A Theory of Fiscal Responsibility and Irresponsibility*

Marina Halac† Pierre Yared‡

April 14, 2024

Abstract

We propose a political economy mechanism that explains the presence of fiscal regimes punctuated by crisis periods. Our model focuses on the interaction between successive deficit-biased governments subject to i.i.d. fiscal shocks. We show that the economy transitions between a fiscally responsible regime and a fiscally irresponsible regime, with transitions occurring during crises when fiscal needs are large. Under fiscal responsibility, governments limit their spending to avoid transitioning to fiscal irresponsibility. Under fiscal irresponsibility, governments spend excessively and precipitate crises that lead to the reinstatement of fiscal responsibility. Regime transitions can only occur if governments' deficit bias is large enough.

Keywords: Private Information, Fiscal Policy, Deficit Bias
JEL Classification: C73, D02, D82, E6, H1, P16

*Part of the analysis in this paper previously appeared in a 2017 working paper titled "Fiscal Rules and Discretion under Limited Enforcement" (NBER WP 25463). We are grateful to Marco Bassetto, Dana Foarta, Juan Ortner, and Chris Phelan for discussions of that working paper, and to Mark Aguiar, Alessandro Dovis, and Hulya Eraslan for discussions of the present paper. We would also like to thank Manuel Amador, Dirk Bergemann, Hector Chade, V.V. Chari, Stephen Coate, Georgy Egorov, Johannes Hörner, Zhen Huo, Navin Kartik, George Mailath, Guido Menzio, Chris Moser, Jesse Schreger, conference and seminar audiences, and especially Alessandro Lizzeri for helpful comments. Simon Boutin and Ferdinand Pieroth provided excellent research assistance.

†Yale University, Department of Economics and CEPR. Email: marina.halac@yale.edu.
‡Columbia University, Graduate School of Business and NBER. Email: pyared@columbia.edu.

1 Introduction

In this paper, we propose a political economy mechanism that explains the presence of fiscal regimes punctuated by crisis periods. Our motivation stems from the empirical observation that countries experience transitions between periods of fiscal responsibility and periods of fiscal irresponsibility, with the transitions in both directions occurring at times of crisis. Figure 1 illustrates these dynamics. Using data put together by Alesina, Favero, and Giavazzi (2019), the figure displays fiscal consolidations for the U.S. and the European Union between 1980 and 2014.

Figure 1 shows not only that countries go through long waves of fiscal consolidations, but also that fiscal consolidations tend to begin and to end with recessions. The early 1990s recessions, which included the European ERM crisis of 1992 and 1993, led to multiple fiscal reforms in Europe and in the U.S. Most notable were the Omnibus Budget Reconciliations Acts of 1990 (under President George H. W. Bush) and 1993 (under President Bill Clinton). These reforms ended with the recessions of the late 1990s. After a hiatus in the 2000s, another round of fiscal consolidations followed the Global Financial Crisis of 2008, with the Budget Control Act of 2011 in the U.S. and tax and spending reforms in the United Kingdom, Ireland, Greece, Portugal, and Spain. Most of these reforms were abandoned following the 2020 COVID-19 recessions.

Similar dynamics to those displayed in Figure 1 have been documented in other regions. Sachs (1990) and Dornbusch and Edwards (1991) describe the experience of Latin American countries, which have historically fluctuated between periods of populism and periods of austerity. Much like in the U.S. and Europe, the fiscal regimes in these countries span multiple political cycles, with left-leaning and right-leaning governments both promoting populism and austerity at different times.[1]

We present a theory that sheds light on these empirical patterns. Our model focuses on the dynamic interaction between successive deficit-biased governments subject to i.i.d. fiscal shocks. We show that the economy endogenously transitions between a *fiscally responsible regime* and a *fiscally irresponsible regime*, with transitions occurring during crises when fiscal needs are large. In the fiscally responsible regime, governments limit their spending in order to avoid transitioning to fiscal irresponsibility. In the fiscally irresponsible regime, governments spend excessively even relative to their biases,

[1] Anecdotal evidence suggests that transitions between these populism and austerity periods also coincide with times of crisis, as in the advanced economies of Figure 1. That said, the experience of Latin American countries is different from that of advanced economies in that issues of inflation, devaluation, and default are very salient (see, e.g., Dovis, Golosov, and Shourideh, 2016). Dovis (2019) studies a model related to ours but which allows for these additional repayment frictions.

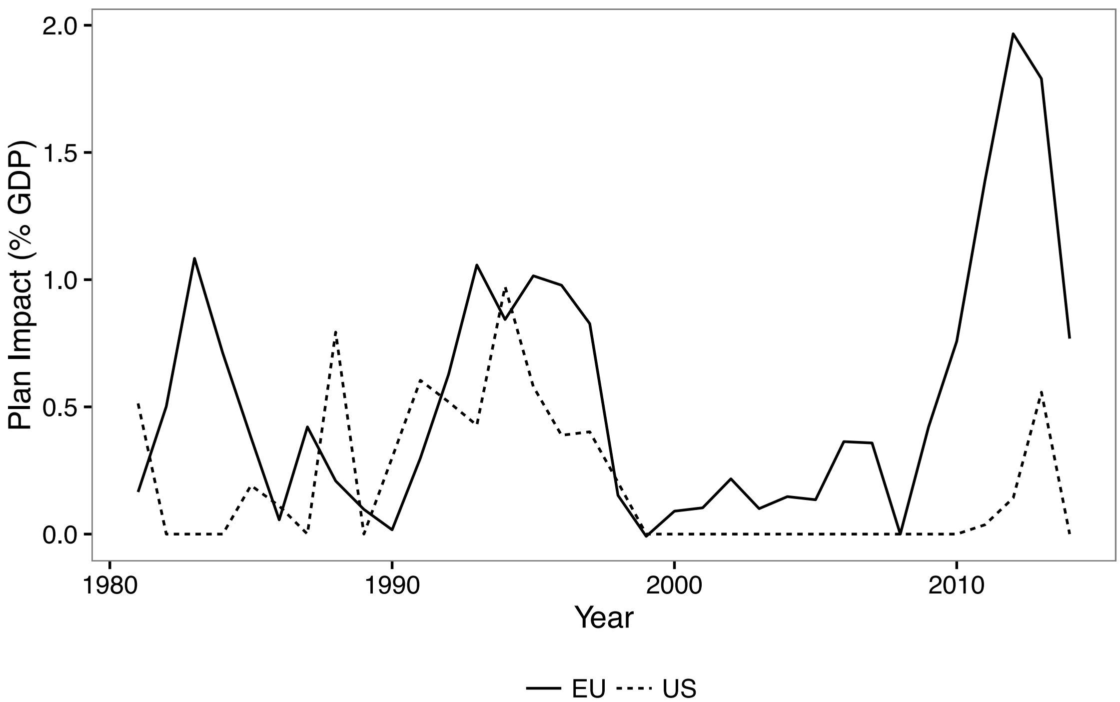

Figure 1: The displayed plan impact measure corresponds to exogenous innovations to tax and spending policy, with higher values indicating greater austerity. See the discussion in Alesina, Favero, and Giavazzi (2019) and their Figure 6.1.

precipitating crises that lead to the reinstatement of fiscal responsibility.

Our environment is an infinite horizon small open economy in which successive governments make borrowing decisions. Prior to the choice of policy at every date, an i.i.d. shock to the social value of deficit-financed government spending is realized. Governments are deficit-biased: for any given shock, the government overvalues current spending relative to future welfare compared to society. This bias captures the fact that governments in power can obtain private benefits from spending, for example by diverting resources towards their preferred spending categories or constituencies (e.g., Aguiar and Amador, 2011). Additionally, we assume that the shock to the value of spending in each period is privately observed by the government in power in that period. As discussed in Section 2, this may reflect governments having superior information about the cost of public goods or aggregate citizen preferences. More broadly, this assumption says that current policy does not depend on past fiscal needs above and beyond what is captured by past policy decisions; such needs are difficult to quantify relative to observable fiscal variables.

An equilibrium in our setting prescribes a level of borrowing for each government in each period. This level of borrowing is a function of the government's observed shock and the history of past borrowing decisions. While multiple equilibria may arise, we focus on the best equilibrium for society, namely the one that maximizes social welfare at the beginning of time. We provide a recursive representation of this equilibrium and study its properties.

To describe the forces underlying our model, consider first what would happen in the absence of either a deficit bias or private information. If governments are not biased towards current spending, then trivially the best equilibrium has each government choosing the first-best policy (i.e. the policy that maximizes social welfare in each period), even if shocks to the value of spending are private information. Moreover, because shocks are i.i.d., fiscal policy (conditional on debt) features no history-dependence.

If governments are deficit-biased but fiscal shocks are publicly observable, the best equilibrium maximizes social welfare subject to a limited commitment constraint. This constraint requires that, for each shock, each government prefer its prescribed level of borrowing and continuation value to any other borrowing level. Since all deviations are public, they are punished (*off path*) with the lowest possible continuation value conditional on the government's choice of debt. The best equilibrium prescribes the first-best policy if this limited commitment constraint does not bind, or the lowest enforceable borrowing level if the constraint binds. In either case, the equilibrium restarts in each period, so fiscal policy again features no history-dependence.

Our setting combines both a deficit bias and private information. Because governments cannot directly condition their policy choices on past shocks under private information, the limited commitment constraint described above is insufficient: a government can now deviate *privately* from its prescribed policy and choose a higher borrowing level without being penalized with a low continuation value. The best equilibrium in this setting therefore maximizes social welfare subject to not only the limited commitment constraint but also a private information constraint: for each government and shock, the government must prefer its prescribed level of borrowing and continuation value to those prescribed for any other shock.

We show that the best equilibrium is characterized by a fiscally responsible low-deficit regime that maximizes social welfare and a fiscally irresponsible high-deficit regime that minimizes social welfare. Transitions between regimes are triggered by high enough fiscal shocks. Furthermore, unlike under observable shocks, temporary transitions may occur *on path*, and fiscal policy therefore is history-dependent.

Our characterization shows that fiscal policy in each regime admits a simple form. The fiscally responsible regime takes the form of a maximally enforced deficit limit. If a government chooses borrowing below the limit, the equilibrium restarts in the fiscally responsible regime in the next period. If instead a government violates the limit, the equilibrium transitions to the fiscally irresponsible regime. Governments may be unconstrained by the deficit limit when experiencing low shocks, but they are constrained under high shocks, and in some cases they break the limit.

The fiscally irresponsible regime takes the mirror form of a maximally enforced surplus limit. If a government chooses borrowing above the limit, the equilibrium returns to the fiscally responsible regime in the next period. If instead a government violates the limit, the equilibrium restarts in the fiscally irresponsible regime. Governments are unconstrained by the surplus limit when experiencing high shocks, but they are constrained under low shocks, and in some cases they break the limit.

A key feature of our environment is that while governments overweigh present spending, they share the same preferences as society for fiscal responsibility in the future. Thus, a maximally enforced deficit limit maximizes social welfare by counteracting the political bias to overborrow: governments are rewarded for choosing low borrowing with a fiscally responsible continuation regime and are punished for choosing high borrowing with a fiscally irresponsible continuation regime. Analogously, a maximally enforced surplus limit—which serves as a punishment—minimizes social welfare by exacerbating the political bias. We show that the promise of returning to fiscal responsibility induces governments to overborrow even relative to their bias, and this reduces social welfare beyond what would be generated in an absorbing Markov outcome. Hence, punishment is always temporary in the best equilibrium for society.

Our analysis can help explain the empirical path of fiscal policy. Periods of fiscal consolidation can be understood as fiscally responsible behavior by governments which realize that deviating from such behavior would set a precedent for deviations by subsequent governments. As such, periods of fiscal consolidation end when shocks are sufficiently severe that the cost of setting this negative precedent is outweighed by the benefit of responding to current economic conditions. Similarly, periods of profligacy can be understood as fiscally irresponsible behavior by governments which derive benefits from current spending and realize that future fiscal consolidations will occur once deficits become large enough following severe shocks. These dynamics imply that fiscal policy depends on the history and cannot be explained by contemporaneous variables alone, and that persistent changes in policy are punctuated by crisis periods.

While the results we have described point to the existence of regimes which sustain each other, they do not tell us whether regime transitions will necessarily occur in equilibrium. To provide insight into this question, we study the conditions for regime transitions in the context of an analytical example. Specifically, we examine a setting in which the social value of government spending takes a log form, which allows us to characterize the best equilibrium as the solution to a factorization algorithm. We show that fiscal regimes can arise only if governments' bias towards current spending is sufficiently large; furthermore, for a range of such biases, regime transitions occur *on path* in the best equilibrium. Intuitively, as governments' deficit bias increases, the threat of fiscal irresponsibility in the future also increases, and this in turn makes it possible to sustain a regime of fiscal responsibility in the present. Using our factorization algorithm, we compute a simulation and illustrate the regime transitions numerically.

A takeaway from our analysis is that governments' deficit bias is a key factor behind the emergence of fiscal regimes. The political economy literature has documented an increase in political biases over the past several decades, and has emphasized the effect of these biases on rising debt levels across advanced economies.[2] Our paper complements this literature by showing that increased biases lead not only to higher long-run debt growth, but also to the presence of persistent regimes in fiscal policy. This finding is consistent with the data and can further inform future empirical work; we suggest some directions in our concluding remarks in Section 6.

Related literature. As noted, our paper contributes to the literature on the political economy of government debt (see fn. 2). Dynamic models related to ours include Battaglini and Coate (2008), Yared (2010), and Dovis, Golosov, and Shourideh (2016). Unlike these papers, we study a setting that features private government information and yields fiscal policy regimes, and we examine how the emergence of regimes depends on the underlying political economy frictions.[3] There is also related research showing that fiscal policy is partly driven by the electoral cycle (e.g., Drazen, 2000; Müller, Storesletten, and Zilibotti, 2016) and is correlated with the government's party identity. In contrast to this work, we show that fiscal regimes can emerge even when parties in power have the same preferences and face no political risk; this is consistent with evidence that fiscal consolidations often span multiple political cycles.

A growing literature studies the impact of private government information for pol-

[2]There is a large literature dating back to Persson and Svensson (1989) and Alesina and Tabellini (1990) that provides political microfoundations for the deficit bias. See Alesina and Passalacqua (2016) and Yared (2019) for surveys. Yared (2019) argues that an increasingly older population, rising political polarization, and rising electoral uncertainty have led to increased political biases across advanced economies.

[3]Battaglini and Coate (2008) and Dovis, Golosov, and Shourideh (2016) analyze Markov perfect equilibria, where policy does not inherit more persistence than payoff relevant state variables. Yared (2010) examines the efficient sustainable equilibrium which admits S,s rules for policy and therefore more persistence than in Markov perfect equilibria, but this persistence dissipates after one period.

icy.[4] Within this literature, our paper relates to work on the tradeoff between commitment and flexibility in policymaking, including Athey, Atkeson, and Kehoe (2005), Amador, Werning, and Angeletos (2006), and Halac and Yared (2014, 2022). Our results pertaining to the fiscally responsible regime use tools developed in Halac and Yared (2022), which examines optimal fiscal rules under private information and limited enforcement. That paper considers a static model with exogenously enforced penalties, whereas here we study a dynamic model in which any punishments must be self-enforced by future equilibrium behavior. Also related are Halac and Yared (2020, 2021), which apply similar tools to monetary policy settings with no state variables.[5]

By studying the welfare-maximizing policy for present-biased governments, our analysis contributes to the literature on hyperbolic discounting that builds on Phelps and Pollak (1968) and Laibson (1994, 1997).[6] Recent related papers include Bernheim, Ray, and Yeltekin (2015), Bisin, Lizzeri, and Yariv (2015), and Lizzeri and Yariv (2017). Bernheim, Ray, and Yeltekin (2015) find that the optimal self-enforcing rule for a consumer with quasi-hyperbolic preferences entails temporary overspending as punishment. However, their setting has no private information, and as such punishment never occurs along the equilibrium path.

We provide a "bang-bang" result and regime dynamics which are reminiscent of the analysis of price wars in Green and Porter (1984) and more broadly related to the dynamics of repeated games in Abreu, Pearce, and Stacchetti (1990) and Sannikov (2007). These papers consider repeated moral hazard settings with finite actions (and thus finite incentive constraints) and a continuum of shocks. Their techniques do not directly apply to our problem which is one of adverse selection and features a continuum of actions and shocks.[7] In the context of government policy, Atkeson and Kehoe (2001) and Atkeson, Chari, and Kehoe (2007) achieve a bang-bang characterization, but that work is also concerned with settings of moral hazard.

Finally, the approach that we use to examine the analytical example of Section 5 resembles the factorization algorithm of Abreu, Pearce, and Stacchetti (1990) but for our adverse selection problem.[8] Exploiting the single-dimensionality of the value set

[4] See Sleet (2004), Ales, Maziero, and Yared (2014), Dovis (2019), Amador and Phelan (2021), and Dovis and Kirpalani (2021), among others.

[5] The claims in Halac and Yared (2020) (which contains no proofs) rely on the results presented in Section 4 of the present paper and their proofs.

[6] See also Strotz (1956), Calvo and Obstfeld (1988), and Barro (1999) for a related treatment of dynamically-inconsistent government preferences.

[7] Athey, Bagwell, and Sanchirico (2004) study related issues in a repeated Bertrand game with private information.

[8] See also Phelan and Stacchetti (2003).

in our setting, we are able to characterize the best equilibrium as a fixed point of an operator function. In contrast to Abreu, Pearce, and Stacchetti (1990), we apply the algorithm starting from the smallest rather than the largest set, providing necessary conditions for the convergence to a fixed point that exceeds the Markov outcome. This method might potentially be useful in other games to determine when the Markov outcome can be improved upon.

2 Model

We present a simple model in which successive governments make borrowing decisions. We describe our setup in Subsection 2.1, define our equilibrium concept in Subsection 2.2, and provide a recursive representation of the welfare-maximizing equilibrium in Subsection 2.3.

2.1 Setup

Consider an infinite horizon small open economy with periods $t = \{0, 1, \ldots\}$ and a different government in each period. At the beginning of each period t, an i.i.d. shock to the economy $\theta_t > 0$ is drawn from a bounded set $\Theta \equiv [\underline{\theta}, \overline{\theta}]$, with a continuously differentiable probability density function $f(\cdot) > 0$ and associated cumulative density function $F(\cdot)$. The realization of this shock is privately observed by the government in power at date t, so we refer to θ_t as this government's *type*.

We denote by $b_t \gtreqless 0$ and $g_t \geq 0$ respectively the government's choices of debt and spending at date t. The government's budget constraint is given by

$$g_t = \tau - Rb_{t-1} + b_t, \tag{1}$$

where $\tau > 0$ is the exogenous tax revenue and $R > 1$ is the exogenous gross interest rate on government bonds.[9]

Social welfare at date t is

$$V_t = \mathbb{E}_t \left[\sum_{k=0}^{\infty} \delta^k \theta_{t+k} U(g_{t+k}) \right],$$

[9]Our analysis also applies if instead of having an exogenous tax revenue, social welfare is an increasing function of the budget deficit.

or equivalently, rewriting it recursively,

$$V_t = \mathbb{E}_t \left[\theta_t U(g_t) + \delta V_{t+1}\right].$$

Here $\delta \in (0,1)$ denotes the social discount factor and $U(\cdot)$ is the utility of government spending, which we assume to be strictly increasing and strictly concave. Observe that a large shock $\theta_t > 0$ implies a large social value of government spending, as would be the case in an economic crisis.

The welfare of the government at date t, when choosing policy following the realization of θ_t, is

$$\alpha\theta_t U(g_t) + \delta V_{t+1}, \tag{2}$$

where $\alpha > 1$ represents the government's deficit bias.

There are three main features of our environment. First, since $\alpha > 1$, government preferences differ from those of society. The bias α captures the fact that a government in power at date t derives private benefits from spending at t, for example by being able to divert resources towards its preferred spending categories or towards a political constituency. The government at date t however shares the same preferences as society from date $t+1$ onward, as it does not receive any private benefits from spending in the future when it is no longer in power.

The second feature of our environment is that the shock θ_t is privately observed by the government at date t (and thus not observed by future governments). One interpretation is that the exact cost of public goods at t is only observable to the government in office, which may be inclined to overspend on these goods. Another possibility is that citizens have heterogeneous preferences or heterogeneous information about the optimal level of public spending, and only the current government sees the aggregate (Sleet, 2004; Piguillem and Schneider, 2016). A final possibility is that future governments observe θ_t but do not condition their behavior directly on this past realization, for example because it is not as easily quantifiable as the history of fiscal variables that do inform their behavior.[10]

The third feature of our environment is that governments have full discretion when choosing policy. At each date t, the government is able to freely choose any level of debt, subject only to feasibility as we describe next. Without any exogenously enforced incentives, it is only the behavior of future governments which can serve as reward and

[10]Under this interpretation and symmetric strategies, the parameter α can be viewed as being inversely related to the probability of reelection in a setting with government turnover; see Aguiar and Amador (2011).

punishment for a government's policy decisions.

We complete the description of our environment with a technical constraint. We require the level of debt in each period t to satisfy $b_t \in [\underline{b}(b_{t-1}), \overline{b}(b_{t-1})]$ for some exogenous bounds $\underline{b}(b_{t-1}) > Rb_{t-1} - \tau$ and $\overline{b}(b_{t-1}) < \tau/(R-1)$. These bounds on debt (which we allow to be a function of the inherited debt) guarantee that payoffs in our model are bounded (see Laibson, 1994). We also assume that the exogenous initial level of debt at time 0 satisfies $b_{-1} < \tau/(R-1)$.

2.2 Equilibrium Definition

We consider the interaction between the successive governments in each period $t = \{0, 1, \ldots\}$. Let $h^{t-1} = \{b_{-1}, b_0, \ldots, b_{t-1}\}$ denote the history of debt through time $t-1$ and $\mathcal{H}^{t-1}$ the set of all possible such histories. A strategy for the government in period t is $\sigma_t\left(h^{t-1}, \theta_t\right)$, specifying, for each history $h^{t-1} \in \mathcal{H}^{t-1}$ and government type $\theta_t \in \Theta$, a feasible level of debt $b_t\left(h^{t-1}, \theta_t\right)$. Note that given the budget constraint (1), a history of debt also pins down the history of spending, and the government's strategy also pins down its choice of spending. Specifically, denoting the available resources at history h^{t-1} by

$$\omega_t(h^{t-1}) \equiv \tau - Rb_{t-1}(h^{t-1}),$$

spending given h^{t-1} and θ_t is $g_t\left(h^{t-1}, \theta_t\right) = \omega_t(h^{t-1}) + b_t\left(h^{t-1}, \theta_t\right)$.

