Fiscal Rules and Discretion under Self-Enforcement*

Marina Halac† Pierre Yared‡

December 15, 2017

Abstract

We study a fiscal policy model in which the government is present-biased towards public spending. Society chooses a fiscal rule to trade off the benefit of committing the government to not overspend against the benefit of granting it flexibility to react to privately observed shocks to the value of spending. Unlike prior work, we characterize rules that are self-enforcing: the government must prefer to comply with the rule rather than face the punishment that follows a breach, where any such punishment must also be self-enforcing. We show that the optimal rule is a maximally enforced deficit limit, which, if violated, leads to the worst punishment for the government. We provide a necessary and sufficient condition for the government to violate the deficit limit following sufficiently high shocks. Punishment takes the form of temporary overspending, after which the optimal rule is restored.

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

*We would like to thank Manuel Amador, Dirk Bergemann, V.V. Chari, Georgy Egorov, Simone Galperti, Zhen Huo, Navin Kartik, Chris Moser, and seminar audiences for comments. George Vojta and Weijie Zhong provided excellent research assistance.

†Columbia University, Graduate School of Business and CEPR. Email: mhalac@columbia.edu.
‡Columbia University, Graduate School of Business and NBER. Email: pyared@columbia.edu.

1 Introduction

Countries impose rules on their governments to constrain their policy decisions. Increasingly prevalent are fiscal rules, in place in 92 countries in 2015, compared to only seven countries in 1990.[1] Governments however do not always respect these rules, and credible enforcement mechanisms are critical to their institution.

Take for example the case of Chile. In 2007, the Chilean government resisted popular pressure to break a fiscal rule that had been in place since 2001, fearing that a breach would set a bad precedent for the future.[2] In 2009, however, faced with the global financial crisis, the government decided to break the rule, citing extraordinary circumstances.[3] No formal sanctions would be imposed, yet the overrule was not without penalties: it led to lax fiscal policy which, as it had been feared, persisted for some time, including under the next administration.[4] It was only in October of 2011 that Chile reinstated its fiscal rule.[5]

In this paper, we study the optimal design of fiscal rules under limited enforcement. How restrictive is an optimal fiscal rule? Should the government violate it when the economy is in distress? What is the optimal structure of penalties for violating the rule?

Our analysis of fiscal rules builds on the approach used in Amador, Werning, and Angeletos (2006) and Halac and Yared (2014).[6] We consider a government that is present-biased towards public spending and privately informed about shocks affecting the value of this spending. Society chooses a fiscal rule to trade off the benefit of committing the government to not overspend against the benefit of granting it flexibility to react to shocks. We depart from prior work by relaxing the assumption that rules can be perfectly enforced. Motivated by real-world rules, we posit that fiscal rules must be *self-enforcing*: the government must prefer to comply with the rule rather than face the punishment that follows a breach. Such a punishment must also be self-enforcing, meaning that the government in the future

[1] See IMF Fiscal Rules Data Set, 2015 and Schaechter et al. (2012).

[2] While some claimed that the rule was too tight, others believed that "loosening the rule when copper prices are still high would send a dangerous message of laxity." See "Corseted: Too Virtuous for its Own Good?," *The Economist*, January 25, 2007, `http://www.economist.com/node/8597095`.

[3] The government changed the surplus target in 2008 and triggered a de facto escape clause in 2009. See Schmidt-Hebbel (2014) and Lledó et al. (2017).

[4] Despite initially criticizing the previous government's fiscal laxity in 2009, the Piñera government chose an even less stringent fiscal path in 2010, while pursuing populist policies. See Velasco and Arenas (2010), Cabalin (2011), and "Government of Bachelet, 'Surprised' by Piñera's Criticisms of Fiscal Deficit," *América Economía*, February 22, 2010, `https://www.americaeconomia.com/economia-mercados/finanzas/gobierno-de-bachelet-sorprendido-por-criticas-de-pinera-deficit-fiscal`.

[5] See Larraín et al. (2011) for a discussion of the methodological and institutional considerations behind this measure.

[6] See also Athey, Atkeson, and Kehoe (2005), Amador and Bagwell (2013), and Ambrus and Egorov (2013).

must have incentives to pursue it. These self-enforcement requirements, combined with the government's private information, introduce new challenges into our problem; we consider a simple framework that allows for a full characterization of the optimal self-enforcing fiscal rule.

Our environment is a small open economy in which the government makes spending and borrowing decisions. In each period, an independent and identically distributed (i.i.d.) shock to the social value of deficit-financed government spending is realized. The government in each period is present-biased: compared to society, the government overvalues current spending relative to the stream of future spending. This preference structure results naturally from the aggregation of heterogeneous, time-consistent citizens' preferences (Jackson and Yariv, 2014, 2015), or as a consequence of turnover in a political economy setting (e.g., Aguiar and Amador, 2011).[7] We assume that the shocks to the value of spending are privately observed by the government, capturing the fact that rules cannot explicitly condition on all contingencies in fiscal policy. Moreover, the government has full discretion regarding the level of spending, capturing the fact that any constraint on spending must be self-enforcing. A fiscal rule in this setting assigns the government a level of spending for each shock, where this assignment must satisfy the government's private information and self-enforcement constraints.

To describe the forces underlying our model, suppose first that fiscal rules could be perfectly enforced. Society then chooses a rule to optimally resolve the tradeoff between commitment and flexibility. Fully committing the government to a spending path would allow to implement the first best policy in the absence of private information, whereas granting the government full flexibility would yield the first best policy in the absence of a present bias. Given both private information and a present bias, however, a tradeoff arises, and the first best is not implementable. Amador, Werning, and Angeletos (2006) show that, under perfect enforcement (and certain distributional assumptions), the solution to this tradeoff is a fiscal rule that takes the form of a deficit limit. In each period, the government spends within the limit if the shock to the value of spending is relatively low, and it spends at the limit if the shock is higher. The problem is essentially static: the rule at each date prescribes future deficit limits that maximize future social welfare, and dynamic incentives are thus not provided to the government.[8]

Our focus is on understanding the optimal fiscal rule when external enforcement is not

[7] See also Persson and Svensson (1989), Alesina and Tabellini (1990), Alesina and Perotti (1995), Lizzeri (1999), Tornell and Lane (1999), Battaglini and Coate (2008), and Caballero and Yared (2010). Our formulation corresponds to the quasi-hyperbolic consumption model; see Laibson (1997).

[8] This feature is due to the assumption that shocks are i.i.d.; see Halac and Yared (2014).

possible, so the rule must be self-enforcing. As is also true under perfect enforcement, a self-enforcing rule must satisfy the government's private information constraints: given a realized shock, the government must prefer its assigned spending level and continuation value to those prescribed for a different shock. In addition, a self-enforcing rule must satisfy the government's self-enforcement constraints: given a realized shock, the government must prefer its assigned spending level and continuation value to a spending level not prescribed for any shock. Any observable deviation at a date t—where the government chooses a spending level corresponding to no shock—is optimally punished by inducing the worst possible continuation from date $t+1$ onward.[9] In fact, if this punishment is severe enough, self-enforcement constraints are non-binding; the government always prefers to abide to the perfect-enforcement deficit limit to avoid the punishment.

Our main result is a characterization of the optimal self-enforcing fiscal rule when self-enforcement constraints are binding. We show that this rule takes the form of a *maximally enforced deficit limit*, which, if violated, leads to the worst punishment for the government. The rule is thus similar to that under perfect enforcement, although it differs in two aspects. First, the deficit limit imposed on the government is more relaxed than the perfect-enforcement limit. Second, the optimal self-enforcing rule may provide dynamic incentives, with the government breaching the deficit limit and triggering punishment under sufficiently high shocks.

To obtain this characterization, we begin by establishing general properties of optimal incentive provision. Society incentivizes the government not to overspend by using continuation values as reward and punishment. We show that in any optimal rule, continuation values must take a *bang-bang* form, so the government receives either the maximal reward or the worst punishment in the future, depending on its behavior in the present.[10] Using milder punishments would be less socially costly; however, the harshest future punishment maximizes the range of shocks under which society can impose fiscal discipline in the present, and we show that this maximizes social welfare.

Given a continuum of shocks with differentiable density, our bang-bang result relies only on generic properties of the distribution function, which guarantee that the information structure is rich enough that steepening incentives always allows to reduce distortions. Under additional distributional assumptions, we are able to further show that optimal bang-bang continuation values must be monotonic in the realized shock. In fact, building on monotonicity, we characterize the spending allocation, establishing that these distributional

[9] This follows from standard arguments; see Abreu (1988).

[10] As we discuss in the review of the literature and in more detail in Subsection 4.1, our bang-bang result is related to, but different from, that in Abreu, Pearce, and Stacchetti (1990).

assumptions are sufficient, as well as necessary, for maximally enforced deficit limits to be uniquely optimal.[11]

Specifically, when the perfect-enforcement rule is not self-enforcing, we show that there are two possible cases for the optimal self-enforcing rule. In the first case, society sets a relaxed deficit limit that satisfies the self-enforcement constraint under all shocks. In the second case, society prescribes a tighter deficit limit with dynamic incentives: the government receives the maximal continuation value if it respects the limit and the worst continuation value if it breaches the limit, where the latter occurs following high enough shocks. We show that the optimal choice for society depends on the distribution of shocks and the government's present bias. Given feasible continuation values, society chooses a tight deficit limit enforced by dynamic incentives if and only if high shocks are sufficiently rare and the government's present bias is sufficiently severe. A deficit limit that is respected under all shocks is not optimal in this case, as this limit would need to be too relaxed to accommodate self-enforcement for high shocks that are unlikely. Society is better off imposing a tighter deficit limit that binds under lower shocks, while letting the government violate the limit and be punished following high shocks.

We complete our characterization of the optimal self-enforcing fiscal rule by examining the structure of punishment. We show that the worst punishment—which, as just described, may be triggered in an optimal rule—takes the form of temporary overspending. This punishment can be implemented with a *maximally enforced surplus limit*. If the shock to the value of spending is relatively high, the government respects the surplus limit and is rewarded with the maximal continuation value; if the shock is relatively low, the government may violate the limit and be punished with the worst continuation value. The logic for this result is analogous to that for the optimal deficit limit, but reverse. A maximally enforced surplus limit minimizes social welfare (and thus maximizes punishment) by incentivizing the government to overspend as much as possible. Since the government is present-biased towards spending, incentivizing overspending relaxes self-enforcement constraints and achieves lower social welfare than incentivizing underspending.

The form of the maximally enforced surplus limit implies that punishment is always temporary. In particular, the economy remains in punishment only if shocks are relatively low; when a high shock occurs, the economy transitions back to a maximally enforced deficit limit, which is associated with the highest continuation value. An optimal fiscal rule with dynamic incentives is therefore self-enforced by transitions in and out of the best and worst equilibria, with the economy fluctuating between periods of fiscal rectitude and periods of

[11] These assumptions are similar to, but stronger than, those in Amador, Werning, and Angeletos (2006). See Section 4.

fiscal profligacy.

Related literature. Our paper fits into the aforementioned literature on mechanism design that studies the tradeoff between commitment and flexibility.[12] This literature is mainly concerned with environments with perfect enforcement, whereas we examine the optimal rule under self-enforcement. Closely related is Amador and Bagwell (2016), which considers the classic problem of regulating a privately informed monopolist but in the absence of transfers. In a static setting with an ex-post participation constraint, the paper shows that optimal regulatory policy takes the form of a threshold which the monopolist satisfies unless it decides to shut down. In contrast to this paper, our analysis does not rely on the perfect enforcement of thresholds, and it considers "money burning" in the form of self-enforcing dynamic punishments.

Also related to our work is the literature on the political economy of fiscal policy.[13] Dovis and Kirpalani (2017), in particular, examine deficit limits that are endogenously enforced by reputational concerns. We depart by studying the optimal self-enforcing fiscal rule under privately observed shocks and no restrictions on its structure. Our work further contributes to the literature on hyperbolic discounting and the benefits of commitment devices.[14] Most closely related is Bernheim, Ray, and Yeltekin (2015), which analyzes the self-enforcement of consumption rules for a consumer with quasi-hyperbolic preferences. The paper shows that optimal punishments take the form of temporary over-consumption; however, the model features no private information, and as a consequence punishment does not occur along the equilibrium path.

