta^i)}{p(\theta^i|\theta^i)} \left(\frac{1}{\beta} - 1\right).$$

Assumption 2 is implied by condition (18) in the three-period example of Section 4 if the level of persistence is low enough, i.e. if $p(\theta^i|\theta^i) \geq 1/2$ is sufficiently close to $1/2$.[28] We discuss the intuition and implications of Assumption 2 in the following subsections.

5.1 Sequentially Optimal Rule

Consider program (16)–(17) defining the sequentially optimal fiscal rule. Given Assumption 2, it can be shown that this rule features pooling: the savings rate $s_t(\theta^t)$ is independent of the realization of θ_t and depends only on θ_{t-1}, which is used in predicting the value of θ_t. This is stated formally in the proposition below.

Proposition 1 (sequential optimum). *For all θ^t and θ^k, the sequential optimum features*

$$s_t(\theta^t) = s_k(\theta^k) \textit{ if } \theta_{t-1} = \theta_{k-1}.$$

[28] The strict inequality in Assumption 2 cannot be satisfied under full persistence ($p(\theta^i|\theta^i) = 1$). However, our problem in that case is rather trivial as the first-best savings rate is the same for both types. If types are close enough, the assumption is satisfied for any persistence level, including those arbitrarily close to 1.

Moreover, $s_t(\theta^t)$ satisfies

$$\mathbb{E}[\theta_t|\theta_{t-1}]U'(1 - s_t(\theta^t)) - \mathbb{E}[\widetilde{\theta}_t|\theta_{t-1}]U'(s_t(\theta^t)) = 0. \tag{21}$$

Proof. See Appendix B. □

The intuition for this result is analogous to that in the three-period example of Section 4, where the optimal static mechanism features pooling when types are sufficiently close. Here the result follows from Assumption 2, which implies that the analog of condition (18) used for the three-period economy holds in the infinite horizon economy.[29]

Given the budget constraint (1), Proposition 1 says that the sequentially optimal rule at date t prescribes a level of debt $b_{t+1}(\theta^t)$ as a function of θ_{t-1} and $b_t(\theta^{t-1})$ only. If $\theta_{t-1} = \theta^H$, the prescribed level of debt is higher than if $\theta_{t-1} = \theta^L$, as a high shock is then more likely and thus the sequentially optimal level of deficit-financed spending at date t is higher. Furthermore, if $b_t(\theta^{t-1})$ is relatively high, then $b_{t+1}(\theta^t)$ must also be relatively high to facilitate the servicing of the debt while simultaneously providing public goods. A useful implication is that the sequential optimum can be implemented with a renegotiated debt limit.

Corollary 1. *The sequentially optimal rule at any date t can be implemented with a history-independent debt limit, $\overline{b}(\theta_{t-1}, b_t(\theta^{t-1}))$.*

It follows from Assumption 2 and equation (21) that under the sequentially optimal rule, both the low type and the high type would like to borrow and spend more than they are allowed to. Thus, the sequentially optimal rule takes the form of a renegotiated debt limit where, given θ_{t-1} and $b_t(\theta^{t-1})$, both types choose the maximum allowable debt. In Section 6, we show that these results extend to an economy with a continuum of types, although in that case some types choose to borrow below the debt limit.

5.2 Ex-ante Optimal Rule

We now consider the ex-ante optimal fiscal rule defined in (12). We first develop a recursive representation of the problem and then characterize the solution.

5.2.1 Recursive Representation

Our recursive representation is similar in spirit to that of Fernandes and Phelan (2000), who develop a recursive technique to study a principal-agent problem under persistent shocks. The

[29]Specifically, this condition is $\dfrac{\theta^H/\widetilde{\theta}^H}{\theta^L/\widetilde{\theta}^L} < \dfrac{1}{\beta}$.

idea of this representation is to solve program (12) by choosing $s_t(\theta^t)$ and $V_{t+1}(\theta^t)$ defined in (10) sequentially for each history θ^{t-1}, where associated with each choice of $V_{t+1}(\theta^t)$ is some continuation welfare $W_{t+1}(\theta^t)$ that is a function of the realized type θ_t and the chosen $V_{t+1}(\theta^t)$. To characterize this value of $W_{t+1}(\theta^t)$, let $W(\theta^i, V)$ for $\theta^i \in \{\theta^L, \theta^H\}$ correspond to the solution to (12) given $\theta_{-1} = \theta^i$ and subject to the following additional constraint:

$$V = \sum_{\theta_0 \in \{\theta^L, \theta^H\}} p(\theta_0 | \theta_{-1} = \theta^{-i}) \big[\theta_0 U(1 - s_0(\theta^0)) + \widetilde{\theta}_0 U(s_0(\theta^0)) + \delta W_1(\theta^0)\big]. \tag{22}$$

Constraint (22) is often referred to as a *threat-keeping constraint*. This constraint says that if θ_{-1} is equal to θ^{-i} as opposed to θ^i, the expected welfare under the savings rate sequence that solves the program must be equal to V. Using this formulation, we rewrite program (12) recursively as follows:

$$W\left(\theta^i, V\right) = \max_{\{s^L, s^H, V^L, V^H\}} \left\{ \begin{array}{c} p(\theta^L|\theta^i)\big(\theta^L U(1 - s^L) + \widetilde{\theta}^L U(s^L) + \delta W(\theta^L, V^L)\big) \\ +p(\theta^H|\theta^i)\big(\theta^H U(1 - s^H) + \widetilde{\theta}^H U(s^H) + \delta W(\theta^H, V^H)\big) \end{array} \right\} \tag{23}$$

s.t.

$$V = \left\{ \begin{array}{c} p(\theta^L|\theta^{-i})\big(\theta^L U(1 - s^L) + \widetilde{\theta}^L U(s^L) + \delta W(\theta^L, V^L)\big) \\ +p(\theta^H|\theta^{-i})\big(\theta^H U(1 - s^H) + \widetilde{\theta}^H U(s^H) + \delta W(\theta^H, V^H)\big) \end{array} \right\}, \tag{24}$$

$$\theta^L U(1 - s^L) + \beta\widetilde{\theta}^L U(s^L) + \beta\delta W(\theta^L, V^L) \geq \theta^L U(1 - s^H) + \beta\widetilde{\theta}^L U(s^H) + \beta\delta V^H, \tag{25}$$

$$\theta^H U(1 - s^H) + \beta\widetilde{\theta}^H U(s^H) + \beta\delta W(\theta^H, V^H) \geq \theta^H U(1 - s^L) + \beta\widetilde{\theta}^H U(s^L) + \beta\delta V^L, \tag{26}$$

$$\underline{V}^L \leq V^L \leq \overline{V}^L, \text{ and } \underline{V}^H \leq V^H \leq \overline{V}^H. \tag{27}$$

(23)–(27) is a recursive representation of (12) starting from some history θ^{t-1}. The program selects savings rates s^i for $i \in \{L, H\}$—which represent the values of $s_t(\theta^t)$—and threats V^i for $i \in \{L, H\}$—which represent the values of $V_{t+1}(\theta^t)$—to maximize social welfare, taking into account that the continuation welfare conditional on $\theta_t = \theta^i$ is equal to $W(\theta^i, V^i)$, and subject to (24)–(27). Constraint (24) is a recursive representation of the threat-keeping constraint stating that if type θ^{-i} deviates and pretends to be θ^i in period $t-1$, his continuation welfare at t is equal to V. As such, V^i, which is chosen in the current period t, is the continuation welfare at $t+1$ to a type θ^{-i} who pretends to be θ^i at t. Constraints (25)–(26) are recursive representations of the incentive compatibility constraints in (11), and constraint (27) guarantees that the values of V^i are within a feasible range.

We next characterize $W(\theta^i, V)$. Clearly, if shocks are i.i.d., it must be that $W(\theta^i, V) = V$. If shocks are persistent, the following lemma holds:

Lemma 1. *If shocks are persistent, $W(\theta^i, V)$ is strictly increasing, strictly concave, and continuously differentiable in V over the range $(\underline{V}^i, \overline{V}^i)$.*

Proof. See Appendix C. $\square$

$\underline{V}^i$ and $\overline{V}^i$ are respectively the lowest and highest values of $V_{t+1}(\theta^t)$ that can be attained in the solution to (12). $\overline{V}^i$ is the value of V that results from the sequence problem (12) given $\theta_{-1} = \theta^i$ and subject to the threat-keeping constraint (22) when this constraint does not bind, so that the solution effectively corresponds to the ex-ante optimum. Values of $V_{t+1}(\theta^t)$ that exceed $\overline{V}^i$ are never attained along the equilibrium path: such high values of $V_{t+1}(\theta^t)$ would tighten the incentive compatibility constraints (25) and (26) while simultaneously reducing the continuation welfare below $W(\theta^i, \overline{V}^i)$. Hence, only threats such that $V^i < \overline{V}^i$ are used, and the optimal level of threats depends on the benefits of relaxing incentive compatibility constraints relative to the costs of reducing continuation welfare $W(\theta^i, V^i)$.

To understand the role of Assumption 2 and the threat-keeping constraint (24), let λ be the Lagrange multiplier on this constraint. The envelope condition implies $W_V(\theta^i, V) = -\lambda$, so that given the strict concavity of $W(\cdot)$, lower values of V are associated with more negative values of λ. It follows that solving (23)–(27) is equivalent to solving the following problem:

$$\max_{\{s^L, s^H, V^L, V^H\}} \left\{ \begin{array}{c} (p(\theta^L|\theta^i) + \lambda p(\theta^L|\theta^{-i}))(\theta^L U(1-s^L) + \widetilde{\theta}^L U(s^L) + \delta W(\theta^L, V^L)) \\ +(p(\theta^H|\theta^i) + \lambda p(\theta^H|\theta^{-i}))(\theta^H U(1-s^H) + \widetilde{\theta}^H U(s^H) + \delta W(\theta^H, V^H)) \end{array} \right\} \tag{28}$$
$$\text{s.t. } (25) - (27).$$

Under persistent types, the role of the threat-keeping constraint is to effectively "twist" the probabilities assigned to each type. If, for example, $\theta^i = \theta^H$, the objective in (28) under-weighs welfare conditional on the low type relative to the high type, and this is done more severely the lower is V. The opposite is true if $\theta^i = \theta^L$. Hence, to satisfy the threat-keeping constraint, the program over-weighs (under-weighs) welfare at t conditional on the type that is less (more) likely to occur from the perspective of a deviating type θ^{-i} at $t-1$.

Assumption 2 facilitates the characterization of $W(\theta^i, V)$ because it guarantees that, along the equilibrium path, the value of V is never chosen to be so low that the objective in (28) would assign a negative weight to some type.[30] Intuitively, if V is very low, then types are sufficiently close that any benefits from providing better incentives today are outweighed by the costs of reduced welfare tomorrow. Note that these relative benefits and costs depend on the persistence of types; in particular, under a high level of persistence $p(\theta^i|\theta^i)$, the future

[30] In the case that the objective assigns a negative weight to some type, we have not been able to prove that $W(\theta^i, V)$ is concave and differentiable, which prevents us from achieving a characterization of the equilibrium dynamics.

welfare costs may be incurred by a type that is unlikely to occur. This explains why the condition in Assumption 2 depends on $p(\theta^i|\theta^i)$. (See Lemma 7 in Appendix C for details.)

5.2.2 Characterization

We now characterize the ex-ante optimal fiscal rule. Consider first the i.i.d. benchmark.

Proposition 2 (ex-ante optimum under i.i.d. shocks). *If shocks are i.i.d., the ex-ante optimum coincides with the sequential optimum.*

Proof. See Appendix C. □

When shocks are i.i.d., the ex-ante optimal rule prescribes the sequentially optimal savings rate in each period, which in this case is constant and which can be implemented with a renegotiated debt limit. This rule is thus history independent and does not provide dynamic incentives for truthtelling. The reason why dynamic incentives are inefficient is as discussed in the three-period example: under i.i.d. shocks, any perturbation in the ex-post optimal rule affects continuation welfare on the equilibrium path (given a truthful report) and off the equilibrium path (given a non-truthful report) equally, and therefore cannot increase ex-ante welfare. This result is analogous to that of Amador, Werning, and Angeletos (2006).

Consider now the case of persistent shocks. Let the solution to (23)–(27) be denoted by

$$\{s^{L*}(\theta^i, V), s^{H*}(\theta^i, V), V^{L*}(\theta^i, V), V^{H*}(\theta^i, V)\}. \tag{29}$$

Lemma 2. *If shocks are persistent, the solution to (23)–(27) has the following properties for all $V \in [\underline{V}^i, \overline{V}^i]$:*

(i) ***(resetting)*** *$V^{L*}(\theta^i, V) = \overline{V}^L$ for $\theta^i \in \{\theta^L, \theta^H\}$;*

(ii) ***(monotonicity of threats)*** *$V^{H*}(\theta^H, V)$ is strictly decreasing in V and $V^{H*}(\theta^i, V)$ satisfies $\overline{V}^H > V^{H*}(\theta^H, \overline{V}^H) > V^{H*}(\theta^L, \overline{V}^L)$;*

(iii) ***(monotonicity of savings rates)*** *$s^{i*}(\theta^H, V)$ is strictly increasing in V and $s^{i*}(\theta^i, V)$ satisfies $s^{i*}(\theta^H, \overline{V}^H) < s^{i*}(\theta^L, \overline{V}^L)$ for $i \in \{L, H\}$.*

Proof. See Appendix C. □

Lemma 2 describes the solution to (23)–(27) given θ^i and V. The first part of the lemma states that the equilibrium at $t+1$ effectively "resets" if the low type is realized at t. This result is analogous to the result in the three-period example that the savings rate at date 1 is sequentially optimal if the low type is realized at date 0. As in that example, the intuition is that setting $V^{L*}(\theta^i, V) = \overline{V}^L$ maximizes the continuation payoff of the low type given a truthful report, and thus maximally relaxes the incentive compatibility constraint of the low type (25) while maximizing social welfare.

The second part of Lemma 2 concerns the magnitude of $V^{H*}(\theta^i, V)$ across different values of V. Comparing $V^{H*}(\theta^H, \overline{V}^H)$ and $V^{H*}(\theta^L, \overline{V}^L)$, the lemma states that the threat used in the ex-ante optimum to induce the low type to report truthfully is more severe if $\theta^i = \theta^L$ than if $\theta^i = \theta^H$. This result is analogous to the result in the three-period example that the savings rate at date 1 following $\theta_0 = \theta^H$ is lower if $\theta_{-1} = \theta^L$ than if $\theta_{-1} = \theta^H$. The intuition is also as in that example: when $\theta^i = \theta^L$, the low shock is more likely, so the benefit of using a more severe threat—namely, relaxing the low type's incentive constraint and curbing his spending—is larger relative to the cost—namely, reducing the high type's continuation welfare. The lemma also considers how $V^{H*}(\theta^H, V)$ varies with V. As discussed, the lower is V, the more "twisted" the effective objective function in (28) is in favor of welfare conditional on a high type, implying that threats are more costly and thus $V^{H*}(\theta^H, V)$ is higher.

The third part of Lemma 2 considers how the optimal savings rate depends on V. Comparing $s^{i*}(\theta^H, \overline{V}^H)$ and $s^{i*}(\theta^L, \overline{V}^L)$, the lemma states that the savings rate conditional on the realization of either type is lower if $\theta^i = \theta^H$ than if $\theta^i = \theta^L$. The intuition is straightforward. Because of Assumption 2, the savings rate of the low (high) type is always below (above) the first-best rate for that type. If $\theta^i = \theta^H$, the objective function in (28) puts a higher weight on welfare conditional on a high type, so the optimal savings rate for this type must be lower and closer to first best. But then for the incentive constraint of the low type (25) to be satisfied, it must be that the savings rate of the low type is also lower, implying that $s^{i*}(\theta^H, \overline{V}^H) < s^{i*}(\theta^L, \overline{V}^L)$ for $i \in \{L, H\}$. The lemma also compares $s^{i*}(\theta^H, V)$ across different values of V. The logic is the same as above: the lower is V, the higher is the weight in (28) assigned to the high type and, consequently, the lower is $s^{i*}(\theta^H, V)$.

We now use Lemma 2 to describe the ex-ante optimal fiscal rule under persistent shocks. Define $\eta_t(\theta^{t-1})$ as the number of periods since the last time that θ^L was realized:[31]

$$\eta_t(\theta^{t-1}) = k \;\text{ if }\; \theta_{t-1-k} = \theta^L \;\text{ and }\; \theta_{t-1-l} = \theta^H \text{ for all } l \in \{0, \dots, k-1\}.$$

Proposition 3 (**ex-ante optimum under persistent shocks**). *If shocks are persistent, the ex-ante optimum has the following features:*

(i) For all θ^t and θ^k with $\theta_t = \theta_k$, $s_t(\theta^t) = s_k(\theta^k)$ if $\eta_t(\theta^{t-1}) = \eta_k(\theta^{k-1})$;

(ii) There exist θ^t,θ^k with $\theta_t = \theta_k$ and $\theta_{t-1} = \theta_{k-1}$ for which $s_t(\theta^t) \neq s_k(\theta^k)$.