An equilibrium is defined as a profile of strategies $\sigma = \left(\sigma_t\left(h^{t-1}, \theta_t\right)\right)_{t=0}^{\infty}$ such that, for each $t \in \{0, 1, \ldots\}$, $\sigma_t\left(h^{t-1}, \theta_t\right)$ maximizes the date-t government's welfare (2) given the continuation strategies $\left(\sigma_{t+k}\left(h^{t+k-1}, \theta_{t+k}\right)\right)_{k=1}^{\infty}$ of all future governments. Given an equilibrium, let $V_t(h^{t-1})$ denote the continuation value to society at date t starting from (on- or off-path) history h^{t-1}. This continuation value can be represented recursively as

$$V_t(h^{t-1}) = \mathbb{E}_t\left[\theta_t U(\omega_t(h^{t-1}) + b_t(h^{t-1}, \theta_t)) + \delta V_{t+1}(h^{t-1}, b_t(h^{t-1}, \theta_t))\right].$$

A profile of strategies $\left(\sigma_t\left(h^{t-1}, \theta_t\right)\right)_{t=0}^{\infty}$ constitutes an equilibrium if and only if it satisfies the governments' private information and limited commitment constraints for all $t \in \{0, 1, \ldots\}$ and all (on- and off-path) histories h^{t-1}. The private information constraint captures the fact that the government at any date t can deviate *privately* by choosing a level of borrowing prescribed for a type different from its own. To guarantee that a government of type θ_t prefers to pursue its prescribed level of borrowing rather than that of any other type $\theta_t' \neq \theta_t$, we must have

$$\alpha\theta_t U(\omega_t(h^{t-1}) + b_t(h^{t-1},\theta_t)) + \delta V_{t+1}(h^{t-1}, b_t(h^{t-1},\theta_t)) \\ \geq \alpha\theta_t U(\omega_t(h^{t-1}) + b_t(h^{t-1},\theta_t')) + \delta V_{t+1}(h^{t-1}, b_t(h^{t-1},\theta_t')) \tag{3}$$

for all $\theta_t, \theta_t' \in \Theta$.

The limited commitment constraint captures the fact that the government at any date t can deviate *publicly* by choosing a level of borrowing not prescribed for any type. To guarantee that a government of type θ_t prefers to pursue its prescribed level of borrowing rather than any other level of borrowing b_t' satisfying $b_t' \neq b_t(h^{t-1},\theta_t')$ for all $\theta_t' \in \Theta$, we must have

$$\alpha\theta_t U(\omega_t(h^{t-1}) + b_t(h^{t-1},\theta_t)) + \delta V_{t+1}(h^{t-1}, b_t(h^{t-1},\theta_t)) \\ \geq \alpha\theta_t U(\omega_t(h^{t-1}) + b_t') + \delta V_{t+1}(h^{t-1}, b_t') \tag{4}$$

for all $\theta_t \in \Theta$ and all b_t' satisfying $b_t' \neq b_t(h^{t-1},\theta_t')$ for all $\theta_t' \in \Theta$.

Since debt is bounded and shocks are i.i.d., there exists an upper bound $\overline{V}(b_t)$ that corresponds to the highest continuation value that can be sustained by equilibrium strategies conditional on debt b_t, with $V_{t+1}(h^{t-1}, b_t) \leq \overline{V}(b_t)$ for all h^{t-1} and b_t. By analogous logic, there also exists a lower bound $\underline{V}(b_t)$, with $V_{t+1}(h^{t-1}, b_t) \geq \underline{V}(b_t)$ for all h^{t-1} and b_t. Given available resources $\omega_t(h^{t-1})$, let $b_t^p(\omega_t(h^{t-1}),\theta_t)$ denote type θ_t's *flexible* level of debt conditional on being punished with this lowest continuation value:

$$b_t^p(\omega_t(h^{t-1}),\theta_t) \in \arg \max_{b_t \in [\underline{b}(b_{t-1}), \overline{b}(b_{t-1})]} \{\alpha\theta_t U(\omega(h^{t-1}) + b_t) + \delta \underline{V}(b_t)\}. \tag{5}$$

Note that satisfying the limited commitment constraint (4) requires that the constraint hold under maximal punishment, namely when $V_{t+1}(h^{t-1}, b_t') = \underline{V}(b_t')$. In fact, given (3), the limited commitment constraint must then hold under maximal punishment for all $b_t' \in [\underline{b}(b_{t-1}), \overline{b}(b_{t-1})]$, and thus necessarily when $b_t' = b_t^p(\omega_t(h^{t-1}),\theta_t)$. Hence, a necessary condition for the limited commitment constraint to be satisfied is

$$\alpha\theta_t U(\omega_t(h^{t-1}) + b_t(h^{t-1},\theta_t)) + \delta V_{t+1}(h^{t-1}, b_t(h^{t-1},\theta_t)) \\ \geq \alpha\theta_t U(\omega_t(h^{t-1}) + b_t^p(h^{t-1},\theta_t)) + \delta \underline{V}(b_t^p(\omega_t(h^{t-1}),\theta_t)) \tag{6}$$

for all $\theta_t \in \Theta$, where the right-hand side is the government's minmax payoff.

Constraints (3) and (6) are clearly necessary for a sequence of debt to be supported by equilibrium strategies. Furthermore, these constraints are also sufficient: if a sequence of debt satisfies (3) and (6), then it can be supported by a strategy profile that specifies the worst feasible continuation equilibrium following any public deviation. Since such a deviation is off path, it is without loss to assume that it is maximally punished.

We define the best equilibrium for society as the equilibrium that maximizes date-0 social welfare $V_0(b_{-1})$ given inital debt b_{-1}.

2.3 Recursive Representation

Given the repeated nature of the game and the fact that shocks are i.i.d., we can represent policies in the best equilibrium recursively (see Abreu, Pearce, and Stacchetti, 1990; Chade, Prokopovych, and Smith, 2008). That is, rather than optimizing over an entire debt sequence, starting from any given date, we can assign each type $\theta \in \Theta$ of the government a level of debt $b(\theta)$ and continuation value $V(b(\theta))$, where these must satisfy the private information and limited commitment constraints, and where the continuation value must itself be drawn from the set of continuation values $\left[\underline{V}(b(\theta)), \overline{V}(b(\theta))\right]$ that satisfy the private information and limited commitment constraints. Let ω be the level of resources associated with initial debt b_{-1} at date 0. Then the best equilibrium for society, which maximizes social welfare at date 0, corresponds to the solution to the following program:

$$\overline{V}(b_{-1}) = \max_{(b(\theta), V(b(\theta)))} \mathbb{E}\left[\theta U(\omega + b(\theta)) + \delta V(b(\theta))\right] \tag{$\mathcal{P}_{\max}$}$$

subject to

$$\alpha\theta U(\omega + b(\theta)) + \delta V(b(\theta)) \geq \alpha\theta U(\omega + b(\theta')) + \delta V(b(\theta')) \text{ for all } \theta, \theta' \in \Theta \tag{7}$$

$$\alpha\theta U(\omega + b(\theta)) + \delta V(b(\theta)) \geq \alpha\theta U(\omega + b^p(\omega,\theta)) + \delta \underline{V}(b^p(\omega,\theta)) \text{ for all } \theta \in \Theta \tag{8}$$

$$b(\theta) \in [\underline{b}(b_{-1}), \bar{b}(b_{-1})] \text{ and } V(b(\theta)) \in [\underline{V}(b(\theta)), \overline{V}(b(\theta))] \text{ for all } \theta \in \Theta. \tag{9}$$

Constraints (7) and (8) are the private information and limited commitment constraints, analogous to (3) and (6), with $b^p(\omega,\theta)$ denoting the government's flexible borrowing level conditional on the lowest continuation value, analogous to (5):

$$b^p(\omega,\theta) \in \arg\max_{b \in [\underline{b}(b_{-1}), \bar{b}(b_{-1})]} \{\alpha\theta U(\omega + b) + \delta \underline{V}(b)\}.$$

Constraint (9) is the feasibility constraint, requiring that the government's debt be within the exogenously specified bounds, and that continuation values be drawn from the set of equilibrium values. We assume that the solution to program ($\mathcal{P}_{\max}$) admits a piecewise continuously differentiable function $b(\theta)$, which allows us to establish our results using perturbations.[11]

Given the continuation value set $\left[\underline{V}(b(\theta)), \overline{V}(b(\theta))\right]$ and the characterization of equilibria that we will derive, it is also useful to write down the program that yields the lowest value for society given initial debt b_{-1}:

$$\underline{V}(b_{-1}) = \min_{(b(\theta),V(b(\theta)))} \mathbb{E}\left[\theta U(\omega + b(\theta)) + \delta V(b(\theta))\right] \qquad (\mathcal{P}_{\min})$$

subject to (7), (8), and (9).

We make the following assumption:

Assumption 1. *Parameters are such that $\overline{V}(\cdot)$ and $\underline{V}(\cdot)$ are continuously differentiable and concave and satisfy $\overline{V}(b) > \underline{V}(b)$ for all finite $b < \tau/(R-1)$.*

The first part, on the differentiability and concavity of the value functions, is guaranteed to hold for example if preferences $U(\cdot)$ satisfy either constant absolute risk aversion (CARA) or constant relative risk aversion (CRRA). The second part of Assumption 1 says that the equilibrium set is not a singleton. Section 5 provides necessary and sufficient conditions on parameters for this to hold in a setting with $U(\cdot) = \log(\cdot)$.

3 Benchmarks

To understand the role of governments' deficit bias and private information, it is instructive to consider what would happen in the absence of either of these frictions. Suppose first that governments are not biased towards current spending and thus $\alpha = 1$. Then trivially the best equilibrium for society has each government choosing the first-best policy, namely the policy that maximizes social welfare in each period. The governments' private information plays no role absent a deficit bias, and social welfare is always at its first-best level.

Suppose next that $\alpha > 1$ but fiscal shocks are publicly observable. Then all deviations are public, implying that they are punished (off path) with the worst feasible continuation equilibrium, and thus that the private information constraint (7) in program ($\mathcal{P}_{\max}$) can be ignored. The best equilibrium for society therefore prescribes either the first-best policy if the limited commitment constraint (8) does not bind, or the lowest enforceable level of debt conditional on the realized shock if (8) binds. In

[11] If the program admits multiple solutions that differ only on a countable set of types, we select the solution that maximizes social welfare for those types.

either case, we show that the highest feasible continuation value is prescribed for all shocks, and hence social welfare is at its highest feasible level $\overline{V}(\cdot)$ at all dates.

Lemma 1. *If $\alpha = 1$ or θ is observable, then $V_{t+1}(h^{t-1}, b_t) = \overline{V}(b_t)$ at every on-path history h^t in the best equilibrium.*

The takeaway is that neither of these benchmark settings can explain the presence of fiscal policy regimes. Since the best equilibrium restarts at each date, fiscal policy (conditional on debt) is independent of the history when governments have either no deficit bias or no private information.

4 Fiscal Policy Regimes

We now study the best equilibrium for society subject to the governments' deficit bias and private information, which corresponds to the solution to program ($\mathcal{P}_{\max}$). Subsection 4.1 presents some preliminaries. Subsection 4.2 shows that the best equilibrium is characterized by two regimes. Subsection 4.3 and Subsection 4.4 examine the form that fiscal policy takes in each of the two regimes, and Subsection 4.5 discusses transitions between the regimes.

4.1 Preliminaries

We provide some preliminaries that allow us to rewrite social welfare in a convenient way. Consider the private information constraint (7). It follows from standard arguments that $(b(\theta), V(b(\theta)))$ satisfies this global private information constraint if and only if $b(\theta)$ is nondecreasing and $(b(\theta), V(b(\theta)))$ satisfies the corresponding local private information constraints.[12] Moreover, the local private information constraints imply that the derivative of government welfare with respect to θ is $\alpha U(\omega + b(\theta))$. Hence, in an equilibrium, government welfare for type $\theta \in \Theta$ satisfies

$$\alpha\theta U(\omega + b(\theta)) + \delta V(b(\theta)) = \alpha\underline{\theta}U(\omega + b(\underline{\theta})) + \delta V(b(\underline{\theta})) + \int_{\underline{\theta}}^{\theta} \alpha U(\omega + b(\widetilde{\theta}))d\widetilde{\theta}. \quad (10)$$

[12] We provide a formal statement in Lemma 2 in Appendix A. Observe that given $b(\theta)$ nondecreasing, satisfaction of (7) requires that $V(b(\theta))$ be nonincreasing in θ.

Following Amador, Werning, and Angeletos (2006), we can substitute (10) into the objective in ($\mathcal{P}_{\max}$) to rewrite social welfare as

$$\alpha\underline{\theta}U(\omega + b(\underline{\theta})) + \delta V(b(\underline{\theta})) + \alpha \int_{\underline{\theta}}^{\overline{\theta}} U(\omega + b(\theta))Q(\theta)d\theta, \tag{11}$$

where

$$Q(\theta) \equiv 1 - F(\theta) - \theta f(\theta)(1 - 1/\alpha).$$

$Q(\theta)$ represents the social value of increasing the level of borrowing prescribed for a government type θ. To interpret it, observe that the first term, $1 - F(\theta)$, resembles that in a *virtual surplus* formulation in mechanism design (Myerson, 1981).[13] This term captures the fact that increasing the borrowing prescribed for a government type θ requires increasing the borrowing prescribed for types higher than θ, so that their welfare increases at the same rate as required by the private information constraint (see (10)). The second term in $Q(\theta)$ reflects the fact that, given the deficit bias $\alpha > 1$, society and the government disagree on the value of current borrowing. Prescribing more borrowing for a government type θ reduces social welfare relative to government welfare by $-\theta f(\theta)(1 - 1/\alpha)$, where $\theta f(\theta)$ is the weight that social welfare places on the current utility from borrowing by type θ, and $(1-1/\alpha)$ is the extent of the disagreement between society and the government.

The formulation above will be useful for our characterization of the best equilibrium in the next sections, which will appeal to properties of the function $Q(\theta)$.

4.2 Two Regimes

We show that if the best equilibrium prescribes interior levels of debt at all on-path histories, then the induced social welfare is always either at its highest or lowest feasible level.

Proposition 1. *Assume $Q'(\theta) \neq 0$ a.e., and suppose $b_t \in (\underline{b}(b_{t-1}), \overline{b}(b_{t-1}))$ at all on-path histories h^t in the best equilibrium. Then $V_{t+1}(h^{t-1}, b_t) \in \left\{\underline{V}(b_t), \overline{V}(b_t)\right\}$ at every such history.*

To prove this result, we first consider program ($\mathcal{P}_{\max}$) which yields the highest

[13] Similar to the standard virtual surplus expression, we can rewrite (11) using the inverse hazard rate:

$$\alpha\underline{\theta}U(\omega + b(\underline{\theta})) + \delta V(b(\underline{\theta})) + \alpha \int_{\underline{\theta}}^{\overline{\theta}} U(\omega + b(\theta)) \left[\frac{1 - F(\theta)}{f(\theta)} - \theta\left(1 - \frac{1}{\alpha}\right)\right] f(\theta)d\theta.$$

feasible welfare $\overline{V}(b_{-1})$ given initial debt b_{-1}. We show that if $(b(\theta), V(b(\theta)))$ is a solution to this program with $b(\theta) \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ for all θ, then the prescribed continuation values satisfy $V(b(\theta)) \in \{\underline{V}(b(\theta)), \overline{V}(b(\theta))\}$ for all θ. The proof uses perturbation arguments developed in Halac and Yared (2022). For intuition, recall that $Q(\theta)$ represents the weight that society assigns to prescribing more borrowing for a government type θ. The condition in Proposition 1 says that the set of types θ for which $Q'(\theta) = 0$ is nowhere dense,[14] which implies that $Q(\theta)$ is either strictly decreasing or strictly increasing over any sufficiently small interval. It follows that society prefers to concentrate borrowing on either lower or higher types in the interval. If $V(b(\theta)) \in (\underline{V}(b(\theta)), \overline{V}(b(\theta)))$ for some θ, we then show that there is a perturbation that strictly increases social welfare. In particular, we improve by compressing borrowing over the interval when $Q'(\theta) < 0$, and by spreading out borrowing when $Q'(\theta) > 0$.[15]

Given the solution to ($\mathcal{P}_{\max}$), we then consider program ($\mathcal{P}_{\min}$) which yields the lowest feasible welfare $\underline{V}(b_{-1})$ given initial debt b_{-1}. We show that analogous perturbation arguments apply to this program; essentially, any perturbation that increases welfare when $Q(\theta)$ is increasing (decreasing) then reduces welfare when $Q(\theta)$ is decreasing (increasing). Hence, we obtain that if $(b(\theta), V(b(\theta)))$ is a solution to ($\mathcal{P}_{\min}$) with $b(\theta) \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ for all θ, then $V(b(\theta)) \in \{\underline{V}(b(\theta)), \overline{V}(b(\theta))\}$ for all θ. Moreover, since the results for ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$) hold for any finite $b_{-1} < \tau/(R-1)$, it follows that if debt is interior at all on-path histories, then continuation values only travel to the extreme points of the feasible set in the best equilibrium.

We emphasize that this "bang-bang" property of continuation values is *necessary* for the maximization of social welfare at time 0: Proposition 1 says that any equilibrium with interior continuation values is strictly dominated. As noted in the Introduction, this result relates to the bang-bang dynamics of repeated games in Abreu, Pearce, and Stacchetti (1990) and Sannikov (2007). Those models however feature moral hazard, whereas our setting is one of adverse selection.

The implication of Proposition 1 is that fiscal policy is characterized by two regimes. At any point in the best equilibrium, governments are either in a regime that maximizes social welfare—which, for reasons that will become evident, we will call the *fiscally responsible regime*—or in a regime that minimizes social welfare—which we will call

[14] Given $f(\theta)$ continuously differentiable, this condition holds generically. Specifically, this condition fails only if $\theta f'(\theta)/f(\theta) = -(2-1/\alpha)/(1-1/\alpha)$ for a positive mass of types θ, but then any arbitrarily small perturbation of α would render the condition true.

[15] This perturbation rules out interior continuation values that are continuously decreasing over the interval. An analogous perturbation can be used to rule out interior continuation values in the form of a step function.

the *fiscally irresponsible regime*. Since $\overline{V}(\cdot) > \underline{V}(\cdot)$ by Assumption 1, the policies in the two regimes are distinct from each other. Thus, if both regimes occur on path in the best equilibrium, then fiscal policy features history dependence: conditional on debt, the policy that is implemented at a given date depends on whether the economy is in the fiscally responsible or fiscally irresponsible regime.

There are a number of questions that Proposition 1 raises. First, what form does fiscal policy take in each of the two regimes? Second, what triggers a transition from one regime to the other? And finally, can regime transitions occur on path in the best equilibrium? We address the first two questions in Subsection 4.3-Subsection 4.5 and the last question in Section 5. To facilitate our analysis, we maintain the following assumption for the rest of the paper:

Assumption 2. *There is $\widehat{\theta} \in \Theta$ such that $Q'(\theta) < 0$ if $\theta < \widehat{\theta}$ and $Q'(\theta) > 0$ if $\theta > \widehat{\theta}$.*

This assumption says that for $\theta < \widehat{\theta}$, society prefers to concentrate borrowing on relatively low government types, whereas for $\theta > \widehat{\theta}$, society prefers to concentrate borrowing on relatively high government types. Note that we allow for $\widehat{\theta}$ to be interior (with $Q(\theta)$ then non-monotonic) or at a boundary of the set Θ (with $Q(\theta)$ then monotonic). Assumption 2 holds for a broad range of distribution functions, including uniform, exponential, log-normal, gamma, and beta for a subset of its parameters, and is analogous to assumptions used in Amador, Werning, and Angeletos (2006) and Halac and Yared (2022).

4.3 Fiscal Responsibility

We study fiscal policy in the fiscally responsible regime by characterizing the solution to program ($\mathcal{P}_{\max}$). Given resources ω associated with initial debt b_{-1}, recall that $b^p(\omega, \theta)$ denotes the government's flexible borrowing level conditional on the lowest continuation value. We analogously define $b^r(\omega, \theta)$ as the government's flexible borrowing level conditional on the highest continuation value:

$$b^r(\omega, \theta) \in \arg \max_{b \in [\underline{b}(b_{-1}), \overline{b}(b_{-1})]} \{\alpha\theta U(\omega + b) + \delta\overline{V}(b)\}.$$

We obtain the following result:

Proposition 2. *If* $(b(\theta), V(b(\theta)))$ *is a solution to* ($\mathcal{P}_{\max}$) *with* $b(\theta) \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ *for all* $\theta \in \Theta$*, then there exist* $\theta^* \in [0, \overline{\theta})$ *and finite* $\theta^{**} > \max\{\theta^*, \underline{\theta}\}$ *such that*

$$(b(\theta), V(b(\theta))) = \begin{cases} \left(b^r(\omega,\theta), \overline{V}(b^r(\omega,\theta))\right) & \text{if } \theta < \theta^* \\ \left(b^r(\omega,\theta^*), \overline{V}(b^r(\omega,\theta^*))\right) & \text{if } \theta \in [\theta^*, \theta^{**}] \\ \left(b^p(\omega,\theta), \underline{V}(b^p(\omega,\theta))\right) & \text{if } \theta > \theta^{**} \end{cases} \quad (12)$$

where

$$\alpha\theta^{**}U(\omega + b^r(\omega,\theta^*)) + \delta\overline{V}(b^r(\omega,\theta^*)) = \alpha\theta^{**}U(\omega + b^p(\omega,\theta^{**})) + \delta\underline{V}(b^p(\omega,\theta^{**})). \quad (13)$$

The left panel of Figure 2 displays an example of the policy described in Proposition 2. We interpret this policy as a *maximally enforced deficit limit*, associated with the borrowing level $b^r(\omega,\theta^*)$. Governments that respect the deficit limit (by borrowing below $b^r(\omega,\theta^*)$) are rewarded with a continuation in the fiscally responsible regime which yields the highest feasible continuation value $\overline{V}(\cdot)$, whereas governments that break the limit are punished with a transition to the fiscally irresponsible regime which yields the lowest feasible continuation value $\underline{V}(\cdot)$. Proposition 2 shows that low types, $\theta < \theta^*$, respect the limit by borrowing at their flexible level conditional on the highest continuation value, $b^r(\omega,\theta)$. High types, $\theta > \theta^{**}$, break the limit by borrowing at their flexible level conditional on the lowest continuation value, $b^p(\omega,\theta)$. Types in between, $\theta \in [\theta^*, \theta^{**}]$, are constrained by the limit but respect it; they borrow at the limit level $b^r(\omega,\theta^*)$ to avoid punishment.[16] This policy thus incentivizes governments to reduce their overborrowing, explaining our term of fiscal responsibility.