Finally, the equilibrium dynamics in our environment, featuring fluctuation between reward and punishment, together with our bang-bang characterization, are related to the analysis of price wars in Green and Porter (1984) and more broadly the dynamics of repeated games in Abreu, Pearce, and Stacchetti (1990). The latter provides conditions under which the bang-bang property is necessary for optimality in settings with repeated moral hazard.

[12] In addition to the work previously cited, see Sleet (2004). More broadly, our paper relates to the literature on delegation in principal-agent settings, including Holmström (1977, 1984), Alonso and Matouschek (2008), and Ambrus and Egorov (2015). Hörner and Guo (2015), Li, Matouschek, and Powell (2017), and Lipnowski and Ramos (2017) study related dynamic problems. The agent in these settings has a bias applying to all periods, rather than a present bias as in our self-control problem.

[13] In addition to Aguiar and Amador (2011) and the work in fn. 7, see Krusell and Rios-Rull (1999), Acemoglu, Golosov, and Tsyvinski (2008), Yared (2010b), Azzimonti (2011), and Song, Storesletten, and Zilibotti (2012). For a recent quantitative analysis of fiscal rules, see Alfaro and Kanczuk (2016), Azzimonti, Battaglini, and Coate (2016), and Hatchondo, Martinez, and Roch (2017), and for a study of coordinated fiscal rules across countries, see Halac and Yared (2017b).

[14] See for example Phelps and Pollak (1968), Laibson (1997), Barro (1999), Krusell and Smith, Jr. (2003), Krusell, Kruscu, and Smith, Jr. (2010), Lizzeri and Yariv (2014), Bisin, Lizzeri, and Yariv (2015), Cao and Werning (2017), Dong (2017), and Moser and de Souza e Silva (2017).

As our environment is one of repeated adverse selection, our characterization is not a direct application of their result; yet, the intuition is related, as we discuss in Subsection 4.1. Athey, Bagwell, and Sanchirico (2004) also address related issues in a repeated Bertrand game with private information.

2 Model

We study a simple model of fiscal policy in which the government in each period makes spending and borrowing decisions. Our environment is similar to that analyzed in Amador, Werning, and Angeletos (2006) and Halac and Yared (2014).[15] As discussed in Section 3, our departure will be in considering fiscal rules that are self-enforcing as opposed to externally enforced.

Consider a small open economy. At the beginning of each period, $t \in \{0, 1, \ldots\}$, the government observes a shock to the economy, $\theta_t > 0$, which is the government's private information or *type*. θ_t is drawn from a bounded set $\Theta \equiv \left[\underline{\theta}, \overline{\theta}\right]$ with a continuously differentiable probability density function $f(\theta_t) > 0$ and associated cumulative density function $F(\theta_t)$.

Following the realization of θ_t in each period t, the government chooses public spending $G_t \geq 0$ and debt B_{t+1} subject to a budget constraint:

$$G_t = \tau + \frac{B_{t+1}}{1+r} - B_t, \tag{1}$$

where $\tau > 0$ is the exogenous fixed tax revenue collected by the government in each period, B_t is the level of debt with which the government enters period t, and r is the exogenous interest rate. B_0 is exogenous and $\lim_{t\to\infty} \frac{B_{t+1}}{(1+r)^t} = 0$, so that all debts must be repaid and all assets must be consumed. Note that cross-subsidization across types is not possible: the net present value of public spending at any point cannot depend on the government's type, unlike in other models such as Atkeson and Lucas (1992) and Thomas and Worrall (1990).[16]

Social welfare at the beginning of date t is

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

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

[16] Other papers that build on these and also feature cross-subsidization are Sleet and Yeltekin (2006, 2008) and Farhi and Werning (2007).

where $\theta_t U(G_t)$ is the social utility from public spending at date t and $\delta \in (0,1)$ is the discount factor. The government's welfare at date t after the realization of its type θ_t, when choosing spending G_t, is

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

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

There are two important features of this environment. First, the government's objective (3) following the realization of its type does not coincide with the social objective (2). In particular, compared to society, the government overweighs the importance of current spending relative to future spending. As mentioned in the Introduction, this structure arises naturally when the government's preferences aggregate citizens' preferences. Jackson and Yariv (2015) show that with any heterogeneity in citizens' preferences, every non-dictatorial aggregation method that respects unanimity must be time inconsistent; moreover, any such method that is time separable must lead to a present bias and thus a representation like that in (2)-(3).[17] Our formulation can also be motivated by political turnover. For instance, preferences such as these may emerge in settings with political uncertainty where policymakers place a higher value on public spending when they hold power and can make spending decisions. A bias of $\beta \in (0,1)$ for a current policymaker can be interpreted in this sense as the probability that the policymaker will be part of the government in the next and all future periods. Due to turnover, policymakers are biased towards present spending relative to future spending and incur excessively high debts; see, e.g., Aguiar and Amador (2011).

The second feature of our environment is that the realization of θ_t—which affects the marginal social utility of public spending—is privately observed by the government at date t. One interpretation is that fiscal rules imposed on the government cannot explicitly condition on the value of θ_t, even if this shock were observable.[18] An alternative interpretation is that the exact cost of public goods is only observable to the policymaker, who may be inclined to overspend on these goods. A third possibility is that citizens have heterogeneous preferences or heterogeneous information regarding the optimal level of public spending, and the government sees an aggregate that the citizens do not see (see Sleet, 2004).

To facilitate an explicit characterization of the optimal self-enforcing fiscal rule, we

[17] Specifically, any such time-separable aggregation method will be utilitarian, and by the results in Jackson and Yariv (2014), it will thus be present-biased.

[18] Halac and Yared (2017a) study a delegation problem in which shocks can be verified by a rule-making body at a cost.

assume:[19]

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

Assumption 1 implies that welfare is separable with respect to the level of debt. To see this, let the spending rate at date t be $g_t \in [0,1]$, corresponding to the fraction of lifetime resources that are spent at t:

$$g_t = \frac{G_t}{(1+r)\tau/r - B_t}. \tag{4}$$

Denote the savings rate at date t by $x_t = 1 - g_t$ and let $W(x_t) \equiv \frac{\delta \mathbb{E}[\theta_t] U(x_t)}{(1-\delta)}$. Using this notation and Assumption 1, social welfare in (2) can be rewritten as

$$\sum_{k=0}^{\infty} \delta^k \mathbb{E}\big[\theta_{t+k} U(g_{t+k}) + W(x_{t+k})\big] + \chi(B_t), \tag{5}$$

for a constant $\chi(B_t)$ that depends on B_t.[20] Analogously, the government's welfare in (3), following the realization of θ_t at date t, can be rewritten as

$$\theta_t U(g_t) + \beta W(x_t) + \beta \sum_{k=1}^{\infty} \delta^k \mathbb{E}\big[\theta_{t+k} U(g_{t+k}) + W(x_{t+k})\big] + \varphi(B_t), \tag{6}$$

for a constant $\varphi(B_t)$ that depends on B_t.[21]

Given the representation in (5) and (6), hereafter we consider the problem of a government that chooses a spending rate g_t and savings rate $x_t = 1 - g_t$ in each period t. Following Bernheim, Ray, and Yeltekin (2015), we impose a lower bound ν and upper bound $1 - \nu$ on g_t so that payoffs (and thus punishments) are bounded; we take $\nu > 0$ small enough that this constraint is otherwise non-binding.

In this environment, the first-best policy that maximizes social welfare (5) at date $t = 0$ is defined by a stochastic sequence of spending and savings rates that satisfy $g_t^{fb} = g^{fb}(\theta_t)$ and $x_t^{fb} = x^{fb}(\theta_t)$, where

$$\theta_t U'(g^{fb}(\theta_t)) = W'(x^{fb}(\theta_t)). \tag{7}$$

On the other hand, the government's flexible policy is defined as the spending and savings rates $g_t^f = g^f(\theta_t)$ and $x_t^f = x^f(\theta_t)$ that maximize the government's welfare (6) at date t,

[19] This assumption is made 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 $\sum_{k=0}^{\infty} \delta^k \mathbb{E}\left[\theta_{t+k}\right] U((1+r)^k[\tau(1+r)/r - B_t])$.

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

holding future government policies $\{g_{t+k}, x_{t+k}\}_{k=1}^{\infty}$ fixed. The flexible policy satisfies

$$\theta_t U'(g^f(\theta_t)) = \beta W'(x^f(\theta_t)). \tag{8}$$

3 Self-Enforcing Rules

A key feature of our analysis is that we allow the government to choose any feasible spending rate g_t in each period t. This is a main distinction from prior work, which restricts the set of available actions by assuming that rules can be perfectly enforced. We instead consider self-enforcing fiscal rules, motivated by the observation that, in practice, governments are free to choose whether to abide to rules, and their choices depend on the consequences of a breach. Hence, in our setting, a government's current spending decision must be self-enforced by the expected behavior of the government in the future following this decision. Specifically, we define a self-enforcing rule as a perfect public equilibrium of the interaction between successive governments. Such a rule must satisfy a sequence of private information and self-enforcement constraints, as we describe next.

3.1 Equilibrium Allocations

We consider the interaction between the governments in each period t. As is standard in the dynamic contracting literature, we restrict attention to public strategies, namely those that depend on the public history and current private information, but not on any privately observed history. Let $h^{t-1} = \{g_0, g_1, ..., g_{t-1}\} \in [0,1]^{t-1}$ denote the public history of spending rates through time $t-1$ and $\mathcal{H}^{t-1}$ the set of all possible such histories. A public strategy for the government in period t is $\sigma_t(h^{t-1}, \theta_t)$, specifying, for each history $h^{t-1} \in \mathcal{H}^{t-1}$ and current government type $\theta_t \in \Theta$, a spending rate $g_t(h^{t-1}, \theta_t)$ and associated savings rate $x_t(h^{t-1}, \theta_t) = 1 - g_t(h^{t-1}, \theta_t)$. A perfect public equilibrium is a profile of public strategies $\sigma = (\sigma_t(h^{t-1}, \theta_t))_{t=0}^{\infty}$ such that for each $t \in \{0, 1, \ldots\}$, $\sigma_t(h^{t-1}, \theta_t)$ maximizes the t-period government's welfare (6) 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. We henceforth refer to perfect public equilibria as simply equilibria.

Let $\theta^{t-1} = \{\theta_0, \theta_1, \ldots, \theta_{t-1}\} \in \Theta^{t-1}$ denote the history of shocks through time $t-1$. Associated with any strategy profile σ is a spending sequence, $\left\{ \left\{ g_t\left(\theta^{t-1}, \theta_t\right) \right\}_{\theta^t \in \Theta^t} \right\}_{t=0}^{\infty}$, as a function of this history. Given a spending sequence and a realized history of shocks θ^{t-1},

denote by $V_t\left(\theta^{t-1}\right)$ the corresponding continuation value (normalized by debt) in period t:

$$V_t(\theta^{t-1}) = \sum_{k=0}^{\infty} \delta^k \mathbb{E}\left[\theta_{t+k} U(g_{t+k}(\theta^{t+k-1}, \theta_{t+k})) + W(x_{t+k}(\theta^{t+k-1}, \theta_{t+k}))\right].$$

The next lemma provides necessary and sufficient conditions for a spending sequence to be supported by equilibrium strategies.