Thus, the ex-ante optimum does not coincide with the sequential optimum, and it exhibits history dependence.

[31] Note that this variable is not defined if the low shock has never been realized.

Proof. See Appendix C. □

Unlike the sequentially optimal rule or the optimal rule under i.i.d. shocks, the ex-ante optimal rule under persistent shocks exhibits history dependence: the mechanism at history θ^t depends not only on the payoff-relevant variables, θ_{t-1} and $b_t(\theta^{t-1})$, but also on when the low shock last occurred, and thus possibly on the entire history of shocks.[32] The logic stems from the fact that the mechanism provides dynamic incentives. Intuitively, because the shock at date $t-2$ predicts the realization of the shock at $t-1$, the shock at $t-2$ affects the relative tightness of incentive compatibility constraints at $t-1$, which in turn affect the policies that are chosen at t in providing dynamic incentives at $t-1$. The result in Lemma 2 that the mechanism resets following a low shock implies in fact that the tightness of current incentive constraints depends on when this resetting began, which explains why prescribed policies depend on the time that passed since a low shock was realized.

The resetting property of the ex-ante optimal rule is related to the "no distortion at the top" result of standard adverse selection models, although here dynamic considerations play an important role.[33] To see this, consider an equilibrium starting from date 0, given $\theta_{-1} = \theta^H$. If $\theta_0 = \theta^L$ is realized, the equilibrium transitions to the ex-ante optimum associated with $\theta_{-1} = \theta^L$; as explained, this is efficient because it maximally relaxes the incentive compatibility constraint of the low type at date 0. If instead $\theta_0 = \theta^H$ is realized, the fiscal rule at date 1 seeks to punish the low type at date 0 who would have lied, while at the same time not harming the truthful high type too much. Now note that there are two ways in which this can be done: on the one hand, spending at date 1 given a low type at date 1 can be made higher and further away from first best; on the other hand, the expected continuation welfare at date 1 from date 2 onward, given a low type at date 1, can be made lower. Because the low type at date 0 is more likely to be a low type at date 1, either of these changes hurt the deviating low type at date 0 more than the truthful high type at date 0. However, while the first option slackens the low type's incentive compatibility constraint at date 1, the second option tightens this constraint. Consequently, it is cheaper from a date 0 perspective to provide incentives to the low type by increasing spending for a single period following a high shock, while resetting the equilibrium thereafter given a low shock.

A natural question regards the equilibrium dynamics for a sequence of consecutive high shocks. These can be described using the second and third parts of Lemma 2, as shown in the following proposition.

[32]It is straightforward to show that for histories for which the low shock has never realized, policies are history dependent in the sense that they are a function of the date t.

[33]A similar resetting property arises in a different context in Hosseini, Jones, and Shourideh (2012); see in particular Section 4 of their paper.

Proposition 4 (dynamics under persistent shocks). *There exists $\widehat{V}$ such that for all θ^t, if $\theta_{t-1} = \theta_t = \theta^H$, then $V_{t+1}(\theta^t) > (<)\,\widehat{V}$ if $V_t(\theta^{t-1}) < (>)\,\widehat{V}$. Thus, the ex-ante optimum features oscillations in savings rates.*

Proof. See Appendix C. □

Combined with Lemma 2, Proposition 4 states that under a sequence of consecutive high shocks, the equilibrium oscillates between periods of high spending and periods of low spending, around some fiscal rule associated with a threat $\widehat{V}$ such that $V^{H*}(\theta^H, \widehat{V}) = \widehat{V}$.[34] In all of these periods, there is a lack of fiscal discipline in that the level of spending exceeds the level under $\theta^i = \theta^H$, $V = \overline{V}^H$.

To understand the oscillatory dynamics, consider again the equilibrium starting from date 0, given $\theta_{-1} = \theta^H$. As explained, if $\theta_0 = \theta^H$ is realized at date 0, the ex-ante optimal rule induces high spending at date 1 given a low type at date 1, where this spending is further away from first best compared to spending at date 0 given a low type at date 0. It follows that at date 1, the low type's incentives to lie and pretend to be a high type are relatively lower, and, in turn, a smaller threat at date 2 following a high type at date 1 is sufficient to provide incentives at date 1. Yet, note that the equilibrium does not reset to the ex-ante optimum at date 2; the reason is simply that if spending at date 2 corresponded to the ex-ante optimal spending level given $\theta_{-1} = \theta^H$, incentive compatibility constraints at dates 0 and 1 could be relaxed by increasing such spending, implying a first-order gain and causing only a second-order loss at date 2.

The economics behind the oscillatory dynamics in Proposition 4 emerge from the self-control nature of the problem. The absence of discipline tomorrow is used to induce discipline today. In turn, the absence of discipline tomorrow allows to increase discipline the day after while preserving incentives tomorrow.

In our analysis of the sequentially optimal rule in Subsection 5.1, we showed that such a rule can be implemented with a history-independent debt limit. That is, in each period, the government is allowed to borrow any amount up to some threshold, where this threshold depends only on the accumulated level of debt and the previous period's shock. A possible implementation of the ex-ante optimal rule is also with the use of debt limits, although, importantly, these limits would now depend on the history. In particular, in the ex-ante optimum, the debt limit up to which the government is allowed to borrow in any period would depend not only on how much the government borrowed up to that period (i.e., the current level of debt), but also on how much the government chose to borrow relative to the debt limits in previous periods.

[34] These oscillatory dynamics may imply convergence towards the fiscal rule associated with $V = \widehat{V}$, although this depends on the slope of the policy function.

5.3 Implications for Debt in the Long Run

The analysis of the previous sections shows that the savings rates induced by the ex-ante optimal and the sequentially optimal rule differ when shocks are persistent. What are the implications for the path of debt? Do these rules imply different levels of debt in the long run?

To characterize the debt dynamics, we consider the case where the market discounts the future at the same rate as the government (prior to the realization of the shock). This benchmark is useful because it highlights the inefficiencies that arise when the government is given discretion to choose policies. Thus, throughout this subsection, we assume

$$\delta = 1/(1+r). \tag{30}$$

Consider first the path of debt under the first-best and full-flexibility policies. The stochastic sequence of debt in the first best, $\mathbf{b}^{fb} \equiv (b_1^{fb}, b_2^{fb}, \ldots)$, is induced by a stochastic sequence of savings rates $\mathbf{s}^{fb} \equiv (s_0^{fb}, s_1^{fb}, \ldots)$ that satisfy $s_t^{fb} = s^{fb}(\theta^i)$ if $\theta_t = \theta^i$, where $s^{fb}(\theta^i)$ is defined in (8). The full-flexibility sequence of debt, $\mathbf{b}^f \equiv (b_1^f, b_2^f, \ldots)$, results when the government chooses its flexible optimum in each period taking into account the behavior of future governments which also choose policies flexibly. This sequence is induced by savings rates $\mathbf{s}^f \equiv (s_0^f, s_1^f, \ldots)$ satisfying $s_t^f = s^f(\theta^i)$ if $\theta_t = \theta^i$, where for $\theta^i \in \{\theta^L, \theta^H\}$,[35]

$$\theta^i U'(1 - s^f(\theta^i)) = \beta\widetilde{\theta}^i U'(s^f(\theta^i)). \tag{31}$$

As might be expected, the debt dynamics under these two policies are starkly different.

Lemma 3. *Assume (30) holds. In the long run,*

(i) Under first best, assets diverge to infinity: $b_t^{fb} \to -\infty$ *as* $t \to \infty$*;*

(ii) Under full flexibility, debt becomes maximal: $b_t^f \to \tau/(1-\delta)$ *as* $t \to \infty$*.*

Proof. See Appendix D. □

The first part of Lemma 3 follows from the self-insurance motive of the government. As is common across a wide class of self-insurance models (e.g., Chamberlain and Wilson, 2000), when condition (30) holds, this motive dominates in the long run, and the first-best policy for the government is to accumulate enough assets to perfectly insure itself with an infinite level of public spending. The second part of Lemma 3, however, shows that if given the option to flexibly choose policy, a present-biased government would not engage in such an accumulation

[35] Multiple equilibria can emerge when the government is given full flexibility in every period. Analogous to our selection of the sequential optimum, we select a unique solution by considering the limit of a finite horizon economy with end date T as T approaches infinity.

of infinite assets. To the contrary, in the long run, the government's present bias would cause the economy to accumulate the maximal amount of debt, pushing public spending towards zero. This stark contrast between the first-best and full-flexibility economies highlights the importance of fiscal rules to alleviate the government's time inconsistency problem.

Consider next the path of debt under the sequentially optimal rule. Let $\mathbf{b}^{so} \equiv (b_1^{so}, b_2^{so}, \ldots)$ be the sequence of debt in the sequential optimum characterized in Proposition 1.

Proposition 5 (long-run debt in sequential optimum). *Assume (30) holds. The sequentially optimal sequence of debt $\mathbf{b}^{so}$ has the following long-run properties:*

(i) Under i.i.d. shocks, the level of debt is constant: $b_t^{so} = b_{t-1}^{so}$ for all t;

(ii) Under persistent shocks, assets diverge to infinity: $b_t^{so} \to -\infty$ as $t \to \infty$.

Proof. See Appendix D. $\square$

In the sequential optimum, the government commits at date $t-1$ to a savings rate at date t. If shocks are i.i.d., the savings rate at t is independent of the government's information at $t-1$, and therefore constant over time. In fact, using condition (21), this rate is equal to the discount factor δ, and hence, given (30), the level of debt is also constant over time. In contrast, when shocks are persistent, the savings rate at date t depends on the information that the government has at $t-1$. In particular, if $\theta_{t-1} = \theta^L$ is realized, the government anticipates lower spending needs at t, and thus commits to saving more at t to insure itself for the future. Analogous to the first-best case above, it is this self-insurance motive, combined with (30), what makes the government eventually accumulate an infinite amount of assets.

Finally, we consider the path of debt under the ex-ante optimal rule. Let $\mathbf{b}^{eo} \equiv (b_1^{eo}, b_2^{eo}, \ldots)$ be the sequence of debt in the ex-ante optimum, induced by a sequence of savings rates $\mathbf{s}^{eo} \equiv (s_0^{eo}, s_1^{eo}, \ldots)$ where s_t^{eo} is characterized by Proposition 2 when shocks are i.i.d. and by Proposition 3 when shocks are persistent. From Proposition 3, the savings rate at date t under persistent shocks is a function only of the current shock and the number of periods since the last low shock. Let $s^{eo}(\theta^i, \eta)$ be the savings rate when the current shock is θ^i and the number of periods since θ^L was last realized is $\eta \geq 0$. Clearly, $s_t^{eo} = s^{eo}(\theta_t, \eta_t)$. Define

$$\begin{aligned}\bar{s} = & \frac{1}{2}\left[p(\theta^i|\theta^i)\log\big(s^{eo}(\theta^L,0)/\delta\big) + (1-p(\theta^i|\theta^i))\log\big(s^{eo}(\theta^H,0)/\delta\big)\right] \\ & + \frac{1}{2}(1-p(\theta^i|\theta^i))\sum_{k=1}^{\infty} p(\theta^i|\theta^i)^{k-1}\left[p(\theta^i|\theta^i)\log\big(s^{eo}(\theta^H,k)/\delta\big) + (1-p(\theta^i|\theta^i))\log\big(s^{eo}(\theta^L,k)/\delta\big)\right].\end{aligned}$$

$\bar{s}$ corresponds to the mean of $\log(s_t/\delta)$ in the long-run ergodic distribution of savings rates implied by the ex-ante optimum.[36]

[36] See Appendix D for details.

Proposition 6 (long-run debt in ex-ante optimum). *Assume (30) holds. The ex-ante optimal sequence of debt* $\mathbf{b}^{eo}$ *has the following long-run properties:*

(i) Under i.i.d. shocks, the level of debt is constant: $b_t^{eo} = b_{t-1}^{eo}$ *for all* t*;*

(ii) Under persistent shocks, if $\bar{s} > 0$*, assets diverge to infinity:* $b_t^{eo} \to -\infty$ *as* $t \to \infty$*. If* $\bar{s} < 0$*, debt becomes maximal:* $b_t^{eo} \to \tau/(1-\delta)$ *as* $t \to \infty$*.*

Both $\bar{s} > 0$ *and* $\bar{s} < 0$ *hold for an open set of parameters* $\{\theta^L, \theta^H, p(\theta^i|\theta^i), \delta, \beta\}$ *satisfying Assumption 2.*

Proof. See Appendix D. □

If shocks are i.i.d., the ex-ante optimal and sequentially optimal rules coincide, so it follows from Proposition 5 that debt is constant. If instead shocks are persistent, the long-run path of debt in the ex-ante optimum may entail accumulating infinite assets or maximal debt, depending on the long-run mean of $\log(s_t/\delta)$. The intuition stems from the interaction of two countervailing forces. On the one hand, there is an operational precautionary motive that pushes the government towards the accumulation of assets, just as in the first best and the sequential optimum. On the other hand, the ex-ante optimum features the provision of dynamic incentives, whereby phases of non-discipline along the equilibrium path are used to sustain discipline in earlier periods. These phases of non-discipline push the government towards the accumulation of debt, and can dominate the self-insurance motive if they involve sufficiently low savings rates. Therefore, while the ex-ante optimal rule yields higher ex-ante welfare than the sequentially optimal rule, it can induce the government to become immiserated in the long run, just as under full flexibility.[37]

To illustrate, Figure 3 considers two economies with different ex-ante optimal debt dynamics. The difference between the two economies is the degree of time inconsistency.[38] As discussed in the three-period example of Section 4, the government's present bias affects the extent to which the ex-ante optimal rule provides dynamic incentives: the larger this bias is, the less effective are threats of future non-discipline to provide incentives in the present, and thus the lower are the incentives that are optimally given. As a result, the figure shows that when the government's present bias is relatively large (that is, β is low), incentives are

[37] A similar force towards immiseration arises in the principal-agent models of Atkeson and Lucas (1992) and Thomas and Worrall (1990). As in our model, this force emerges as a consequence of dynamic incentive provision. However, note that those papers consider i.i.d. shocks, whereas in our self-control setting dynamic incentives are provided only when shocks are persistent. Moreover, in our model, dynamic incentives need not always result in immiseration, as the countervailing self-insurance motive pushes towards asset accumulation.

[38] We consider $\theta^L = 2$, $\theta^H = 3$, $p(\theta^i|\theta^i) = 0.55$, $\delta = 0.7$, $\beta' = 0.01$, $\beta'' = 0.4$, $b_0 = 0$, $\tau = 1$. Recall that the sequential optimum is independent of β. For the ex-ante optimum, the comparison across different values of β takes into account that, given other parameters, β must be low enough for Assumption 2 to be satisfied.

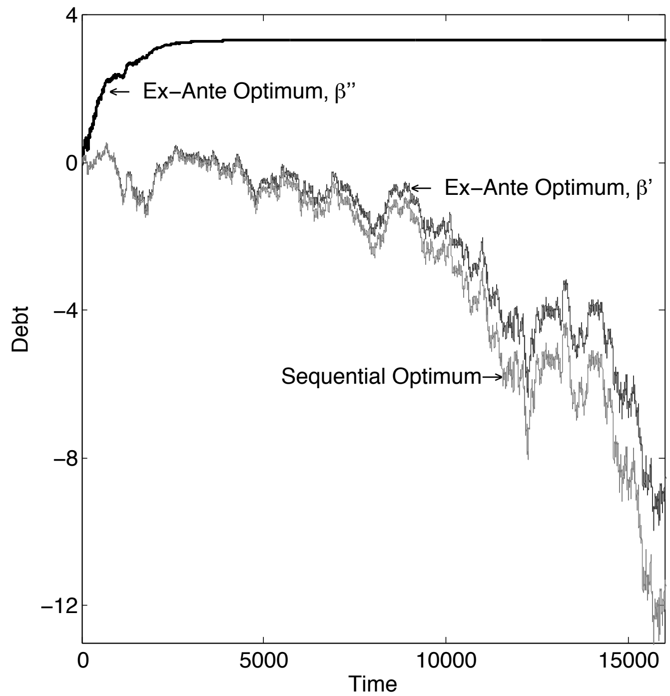

Figure 3: Path of debt in the ex-ante and sequentially optimal rules with $\delta = \frac{1}{1+r}$, $\beta'' > \beta'$.

low and the self-insurance motive dominates, inducing the government to accumulate infinite assets over time. Instead, when the government's present bias is relatively small (that is, β is high), the ex-ante optimal mechanism provides strong dynamic incentives, so periods of non-discipline are severe enough that the government becomes immiserated in the long run.

6 Extension to Continuum of Shocks

In this section, we extend our analysis to a setting with a continuum of shocks. We consider this setting not only to explore the robustness of our main results, but also because economies with a continuum of shocks are the main focus of the mechanism design literature that studies the tradeoff between commitment and flexibility.