Observe that the maximally enforced deficit limit can take one of two forms. One possible form has $\theta^{**} \geq \overline{\theta}$. In this case, the government respects the deficit limit under all shocks, so the economy remains in the fiscally responsible regime associated with welfare $\overline{V}(\cdot)$ in the following period.[17] The other possible form has $\theta^{**} < \overline{\theta}$. In this case, the government breaks the deficit limit under high enough shocks, $\theta > \theta^{**}$, so the economy transitions to the fiscally irresponsible regime associated with welfare $\underline{V}(\cdot)$ if such a shock realizes, and remains in the fiscally responsible regime otherwise. We provide conditions for regime transitions, as well as a concrete example, in Section 5.

To describe the proof of Proposition 2, recall from Proposition 1 that any (inte-

[16] By (13), the limited commitment constraint holds with equality for type θ^{**}, and one can verify (see Lemma 5 in Appendix B) that this constraint and the private information constraint are satisfied for all types.

[17] This is always the case if $\widehat{\theta}$ defined in Assumption 2 is equal to $\overline{\theta}$, i.e., if $Q(\theta)$ is monotonically decreasing. See Proposition 4 in Halac and Yared (2022).

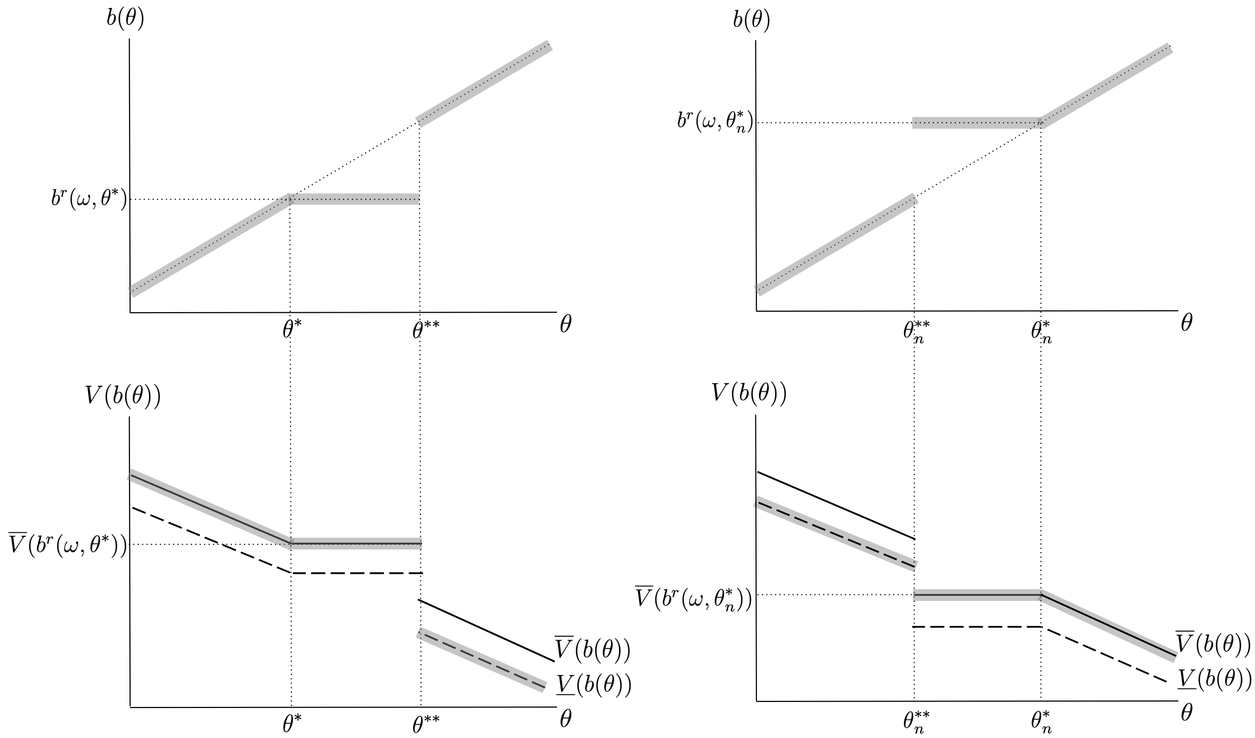

Figure 2: Examples of a maximally enforced deficit limit (left panel) and a maximally enforced surplus limit (right panel). The thick grey lines depict the prescribed levels of debt in the top graphs and the prescribed continuation values in the bottom graphs. The solid and dashed black lines in the bottom graphs depict $\overline{V}(b(\theta))$ and $\underline{V}(b(\theta))$ respectively. The figure is drawn under $U(\cdot) = \log(\cdot)$, in which case $\overline{V}(b(\theta)) - \underline{V}(b(\theta))$ equals a constant independent of $b(\theta)$ and thus $b^r(\omega, \theta) = b^p(\omega, \theta)$ for all ω and θ; see Section 5.

rior) solution to program ($\mathcal{P}_{\max}$) prescribes a continuation value $V(b(\theta))$ equal to either $\overline{V}(b(\theta))$ or $\underline{V}(b(\theta))$. We show that under Assumption 2, the prescribed continuation values are monotonic, with the government either never receiving the lowest continuation value $\underline{V}(b(\theta))$ or receiving such continuation value only under high enough shocks. Intuitively, Assumption 2 says that for types $\theta > \widehat{\theta}$, society prefers to concentrate borrowing on relatively high types; this is achieved by using high-powered incentives that specify the lowest continuation value for high levels of borrowing. In contrast, for types $\theta < \widehat{\theta}$, society prefers to concentrate borrowing on relatively low types; this is achieved by using flat incentives that prescribe the highest continuation value. Hence, we obtain that either all government types receive $\overline{V}(b(\theta))$, or only types above an interior point θ^{**} are punished with $\underline{V}(b(\theta))$. We further establish that the prescribed debt $b(\theta)$ is continuous for all $\theta \leq \theta^{**}$, and therefore that the solution must take the form of a maximally enforced deficit limit.

4.4 Fiscal Irresponsibility

In principle, different continuation equilibria could serve as punishment for a government breaking the maximally enforced deficit limit under fiscal responsibility. In fact, the result in Proposition 2 holds independently of the exact structure of $\underline{V}(\cdot)$. However, as we have noted, the best equilibrium uses the worst feasible punishment, as such a punishment maximally relaxes the constraints in program ($\mathcal{P}_{\max}$) and thus maximizes social welfare. We therefore study fiscal policy in the fiscally irresponsible regime by characterizing the solution to program ($\mathcal{P}_{\min}$) which minimizes social welfare. We obtain the following result:

Proposition 3. *If $(b(\theta), V(b(\theta)))$ is a solution to ($\mathcal{P}_{\min}$) with $b(\theta) \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ for all $\theta \in \Theta$, then there exist finite $\theta_n^* > \underline{\theta}$ and $\theta_n^{**} \in [\underline{\theta}, \min\{\theta_n^*, \overline{\theta}\})$ such that*

$$(b(\theta), V(b(\theta))) = \begin{cases} \left(b^r(\omega, \theta), \overline{V}(b^r(\omega, \theta))\right) & \text{if } \theta > \theta_n^* \\ \left(b^r(\omega, \theta_n^*), \overline{V}(b^r(\omega, \theta_n^*))\right) & \text{if } \theta \in [\theta_n^{**}, \theta_n^*] \\ \left(b^p(\omega, \theta), \underline{V}(b^p(\omega, \theta))\right) & \text{if } \theta < \theta_n^{**} \end{cases} \quad (14)$$

where

$$\alpha\theta_n^{**} U(\omega + b^r(\omega, \theta_n^*)) + \delta\overline{V}(b^r(\omega, \theta_n^*)) = \alpha\theta_n^{**} U(\omega + b^p(\omega, \theta_n^{**})) + \delta\underline{V}(b^p(\omega, \theta_n^{**})). \quad (15)$$

The right panel of Figure 2 displays an example of the policy described in Proposition 3. We interpret this policy as a *maximally enforced surplus limit*, associated with the borrowing level $b^r(\omega, \theta_n^*)$. Governments that respect the surplus limit (by borrowing above $b^r(\omega, \theta^*)$) are rewarded with a transition to the fiscally responsible regime which yields the highest feasible continuation value $\overline{V}(\cdot)$, whereas governments that break the limit are punished with a continuation in the fiscally irresponsible regime which yields the lowest feasible continuation value $\underline{V}(\cdot)$. Proposition 3 shows that high types, $\theta > \theta_n^*$, respect the limit by borrowing at their flexible level conditional on the highest continuation value, $b^r(\omega, \theta)$. Low types, $\theta < \theta_n^{**}$, break the limit by borrowing at their flexible level conditional on the lowest continuation value, $b^p(\omega, \theta)$. Types in between, $\theta \in [\theta_n^{**}, \theta_n^*]$, are constrained by the limit but respect it; they borrow at the limit level $b^r(\omega, \theta_n^*)$ to avoid punishment.[18] This policy thus incentivizes governments to increase their overborrowing, explaining our term of fiscal irresponsibility.

To see why inducing overborrowing minimizes social welfare, consider a given gov-

[18] By (15), the limited commitment constraint holds with equality for type θ_n^{**}, and one can verify that this constraint and the private information constraint are satisfied for all types.

ernment type. There are two ways in which the social welfare derived from this government can be made inefficiently low: either by inducing the government to borrow too little or by inducing it to borrow too much. Because governments are biased towards overborrowing, the latter option relaxes the limited commitment constraint and is a more efficient means of reducing welfare. Thus, in the fiscally irresponsible regime, all government types borrow above the socially optimal level; in fact, they all borrow weakly above, and some strictly above, their own preferred level.

Importantly, observe that the fiscally irresponsible regime is always temporary. This follows from the fact that the maximally enforced surplus limit described in Proposition 3 specifies $\theta_n^{**} < \overline{\theta}$. Hence, governments respect the surplus limit for all shocks $\theta \in [\theta_n^{**}, \overline{\theta}]$, implying that the best equilibrium transitions back to the fiscally responsible regime with strictly positive probability.

The proof of Proposition 3 uses analogous arguments as that of Proposition 2. One step in the proof that requires additional care is establishing that the maximally enforced surplus limit indeed specifies $\theta_n^{**} < \overline{\theta}$. In particular, we show that the fiscally irresponsible regime is not an absorbing Markov equilibrium in which $\underline{V}(\cdot)$ is sustained at all dates, with all government types $\theta \in \Theta$ choosing their flexible debt level conditional on the lowest continuation value, $b^p(\omega, \theta)$. We prove that a surplus limit that is respected by high enough types achieves lower social welfare than the Markov outcome.

The intuition for this result is as follows. Suppose $\theta_n^{**} = \overline{\theta}$, so that all government types $\theta \in \Theta$ choose their flexible debt level $b^p(\omega, \theta)$ and receive continuation value $\underline{V}(\cdot)$. Consider a perturbation where we reduce θ_n^{**} and make an arbitrarily small set of types $[\theta_n^{**}, \overline{\theta}]$ just willing to increase their borrowing above $b^p(\omega, \theta)$ to be rewarded with a continuation value $\overline{V}(\cdot)$. We show that making these high types indifferent on the margin implies that society is made strictly worse off with the perturbation. The key point is that the government overweighs borrowing in the present, while sharing the same preferences as society for increasing the continuation value in the future. Hence, if the government's cost of increasing overborrowing equals the benefit of increasing the continuation value, then the *social* cost of increasing overborrowing outweighs that benefit. It follows that a surplus limit as described in Proposition 3 serves as a more severe punishment than an absorbing regime taking the form of a Markov equilibrium.

4.5 Regime Transitions

The results in Proposition 1-Proposition 3 have implications for the dynamics of fiscal policy. Starting in a fiscally responsible regime at date t, the best equilibrium takes the form of a maximally enforced deficit limit, which aims to counteract the governments' deficit bias and limit overborrowing. If a shock $\theta_t \leq \theta^{**}$ is realized, the government at date t respects the deficit limit and the equilibrium restarts in the fiscally responsible regime at $t+1$. If instead $\theta_t > \theta^{**}$, the government at date t breaks the deficit limit and the equilibrium transitions to the fiscally irresponsible regime at $t+1$.

Starting in a fiscally irresponsible regime at date t, the best equilibrium takes the form of a maximally enforced surplus limit, which induces governments to succumb to their deficit bias and overborrow. If a shock $\theta_t \geq \theta_n^{**}$ is realized, the government at date t respects the surplus limit and the equilibrium transitions to the fiscally responsible regime at $t+1$. If instead $\theta_t < \theta_n^{**}$, the government at date t breaks the surplus limit and the equilibrium restarts in the fiscally irresponsible regime at $t+1$.

The characterization sheds light on the empirical path of fiscal policy discussed in the Introduction. Periods of fiscal consolidation can be understood as fiscally responsible behavior by governments which realize that deviating from such behavior would set a precedent for deviations by subsequent governments. As such, periods of fiscal consolidation end when shocks are sufficiently severe that the cost of setting this negative precedent is outweighed by the benefit of responding to current economic conditions. Analogously, periods of profligacy can be understood as fiscally irresponsible behavior by governments which derive benefits from current spending and realize that future fiscal consolidations will occur once deficits become sufficiently large. As such, periods of profligacy end when shocks are severe enough to demand such large deficits. Intuitively, it is in extreme situations that we see governments coordinate to change the equilibrium trajectory of policy.

Despite shocks being i.i.d., we find that fiscal policy (conditional on debt) is history-dependent and cannot be explained by contemporaneous variables alone. Governments' borrowing choices depend not only on current economic conditions and their inherited level of debt, but also on the regime in which they find themselves. Moreover, we find that persistent changes in fiscal policy are punctuated by crisis periods, as transitions between regimes occur when shocks to the value of spending are sufficiently high. These findings are consistent with the empirical patterns described in the Introduction as well as with econometric evidence that we discuss in Section 6.

5 Analytical Example

Our results in Proposition 1-Proposition 3 hold under Assumption 1, which guarantees that the value functions are continuously differentiable and concave with $\overline{V}(\cdot) > \underline{V}(\cdot)$.

Moreover, these results characterize the best equilibrium under interior solutions for debt. In this section, we describe an analytical example in which Assumption 1 holds and in which debt is interior along the equilibrium path, so that Proposition 1-Proposition 3 hold in all periods. Applying a factorization algorithm, we show that $\overline{V}(\cdot) > \underline{V}(\cdot)$ if and only if the governments' deficit bias α is large enough. This means that governments must be sufficiently biased towards present spending for the equilibrium to feature fiscal policy regimes.

5.1 Primitives and Preliminaries

Take a utility of government spending $U(\cdot) = \log(\cdot)$.[19] Let $s_t \in [0,1]$ denote the saving rate at time t, defined as the fraction of lifetime resources that are not spent at t:

$$g_t = (1 - s_t)R\left(\frac{\tau}{R-1} - b_{t-1}\right).$$

Then social welfare at date t can be written as

$$V_t = \mathbb{E}_t\left[\sum_{k=0}^{\infty}\delta^k\left(\theta_{t+k}U(1-s_{t+k}) + \frac{\delta}{1-\delta}\theta_{t+k}U(s_{t+k})\right)\right] + \chi(b_{t-1}), \tag{16}$$

where $\chi(b_{t-1})$ is a constant that depends on b_{t-1}.[20] Observe that under this parameterization, a choice of debt b_t is equivalent to a choice of saving rate s_t. Moreover, the (exogenous) bounds on feasible debt levels $\left[\underline{b}(b_{t-1}), \overline{b}(b_{t-1})\right]$ are replaced with bounds on saving rates $[\underline{s}, \overline{s}]$, where $\underline{s} > 0$ and $\overline{s} < 1$.

The representation in (16) has two main implications. The first implication is that welfare is separable in the inherited level of debt. This separability means that the continuation equilibria characterizing the highest and lowest feasible continuation values, $\overline{V}(b)$ and $\underline{V}(b)$, admit future sequences of saving rates that are independent of initial debt. Using (16), we can then show (see Appendix C) that these values are continuously differentiable and concave, and that they satisfy

$$\overline{V}(b) - \underline{V}(b) = P^* \tag{17}$$

for any initial debt b and some $P^* \geq 0$ that is independent of b. By (17), a government's flexible borrowing level given resources ω and type θ is $b^f(\theta) \equiv b^r(\omega, \theta) = b^p(\omega, \theta)$,

[19]Log preferences are used in previous work studying economies with hyperbolic discounting, including Barro (1999) and, in the context of fiscal rules, Halac and Yared (2014).

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

independently of the value of P^*. We let $g^f(\theta) \equiv b^f(\theta) + \omega$ denote the corresponding flexible spending level, where we omit the dependence on ω to reduce notation.

The second implication of the representation in (16) is that conditional on a finite future punishment $\overline{V}(b) - \underline{V}(b) = P^*$, the solutions to ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$) prescribe levels of debt with corresponding saving rates that are strictly between 0 and 1. This follows from the fact that, given log preferences, $\lim_{s\to 0} U(s) = \lim_{s\to 1} U(1-s) = -\infty$, and thus enforcement constraints cannot be satisfied for $s \in \{0, 1\}$ and finite P^*. Therefore, for any given finite P^*, there exist bounds $0 < \underline{s} < \overline{s} < 1$ such that the solutions to ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$) yield debt levels with corresponding saving rates in $(\underline{s}, \overline{s})$.

These properties of the value functions and of the solutions to ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$) imply that, if $\overline{V}(b) - \underline{V}(b)$ is strictly positive and finite for all finite $b < \tau/(R-1)$, and if the bounds $[\underline{s}, \overline{s}]$ are sufficiently wide, then Assumption 1 holds and the characterization in Proposition 1-Proposition 3 applies to this environment. In fact, since programs ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$) can be represented as independent of debt with a choice of saving rate $s(\theta)$ for each government type θ, in this case the characterization yields maximally enforced deficit and surplus limits with thresholds $\{\theta^*, \theta^{**}\}$ and $\{\theta^*_n, \theta^{**}_n\}$ that are also independent of initial debt. We are thus simply left to consider the conditions under which a solution with $\overline{V}(b) - \underline{V}(b) = P^*$ strictly positive and finite exists.

To facilitate the analysis, we will take our environment to have a distribution of shocks satisfying $f(\underline{\theta}) = f(\overline{\theta}) = 0$. This ensures that $Q(\theta)$ is continuous at $\underline{\theta}$ and $\overline{\theta}$ with $Q(\underline{\theta}) = 1$ and $Q(\overline{\theta}) = 0$.[21] We in turn obtain that under maximally enforced deficit and surplus limits as described in Proposition 2 and Proposition 3, social welfare is everywhere differentiable with respect to the thresholds $\{\theta^*, \theta^{**}, \theta^*_n, \theta^{**}_n\}$. Moreover, for $\alpha > 1$ and $\widehat{\theta}$ defined in Assumption 2, we obtain $\widehat{\theta} \in \big(\underline{\theta}, \overline{\theta}\big)$.

5.2 Conditions for Fiscal Policy Regimes

The results presented below provide conditions for fiscal policy regimes to arise in the best equilibrium and for regime transitions to occur along the equilibrium path. We prove these results by developing a factorization algorithm which we describe in the next subsection.

[21]Note that the analog of this assumption would hold over an unbounded support, since in that circumstance $\lim_{\theta\to 0} Q(\theta) = 1$ and $\lim_{\theta\to\infty} Q(\theta) = 0$.

Proposition 4. *Consider a setting with $U(\cdot) = \log(\cdot)$ and $f(\underline{\theta}) = f(\overline{\theta}) = 0$. There exist bounds $0 < \underline{s} < \overline{s} < 1$ on saving rates such that the best equilibrium is unique, prescribes $s_t \in (\underline{s}, \overline{s})$ at all on-path histories, and remains the unique best equilibrium for any strictly interior bounds wider than $[\underline{s}, \overline{s}]$. Moreover, there exist $\widetilde{\delta} \in (0,1)$ and $\widetilde{\alpha} \in (1, \infty)$ such that if $\delta > \widetilde{\delta}$, then $\overline{V}(\cdot) > \underline{V}(\cdot)$ in the equilibrium if and only if $\alpha > \widetilde{\alpha}$. If $\delta \leq \widetilde{\delta}$, then $\overline{V}(\cdot) = \underline{V}(\cdot)$ for all $\alpha \geq 1$.*

This proposition says that a pre-condition for fiscal regimes is that the discount factor $\delta \in (0,1)$ be high enough.[22] If δ is too low, the unique equilibrium is Markov, with $\overline{V}(\cdot) = \underline{V}(\cdot)$ and all government types $\theta \in \Theta$ choosing their flexible spending level $g^f(\theta)$ in all periods. Governments must be sufficiently patient for the dynamic incentives provided by future play to deter them from spending flexibly in the present.

Given a sufficiently high discount factor, $\delta > \widetilde{\delta}$, the more interesting part of Proposition 4 is that the existence of fiscal regimes also requires the governments' deficit bias α to be sufficiently large.[23] Observe that if $\alpha = 1$, the unique equilibrium has all governments choosing their flexible spending level (which in this case also corresponds to the first-best spending level) and therefore $\overline{V}(\cdot) = \underline{V}(\cdot)$. In other words, punishments are infeasible when $\alpha = 1$, since preferences are dynamically consistent across governments and future governments cannot credibly punish current ones. What Proposition 4 states is that for a small enough bias, $\alpha \in (1, \widetilde{\alpha}]$, it is also the case that $\overline{V}(\cdot) = \underline{V}(\cdot)$. Dynamic incentives can be provided if and only if $\alpha > \widetilde{\alpha}$.