Lemma 1. *A spending sequence* $\left\{ \left\{ g_t\left(\theta^{t-1}, \theta_t\right) \right\}_{\theta^t \in \Theta^t} \right\}_{t=0}^{\infty}$ *is supported by equilibrium strategies if and only if, for all* $t \in \{0, 1, \ldots\}$,

$$\begin{aligned} &\theta_t U(g_t(\theta^{t-1}, \theta_t)) + \beta W(x_t(\theta^{t-1}, \theta_t)) + \beta\delta V_{t+1}(\theta^{t-1}, \theta_t) \\ &\quad \geq \theta_t U(g_t(\theta^{t-1}, \theta_t')) + \beta W(x_t(\theta^{t-1}, \theta_t')) + \beta\delta V_{t+1}(\theta^{t-1}, \theta_t') \quad \forall \theta_t, \theta_t' \neq \theta_t, \end{aligned} \tag{9}$$

and

$$\begin{aligned} &\theta_t U(g_t(\theta^{t-1}, \theta_t)) + \beta W(x_t(\theta^{t-1}, \theta_t)) + \beta\delta V_{t+1}(\theta^{t-1}, \theta_t) \\ &\quad \geq \theta_t U(g^f(\theta_t)) + \beta W(x^f(\theta_t)) + \beta\delta \underline{V} \quad \forall \theta_t, \end{aligned} \tag{10}$$

where $\underline{V}$ *is the lowest value of* $V_t\left(\theta^{t-1}\right)$ *supported by equilibrium strategies.*

Constraint (9) is the private information constraint, capturing the fact that the government at date t can misrepresent its type. This constraint guarantees that a government of type θ_t prefers to pursue its equilibrium policy rather than that of another type θ_t'; given (9), this unobservable deviation would entail a spending rate $g_t(\theta^{t-1}, \theta_t')$ and continuation value $V_{t+1}(\theta^{t-1}, \theta_t')$. Constraint (10) is the self-enforcement constraint, capturing the fact that the government at date t can freely choose any feasible spending rate g_t. This constraint guarantees that a government of type θ_t prefers to pursue its equilibrium policy rather than deviating to its flexible policy $g^f(\theta_t)$ and being maximally punished with continuation value $\underline{V}$.[22] Note that the right-hand side of (10) is the government's minmax payoff.

Constraints (9) and (10) are clearly necessary along an equilibrium path. Furthermore, if a spending sequence satisfies these constraints, then it can be supported by an equilibrium in which, following an observable deviation in period t, the government in each following period reverts to the worst equilibrium, delivering a continuation value $\underline{V}$. Since an observable deviation is off the equilibrium path, it is without loss to assume that it is maximally punished (Abreu, 1988).

[22] The set of continuation values is closed by standard arguments since the set of feasible spending rates is compact.

3.2 Recursive Representation

A self-enforcing fiscal rule is a perfect public equilibrium as described in the previous section, inducing a sequence of spending that satisfies the private information constraint (9) and the self-enforcement constraint (10). A self-enforcing fiscal rule is optimal if it maximizes social welfare (5) at time $t = 0$ subject to these constraints.

Given the repeated nature of our model and following the logic of Abreu, Pearce, and Stacchetti (1990), we can represent the optimal self-enforcing rule recursively: rather than optimizing over an entire spending sequence, we choose current spending and savings rates $g(\theta)$ and $x(\theta)$ as a function of the shock θ, along with a continuation value $V(\theta)$ which is itself drawn from the set of values that satisfy the private information and self-enforcement constraints. That is, letting $\left[\underline{V}, \overline{V}\right]$ correspond to the set of continuation values supported by equilibrium strategies,[23] an optimal self-enforcing fiscal rule solves:

$$\overline{V} = \max_{\{g(\theta), x(\theta), V(\theta)\}_{\theta \in \Theta}} \mathbb{E}\left[\theta U(g(\theta)) + W(x(\theta)) + \delta V(\theta)\right] \tag{11}$$

subject to

$$\theta U(g(\theta)) + \beta W(x(\theta)) + \beta\delta V(\theta) \geq \theta U(g(\theta')) + \beta W(x(\theta')) + \beta\delta V(\theta') \quad \forall \theta, \theta' \neq \theta, \tag{12}$$

$$\theta U(g(\theta)) + \beta W(x(\theta)) + \beta\delta V(\theta) \geq \theta U(g^f(\theta)) + \beta W(x^f(\theta)) + \beta\delta \underline{V} \quad \forall \theta, \tag{13}$$

$$g(\theta) + x(\theta) = 1, \tag{14}$$

$$V(\theta) \in \left[\underline{V}, \overline{V}\right]. \tag{15}$$

The private information constraint (12) is a recursive formulation of (9), and the self-enforcement constraint (13) is a recursive formulation of (10). Constraints (14) and (15) are feasibility constraints. We say that a rule is incentive compatible if it satisfies (12)-(13), and it is incentive compatible and feasible, or incentive feasible for short, if it satisfies (12)-(15).

Analogous to the problem in (11), the worst punishment $\underline{V}$ corresponds to the solution to:

$$\underline{V} = \min_{\{g(\theta), x(\theta), V(\theta)\}_{\theta \in \Theta}} \mathbb{E}\left[\theta U(g(\theta)) + W(x(\theta)) + \delta V(\theta)\right] \tag{16}$$

subject to (12)-(15).

[23] The set of equilibrium continuation values can be guaranteed to be convex by introducing a public randomization device. Such a device however will not be used in the optimum, which has a bang-bang nature. We thus omit the details to ease the exposition.

Throughout our analysis, we assume that the solutions to both problems (11) and (16) admit piecewise continuously differentiable functions $g(\theta)$, which allows us to establish our results by the use of perturbations.[24] An alternative approach to using perturbation arguments would be to use Lagrangian methods, as in Amador, Werning, and Angeletos (2003). Since our problem is not globally concave, however, a Lagrangian approach would not establish uniqueness of the solution. Moreover, such an approach would make it difficult to identify general properties that must hold in any solution, as well as to obtain necessary and sufficient conditions for the optimality of maximally enforced deficit limits. We are able to provide these results using perturbations.

The next lemma follows from standard arguments; see Fudenberg and Tirole (1991):

Lemma 2. *$\{g(\theta), x(\theta), V(\theta)\}_{\theta\in\Theta}$ satisfies the private information constraint (12) if and only if: (i) $g(\theta)$ is nondecreasing, and (ii) the following local private information constraints are satisfied:*

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

$$\frac{dg(\theta)}{d\theta}\left(\theta U'(g(\theta)) - \beta W'(x(\theta))\right) + \beta\delta\frac{dV(\theta)}{d\theta} = 0.$$

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

$$\lim_{\theta\uparrow\theta'}\left\{\theta' U(g(\theta)) + \beta W(x(\theta)) + \beta\delta V(\theta)\right\} = \lim_{\theta\downarrow\theta'}\left\{\theta' U(g(\theta)) + \beta W(x(\theta)) + \beta\delta V(\theta)\right\}.$$

By the local private information constraints, government welfare is piecewise continuously differentiable. Moreover, these constraints imply that the derivative of government welfare with respect to θ is $U(g(\theta))$. Hence, in an incentive compatible rule, government welfare for type $\theta \in \Theta$ satisfies

$$\theta U(g(\theta)) + \beta W(x(\theta)) + \beta\delta V(\theta) = \underline{\theta} U(g(\underline{\theta})) + \beta W(x(\underline{\theta})) + \beta\delta V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} U(g(\widetilde{\theta}))d\widetilde{\theta}. \quad (17)$$

Substitution of (17) into the social welfare function in (11) implies that social welfare (normalized by debt) can be written as

$$\frac{1}{\beta}\underline{\theta} U(g(\underline{\theta})) + W(x(\underline{\theta})) + \delta V(\underline{\theta}) + \frac{1}{\beta}\int_{\underline{\theta}}^{\overline{\theta}} U(g(\theta))Q(\theta)d\theta, \quad (18)$$

[24] Also, if the problem in (11) (respectively, (16)) admits multiple solutions that differ only on a countable set of types, we select the solution that maximizes (respectively, minimizes) social welfare for those types.

where

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

This formulation will be useful for our characterization of the optimal self-enforcing fiscal rule in the next section, which will appeal to properties of the function $Q(\theta)$. For intuition, note that since the government is biased towards overspending relative to society, higher levels of spending can be attributed to larger spending distortions. In this sense, $Q(\theta)$ represents the weight that society places on allowing spending distortions by a government of type θ: the higher $Q(\theta)$, the lower the social welfare cost of distorting type θ's spending. The shape of this function will tell us how society wishes to allocate distortions across different government types.

4 Maximally Enforced Deficit Limits

We characterize the optimal self-enforcing fiscal rule by solving the recursive problem (11)-(15). We begin in this section by taking the set of feasible continuation values $\left[\underline{V}, \overline{V}\right]$ as given, where we assume, for the problem to be interesting, that $\overline{V} > \underline{V}$.[25] We show that the unique optimal rule is a deficit limit with maximal enforcement. The continuation equilibrium induced by this rule is examined in Section 5.

We define a maximally enforced deficit limit as follows:

Definition 1. $\{g(\theta), x(\theta), V(\theta)\}_{\theta\in\Theta}$ *is a maximally enforced deficit limit if there exist* $\theta^* \in \left[0, \overline{\theta}\right)$ *and finite* $\theta^{**} > \max\{\theta^*, \underline{\theta}\}$ *such that*

$$\{g(\theta), V(\theta)\} = \begin{cases} \left\{\min\{g^f(\theta), g^f(\theta^*)\}, \overline{V}\right\} & \text{if } \theta \leq \theta^{**}, \\ \left\{g^f(\theta), \underline{V}\right\} & \text{if } \theta > \theta^{**}, \end{cases} \tag{19}$$

where

$$\theta^{**}U(g^f(\theta^*)) + \beta W(x^f(\theta^*)) + \beta\delta\overline{V} = \theta^{**}U(g^f(\theta^{**})) + \beta W(x^f(\theta^{**})) + \beta\delta\underline{V}. \tag{20}$$

Figure 1 illustrates the spending allocation (top panel) and continuation value (bottom panel) under a maximally enforced deficit limit with $\theta^* > \underline{\theta}$ and $\theta^{**} < \overline{\theta}$. Under this rule, types $\theta \in [\underline{\theta}, \theta^*)$ and $\theta \in \left(\theta^{**}, \overline{\theta}\right]$ choose their flexible spending rate $g^f(\theta)$ and types

[25] If it were the case that $\overline{V} = \underline{V}$, then the unique equilibrium would entail all government types choosing their flexible spending rate $g^f(\theta)$ at all dates. In the Online Appendix, we provide a sufficient condition for $\overline{V} > \underline{V}$ to hold under the assumptions maintained for our main result in Proposition 2. This condition amounts to the discount factor $\delta \in (0,1)$ being high enough.

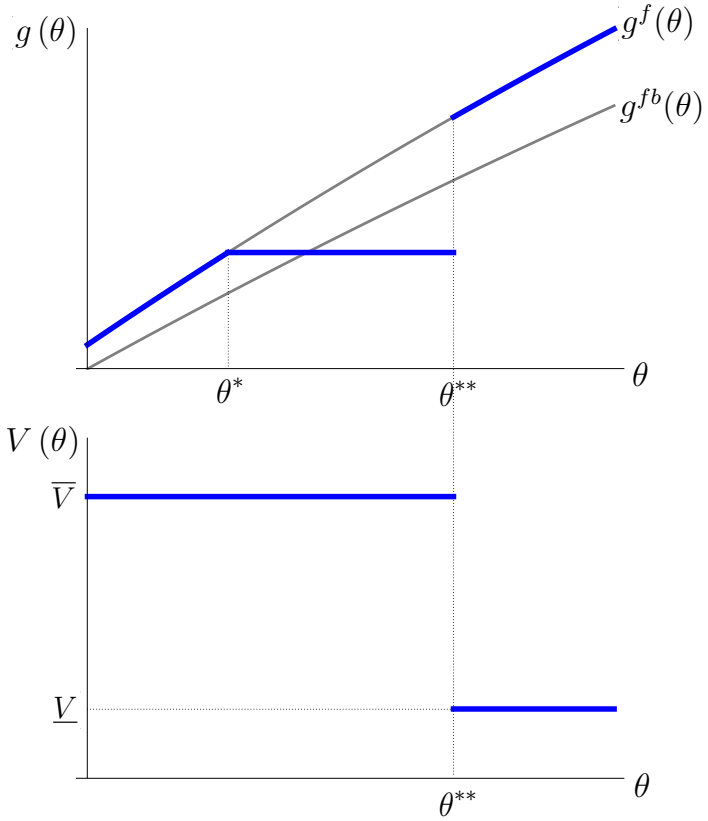

Figure 1: Spending allocation (top panel) and continuation value (bottom panel) under a maximally enforced deficit limit.