The main complication that emerges in this extension is that, under multiple shocks, one must ensure that not only local but also global incentive compatibility constraints are satisfied. This does not complicate the analysis of the sequential optimum, as under log utility that problem reduces to a two-period problem. However, this does complicate the analysis of the ex-ante optimum, as the recursive method described in Subsection 5.2.1 no longer applies. Nonetheless, we show in this section that the main insights from the economy with two shocks continue to hold under a continuum of shocks.

6.1 Environment

Consider the benchmark environment described in Section 2 but with the government's type, $\theta_t > 0$, now being drawn from a continuous support $\Theta \equiv [\underline{\theta}, \overline{\theta}]$. Let $p(\theta_t|\theta_{t-1})$ and $\widetilde{\theta}_t$ be as previously defined, and assume that $p(\theta_t|\theta_{t-1})$ is strictly positive for all θ_t and θ_{t-1} and continuously differentiable with respect to θ_t and θ_{t-1}. Assumption 3 below, which holds in the two-shock economy, ensures that θ_t is mean reverting.

Assumption 3. *$\theta_t/\widetilde{\theta}_t$ is strictly increasing in θ_t.*

We also make a technical assumption regarding the distribution of shocks. Define $\theta_p(\theta_{t-1}) = \max\{\underline{\theta}, \theta'\}$ where θ' is the lowest $\theta \in \Theta$ such that for all $\theta'' \geq \theta$,

$$\frac{\widetilde{\theta}''}{\mathbb{E}[\widetilde{\theta}_t|\theta_t \geq \theta'', \theta_{t-1}]} \frac{\mathbb{E}[\theta_t|\theta_t \geq \theta'', \theta_{t-1}]}{\theta''} \leq \frac{1}{\beta}, \tag{32}$$

where $\widetilde{\theta}''$ is the value of (4) associated with $\theta_t = \theta''$. Note that if shocks are i.i.d., $\theta_p(\theta_{t-1}) = \theta_p$, independent of θ_{t-1}. Given our assumptions, $\theta_p(\theta_{t-1})$ is a continuously differentiable function of θ_{t-1}. Using this definition, we assume:

Assumption 4. *For all $\theta_t \leq \theta_p(\theta_{t-1})$,*

$$\frac{d\log p(\theta_t|\theta_{t-1})}{d\log\theta_t} \geq -\frac{2-\beta}{1-\beta} + \frac{1}{1-\beta}\frac{d\log\widetilde{\theta}_t}{d\log\theta_t} + \frac{d\log(d\widetilde{\theta}_t/d\theta_t)}{d\log\theta_t}. \tag{33}$$

Assumption 4 is isomorphic to Assumption A in Amador, Werning, and Angeletos (2006)'s study of an economy with i.i.d. shocks, where the main difference is that our condition incorporates the persistence of shocks through $d\log\widetilde{\theta}_t/d\log\theta_t$ and $d\log(d\widetilde{\theta}_t/d\theta_t)/d\log\theta_t$. This assumption is satisfied if the severity of the time-inconsistency problem is sufficiently low (i.e., β is sufficiently high) and the first and second derivatives of the density function $p(\theta_t|\theta_{t-1})$ with respect to both elements are bounded.[39]

Definitions for the ex-ante optimal and sequentially optimal fiscal rules analogous to those provided for the two-shock economy apply in this setting. We thus use the analysis of Section 3 to characterize the equilibrium.

6.2 Sequentially Optimal Rule

An analogous program to (16)–(17) defines the sequentially optimal fiscal rule. The incentive compatibility constraints (17) effectively imply that the problem is static and that global

[39] This follows from the fact that, given Assumption 3, $d\log\widetilde{\theta}_t/d\log\theta_t < 1$.

incentive constraints can be ignored. As such, analogous techniques to those used in the analysis of the two-period problem of Amador, Werning, and Angeletos (2006) apply here and can be used to characterize the sequential optimum. Let $s^f(\theta_t)$ be the flexible optimum of a government of type θ_t that is awarded full discretion as defined in (31).

Proposition 7 (sequential optimum under continuum of shocks). *Let $\underline{s}(\theta_{t-1})$ be defined by $\underline{s}(\theta_{t-1}) = s^f(\theta_p(\theta_{t-1}))$ if $\theta_p(\theta_{t-1}) > \underline{\theta}$, and*

$$\mathbb{E}[\theta_t|\theta_{t-1}]U'(1-\underline{s}(\theta_{t-1})) - \mathbb{E}[\widetilde{\theta}_t|\theta_{t-1}]U'(\underline{s}(\theta_{t-1})) = 0 \tag{34}$$

otherwise. For all θ^t, the sequential optimum features

$$s_t(\theta^t) = \max\{s^f(\theta_t), \underline{s}(\theta_{t-1})\}.$$

Proof. See Appendix E. □

Corollary 2. *The sequentially optimal rule at any date t can be implemented with a history-independent debt limit, $\overline{b}(\theta_{t-1}, b_t(\theta^{t-1}))$.*

This proposition and corollary state that if $\theta_p(\theta_{t-1}) > \underline{\theta}$, all types θ_t below $\theta_p(\theta_{t-1})$ are awarded full discretion, so they can choose their flexible optimal savings rate, and all types above $\theta_p(\theta_{t-1})$ are awarded no discretion, so they must choose the same savings rate as type $\theta_p(\theta_{t-1})$. If instead $\theta_p(\theta_{t-1}) = \underline{\theta}$, no type is given discretion, and all types are assigned a savings rate $\underline{s}(\theta_{t-1})$ satisfying (34).

The dependence of the minimum savings rate $\underline{s}(\theta_{t-1})$ on θ_{t-1} captures the fact that the shock at date $t-1$ provides information regarding the tradeoff between commitment and flexibility at date t. Note that if shocks are i.i.d., $\underline{s}(\theta_{t-1})$ and the associated debt limit $\overline{b}(\theta_{t-1}, b_t(\theta^{t-1}))$ are independent of θ_{t-1}. Moreover, note that as β approaches 1, so that the time-inconsistency problem vanishes, $\theta_p(\theta_{t-1})$ approaches $\overline{\theta}$, and thus the sequentially optimal rule provides full discretion to all types.

As in the two-type case, the sequentially optimal rule does not provide dynamic incentives. It therefore takes the form of a set of allowable savings rates, from which the government chooses the one that is closest to its flexible optimum. To understand why very high types are given no discretion, note that because of the bounded distribution of shocks, allowing flexibility for these types has no ex-ante welfare gain—such very high types would be overborrowing under *any* realized shock, so there is no tradeoff between commitment and flexibility for them. $\theta_p(\theta_{t-1})$ can be interpreted as the type above which the value of flexibility is exceeded by the value of commitment.

To understand why all types below the cutoff $\theta_p(\theta_{t-1})$ are given full discretion, consider the alternative of having a mechanism that admits "holes," namely, where some interior interval of savings rates are not allowed. Given such a hole, the types whose flexible optimum is inside the hole would choose savings rates at the boundaries of the hole. Specifically, those inside the hole whose type is relatively high would reduce their savings by choosing the lower boundary—which is socially costly—and those inside the hole whose type is relatively low would increase their savings by choosing the upper boundary—which is socially beneficial. The resulting total change in welfare then depends on the slope of the density function; under Assumption 4, which effectively puts a lower bound on the elasticity of $p(\theta_t|\theta_{t-1})$ with respect to θ_t, the total welfare change of introducing a hole is always negative.

6.3 Ex-ante Optimal Rule

We next consider the ex-ante optimal fiscal rule. For this analysis, we assume that the mechanism at time t admits savings rates that are piecewise continuously differentiable with respect to the history θ^t.[40] As in the two-type case, we begin by considering the i.i.d. benchmark.

Proposition 8 (ex-ante optimum under continuum of i.i.d. shocks). *Suppose $p(\theta_t|\theta_{t-1})$ is independent of θ_{t-1}. Then the ex-ante optimum coincides with the sequential optimum.*

Proof. See Appendix E. □

The intuition for this result is analogous to that of the two-type case. Under i.i.d. shocks, any ex-post suboptimality affects welfare on and off the equilibrium path equally, and hence cannot enhance efficiency. Consequently, dynamic incentives are not provided and the ex-ante optimum coincides with the sequential optimum. It of course follows that when shocks are i.i.d., the ex-ante optimum can be implemented with a renegotiated debt limit.[41]

Consider now the case of persistent shocks, where $p(\theta_t|\theta_{t-1})$ depends on θ_{t-1}. We define persistence with the following condition.

Condition 1 (mechanism relevance of past information). *There is a positive measure of types θ_{t-1} and θ_t such that $\theta_p(\theta_{t-1}) > \underline{\theta}$, $s^f(\theta_t) > s^f(\theta_p(\theta_{t-1}))$, and $\underline{s}'(\theta_t) \neq 0$.*

Condition 1 concerns the sequence of minimum savings policies implied by the sequential optimum described in Proposition 7. It states that in the sequentially optimal fiscal rule, there is a positive measure of types θ_{t-1} and θ_t with the property that, given θ_{t-1}, the government of type θ_t at date t has full discretion, so it spends above first-best level, and, moreover,

[40] This assumption is without loss in our setting in the case of i.i.d. shocks, and it is used in other settings in related work such as Athey, Atkeson, and Kehoe (2005).

[41] Note that in contrast with the two-type case, Proposition 8 requires Assumption 4 on the distribution function. The reason is that, if this assumption is not satisfied, the sequential optimum can admit a hole, in which case dynamic incentives may be provided to those types who bunch at the lower boundary of the hole to induce them to borrow less.

such government has information which is locally relevant regarding the sequentially optimal mechanism at $t+1$ (i.e., $\underline{s}'(\theta_t) \neq 0$). The analog of Condition 1 is trivially satisfied in a two-type economy, as the low type always spends above first best in the sequential optimum, and this type also has information about the future sequentially optimal mechanism. In the setting with a continuum of types, Condition 1 is always satisfied if $\underline{s}'(\theta_t) \neq 0$ for some positive measure of types θ_t, so that current information is relevant to future mechanisms, and if β is sufficiently close to 1, so that such types θ_t with mechanism-relevant information have full discretion at t starting from some θ_{t-1}.

While an explicit characterization of the ex-ante optimal rule becomes more complicated under a continuum of persistent shocks, we show that under Condition 1, this rule does not coincide with the sequentially optimal rule, and it exhibits history dependence.

Proposition 9 (ex-ante optimum under continuum of persistent shocks). *Suppose Condition 1 is satisfied. The ex-ante optimum has the following features:*

(i) It does not coincide with the sequential optimum, and

(ii) It exhibits history dependence: there exist θ^t and θ^k with $\theta_t = \theta_k$ and $\theta_{t-1} = \theta_{k-1}$ for which $s_t(\theta^t) \neq s_k(\theta^k)$.

Proof. See Appendix E. □

The intuition for Proposition 9 is familiar from the two-type economy. For the first part of the proposition, suppose by contradiction that the ex-ante optimum coincided with the sequential optimum. We show that ex-ante welfare can then be improved with a perturbation that induces ex-post suboptimality. For concreteness, suppose that $\theta_p(\theta_{-1}) > \underline{\theta}$ for all θ_{-1} and $\theta_p'(\theta_{-1}) > 0$, so the sequentially optimal rule at date 0 becomes more relaxed the higher is the shock θ_{-1}. Consider perturbing the mechanism by assigning at date 1 the fiscal rule associated with a cutoff $\theta_p(\theta_0 + \mu(\theta_0))$, for some $\mu(\theta_0) > 0$ arbitrarily small. This relaxes incentive compatibility constraints at date 0, allowing to increase the savings rates assigned to all types $\theta_0 < \theta_p(\theta_{-1})$, from $s^f(\theta_0)$ to $s^f(\theta_0) + \varepsilon(\theta_0)$, for some $\varepsilon(\theta_0) > 0$ that satisfies incentive compatibility. By envelope arguments analogous to those of Section 4, we can show that ex-ante welfare increases as a result: the first-order gain of bringing savings closer to first best at date 0 outweighs the second-order loss of assigning suboptimal rules at date 1. Therefore, as in the two-type case, the ex-ante optimal mechanism uses a threat of lack of fiscal discipline in the future to induce discipline today.

The logic for the second part of Proposition 9 is also analogous to that used for the two-type economy. History dependence arises because the mechanism provides dynamic incentives. That is, because the shock at date $t-2$ predicts the realization of the shock at $t-1$, the shock at $t-2$ affects the relative tightness of incentive compatibility constraints at $t-1$, and therefore it affects the policies that are chosen at t to provide dynamic incentives to the government at $t-1$.

7 Concluding Remarks

This paper has studied the role of persistence in determining the optimal structure of fiscal rules. We showed that when the shocks to the economy are i.i.d., the ex-ante optimal fiscal rule coincides with the sequentially optimal rule, taking the simple form of a renegotiated debt limit. In contrast, when the shocks are persistent, the ex-ante optimal rule is no longer sequentially optimal; this rule now provides dynamic incentives and exhibits history dependence. The ex-ante optimal mechanism features rich dynamics, with high shocks leading to an erosion of future fiscal discipline while low shocks reinstate discipline, and with policy distortions oscillating over time given a sequence of high shocks. Moreover, while the sequentially optimal rule leads to the accumulation of infinite assets over time, the ex-ante optimal rule may induce the government to accumulate maximal debt and become immiserated in the long run.

We believe our paper leaves interesting questions for future research. First, we have considered an environment in which the government has the ability to commit within the period to a fiscal rule. A natural question is whether, in the absence of commitment power, such a rule can be self-enforced. One can show that in our two-type economy, under both the sequentially optimal and the ex-ante optimal rule, social welfare in each period on the equilibrium path is strictly higher than welfare under full flexibility. In fact, full flexibility to choose policy is the worst punishment that can be imposed on the government. It follows from standard folk theorems then that either of these rules can be self-enforced by the threat of reversion to full flexibility if society's discount factor (δ) is high enough. It would be interesting to explore the additional distortions that emerge when the discount factor is relatively low.

Another possible extension would be to consider more general time-inconsistent preferences with hyperbolic discounting. The nature of the problem would change compared to our quasi-hyperbolic setting, as the preferences of the current government regarding future policies would no longer coincide with those of society. As the government's bias extends to future periods, in fact, the problem becomes one of delegation. Finally, our model can be enriched to consider more general social preferences for public spending, as well as a micro-founded economic environment taking into account the endogeneity of interest rates and the distortionary effects of taxation. A natural future direction would be to quantitatively assess the properties of optimal fiscal rules and how they depend on the economic structure.

Appendix

A Sequentially Optimal Rule

In this section, we consider the problem defined by (13)–(15) for a finite horizon T-period economy and show that as $T \to \infty$, this program is equivalent to (16)–(17).

Define

$$J_T^T(\theta_{T-1}, b_T) = \sum_{\theta_T \in \{\theta^L, \theta^H\}} p(\theta_T|\theta_{T-1})\,\theta_T U(\tau - b_T) \tag{A.1}$$

as the expected social welfare at date T conditional on θ_{T-1} and b_T, given the budget constraint (1) and given the end date T. Using $J_T^T(\theta_{T-1}, b_T)$, define the sequential optimum recursively as a solution to:

$$J_t^T(\theta_{t-1}, b_t) = \max_{\{g_t(\theta_t), b_{t+1}(\theta_t)\}_{\theta_t \in \{\theta^L, \theta^H\}}} \sum_{\theta_t \in \{\theta^L, \theta^H\}} p(\theta_t|\theta_{t-1}) \left(\theta_t U(g_t(\theta_t)) + \delta J_{t+1}^T(\theta_t, b_{t+1}(\theta_t))\right) \tag{A.2}$$

s.t.

$$g_t(\theta_t) = \tau + b_{t+1}(\theta_t)/(1+r) - b_t \text{ and} \tag{A.3}$$

$$\theta_t U(g_t(\theta_t)) + \beta\delta J_{t+1}^T(\theta_t, b_{t+1}(\theta_t)) \geq \theta_t U(g_t(\widehat{\theta}_t)) + \beta\delta J_{t+1}^T(\theta_t, b_{t+1}(\widehat{\theta}_t)) \text{ for } \widehat{\theta}_t \neq \theta_t. \tag{A.4}$$

The only difference between (13)–(15) and (A.2)–(A.4) is that the latter takes into account the finite horizon, with $J_t^T(\theta_{t-1}, b_t)$ being welfare at t conditional on the end date T.

For $t < T$, define $\widetilde{\theta}_t^T$ analogously to (4):

$$\widetilde{\theta}_t^T = \sum_{k=1}^{T-t} \delta^k \mathbb{E}[\theta_{t+k}|\theta_t],$$

where it is clear that $\lim_{T\to\infty} \widetilde{\theta}_t^T = \widetilde{\theta}_t$, for $\widetilde{\theta}_t$ defined in (4).