The intuition for this result stems from the concavity of the value functions. For α close to 1, the highest continuation value $\overline{V}(\cdot)$ is close to its first-best level. By concavity, this means that a small difference in continuation values between two regimes would require a large difference in spending. However, governments are not willing to choose a spending level that is far from first best when their deficit bias is small. Hence, strong enough future punishments cannot be credibly imposed as to provide dynamic incentives, and the Markov equilibrium continues to be the unique one when $\alpha \in (1, \widetilde{\alpha}]$.

As α increases above $\widetilde{\alpha}$, two things happen. First, welfare moves away from first best, so concavity implies that a given difference in continuation values can be achieved with smaller differences in spending. Second, governments are more willing to spend above the first best level as they are more severely biased towards the present. These two effects imply that for α large enough, strong future punishments can be credibly imposed to deter governments from spending at their flexible level in the present.

[22]The proof of Proposition 4 provides an expression for the cutoff $\widetilde{\delta} \in (0,1)$.

[23]The proof of Proposition 4 shows that the cutoff $\widetilde{\alpha}$ is a decreasing function of δ. That is, the higher is $\delta > \widetilde{\delta}$, the larger is the range of biases α under which the equilibrium admits fiscal regimes.

We thus obtain that for $\delta > \widetilde{\delta}$ and $\alpha > \widetilde{\alpha}$, the best equilibrium for society is characterized by fiscally responsible and fiscally irresponsible regimes as described in Proposition 1-Proposition 3. Does the economy transition between the two regimes along the equilibrium path? The following corollary guarantees that the answer is yes for a range of values of the governments' deficit bias.[24]

Corollary 1. *Take the setting of Proposition 4 with $\delta > \widetilde{\delta}$. There exists $\widetilde{\widetilde{\alpha}} > \widetilde{\alpha}$ such that if $\alpha \in (\widetilde{\alpha}, \widetilde{\widetilde{\alpha}})$, then the best equilibrium features regime transitions on path.*

Given an equilibrium with regimes, recall from Subsection 4.3 that whether or not regime transitions occur on path depends on the tightness of the maximally enforced deficit limit that is implemented in the fiscally responsible regime. Transitions do not occur if the deficit limit is lax enough that governments respect it under all shocks. Instead, if the deficit limit is tighter, the economy (temporarily) transitions to the fiscally irresponsible regime when high enough shocks are realized. Corollary 1 says that we must be in the latter scenario if α is close to the cutoff $\widetilde{\alpha}$. Intuitively, in this case, the punishment $\overline{V}(\cdot) - \underline{V}(\cdot)$ that can be sustained in equilibrium is small, so the deficit limit would have to be very lax for governments to be willing to always respect it. Since $f(\overline{\theta}) = 0$, it is socially beneficial to tighten the deficit limit to improve fiscal discipline, and to let the economy transition to the fiscally irresponsible regime following high enough shocks which are unlikely.[25]

The results of this section highlight the role of governments' deficit bias. As noted in the Introduction, the political economy literature argues that political biases have increased over the last several decades, and that higher biases have resulted in rising debt levels across advanced economies (see the papers cited in fn. 2). Our results indicate that increased biases may not only lead to higher long-run debt growth, but also to the emergence of regimes in fiscal policy.

5.3 Factorization Algorithm

To prove Proposition 4, we develop a factorization algorithm. We consider a candidate interior equilibrium with fiscal regimes given by maximally enforced deficit and surplus limits as described in Proposition 2 and Proposition 3, parameterized by the thresholds $\{\theta^*, \theta^{**}, \theta_n^*, \theta_n^{**}\}$. This equilibrium can be represented by a system of equations. Specifically, we show in Appendix C that integrating conditions (13) and (15)

[24] Recall that by assuming $f(\overline{\theta}) = 0$, the environment that we consider in Proposition 4 and Corollary 1 takes $\widehat{\theta}$ defined in Assumption 2 to be interior. As pointed out in Subsection 4.3, the best equilibrium features no regime transitions if $\widehat{\theta} = \overline{\theta}$.

[25] See Proposition 4 in Halac and Yared (2022) for a related result.

and substituting with (17) yields

$$\delta P^* = \alpha \int_{\theta^*}^{\theta^{**}} \left[U(g^f(\theta)) - U(g^f(\theta^*))\right] d\theta, \tag{18}$$

$$\delta P^* = \alpha \int_{\theta_n^*}^{\theta_n^{**}} \left[U(g^f(\theta)) - U(g^f(\theta_n^*))\right] d\theta. \tag{19}$$

Moreover, writing the values $\overline{V}(b)$ and $\underline{V}(b)$ with the representation of welfare given in (11), computing the difference and again using (17), we obtain

$$P^* = \delta P^* + \alpha \left[\begin{array}{c} \int_{\theta^*}^{\theta^{**}} \big(U(g^f(\theta^*)) - U(g^f(\theta))\big) Q(\theta) d\theta \\ -\int_{\theta_n^{**}}^{\theta_n^*} \big(U(g^f(\theta_n^*)) - U(g^f(\theta))\big) Q(\theta) d\theta \end{array} \right]. \tag{20}$$

Equations (18) and (19) define the limited commitment constraints for a maximally enforced deficit limit in the fiscally responsible regime and for a maximally enforced surplus limit in the fiscally irresponsible regime, respectively. Equation (20) defines the value of punishment. Using this representation, consider the following program:

$$T(P) = \max_{\theta^*, \theta^{**}, \theta_n^*, \theta_n^{**}} \left\{ \delta P + \alpha \left[\begin{array}{c} \int_{\theta^*}^{\theta^{**}} \big(U(g^f(\theta^*)) - U(g^f(\theta))\big) Q(\theta) d\theta \\ -\int_{\theta_n^{**}}^{\theta_n^*} \big(U(g^f(\theta_n^*)) - U(g^f(\theta))\big) Q(\theta) d\theta \end{array} \right] \right\} \tag{21}$$

subject to

$$\delta P \geq \alpha \int_{\theta^*}^{\theta^{**}} \left[U(g^f(\theta)) - U(g^f(\theta^*))\right] d\theta \tag{22}$$

$$\delta P \geq \alpha \int_{\theta_n^{**}}^{\theta_n^*} \left[U(g^f(\theta_n^*)) - U(g^f(\theta))\right] d\theta. \tag{23}$$

Given a level of punishment P that can be inflicted on the government in the future, the program computes the largest punishment $T(P)$ that can be inflicted in the present. We show in Appendix C that there exists a solution that satisfies constraints (22)-(23) with equality. Moreover, we argue that the best equilibrium for society is unique and is characterized by the largest value of P^* that satisfies $T(P^*) = P^*$. This fixed point represents an equilibrium in which the largest punishment in the future supports the largest punishment in the present.

Figure 3 depicts the function $T(P)$. We prove that $T(P)$ is increasing and concave and satisfies $T(0) = 0$ and $\lim_{P\to\infty} T'(P) < 1$. The fact that $T(0) = 0$ is intuitive.

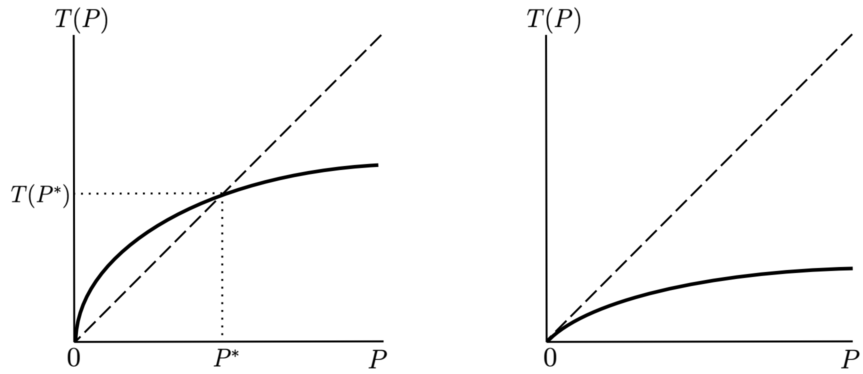

Figure 3: Representation of the function $T(P)$. The left panel depicts a scenario in which $T'(0) > 1$, and the right panel in which $T'(0) < 1$. The dashed line is the 45 degree line.

The algorithm always admits a fixed point at 0, corresponding to the Markov outcome, where $\overline{V}(\cdot) = \underline{V}(\cdot)$ is supported by governments choosing their flexible spending level $g^f(\theta)$ at all dates. If the largest punishment that is inflicted in the future is zero, then the largest punishment that is inflicted in the present is also zero.

The fact that $T(P)$ is concave with $\lim_{P\to\infty} T'(P) < 1$ means that there is at most one point $P^* > 0$ satisfying $T(P^*) = P^*$. The left panel of Figure 3 depicts a scenario in which $T'(0) > 1$ and thus such a fixed point exists. In this case, $\overline{V}(\cdot) > \underline{V}(\cdot)$, implying that Assumption 1 holds and the best equilibrium admits regimes as described in Proposition 1-Proposition 3. The right panel of Figure 3 depicts the other possible scenario, in which $T'(0) \leq 1$ and thus the unique fixed point is the Markov outcome.

$T(P)$ is analogous to the factorization algorithm introduced by Abreu, Pearce, and Stacchetti (1990), but for our problem of adverse selection. The analog of P in their work would be the set of continuation values for the players. In our environment, which features a single player in any given period, the set of continuation values is one-dimensional; hence, for our purposes, it is sufficient to consider the value of $\overline{V}(\cdot) - \underline{V}(\cdot) = P$. Another difference with Abreu, Pearce, and Stacchetti (1990) is how we apply the factorization algorithm. In their setting, one starts with the largest set of continuation values and the algorithm is applied repeatedly to obtain the largest fixed point. However, beyond computing a fixed point, we are interested in understanding whether there is a fixed point that features regimes, namely a non-Markov equilibrium. Given our characterization of $T(P)$, we are able to obtain a condition for such a fixed point by applying the algorithm from below. Starting from the Markov outcome with $P = 0$, we obtain a sufficient condition for the algorithm to converge to another fixed point when P is raised above 0, namely a condition for $T'(0) > 1$. We show that such a higher fixed point, if it exists, must be the unique best equilibrium.

5.4 Numerical Simulation

The factorization algorithm presented above facilitates computation. We next apply it to illustrate our results with a numerical simulation. We select parameter values under which the best equilibrium admits fiscal regimes as described in Proposition 1-Proposition 3. We compare the path for the spending rate in the best equilibrium to the paths for the first-best spending rate (i.e., assuming governments are not deficit-biased) and for the flexible spending rate (i.e., assuming a Markov equilibrium).

We take a governments' deficit bias $\alpha = 1.151$.[26] For the distribution of shocks, we use a log normal distribution with mean 0 and variance $\sigma = 0.175$, truncated to have support $[\underline{\theta}, \overline{\theta}] = [0.01, 100.01]$. Finally, we take a social discount factor $\delta = 0.943$. Given the value of α and a gross interest rate on government bonds $R = 1.05$, we choose the values for σ and δ so that the mean and variance of the flexible spending rate match the mean and variance of the spending rate for the U.S. over 1970-2020.[27]

Figure 4 depicts a simulated path for the spending rate in the best equilibrium along with the paths for the first-best and flexible spending rates. Shaded regions indicate periods of fiscal irresponsibility in the best equilibrium; non-shaded regions correspond to periods of fiscal responsibility. The two horizontal lines indicate the threshold spending rates in the two regimes. The bottom horizontal line is the spending rate corresponding to the maximally enforced deficit limit under fiscal responsibility (that is, the flexible spending rate of type θ^*). The top horizontal line is the spending rate corresponding to the maximally enforced surplus limit under fiscal irresponsibility (that is, the flexible spending rate of type θ_n^*).

Because governments are deficit-biased, we can see in Figure 4 that both the best-equilibrium spending rate and the flexible spending rate exceed the first-best spending rate at all dates. The comparison of the best-equilibrium and flexible rates illustrates how governments are provided incentives to reduce their overspending when the economy is in the fiscally responsible regime. Specifically, observe that in this regime, the best-equilibrium rate coincides with the flexible rate if the latter is below the deficit-limit threshold given by the bottom horizontal line in Figure 4. If instead the flexible rate exceeds this threshold, but not by much, then the government chooses to constrain its spending in order to avoid breaking the deficit limit and transitioning to the fiscally

[26] If we interpret $1/\alpha$ as a reelection probability in a setting with turnover (see fn. 10), then $\alpha = 1.151$ implies that on average the government is in power 7.6 years, which is in line with the average length of time the same party has held presidency in the U.S. from 1944 to 2020.

[27] We compute this U.S. spending rate, as defined in our model, assuming a tax revenue to GDP that is equal to the average level at all dates. We use data on federal debt, receipts, and outlays from the Federal Reserve Bank of St. Louis.

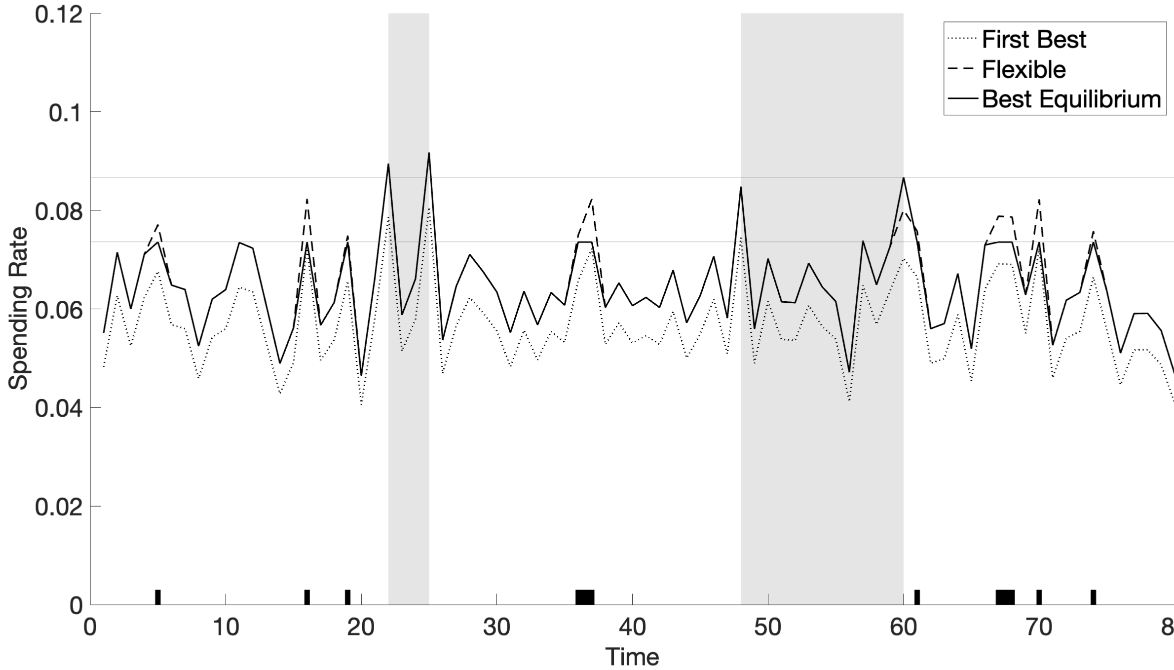

Figure 4: We take $U(\cdot) = \log(\cdot)$, $\alpha = 1.151$, $\delta = 0.943$, and $f(\theta)$ log normal with mean 0, variance $\sigma = 0.175$, and support $[\underline{\theta}, \overline{\theta}] = [0.01, 100.01]$. The bottom horizontal line depicts the flexible spending rate of type θ^*, given by 0.0736; the top horizontal line depicts the flexible spending rate of type θ_n^*, given by 0.0867. The shaded areas correspond to periods of fiscal irresponsibility.

irresponsible regime. This occurs in the figure at all dates marked with a black rectangle on the horizontal axis: the best-equilibrium rate at these dates is exactly at the deficit-limit threshold while the flexible rate exceeds the threshold. For high enough shocks, the flexible rate is sufficiently above the threshold that the government decides to break the deficit limit by spending at the flexible level. This occurs at the dates that shaded regions begin, as the economy then transitions to fiscal irresponsibility.

The comparison of the best-equilibrium and flexible rates also illustrates how governments are provided incentives to overspend when the economy is in the fiscally irresponsible regime. Observe that in this regime, the best-equilibrium rate and the flexible rate coincide if the latter is above the surplus-limit threshold given by the top horizontal line in Figure 4. If instead the flexible rate is below this threshold, but not by much, then the government chooses to increase its spending in order to respect the surplus limit and trigger a transition to the fiscally responsible regime. This occurs at the date that the second shaded region ends: the best-equilibrium rate at this date is exactly at the surplus-limit threshold while the flexible rate is below the threshold.

For low enough shocks, the flexible rate is sufficiently below the threshold that the government decides to break the surplus limit by spending at the flexible level. This occurs at all dates strictly interior to the shaded regions, as the economy then remains in the fiscally irresponsible regime.

6 Concluding Remarks

We have studied an equilibrium model of fiscal policy that generates persistent regimes. An important insight from our analysis is that the same deficit bias that can lead governments to overaccumulate debt is also a force that can lead the economy to fluctuate between periods of fiscal responsibility and irresponsibility, with transitions occurring during crises when fiscal needs are large. These fluctuations emerge in our setting not because of fluctuations in power across heterogeneous governments, but because of the dynamic strategic interaction between identical governments with the same bias. We find that transitions between fiscal regimes occur only if the governments' bias is sufficiently large: the threat of fiscal irresponsibility in the future is then severe enough to sustain fiscal responsibility in the present.

As discussed in the Introduction, the dynamics that we stress are consistent with patterns documented in the U.S., the European Union, and other regions. These dynamics are also consistent with, and have implications for, econometric analyses of fiscal policy. Cassou, Shadmani, and Vázquez (2017), Aldama and Creel (2019), and Elenev et al. (2021) find that U.S. fiscal policy exhibits history dependence and is characterized by two distinct regimes. Our results suggest that transitions between regimes should be tied to large negative economic shocks; this is in line with Cassou, Shadmani, and Vázquez (2017), which finds evidence of regime transitions being more likely following negative output gaps. Additionally, our results say that the coefficients determining the likelihood of regime transitions should be tied to the parameters driving debt growth—since both regimes and debt growth depend on governments' deficit bias in our model.

These implications for the path of fiscal variables can also be relevant for forecasting. For example, the current debate on the U.S. fiscal capacity has as a key ingredient the specification for the stochastic process for primary surpluses in the future.[28] Jiang et al. (2022) argue that the market value of U.S. government debt is not consistent with the primary surplus process if this process in the future remains the same as in

[28] This is evident in the following blog post by John Cochrane: `https://johnhcochrane.blogspot.com/2020/07/the-surplus-process.html`. See also Jiang et al. (2023).

the past; a change in regimes must occur instead. Our analysis provides a rationale for regime switching and can inform empirical work on this topic. We emphasize the need for future primary surpluses to be modeled as a function of not only contemporaneous economic variables but also the history of these economic variables, as well as political variables which determine governments' deficit bias and thus impact the value of the primary surplus. Moreover, our model suggests that changes in political variables can affect the process for primary surpluses not only directly but also indirectly by changing the regime-switching framework. These considerations can be useful to understand how the market value of U.S. government debt would respond to different economic and political shocks.

Relatedly, in future work, we believe it would be interesting to further investigate the quantitative implications of our model. For instance, one could study the degree to which the introduction of a deficit bias—which, as noted, affects both the long-term trend and the time-path of public debt—is helpful to fit the data relative to models without a deficit bias. This would require extending our setting in the direction of prior quantitative work on fiscal policy (e.g., Chari, Christiano, and Kehoe, 1991; Aiyagari, Marcet, Sargent, and Seppälä, 2002; Bhandari, Evans, Golosov, and Sargent, 2017; Debortoli, Nunes, and Yared, 2017), in particular using a richer structure of fiscal instruments, considering public spending shocks that are persistent,[29] and supplementing public spending shocks with other relevant types of shocks such as productivity shocks and discount rate shocks.

There are also potentially interesting directions for future research on the theory side. One possible direction is to study governments whose deficit bias applies to multiple periods. While in our formulation the preferences of the government regarding future policies coincide with those of society, a government's bias that extends to future periods would make the problem closer to one of repeated delegation. Another potential direction would be to consider political parties. Our model considers governments with the same bias and thus abstracts from the presence of political parties that have different preferences on policy. By taking parties into account, the analysis could provide a more nuanced interpretation of how successive governments are able to coordinate on regimes of fiscal responsibility and fiscal irresponsibility over time.

Finally, while we have focused on fiscal policy, the insights of this paper may be applied to other settings. For example, consider an individual who suffers from a self-

[29] Halac and Yared (2014) characterize the optimal mechanism under persistent shocks in a related model of fiscal policy with full enforcement. Introducing persistent shocks to our setting with limited, endogenous enforcement would add new challenges, as it would require a characterization of the worst as well as the best equilibrium.

control problem and wishes to curb his consumption of a temptation good, such as television or alcohol, while at the same time responding to consumption shocks over time. Our results suggest that the best self-enforcing consumption plan takes the form of a consumption threshold. The individual may violate the threshold when his value of consumption is high enough, and violation is punished by future selves with temporary over-consumption. Moreover, transitions in and out of periods of self-enforcing binging occur only if the individual's present bias is high enough.

References

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58, 1041–1063.

Aguiar, M. and M. Amador (2011): "Growth under the Shadow of Expropriation," *Quarterly Journal of Economics*, 126, 651–697.

Aiyagari, S. R., A. Marcet, T. J. Sargent, and J. Seppälä (2002): "Optimal Taxation without State-Contingent Debt," *Journal of Political Economy*, 110, 1220–1254.

Aldama, P. and J. Creel (2019): "Fiscal Policy in the US: Sustainable After All?" *Economic Modelling*, 81, 471–479.

Ales, L., P. Maziero, and P. Yared (2014): "A Theory of Political and Economic Cycles," *Journal of Economic Theory*, 153, 224–251.