$\theta \in [\theta^*, \theta^{**}]$ choose type θ^*'s flexible spending rate $g^f(\theta^*)$. Furthermore, types $\theta \leq \theta^{**}$ are maximally rewarded with continuation value $\overline{V}$ whereas types $\theta > \theta^{**}$ are maximally punished with continuation value $\underline{V}$. As shown in equation (20), the self-enforcement constraint holds with equality for type θ^{**}. It is immediate that this rule satisfies the private information constraint (12) and the self-enforcement constraint (13).

The fiscal rule described in Definition 1 can be implemented using a maximum deficit limit, spending limit, or debt limit, where this limit would be associated with the spending rate $g^f(\theta^*)$. If the limit is satisfied, the government receives maximal reward $\overline{V}$; if the limit is breached, the government receives maximal punishment $\underline{V}$. Note that the limit is breached along the equilibrium path if and only if $\theta^{**} < \overline{\theta}$; we will provide conditions under which this inequality holds in an optimal deficit limit.

To establish our results, we proceed as follows. First, we show in Subsection 4.1 that any solution to (11)-(15) must feature bang-bang continuation values, so the rule provides high-powered incentives for the government not to overspend. This result relies only on generic properties of the function $Q(\theta)$ that weighs spending distortions in the social welfare

representation in (18). Next, we show in Subsection 4.2 that under additional assumptions on $Q(\theta)$, optimal bang-bang incentives must be monotonic, with higher types receiving weakly lower continuation value than lower types. This facilitates our characterization of optimal spending allocations in Subsection 4.3, which shows that any solution to (11)-(15) is a maximally enforced deficit limit. We further establish that the optimal limit is unique, and provide a necessary and sufficient condition for the government to violate the limit following high enough shocks. Finally, in Subsection 4.4, we show that our assumptions on $Q(\theta)$ are not only sufficient but also necessary for any solution to (11)-(15) to be a maximally enforced deficit limit.

4.1 Bang-Bang Incentives

Society can use continuation values as rewards and punishments to discipline the government and reduce overspending. The next proposition shows that in any optimal rule, these rewards and punishments are extreme. That is, continuation values are bang-bang: given a feasible set $[\underline{V}, \overline{V}]$, along the equilibrium path $V(\theta)$ only travels to the extreme points in this set.

Proposition 1 (necessity of bang-bang)**.** *Suppose $Q(\theta)$ satisfies the following generic properties:*

(i) $Q'(\theta) \neq 0$ *almost everywhere;*

(ii) If $\underline{\theta} \leq \theta^L < \theta^H \leq \overline{\theta}$ *and* $Q(\theta^L) = Q(\theta^H) = \widehat{Q}$*, then* $\int_{\theta^L}^{\theta^H} Q(\theta)d\theta \neq \int_{\theta^L}^{\theta^H} \widehat{Q}d\theta$.

Then in any solution to (11)-(15), for all $\theta \in (\underline{\theta}, \overline{\theta})$*,* $V(\theta) \in \left\{\underline{V}, \overline{V}\right\}$ *and* $V(\theta)$ *is left- or right-continuous at* θ.

Proposition 1 shows that the bang-bang property is *necessary* for social welfare maximization. As in other repeated game settings, an optimal fiscal rule using only extreme continuation values always exists in our framework; this is true simply because an interior continuation value $V(\theta)$ can be assigned in expectation by randomizing over $\underline{V}$ and $\overline{V}$. Proposition 1 proves a stronger result: any rule that prescribes interior continuation values is strictly dominated by one with high-powered incentives.

This proposition is consistent with the work of Abreu, Pearce, and Stacchetti (1990), who provide general conditions under which the bang-bang property is necessary for optimality in environments of repeated moral hazard (see Theorem 7, p.1055 in their paper). There are important differences, however, between our result and theirs. On the one hand,

establishing the necessity of bang-bang continuation values is facilitated in our self-control setting by the fact that the set of continuation values is one-dimensional, unlike in their model. On the other hand, new difficulties arise when establishing the bang-bang property in our setting of repeated adverse selection as opposed to moral hazard. Under moral hazard, local perturbations that spread out continuation values around a given public realization of uncertainty are always incentive feasible and socially beneficial, as they induce more desirable actions.[26] In contrast, under adverse selection, local perturbations that spread out continuation values can reduce social welfare (for example, if $Q'(\theta) < 0$); moreover, even when such perturbations are socially beneficial (for example, if $Q'(\theta) > 0$), they may violate the monotonicity of the spending schedule required by incentive compatibility.

Despite these differences, the intuition for our bang-bang result is related to that in Abreu, Pearce, and Stacchetti (1990) in that it stems from the richness of the information structure.[27] Given a continuum of types, this richness is guaranteed by conditions (i) and (ii) in Proposition 1, and it implies that society can always reduce distortions by steepening incentives. Condition (i) states that over any sufficiently small interval, $Q(\theta)$ is either strictly decreasing or strictly increasing; as a result, society benefits from moving spending distortions towards either lower types or higher types, respectively. Condition (ii) extends this property: if $Q(\theta)$ takes the same value at two endpoints of an interval—so that society weighs spending distortions at these endpoints equally—this condition requires that the average weight on distortions over the interval be different from that of the endpoints. That is, society strictly prefers to concentrate distortions within the interval or at an endpoint. We note that given $f(\theta)$ continuously differentiable, (i) and (ii) are generic properties of $\{f(\theta), \beta\}$ pairs which characterize $Q(\theta)$ for $\theta \in \Theta$.[28]

It is also worth noting that the result in Proposition 1 does not rely on the presence of self-enforcement constraints and would continue to hold absent (13). As such, our bang-bang characterization of continuation values applies more generally to other self-control problems. One example is the problem of Amador, Werning, and Angeletos (2006) under our Assumption 1 on preferences; this assumption ensures that spending is interior, as

[26] For example, in a standard principal-agent model with linear preferences over consumption and continuous disutility of effort, spreading out continuation values around a given realization of output allows to induce higher effort.

[27] See Yared (2010a) for a related discussion in a repeated adverse selection model.

[28] Condition (i) fails only if $\theta f'(\theta)/f(\theta) = -(2-\beta)/(1-\beta)$ for a positive mass of types, but then any arbitrarily small perturbation of β would render the condition true. Similarly, condition (ii) fails only if $Q(\theta^L) = Q(\theta^H) = \widehat{Q}$ and $\int_{\theta^L}^{\theta^H} Q(\theta)d\theta = \int_{\theta^L}^{\theta^H} \widehat{Q}d\theta$ for some $\theta^L, \theta^H \in \Theta$. However, the condition would then hold for any arbitrarily small perturbation of $f(\theta)$ over (θ^L, θ^H) which preserves the value of $\int_{\theta^L}^{\theta^H} f(\theta)d\theta$ but not that of $\int_{\theta^L}^{\theta^H} F(\theta)d\theta$.

required for our characterization.[29] We conjecture that the arguments may also extend to the monetary policy model of Athey, Atkeson, and Kehoe (2005).

The formal proof of Proposition 1 in the Appendix makes use of perturbation arguments.[30] We establish the result in three steps; we next provide a summary which may give intuition and serve as a guide to follow the proof.

Step 1 shows that $V(\theta)$ is a step function, so dynamic incentives are not provided locally to the government. Suppose instead that $V(\theta)$ was strictly increasing or strictly decreasing, and thus strictly interior, over an interval $[\theta^L, \theta^H]$. By Lemma 2, $g(\theta)$ must be strictly increasing over the interval, and by condition (i) in Proposition 1, we can take a subinterval such that $Q'(\theta) > 0$ or $Q'(\theta) < 0$ for all $\theta \in [\theta^L, \theta^H]$.[31] We then show that there exists an incentive feasible perturbation that strictly increases social welfare. To illustrate, take $g(\theta) < g^f(\theta)$, and hence $V'(\theta) < 0$, for all $\theta \in [\theta^L, \theta^H]$. If $Q'(\theta) < 0$, we construct a *flattening perturbation* that rotates the increasing $g(\theta)$ schedule clockwise over $[\theta^L, \theta^H]$, which entails reducing dynamic incentives with a counterclockwise rotation of the decreasing $V(\theta)$ function. This perturbation is socially beneficial because, given $Q'(\theta) < 0$, society prefers to concentrate spending distortions on lower rather than higher types. If instead $Q'(\theta) > 0$, we construct a *steepening perturbation* that drills a hole in the $g(\theta)$ schedule by making allocations in (θ^L, θ^H) no longer available, which entails increasing dynamic incentives by moving interior continuation values towards $V(\theta^L)$ or $V(\theta^H)$. This perturbation is socially beneficial because, given $Q'(\theta) > 0$, society prefers to concentrate spending distortions on higher rather than lower types. We obtain that $V(\theta)$ must be a step function, and we also show that $V(\theta)$ must be left- or right-continuous at each $\theta \in (\underline{\theta}, \overline{\theta})$.

Step 2 of the proof establishes that $V(\theta) \in \{\underline{V}, \overline{V}\}$ at any point θ at which $\frac{dg(\theta)}{d\theta} > 0$. Since, by Step 1, local dynamic incentives are not provided, we show that any such point θ must belong to an interval with flexible spending and constant continuation value $V \in [\underline{V}, \overline{V}]$. However, if $V \in (\underline{V}, \overline{V})$, we can construct incentive feasible perturbations similar to those above: if $Q'(\theta) < 0$ over the interval, we perform a flattening perturbation that rotates the spending schedule clockwise; if $Q'(\theta) > 0$, we perform a steepening perturbation that drills a hole in the spending schedule. By the logic in Step 1, these perturbations strictly increase social welfare, so $V \in (\underline{V}, \overline{V})$ cannot be optimal.

[29] Ambrus and Egorov (2013) provide examples in the setting of Amador, Werning, and Angeletos (2006) in which spending is at a corner and, as a result, interior punishments can be optimal.

[30] Some of the arguments we use for regions where $Q'(\theta) < 0$ are similar to those employed by Athey, Atkeson, and Kehoe (2005) in their analysis of optimal inflation rules. Unlike in their work, where rules are perfectly enforced, our arguments take into account the constraints due to self-enforcement.

[31] Note that given $f(\theta)$ continuously differentiable, condition (i) in Proposition 1 implies that the set of types θ such that $Q'(\theta) = 0$ is nowhere dense.

Step 3 completes the proof by showing that $V(\theta) \in \{\underline{V}, \overline{V}\}$ at any point θ at which $\frac{dg(\theta)}{d\theta} = 0$. By the results in Step 1 and Step 2, if $V(\theta) \in (\underline{V}, \overline{V})$ for such a point, then this point belongs to a stand-alone segment $[\theta^L, \theta^H]$ with $g(\theta) = g$ and $V(\theta) = V$ for all $\theta \in [\theta^L, \theta^H]$, for some $g > 0$ and $V \in (\underline{V}, \overline{V})$, and with $g(\theta)$ jumping at each boundary (unless $\theta^L = \underline{\theta}$ or $\theta^H = \overline{\theta}$). We then show that there exists an incentive feasible perturbation that changes g and V slightly and strictly increases social welfare. This perturbation uses the fact that condition (ii) in Proposition 1 holds. For example, if it is the case that $\int_{\theta^L}^{\theta^H} Q(\theta)d\theta > Q(\theta^L)$, we perform a segment-shifting steepening perturbation: we increase g and change V so as to leave the government welfare of type θ^H unchanged, thus increasing spending for types $\theta \in (\theta^L, \theta^H]$ and letting type θ^L jump down to a lower spending rate. This perturbation is socially beneficial because, given $\int_{\theta^L}^{\theta^H} Q(\theta)d\theta > Q(\theta^L)$, society prefers to concentrate spending distortions on $\theta \in (\theta^L, \theta^H]$ compared to θ^L. Analogous perturbations apply to other instances of condition (ii): if $\int_{\theta^L}^{\theta^H} Q(\theta)d\theta < Q(\theta^L)$, we perform a segment-shifting flattening perturbation where we reduce g; if $Q(\theta^H) \neq \int_{\theta^L}^{\theta^H} Q(\theta)d\theta$, we construct segment-shifting steepening and flattening perturbations where we keep the government welfare of type θ^L unchanged.