With some abuse of notation, define the finite horizon savings rate analogously to (5):

$$g_t(\theta_t) = (1 - s_t(\theta_t))(\tau((1+r)^{T-t+1} - 1)/(r(1+r)^{T-t}) - b_t).$$

We now show that, given T, the program defined in (A.2)–(A.4) for any t is equivalent to

$$\max_{\{s_t(\theta_t)\}_{\theta_t \in \{\theta^L, \theta^H\}}} \sum_{\theta_t \in \{\theta^L, \theta^H\}} p(\theta_t|\theta_{t-1})\left(\theta_t U(1 - s_t(\theta_t)) + \widetilde{\theta}_t^T U(s_t(\theta_t))\right) \tag{A.5}$$

s.t.

$$\theta_t U(1 - s_t(\theta_t)) + \beta\widetilde{\theta}_t^T U(s_t(\theta_t)) \geq \theta_t U(1 - s_t(\widehat{\theta}_t)) + \beta\widetilde{\theta}_t^T U(s_t(\widehat{\theta}_t)) \text{ for } \widehat{\theta}_t \neq \theta_t. \tag{A.6}$$

We show this by induction. First note that (A.2)–(A.4) is equivalent to (A.5)–(A.6) for $t = T-1$. This follows from the fact that $J_T^T(\theta_{T-1}, b_T)$ satisfies

$$\delta J_T^T(\theta_{T-1}, b_T) = \widetilde{\theta}_{T-1}^T U(s_{T-1}(\theta_{T-1})) + \chi_{T-1}^T(b_{T-1}), \tag{A.7}$$

for some constant $\chi^T_{T-1}(b_{T-1})$ which depends only on b_{T-1}. Substitution of (A.7) into the program (A.2)–(A.4) at $t = T-1$ implies that (A.2)–(A.4) is equivalent to (A.5)–(A.6) at $t = T-1$. Moreover, it also implies that

$$\delta J^T_{T-1}(\theta_{T-2}, b_{T-1}) = \widetilde{\theta}^T_{T-2} U(s_{T-2}(\theta_{T-2})) + \chi^T_{T-2}(b_{T-2}) \tag{A.8}$$

for some constant $\chi^T_{T-2}(b_{T-2})$ which depends only on b_{T-2}. Taking (A.8) into account, forward iteration of this reasoning implies that (A.2)–(A.4) is equivalent to (A.5)–(A.6) for all t. Taking $T \to \infty$, it is clear that (A.5)–(A.6) converges to (16)–(17), completing the argument.

B Proofs for Subsection 5.1

B.1 Proofs of Proposition 1 and Corollary 1

Consider the problem defined by (16)–(17). We first show that Assumption 2 implies that the solution to this problem admits pooling. We use the definition of the first-best benchmark given in (8) and we define the full flexibility benchmark as a sequence of savings rates $\mathbf{s}^f \equiv (s^f_0, s^f_1, \ldots)$ satisfying $s^f_t = s^f(\theta^i)$ if $\theta_t = \theta^i$, where for $\theta^i \in \{\theta^L, \theta^H\}$,

$$\theta^i U'(1 - s^f(\theta^i)) = \beta \widetilde{\theta}^i U'(s^f(\theta^i)). \tag{B.1}$$

Assumption 2 together with the concavity of $U(\cdot)$ imply

$$s^f(\theta^H) < s^f(\theta^L) < s^{fb}(\theta^H) < s^{fb}(\theta^L). \tag{B.2}$$

Hence, (17) must bind for some type in the solution to (16)–(17); otherwise, the solution would satisfy (8), but given (B.2) and the strict concavity of $U(\cdot)$, (17) would then be violated.

Note furthermore that the solution must admit $s_t(\theta^H) \leq s_t(\theta^L)$. This follows since pooling both types to the same savings rate must weakly lower welfare. Specifically, a perturbation which assigns the low type's savings rate to the high type must weakly lower welfare:

$$\frac{\theta^H}{\widetilde{\theta}^H}(U(1 - s_t(\theta^H)) - U(1 - s_t(\theta^L))) + (U(s_t(\theta^H)) - U(s_t(\theta^L))) \geq 0. \tag{B.3}$$

Analogously, a perturbation which assigns the high type's savings rate to the low type must weakly lower welfare:

$$\frac{\theta^L}{\widetilde{\theta}^L}(U(1 - s_t(\theta^L)) - U(1 - s_t(\theta^H))) + (U(s_t(\theta^L)) - U(s_t(\theta^H))) \geq 0. \tag{B.4}$$

Adding (B.3) and (B.4) gives

$$\left(\frac{\theta^H}{\widetilde{\theta}^H} - \frac{\theta^L}{\widetilde{\theta}^L}\right)(U(1 - s_t(\theta^H)) - U(1 - s_t(\theta^L))) \geq 0,$$

which requires that $s_t(\theta^H) \leq s_t(\theta^L)$.

Finally, note that it must be that (17) is an equality for $\theta_t = \theta^L$. Suppose instead that it were a strict inequality. Then given the above argument, (17) must be an equality for $\theta_t = \theta^H$. Substitution of (17) into the left hand side of (B.3) then implies

$$(1-\beta)\left[U(s_t(\theta^H)) - U(s_t(\theta^L))\right] \geq 0, \tag{B.5}$$

which implies $s_t(\theta^H) \geq s_t(\theta^L)$. Given that $s_t(\theta^H) \leq s_t(\theta^L)$, this means that $s_t(\theta^H) = s_t(\theta^L)$, and thus (17) is an equality for $\theta_t = \theta^L$.

To establish that there is pooling, suppose that there is separation so that (17) is ignored for the high type, and consider the first order conditions to (16)–(17). Let ϕ correspond to the Lagrange multiplier on (17) for the low type. First order conditions yield:

$$\frac{1 + \dfrac{\phi}{p(\theta^L|\theta_{t-1})}}{1 + \beta\dfrac{\phi}{p(\theta^L|\theta_{t-1})}} = \frac{\widetilde{\theta}^L U'(s_t(\theta^L))}{\theta^L U'(1-s_t(\theta^L))}, \tag{B.6}$$

$$\frac{1 - \dfrac{\theta^L}{\theta^H}\dfrac{\phi}{p(\theta^H|\theta_{t-1})}}{1 - \beta\dfrac{\widetilde{\theta}^L}{\widetilde{\theta}^H}\dfrac{\phi}{p(\theta^H|\theta_{t-1})}} = \frac{\widetilde{\theta}^H U'(s_t(\theta^H))}{\theta^H U'(1-s_t(\theta^H))}. \tag{B.7}$$

Equation (B.6) implies $s_t(\theta^L) < s^{fb}(\theta^L)$. Moreover, note that Assumption 2 implies

$$\frac{\theta^L}{\widetilde{\theta}^L} > \beta\frac{\theta^H}{\widetilde{\theta}^H}, \tag{B.8}$$

and thus (B.7) implies $s_t(\theta^H) > s^{fb}(\theta^H)$. Therefore, from (B.2), this means that $s^f(\theta^L) < s_t(\theta^H) < s_t(\theta^L)$, but this violates (17) since the low type can make itself strictly better off by claiming to be a high type. Therefore, (17) must hold with equality for both types and the mechanism features pooling.

Since the equilibrium features pooling, it follows that the optimal pooling level necessarily satisfies first order condition (21) where the pooling level depends only on θ_{t-1}. Since the solution to (21) admits a savings rate between $s^{fb}(\theta^H)$ and $s^{fb}(\theta^L)$, and since this savings rate is then above both $s^f(\theta^H)$ and $s^f(\theta^L)$, it follows that the equilibrium can be implemented with a required minimum savings rate that solves (21), which both types would choose. From (1), it is clear that requiring such a minimum savings rate is equivalent to imposing a maximum debt limit, which proves the corollary.■

C Proofs for Subsection 5.2

C.1 Proof of Lemma 1

In order to prove this result, we begin by considering a relaxed program which effectively lets (1) be a weak inequality so as to allow for money burning:

$$Q\left(\theta^i, Z\right) = \max_{\{u^L, y^L, u^H, y^H, Z^L, Z^H\}} \left\{ \begin{array}{c} p(\theta^L|\theta^i)(\theta^L u^L + \widetilde{\theta}^L y^L + \delta Q(\theta^L, Z^L)) \\ +p(\theta^H|\theta^i)(\theta^H u^H + \widetilde{\theta}^H y^H + \delta Q(\theta^H, Z^H)) \end{array} \right\} \tag{C.1}$$

s.t.

$$U^{-1}(u^L) + U^{-1}(y^L) \leq 1, \tag{C.2}$$

$$U^{-1}(u^H) + U^{-1}(y^H) \leq 1, \tag{C.3}$$

$$Z = \left\{ \begin{array}{c} p(\theta^L|\theta^{-i})(\theta^L u^L + \widetilde{\theta}^L y^L + \delta Q(\theta^L, Z^L)) \\ +p(\theta^H|\theta^{-i})(\theta^H u^H + \widetilde{\theta}^H y^H + \delta Q(\theta^H, Z^H)) \end{array} \right\}, \tag{C.4}$$

$$\theta^L u^L + \beta\widetilde{\theta}^L y^L + \beta\delta Q(\theta^L, Z^L) \geq \theta^L u^H + \beta\widetilde{\theta}^L y^H + \beta\delta Z^H, \tag{C.5}$$

$$\theta^H u^H + \beta\widetilde{\theta}^H y^H + \beta\delta Q(\theta^H, Z^H) \geq \theta^H u^L + \beta\widetilde{\theta}^H y^L + \beta\delta Z^L, \tag{C.6}$$

$$\underline{Z}^L \leq Z^L \leq \overline{Z}^L, \text{ and } \underline{Z}^H \leq Z^H \leq \overline{Z}^H. \tag{C.7}$$

(C.1)–(C.7) is identical to (23)–(27) if (C.2) and (C.3) hold with equality, in which case $u^i = U(1 - s^i)$ and $y^i = U(s^i)$. We will eventually establish that in the solution, (C.2) and (C.3) hold with equality, and thus (C.1)–(C.7) is identical to (23)–(27). Note that, in this program, $\underline{Z}^i$ and $\overline{Z}^i$ are respectively the lowest and highest values of Z given θ^i for which a solution to the program exists. We establish some preliminary lemmas regarding the solution to the relaxed program.

Lemma 4. *$Q(\theta^i, Z)$ is weakly concave in $Z \in [\underline{Z}^i, \overline{Z}^i]$.*

Proof. Consider the sequence problem defined by (C.1)–(C.7), and let

$$\gamma^*(\theta^i, Z) = \left\{ \left\{ u(\theta^t), y(\theta^t) \right\}_{\theta^t \in \Theta^t} \right\}_{t=0}^{\infty}$$

correspond to the stochastic sequence of u's and y's defined by forward iteration which solves (C.1)–(C.7) for some given $\theta_{-1} = \theta^i$ and $Z_0 = Z$. Suppose that $Z' \in [\underline{Z}^i, \overline{Z}^i]$ and $Z'' \in [\underline{Z}^i, \overline{Z}^i]$ with $Z'' > Z'$, and consider the solution to the sequence problem for $Z = \kappa Z' + (1-\kappa) Z'' \equiv Z^\kappa$. Define a potential solution

$$\gamma^\kappa = \kappa\gamma^*(\theta^i, Z') + (1 - \kappa)\gamma^*(\theta^i, Z'').$$

γ^κ is a convex combination of the stochastic u and y sequences under Z' and Z''. Because the set of u and y sequences that satisfy the sequence analogue of the constraint set in (C.2)–(C.7) is convex, it follows that γ^κ satisfies the constraints of the problem for $Z = Z^\kappa$. Therefore, the value of $Q(\theta^i, Z^\kappa)$ must be weakly greater than the welfare under γ^κ; that is:

$$Q(\theta^i, Z^\kappa) \geq \kappa Q(\theta^i, Z') + (1 - \kappa) Q(\theta^i, Z''). \tag{C.8}$$

It follows from (C.8) that $Q(\theta^i, Z)$ is weakly concave. □

Define $Z^i_{\max} = \arg\max_{Z \in [\underline{Z}^i, \overline{Z}^i]} Q(\theta^i, Z)$. We now use Lemma 4 to characterize $Z^i_{\max}$ and $Q(\theta^i, Z^i_{\max})$.

Lemma 5. *The solution to* (C.1)–(C.7) *for* $Z = Z^i_{\max}$ *has the following properties:*

(i) (C.5) *holds with equality and* (C.6) *holds as a strict inequality,*

(ii) (C.2) *and* (C.3) *both hold with equality.*

Proof. **Proof of part (i).** We establish this result in three steps. We take into account that the solution to (C.1)–(C.7) for $Z^i = Z^i_{max}$ is equivalent to the solution to (C.1)–(C.7) ignoring (C.4).

Step 1. Either (C.5) or (C.6) hold with equality. Suppose by contradiction that this is not the case. Then (C.5) and (C.6) can be ignored, and the solution admits the first-best allocation defined by (8) and $Z^i = Z^i_{\max}$ for $i = \{L, H\}$, as this maximizes welfare. This implies that the infinite repetition of the first best allocation is incentive compatible, which means that in fact $Q(\theta^i, Z^i) = Z^{-i}$ for $i = \{L, H\}$. But the same arguments as in the proof of Proposition 1 then imply that (C.5) is violated, leading to a contradiction.

Step 2. Suppose that in the solution, (C.6) holds with equality. Then there is bunching, with $u^L = u^H$ and $y^H = y^L$, and (C.2) and (C.3) both hold with equality. To see why, note first that the solution admits $u^H \geq u^L$ and $y^L \geq y^H$. This follows from the fact that a perturbation that changes the high type's allocation to the low type's (so that there is bunching) and sets $Z^i = Z^i_{\max}$ for $i = \{L, H\}$ must weakly lower welfare:

$$\theta^H(u^H - u^L) + \widetilde{\theta}^H(y^H - y^L) \geq 0. \tag{C.9}$$

It can be verified that this perturbation satisfies (C.2)–(C.3) and (C.5)–(C.7). Analogously, a perturbation that changes the low type's allocation to the high type's must weakly lower welfare:

$$\theta^L(u^L - u^H) + \widetilde{\theta}^L(y^L - y^H) \geq 0. \tag{C.10}$$

Adding (C.9) and (C.10) gives

$$\left(\frac{\theta^H}{\widetilde{\theta}^H} - \frac{\theta^L}{\widetilde{\theta}^L}\right)(u^H - u^L) \geq 0 \text{ and } \left(\frac{\widetilde{\theta}^L}{\theta^L} - \frac{\widetilde{\theta}^H}{\theta^H}\right)\left(y^L - y^H\right) \geq 0,$$

which requires $u^H \geq u^L$ and $y^L \geq y^H$.

Next, note that a perturbation that changes the high type's allocation and continuation allocation to be the same as the low type's must also weakly lower welfare:

$$\theta^H(u^H - u^L) + \widetilde{\theta}^H(y^H - y^L) + \delta(Q(\theta^H, Z^H) - Z^L) \geq 0.$$

Substitution of (C.6) holding with equality into the above condition gives

$$\left(1 - \frac{1}{\beta}\right)\theta^H(u^H - u^L) \geq 0,$$

which means that $u^H \leq u^L$. Combined with the result above, this implies $u^H = u^L$.

Finally, note that conditional on $u^H = u^L$, setting $y^H = y^L$, so that (C.2) and (C.3) hold with equality and $Z^i = Z^i_{\max}$ for $i = \{L, H\}$, yields the highest feasible welfare. Moreover, such a solution satisfies (C.5) and (C.6), so it is incentive compatible. Therefore, there is bunching and the resource constraints hold with equality.

Step 3. A solution where (C.6) holds with equality is suboptimal. From Step 2, such a solution admits bunching, implying that $Q(\theta^H, Z^H_{\max}) = Z^L_{\max}$. By construction, it must also be that $Q(\theta^L, Z^L_{\max}) = Z^H_{\max}$. Thus, the solution admits the same repeated level of bunching which is independent of the previous period's shock. However, this level of bunching yields a welfare strictly below that achieved in the sequential optimum described in Proposition 1, where the level of bunching depends on the previous period's shock. Therefore, (C.6) holds as a strict inequality in the solution, and by Step 1, (C.5) holds with equality.

Proof of part (ii). By the previous steps, the solution is such that (C.5) holds as an equality and (C.6) as an inequality. We now show that (C.2) and (C.3) hold with equality. Suppose first that (C.2) is a strict inequality. Consider a perturbation that decreases u^L and increases y^L while holding $\theta^L u^L + \beta\widetilde{\theta}^L y^L$ constant. This perturbation clearly increases welfare. Moreover, this perturbation leaves (C.5) unaffected, and for a small enough perturbation, (C.6) continues to hold. Therefore, this perturbation is incentive compatible and strictly increases welfare. Suppose next that (C.3) is an inequality. Consider a perturbation that decreases u^H and increases y^H while holding $\theta^L u^H + \beta\widetilde{\theta}^L y^H$ constant. It follows from Assumption 2, which implies (B.8), that this perturbation increases welfare. Moreover, this perturbation leaves (C.5) unaffected, and for a small enough perturbation, (C.6) continues to hold. Therefore, this perturbation is incentive compatible and strictly increases welfare.