Alesina, A., C. Favero, and F. Giavazzi (2019): *Austerity: When It Works and When It Doesn't*, Princeton: Princeton University Press.

Alesina, A. and A. Passalacqua (2016): "The Political Economy of Government Debt," in *Handbook of Macroeconomics*, North Holland, vol. 2, 2599–2651.

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

Amador, M. and C. Phelan (2021): "Reputation and Sovereign Default," *Econometrica*, 89, 1979–2010.

Amador, M., I. Werning, and G.-M. Angeletos (2006): "Commitment Vs. Flexibility," *Econometrica*, 74, 365–396.

Athey, S., A. Atkeson, and P. J. Kehoe (2005): "The Optimal Degree of Discretion in Monetary Policy," *Econometrica*, 73, 1431–1475.

Athey, S., K. Bagwell, and C. Sanchirico (2004): "Collusion and Price Rigidity," *Review of Economic Studies*, 71, 317–349.

Atkeson, A., V. V. Chari, and P. J. Kehoe (2007): "On the Optimal Choice of a Monetary Policy Instrument," NBER Working Paper.

Atkeson, A. and P. J. Kehoe (2001): "The Advantage of Transparent Instruments of Monetary Policy," NBER Working Paper.

Barro, R. J. (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, 201–236.

Bernheim, B. D., D. Ray, and S. Yeltekin (2015): "Poverty and Self-Control," *Econometrica*, 83, 1877–1911.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2017): "Fiscal Policy and Debt Management with Incomplete Markets," *Quarterly Journal of Economics*, 132, 617–663.

Bisin, A., A. Lizzeri, and L. Yariv (2015): "Government Policy with Time Inconsistent Voters," *American Economic Review*, 105, 1711–1737.

Calvo, G. A. and M. Obstfeld (1988): "Optimal Time-Consistent Fiscal Policy with Finite Lifetimes," *Econometrica*, 56, 411–432.

Cassou, S. P., H. Shadmani, and J. Vázquez (2017): "Fiscal Policy Asymmetries and the Sustainability of US Government Debt Revisited," *Empirical Economics*, 53, 1193–1215.

Chade, H., P. Prokopovych, and L. Smith (2008): "Repeated Games with Present-Biased Preferences," *Journal of Economic Theory*, 139, 157–175.

Chari, V. V., L. J. Christiano, and P. J. Kehoe (1991): "Optimal Fiscal and Monetary Policy: Some Recent Results," *Journal of Money, Credit and Banking*, 23, 519–539.

Debortoli, D., R. Nunes, and P. Yared (2017): "Optimal Time-Consistent Government Debt Maturity," *Quarterly Journal of Economics*, 132, 55–102.

Dornbusch, R. and S. Edwards, eds. (1991): *The Macroeconomics of Populism in Latin America*, University of Chicago Press.

Dovis, A. (2019): "Efficient Sovereign Default," *Review of Economic Studies*, 86, 282–312.

Dovis, A., M. Golosov, and A. Shourideh (2016): "Political Economy of Sovereign Debt: A Theory of Cycles of Populism and Austerity," NBER Working Paper.

Dovis, A. and R. Kirpalani (2021): "Rules without Commitment: Reputation and Incentives," *Review of Economic Studies*, 88, 2833–2856.

Drazen, A. (2000): "The Political Business Cycle after 25 Years," *NBER Macroeconomics Annual*, 15, 75–117.

Elenev, V., T. Landvoigt, S. V. Nieuwerburgh, and P. Shultz (2021): "Can Monetary Policy Create Fiscal Capacity?" NBER Working Paper.

Fudenberg, D. and J. Tirole (1991): *Game Theory*, Cambridge, Mass.: MIT Press.

Green, E. and R. Porter (1984): "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, 52, 87–100.

Halac, M. and P. Yared (2014): "Fiscal Rules and Discretion under Persistent Shocks," *Econometrica*, 82, 1557–1614.

——— (2020): "Inflation Targeting under Political Pressure," in *Independence, Credibility, and Communication of Central Banking*, ed. by E. Pasten and R. Reis, Central Bank of Chile.

——— (2021): "Instrument-Based vs. Target-Based Rules," *Review of Economic Studies*, 89, 312–345.

——— (2022): "Fiscal Rules and Discretion under Limited Enforcement," *Econometrica*, 90, 2093–2127.

Jiang, Z., H. Lustig, S. V. Nieuwerburgh, and M. Z. Xiaolan (2022): "Measuring U.S. Fiscal Capacity using Discounted Cash Flow Analysis," *Brookings Papers on Economic Activity*, Fall, 157–209.

——— (2023): "The U.S. Public Debt Valuation Puzzle," Working Paper.

Laibson, D. (1994): "Self-Control and Saving," Department of Economics, Harvard University.

——— (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–477.

Lizzeri, A. and L. Yariv (2017): "Collective Self-Control," *American Economic Journal: Microeconomics*, 9, 213–244.

Müller, A., K. Storesletten, and F. Zilibotti (2016): "The Political Color of Fiscal Responsibility," *Journal of the European Economic Association*, 14, 252–302.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

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.

Phelan, C. and E. Stacchetti (2003): "Sequential Equilibria in a Ramsey Tax Model," *Econometrica*, 69, 1491–1518.

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

Piguillem, F. and A. Schneider (2016): "Coordination, Efficiency and Policy Discretion," Working Paper.

Sachs, J. (1990): "Social Conflict and Populist Policies in Latin America," in *Labour Relations and Economic Performance. International Economic Association Series.*, ed. by R. Brunetta and C. Dell'Aringa, Palgrave Macmillan, London.

Sannikov, Y. (2007): "Games with Imperfectly Observable Actions in Continuous Time," *Econometrica*, 75, 1285–1329.

Sleet, C. (2004): "Optimal Taxation with Private Government Information," *Review of Economic Studies*, 71, 1217–1239.

Strotz, R. H. (1956): "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 23, 165–180.

Yared, P. (2010): "Politicians, Taxes and Debt," *Review of Economic Studies*, 77, 806–840.

——— (2019): "Rising Government Debt: Causes and Solutions for a Decades-Old Trend," *Journal of Economic Perspectives*, 33, 115–140.

A Preliminaries

In Subsection 4.1, we claimed that the private information constraint (7) can be replaced with a monotonicity constraint and local private information constraints. The lemma below provides a formal statement; for a proof, see Fudenberg and Tirole (1991).

Lemma 2. $(b(\theta), V(b(\theta)))$ *satisfies the private information constraint* (7) *if and only if* $b(\theta)$ *is nondecreasing and the following local private information constraints are satisfied:*

1. *At any point* θ *at which* $b(\cdot)$*, and thus* $V(\cdot)$*, are differentiable,*

$$\frac{db(\theta)}{d\theta}\left(\alpha\theta U'(\omega + b(\theta)) + \delta V'(b(\theta))\right) = 0.$$

2. *At any point* θ *at which* $b(\cdot)$ *is not differentiable,*

$$\lim_{\theta' \uparrow \theta} \left\{\alpha\theta U(\omega + b(\theta')) + \delta V(b(\theta'))\right\} = \lim_{\theta' \downarrow \theta} \left\{\alpha\theta U(\omega + b(\theta')) + \delta V(b(\theta'))\right\}.$$

B Proofs for Section 3 and Section 4

We introduce some terminology. When studying programs ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$), we will say that an allocation $(b(\theta), V(b(\theta)))$ is *incentive feasible* if it satisfies constraints (7)-(9), and it is *optimal* if it is a solution to the program.

B.1 Proof of Lemma 1

Take first the case in which $\alpha = 1$. Then trivially the best equilibrium has each government choosing the first-best policy, and thus social welfare is at its first-best level at each history.

Take next the case in which θ is observable. Suppose by contradiction that there is an on-path history in the best equilibrium at which $V(b(\theta)) < \overline{V}(b(\theta))$. Since increasing $V(b(\theta))$ relaxes the limited commitment constraint (8) (and since the private information constraint (7) can be ignored in this case), it follows that there is a perturbation that increases $V(b(\theta))$ that is feasible. Since such a perturbation increases the objective in ($\mathcal{P}_{\max}$), we reach a contradiction.

B.2 Proof of Proposition 1

Assume $Q'(\theta) \neq 0$ a.e. We prove the proposition by establishing Lemma 3 and Lemma 4.

Lemma 3. *If $(b(\theta), V(b(\theta)))$ is a solution to ($\mathcal{P}_{\max}$) with $b(\theta) \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ for all $\theta \in \Theta$, then $V(b(\theta)) \in \{\underline{V}(b(\theta)), \overline{V}(b(\theta))\}$ for all $\theta \in \Theta$.*

Proof. Take any solution $(b(\theta), V(b(\theta)))$ to ($\mathcal{P}_{\max}$) with $b(\theta) \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ for all $\theta \in \Theta$. We proceed in three steps.

Step 1. We show that $V(b(\theta))$ is left-continuous at each $\theta \in (\underline{\theta}, \overline{\theta}]$ and $V(b(\underline{\theta})) = \overline{V}(b(\underline{\theta}))$.

For the first claim, suppose by contradiction that there is $\theta \in (\underline{\theta}, \overline{\theta}]$ at which $V(b(\theta))$ is not left-continuous. Denote $(b(\theta^-), V(b(\theta^-))) \equiv \lim_{\theta' \uparrow \theta}(b(\theta'), V(b(\theta')))$. By Lemma 2,

$$0 < \alpha\theta \left(U(\omega + b(\theta)) - U(\omega + b(\theta^-))\right) = \delta \left(V(b(\theta^-)) - V(b(\theta))\right).$$

Given $\alpha > 1$, this implies

$$\theta \left(U(\omega + b(\theta)) - U(\omega + b(\theta^-))\right) < \delta \left(V(b(\theta^-)) - V(b(\theta))\right).$$

It follows that a perturbation that assigns $(b(\theta^-), V(b(\theta^-)))$ to type θ is incentive feasible, strictly increases social welfare from type θ, and does not affect social welfare from types other than θ. This contradicts the optimality of the original allocation, proving the claim.

For the second claim, suppose by contradiction that $V(b(\underline{\theta})) < \overline{V}(b(\underline{\theta}))$. We perform a perturbation where we change $b(\underline{\theta}) \in (\underline{b}, \overline{b})$ by $-db(\underline{\theta}) < 0$ arbitrarily close to zero and change $V(b(\underline{\theta}))$ so as to keep type $\underline{\theta}$ equally well off:

$$db(\underline{\theta})\alpha\underline{\theta}U'(\omega + b(\underline{\theta})) + \delta dV(b(\underline{\theta})) = 0.$$

This perturbation is incentive feasible and does not affect social welfare from types $\theta \in (\underline{\theta}, \overline{\theta}]$. The change in social welfare from type $\underline{\theta}$ is equal to

$$-\left[db(\underline{\theta})\underline{\theta}U'(\omega + b(\underline{\theta})) + \delta dV(b(\underline{\theta}))\right] = db(\underline{\theta})\underline{\theta}U'(\omega + b(\underline{\theta}))\,(\alpha - 1) > 0.$$

This contradicts the optimality of the original allocation, proving the claim.

Step 2. We show that $V(b(\theta))$ is a step function over any interval $[\theta^L, \theta^H]$ with $V(b(\theta)) \in (\underline{V}(b(\theta)), \overline{V}(b(\theta)))$ for $\theta \in [\theta^L, \theta^H]$.

By the private information constraints, $V(b(\theta))$ is piecewise continuously differentiable and nonincreasing. Suppose by contradiction that there is an interval $[\theta^L, \theta^H]$ over which $V(b(\theta))$ is continuously strictly decreasing in θ with $0 < V(b(\theta)) < \overline{V}(b(\theta))$. By Lemma 2, $b(\theta)$ must be continuously strictly increasing over the interval, and without loss we can take an interval over which $b(\theta)$ is continuously differentiable. Moreover, by the generic property that $Q'(\theta) \neq 0$ a.e., we can take an interval with either $Q'(\theta) > 0$ or $Q'(\theta) < 0$ for all $\theta \in [\theta^L, \theta^H]$. We consider each possibility in turn.

Case 1: Suppose $Q'(\theta) < 0$ for all $\theta \in [\theta^L, \theta^H]$. We show that there is an incentive feasible flattening perturbation that rotates the increasing borrowing schedule $b(\theta)$ clockwise over $[\theta^L, \theta^H]$ and strictly increases social welfare. Define

$$\overline{U} = \frac{1}{\left(\theta^H - \theta^L\right)} \int_{\theta^L}^{\theta^H} U(\omega + b(\theta))d\theta.$$

For given $\kappa \in [0, 1]$, let $\widetilde{b}(\theta, \kappa)$ be the solution to

$$U(\omega + \widetilde{b}(\theta, \kappa)) = \kappa\overline{U} + (1 - \kappa)U(\omega + b(\theta)), \tag{24}$$

which clearly exists. Define $\widetilde{V}(\widetilde{b}(\theta))$ as the solution to

$$\begin{aligned}\alpha\theta U(\omega+&\widetilde{b}(\theta, \kappa)) + \delta\widetilde{V}(\widetilde{b}(\theta, \kappa)) \\ &= \alpha\theta^L U(\omega + b(\theta^L)) + \delta V(b(\theta^L)) + \int_{\theta^L}^{\theta} \alpha U(\omega + \widetilde{b}(\widetilde{\theta}, \kappa))d\widetilde{\theta}.\end{aligned} \tag{25}$$

The original allocation corresponds to $\kappa = 0$. We consider a perturbation where we increase κ marginally above zero if and only if $\theta \in [\theta^L, \theta^H]$. Note that differentiating

(24) and (25) with respect to κ yields

$$\frac{d\widetilde{b}(\theta,\kappa)}{d\kappa} = \frac{\overline{U} - U(\omega + b(\theta))}{U'(\omega + \widetilde{b}(\theta,\kappa))}, \tag{26}$$

$$\frac{d\widetilde{b}(\theta,\kappa)}{d\kappa}(\alpha\theta U'(\omega + \widetilde{b}(\theta,\kappa)) + \delta\widetilde{V}'(\widetilde{b}(\theta,\kappa))) = \int_{\theta^L}^{\theta} \frac{d\widetilde{b}(\widetilde{\theta},\kappa)}{d\kappa}\alpha U'(\omega + \widetilde{b}(\widetilde{\theta},\kappa))d\widetilde{\theta}. \tag{27}$$

Substituting (26) in (27) yields that for a type $\theta \in [\theta^L, \theta^H]$, the change in government welfare from a marginal increase in κ, starting from $\kappa = 0$, is equal to

$$D(\theta) \equiv \int_{\theta^L}^{\theta} \alpha \left(\overline{U} - U(\omega + b(\widetilde{\theta}))\right) d\widetilde{\theta}.$$

We begin by showing that the perturbation satisfies constraints (7)-(9). For (7), note that $D(\theta^L) = D(\theta^H) = 0$, so the perturbation leaves the government welfare of types θ^L and θ^H (and that of types $\theta < \theta^L$ and $\theta > \theta^H$) unchanged. Using Lemma 2 and the representation in (10), it then follows from (25) and the fact that $\widetilde{b}(\theta,\kappa)$ is nondecreasing that the perturbation satisfies (7) for all $\theta \in \Theta$ and any $\kappa \in [0, 1]$.

To prove that the perturbation satisfies (8), we show that the government welfare of types $\theta \in [\theta^L, \theta^H]$ weakly rises when κ increases marginally. Since $D(\theta^L) = D(\theta^H) = 0$, it is sufficient to show that $D(\theta)$ is concave over (θ^L, θ^H) to prove that $D(\theta) \geq 0$ for all θ in this interval. Indeed, we can verify that $D''(\theta) = -\alpha U'(\omega + b(\theta))\frac{db(\theta)}{d\theta} < 0$.

Lastly, observe that (9) is satisfied for $\kappa > 0$ small enough. This follows from $\underline{V}(b(\theta))$ being continuous and from $V(b(\theta)) \in (\underline{V}(b(\theta)), \overline{V}(b(\theta)))$ for $\theta \in [\theta^L, \theta^H]$ in the original allocation.

We next show that the perturbation strictly increases social welfare. Using the representation in (11), the change in social welfare from an increase in κ is equal to

$$\alpha \int_{\theta^L}^{\theta^H} \frac{d\widetilde{b}(\theta,\kappa)}{d\kappa} U'(\omega + \widetilde{b}(\theta,\kappa))Q(\theta)d\theta.$$

Substituting with (26) yields that at $\kappa = 0$, this is equal to

$$\alpha \int_{\theta^L}^{\theta^H} \left(\overline{U} - U\left(\omega + b(\theta)\right)\right) Q(\theta)d\theta.$$

This is an integral over the product of two terms. The first term is strictly decreasing in θ since $b(\theta)$ is strictly increasing over $[\theta^L, \theta^H]$. The second term is also strictly decreasing in θ; this follows from $Q'(\theta) < 0$ for all $\theta \in [\theta^L, \theta^H]$. Therefore, the two terms are positively correlated with one another, and thus the change in social welfare is strictly greater than

$$\alpha \int_{\theta^L}^{\theta^H} \left(\overline{U} - U(\omega + b(\theta))\right) d\theta \int_{\theta^L}^{\theta^H} Q(\theta) d\theta,$$

which is equal to 0. Hence, we obtain that if $V(b(\theta))$ is strictly interior and $Q'(\theta) < 0$ over a given interval, then $V(b(\theta))$ must be a step function over the interval.

<u>Case 2:</u> Suppose $Q'(\theta) > 0$ for all $\theta \in [\theta^L, \theta^H]$. Recall that $b(\theta)$ is continuously strictly increasing over $[\theta^L, \theta^H]$. We begin by showing that the limited commitment constraint cannot bind for all $\theta \in [\theta^L, \theta^H]$. Suppose by contradiction that it does. Using the representation of government welfare in (10), this implies

$$\int_{\theta}^{\theta^H} \alpha(U(\omega + b^p(\omega, \widetilde{\theta})) - U(\omega + b(\widetilde{\theta})))d\widetilde{\theta} = 0$$

for all $\theta \in [\theta^L, \theta^H]$, which requires $(b(\theta), V(b(\theta))) = (b^p(\omega, \theta), \underline{V}(b^p(\omega, \theta)))$ for all $\theta \in (\theta^L, \theta^H)$. However, this contradicts the assumption that $V(b(\theta)) \in (\underline{V}(b(\theta)), \overline{V}(b(\theta)))$ for all $\theta \in [\theta^L, \theta^H]$. Hence, the limited commitment constraint cannot bind for all types in the interval, and without loss we can take an interval with this constraint being slack for all $\theta \in [\theta^L, \theta^H]$.

We next show that there is a steepening perturbation that is incentive feasible and strictly increases social welfare. Consider drilling a hole around a type θ^M within $[\theta^L, \theta^H]$ so that we marginally remove the allocation around this type. That is, θ^M can no longer choose $(b(\theta^M), V(b(\theta^M)))$ and is indifferent between jumping to the lower or upper limit of the hole. With some abuse of notation, denote the limits of the hole by θ^L and θ^H, where the perturbation marginally increases θ^H from θ^M. Since the limited commitment constraint is slack for all $\theta \in [\theta^L, \theta^H]$, the perturbation is incentive feasible. The change in social welfare is

$$\alpha \int_{\theta^M}^{\theta^H} \frac{db(\theta^H)}{d\theta^H} U'(\omega + b(\theta^H)) Q(\theta) d\theta + \alpha \frac{d\theta^M}{d\theta^H} \left(U(\omega + b(\theta^L)) - U(\omega + b(\theta^H))\right) Q(\theta^M). \tag{28}$$

By indifference of type θ^M,

$$\alpha\theta^M U(\omega + b(\theta^L)) + \delta V(b(\theta^L)) = \alpha\theta^M U(\omega + b(\theta^H)) + \delta V(b(\theta^H)).$$

Differentiating this indifference condition with respect to θ^H yields

$$\frac{d\theta^M}{d\theta^H} = \frac{db(\theta^H)}{d\theta^H} U'(\omega + b(\theta^H)) \frac{\alpha(\theta^H - \theta^M)}{U(\omega + b(\theta^H)) - U(\omega + b(\theta^L))},$$

where we have used the private information constraint $\alpha \frac{db(\theta^H)}{d\theta^H}(\theta^H U'(\omega + b(\theta^H)) + \delta V'(b(\theta^H))) = 0$. Substituting back into (28), the change in social welfare is

$$\alpha \frac{db(\theta^H)}{d\theta^H} U'(\omega + b(\theta^H)) \int_{\theta^M}^{\theta^H} (Q(\theta) - Q(\theta^M)) d\theta.$$

Since $\frac{db(\theta^H)}{d\theta^H} > 0$, $U'(\omega + b(\theta^H)) > 0$, and $Q'(\theta) > 0$, this expression is strictly positive. Hence, we obtain that if $V(b(\theta))$ is strictly interior and $Q'(\theta) > 0$ over a given interval, then $V(b(\theta))$ must be a step function over the interval.

Step 3. We show that $V(b(\theta)) \in \{\underline{V}(b(\theta)), \overline{V}(b(\theta))\}$ for all $\theta \in \Theta$.

Suppose by contradiction that $V(b(\theta)) \in (\underline{V}(b(\theta)), \overline{V}(b(\theta)))$ for some $\theta \in \Theta$. By the previous steps and Lemma 2, type θ belongs to a stand-alone segment $(\theta^L, \theta^H]$, such that $b(\theta) = b$ and $V(b(\theta)) = V$ for all $\theta \in (\theta^L, \theta^H]$, $b \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ and $V \in (\underline{V}(b), \overline{V}(b))$ (by assumption), $b(\theta)$ jumps at θ^L, and $b(\theta)$ jumps at θ^H unless $\theta^H = \overline{\theta}$.