4.2 Monotonic Incentives

To further characterize the schedule of continuation values $V(\theta)$, and solve for the optimal rule in the next subsection, we make the following assumption:

Assumption 2. *There exists $\widehat{\theta} \in \Theta$ such that*

$$\theta \frac{f'(\theta)}{f(\theta)} > -\frac{2-\beta}{1-\beta} \quad \text{if } \theta < \widehat{\theta}, \text{ and} \tag{21}$$

$$\theta \frac{f'(\theta)}{f(\theta)} < -\frac{2-\beta}{1-\beta} \quad \text{if } \theta > \widehat{\theta}. \tag{22}$$

Assumption 2 states that $Q(\theta)$ is strictly decreasing for $\theta < \widehat{\theta}$ and strictly increasing for $\theta > \widehat{\theta}$, with a minimum value $Q(\widehat{\theta}) \leq Q(\overline{\theta}) < 0$. Note that the assumption allows for either of the inequalities in (21) or (22) to hold for all $\theta \in \Theta$; in this case $\widehat{\theta}$ is defined as either the upper bound or the lower bound of the set Θ, respectively. Assumption 2 satisfies properties (i) and (ii) in Proposition 1 and holds for a broad range of distributions, including the uniform, exponential, log-normal, and gamma distributions. This assumption is similar to, but stronger than, the distributional assumption used in Amador, Werning, and Angeletos (2006); we will show in Subsection 4.4 that Assumption 2 is necessary for our characterization of optimal self-enforcing rules.

We maintain Assumption 2 for the remainder of our analysis. The next lemma shows that, given the implied shape of $Q(\theta)$, optimal incentives are monotonic.

Lemma 3. *In any solution to* (11)-(15)*, $V(\theta)$ is weakly decreasing, left-continuous at $\theta = \overline{\theta}$, and right-continuous at $\theta = \underline{\theta}$ with $V(\underline{\theta}) = \overline{V}$.*

Together with the bang-bang property established in Proposition 1, this lemma implies that either the continuation value is at its maximum level $\overline{V}$ for all types $\theta \in \Theta$, or it jumps down from $\overline{V}$ to $\underline{V}$ at some interior point $\theta \in \Theta$. The intuition for this result is related to the shape of the $Q(\theta)$ function that tells us how society wishes to allocate spending distortions across types. If $\theta > \widehat{\theta}$, society benefits from concentrating spending distortions on relatively high types; this is achieved by using high-powered incentives that punish high levels of spending by lowering the continuation value from $\overline{V}$ to $\underline{V}$. In contrast, if $\theta < \widehat{\theta}$, society benefits from concentrating spending distortions on relatively low types; this is achieved by using low-powered incentives that keep the continuation value constant at $\overline{V}$.

The proof of Lemma 3 in the Appendix proceeds in three steps, which we briefly summarize here. Step 1 shows that any interval with continuation value $\underline{V}$ must lie above $\widehat{\theta}$. To see why, note that the self-enforcement constraint requires that all types in the interval spend flexibly. Then if $Q'(\theta) < 0$ for such types, we can perform an incentive feasible flattening perturbation that rotates the spending schedule $g(\theta)$ clockwise; by the logic in Step 1 in the proof of Proposition 1, this perturbation is socially beneficial. Hence, we must have $Q'(\theta) > 0$, and by Assumption 2 the interval must lie above $\widehat{\theta}$.

Step 2 shows that if $V(\theta^{**}) = \underline{V}$ for some type $\theta^{**} \geq \widehat{\theta}$, then $V(\theta) = \underline{V}$ for all $\theta \geq \theta^{**}$. Suppose by contradiction that $V(\theta) = \overline{V}$ for some $\theta > \theta^{**}$. Extending the continuity result established in Proposition 1 to type $\overline{\theta}$ yields that there exists an interval $[\theta^L, \theta^H]$, $\theta^L > \theta^{**}$, with $V(\theta) = \overline{V}$ for all θ in the interval. However, if $\frac{dg(\theta)}{d\theta} > 0$ for $\theta \in [\theta^L, \theta^H]$, we can perform an incentive feasible steepening perturbation that drills a hole in the $g(\theta)$ schedule; by the logic in Step 1 in the proof of Proposition 1 and $Q'(\theta) > 0$, this perturbation is socially beneficial. Moreover, if $g(\theta)$ is constant over a stand-alone segment $[\theta^L, \theta^H]$, we can perform an incentive feasible segment-shifting steepening perturbation that slightly reduces spending and changes the continuation value so as to leave type θ^L equally well off; by the logic in Step 3 in the proof of Proposition 1 and $Q'(\theta) > 0$, this perturbation is socially beneficial. Thus, a segment $[\theta^L, \theta^H]$ with continuation value $\overline{V}$ and $\theta^L > \theta^{**}$ cannot exist, proving the claim.

Given any type θ^{**} as just defined, Step 3 completes the proof by showing that $\theta^{**} > \underline{\theta}$. If this were not the case, we would have $V(\theta) = \underline{V}$ for all types $\theta \in \Theta$. However, a global perturbation that increases the continuation value for all types would then be incentive

feasible and strictly increase social welfare.

4.3 Optimal Self-Enforcing Fiscal Rule

The following proposition states the main result of the paper:

Proposition 2 (optimal self-enforcing fiscal rule)**.** *If* $\{g(\theta), x(\theta), V(\theta)\}_{\theta\in\Theta}$ *is a solution to* (11)-(15)*, then there exist* $\theta^* \in [0, \overline{\theta})$ *and finite* $\theta^{**} > \max\{\theta^*, \underline{\theta}\}$ *satisfying* (19)-(20)*. Hence, any optimal self-enforcing rule is a maximally enforced deficit limit.*

By Proposition 1 and Lemma 3, either all government types $\theta \in \Theta$ receive the highest continuation value $\overline{V}$, or the continuation value jumps down from $\overline{V}$ to $\underline{V}$ at a point $\theta^{**} \in (\underline{\theta}, \overline{\theta})$. To establish Proposition 2, we take θ^{**} as given and solve for the optimal allocation above and below this point. If $\theta \in (\theta^{**}, \overline{\theta}]$, the allocation is characterized by the binding self-enforcement constraint with $\{g(\theta), V(\theta)\} = \{g^f(\theta), \underline{V}\}$. If $\theta \in [\underline{\theta}, \theta^{**}]$, we show that the spending schedule $g(\theta)$ must be continuous, so the allocation takes the form of bounded discretion. Since a minimum spending requirement would reduce social welfare, a maximum spending limit $g^f(\theta^*)$ is optimal for types in this range.

The proof that $g(\theta)$ is continuous over $[\underline{\theta}, \theta^{**}]$ builds on perturbation arguments similar to those used in the previous subsections. Given the absence of dynamic incentives, a discontinuity would entail a hole in some range $[\theta^L, \theta^H]$; that is, $g(\theta)$ would jump at a point $\theta^M \in (\theta^L, \theta^H)$ such that types $\theta \in (\theta^L, \theta^M)$ spend at $g^f(\theta^L) < g^f(\theta)$ and types $\theta \in (\theta^M, \theta^H)$ spend at $g^f(\theta^H) > g^f(\theta)$. However, if $\theta^M < \widehat{\theta}$, an incentive feasible perturbation that slightly closes the hole by raising θ^L strictly increases social welfare: given $Q'(\theta) < 0$, the benefit of lowering spending for types right above θ^M outweighs the cost of raising spending for types right above θ^L. Moreover, if $\theta^M \geq \widehat{\theta}$, then $(\theta^M, \theta^H]$ is a stand-alone segment over which $g(\theta) = g$ and $Q'(\theta) > 0$, and by the logic in Step 2 in the proof of Lemma 3, an incentive feasible segment-shifting steepening perturbation strictly increases social welfare. It follows that $g(\theta)$ cannot be discontinuous below θ^{**}.

Proposition 2 proves that any optimal rule must be a maximally enforced deficit limit, but it is silent on whether this limit is violated along the equilibrium path, that is, whether $\theta^{**} < \overline{\theta}$. To address this issue, define $\theta_e \in [0, \overline{\theta}]$ as the threshold type corresponding to the optimal deficit limit under perfect enforcement, as in the work of Amador, Werning, and Angeletos (2006). Given Assumption 2, θ_e is the unique solution to[32]

$$\int_{\theta_e}^{\overline{\theta}} Q(\theta) d\theta = 0. \tag{23}$$

[32]Note that $Q(\theta) = 1$ for $\theta < \underline{\theta}$.

Clearly, if a deficit limit associated with maximum spending rate $g^f(\theta_e)$ is self-enforcing, then by Proposition 2 it is optimal:

Corollary 1. *If*

$$\overline{\theta}U(g^f(\theta_e)) + \beta W(x^f(\theta_e)) + \beta\delta\overline{V} \geq \overline{\theta}U(g^f(\overline{\theta})) + \beta W(x^f(\overline{\theta})) + \beta\delta\underline{V}, \tag{24}$$

the unique solution to (11)-(15) *is a maximally enforced deficit limit with* $\theta^* = \theta_e$ *and* $\theta^{**} \geq \overline{\theta}$.

When condition (24) holds, the highest type $\overline{\theta}$, and therefore all types $\theta \in \Theta$, prefer to respect the perfect-enforcement limit $g^f(\theta_e)$ and receive maximal reward $\overline{V}$ rather than spend above the limit and receive maximal punishment $\underline{V}$. The optimal self-enforcing rule therefore coincides with that under perfect enforcement and features no dynamic incentives.

Our interest is in characterizing the optimal rule when condition (24) does not hold, so the perfect-enforcement limit $g^f(\theta_e)$ is not self-enforcing. In this case, there exists a unique type $\theta_b > \theta_e$ corresponding to the tightest deficit limit that all types $\theta \in \Theta$ would respect:

$$\overline{\theta}U(g^f(\theta_b)) + \beta W(x^f(\theta_b)) + \beta\delta\overline{V} = \overline{\theta}U(g^f(\overline{\theta})) + \beta W(x^f(\overline{\theta})) + \beta\delta\underline{V}. \tag{25}$$

The next proposition provides a necessary and sufficient condition for punishments to be used along the equilibrium path when (24) is not satisfied:

Proposition 3 (use of punishment)**.** *Suppose* (24) *does not hold, i.e. the optimal deficit limit under perfect enforcement is not self-enforcing. If*

$$\int_{\theta_b}^{\overline{\theta}} \left(Q(\theta) - Q(\overline{\theta})\right) d\theta \geq 0, \tag{26}$$

the unique solution to (11)-(15) *is a maximally enforced deficit limit with* $\theta^* = \theta_b$ *and* $\theta^{**} = \overline{\theta}$. *Otherwise, the unique solution to* (11)-(15) *is a maximally enforced deficit limit with* $\theta^* \in (\theta_e, \theta_b)$ *and* $\theta^{**} < \overline{\theta}$.

When the perfect-enforcement limit $g^f(\theta_e)$ is not self-enforcing, society faces the following tradeoff. On the one hand, society can raise the value of θ^* to the point that the associated limit $g^f(\theta^*)$ satisfies the self-enforcement constraint of type $\overline{\theta}$ and thus all types $\theta \in \Theta$. This option entails setting $\theta^* = \theta_b$ and $\theta^{**} = \overline{\theta}$ and has the benefit of avoiding socially costly punishments along the equilibrium path, albeit at the cost of allowing significant overspending within the relaxed deficit limit. On the other hand, society can impose

a tighter deficit limit $g^f(\theta^*)$ which does not satisfy the self-enforcement constraint of all types. This option sets $\theta^* < \theta_b$ and $\theta^{**} < \overline{\theta}$ and induces higher discipline on types $\theta \leq \theta^{**}$, but at the cost of imposing punishments on path whenever a shock $\theta > \theta^{**}$ is realized.