Finally, we show that in the solution, $u^L < u^H$ and $y^L > y^H$. From Step 2 in the proof of part (i), $u^L \leq u^H$ and $y^L \geq y^H$. Since (C.2) and (C.3) hold with equality, if it is the case that $u^L = u^H$, then it must be that $y^L = y^H$, and vice versa, so that there would be bunching. However, by Step 3 of the proof of part (i), a solution that admits bunching is suboptimal. It thus follows that $u^L < u^H$ and $y^L > y^H$. $\square$

We now use the fact that $Z^i_{\max}$ characterizes the global optimum of $Q(\theta^i, Z)$ to characterize $Q(\theta^i, Z)$.

Lemma 6. *$Q(\theta^i, Z)$ has the following properties.*

(i) $\exists Z^i_{\min} \in [\underline{Z}^i, Z^i_{\max})$ s.t. $Q(\theta^i, Z)$ is continuously differentiable in Z if $Z \in (Z^i_{\min}, Z^i_{\max}]$,

(ii) If $Z \in (Z^i_{\min}, Z^i_{\max})$, $Q(\theta^i, Z)$ is strictly concave in Z.

Proof. We establish these results in four steps.

Step 1. We first establish that $Q(\theta^i, Z)$ is continuously differentiable in Z at $Z = Z^i_{\max}$ with a derivative equal to zero. Consider the solution to (C.1)–(C.7) for $Z = Z^i_{\max}$. Let s^L and s^H correspond to the associated savings rates in the solution, where by the proof of part (ii) of Lemma 5, $s^H < s^L$. Define $s^L_\varepsilon = s^L + \varepsilon$ and s^H_ε as the solution to

$$\begin{aligned}&\theta^L(U(1-s^L) - U(1-s^H)) + \beta\widetilde{\theta}^L(U(s^L) - U(s^H))\\ &= \theta^L(U(1-s^L_\varepsilon) - U(1-s^H_\varepsilon)) + \beta\widetilde{\theta}^L(U(s^L_\varepsilon) - U(s^H_\varepsilon)).\end{aligned} \tag{C.11}$$

Let

$$Z_\varepsilon = Z^i_{\max} + \left\{ \begin{array}{c} p(\theta^L|\theta^{-i})(\theta^L(U(1-s^L_\varepsilon) - U(1-s^L)) + \widetilde{\theta}^L(U(s^L_\varepsilon) - U(s^L))) \\ +p(\theta^H|\theta^{-i})(\theta^H(U(1-s^H_\varepsilon) - U(1-s^H)) + \widetilde{\theta}^H(U(s^H_\varepsilon) - U(s^H))) \end{array} \right\} \quad \text{(C.12)}$$

and

$$Q_\varepsilon = Q(\theta^i, Z^i_{\max}) + \left\{ \begin{array}{c} p(\theta^L|\theta^i)(\theta^L(U(1-s^L_\varepsilon) - U(1-s^L)) + \widetilde{\theta}^L(U(s^L_\varepsilon) - U(s^L))) \\ +p(\theta^H|\theta^i)(\theta^H(U(1-s^H_\varepsilon) - U(1-s^H)) + \widetilde{\theta}^H(U(s^H_\varepsilon) - U(s^H))) \end{array} \right\}. \quad \text{(C.13)}$$

From the definition of $Z^i_{\max}$, it must be that the derivative of Q_ε with respect to ε at $\varepsilon = 0$ is zero, as the objective attains a local maximum, where we have taken into account that the perturbation continues to satisfy (C.2)–(C.3) and (C.5)–(C.7). Note that it cannot be that the derivative of Z_ε with respect to ε at $\varepsilon = 0$ is also zero. Given that $p(\theta^L|\theta^{-i}) \neq p(\theta^L|\theta^i)$ and $p(\theta^H|\theta^{-i}) \neq p(\theta^H|\theta^i)$, this could be true if s^L and s^H are chosen at the first-best level, but this is also ruled out in Step 1 in the proof of part (i) of Lemma 5. Therefore, it cannot also be that Z_ε is at its local maximum. It thus follows that one can use this perturbation to apply Lemma 1 of Benveniste and Scheinkman (1979) to show that $Q(\theta^i, Z)$ is continuously differentiable in Z at $Z = Z^i_{\max}$ and has a derivative of zero.

Step 2. Define $Z^i_{\min}$ as follows. If

$$\lim_{\varepsilon>0,\varepsilon\to 0} \frac{Q(\theta^i, \underline{Z}^i + \varepsilon) - Q(\theta^i, \underline{Z}^i)}{\varepsilon} < \frac{1 - p(\theta^i|\theta^i)}{p(\theta^i|\theta^i)},$$

then $Z^i_{\min} = \underline{Z}^i$. Otherwise, $Z^i_{\min}$ is the highest point in $[\underline{Z}^i, \overline{Z}^i]$ such that

$$\lim_{\varepsilon>0,\varepsilon\to 0} \frac{Q(\theta^i, Z^i_{\min} + \varepsilon) - Q(\theta^i, Z^i_{\min})}{\varepsilon} \leq \frac{1 - p(\theta^i|\theta^i)}{p(\theta^i|\theta^i)} \leq \lim_{\varepsilon>0,\varepsilon\to 0} \frac{Q(\theta^i, Z^i_{\min}) - Q(\theta^i, Z^i_{\min} - \varepsilon)}{\varepsilon}, \quad \text{(C.14)}$$

where such a point necessarily exists given Step 1 and the concavity of $Q(\cdot)$. Note further that (C.1)–(C.7) can be represented by

$$\max_{\{u^L,y^L,u^H,y^H,Z^L,Z^H\}} \left\{ \begin{array}{c} (p(\theta^L|\theta^i) + \lambda p(\theta^L|\theta^{-i}))(\theta^L u^L + \widetilde{\theta}^L y^L + \delta Q(\theta^L, Z^L)) \\ +(p(\theta^H|\theta^i) + \lambda p(\theta^H|\theta^{-i}))(\theta^H u^H + \widetilde{\theta}^H y^H + \delta Q(\theta^H, Z^H)) \end{array} \right\} \quad \text{(C.15)}$$
$$\text{s.t. (C.2)} - \text{(C.3)} \quad \text{and} \quad \text{(C.5)} - \text{(C.7)}.$$

(C.15) corresponds (C.1)–(C.7) where λ is the Lagrange multiplier on the threat-keeping constraint (C.4). By the envelope condition,

$$\lim_{\varepsilon>0,\varepsilon\to 0} \frac{Q(\theta^i, Z + \varepsilon) - Q(\theta^i, Z)}{\varepsilon} \leq -\lambda \leq \lim_{\varepsilon>0,\varepsilon\to 0} \frac{Q(\theta^i, Z) - Q(\theta^i, Z - \varepsilon)}{\varepsilon}.$$

It thus follows from concavity and the definition of $Z^i_{\min}$ above that if $Z \in (Z^i_{\min}, Z^i_{\max}]$, then

$$p(\theta^L|\theta^i) + \lambda p(\theta^L|\theta^{-i}) > 0 \text{ and } p(\theta^H|\theta^i) + \lambda p(\theta^H|\theta^{-i}) > 0, \quad \text{(C.16)}$$

so that the objective in (C.15) puts positive weight on the realization of both states.

Step 3. Analogous arguments to those used in the proof of Lemma 5 imply that given (C.16), the solution to (C.15) has the following properties: (C.5) holds with equality and (C.6) as a strict inequality; (C.2) and (C.3) hold with equality; the solution admits $s^L > s^H$. Given these observations, a perturbation as the one used in Step 1 can be used here, and the same arguments from Step 1 imply that $Q(\theta^i, Z)$ is continuously differentiable in Z.

Step 4. We now show that $Q(\theta^i, Z)$ is strictly concave in this range. Consider the argument of concavity in the proof of Lemma 4. The potential solution yields welfare $\kappa Q(\theta^i, Z') + (1 - \kappa)Q(\theta^i, Z'')$, while leaving the resource constraints (C.2) and (C.3) as strict inequalities in the initial date. Given the arguments in Step 3, welfare can be strictly increased by letting (C.2) and (C.3) hold as equalities, implying that (C.8) must be a strict inequality. Thus, $Q(\theta^i, Z)$ is strictly concave. □

We now characterize the solution to the program for $Z \in (Z^i_{\min}, Z^i_{\max})$, where we denote by $Z^{L*}(\theta^i, Z)$ and $Z^{H*}(\theta^i, Z)$ the optimal values of Z^L and Z^H respectively given θ^i and Z.

Lemma 7. *If $Z \in (Z^i_{\min}, Z^i_{\max}]$, the solution to (C.1)–(C.7) has the following properties:*

(i) (C.2) *and* (C.3) *hold with equality,*

(ii) $Z^{L*}(\theta^i, Z) = Z^L_{\max}$ *for* $\theta^i \in \{\theta^L, \theta^H\}$,

(iii) $Z^{H*}(\theta^i, Z) < Z^H_{\max}$ *for* $\theta^i \in \{\theta^L, \theta^H\}$,

(iv) $Z^{H*}(\theta^H, Z)$ *is strictly decreasing in* Z*, and*

(v) $Z^{H*}(\theta^H, Z^H_{\max}) > Z^{H*}(\theta^L, Z^L_{\max}) > Z^H_{\min}$.

Proof. **Proof of part (i).** This follows from Step 3 in the proof of Lemma 6.

Proof of part (ii). Consider the problem as represented in (C.15). Suppose it were the case that $Z^{L*}(\theta^i, Z) < Z^L_{\max}$. Then a perturbation which moves Z^L in the direction of $Z^L_{\max}$ strictly increases welfare by increasing $Q(\theta^L, Z^L)$. Furthermore, it relaxes (C.5) and, since (C.6) is a strict inequality in the solution to the program (from Step 3 in the proof of Lemma 6), (C.6) continues to hold.

Proof of part (iii). Consider the solution to (C.15) given that (C.2) and (C.3) hold with equality. Let s^L and s^H correspond to the savings rates in the low and high shocks, respectively. Let ϕ be the Lagrange multiplier on (C.5). First order conditions with respect to s^L and s^H yield

$$\frac{U'(1-s^L)}{U'(s^L)} = \left(\frac{\widetilde{\theta}^L}{\theta^L} \frac{1 + \beta \dfrac{\phi}{p(\theta^L|\theta^i) + \lambda p(\theta^L|\theta^{-i})}}{1 + \dfrac{\phi}{p(\theta^L|\theta^i) + \lambda p(\theta^L|\theta^{-i})}} \right), \tag{C.17}$$

$$\frac{U'(1-s^H)}{U'(s^H)} = \left(\frac{\widetilde{\theta}^H - \beta \widetilde{\theta}^L \dfrac{\phi}{p(\theta^H|\theta^i) + \lambda p(\theta^H|\theta^{-i})}}{\theta^H - \theta^L \dfrac{\phi}{p(\theta^H|\theta^i) + \lambda p(\theta^H|\theta^{-i})}} \right). \tag{C.18}$$

Note that $\phi > 0$; otherwise, the same arguments as those used in Step 1 in the proof of part (i) of Lemma 5 would imply that the repeated first-best allocation is incentive compatible, leading to a contradiction. Finally, first order conditions with respect to Z^H yield

$$\begin{aligned}\lim_{\varepsilon>0,\varepsilon\to 0}\frac{Q(\theta^H,Z^H)-Q(\theta^H,Z^H-\varepsilon)}{\varepsilon} &\geq \beta\frac{\phi}{p(\theta^H|\theta^i)+\lambda p(\theta^H|\theta^{-i})} \\ &\geq \lim_{\varepsilon>0,\varepsilon\to 0}\frac{Q(\theta^H,Z^H+\varepsilon)-Q(\theta^H,Z^H)}{\varepsilon},\end{aligned} \tag{C.19}$$

where we have taken into account that Z^H may be below $Z^H_{\min}$, in which case $Q(\theta^H, Z^H)$ may not necessarily be differentiable. Given the definition of $Z^H_{\max}$ and the fact that $Q_Z(\theta^H, Z^H_{\max}) = 0$, it follows from (C.19) that $Z^{H*}(\theta^i, Z) < Z^H_{\max}$, since $\phi > 0$.

Proof of part (iv). Suppose that $\theta^i = \theta^H$ and consider $Z' \in (Z^H_{\min}, Z^H_{\max}]$ and $Z'' \in (Z^H_{\min}, Z^H_{\max}]$ with $Z'' > Z'$. Given part (ii) of Lemma 6, it follows that Z' is associated with multiplier λ' in (C.15) and Z'' is associated with multiplier λ'' in (C.15) where $\lambda' < \lambda'' < 0$. We will establish that $Z^{H*}(\theta^H, Z'') < Z^{H*}(\theta^H, Z')$. Suppose instead that $Z^{H*}(\theta^H, Z'') \geq Z^{H*}(\theta^H, Z')$. Let ϕ' and ϕ'' correspond to the Lagrange multipliers on (C.5) for the program for $Z = Z'$ and $Z = Z''$, respectively. From (C.19) together with the concavity of $Q(\cdot)$, it must be that

$$\frac{\phi'}{p(\theta^H|\theta^H)+\lambda' p(\theta^H|\theta^L)} \geq \frac{\phi''}{p(\theta^H|\theta^H)+\lambda'' p(\theta^H|\theta^L)}. \tag{C.20}$$

Substituting into (C.18), taking into account (B.8), this implies that the solution $s^{H*}(\cdot)$ must satisfy

$$s^{fb}(\theta^H) < s^{H*}(\theta^H, Z'') \leq s^{H*}(\theta^H, Z'), \tag{C.21}$$

where $s^{fb}(\theta^H)$ is defined in (8). Moreover, since $p(\theta^H|\theta^H)/p(\theta^H|\theta^L) > p(\theta^L|\theta^H)/p(\theta^L|\theta^L)$, it follows that

$$\frac{\phi'}{p(\theta^L|\theta^H)+\lambda' p(\theta^L|\theta^L)} > \frac{\phi''}{p(\theta^L|\theta^H)+\lambda'' p(\theta^L|\theta^L)}. \tag{C.22}$$

Substituting into (C.17) implies that the solution $s^{L*}(\cdot)$ must satisfy

$$s^{L*}(\theta^H, Z') < s^{L*}(\theta^H, Z'') < s^{fb}(\theta^L), \tag{C.23}$$

where $s^{fb}(\theta^L)$ is defined in (8). From Step 3 of the proof of Lemma 6,

$$s^{H*}(\theta^H, Z') < s^{L*}(\theta^H, Z'). \tag{C.24}$$

Note that $\theta^L U(1-s) + \beta\widetilde{\theta}^L U(s)$ is strictly decreasing in s for $s \geq s^{fb}(\theta^H) > s^f(\theta^L)$, where we have appealed to (B.2). Therefore, conditions (C.21), (C.23), and (C.24) combined together imply that

$$\begin{aligned} &\theta^L(U(1-s^{L*}(\theta^H,Z''))-U(1-s^{H*}(\theta^H,Z''))) + \beta\widetilde{\theta}^L(U(s^{L*}(\theta^H,Z''))-U(s^{H*}(\theta^H,Z''))) \\ < \;&\theta^L(U(1-s^{L*}(\theta^H,Z'))-U(1-s^{H*}(\theta^H,Z'))) + \beta\widetilde{\theta}^L(U(s^{L*}(\theta^H,Z'))-U(s^{H*}(\theta^H,Z'))).\end{aligned}$$

Given that (C.5) holds with equality and $Z^{L*}(\theta^i, Z) = Z^L_{\max}$ from part (ii), this implies

$$\beta\delta(Z^{H*}(\theta^H, Z'') - Q(\theta^L, Z^L_{\max})) < \beta\delta(Z^{H*}(\theta^H, Z') - Q(\theta^L, Z^L_{\max})),$$

which contradicts the fact that $Z^{H*}(\theta^H, Z'') \geq Z^{H*}(\theta^H, Z')$.