We first show that the limited commitment constraint must be slack for all $\theta \in (\theta^L, \theta^H)$. Express this constraint as the difference between the left-hand and right-hand sides of (8), so that it must be weakly positive and it equals zero if it binds. By the private information constraints, the derivative of the limited commitment constraint with respect to θ is $\alpha U(\omega + b(\theta)) - \alpha U(\omega + b^p(\omega, \theta))$. Since $b(\theta)$ is constant over $(\theta^L, \theta^H]$ and $b^p(\omega, \theta)$ is nondecreasing, it follows that the limited commitment constraint is weakly concave over the interval. Then, if the constraint binds at any interior point $\theta' \in (\theta^L, \theta^H)$, it must bind at all $\theta \in (\theta^L, \theta^H)$. However, by the arguments used in Case 2 in Step 2 above, that would require $b = b^p(\omega, \theta)$ and $V = \underline{V}(b)$ for $\theta \in (\theta^L, \theta^H)$, contradicting the assumption that V is strictly interior.

We next show that there is an incentive feasible perturbation that strictly increases social welfare. We consider segment-shifting perturbations that marginally change the constant borrowing level b and continuation value V. There are two cases:

<u>Case 1:</u> Suppose $\int_{\theta^L}^{\theta^H} Q(\theta^L) d\theta < \int_{\theta^L}^{\theta^H} Q(\theta) d\theta$. Consider a perturbation that marginally changes the borrowing level by $db > 0$ and changes V in order to keep type θ^H equally well off. For arbitrarily small $db > 0$, this perturbation makes the lowest types in

$(\theta^L, \theta^H]$, arbitrarily close to θ^L, jump either to the allocation of type θ^L or to their flexible allocation under maximal punishment $(b^p(\omega,\theta), \underline{V}(b^p(\omega,\theta)))$, where we let the perturbation introduce the latter. In the limit as db goes to zero, the change in social welfare is[30]

$$\alpha \int_{\theta^L}^{\theta^H} U'(\omega+b)Q(\theta)d\theta + \alpha\frac{d\theta^L}{db}(U(\omega+b(\theta^L)) - U(\omega+b))Q(\theta^L). \tag{29}$$

The perturbation satisfies

$$db\ \alpha\theta^H U'(\omega+b) + \delta dV = 0, \tag{30}$$

and the following indifference condition for type θ^L:

$$\alpha\theta^L U(\omega+b) + \delta V(b) = \alpha\theta^L U(\omega+b(\theta^L)) + \delta V(b(\theta^L)).$$

To verify that the perturbation is incentive feasible for db arbitrarily close to zero, note that the limited commitment constraint is slack for all $\theta \in (\theta^L, \theta^H)$, V is strictly interior, and the government welfare of types θ^L and θ^H remains unchanged with the perturbation.

To verify that the perturbation strictly increases social welfare, note that differentiating the indifference condition of type θ^L and substituting with (30) yields

$$\frac{d\theta^L}{db} = -U'(\omega+b)\frac{\left(\theta^H - \theta^L\right)}{U(\omega+b(\theta^L)) - U(\omega+b)}.$$

Substituting back into (29), the change in social welfare is

$$\alpha U'(\omega+b)\int_{\theta^L}^{\theta^H}(Q(\theta) - Q(\theta^L))d\theta.$$

Since $U'(\omega+b) > 0$ and by assumption $\int_{\theta^L}^{\theta^H} Q(\theta^L)d\theta < \int_{\theta^L}^{\theta^H} Q(\theta)d\theta$, the above expression is strictly positive. The claim follows.

Case 2: Suppose $\int_{\theta^L}^{\theta^H} Q(\theta^L)d\theta \geq \int_{\theta^L}^{\theta^H} Q(\theta)d\theta$. By the generic property in Proposition 1, there exists $\theta^h \in (\theta^L, \theta^H]$ such that $\int_{\theta^L}^{\theta^h} Q(\theta^L)d\theta > \int_{\theta^L}^{\theta^h} Q(\theta)d\theta$. Then consider a perturbation where, for $\theta \in (\theta^L, \theta^h]$, we change the borrowing level by $-db < 0$ arbitrarily close to zero and change V in order to keep type θ^h equally well off. This perturbation

[30] The arguments that follow are unchanged if $(b(\theta^L), V(b(\theta^L)))$ is replaced with $(b^p(\omega,\theta^L), \underline{V}(b^p(\omega,\theta^L)))$ for the cases where the limited commitment constraint binds.

makes types arbitrarily close to θ^L jump up to the allocation of the stand-alone segment. Arguments analogous to those in Case 1 above imply that the perturbation is incentive feasible. Moreover, following analogous steps as in that case yields that the implied change in social welfare is

$$-\alpha U'(\omega + b) \int_{\theta^L}^{\theta^h} (Q(\theta) - Q(\theta^L)) d\theta.$$

Since $U'(\omega + b) > 0$ and by assumption $\int_{\theta^L}^{\theta^h} Q(\theta^L) d\theta > \int_{\theta^L}^{\theta^h} Q(\theta) d\theta$, the above expression is strictly positive. The claim follows. □

Lemma 4. *If $(b(\theta), V(b(\theta)))$ is a solution to $(\mathcal{P}_{\min})$ with $b(\theta) \in (\underline{b}(b_{-1}), \bar{b}(b_{-1}))$ for all $\theta \in \Theta$, then $V(b(\theta)) \in \{\underline{V}(b(\theta)), \overline{V}(b(\theta))\}$ for all $\theta \in \Theta$.*

Proof. The proof of this lemma is analogous to the proof of Lemma 3. We therefore describe this proof only briefly here, focusing on the steps that are different.

Take any solution $(b(\theta), V(b(\theta)))$ to $(\mathcal{P}_{\min})$ with $b(\theta) \in (\underline{b}(b_{-1}), \bar{b}(b_{-1}))$ for all $\theta \in \Theta$. By analogous arguments to those in the first part of Step 1 in the proof of Lemma 3, we can establish that $V(b(\theta))$ must be right-continuous at each $\theta \in [\underline{\theta}, \overline{\theta})$. Moreover, by arguments analogous to those in the second part of Step 1, we can establish that $V(b(\overline{\theta})) = \overline{V}(b(\overline{\theta}))$. Specifically, if $V(b(\overline{\theta})) \in (\underline{V}(b(\overline{\theta})), \overline{V}(b(\overline{\theta})))$, then a perturbation that marginally increases $b(\overline{\theta}) \in (\underline{b}, \bar{b})$ and changes $V(b(\overline{\theta}))$ so as to keep type $\overline{\theta}$'s welfare unchanged is incentive feasible and strictly reduces social welfare. Such a perturbation is also incentive feasible (and welfare reducing) if $V(b(\overline{\theta})) = \underline{V}(b(\overline{\theta}))$, as in this case $b(\overline{\theta}) = b^p(\omega, \overline{\theta})$ by the limited commitment constraint (8) and thus the perturbation requires setting $V(b(\overline{\theta})) > \underline{V}(b(\overline{\theta}))$. It follows that $V(b(\overline{\theta})) = \overline{V}(b(\overline{\theta}))$.

The claims in Step 2 and Step 3 in the proof of Lemma 3 also apply when solving program $(\mathcal{P}_{\min})$. The reason is that perturbations that apply whenever $Q'(\theta) > 0$ in the maximization of social welfare now apply whenever $Q'(\theta) < 0$ in the minimization of social welfare, and vice versa. Hence, the arguments in these steps, together with those in Step 1 just described, imply that $V(b(\theta)) \in \{\underline{V}(b(\theta)), \overline{V}(b(\theta))\}$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$ in any solution to $(\mathcal{P}_{\min})$ with $b(\theta) \in (\underline{b}(b_{-1}), \bar{b}(b_{-1}))$ for all $\theta \in \Theta$. □

B.3 Proof of Proposition 2

We first prove Lemma 5 and Lemma 6 and then proceed with the proof of the proposition.

Lemma 5. *If $(b(\theta), V(b(\theta)))$ is a maximally enforced deficit limit, it satisfies the private information constraint* (7) *and the limited commitment constraint* (8)*.*

Proof. We proceed in three steps.

Step 1. Suppose $\theta^* \geq \underline{\theta}$. We show that (7) and (8) are satisfied for $\theta \in [\underline{\theta}, \theta^*]$.

The claim follows immediately from the fact that all types $\theta \in [\underline{\theta}, \theta^*]$ are assigned their flexible debt levels with the highest continuation value. Thus, given $\theta \in [\underline{\theta}, \theta^*]$, type θ's welfare cannot be increased, so (7) and (8) are trivially satisfied.

Step 2. We show that (7) and (8) are satisfied for $\theta \in (\theta^*, \theta^{**}]$.

Take first the limited commitment constraint (8). We can rewrite it for $\theta \in (\theta^*, \theta^{**}]$ as

$$\alpha\theta U(\omega + b^r(\omega, \theta^*)) + \delta\overline{V}(b^r(\omega, \theta^*)) - \alpha\theta U(\omega + b^p(\omega, \theta)) - \delta\underline{V}(b^p(\omega, \theta)) \geq 0. \quad (31)$$

Differentiating the left-hand side with respect to θ, given θ^* and the definition of $b^p(\omega, \theta)$, yields

$$\alpha U(\omega + b^r(\omega, \theta^*)) - \alpha U(\omega + b^p(\omega, \theta)),$$

which is weakly decreasing in θ, since $b^p(\omega, \theta)$ is nondecreasing. This means that the left-hand side of (31) is weakly concave. Since (31) holds as a strict inequality for $\theta = \theta^*$ and as an equality for $\theta = \theta^{**}$ (by (13)), this weak concavity implies that (31) holds as a strict inequality for all $\theta \in (\theta^*, \theta^{**})$. Thus, constraint (8) is satisfied for all $\theta \in (\theta^*, \theta^{**}]$.

Take next the private information constraint (7). This constraint is trivially satisfied for all $\theta \in (\theta^*, \theta^{**}]$ given $\theta' \in [\theta^*, \theta^{**}]$, since all types $\theta \in [\theta^*, \theta^{**}]$ are prescribed the same level of debt and continuation value. We next show that the constraint is also satisfied given $\theta' > \theta^{**}$ and $\theta' < \theta^*$:

Step 2a: We show that (7) is satisfied for all $\theta \in (\theta^*, \theta^{**}]$ given $\theta' > \theta^{**}$. Note that $(b(\theta'), V(b(\theta'))) = (b^p(\omega, \theta'), \underline{V}(b^p(\omega, \theta')))$ for all $\theta' > \theta^{**}$, and by the definition of $b^p(\omega, \theta)$,

$$\alpha\theta U(\omega + b^p(\omega, \theta)) + \delta\underline{V}(b^p(\omega, \theta)) \geq \alpha\theta U(\omega + b^p(\omega, \theta')) + \delta\underline{V}(b^p(\omega, \theta'))$$

for all $\theta' \in \Theta$. Thus, the fact that the limited commitment constraint (8) is satisfied for all $\theta \in (\theta^*, \theta^{**}]$ implies that (7) is satisfied for all such types given $\theta' > \theta^{**}$.

Step 2b: We show that (7) is satisfied for all $\theta \in (\theta^*, \theta^{**}]$ given $\theta' < \theta^*$. Suppose by contradiction that this is not the case, that is,

$$\alpha\theta(U(\omega + b^r(\omega, \theta^*)) - U(\omega + b^r(\omega, \theta'))) < \delta\left(\overline{V}(b^r(\omega, \theta')) - \overline{V}(b^r(\omega, \theta^*))\right) \quad (32)$$

for some $\theta \in (\theta^*, \theta^{**}]$ and $\theta' < \theta^*$. By Step 1, (7) holds for type θ^* given $\theta' < \theta^*$:

$$\alpha\theta^*(U(\omega + b^r(\omega, \theta^*)) - U(\omega + b^r(\omega, \theta'))) \geq \delta\left(\overline{V}(b^r(\omega, \theta')) - \overline{V}(b^r(\omega, \theta^*))\right). \quad (33)$$

Combining (32) and (33) yields

$$\alpha(\theta^* - \theta)(U(\omega + b^r(\omega, \theta^*)) - U(\omega + b^r(\omega, \theta'))) > 0,$$

which is a contradiction since $\theta > \theta^*$ and $b^r(\omega, \theta') \leq b^r(\omega, \theta^*)$. The claim follows.

Step 3. Suppose $\theta^{**} < \overline{\theta}$. We show that (7) and (8) are satisfied for $\theta \in (\theta^{**}, \overline{\theta}]$.

Constraint (8) is satisfied as an equality for all $\theta \in (\theta^{**}, \overline{\theta}]$. It is immediate that constraint (7) is satisfied for all $\theta \in (\theta^{**}, \overline{\theta}]$ given $\theta' \in (\theta^{**}, \overline{\theta}]$, since all such types are prescribed their flexible debt level with the lowest continuation value. Consider next constraint (7) for $\theta \in (\theta^{**}, \overline{\theta}]$ given $\theta' \in [\theta^*, \theta^{**}]$. Note that $(b(\theta'), V(b(\theta'))) = (b^r(\omega, \theta^*), \overline{V}(b^r(\omega, \theta^*)))$ for all $\theta' \in [\theta^*, \theta^{**}]$. Thus, satisfaction of this constraint is ensured if (31) is violated for $\theta \in (\theta^{**}, \overline{\theta}]$. The latter is true since, as shown above, the left-hand side of (31) is weakly concave and (31) holds as an equality for $\theta = \theta^{**}$ and a strict inequality for $\theta \in (\theta^*, \theta^{**})$.

Finally, consider constraint (7) for $\theta \in (\theta^{**}, \overline{\theta}]$ given $\theta' < \theta^*$. Since (7) is satisfied given $\theta' \in [\theta^*, \theta^{**}]$, satisfaction of this constraint given $\theta' < \theta^*$ is ensured if

$$\alpha\theta(U(\omega + b^r(\omega, \theta^*)) - U(\omega + b^r(\omega, \theta'))) \geq \delta(\overline{V}(b^r(\omega, \theta')) - \overline{V}(b^r(\omega, \theta^*)))$$

for $\theta \in (\theta^{**}, \overline{\theta}]$. The latter follows from the same logic as in Step 2b above. □

Lemma 6. *If $(b(\theta), V(b(\theta)))$ is a solution to $(\mathcal{P}_{\max})$ with $b(\theta) \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ for all $\theta \in \Theta$, then either $V(b(\theta)) = \overline{V}(b(\theta))$ for all $\theta \in \Theta$, or there exists $\theta^{**} \in (\underline{\theta}, \overline{\theta})$ such that $V(b(\theta)) = \overline{V}(b(\theta))$ for all $\theta \in [\underline{\theta}, \theta^{**}]$ and $V(b(\theta)) = \underline{V}(b(\theta))$ for all $\theta \in (\theta^{**}, \overline{\theta}]$.*

Proof. Take any solution $(b(\theta), V(b(\theta)))$ to $(\mathcal{P}_{\max})$ with $b(\theta) \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ for all $\theta \in \Theta$. We proceed in three steps.

Step 1. We show that if $V(b(\theta^{**})) = \underline{V}(b(\theta^{**}))$, then $\theta^{**} \geq \widehat{\theta}$.

By Lemma 3 and Step 1 in the proof of that lemma, if $V(b(\theta^{**})) = \underline{V}(b(\theta^{**}))$ for some $\theta^{**} \in \Theta$, then $V(b(\theta)) = \underline{V}(b(\theta))$ over an interval $(\theta^L, \theta^H]$ that contains θ^{**}. Take the largest such interval. We establish that $\theta^L \geq \widehat{\theta}$.

Suppose by contradiction that $\theta^L < \widehat{\theta}$. Note that constraint (8) requires $b(\theta) = b^p(\omega, \theta)$ for all $\theta \in (\theta^L, \theta^H]$, and the strict concavity of $\underline{V}(\cdot)$ implies that $b^p(\omega, \theta)$ is strictly increasing over a subset of $(\theta^L, \theta^H]$ below $\widehat{\theta}$. Without loss, take such a subset with $b^p(\omega, \theta)$ being continuously differentiable. Then we can perform a flattening perturbation that rotates the borrowing schedule clockwise over this subset, analogous to the perturbation used in Step 2 (Case 1) in the proof of Lemma 3. By the arguments in that step, this perturbation is incentive feasible. In particular, note that since the perturbation weakly increases the government welfare of all types θ in the subset while simultaneously changing their borrowing allocation, it follows from the definition of $b^p(\omega, \theta)$ that the perturbation must necessarily increase $V(b(\theta))$ above $\underline{V}(b(\theta))$. Furthermore, by $Q'(\theta) < 0$ for all types θ in the subset (by the subset being below $\widehat{\theta}$ and Assumption 2), the perturbation strictly increases social welfare, yielding a contradiction.

Step 2. We show that if $V(b(\theta^{**})) = \underline{V}(b(\theta^{**}))$, then $V(b(\theta)) = \underline{V}(b(\theta))$ for all $\theta \geq \theta^{**}$.

Suppose by contradiction that $V(b(\theta^{**})) = \underline{V}(b(\theta^{**}))$ for $\theta^{**} \in \Theta$ and $V(b(\theta)) > \underline{V}(b(\theta))$ for some $\theta > \theta^{**}$. By Step 1, $\theta^{**} \geq \widehat{\theta}$. Moreover, by Lemma 3 and Step 1 in the proof of that lemma, there exist $\theta^H > \theta^L \geq \theta^{**}$ such that $V(b(\theta)) = \overline{V}(b(\theta))$ for all $\theta \in (\theta^L, \theta^H]$.

We begin by establishing that $b(\theta) = b$ for all $\theta \in (\theta^L, \theta^H]$ and some $b \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$. Suppose by contradiction that $b(\theta)$ is strictly increasing at some point $\theta' \in (\theta^L, \theta^H]$. Note that the private information constraint (7) implies $b(\theta) = b^r(\omega, \theta)$, and thus a slack limited commitment constraint (8), in the neighborhood of such a type θ'. Then we can perform an incentive feasible steepening perturbation that drills a hole in the $b(\theta)$ schedule in this neighborhood, as that described in Step 2 (Case 2) in the proof of Lemma 3. By the arguments in that step, this perturbation strictly increases social welfare, yielding a contradiction.

We next show that a segment $(\theta^L, \theta^H]$ with $b(\theta) = b$ and $V(b(\theta)) = \overline{V}(b)$ for all $\theta \in (\theta^L, \theta^H]$ and $\theta^L \geq \theta^{**}$ cannot exist. Suppose by contradiction that it does. Take θ^L to be the lowest point weakly above θ^{**} at which $V(b(\theta))$ jumps, and take θ^H to be the lowest point above θ^L at which $V(b(\theta))$ jumps again, or $\theta^H = \overline{\theta}$ if $V(b(\theta))$ does not jump above θ^L. Then $(\theta^L, \theta^H]$ is a stand-alone segment with constant borrowing b and continuation value $V = \overline{V}(b)$. By arguments analogous to those in Step 3 of the proof of Lemma 3, the limited commitment constraint must be slack for all $\theta \in (\theta^L, \theta^H)$. We then show that there is an incentive feasible segment-shifting perturbation that is socially beneficial. There are three cases to consider:

Case 1: Suppose $\alpha\theta^H U(\omega + b) + \delta\overline{V}(b) \leq \alpha\theta^H U(\omega + b') + \delta\overline{V}(b')$ for $b' = b + \varepsilon$, $\varepsilon > 0$ arbitrarily small. Then we perform a segment-shifting perturbation as that in Step 3 (Case 1) in the proof of Lemma 3, where we marginally increase b and reduce $V(b)$ marginally below $\overline{V}(b)$ so as to keep type θ^H's welfare unchanged. This perturbation is incentive feasible. Moreover, since $\theta^L \geq \theta^{**}$ and Assumption 2 imply $\int_{\theta^L}^{\theta^H} Q(\theta^L)d\theta < \int_{\theta^L}^{\theta^H} Q(\theta)d\theta$, this perturbation strictly increases social welfare, yielding a contradiction.

Case 2: Suppose $\alpha\theta^H U(\omega + b) + \delta\overline{V}(b) > \alpha\theta^H U(\omega + b') + \delta\overline{V}(b')$ for $b' = b + \varepsilon$, $\varepsilon > 0$ arbitrarily small, and $\theta^H < \overline{\theta}$. Then we perform a segment-shifting perturbation that marginally changes the borrowing level by $-db < 0$ and reduces V marginally below $\overline{V}(b)$ so as to keep type θ^L's welfare unchanged. This perturbation is incentive feasible. Denote by $(b(\theta^h), V(b(\theta^h)))$ the allocation above θ^H over which type θ^H is initially indifferent. Note that analogous to the perturbation in Step 3 in the proof of Lemma 3, this perturbation makes the highest types in $(\theta^L, \theta^H]$, arbitrarily close to θ^H, jump either to $(b(\theta^h), V(b(\theta^h)))$ or to their flexible allocation under the maximal punishment $(b^p(\omega,\theta), \underline{V}(b^p(\omega,\theta)))$, where we let the perturbation introduce the latter. In the limit as db goes to zero, the change in social welfare is[31]

$$-\alpha \int_{\theta^L}^{\theta^H} U'(\omega + b)Q(\theta)d\theta + \alpha\frac{d\theta^H}{db}(U(\omega + b(\theta^h)) - U(\omega + b))Q(\theta^H). \tag{34}$$

The perturbation satisfies

$$db\ \alpha\theta^L U'(\omega + b) + \delta dV = 0, \tag{35}$$

and the following indifference condition for type θ^H:

$$\alpha\theta^H U(\omega + b) + \delta V(b) = \alpha\theta^H U(\omega + b(\theta^h)) + \delta V(b(\theta^h)).$$

[31] The arguments that follow are unchanged if $(b(\theta^h), V(b(\theta^h)))$ is replaced with $(b^p(\omega,\theta^H), \underline{V}(b^p(\omega,\theta^H)))$ for the cases where the limited commitment constraint binds.

Differentiating this indifference condition and substituting with (35) yields

$$\frac{d\theta^H}{db} = U'(\omega + b)\frac{(\theta^H - \theta^L)}{U(\omega + b(\theta^h)) - U(\omega + b)}.$$

Substituting back into (34), the change in social welfare is

$$-\alpha U'(\omega + b)\int_{\theta^L}^{\theta^H}(Q(\theta) - Q(\theta^H))d\theta.$$

Since $U'(\omega + b) > 0$ and $Q'(\theta) > 0$ over $[\theta^L, \theta^H]$ given $\theta^L \geq \theta^{**}$, this expression is strictly positive. Thus, the perturbation strictly increases social welfare, yielding a contradiction.