Proposition 3 shows that which of these two options is optimal for society depends on whether condition (26) holds or not. The intuition for this condition is familiar by now: it tells us how society wishes to allocate spending distortions, and in particular whether society prefers to concentrate distortions on types $\theta \in [\theta_b, \overline{\theta})$ versus $\overline{\theta}$. A relaxed deficit limit that avoids punishments concentrates distortions on $\theta \in [\theta_b, \overline{\theta})$, whereas tightening the deficit limit by the use of punishments moves distortions towards $\overline{\theta}$.

Taking feasible continuation values as given, condition (26) is determined by the distribution of shocks and the government's present bias. Specifically, (26) holds if high shocks are relatively frequent (so $f(\overline{\theta})$ and thus $-Q(\overline{\theta}) > 0$ are high) and the government's present bias is not too severe (so $\int_{\theta_b}^{\overline{\theta}} Q(\theta)d\theta < 0$ is high). In this case—which would arise, for example, if $\widehat{\theta} = \overline{\theta}$, such as under a uniform distribution of shocks—imposing punishments on high types is very costly, and, moreover, society can satisfy these types' self-enforcement constraints without relaxing the deficit limit by much. The optimal rule therefore sets a relaxed deficit limit that all types respect, namely with $\theta^* = \theta_b$ and $\theta^{**} = \overline{\theta}$.

In contrast, if high shocks are relatively rare and the government's present bias is severe, condition (26) does not hold. In this case—which would arise under $\widehat{\theta} < \overline{\theta}$ and $f(\overline{\theta})$ close enough to zero—the cost of punishing high types is relatively low, and, moreover, society would need to raise the limit $g^f(\theta^*)$ significantly if it were to satisfy the self-enforcement constraints of all types. Hence, it is optimal to set a tighter deficit limit, and let a positive mass of government types $\theta \in (\theta^{**}, \overline{\theta}]$ violate the limit and receive maximal punishment on the equilibrium path.[33]

Given the properties of $Q(\theta)$, Proposition 3 shows that there exist a unique threshold θ^* and associated θ^{**} satisfying (20) that optimally resolve the tradeoff between imposing fiscal discipline and avoiding punishments on path. As a result, the deficit limit that maximizes social welfare is unique.

4.4 Discussion of Distributional Assumption

The results in Subsection 4.3 show that Assumption 2 is sufficient to obtain the unique optimality of maximally enforced deficit limits for any given continuation values $\{\underline{V}, \overline{V}\}$. In this section, we explore the necessity of this distributional assumption for our findings.

[33] While Proposition 3 conditions on a given set of feasible continuation values $[\overline{V}, \underline{V}]$, one can derive from this result sufficient conditions for punishment to occur or not along the equilibrium path regardless of these values. See Subsection 5.2 for a discussion.

Definition 2. *Assumption 2 is* weakly violated *if there exist* $\theta^L, \theta^H \in \Theta$, $\theta^H > \theta^L$, *and* $\Delta > 0$ *such that (i)* $Q'(\theta) \geq 0$ *for* $\theta \in [\theta^L, \theta^L + \Delta]$ *and (ii)* $Q'(\theta) \leq 0$ *for* $\theta \in [\theta^H - \Delta, \theta^H]$. *Assumption 2 is* strictly violated *if the inequalities in (i) and (ii) are strict.*

We find that both weak and strict violations of Assumption 2 would affect our results:

Proposition 4 (necessity of distributional assumption)**.** *If Assumption 2 is weakly violated, then there exist continuation values* $\{\underline{V}, \overline{V}\}$ *for which not every solution to* (11)-(15) *is a maximally enforced deficit limit. Moreover, if Assumption 2 is strictly violated, then there exist continuation values* $\{\underline{V}, \overline{V}\}$ *for which no solution to* (11)-(15) *is a maximally enforced deficit limit.*

This proposition shows that Assumption 2 is not only sufficient but also necessary for maximally enforced deficit limits to be uniquely optimal given any continuation values $\{\underline{V}, \overline{V}\}$. In this sense, our analysis imposes the minimal structure that guarantees the unique optimality of this class of rules, which resemble fiscal rules commonly used in practice in the form of a deficit limit or a spending or debt limit. Note that both weak and strict violations of Assumption 2 are possible even when $Q(\theta)$ satisfies properties (i) and (ii) in Proposition 1. Under these violations, an optimal self-enforcing rule may thus feature bang-bang incentives yet induce an allocation that would not be implementable by a deficit limit. We prove the first part of Proposition 4 by construction and the second part by contradiction.[34]

5 Punishment and Dynamics

Section 4 characterized the optimal self-enforcing fiscal rule taking the set of feasible continuation values as given. In this section, we complete our characterization by describing the allocations that underly the continuation values used in an optimal rule.

5.1 Worst Punishment

The worst punishment for the government, given by the continuation value $\underline{V}$, determines the tightness of the self-enforcement constraints in (13). Moreover, as shown in Proposition 3, this punishment is induced along the equilibrium path whenever the optimal deficit

[34] As noted in Subsection 4.2, Assumption 2 is stronger than the distributional assumption used in Amador, Werning, and Angeletos (2006). Define θ_a as the lowest value such that $\int_{\theta}^{\overline{\theta}} Q(\widetilde{\theta}) d\widetilde{\theta} \leq 0$ for all $\theta \geq \theta_a$. Then using our notation and taking $f(\theta)$ to be differentiable, Amador, Werning, and Angeletos (2006) assume $Q'(\theta) \leq 0$ for all $\theta \leq \theta_a$. One can verify that there are distributions $F(\theta)$ satisfying this assumption for which Assumption 2 is strictly violated, and therefore for which there exist continuation values $\{\underline{V}, \overline{V}\}$ such that no solution to (11)-(15) is a maximally enforced deficit limit.

limit under perfect enforcement is not self-enforcing and condition (26) is not satisfied. The next proposition describes the allocation underlying $\underline{V}$, which solves the program in (16).

Proposition 5 (characterization of punishment)**.** *If* $\{g(\theta), x(\theta), V(\theta)\}_{\theta\in\Theta}$ *is a solution to* (16) *subject to* (12)-(15)*, then there exist finite* $\theta_p^* > \underline{\theta}$ *and* $\theta_p^{**} \in \left[\underline{\theta}, \min\left\{\theta_p^*, \overline{\theta}\right\}\right)$ *satisfying*

$$\{g(\theta), V(\theta)\} = \begin{cases} \left\{\max\{g^f(\theta), g^f(\theta_p^*)\}, \overline{V}\right\} & \text{if } \theta \geq \theta_p^{**}, \\ \{g^f(\theta), \underline{V}\} & \text{if } \theta < \theta_p^{**}, \end{cases}$$

where

$$\theta_p^{**} U(g^f(\theta_p^*)) + \beta W(x^f(\theta_p^*)) + \beta\delta\overline{V} = \theta_p^{**} U(g^f(\theta_p^{**})) + \beta W(x^f(\theta_p^{**})) + \beta\delta\underline{V}.$$

Hence, the worst punishment is a maximally enforced surplus limit. Moreover, the worst punishment is unique.

In the absence of self-enforcement constraints, the worst punishment would entail forcing all government types to choose either the highest or the lowest feasible spending rate so that the value $\underline{V}$ is minimized. However, such a harsh punishment would not be self-enforcing. Proposition 5 shows that the worst punishment that is self-enforcing takes the form of a maximally enforced surplus limit, associated with a minimum spending rate $g^f(\theta_p^*)$. Government types that respect the surplus limit by spending above $g^f(\theta_p^*)$ are maximally rewarded with continuation value $\overline{V}$; government types that violate the surplus limit by spending below $g^f(\theta_p^*)$ are maximally punished with continuation value $\underline{V}$. Because a positive mass of types $\theta \geq \theta_p^{**}$ respect the limit, the equilibrium transitions back to the allocation associated with value $\overline{V}$ with strictly positive probability.

A maximally enforced surplus limit minimizes social welfare by incentivizing overspending. Intuitively, given a government type θ, there are two ways in which society can induce low welfare: either by inducing spending below the first-best rate $g^{fb}(\theta)$, or by inducing spending above this rate. Since the government is biased towards overspending in the present, the latter relaxes self-enforcement constraints, and it is thus a more efficient means of reducing welfare. As a result, in the worst-punishment allocation, all types spend above their first-best rate, and in fact weakly above their flexible rate. Moreover, consistent with Proposition 1, overspending is incentivized by the use of bang-bang continuation values: society maximally rewards government types that spend above the surplus limit and maximally punishes those that spend below the limit.[35] High-powered incentives allow society to maximize distortions.

[35] These features are related to the results of Bernheim, Ray, and Yeltekin (2015), who study self-enforcing

The proof of Proposition 5 follows from analogous arguments to those used to obtain Proposition 2. The main difference, of course, is that instead of maximizing the social welfare function in (18), the worst punishment minimizes this function. As a result, the arguments that apply to types $\theta < \widehat{\theta}$ when maximizing welfare—which make use of the fact that $Q(\theta)$ in (18) is strictly decreasing in this region—instead apply to types $\theta > \widehat{\theta}$ when minimizing welfare—since $-Q(\theta)$ is strictly decreasing in this region. The same is true with regards to the arguments that apply to types $\theta > \widehat{\theta}$ in the maximization of welfare, which now apply to types $\theta < \widehat{\theta}$ in the minimization of welfare. This explains why the worst punishment is a maximally enforced surplus limit while the optimal rule is a maximally enforced deficit limit. One step that requires additional care is establishing that the optimal surplus limit is indeed respected by some government types; that is, the worst punishment is not an absorbing state with $\{g(\theta), V(\theta)\} = \{g^f(\theta), \underline{V}\}$ for all $\theta \in \Theta$. We prove this by showing that a surplus limit that is respected by types θ arbitrarily close to $\overline{\theta}$ achieves lower social welfare than assigning flexible spending and maximal punishment to all types: the cost of increasing overspending outweighs the benefit of increasing the continuation value for high enough types.

5.2 Bang-Bang Dynamics

Given our characterization in Proposition 2 and Proposition 5, the maximal reward and the worst punishment, $\{\overline{V}, \underline{V}\}$, can be shown to solve the following system:

$$\overline{V} = \max_{\theta^*, \theta^{**} > \theta^*} \left\{ \frac{1}{\beta(1-\delta)} \left(\int_0^{\theta^*} U(g^f(\theta)) Q(\theta) d\theta + \int_{\theta^*}^{\theta^{**}} U(g^f(\theta^*)) Q(\theta) d\theta + \int_{\theta^{**}}^{\overline{\theta}} U(g^f(\theta)) Q(\theta) d\theta \right) \right\} \tag{27}$$

$$\text{subject to } \int_{\theta^*}^{\theta^{**}} (U(g^f(\theta)) - U(g^f(\theta^*))) d\theta = \beta\delta(\overline{V} - \underline{V});$$

$$\underline{V} = \min_{\theta_p^*, \theta_p^{**} < \theta_p^*} \left\{ \frac{1}{\beta(1-\delta)} \left(\int_0^{\theta_p^{**}} U(g^f(\theta)) Q(\theta) d\theta + \int_{\theta_p^{**}}^{\theta_p^*} U(g^f(\theta_p^*)) Q(\theta) d\theta + \int_{\theta_p^*}^{\overline{\theta}} U(g^f(\theta)) Q(\theta) d\theta \right) \right\} \tag{28}$$

$$\text{subject to } \int_{\theta_p^{**}}^{\theta_p^*} (U(g^f(\theta_p^*)) - U(g^f(\theta))) d\theta = \beta\delta(\overline{V} - \underline{V}).$$

The objective functions above are written using an analogous representation of social

consumption rules in an environment without shocks or private information. They find that the worst punishment takes the form of over-consumption followed by maximal reward.