Proof of part (v). First note that $Z^{H*}(\theta^H, Z^H_{\max}) > Z^{H*}(\theta^L, Z^L_{\max})$ follows from analogous arguments to those used in the proof of part (iv), taking into account that the associated value of λ is 0 for $Z = Z^i_{\max}$, $i = \{L, H\}$. We are then left to show that $Z^{H*}(\theta^L, Z^L_{\max}) > Z^H_{\min}$. Suppose instead that $Z^{H*}(\theta^L, Z^L_{\max}) \leq Z^H_{\min}$. By the definition of $Z^H_{\min}$ in Lemma 6, using (C.19) and the fact that $\lambda = 0$ for $Z = Z^i_{\max}$, it follows that

$$\phi \geq \frac{1}{\beta}\frac{(1 - p(\theta^i|\theta^i))^2}{p(\theta^i|\theta^i)}. \tag{C.25}$$

Note that by the arguments in the proof of part (iv), $s^{H*}(\theta^L, Z^L_{\max}) < s^{fb}(\theta^L)$. From (C.18), taking into account that $\lambda = 0$, this means that

$$\frac{\theta^H}{\theta^L} - \frac{\widetilde{\theta}^H}{\widetilde{\theta}^L} > \phi\frac{(1-\beta)}{1 - p(\theta^i|\theta^i)}. \tag{C.26}$$

Combining (C.25) with (C.26) implies that

$$\frac{\theta^H}{\theta^L} - \frac{\widetilde{\theta}^H}{\widetilde{\theta}^L} > \frac{1 - p(\theta^i|\theta^i)}{p(\theta^i|\theta^i)}\left(\frac{1}{\beta} - 1\right). \tag{C.27}$$

But this contradicts Assumption 2, completing the proof. $\square$

By Lemma 7, the relaxed program in (C.1)–(C.7) is equivalent to the original program in (23)–(27) if we define $\underline{V}^i = Z^i_{\min}$ and $\overline{V}^i = Z^i_{\max}$. This is because (C.2) and (C.3) both hold with equality, and for $Z \in (Z^i_{\min}, Z^i_{\max}]$, $Z^{L*}(\theta^i, Z) \in (Z^L_{\min}, Z^L_{\max}]$ and $Z^{H*}(\theta^i, Z) \in (Z^H_{\min}, Z^H_{\max}]$, so that $\underline{V}^i$ and $\overline{V}^i$ effectively correspond to the minimum and maximum values of V that would ever be reached given that the equilibrium begins with the ex-ante optimum with $Z_0 = Z^i_{\max}$. It follows then from Lemma 6 that $W(\theta^i, V)$ is strictly increasing in V, strictly concave, and continuously differentiable in V over the range $(\underline{V}^i, \overline{V}^i)$.■

C.2 Proof of Proposition 2

If shocks are i.i.d., $p(\theta^L|\theta^i) = p(\theta^H|\theta^i) = 0.5$ and $W(\theta^i, V) = V$. Moreover, $\overline{V}^L = \overline{V}^H = \overline{V}$ and $\underline{V}^L = \underline{V}^H = \underline{V}$ since the value and solution to (23)–(27) is independent of θ^i. Now consider the program starting from $V = \overline{V}$, which is the solution to the ex-ante optimum. Let us denote the solution as in (29).

Analogous arguments to those used in the proof of Lemma 5 imply that in the solution, either (25) and (26) both hold as equalities with $s^{L*}(\theta^i, \overline{V}) = s^{H*}(\theta^i, \overline{V})$, so that there is bunching, or alternatively (25) holds as an equality and (26) as a strict inequality. We consider the latter case first and rule it out. Note that the same arguments as those used in the proof of Proposition 1 imply that if there is separation, the solution admits

$$s^{fb}(\theta^H) < s^{H*}(\theta^i, \overline{V}) < s^{L*}(\theta^i, \overline{V}) < s^{fb}(\theta^L). \tag{C.28}$$

This means that if the solution admits $V^{L*}(\theta^i, \overline{V}) = V^{H*}(\theta^i, \overline{V}) = \overline{V}$, then from (B.2), (25) would be violated. Therefore, for (25) to hold as an equality and (26) as a strict inequality, it is necessary that $V^{L*}(\theta^i, \overline{V}) < \overline{V}$ or $V^{H*}(\theta^i, \overline{V}) < \overline{V}$. It is clear that setting $V^{L*}(\theta^i, \overline{V}) = \overline{V}$ maximizes social welfare while fully relaxing (25), so the only possibility to consider is $V^{H*}(\theta^i, \overline{V}) < \overline{V}$. Consider an increase in V^H and an increase in s^H that leaves the following term unchanged:

$$\theta^L U(1 - s^H) + \beta \widetilde{\theta}^L U(s^H) + \beta\delta V^H.$$

Such a perturbation leaves (25) unchanged, and since (26) was satisfied with strict inequality, (26) continues to hold for a small enough perturbation. One can show that given (B.8) and (C.28), the change in welfare from an arbitrarily small perturbation takes the same sign as $\theta^L/\beta - \theta^H$ which is positive, so that the perturbation strictly increases welfare.

This implies that the equilibrium admits bunching with $s^{L*}(\theta^i, \overline{V}) = s^{H*}(\theta^i, \overline{V})$. Conditional on bunching, the optimal mechanism assigns $V^{L*}(\theta^i, \overline{V}) = \overline{V}$ and $V^{H*}(\theta^i, \overline{V}) = \overline{V}$ since (25) and (26) are trivially satisfied without the use of dynamic incentives. Therefore, the ex-ante optimum corresponds to the sequential optimum and Proposition 1 applies.■

C.3 Proof of Lemma 2

Parts (i) and (ii) follow directly from Lemma 7 in the solution to the relaxed problem. We are left then to prove part (iii). Let ϕ correspond to the Lagrange multiplier on (25). First order conditions with respect to s^L, s^H, and V^H, respectively, yield

$$\frac{U'(1-s^L)}{U'(s^L)} = \left(\frac{\widetilde{\theta}^L}{\theta^L} \frac{1 + \beta \dfrac{\phi}{p(\theta^L|\theta^i) + \lambda p(\theta^L|\theta^{-i})}}{1 + \dfrac{\phi}{p(\theta^L|\theta^i) + \lambda p(\theta^L|\theta^{-i})}} \right), \tag{C.29}$$

$$\frac{U'(1-s^H)}{U'(s^H)} = \left(\frac{\widetilde{\theta}^H - \beta\widetilde{\theta}^L \dfrac{\phi}{p(\theta^H|\theta^i) + \lambda p(\theta^H|\theta^{-i})}}{\theta^H - \theta^L \dfrac{\phi}{p(\theta^H|\theta^i) + \lambda p(\theta^H|\theta^{-i})}} \right), \tag{C.30}$$

$$W_V(\theta^i, V^H) = \beta \frac{\phi}{p(\theta^H|\theta^i) + \lambda p(\theta^H|\theta^{-i})}. \tag{C.31}$$

Consider first the case with $\theta^i = \theta^H$. Consider $V'' > V'$, so that from the envelope condition and the strict concavity of $W(\theta^i, V)$, the associated multipliers satisfy $\lambda' < \lambda'' < 0$. Since $V^{H*}(\theta^H, V)$ is strictly decreasing in V, this means from (C.31) that

$$\frac{\phi''}{p(\theta^H|\theta^i) + \lambda'' p(\theta^H|\theta^{-i})} > \frac{\phi'}{p(\theta^H|\theta^i) + \lambda' p(\theta^H|\theta^{-i})}.$$

Combining this inequality with (C.30) implies that $s^{H*}(\theta^i, V'') > s^{H*}(\theta^i, V')$. From parts (i) and (ii), we know that $V^{L*}(\theta^i, V'') = V^{L*}(\theta^i, V')$ and $V^{H*}(\theta^i, V'') < V^{H*}(\theta^i, V')$. Since (25) binds, this implies that

$$\begin{aligned}
& \theta^L \left(U(1 - s^{L*}(\theta^i, V'')) - U(1 - s^{L*}(\theta^i, V'))\right) + \beta\widetilde{\theta}^L \left(U(s^{L*}(\theta^i, V'')) - U(s^{L*}(\theta^i, V'))\right) \\
= \; & \theta^L \left(U(1 - s^{H*}(\theta^i, V'')) - U(1 - s^{H*}(\theta^i, V'))\right) + \beta\widetilde{\theta}^L \left(U(s^{H*}(\theta^i, V'')) - U(s^{H*}(\theta^i, V'))\right) \\
& +\beta \left(V^{H*}(\theta^i, V'') - V^{H*}(\theta^i, V')\right) \\
< \; & 0.
\end{aligned}$$

In order for this inequality to hold, it must be that $s^{L*}(\theta^i, V'') > s^{L*}(\theta^i, V')$, where we have used the arguments in the proof of part (iv) of Lemma 7 which require

$$s^{fb}(\theta^L) > s^{L*}(\theta^i, V) > s^{H*}(\theta^i, V) > s^{fb}(\theta^H) > s^f(\theta^L).$$

Finally, the claim that $s^{i*}(\theta^H, \overline{V}^H) < s^{i*}(\theta^L, \overline{V}^L)$ follows from analogous arguments to those above, taking into account that the associated value of λ is 0 for $V = \overline{V}^i$ for $i = \{L, H\}$.■

C.4 Proof of Proposition 3

Proof of part (i). From part (i) of Lemma 2, if $\eta_t(\theta^{t-1}) = 0$, then $s_t(\theta^{t-1}, \theta^i) = s^{i*}(\theta^L, \overline{V}^L)$ for $i = \{L, H\}$. Moreover, if $\eta_t(\theta^{t-1}) = 1$, then $s_t(\theta^{t-1}, \theta^i) = s^{i*}(\theta^H, V^{H*}(\theta^L, \overline{V}^L))$ for $i = \{L, H\}$. Forward iteration implies that if $\eta_t(\theta^{t-1}) = k$ for $k > 1$, then

$$s_t(\theta^{t-1}, \theta^i) = s^{i*}(\theta^H, V^{H*^{k-1}}(\theta^H, V^{H*}(\theta^L, \overline{V}^L))),$$

where with some abuse of notation, $V^{H*^{k-1}}(\cdot)$ corresponds to $k-1$ iterations of the operator $V^{H*}(\theta^H, \cdot)$. Therefore, $\eta_t(\theta^{t-1})$ determines $s_t(\theta^t)$ conditional on θ_t.

Proof of part (ii). Suppose that $\theta_{t-1} = \theta_{k-1} = \theta^H$, and that $\eta_t(\theta^{t-1}) = 1$ while $\eta_k(\theta^{k-1}) = 2$. From part (ii) of Lemma 2, it must be that $V_t(\theta^{t-1}) < V_k(\theta^{k-1})$. From part (iii) of Lemma 2, this implies that $s_t(\theta^t) < s_k(\theta^k)$ if $\theta_t = \theta_k$.■

C.5 Proof of Proposition 4

From part (ii) of Lemma 2, $V^{H*}(\theta^H, V)$ is strictly decreasing in V. We can establish that there exists a unique $\widehat{V}$ with the property that $V^{H*}(\theta^H, \widehat{V}) = \widehat{V}$. Note that $\overline{V}^H > V^{H*}(\theta^H, \overline{V}^H)$, and it can be shown that $\underline{V}^H < V^{H*}(\theta^H, \underline{V}^H)$. To see why, consider the solution to the program given $V = \underline{V}^H$, taking into account that this corresponds to the solution to (28) as λ approaches $-(1 - p(\theta^i|\theta^i))/p(\theta^i|\theta^i)$. The value of the objective for $\lambda = -(1 - p(\theta^i|\theta^i))/p(\theta^i|\theta^i)$ is weakly exceeded by

$$\theta^H U(1 - s^{fb}(\theta^H)) + \widetilde{\theta}^H U(s^{fb}(\theta^H)) + \delta p(\theta^H|\theta^H) W(\theta^H, \overline{V}^H),$$

which is the unconstrained value of the objective function uniquely attained under $s^{H*}(\theta^H, \underline{V}^H) = s^{fb}(\theta^H)$ and $V^{H*}(\theta^H, \underline{V}^H) = \overline{V}^H$. Note that a solution which satisfies the constraints of the problem and achieves this unconstrained maximum exists. For instance, a potential candidate solution assigns $s^{H*}(\theta^H,\underline{V}^H) = s^{L*}(\theta^H,\underline{V}^H) = s^{fb}(\theta^H)$ and $V^{H*}(\theta^H,\underline{V}^H) = V^{L*}(\theta^H,\underline{V}^H) = \overline{V}^H$ and satisfies all the constraints of the problem. Therefore, $V^{H*}(\theta^H,\underline{V}^H) = \overline{V}^H > \underline{V}^H$. Given that $V^{H*}(\theta^H,V)$ monotonically declines in V, there thus exists a unique $\widehat{V}$ with the property that $V^{H*}(\theta^L,\widehat{V}) = \widehat{V}$.

Given that $V^{H*}(\theta^H,V)$ is monotonically declining in V, it follows that if $V < (>)\, \widehat{V}$, then $V^{H*}(\theta^H,V) > (<)\, \widehat{V}$. Therefore, for $\theta_{t-1} = \theta_t = \theta^H$, if $V_t(\theta^{t-1}) < (>)\, \widehat{V}$, then $V_{t+1}(\theta^t) = V^{H*}(\theta^H,V_t(\theta^{t-1})) > (<)\, \widehat{V}$.■

D Proofs for Subsection 5.3

D.1 Proof of Lemma 3

Proof of part (i). From (1) and (5), given (30), we have

$$(\tau/(1-\delta) - b_{t+1}) = (\tau/(1-\delta) - b_t)(s_t/\delta). \tag{D.1}$$

Taking the log of both sides of (D.1), taking the sum over t between 0 and T, and dividing by T yields:

$$\frac{\log\left(\tau/(1-\delta) - b_{T+1}\right)}{T} = \frac{\log(\tau/(1-\delta) - b_0)}{T} + \frac{1}{T}\sum_{t=0}^{T}\log(s_t/\delta),$$

so that

$$\lim_{T\to\infty}\frac{\log\left(\tau/(1-\delta) - b_{T+1}\right)}{T} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T}\log(s_t/\delta). \tag{D.2}$$

Now consider the stochastic sequence $\mathbf{s}^{fb}$ which satisfies (8). Since this sequence is ergodic, using the Birkhoff theorem,[42] (D.2) reduces to:

$$\lim_{T\to\infty}\frac{\log(\tau/(1-\delta) - b^{fb}_{T+1})}{T} = \left[\frac{1}{2}\log\left(\frac{1}{\delta}\frac{\widetilde{\theta}^H}{\theta^H+\widetilde{\theta}^H}\right) + \frac{1}{2}\log\left(\frac{1}{\delta}\frac{\widetilde{\theta}^L}{\theta^L+\widetilde{\theta}^L}\right)\right], \tag{D.3}$$

where the term on the right hand side corresponds to the mean of $\log(s_t/\delta)$ in the long-run invariant distribution of first-best savings rates, taking into account that the symmetry of $p(\theta^i|\theta^i)$ implies that θ^L and θ^H must occur with equal probability in the long run. Note that

$$\begin{aligned}\frac{1}{2}\log\left(\frac{1}{\delta}\frac{\widetilde{\theta}^H}{\theta^H+\widetilde{\theta}^H}\right) + \frac{1}{2}\log\left(\frac{1}{\delta}\frac{\widetilde{\theta}^L}{\theta^L+\widetilde{\theta}^L}\right) &= -\left[\frac{1}{2}\log\left(\delta\frac{\theta^H+\widetilde{\theta}^H}{\widetilde{\theta}^H}\right) + \frac{1}{2}\log\left(\delta\frac{\theta^L+\widetilde{\theta}^L}{\widetilde{\theta}^L}\right)\right]\\ &> -\log\left(\frac{1}{2}\delta\frac{\theta^H+\widetilde{\theta}^H}{\widetilde{\theta}^H} + \frac{1}{2}\delta\frac{\theta^L+\widetilde{\theta}^L}{\widetilde{\theta}^L}\right)\\ &\geq -\log\left(\delta\frac{\theta^H+\widetilde{\theta}^H+\theta^L+\widetilde{\theta}^L}{\widetilde{\theta}^H+\widetilde{\theta}^L}\right) = 0,\end{aligned} \tag{D.4}$$

[42] See Durrett (2004), p. 337.

where the second inequality follows from Jensen's inequality and the third inequality follows from algebraic manipulation taking into account that $\theta^H > \theta^L$. Combining (D.3) with (D.4), we obtain $\lim_{T\to\infty} \log(\tau/(1-\delta) - b^{fb}_{T+1}) = \infty$, which implies $\lim_{T\to\infty} b^{fb}_{T+1} = -\infty$.

Proof of part (ii). By analogous reasoning as in the proof of part (i), taking into account that $\mathbf{s}^f$ satisfies (31), we have

$$\lim_{T\to\infty} \frac{\log(\tau/(1-\delta) - b^f_{T+1})}{T} = \left[\frac{1}{2}\log\left(\frac{1}{\delta}\frac{\beta\widetilde{\theta}^H}{\theta^H + \beta\widetilde{\theta}^H}\right) + \frac{1}{2}\log\left(\frac{1}{\delta}\frac{\beta\widetilde{\theta}^L}{\theta^L + \beta\widetilde{\theta}^L}\right)\right]. \tag{D.5}$$

Note that

$$\frac{1}{\delta}\frac{\beta\widetilde{\theta}^H}{\theta^H + \beta\theta^H} \leq \frac{1}{\delta}\frac{\beta\widetilde{\theta}^L}{\theta^L + \beta\widetilde{\theta}^L} < \frac{1}{\delta}\frac{\widetilde{\theta}^H}{\theta^H + \widetilde{\theta}^H} < 1, \tag{D.6}$$

where we have used (B.8) and the fact that $\theta^H > \theta^L$. Combining (D.5) with (D.6), we obtain $\lim_{T\to\infty} \log(\tau/(1-\delta) - b^f_{T+1}) = -\infty$, which implies $\lim_{T\to\infty} b^f_{T+1} = \tau/(1-\delta)$.■

D.2 Proof of Proposition 5

Proof of part (i). If shocks are i.i.d., then (21) implies that $s^{so}_t = \delta$ for all t, where s^{so}_t is the savings rate in the sequential optimum. From (D.1), this implies that $b^{so}_{t+1} = b^{so}_t$ for all t.