Case 3: Suppose $\alpha\theta^H U(\omega + b) + \delta\overline{V}(b) > \alpha\theta^H U(\omega + b') + \delta\overline{V}(b')$ for $b' = b + \varepsilon$, $\varepsilon > 0$ arbitrarily small, and $\theta^H = \overline{\theta}$. Then we perform a segment-shifting perturbation as that in Case 2 above, where we marginally reduce b and decrease V marginally below $\overline{V}(b)$ so as to keep type θ^L's welfare unchanged. This perturbation is incentive feasible. Note that analogous to Case 2, this perturbation makes the highest types in $(\theta^L, \theta^H]$, arbitrarily close to θ^H, either jump to their flexible allocation under maximal punishment $(b^p(\omega, \theta), \underline{V}(b^p(\omega, \theta)))$ or remain with the perturbed allocation. In the former case, the same arguments as in Case 2 apply, yielding that the perturbation strictly increases social welfare by $-\alpha U'(\omega + b)\int_{\theta^L}^{\theta^H}(Q(\theta) - Q(\theta^H)) > 0$. In the latter case, those arguments imply that the change in social welfare is equal to $-\alpha U'(\omega + b)\int_{\theta^L}^{\theta^H} Q(\theta)d\theta$, which is strictly positive since $Q(\overline{\theta}) \leq 0$ and $Q'(\theta) > 0$ over $[\theta^L, \theta^H]$. Hence, the perturbation strictly increases social welfare, yielding a contradiction.

Step 3. We show that $V(b(\theta))$ is right-continuous at $\underline{\theta}$.

Suppose by contradiction that this is not the case. Then by the previous steps, Lemma 3, and Step 1 in the proof of that lemma, $V(b(\theta)) = \underline{V}(b(\theta))$ for all $\theta \in (\underline{\theta}, \overline{\theta}]$ and $V(b(\theta))$ jumps down at $\underline{\theta}$ from $\overline{V}(b(\theta))$. Note that constraint (8) implies $b(\theta) = b^p(\omega, \theta)$ for all $\theta \in (\underline{\theta}, \overline{\theta}]$, and indifference of $\underline{\theta}$ requires

$$\alpha\underline{\theta}U(\omega + b(\underline{\theta})) + \delta\overline{V}(b(\underline{\theta})) = \lim_{\theta\downarrow\underline{\theta}}\{\alpha\underline{\theta}U(\omega + b^p(\omega, \theta)) + \delta\underline{V}(b^p(\omega, \theta))\}.$$

Take $\Delta \in (0, \min_{\theta\in\Theta}\{\overline{V}(b(\theta)) - \underline{V}(b(\theta))\})$. Consider a global perturbation that assigns $V(b(\theta)) = \underline{V}(b(\theta)) + \Delta$ to all $\theta \in (\underline{\theta}, \overline{\theta}]$ and assigns type $\underline{\theta}$ the limit allocation to its right. This perturbation keeps borrowing unchanged for types $\theta \in (\underline{\theta}, \overline{\theta}]$ and is incentive feasible. Moreover, using the representation in (11), the change in social welfare from this perturbation is equal to $\delta\Delta$. Thus, the perturbation strictly increases social welfare, yielding a contradiction. □

Proof of Proposition 2. We now proceed to prove the proposition. Take any solution $(b(\theta), V(b(\theta)))$ to $(\mathcal{P}_{\max})$ with $b(\theta) \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ for all $\theta \in \Theta$. By Lemma 6 and the limited commitment constraint (8), there exists $\theta^{**} > \underline{\theta}$ such that $(b(\theta), V(b(\theta))) = (b^p(\omega, \theta), \underline{V}(b^p(\omega, \theta)))$ for all $\theta > \theta^{**}$ and $V(b(\theta)) = \overline{V}(b(\theta))$ for all $\theta \leq \theta^{**}$ (where it is possible that $\theta^{**} > \overline{\theta}$). Moreover, since the limited commitment constraint holds with equality at θ^{**}, this type's allocation satisfies

$$\alpha\theta^{**}U(\omega + b(\theta^{**})) + \delta\overline{V}(b(\theta^{**})) = \alpha\theta^{**}U(\omega + b^p(\omega, \theta^{**})) + \delta\underline{V}(b^p(\omega, \theta^{**})). \tag{36}$$

These results characterize the allocation for types $\theta \geq \theta^{**}$. To characterize the allocation for types $\theta < \theta^{**}$, we proceed in three steps.

Step 1. We show that $b(\theta)$ is continuous over $[\underline{\theta}, \theta^{**}]$.

By Step 1 in the proof of Lemma 6, $\theta^{**} \geq \widehat{\theta}$. There are two cases to consider:

Case 1: Suppose by contradiction that $b(\theta)$ has a point of discontinuity below $\widehat{\theta}$: there is a type $\theta^M < \widehat{\theta}$ which is indifferent between choosing $\lim_{\theta\uparrow\theta^M} b(\theta)$ and $\lim_{\theta\downarrow\theta^M} b(\theta) > \lim_{\theta\uparrow\theta^M} b(\theta)$. Note that given $V(b(\theta)) = \overline{V}(b(\theta))$ for all $\theta \in [\underline{\theta}, \theta^{**}]$ and $\theta^{**} \geq \widehat{\theta}$, there must be a hole with types $\theta \in [\theta^L, \theta^M)$ bunched at $b^r(\omega, \theta^L)$ and types $\theta \in (\theta^M, \theta^H]$ bunched at $b^r(\omega, \theta^H)$, for some $\theta^L < \theta^M < \theta^H$. Now consider perturbing the allocation by marginally increasing θ^L, in an effort to slightly close the hole. This perturbation leaves the government welfare of types strictly above θ^M unchanged and is incentive feasible. The change in social welfare is[32]

$$\alpha \int_{\theta^L}^{\theta^M} \frac{db^r(\omega, \theta^L)}{d\theta^L} U'(\omega + b^r(\omega, \theta^L))Q(\theta)d\theta + \alpha\frac{d\theta^M}{d\theta^L}\left(U(\omega + b^r(\omega, \theta^L)) - U(\omega + b^r(\omega, \theta^H))\right)Q(\theta^M). \tag{37}$$

By indifference of type θ^M,

$$\alpha\theta^M U(\omega + b^r(\omega, \theta^L)) + \delta V(b^r(\omega, \theta^L)) = \alpha\theta^M U(\omega + b^r(\omega, \theta^H)) + \delta V(b^r(\omega, \theta^H)).$$

[32] Note that this welfare representation is valid even if $\theta^L < \underline{\theta}$, as we can apply the envelope condition in (10) from any positive $\theta' < \underline{\theta}$. For $\theta < \underline{\theta}$, we have $Q(\theta) = 1$.

Differentiating this indifference condition with respect to θ^L yields

$$\frac{d\theta^M}{d\theta^L} = \frac{db^r(\omega,\theta^L)}{d\theta^L} U'(\omega + b^r(\omega,\theta^L)) \frac{(\theta^M - \theta^L)}{U(\omega + b^r(\omega,\theta^H)) - U(\omega + b^r(\omega,\theta^L))},$$

where we have used the fact that $\alpha\theta^L U'(\omega+b^r(\omega,\theta^L)) = -\delta\overline{V}'(b^r(\omega,\theta^L))$. Substituting back into (37), the change in social welfare is

$$\alpha\frac{db^r(\omega,\theta^L)}{d\theta^L} U'(\omega + b^r(\omega,\theta^L)) \int_{\theta^L}^{\theta^M} (Q(\theta) - Q(\theta^M))d\theta.$$

Since $\frac{db^r(\omega,\theta^L)}{d\theta^L} > 0$, $U'(\omega + b^r(\omega,\theta^L)) > 0$, and $Q'(\theta) < 0$ over $\theta \in [\theta^L, \theta^M]$ given $\theta^M < \widehat{\theta}$, this expression is strictly positive. Thus, the perturbation strictly increases social welfare, showing that $b(\theta)$ cannot jump at a point below $\widehat{\theta}$.

Case 2: Suppose by contradiction that $b(\theta)$ is discontinuous at a point $\theta \in [\widehat{\theta}, \theta^{**}]$. Note that since $V(b(\theta)) = \overline{V}(b(\theta))$ for all $\theta \in [\widehat{\theta}, \theta^{**}]$, we can apply the same logic as in Step 2 in the proof of Lemma 6 to show that $\frac{db(\theta)}{d\theta} = 0$ over any continuous interval in $[\widehat{\theta}, \theta^{**}]$. Hence, if $b(\theta)$ jumps at a point $\theta \in [\widehat{\theta}, \theta^{**}]$, then there exists a stand-alone segment $(\theta^L, \theta^H]$ with constant borrowing $b \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ and continuation value $V = \overline{V}(b)$, satisfying $\theta^L \geq \widehat{\theta}$. However, using again the arguments in Step 2 in the proof of Lemma 6, we can then perform an incentive feasible segment-shifting perturbation that strictly increases social welfare. Thus, $b(\theta)$ cannot jump at a point $\theta \in [\widehat{\theta}, \theta^{**}]$.

Step 2. We show that $b(\theta) \leq b^r(\omega, \theta)$ for all $\theta \in [\underline{\theta}, \theta^{**}]$.

By Step 1 above, the allocation over $[\underline{\theta}, \theta^{**}]$ must take the form of bounded discretion, with either a minimum borrowing level or a maximum borrowing level or both. We next show that a binding minimum borrowing requirement is strictly suboptimal. Suppose by contradiction that this is not the case, namely there exist $\theta^\star > \underline{\theta}$ and an optimal allocation prescribing $(b(\theta), V(b(\theta))) = (b^r(\omega,\theta^\star), \overline{V}(b^r(\omega,\theta^\star)))$ for all $\theta \in [\underline{\theta}, \theta^\star]$, where $b^r(\omega,\theta) < b^r(\omega,\theta^\star)$ for all $\theta \in [\underline{\theta}, \theta^\star)$. Consider a perturbation where we remove this minimum borrowing requirement, that is, we set $(b(\theta), V(b(\theta))) = (b^r(\omega,\theta), \overline{V}(b^r(\omega,\theta)))$ for all $\theta \in [\underline{\theta}, \theta^\star)$. Clearly, this perturbation is incentive feasible, and it keeps the allocation of types $\theta \in [\theta^\star, \overline{\theta}]$, and thus the social welfare from these types, unchanged. The change in social welfare from each type $\theta \in [\underline{\theta}, \theta^\star)$ is

$$\theta U(\omega + b^r(\omega,\theta)) + \delta\overline{V}(b^r(\omega,\theta)) - \theta U(\omega + b^r(\omega,\theta^\star)) - \delta\overline{V}(b^r(\omega,\theta^\star)).$$

Note that by the definition of $b^r(\omega,\theta)$,

$$\delta\overline{V}(b^r(\omega,\theta)) - \delta\overline{V}(b^r(\omega,\theta^\star)) \geq \alpha\left(\theta U(\omega + b^r(\omega,\theta^\star)) - \theta U(\omega + b^r(\omega,\theta))\right).$$

Substituting back into the previous expression, we obtain that the change in social welfare from each $\theta \in [\underline{\theta}, \theta^\star)$ is greater than

$$(\alpha - 1)\left(\theta U(\omega + b^r(\omega,\theta^\star)) - \theta U(\omega + b^r(\omega,\theta))\right),$$

which is strictly positive. Thus, the perturbation strictly increases social welfare, implying that a binding minimum borrowing requirement is strictly suboptimal.

Step 3. We show that $b(\theta) < b^r(\omega,\theta)$ for some $\theta \in \Theta$.

By Step 1 and Step 2, the allocation for types $\theta \in [\underline{\theta}, \theta^{**}]$ is as described in Proposition 2 for some $\theta^* \geq 0$. That is, equation (36) necessarily holds for $b(\theta^{**}) = b^r(\omega,\theta^*)$. All that remains to be shown is that $\theta^* < \overline{\theta}$. Suppose by contradiction that this is not true, which implies $(b(\theta), V(b(\theta))) = (b^r(\omega,\theta), \overline{V}(b^r(\omega,\theta)))$ for all $\theta \in \Theta$. Consider an incentive feasible perturbation that assigns $(b(\theta), V(b(\theta))) = (b^r(\omega,\overline{\theta}-\varepsilon), \overline{V}(b^r(\omega,\overline{\theta}-\varepsilon)))$ to all $\theta \in [\overline{\theta}-\varepsilon, \overline{\theta}]$, where $\varepsilon > 0$ is chosen to be small enough as to continue to satisfy the limited commitment constraint (8) for all types $\theta \in \Theta$. Using the representation in (11), the change in social welfare is

$$\alpha \int_{\overline{\theta}-\varepsilon}^{\overline{\theta}} (U(\omega + b^r(\omega,\overline{\theta}-\varepsilon)) - U(\omega + b^r(\omega,\theta)))Q(\theta)d\theta.$$

For $\varepsilon > 0$ arbitrarily small, $b^r(\omega,\overline{\theta}-\varepsilon) < b^r(\omega,\theta)$ and $Q(\theta) < 0$ for all $\theta \in (\overline{\theta}-\varepsilon, \overline{\theta})$. Thus, the perturbation strictly increases social welfare, proving the claim.

B.4 Proof of Proposition 3

The proof of this proposition is analogous to the proof of Proposition 2. We therefore describe this proof only briefly here, focusing on the steps that are different.

Analogous arguments to those in Lemma 5 imply that if $(b(\theta), V(b(\theta)))$ is a maximally enforced surplus limit, then it satisfies the private information and limited commitment constraints in program $(\mathcal{P}_{\text{min}})$. Consider next the proof of Lemma 6. Step 1 and Step 2 in that proof can be applied isomorphically to $(\mathcal{P}_{\text{min}})$ in the sense that the arguments applying to types $\theta < \widehat{\theta}$ in the maximization of social welfare now apply to types $\theta > \widehat{\theta}$ in the minimization of social welfare, and vice versa. Combined with the claims above, these steps thus imply the following: in any solution $(b(\theta), V(b(\theta)))$ to $(\mathcal{P}_{\min})$ with $b(\theta) \in (\underline{b}(b_{-1}), \overline{b}(b_{-1}))$ for all $\theta \in \Theta$, there exists $\theta_n^{**} \leq \overline{\theta}$ such that $V(b(\theta)) = \underline{V}(b(\theta))$ for all $\theta \in [\underline{\theta}, \theta_n^{**})$ and $V(b(\theta)) = \overline{V}(b(\theta))$ for all $\theta \in [\theta_n^{**}, \overline{\theta}]$.

The analog of Step 3 in the proof of Lemma 6 consists of showing that $\theta_n^{**} < \overline{\theta}$. To see why this must be true, suppose by contradiction that $\theta_n^{**} = \overline{\theta}$, namely that $V(b(\theta)) = \underline{V}(b(\theta))$ for all $\theta \in [\underline{\theta}, \overline{\theta})$ and $V(b(\theta))$ jumps at $\overline{\theta}$ to $\overline{V}(b(\overline{\theta}))$. Note that the limited commitment constraint (8) implies $b(\theta) = b^p(\omega, \theta)$ for all $\theta \in [\underline{\theta}, \overline{\theta})$, and using the representation in (11), social welfare is equal to

$$\alpha \underline{\theta} U(\omega + b^p(\omega, \underline{\theta})) + \delta \underline{V}(b^p(\omega, \underline{\theta})) + \alpha \int_{\underline{\theta}}^{\overline{\theta}} U(\omega + b^p(\omega, \theta)) Q(\theta) d\theta. \tag{38}$$

Consider a global perturbation in which all types $\theta \in \Theta$ are assigned the allocation corresponding to a maximally enforced surplus limit $\{\theta_n^*, \theta_n^{**}\}$, with $\theta_n^{**} \in (\underline{\theta}, \overline{\theta})$, $Q(\theta_n^{**}) < 0$, and $\theta_n^* \geq \overline{\theta}$ (and with equation (15) being satisfied). Note that this is feasible since $Q(\overline{\theta}) < 0$ and $Q(\cdot)$ is continuous. Using the representation in (38) and taking into account that the perturbation keeps the allocation of types $\theta \in [\underline{\theta}, \theta_n^{**})$ unchanged, we find that the change in social welfare from the perturbation is equal to

$$\alpha \int_{\theta_n^{**}}^{\overline{\theta}} (U(\omega + b^r(\omega, \theta_n^*)) - U(\omega + b^p(\omega, \theta)) Q(\theta) d\theta. \tag{39}$$

Note that $b^r(\omega, \theta_n^*) > b^p(\omega, \theta)$ and $Q(\theta) < 0$ for all $\theta \in [\theta_n^{**}, \overline{\theta}]$ (by construction, Assumption 2, and the surplus limit satisfying the private information and limited commitment constraints). Hence, the perturbation strictly reduces social welfare, implying that $\theta_n^{**} < \overline{\theta}$ must hold in any solution to $(\mathcal{P}_{\min})$.

Given the claims above, the next step to prove Proposition 3 is to show that $b(\theta)$ is continuous for $\theta \geq \theta_n^{**}$. Here analogous arguments to those in the proof of Proposition 2 apply. The optimality of a surplus limit that is binding (i.e., with $\theta_n^* > \underline{\theta}$) also follows from analogous arguments as in that proof. Finally, note that the optimal surplus limit must satisfy $\theta_n^{**} \geq \underline{\theta}$: otherwise, if $\theta_n^{**} < \underline{\theta}$, then a perturbation that tightens the limit by raising θ_n^{**} to $\underline{\theta}$ (and raising θ_n^* so as to satisfy the indifference condition in (15)) is incentive feasible and strictly reduces social welfare.

C Proofs for Section 5

C.1 Preliminaries

We begin by establishing the properties discussed in Subsection 5.1. Take $U(\cdot) = \log(\cdot)$. We consider a representation of equilibrium using saving rates, as defined in the text. Note that for any period $t \in \{0, 1, \ldots\}$ and history of debt $h^{t-1} = \{b_{-1}, b_0, \ldots, b_{t-1}\}$, there is a corresponding history of initial debt and subsequent saving rates, $\widetilde{h}^{t-1} = \{b_{-1}, s_0, \ldots, s_{t-1}\}$. Moreover, a strategy for the government in period t can be equivalently defined as specifying either a debt level $b_t(h^{t-1}, \theta_t)$ for each history h^{t-1} and government type θ_t, or a saving rate $s_t(\widetilde{h}^{t-1}, \theta_t)$ for each history $\widetilde{h}^{t-1}$ and government type θ_t. It is thus without loss to redefine strategies and payoffs to condition on $\widetilde{h}^{t-1}$, with $V_t(\widetilde{h}^{t-1})$ denoting the continuation value at $\widetilde{h}^{t-1}$.

Observe that (16) implies that the continuation value at any given history is separable in the inherited level of debt. As a consequence, the private information and limited commitment constraints (3) and (6) are independent of initial debt, implying that whether or not a profile of saving rate strategies constitutes an equilibrium is also independent of initial debt. Let $\widetilde{V}_t(\widetilde{h}^{t-1})$ denote the continuation value normalized by the level of debt starting from a history $\widetilde{h}^{t-1}$:

$$\widetilde{V}_t(\widetilde{h}^{t-1}) = V_t(\widetilde{h}^{t-1}) - \frac{\mathbb{E}_t[\theta_t]}{1-\delta} \log\left(\frac{R\tau}{R-1} - Rb_{t-1}(\widetilde{h}^{t-1})\right).$$

We obtain that the highest and lowest normalized continuation values at time 0—namely the highest and lowest values of $\widetilde{V}_0(b_{-1})$—are independent of the initial debt b_{-1}. We can thus represent these values by $\widetilde{\overline{V}}$ and $\widetilde{\underline{V}}$, and, using the definition of normalized welfare, we have

$$\overline{V}(b) = \widetilde{\overline{V}} + \frac{\mathbb{E}_t[\theta_t]}{1-\delta} \log\left(\frac{R\tau}{R-1} - Rb\right), \tag{40}$$

$$\underline{V}(b) = \widetilde{\underline{V}} + \frac{\mathbb{E}_t[\theta_t]}{1-\delta} \log\left(\frac{R\tau}{R-1} - Rb\right). \tag{41}$$

It follows that $\overline{V}(\cdot)$ and $\underline{V}(\cdot)$ are continuously differentiable, strictly decreasing, and strictly concave. Moreover, given the constants $\widetilde{\overline{V}} \geq \widetilde{\underline{V}}$, we have

$$\overline{V}(b) - \underline{V}(b) = \widetilde{\overline{V}} - \widetilde{\underline{V}} \equiv P^*, \tag{42}$$

where $P^* \geq 0$ is independent of b, and is finite given that the set of feasible policies is closed and payoffs are thus bounded.

Lastly, observe that by the arguments above, programs ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$) characterizing the values $\overline{V}(b_{-1})$ and $\underline{V}(b_{-1})$ can be represented using saving rates $s(\theta)$ and normalized continuation values $\widetilde{V}(s(\theta))$, where the lowest and highest such values, $\underline{\widetilde{V}}$ and $\overline{\widetilde{V}}$, are independent of initial debt. It follows that if $\overline{\widetilde{V}} > \underline{\widetilde{V}}$, then the interior solutions to ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$) are given by thresholds $\{\theta^*, \theta^{**}\}$ and $\{\theta_n^*, \theta_n^{**}\}$ that are also independent of initial debt.

C.2 Proof of Proposition 4

Note that by (40)-(41) and the definitions of $b^r(\omega,\theta)$ and $b^p(\omega,\theta)$, we have $b^r(\omega,\theta) = b^p(\omega,\theta)$ for all ω and θ, independent of the value of $P^* = \overline{V}(b_{-1}) - \underline{V}(b_{-1})$. Let $b^f(\theta) \equiv b^r(\omega,\theta) = b^p(\omega,\theta)$ and $g^f(\theta) \equiv b^f(\theta) + \omega$, where we omit the dependence on ω to reduce notation. We proceed in four steps.

Step 1. Suppose there is an interior equilibrium with fiscal regimes characterized by maximally enforced deficit and surplus limits as described in Proposition 2 and Proposition 3, with cutoffs $\{\theta^*, \theta^{**}, \theta_n^*, \theta_n^{**}\}$ and value of punishment $P^* = \overline{V}(b_{-1}) - \underline{V}(b_{-1})$. We show that this equilibrium satisfies the system of equations (18)-(20).