welfare as in (18), except that we use the form of the optimal rules obtained in Proposition 2 and Proposition 5, and we take the reference type to be arbitrarily close to zero as opposed to equal to $\underline{\theta}$ as in (18).[36] Similarly, the binding self-enforcement constraints for types θ^{**} and θ_p^{**} under the optimal deficit limit and surplus limit, respectively, are written using an analogous representation of government welfare as in (17). Programs (27) and (28) show that the larger the difference between the maximal reward and the worst punishment, $(\overline{V} - \underline{V})$, the more society can relax self-enforcement constraints, which in turn allows to sustain a higher reward and a worse punishment. An optimal self-enforcing rule yields the highest value of $\overline{V}$ and lowest value of $\underline{V}$ that solve this fixed point problem.[37]

As discussed in Subsection 4.1, the optimal self-enforcing fiscal rule features a bang-bang property, so that along the equilibrium path continuation values only travel to extreme points in the feasible set $[\underline{V}, \overline{V}]$. Given this bang-bang form, the dynamics induced by the optimal rule depend on the tightness of self-enforcement constraints and the distribution of shocks. By Corollary 1 and Proposition 3, if either the optimal deficit limit under perfect enforcement is self-enforcing or condition (26) holds, the equilibrium features no dynamics: the same deficit limit $g^f(\theta^*)$ is imposed in every date. In contrast, if the perfect-enforcement deficit limit is not self-enforcing and condition (26) does not hold, Proposition 3 and Proposition 5 imply that the economy transitions in and out of the best equilibrium associated with value $\overline{V}$ and the worst punishment associated with value $\underline{V}$. Starting from the initial period, the government is subject to a deficit limit $g^f(\theta^*)$. If the realized shock is $\theta \leq \theta^{**}$, the government respects the deficit limit and the best equilibrium restarts in the second period; if the realized shock is $\theta > \theta^{**}$, the government violates the deficit limit and the equilibrium transitions to punishment in the second period. Starting from a punishment period, the government is subject to a surplus limit $g^f(\theta_p^*)$. If the realized shock is $\theta \geq \theta_p^{**}$, the government respects the surplus limit and the equilibrium transitions to the best equilibrium in the next period; if the realized shock is $\theta < \theta_p^{**}$, the government violates the surplus limit and the equilibrium remains in punishment in the next period.

The primitives of the model determine whether self-enforcement constraints bind and

[36] Noting that $f(\theta) = 0$ (and thus $Q(\theta) = 1$) for $\theta < \underline{\theta}$ and $\lim_{\theta \downarrow 0}\{\theta U(g^f(\theta)) + \delta W(x^f(\theta))\} = 0$.

[37] The first-order conditions of programs (27) and (28) are intuitive. For example, suppose the perfect-enforcement rule is not self-enforcing and (26) does not hold, so that by Proposition 3, $\theta^{**} < \overline{\theta}$ in an optimal self-enforcing rule. Then the first-order condition of (27) yields

$$\int_{\theta^*}^{\theta^{**}} (Q(\theta) - Q(\theta^{**}))d\theta = 0.$$

Analogous to (23) and (26), this condition says that an optimal maximally enforced deficit limit balances spending distortions across government types: θ^* and θ^{**} are such that the average distortion on types $\theta \in [\theta^*, \theta^{**})$ is weighted equally as that on type θ^{**}.

condition (26) does not hold, as required for dynamic incentives to be provided to the government. Self-enforcement constraints will bind so long as the highest shock $\overline{\theta}$ is sufficiently extreme. Intuitively, as $\overline{\theta}$ increases, the difference between type $\overline{\theta}$'s per-period welfare from flexible spending, $\overline{\theta}U(g^f(\overline{\theta}))+\beta W(x^f(\overline{\theta}))$, and this type's per-period welfare from spending at the perfect-enforcement limit, $\overline{\theta}U(g^f(\theta_e))+\beta W(x^f(\theta_e))$, becomes larger, and for high enough $\overline{\theta}$ condition (24) cannot hold. Regarding (26), as discussed in Subsection 4.3, it depends on the distribution of shocks and the government's present bias whether this condition holds for a given set of continuation values. In particular, the condition will fail provided that $-Q(\overline{\theta})>0$ is low enough—which holds if high shocks are unlikely—and β is low enough—which means that tight rules are not self-enforcing under high shocks. For any continuation values $\{\underline{V},\overline{V}\}$ and parameter $\beta<1$, a sufficient condition for (26) to fail is that $f(\theta)$ approach zero as θ approaches $\overline{\theta}$.

In sum, our analysis shows that the presence of shocks that are sufficiently extreme and rare (relative to the government's present bias) is a sufficient condition for punishments to occur along the equilibrium path. The optimal deficit limit under perfect enforcement is not self-enforcing in this case, and society would have to set an excessively lax deficit limit to ensure that the government always respects it. Instead, it is optimal to impose a tighter limit which the government respects only under non-extreme shocks, and to temporarily transition to maximal punishment when extreme shocks occur.

6 Concluding Remarks

This paper has studied the role of self-enforcement in determining the optimal structure of fiscal rules. Under perfect enforcement, an optimal rule takes the form of a deficit limit which is never breached in equilibrium. Under self-enforcement, an optimal rule is a maximally enforced deficit limit, which, if violated, leads to the worst punishment for the government. We established necessary and sufficient conditions under which the government violates the deficit limit along the equilibrium path, following bad enough shocks to the economy. In this case, the economy fluctuates between periods of reward—with a maximally enforced deficit limit that constrains overspending—and periods of punishment—with a maximally enforced surplus limit that incentivizes overspending. The government enters a reward phase whenever it satisfies the imposed deficit or surplus limit, and transitions into a punishment phase whenever it violates the limit. These transitions ensure that both the maximally enforced deficit limit and the maximally enforced surplus limit are self-enforcing.

Our focus has been on a situation in which external enforcement is absent and the

optimal fiscal rule under perfect enforcement is not self-enforcing. It is worth noting that our analysis also applies to situations in which external enforcement is possible but limited. Society may be able to impose sanctions on the government beyond those that are self-enforcing, yet these may not be severe enough to guarantee perfect enforcement. Our analysis in Section 4 applies to this case without change, taking $\underline{V}$ to be the harshest possible punishment that society can impose. An optimal rule is a maximally enforced deficit limit, which, if violated, leads to the worst punishment, including externally enforced penalties. Provided that these penalties are sufficiently small, the perfect-enforcement deficit limit may not be implementable, and punishments would arise along the equilibrium path under the conditions derived in Section 4.

We believe there are potentially interesting directions for future research. We would like to explore in future work the extent of our bang-bang characterization in Proposition 1; as we have noted, our result does not rely on self-enforcement and may apply to other models of repeated adverse selection. Another possible direction would be to consider an extension of our model with more general time-inconsistent preferences. The problem would change compared to our quasi-hyperbolic setting: 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. Additionally, our setting can be enriched to consider a group of countries or subnational regions subject to a common fiscal rule which must be self-enforcing. The properties of an optimal common rule would depend on governments' self-enforcement constraints and the nature of collective punishments.

Finally, while we have focused on fiscal policy, the insights of this paper may be applied to other settings featuring a commitment-versus-flexibility tradeoff and self-enforcement. For example, consider an individual who suffers from a self-control problem and establishes rules for himself to curb his consumption of a temptation good such as television or alcohol. Such an individual values rules that impose discipline on himself, but also benefits from having flexibility to increase his consumption of the temptation good when he finds this consumption to be highly valuable. Furthermore, in the absence of external enforcement, any rule the individual imposes on himself must be self-enforcing. In this context, our results imply that an optimal rule takes the form of a consumption threshold, which the individual may violate when his value of consumption is high enough. Violation of the threshold leads to punishment in the form of over-consumption. That is, to limit his consumption of the temptation good, the individual would transition in and out of periods of self-enforcing binging.

References

Abreu, D. (1988): "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 56, 383–96.

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58, 1041–63.

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

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

Alesina, A. and R. Perotti (1995): "The Political Economy of Budget Deficits," *IMF Staff Papers*, 42, 1–31.

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

Alfaro, L. and F. Kanczuk (2016): "Fiscal Rules and Sovereign Default," Working Paper.

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

Amador, M. and K. Bagwell (2013): "The Theory of Delegation with an Application to Tariff Caps," *Econometrica*, 81, 1541–600.

——— (2016): "Regulating a Monopolist With Uncertain Costs Without Transfers," Working Paper.

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

——— (2006): "Commitment Vs. Flexibility," *Econometrica*, 74, 365–396.

Ambrus, A. and G. Egorov (2013): "Comment on 'Commitment vs. Flexibility'," *Econometrica*, 81, 2113–24.

——— (2015): "Delegation and Nonmonetary Incentives," Working Paper.

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

Athey, S., K. Bagwell, and C. Sanchirico (2004): "Collusion and Price Rigidity," *Review of Economic Studies*, 71, 317–49.

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

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

Azzimonti, M., M. Battaglini, and S. Coate (2016): "The Costs and Benefits of Balanced Budget Rules: Lessons from a Political Economy Model of Fiscal Policy," *Journal of Public Economics*, 136, 45–61.

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

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

Bernheim, B. D., D. Ray, and S. Yeltekin (2015): "Poverty and Self-Control," *Econometrica*, 83, 1877–911.

Bisin, A., A. Lizzeri, and L. Yariv (2015): "Government Policy with Time Inconsistent Voters," *American Economic Review*, 105, 1711–37.

Cabalin, C. (2011): "Why is Piñera's Government So Unpopular in Chile?" *The Guardian*, August 5.

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

Cao, D. and I. Werning (2017): "Saving and Dissaving with Hyperbolic Discounting," Working Paper.

Dong, M. (2017): "Bright-line Personal Rules: Addictive Good Consumption by A Hyperbolic Discounting Consumer," Working Paper.

Dovis, A. and R. Kirpalani (2017): "Fiscal Rules, Bailouts, and Reputation in Federal Governments," Working Paper.

Farhi, E. and I. Werning (2007): "Inequality and Social Discounting," *Journal of Political Economy*, 115, 365–402.

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, 1557–614.

——— (2017a): "Commitment vs. Flexibility with Costly Verification," Working Paper.

——— (2017b): "Fiscal Rules and Discretion in a World Economy," Working Paper.

Hatchondo, J. C., L. Martinez, and F. Roch (2017): "Fiscal Rules and the Sovereign Default Premium," Working Paper.

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

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

Hörner, J. and Y. Guo (2015): "Dynamic Mechanisms without Money," Working Paper.

Jackson, M. O. and L. Yariv (2014): "Present Bias and Collective Dynamic Choice in the Lab," *American Economic Review*, 104, 4184–204.

——— (2015): "Collective Dynamic Choice: The Necessity of Time Inconsistency," *American Economic Journal: Microeconomics*, 7, 150–78.

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

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

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

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

Larraín, F., R. Costa, R. Cerda, M. Villena, and A. Tomaselli (2011): "Una Política Fiscal de Balance Estructural de Segunda Generación para Chile," Estudios de Finanzas Públicas, Dirección de Presupuestos, Gobierno de Chile.

Li, J., N. Matouschek, and M. Powell (2017): "Power Dynamics in Organizations," *American Economic Journal: Microeconomics*, Forthcoming.

Lipnowski, E. and J. Ramos (2017): "Repeated Delegation," Working Paper.

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

Lizzeri, A. and L. Yariv (2014): "Collective Self-Control," Working Paper.

Lledó, V., S. Yoon, X. Fang, S. Mbaye, and Y. Kim (2017): "Fiscal Rules at a Glance," *International Monetary Fund*, March.

Moser, C. and P. O. de Souza e Silva (2017): "Optimal Paternalistic Savings Policies," Working Paper.

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

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

Schaechter, A., T. Kinda, N. Budina, and A. Weber (2012): "Fiscal Rules in Response to the Crisis – Toward the 'Next-Generation' Rules. A New Dataset," IMF Working Paper WP/12/187.

Schmidt-Hebbel, K. (2014): "Chile's Fiscal Policy Rule," Presentation for the Sixth Annual Meeting of OECD Parliamentary Budget Officials and Independent Fiscal Institutions.