Proof of part (ii). By analogous reasoning as in the proof of part (i) of Lemma 3, and taking into account that $\mathbf{s}^{so}$ satisfies (21), we have

$$\lim_{T\to\infty} \frac{\log(\tau/(1-\delta) - b^{so}_{T+1})}{T} = \left[\begin{array}{l} \frac{1}{2}\log\left(\frac{1}{\delta}\frac{p(\theta^i|\theta^i)\widetilde{\theta}^H + (1-p(\theta^i|\theta^i))\widetilde{\theta}^L}{p(\theta^i|\theta^i)(\theta^H + \widetilde{\theta}^H) + (1-p(\theta^i|\theta^i))(\theta^L + \widetilde{\theta}^L)}\right) \\ +\frac{1}{2}\log\left(\frac{1}{\delta}\frac{(1-p(\theta^i|\theta^i))\widetilde{\theta}^H + p(\theta^i|\theta^i)\widetilde{\theta}^L}{(1-p(\theta^i|\theta^i))(\theta^H + \widetilde{\theta}^H) + p(\theta^i|\theta^i)(\theta^L + \widetilde{\theta}^L)}\right) \end{array}\right]. \tag{D.7}$$

Note that by Jensen's inequality, the right hand side of (D.7) is strictly larger than

$$-\log\left(\begin{array}{l} \frac{1}{2}\left(\frac{1}{\delta}\frac{p(\theta^i|\theta^i)(\theta^H + \widetilde{\theta}^H) + (1-p(\theta^i|\theta^i))(\theta^L + \widetilde{\theta}^L)}{p(\theta^i|\theta^i)\widetilde{\theta}^H + (1-p(\theta^i|\theta^i))\widetilde{\theta}^L}\right) \\ +\frac{1}{2}\left(\frac{1}{\delta}\frac{(1-p(\theta^i|\theta^i))(\theta^H + \widetilde{\theta}^H) + p(\theta^i|\theta^i)(\theta^L + \widetilde{\theta}^L)}{(1-p(\theta^i|\theta^i))\widetilde{\theta}^H + p(\theta^i|\theta^i)\widetilde{\theta}^L}\right) \end{array}\right),$$

which by algebraic manipulation, taking into account that $\theta^H > \theta^L$, can be shown to be weakly larger than

$$-\log\left(\delta\frac{\theta^H + \widetilde{\theta}^H + \theta^L + \widetilde{\theta}^L}{\widetilde{\theta}^H + \widetilde{\theta}^L}\right) = 0.$$

Combining this with (D.7), we obtain $\lim_{T\to\infty} \log(\tau/(1-\delta) - b^{so}_{T+1}) = \infty$, which implies $\lim_{T\to\infty} b^{so}_{T+1} = -\infty$.■

D.3 Proof of Proposition 6

Proof of part (i). This follows directly from Proposition 2 and Proposition 5.

Proof of part (ii). By analogous reasoning as in the proof of part (i) of Lemma 3, equation (D.2) applies. The sequence $\mathbf{s}^{eo}$ described in Proposition 3 is ergodic so that $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T}\log(s_t/\delta)$ is equal to the mean of $\log(s_t/\delta)$ in the ergodic distribution of savings rates. Note that given Lemma 2 and Proposition 3, in such an ergodic distribution,

$$\begin{aligned}
\Pr[\eta_t = 0] &= \Pr\left[\eta_{t-1} = 0\right] p(\theta^i|\theta^i) + \sum_{k=1}^{\infty} \Pr\left[\eta_{t-1} = k\right](1 - p(\theta^i|\theta^i)),\\
\Pr[\eta_t = 1] &= \Pr\left[\eta_{t-1} = 0\right](1 - p(\theta^i|\theta^i)), \text{ and}\\
\Pr[\eta_t = k] &= \Pr\left[\eta_{t-1} = k-1\right] p(\theta^i|\theta^i) \text{ for } k > 1.
\end{aligned}$$

Since $\Pr[\eta_t = k] = \Pr\left[\eta_{t-1} = k\right]$ for all $k \geq 0$ and since $\sum_{k=0}^{\infty}\Pr[\eta_t = k] = 1$, it thus follows by substituting above that $\Pr[\eta_t = 0] = 1/2$ and $\Pr\left[\eta_{t-1} = k\right] = (1-p(\theta^i|\theta^i))p(\theta^i|\theta^i)^{k-1}/2$ for $k > 0$. Moreover, conditional on $\eta_t = 0$, the probability of $\theta_t = \theta^L$ is $p(\theta^i|\theta^i)$ and conditional on $\eta_t > 0$, the probability of $\theta_t = \theta^H$ is $p(\theta^i|\theta^i)$. This means that $\bar{s}$ given in the text corresponds to the mean of $\log(s_t/\delta)$ in the invariant distribution of savings rates. Therefore,

$$\lim_{T\to\infty} \frac{\log(\tau/(1-\delta) - b^{eo}_{T+1})}{T} = \bar{s},$$

and it follows that if $\bar{s} > 0$, then $\lim_{T\to\infty}\log(\tau/(1-\delta) - b^{eo}_{T+1}) = \infty$, which means that $\lim_{T\to\infty} b^{eo}_{T+1} = -\infty$. If instead $\bar{s} < 0$, then $\lim_{T\to\infty}\log(\tau/(1-\delta) - b^{eo}_{T+1}) = -\infty$, which means that $\lim_{T\to\infty} b^{eo}_{T+1} = \tau/(1-\delta)$.

Finally, to show that $\bar{s} > 0$ and $\bar{s} < 0$ both hold for an open set of parameters, consider the values for $\left\{\theta^L, \theta^H, p(\theta^i|\theta^i), \delta, \beta\right\}$ used in Figure 3 (see fn. 38). It is easy to verify using computational methods that if $\beta = 0.01$, then $\bar{s} > 0$; by continuity this is also true for a neighborhood of parameter vectors. Similarly, one can verify that if $\beta = 0.4$, then $\bar{s} < 0$, and again by continuity this is true for a neighborhood of parameters. Note that all these values satisfy Assumption 2. ■

E Proofs for Section 6

E.1 Proofs of Proposition 7 and Corollary 2

In order to prove these results, consider the program isomorphic to (16)–(17) under a continuum of shocks. Define a function $f(\theta_t) = \theta_t/\widetilde{\theta}_t$ for $\widetilde{\theta}_t$ which depends on θ_t as defined in (4). Let $\omega_t = f(\theta_t) \in \Omega \equiv [\underline{\omega}, \overline{\omega}]$, where it is clear that from Assumption 3, there is a one to one mapping from θ_t to ω_t. Let $h(\omega_t|\omega_{t-1})$ correspond to the value of $\widetilde{\theta}_t p(\theta_t|\theta_{t-1})\left(f'(\theta_t)\right)^{-1}/\mathbb{E}[\widetilde{\theta}_t p(\theta_t|\theta_{t-1})|\theta_{t-1}]$, so that $h(\omega_t|\omega_{t-1}) > 0$,

$$h(\omega_t|\omega_{t-1})d\omega_t = \widetilde{\theta}_t p(\theta_t|\theta_{t-1})/\mathbb{E}[\widetilde{\theta}_t p(\theta_t|\theta_{t-1})|\theta_{t-1}]d\theta_t,$$

and $\int_{\underline{\omega}}^{\overline{\omega}} h(\omega_t|\omega_{t-1})d\omega_t = 1$, where we have used the fact that $d\omega_t = f'(\theta_t)d\theta_t$. Therefore, $h(\omega_t|\omega_{t-1})$ is effectively a density function. Define $H(\omega_t|\omega_{t-1})$ as the associated c.d.f.

Using this formulation, (16)–(17) can be rewritten as

$$\max_{\{s_t(\omega_t)\}_{\omega_t\in\Omega}} \int_{\underline{\omega}}^{\overline{\omega}} h(\omega_t|\omega_{t-1})\big(\omega_t U(1-s_t(\omega_t)) + U(s_t(\omega_t))\big)d\omega_t \tag{E.1}$$

$$\text{s.t.}$$

$$\omega_t U(1-s_t(\omega_t)) + \beta U(s_t(\omega_t)) \geq \omega_t U(1-s_t(\widehat{\omega}_t)) + \beta U(s_t(\widehat{\omega}_t))\ \forall\omega_t \text{ and } \forall\widehat{\omega}_t \neq \omega_t. \tag{E.2}$$

(E.1)–(E.2) is identical to (16)–(17), where we have used the one to one mapping from θ_t to ω_t to write the program as one of choosing a savings rate conditional on the report $\widehat{\omega}_t$.

Now consider a relaxed version of (E.1)–(E.2) which allows (1) to be an inequality:

$$\max_{\{u_t(\omega_t),y_t(\omega_t)\}_{\omega_t\in\Omega}} \int_{\underline{\omega}}^{\overline{\omega}} h(\omega_t|\omega_{t-1})\big(\omega_t u_t(\omega_t) + y_t(\omega_t)\big)d\omega_t \tag{E.3}$$

$$\text{s.t.}$$

$$U^{-1}(u_t(\omega_t)) + U^{-1}(y_t(\omega_t)) \leq 1\ \forall\omega_t \tag{E.4}$$

$$\omega_t u_t(\omega_t) + \beta y_t(\omega_t) \geq \omega_t u_t(\widehat{\omega}_t) + \beta y_t(\widehat{\omega}_t)\ \forall\omega_t \text{ and } \forall\widehat{\omega}_t \neq \omega_t. \tag{E.5}$$

(E.3)–(E.5) is identical to (E.1)–(E.2) if the solution admits (E.4) holding with equality $\forall\omega_t$. (E.3)–(E.5) corresponds to the problem analyzed in Section 3.2 of Amador, Werning, and Angeletos (2006) so that their analysis applies here as well.

The envelope condition which characterizes (E.5) implies that

$$\frac{\omega_t}{\beta}u(\omega_t) + y(\omega_t) = \int_{\underline{\omega}}^{\omega_t} \frac{1}{\beta}u(\omega')d\omega' + \frac{\underline{\omega}}{\beta}u(\underline{\omega}) + y(\underline{\omega}). \tag{E.6}$$

Standard arguments also require $u(\omega_t)$ to be a non-decreasing function of ω_t. Thus, (E.6) and monotonicity are necessary for incentive compatibility. Substituting (E.6) into the objective function and the resource constraint and integrating by parts allows us to rewrite the problem as:

$$\max_{\{u_t(\omega_t),y(\underline{\omega})\}} \left\{\frac{\underline{\omega}}{\beta}u(\underline{\omega}) + y(\underline{\omega}) + \frac{1}{\beta}\int_{\underline{\omega}}^{\overline{\omega}} (1-G(\omega_t|\omega_{t-1}))u(\omega_t)d\omega_t\right\} \tag{E.7}$$

$$\text{s.t.}$$

$$U(1-U^{-1}(u_t(\omega_t))) + \frac{\omega_t}{\beta}u(\omega_t) - \frac{\underline{\omega}}{\beta}u(\underline{\omega}) - y(\underline{\omega}) - \frac{1}{\beta}\int_{\underline{\omega}}^{\omega_t} u(\omega')d\omega' \geq 0 \tag{E.8}$$

$$\text{and } u_t(\omega_t) \text{ non-decreasing} \tag{E.9}$$

for

$$G(\omega_t|\omega_{t-1}) = H(\omega_t|\omega_{t-1}) + \omega_t(1-\beta)h(\omega_t|\omega_{t-1}).$$

The above program can be solved using Lagrangian methods. Following Amador, Werning, and Angeletos (2006), define $\omega_p(\omega_{t-1}) = \max\{\underline{\omega}, \omega'\}$ where ω' is the lowest $\omega \in \Omega$ such that

$\forall \omega'' \geq \omega$,

$$\int_{\omega''}^{\overline{\omega}} (1 - G(\omega'''|\omega_{t-1}))d\omega''' \leq 0. \tag{E.10}$$

Consider the following condition.

Condition 2. *$\forall \omega_{t-1}$ and $\forall \omega_t \leq \omega_p(\omega_{t-1})$, $G(\omega_t|\omega_{t-1})$ is non-decreasing in ω_t.*

Proposition 2 in Amador, Werning, and Angeletos (2006) states that the solution to (E.7)–(E.9) admits $s_t(\omega_t) = s_t(\omega_p(\omega_{t-1}))$ if $\omega_t \geq \omega_p(\omega_{t-1})$. Moreover, Proposition 3 in that paper states that if Condition 2 holds, then the solution to (E.7)–(E.9) admits (E.8) holding with equality, so that (E.4) also holds with equality. Furthermore, $s_t(\omega_t) = s^f(\omega_t)$ if $\omega_t < \omega_p(\omega_{t-1})$, where with some abuse of notation, $s^f(\omega_t)$ is defined as the flexible optimum given by

$$\omega_t U'(1 - s^f(\omega_t)) = \beta U'(s^f(\omega_t)).$$

Let $\underline{s}(\omega_{t-1})$ be defined by $\underline{s}(\omega_{t-1}) = s^f(\omega_p(\omega_{t-1}))$ if $\omega_p(\omega_{t-1}) > \underline{\omega}$, and

$$\int_{\underline{\omega}}^{\overline{\omega}} h(\omega_t|\omega_{t-1})\big(\omega_t U'(1 - \underline{s}(\omega_{t-1})) - U'(\underline{s}(\omega_{t-1}))\big)d\omega_t = 0$$

otherwise. $\forall \omega^t$, it follows then that the sequential optimum features

$$s_t(\omega^t) = \max\{s^f(\omega_t), \underline{s}(\omega_{t-1})\}.$$

This therefore means that the sequentially optimal rule at any date t can be implemented with a debt limit, which depends only on ω_{t-1} and the current level of debt.

In order to complete the argument, we must verify that $\theta_p(\theta_{t-1}) = f^{-1}(\omega_p(f(\theta_{t-1})))$ and that Assumption 4 is identical to Condition 2. To show that $\theta_p(\theta_{t-1}) = f^{-1}(\omega_p(f(\theta_{t-1})))$, note that (E.10) can be rewritten as

$$[1 - H(\omega''|\omega_{t-1})]\omega'' \left(\frac{1}{\beta} - \frac{\left(\int_{\omega''}^{\overline{\omega}} \omega''' h(\omega'''|\omega_{t-1})d\omega'''\right) / \left(\int_{\omega''}^{\overline{\omega}} h(\omega'''|\omega_{t-1})d\omega'''\right)}{\omega''} \right) \geq 0.$$

Letting $\theta'' = f^{-1}(\omega'')$ and $\theta''' = f^{-1}(\omega''')$ with associated values $\widetilde{\theta}''$ and $\widetilde{\theta}'''$, the above condition becomes

$$\left(\frac{1}{\beta} - \frac{\left(\int_{\theta''}^{\overline{\theta}} \theta_t''' p(\theta_t'''|\theta_{t-1})d\theta'''\right) / \left(\int_{\theta''}^{\overline{\theta}} \widetilde{\theta}_t''' p(\theta_t'''|\theta_{t-1})d\theta'''\right)}{\theta''/\widetilde{\theta}''} \right) \geq 0,$$

which becomes (32). This establishes that $\theta_p(\theta_{t-1}) = f^{-1}(\omega_p(f(\theta_{t-1})))$.