To obtain (18), take condition (13). Integrating its left- and right-hand sides, we can rewrite this condition as

$$\alpha\theta^* U(g^f(\theta^*)) + \delta\overline{V}(b^f(\theta^*)) + \alpha\int_{\theta^*}^{\theta^{**}} U(g^f(\theta^*))d\theta = \alpha\theta^* U(g^f(\theta^*)) + \delta\underline{V}(b^f(\theta^*)) + \alpha\int_{\theta^*}^{\theta^{**}} U(g^f(\theta))d\theta.$$

By (42), this equality simplifies to (18). Analogous steps yield that condition (15) can be rewritten as (19). Finally, to obtain (20), we can use the representation of welfare in (10) to write

$$\begin{aligned}
\overline{V}(b) &= \lim_{\theta'\to 0}\left[\alpha\theta' U(g^f(\theta')) + \delta\overline{V}(b^f(\theta'))\right] + \int_0^{\theta^*} \alpha U(g^f(\theta))Q(\theta)d\theta \\
&\quad + \int_{\theta^*}^{\theta^{**}} \alpha U(g^f(\theta^*))Q(\theta)d\theta + \int_{\theta^{**}}^{\overline{\theta}} \alpha U(g^f(\theta))Q(\theta)d\theta, \\
\underline{V}(b) &= \lim_{\theta'\to 0}\left[\alpha\theta' U\left(g^f(\theta')\right) + \delta\underline{V}(b^f(\theta'))\right] + \int_0^{\theta_n^{**}} \alpha U(g^f(\theta))Q(\theta)d\theta \\
&\quad + \int_{\theta_n^{**}}^{\theta_n^*} \alpha U(g^f(\theta_n^*))Q(\theta)d\theta + \int_{\theta_n^*}^{\overline{\theta}} \alpha U(g^f(\theta))Q(\theta)d\theta,
\end{aligned}$$

where $Q(\theta) = 1$ for $\theta < \underline{\theta}$ and $Q(\theta) = 0$ for $\theta > \overline{\theta}$. Subtracting the bottom equation from the top one and again using (42) yields (20).

Step 2. Consider the program given by (21)-(23). Let P^* be the largest value of P that admits $T(P) = P$ and suppose $P^* > 0$. We show that $P^* \geq \overline{V}(b_{-1}) - \underline{V}(b_{-1})$. Moreover, there exists a sufficiently large feasible set $[\underline{b}(b_{-1}), \overline{b}(b_{-1})]$ such that $\{\overline{V}(b_{-1}), \underline{V}(b_{-1})\}$ are supported by interior allocations and $P^* = \overline{V}(b_{-1}) - \underline{V}(b_{-1})$.

Step 2a. Consider a solution to (21)-(23) for some $P > 0$. We show that there exists such a solution that admits (22) and (23) with equality.

Suppose first that (22) holds as a strict inequality in the solution. The derivative of (21) with respect to θ^{**} takes the same sign as $-Q(\theta^{**})$. By Assumption 2 and $f(\overline{\theta}) = 0$, we have $-Q(\theta)$ strictly negative for low θ and strictly positive for high θ given $\theta < \overline{\theta}$, and we have $Q(\theta) = 0$ given $\theta \geq \overline{\theta}$. Thus, the solution admits either $\theta^* = \theta^{**}$ or $\theta^{**} \geq \overline{\theta}$.

We show that the solution cannot have $\theta^* = \theta^{**}$. Suppose $\theta^* = \theta^{**}$ and consider first the case that $Q(\theta^{**}) < 0$. We can perform a perturbation that increases θ^{**} until either constraint (22) holds as an equality or $\theta^{**} = \overline{\theta}$. This perturbation increases the value of $\int_{\theta^*}^{\theta^{**}} \left[U(g^f(\theta^*)) - U(g^f(\theta))\right] Q(\theta)d\theta$ in the objective while satisfying all constraints, thus yielding a contradiction. Take next the case that $Q(\theta^{**}) \geq 0$. We can first perform a perturbation that changes θ^{**} and θ^* by the same amount $\Delta \gtrless 0$, which does not affect the objective nor the constraints (since $\theta^{**} = \theta^*$), and then we can perform the same perturbation as above starting from the new values. By choosing Δ such that $Q(\theta^{**} + \Delta) < 0$, we obtain again that the perturbation increases the value of the objective while satisfying all constraints, thus yielding a contradiction.

It follows from the above claims that the solution must have $\theta^* < \theta^{**}$ and $\theta^{**} \geq \overline{\theta}$. Then the derivative of (21) with respect to θ^* implies

$$\int_{\theta^*}^{\theta^{**}} Q(\theta)d\theta = \int_{\theta^*}^{\overline{\theta}} Q(\theta)d\theta = 0,$$

which yields a unique interior value of θ^* given Assumption 2. Since the optimal value of θ^* is independent of θ^{**} and $\theta^{**} \geq \overline{\theta}$, the objective in (21) is invariant to increases in θ^{**}. Moreover, the right-hand side of (23) is invariant to θ^{**}, while the right-hand side of (22) is rising in θ^{**}. Therefore, there exists a solution to (21)-(23) that admits (22) as an equality.

Suppose next that (23) holds as a strict inequality in the solution. The derivative of (21) with respect to θ_n^{**} takes the same sign as $Q(\theta_n^{**})$. By Assumption 2 and

$f(\overline{\theta}) = 0$, $Q(\theta)$ is strictly positive for low θ and strictly negative for high θ given $\theta < \overline{\theta}$, and $Q(\theta) = 0$ given $\theta > \overline{\theta}$. Thus, the optimal value of θ_n^{**} is interior and satisfies $Q(\theta_n^{**}) = 0$. Now consider the optimal value of θ_n^*. The derivative of (21) with respect to θ_n^* is proportional to $-\int_{\theta_n^{**}}^{\theta_n^*} Q(\theta)d\theta$. Since $Q(\theta) < 0$ for all $\theta \in \left(\theta_n^{**}, \overline{\theta}\right)$, it follows that $-\int_{\theta_n^{**}}^{\theta_n^*} Q(\theta)d\theta > 0$ for all θ_n^*, and thus the value of θ_n^* that maximizes the objective is unbounded from above. Since the right-hand side of (23) approaches ∞ as $\theta_n^* \to \infty$, it follows that (23) must hold with equality in the solution.

Step 2b. We show that if the largest value P^* that admits $T(P^*) = P^*$ satisfies $P^* > 0$, then $P^* \geq \overline{V}(b_{-1}) - \underline{V}(b_{-1})$, with equality if $[\underline{b}(b_{-1}), \overline{b}(b_{-1})]$ is sufficiently large that $\{\overline{V}(b_{-1}), \underline{V}(b_{-1})\}$ are supported by interior allocations.

Fix $P > 0$ and consider programs ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$), defining $\overline{V}(b_{-1})$ and $\underline{V}(b_{-1})$ respectively, with the constraint that the highest and lowest feasible continuation values satisfy $\overline{V}(b(\theta)) - \underline{V}(b(\theta)) = P$. By the arguments in Subsection C.1, we can represent these programs using saving rates $s(\theta)$ and normalized continuation values $\widetilde{V}(s(\theta))$, where the lowest and highest normalized continuation values $\underline{\widetilde{V}}$ and $\overline{\widetilde{V}}$ are independent of initial debt and satisfy $\overline{\widetilde{V}} - \underline{\widetilde{V}} = P$. Letting $\widetilde{P}(s(\theta)) \equiv \widetilde{V}(s(\theta)) - \underline{\widetilde{V}}$, the limited commitment constraint (8) using such a representation can be written as

$$\alpha\theta \log(1 - s(\theta)) + \frac{\delta}{1-\delta}\mathbb{E}[\theta] \log(s(\theta)) + \delta \widetilde{P}(s(\theta))$$
$$\geq \alpha\theta \log(1 - s^f(\theta)) + \frac{\delta}{1-\delta}\mathbb{E}[\theta] \log(s^f(\theta)), \tag{43}$$

where $s(\theta) \in [\underline{s}, \overline{s}]$, $\widetilde{P}(s(\theta)) \in [0, P]$, and $s^f(\theta)$ denotes the savings rate associated with flexible spending $g^f(\theta)$.

We claim that if the feasible set $[\underline{s}, \overline{s}]$ is large enough, then the solutions to programs ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$) conditional on $\overline{\widetilde{V}} - \underline{\widetilde{V}} = P$ must be interior. Suppose by contradiction that this is not the case. Observe that given $\widetilde{P}(s(\theta)) \in [0, P]$, the left-hand side of constraint (43) approaches $-\infty$ as $s(\theta)$ approaches either 0 or 1. Thus, if the allocation is at the boundaries of the set $[\underline{s}, \overline{s}]$, the constraint is violated for $[\underline{s}, \overline{s}]$ large enough, yielding a contradiction.

It follows that for sufficiently large $[\underline{s}, \overline{s}]$ and $\overline{\widetilde{V}} - \underline{\widetilde{V}} = P > 0$, programs ($\mathcal{P}_{\max}$) and ($\mathcal{P}_{\min}$) admit interior allocations and Proposition 2 and Proposition 3 hold. Thus, conditional on $P > 0$, the highest value $\overline{V}(b_{-1})$ must be bounded from above by the the solution to ($\mathcal{P}_{\max}$) described in Proposition 2, and the lowest value $\underline{V}(b_{-1})$ must be bounded from below by the solution to ($\mathcal{P}_{\min}$) described in Proposition 3. The claim then follows from Step 1, the definition of $T(P)$, and Step 2a.

Step 3. We show that $T(P)$ has the following properties: $T'(P) > 0$; $T''(P) < 0$; $\lim_{P\to\infty} T'(P) < 1$; and

$$\lim_{P\to 0} T'(P) > (<) 1 \text{ if } 1 + 2\frac{\delta}{1-\delta} Q(\widehat{\theta}) < (>) 0. \tag{44}$$

Step 3a. We show that any solution to (21)-(23) for $P > 0$ is interior. Consider first $\{\theta^*, \theta^{**}\}$. Suppose that $\theta^* = 0$. Then (22) would be violated since $U(g^f(0)) = -\infty$, unless $\theta^{**} = \theta^*$, but in that case (22) would be a strict inequality, violating Step 2a. Therefore, $\theta^* > 0$. Analogous arguments imply that θ^{**} is finite. Since $\theta^* < \theta^{**}$ (by (22) binding in the solution), it follows that both θ^* and θ^{**} are interior.

Consider next $\{\theta_n^*, \theta_n^{**}\}$. Suppose that $\theta_n^* = \infty$. Then (23) would be violated since $U(g^f(\infty)) = \infty$, unless $\theta_n^{**} = \theta_n^*$, but in that case (23) would be a strict inequality, violating Step 2a. Therefore, θ_n^* is finite. Suppose next that $\theta_n^{**} = 0$. Then necessarily $\theta_n^* > 0$, since otherwise (23) would be a strict inequality, violating Step 2a. Consider an increase in θ_n^{**} by $\varepsilon > 0$ arbitrarily small. The change in the objective in (21) is proportional to $Q(\theta_n^{**}) = 1 > 0$. Constraint (22) is unchanged, whereas constraint (23) is relaxed as its right-hand side decreases. Therefore, $\theta_n^{**} > 0$. Since $\theta_n^{**} < \theta_n^*$ (by (23) binding in the solution), it follows that both θ_n^{**} and θ_n^* are interior.

Step 3b. We show that the solution to (21)-(23) is unique. Let $\mu^R \geq 0$ and $\mu^P \geq 0$ denote the Lagrange multipliers on (22) and (23). By Step 3a, the solution is characterized by the following first-order conditions:

$$\int_{\theta^*}^{\theta^{**}} Q(\theta) d\theta = -\mu^R \int_{\theta^*}^{\theta^{**}} 1 d\theta \tag{45}$$

$$Q(\theta^{**}) = -\mu^R \tag{46}$$

$$\int_{\theta_n^{**}}^{\theta_n^*} Q(\theta) d\theta = -\mu^P \int_{\theta_n^{**}}^{\theta_n^*} 1 d\theta \tag{47}$$

$$Q(\theta_n^{**}) = -\mu^P. \tag{48}$$

Conditions (45) and (46) imply

$$\int_{\theta^*}^{\theta^{**}} \left[Q(\theta) - Q(\theta^{**})\right] d\theta = 0. \tag{49}$$

Observe that the derivative of the left-hand side with respect to θ^* is $-(Q(\theta^*) - Q(\theta^{**}))$, and the derivative of the left-hand side with respect to θ^{**} is $-\int_{\theta^*}^{\theta^{**}} Q'(\theta^{**}) d\theta$. Both of these are negative given Assumption 2 and (49), so condition (49) defines a decreasing relationship between θ^* and θ^{**}. Now consider constraint (22) which holds with equality by Step 2a. The right-hand side of (22) is increasing in θ^{**} but decreasing in θ^*, so (22) defines an increasing relationship between θ^* and θ^{**}. It follows that the values of θ^* and θ^{**} are uniquely pinned down by (22) and (49).

Conditions (47) and (48) imply

$$\int_{\theta_n^{**}}^{\theta_n^*} \left[Q(\theta_n^{**}) - Q(\theta)\right] d\theta = 0. \tag{50}$$

By analogous arguments to those used above, this condition defines a decreasing relationship between θ_n^{**} and θ_n^*, whereas constraint (23), which holds with equality by Step 2a, defines an increasing relationship between θ_n^* and θ_n^{**}. It follows that the values of θ_n^* and θ_n^{**} are uniquely pinned down by (23) and (50).

Step 3c. We show that $T'(P) > 0$. By the Envelope condition,

$$T'(P) = \delta\left(1 + \mu^R + \mu^P\right) = \delta\left(1 - Q(\theta^{**}) - Q(\theta_n^{**})\right), \tag{51}$$

where the second equality follows from (46) and (48). Given Assumption 2, conditions (49) and (50) imply that $\theta^* < \widehat{\theta} < \theta^{**}$, $\theta_n^* > \widehat{\theta} > \theta_n^{**}$, $Q(\theta^{**}) < 0$, and $Q(\theta_n^{**}) < 0$. Therefore, (51) implies $T'(P) > 0$.

Step 3d. We show that $T''(P) < 0$. From Step 3c,

$$T''(P) = \delta\left(-Q'(\theta^{**})\frac{d\theta^{**}}{dP} - Q'(\theta_n^{**})\frac{d\theta_n^{**}}{dP}\right)$$

(where recall that $Q(\theta)$ is differentiable everywhere given $f(\underline{\theta}) = f(\overline{\theta}) = 0$). Assumption 2, (49) and (50) imply that $Q'(\theta^{**}) > 0$ and $Q'(\theta_n^{**}) < 0$. To prove that $T''(P) < 0$, it is therefore sufficient to prove that $d\theta^{**}/dP > 0$ and $d\theta_n^{**}/dP < 0$.

Consider first $d\theta^{**}/dP$. A higher value of P means that a strictly higher value of θ^{**} is required to satisfy (22) with equality for every value of θ^*. Given the decreasing relationship between θ^* and θ^{**} defined by condition (49), it follows that a higher value of P is associated with a lower value of θ^* and a higher value of θ^{**}. Thus, $d\theta^{**}/dP > 0$.

Consider next $d\theta_n^{**}/dP$. A higher value of P means that a lower value of θ_n^{**} is required to satisfy (23) with equality for every value of θ_n^*. Given the decreasing relationship between θ_n^* and θ_n^{**} defined by condition (50), it follows that a higher value of P is associated with a higher value of θ_n^* and a lower value of θ_n^{**}. Thus, $d\theta_n^{**}/dP < 0$.

Step 3e. We show that $\lim_{P\to\infty} T'(P) < 1$. Using (51), observe that if μ^R and μ^P each approach 0 as $P \to \infty$, then $T'(P)$ approaches $\delta < 1$. To prove the claim, it is thus sufficient to prove that μ^R and μ^P each approach 0 as $P \to \infty$.

We first show that $\mu^R = 0$ for P sufficiently large. Consider the solution to the relaxed problem in (21) that ignores constraint (22). The first-order conditions (45) and (46) imply $\theta^{**} \geq \overline{\theta}$ and $\theta^* = \theta_e$ for θ_e defined by

$$\int_{\theta_e}^{\overline{\theta}} Q(\theta) d\theta = 0. \tag{52}$$

It follows that if

$$\delta P \geq \alpha \int_{\theta_e}^{\overline{\theta}} \left[U(g^f(\theta)) - U(g^f(\theta_e))\right] d\theta,$$

then this solution satisfies (22) for some $\theta^{**} \geq \overline{\theta}$. Therefore, the solution to the relaxed problem solves the original problem, implying $\mu^R = 0$ for P sufficiently large.

We next show that $\mu^P \to 0$ as $P \to \infty$. From Assumption 2 and (50), $Q(\theta_n^{**}) < 0$, implying that θ_n^{**} is strictly bounded from below. It follows that for (23) to hold as an equality, it must be that $\theta_n^* \to \infty$ as $P \to \infty$. We can then rewrite condition (50) as

$$\int_{\theta_n^{**}}^{\theta_n^*} \left[Q(\theta_n^{**}) - Q(\theta)\right] d\theta = \left(\theta_n^* - \theta_n^{**}\right) Q(\theta_n^{**}) - \int_{\theta_n^{**}}^{\overline{\theta}} Q(\theta) d\theta = 0.$$

Substituting with (48) and rearranging terms yields

$$\mu^P = -\frac{\int_{\theta_n^{**}}^{\overline{\theta}} Q(\theta) d\theta}{\theta_n^* - \theta_n^{**}}.$$

As $P \to \infty$ and thus $\theta_n^* \to \infty$, the numerator is bounded whereas the denominator grows unboundedly. Thus, $\mu^P \to 0$ as $P \to \infty$.

Step 3f. We show that $T(P)$ has the property stated in (44). Given Assumption 2, conditions (49) and (50) require $\theta^* < \widehat{\theta} < \theta^{**}$ and $\theta_n^* > \widehat{\theta} > \theta_n^{**}$. Therefore, satisfaction of (22) and (23) implies that θ^*, θ^{**}, θ_n^* and θ_n^{**} each approach $\widehat{\theta}$ as $P \to 0$. Using (51), it thus follows that

$$\lim_{P\to 0} T'(P) = \delta(1 - 2Q(\widehat{\theta})),$$

and therefore $\lim_{P\to 0} T'(P) > (<) 1$ if $1 + 2\frac{\delta}{1-\delta} Q(\widehat{\theta}) < (>) 0$.

Step 4. We prove the claim in the proposition.

Observe that $T(0) = 0$, and by continuity, $\lim_{P\to 0} T(P) = 0$. Consider the condition given in Step 3:

$$1 + 2\frac{\delta}{1-\delta} Q(\widehat{\theta}) < 0. \tag{53}$$

Suppose first that (53) holds. Then by Step 3, the shape of $T(P)$ implies that there exists a unique value $P^* > 0$ such that $T(P^*) = P^*$. By Step 2b, for a sufficiently large feasible set $[\underline{s}, \overline{s}]$, the values $\overline{V}(\cdot)$ and $\underline{V}(\cdot)$ are supported by interior allocations and satisfy $\overline{V}(\cdot) - \underline{V}(\cdot) = P^*$, implying $\overline{V}(\cdot) > \underline{V}(\cdot)$.

Suppose next that (53) does not hold. Then by Step 3, the largest value P^* that admits $T(P^*) = P^*$ is $P^* = 0$. By Step 2b, $P^* \geq \overline{V}(\cdot) - \underline{V}(\cdot)$ for all $P^* > 0$. Hence, by continuity, we must have $\overline{V}(\cdot) = \underline{V}(\cdot)$.

Given the above claims, all is left to show is that (53) holds if and only if $\delta > \widetilde{\delta}$ and $\alpha > \widetilde{\alpha}$. Observe that the left-hand side of this condition is strictly decreasing in α. Let $\widetilde{\delta} \in (0, 1)$ be the discount factor that sets the left-hand-side equal to 0 when $\alpha \to \infty$:

$$1 + 2\frac{\widetilde{\delta}}{1-\widetilde{\delta}}(1 - F(\widehat{\theta}) - \widehat{\theta} f(\widehat{\theta})) = 0,$$

for $\widehat{\theta}$ satisfying $\widehat{\theta} f'(\widehat{\theta})/f(\widehat{\theta}) = -2$. Then if $\delta \leq \widetilde{\delta}$, condition (53) is violated for all $\alpha \geq 1$. If instead $\delta > \widetilde{\delta}$, we can find a finite value $\widetilde{\alpha} \geq 1$ such that the left-hand side of (53) equals 0, so that condition (53) holds for $\alpha > \widetilde{\alpha}$ and is violated for $\alpha \leq \widetilde{\alpha}$. Observe that since the left-hand side of condition (53) is increasing in δ, such a value $\widetilde{\alpha}$ is decreasing in δ.

C.3 Proof of Corollary 1

By Proposition 4, given $\delta > \widetilde{\delta}$ and $\alpha > \widetilde{\alpha}$, the best equilibrium has fiscal regimes characterized by maximally enforced deficit and surplus limits as described in Proposition 2 and Proposition 3, with cutoffs $\{\theta^*, \theta^{**}, \theta_n^*, \theta_n^{**}\}$ and value of punishment $P = \overline{V}(b_{-1}) - \underline{V}(b_{-1})$. Moreover, as claimed in Step 3f in the proof of Proposition 4, we must have $\theta^* < \widehat{\theta} < \theta^{**}$, where $\widehat{\theta} \in (\underline{\theta}, \overline{\theta})$ in this environment, and where θ^* and θ^{**} each approach $\widehat{\theta}$ as $P \to 0$. The corollary then follows from the fact that, given $\delta > \widetilde{\delta}$ and $\alpha > \widetilde{\alpha}$, we have $P \to 0$ as $\alpha \to \widetilde{\alpha}$. Thus, there exists $\widetilde{\widetilde{\alpha}} > \widetilde{\alpha}$ such that for $\alpha \in (\widetilde{\alpha}, \widetilde{\widetilde{\alpha}})$, P is sufficiently close to 0, and thus θ^* and θ^{**} are sufficiently close to $\widehat{\theta}$, that we must have $\underline{\theta} < \theta^* < \widehat{\theta} < \theta^{**} < \overline{\theta}$.