To show that Assumption 4 is identical to Condition 2, note that $G(\omega_t|\omega_{t-1})$ is continuously differentiable in ω_t and first order conditions imply that Condition 2 reduces to

$$\frac{d \log h(\omega_t|\omega_{t-1})}{d \log \omega_t} \geq -\frac{2 - \beta}{1 - \beta}. \tag{E.11}$$

We can show that (33) implies (E.11). Given the definition of $h(\cdot)$, note that the left hand side of (E.11) can be expanded so that (E.11) becomes

$$\frac{\dfrac{d\log\widetilde{\theta}_t}{d\log\theta_t}+\dfrac{d\log p(\theta_t|\theta_{t-1})}{d\log\theta_t}-\dfrac{d\log\left(d\widetilde{\theta}_t/d\theta_t\right)}{d\log\theta_t}}{1-\dfrac{d\log\widetilde{\theta}_t}{d\log\theta_t}}\geq-\frac{2-\beta}{1-\beta},$$

which is equivalent to (33).■

E.2 Proof of Proposition 8

Consider the case of i.i.d. shocks, and with some abuse of notation, let $W_t(\cdot)$ correspond to the continuous type analogue of the continuation value defined in (9) divided by $\delta\mathbb{E}[\theta_t]$, where we have taken into account that shocks are i.i.d. Let the range $[\underline{W},\overline{W}]$ correspond to the feasible range of such continuation values. To write the period zero problem, we pursue an analogous strategy as in the proof of Proposition 7 by considering the relaxed problem which allows the resource constraint to hold as an inequality:

$$\max_{\{u_t(\omega_t),y_t(\omega_t)\}_{\omega_t\in\Omega}}\int_{\underline{\omega}}^{\overline{\omega}}h(\omega_t|\omega_{t-1})\big(\omega_t u_t(\omega_t)+y_t(\omega_t)+W_t(\omega_t)\big)d\omega_t \tag{E.12}$$

s.t.

$$U^{-1}(u_t(\omega_t))+U^{-1}(y_t(\omega_t))\leq 1\ \forall\omega_t, \tag{E.13}$$

$$\omega_t u_t(\omega_t)+\beta y_t(\omega_t)+\beta W_t(\omega_t)\geq\omega_t u_t(\widehat{\omega}_t)+\beta y_t(\widehat{\omega}_t)+\beta W_t(\widehat{\omega}_t)\ \forall\omega_t\text{ and }\forall\widehat{\omega}_t\neq\omega_t, \tag{E.14}$$

$$\text{and }\underline{W}\leq W_t(\omega_t)\leq\overline{W}\ \forall\omega_t. \tag{E.15}$$

The same envelope condition in (E.6) applies. Together with the monotonicity of $u_t(\omega_t)$, it implies incentive compatibility. Substituting (E.6) into (E.12), (E.13), and (E.15), the program can be rewritten as

$$\max_{\left\{u_t(\omega_t),y(\underline{\omega}),\underline{W}\leq W(\underline{\omega})\leq\overline{W}\right\}}\left\{\frac{\underline{\omega}}{\beta}u(\underline{\omega})+y(\underline{\omega})+W(\underline{\omega})+\frac{1}{\beta}\int_{\underline{\omega}}^{\overline{\omega}}(1-G(\omega_t|\omega_{t-1}))u(\omega_t)d\omega_t\right\} \tag{E.16}$$

s.t.

$$U(1-U^{-1}(u_t(\omega_t)))+\overline{W}+\frac{\omega_t}{\beta}u(\omega_t)-\frac{\underline{\omega}}{\beta}u(\underline{\omega})-y(\underline{\omega})-W(\underline{\omega})-\frac{1}{\beta}\int_{\underline{\omega}}^{\omega}u(\theta')d\theta'\geq 0 \tag{E.17}$$

$$\text{and }u_t(\omega_t)\text{ non-decreasing.} \tag{E.18}$$

(E.16)–(E.18) is identical to (E.7)–(E.9) since the term $y(\underline{\omega})$ in (E.7)–(E.9) is replaced with $y(\underline{\omega})+W(\underline{\omega})$ in (E.16)–(E.18) and the term $U(1-U^{-1}(u_t(\omega_t)))$ is replaced with $U(1-U^{-1}(u_t(\omega_t)))+\overline{W}$. Therefore, the same arguments as those of Amador, Werning, and Angeletos (2006) imply that the solution to (E.16)–(E.18) admits (E.17) holding with equality, which then means that (E.13) holds with equality and $W(\omega_t)=\overline{W}$. It thus follows that the solution admits a static mechanism, and therefore the optimal mechanism is characterized as in Proposition 7.■

E.3 Proof of Proposition 9

Proof of part (i). Suppose by contradiction that the ex-ante optimum coincides with the sequential optimum, so the solution is characterized by Proposition 7. With some abuse of notation, let $V^{\theta_0}(\widehat{\theta}_0)$ correspond to the expected date 1 welfare in the sequential optimum to a type θ_0 who lies and claims to be a type $\widehat{\theta}_0$. By lying, this type receives a mechanism associated with $\widehat{\theta}_0$, evaluated using probabilities $p(\theta_1|\theta_0)$. Given the description of the solution in Proposition 7, it is straightforward to show that $V^{\theta_0}(\widehat{\theta}_0)$ is continuously differentiable in $\widehat{\theta}_0$ with $V^{\theta_0\prime}(\widehat{\theta}_0) = 0$ for $\widehat{\theta}_0 = \theta_0$.

We now consider a perturbation that affects types $\theta_0 < \theta_p(\theta_{-1})$. To facilitate the construction of the perturbation, note first that one implementation of the sequentially optimal mechanism is as follows. The government can choose any savings rate above $s^f(\theta_p(\theta_{-1}))$. If the chosen savings rate is $s_0 > s^f(\theta_p(\theta_{-1}))$, then the mechanism at date 1 corresponds to the sequentially optimal mechanism for type $s^{f^{-1}}(s_0)$, where $s^{f^{-1}}(\cdot)$ is a function that uses (B.1) to derive the type from the chosen flexible savings rate. If instead the chosen savings rate is $s_0 = s^f(\theta_p(\theta_{-1}))$, then the government reports its type $\widehat{\theta}_0$, and the mechanism at date 1 corresponds to the sequentially optimal mechanism for type $\widehat{\theta}_0$.

Given this implementation, consider the following perturbation. If the chosen savings rate is $s_0 = s^f(\theta_p(\theta_{-1}))$, the mechanism is unchanged. If instead the chosen savings rate is $s_0 > s^f(\theta_p(\theta_{-1}))$, the mechanism at date 1 corresponds to the sequentially optimal mechanism for type $s^{f^{-1}}(s_0)+\varepsilon\mu(s^{f^{-1}}(s_0))$, for $\varepsilon > 0$ arbitrarily small and for some continuously differentiable function $\mu(\cdot)$ satisfying $\mu(\cdot) > 0$, $\mu'(\cdot) > -1$, and $\lim_{\theta_0\to\theta_p(\theta_{-1})}\mu(\theta_0) = 0$. Note that if $\varepsilon = 0$, the original sequentially optimal mechanism is in place. If instead $\varepsilon > 0$, arbitrarily small, then all types $\theta_0 \geq \theta_p(\theta_{-1})$ do not change their behavior, but the chosen savings rate $s_0^\varepsilon(\theta_{-1},\theta_0)$ for types $\theta_0 < \theta_p(\theta_{-1})$ must satisfy the following first order condition:

$$
\begin{aligned}
&-\theta_0 U'(1 - s_0^\varepsilon(\theta_{-1},\theta_0)) + \beta\widetilde{\theta}_0 U'(s_0^\varepsilon(\theta_{-1},\theta_0)) \qquad \text{(E.19)}\\
&= -\beta\delta V^{\theta_0\prime}(s^{f^{-1}}(s_0^\varepsilon(\theta_{-1},\theta_0)) + \varepsilon\mu(s^{f^{-1}}(s_0^\varepsilon(\theta_{-1},\theta_0))))\frac{d[s^{f^{-1}}(s_0^\varepsilon(\theta_{-1},\theta_0)) + \varepsilon\mu(s^{f^{-1}}(s_0^\varepsilon(\theta_{-1},\theta_0)))]}{ds_0^\varepsilon(\theta_{-1},\theta_0)}.
\end{aligned}
$$

As ε approaches 0, $s_0^\varepsilon(\theta_{-1},\theta_0)$ approaches $s_0^f(\theta_0)$. We can show that it must approach it from above; that is, for sufficiently small $\varepsilon > 0$, it must be that $s_0^\varepsilon(\theta_{-1},\theta_0) \geq s_0^f(\theta_0)$. If instead $s_0^\varepsilon(\theta_{-1},\theta_0) < s_0^f(\theta_0)$, then type θ_0 could make itself strictly better off by increasing its savings rate to $s_0^f(\theta_0)$, as this maximizes its immediate welfare and raises its continuation utility.

Suppose that Condition 1 holds for θ_{-1} and some θ_0. Together with (E.19), this implies that $s_0^\varepsilon(\theta_{-1},\theta_0) > s_0^f(\theta_0)$ for a positive measure of types θ_0. Using (E.19), the change in welfare as ε approaches 0 for any such θ_0 has the same sign as

$$
[-\theta_0 U'(1 - s_0^f(\theta_0)) + \widetilde{\theta}_0 U'(s_0^f(\theta_0))]\frac{ds_0^\varepsilon(\theta_{-1},\theta_0)}{d\varepsilon}\Big|_{\varepsilon=0} + \delta V^{\theta_0\prime}(\theta_0)\frac{1}{s^{f\prime}(\theta_0)}\frac{ds_0^\varepsilon(\theta_{-1},\theta_0)}{d\varepsilon}\Big|_{\varepsilon=0} > 0,
$$

where we have used (B.1) and the fact that $V^{\theta_0\prime}(\theta_0) = 0$. Therefore, the perturbation strictly increases welfare. Note that if Condition 1 does not hold for θ_{-1}, then it necessarily holds at some θ_{t-1}, and the same perturbation can be performed at that date while continuing to satisfy all incentive compatibility constraints at $t - k$ for $k > 1$.

Proof of part (ii). Suppose by contradiction that the mechanism does not exhibit history dependence. Let $V^{\theta}(\widehat{\theta})$ correspond to the continuation value to a type θ who reports $\widehat{\theta}$ under this history-independent mechanism, where by assumption, $V^{\theta}(\widehat{\theta})$ is piecewise continuously differentiable. Given that the continuation mechanism is independent of the date, and given that the ex-ante optimal mechanism is chosen at date 0, it follows that $\forall\theta_{-1}$, the continuation value at date 0 is $V^{\theta}(\theta)$ if $\theta = \theta_{-1}$, and by optimality, $V^{\theta\prime}(\theta) = 0$. Thus, the first order conditions which guarantee truthtelling whenever the mechanism is differentiable imply that

$$[-\theta_t U'(1 - s_t(\theta_{t-1}, \theta_t)) + \beta\widetilde{\theta}_t U'(s_t(\theta_{t-1}, \theta_t))]\frac{ds_t(\theta_{t-1}, \theta_t)}{d\theta_t} = 0. \tag{E.20}$$

This requires that either $s_t(\theta_{t-1}, \theta_t) = s^f(\theta_t)$ or $ds_t(\theta_{t-1}, \theta_t)/d\theta_t = 0$. Therefore, (E.20) effectively corresponds to the first order condition to static incentive compatibility constraints. As such, the optimal mechanism is not dynamic. The solution to the program subject to a sequence of static incentive compatibility constraints coincides with the solution to the sequential optimum described in Proposition 7. However, part (i) shows that this mechanism is suboptimal.■

References

Acemoglu, D., M. Golosov, and A. Tsyvinski (2008): "Political Economy of Mechanisms," *Econometrica*, 76, 619–41.

Aguiar, M., and M. Amador (2011): "Growth under the Shadow of Expropriation," *Quarterly Journal of Economics*, 126, 651–97.

Ales, L., P. Maziero, and P. Yared (2012): "A Theory of Political and Economic Cycles," Working Paper.

Alesina, A., and R. Perotti (1994): "The Political Economy of Budget Deficits," NBER Working Paper 4637.

Alesina, A., and G. Tabellini (1990): "A Positive Theory of Fiscal Deficits and Government Debt," *Review of Economic Studies*, 57, 403–14.

Alonso, R., and N. Matouschek (2008): "Optimal Delegation," *Review of Economic Studies*, 75, 259–93.

Amador, M., and K. Bagwell (2011): "The Theory of Delegation with an Application to Tariff Caps," Working Paper.

Amador, M., I. Werning, and G.-M. Angeletos (2003): "Commitment Vs. Flexibility," NBER Working Paper.

——— (2006): "Commitment Vs. Flexibility," *Econometrica*, 74, 365–96.

Ambrus, A., and G. Egorov (2009): "Delegation and Nonmonetary Incentives," Working Paper.

——— (2012): "Commitment-Flexibility Trade-off and Withdrawal Penalties," Working Paper.

Athey, S., A. Atkeson, and P. J. Kehoe (2005): "The Optimal Degree of Discretion in Monetary Policy," *Econometrica*, 73, 1431–75.

Atkeson, A., and R. Lucas (1992): "On Efficient Distribution with Private Information," *Review of Economic Studies*, 59(3), 427–53.

Azzimonti, M. (2011): "Barriers to Investment in Polarized Societies," *American Economic Review*, 101, 2182–204.

Azzimonti, M., M. Battaglini, and S. Coate (2010): "Analyzing the Case for a Balanced Budget Ammendment to the U.S. Constitution," Working Paper.

Barro, R. (1990): "On the Predictability of Tax-Rate Changes," in *Macroeconomic Policy*, pp. 268–97. Harvard University Press; Cambridge.

——— (1999): "Ramsey Meets Laibson in the Neoclassical Growth Model," *Quarterly Journal of Economics*, 114, 1125–52.

Battaglini, M., and S. Coate (2008): "A Dynamic Theory of Public Spending, Taxation, and Debt," *American Economic Review*, 98, 201–36.

Battaglini, M., and R. Lamba (2012): "Optimal Dynamic Contracting," Working Paper.

Benveniste, L. M., and J. A. Scheinkman (1979): "On the Differentiability of the Value Function in Dynamic Models of Economics," *Econometrica*, 447, 727–32.

Bernheim, B. D., D. Ray, and S. Yeltekin (1999): "Self-Control, Saving, and the Low Asset Trap," Working Paper.

Bond, P., and G. Sigurdsson (2011): "Commitment Contracts," Working Paper.

Caballero, R., and P. Yared (2010): "Future Rent-Seeking and Current Public Savings," *Journal of International Economics*, 82, 124–36.

Chamberlain, G., and C. A. Wilson (2000): "Optimal Intertemporal Consumption under Uncertainty," *Review of Economic Dynamics*, 3(3), 365–95.

Durrett, R. (2004): *Probability: Theory and Examples*. Cengage Learning, third edn.

Farhi, E., and I. Werning (2010): "Insurance and Taxation over the Life Cycle," Working Paper.

Farrell, J., and E. Maskin (1989): "Renegotiation in Repeated Games," *Games and Economic Behavior*, 4, 327–60.

Fernandes, A., and C. Phelan (2000): "A Recursive Formulation for Repeated Agency with History Dependence," *Journal of Economic Theory*, 19, 223–47.

Golosov, M., M. Troshkin, and A. Tsyvinski (2011): "Optimal Dynamic Taxes," Working Paper.

Golosov, M., and A. Tsyvinski (2006): "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy*, 114, 257–79.

Halac, M. (2012): "Relational Contracts and the Value of Relationships," *American Economic Review*, 102, 750–79.

Holmström, B. (1977): "On Incentives and Control in Organizations," Ph.D. Thesis, Stanford University.

——— (1984): "On the Theory of Delegation," in *Bayesian Models in Economic Theory*, ed. by M. Boyer, and R. Kihlstrom. North-Holland; New York.

Hosseini, R., L. E. Jones, and A. Shourideh (2012): "Optimal Contracting and Dynamic Altruism: Consumption and Family Size," Working Paper.

International Monetary Fund (2009): "Fiscal Rules—Anchoring Expectations for Sustainable Public Finances," IMF Working Paper.

Kapicka, M. (2010): "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First Order Approach," Working Paper.

Kocherlakota, N. (1996): "Reconsideration-Proofness: A Refinement for Infinite Horizon Time Inconsistency," *Games and Economic Behavior*, 15, 33–54.

Krusell, P., B. Kruscu, and A. Smith, Jr. (2010): "Temptation and Taxation," *Econometrica*, 78, 2063–84.

Krusell, P., and J.-V. Rios-Rull (1999): "On the Size of U.S. Government: Political Economy in the Neoclassical Growth Model," *American Economic Review*, 89, 1156–81.

Krusell, P., and A. Smith, Jr. (2003): "Consumption-Savings Decisions with Quasi-Geometric Discounting," *Econometrica*, 71, 365–75.

Kydland, F., and E. Prescott (1977): "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85, 473–91.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–77.

Lizzeri, A. (1999): "Budget Deficits and Redistributive Politics," *American Economic Review*, 66, 909–28.

Pavan, A., I. Segal, and J. Toikka (2010): "Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization and Information Disclosure," Working Paper.

Persson, T., and L. Svensson (1989): "Why a Stubborn Conservative Would Run a Deficit: Policy with Time-Inconsistent Preferences," *Quarterly Journal of Economics*, 104, 325–45.

Phelps, E., and R. Pollak (1968): "On Second Best National Savings and Game-Equilibrium Growth," *Review of Economic Studies*, 35, 185–99.

Sleet, C. (2004): "Optimal Taxation with Private Government Information," *Review of Economic Studies*, 71, 1217–39.

Song, Z., K. Storesletten, and F. Zilibotti (2012): "Rotten Parents and Disciplined Children: A Politico-Economic Theory of Public Expenditure and Debt," *Econometrica*, forthcoming.

Strulovici, B. (2011): "Renegotiation-Proof Contracts with Moral Hazard and Persistent Private Information," Working Paper.

Thomas, J., and T. Worrall (1990): "Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem," *Journal of Economic Theory*, 51(2), 367–90.

Tornell, A., and P. Lane (1999): "The Voracity Effect," *American Economic Review*, 89, 22–46.

Williams, N. (2011): "Persistent Private Information," *Econometrica*, 79, 1233–74.

Yared, P. (2010): "Politicians, Taxes and Debt," *Review of Economic Studies*, 77, 806–40.