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

Sleet, C. and S. Yeltekin (2006): "Credibility and Endogenous Societal Discounting," *Review of Economic Dynamics*, 9, 410–37.

——— (2008): "Politically Credible Social Insurance," *Journal of Monetary Economics*, 55, 129–51.

Song, Z., K. Storesletten, and F. Zilibotti (2012): "Rotten Parents and Disciplined Children: A Politico-Economic Theory of Public Expenditure and Debt," *Econometrica*, 80, 2785–803.

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

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

Velasco, A. and A. Arenas (2010): "Fiscal policy: The Unbearable Laxity of Being," *La Tercera*, Santiago, Chile, October 29.

Yared, P. (2010a): "A Dynamic Theory of War and Peace," *Journal of Economic Theory*, 145, 1921–50.

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

A Proofs

This Appendix contains the proofs of Proposition 1, Lemma 3, and Proposition 2. See the Online Appendix for the remaining proofs and supplementary discussions.

A.1 Preliminaries

We describe two functions that we will use in our proofs.

Lemma 4. *Given $\theta^L \leq \theta^M \leq \theta^H$, define the functions*

$$\begin{aligned} S^L(\theta^L, \theta^M) &= \int_{\theta^L}^{\theta^M} \left(\beta\theta - \theta^L\right) f(\theta)\, d\theta + (1-\beta)\,\theta^M f(\theta^M)(\theta^M - \theta^L), \\ S^H(\theta^H, \theta^M) &= \int_{\theta^M}^{\theta^H} \left(\beta\theta - \theta^H\right) f(\theta) d\theta + (1-\beta)\,\theta^M f(\theta^M)(\theta^H - \theta^M). \end{aligned}$$

Then $S^L(\theta^L, \theta^M) > 0$ if $Q'(\theta) < 0$ for all $\theta \in (\theta^L, \theta^M)$, $S^L(\theta^L, \theta^M) < 0$ if $Q'(\theta) > 0$ for all $\theta \in (\theta^L, \theta^M)$, $S^H(\theta^H, \theta^M) > 0$ if $Q'(\theta) > 0$ for all $\theta \in (\theta^M, \theta^H)$, and $S^H(\theta^H, \theta^M) < 0$ if $Q'(\theta) < 0$ for all $\theta \in (\theta^M, \theta^H)$.

Proof. Consider the claims about $S^L(\theta^L, \theta^M)$. Note that $S^L(\theta, \theta^M)|_{\theta=\theta^M} = 0$, and hence $S^L(\theta^L, \theta^M) = -\int_{\theta_L}^{\theta^M} \frac{dS^L(\theta,\theta^M)}{d\theta} d\theta$. Moreover,

$$\frac{dS^L(\theta, \theta^M)}{d\theta} = -\int_{\theta}^{\theta^M} f(\widetilde{\theta}) d\widetilde{\theta} + (1-\beta)\,\theta f(\theta) - (1-\beta)\,\theta^M f(\theta^M),$$

and thus $\frac{dS^L(\theta,\theta^M)}{d\theta}|_{\theta=\theta^M} = 0$. Therefore, $S^L(\theta^L, \theta^M) = \int_{\theta_L}^{\theta^M} \int_{\theta}^{\theta^M} \frac{d^2 S^L(\widetilde{\theta},\theta^M)}{d\widetilde{\theta}^2} d\widetilde{\theta} d\theta$, where

$$\frac{d^2 S^L(\theta, \theta^M)}{d\theta^2} = (2-\beta)\, f(\theta) + (1-\beta)\,\theta f'(\theta).$$

Note that $\frac{d^2 S^L(\theta,\theta^M)}{d\theta^2} > 0$ if $Q'(\theta) < 0$, $\frac{d^2 S^L(\theta,\theta^M)}{d\theta^2} = 0$ if $Q'(\theta) = 0$, and $\frac{d^2 S^L(\theta,\theta^M)}{d\theta^2} < 0$ if $Q'(\theta) > 0$. The claims about $S^L(\theta^L, \theta^M)$ follow.

The proof for the claims about $S^H(\theta^H, \theta^M)$ is analogous and thus omitted. □

A.2 Proof of Proposition 1

We proceed in three steps.

Step 1. We show that in any solution to (11)-(15), $V(\theta)$ is a step function and $g(\theta)$ and $V(\theta)$ are left- or right-continuous at each $\theta \in (\underline{\theta}, \overline{\theta}]$.

Note first that by the local private information constraints, $V(\theta)$ is piecewise continuously differentiable. Suppose by contradiction that $V(\theta)$ is not a step function. Then there is an interval over which $\frac{dV(\theta)}{d\theta}$ exists and is not zero, and therefore a sub-interval $[\theta^L, \theta^H]$ over which $\frac{dV(\theta)}{d\theta}$ is either strictly positive or strictly negative, with $\underline{V} < V(\theta) < \overline{V}$. By the private information constraint (12), $g(\theta)$ must be continuous and strictly increasing over $[\theta^L, \theta^H]$. By property (i) in Proposition 1, it suffices to consider the following two cases:

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

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

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

$$U\left(\widetilde{g}\left(\theta, \kappa\right)\right) = \kappa \overline{U} + (1-\kappa)\, U\left(g\left(\theta\right)\right), \tag{29}$$

which clearly exists. Let $\widetilde{x}(\theta,\kappa) = 1 - \widetilde{g}(\theta,\kappa)$ and define $\widetilde{V}(\theta,\kappa)$ as the solution to

$$\begin{aligned}&\theta U\left(\widetilde{g}\left(\theta,\kappa\right)\right) + \beta W\left(\widetilde{x}\left(\theta,\kappa\right)\right) + \beta\delta\widetilde{V}\left(\theta,\kappa\right)\\ &\qquad = \theta U(g(\theta^L)) + \beta W(x(\theta^L)) + \beta\delta V(\theta^L) + \int_{\theta^L}^{\theta} U(\widetilde{g}(\widetilde{\theta},\kappa))d\widetilde{\theta}. \qquad (30)\end{aligned}$$

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 (29) and (30) with respect to κ yields

$$\frac{d\widetilde{g}\left(\theta,\kappa\right)}{d\kappa} = \frac{\overline{U} - U\left(g\left(\theta\right)\right)}{U'\left(\widetilde{g}\left(\theta,\kappa\right)\right)}, \qquad (31)$$

$$\frac{d\widetilde{g}\left(\theta,\kappa\right)}{d\kappa}\left(\theta U'\left(\widetilde{g}\left(\theta,\kappa\right)\right) - \beta W'\left(\widetilde{x}\left(\theta,\kappa\right)\right)\right) + \beta\delta\frac{d\widetilde{V}\left(\theta,\kappa\right)}{d\kappa} = \int_{\theta^L}^{\theta}\frac{d\widetilde{g}(\widetilde{\theta},\kappa)}{d\kappa}U'(\widetilde{g}(\widetilde{\theta},\kappa))d\widetilde{\theta}. \qquad (32)$$

Substituting (31) in (32) yields that for a type $\theta \in [\theta^L, \theta^H]$, the change in government welfare from a marginal increase in κ is equal to

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

We begin by showing that the perturbation satisfies constraints (12)-(15). For the private information constraint (12), 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 (17), it then follows from equation (30) and the fact that $\widetilde{g}(\theta,\kappa)$ is nondecreasing that the perturbation satisfies constraint (12) for all $\theta \in \Theta$ and any $\kappa \in [0,1]$.

To prove that the perturbation satisfies the self-enforcement constraint (13), 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 verify:

$$\begin{aligned}D'\left(\theta\right) &= \overline{U} - U\left(g\left(\theta\right)\right),\\ D''\left(\theta\right) &= -U'\left(g\left(\theta\right)\right)\frac{dg\left(\theta\right)}{d\theta} < 0.\end{aligned}$$

Lastly, observe that constraint (14) is satisfied by construction, and constraint (15) is satisfied if $\kappa > 0$ is small enough. The latter claim follows from the fact that, by assumption, $V(\theta) \in (\underline{V}, \overline{V})$ for $\theta \in [\theta^L, \theta^H]$ in the original rule. Hence, for $\kappa > 0$ small enough, there

exist $\widetilde{V}(\theta,\kappa) \in [\underline{V}, \overline{V}]$ satisfying (29)-(30) for all $\theta \in [\theta^L, \theta^H]$.

We next show that the perturbation strictly increases social welfare. The change in social welfare from an increase in κ is equal to

$$\int_{\theta^L}^{\theta^H} \left[\frac{d\widetilde{g}(\theta,\kappa)}{d\kappa} \left(\theta U'\left(\widetilde{g}(\theta,\kappa)\right) - W'\left(\widetilde{x}(\theta,\kappa)\right) \right) + \delta \frac{d\widetilde{V}(\theta,\kappa)}{d\kappa} \right] f(\theta)\, d\theta.$$

Substituting with (32) yields

$$\int_{\theta^L}^{\theta^H} \left[\frac{d\widetilde{g}(\theta,\kappa)}{d\kappa} \theta U'\left(\widetilde{g}(\theta,\kappa)\right) \left(1 - \frac{1}{\beta}\right) + \frac{1}{\beta} \int_{\theta^L}^{\theta} \frac{d\widetilde{g}(\widetilde{\theta},\kappa)}{d\kappa} U'(\widetilde{g}(\widetilde{\theta},\kappa)) d\widetilde{\theta} \right] f(\theta)\, d\theta.$$

Substituting with (31) and integrating, this expression simplifies to

$$\frac{1}{\beta} \int_{\theta^L}^{\theta^H} \left(\overline{U} - U\left(g(\theta)\right) \right) \left(F(\theta^H) - F(\theta) - (1-\beta)\theta f(\theta) \right) d\theta.$$

This is an integral over the product of two terms. The first term is strictly decreasing in θ since $g(\theta)$ 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, these two terms are positively correlated with one another, and thus the change in social welfare is strictly greater than

$$\frac{1}{\beta} \int_{\theta^L}^{\theta^H} \left(\overline{U} - U\left(g(\theta)\right) \right) d\theta \int_{\theta^L}^{\theta^H} \left(F(\theta^H) - F(\theta) - (1-\beta)\theta f(\theta) \right) d\theta,$$

which is equal to 0. It follows that the change in social welfare from the perturbation is strictly positive. Hence, in any solution to (11)-(15), if $V(\theta)$ is continuous and $Q'(\theta) < 0$ over a given interval, then $V'(\theta) = 0$ over the interval.

Case 2: Suppose $Q'(\theta) > 0$ for all $\theta \in [\theta^L, \theta^H]$. Recall that $g(\theta)$ is continuous and strictly increasing over $[\theta^L, \theta^H]$. We begin by showing that the self-enforcement constraint cannot bind for all $\theta \in [\theta^L, \theta^H]$. Suppose it did. Using the representation of government welfare in (17), this implies

$$\int_{\theta}^{\theta^H} \left(U(g^f(\widetilde{\theta})) - U(g(\widetilde{\theta})) \right) d\widetilde{\theta} = 0 \tag{33}$$

for all $\theta \in [\theta^L, \theta^H]$. Note that the allocation must then satisfy $\{g(\theta), V(\theta)\} = \{g^f(\theta), \underline{V}\}$ for all $\theta \in (\theta^L, \theta^H)$: given $g(\theta)$ continuous, implicit differentiation of (33) with respect to θ given (33) holding with equality implies $g(\theta) = g^f(\theta)$; given this, the binding self-enforcement constraint yields $V(\theta) = \underline{V}$. However, this contradicts the assumption that

$V(\theta) \in (\underline{V}, \overline{V})$ for all $\theta \in [\theta^L, \theta^H]$. Hence, we obtain that the self-enforcement constraint cannot bind for all types in the interval, and without loss we take the interval to be such that this constraint is slack for all $\theta \in [\theta^L, \theta^H]$.

We show that there exists a steepening perturbation that is incentive feasible and strictly increases social welfare. Specifically, 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, type θ^M can no longer choose $g(\theta^M)$ and $V(\theta^M)$ 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 self-enforcement constraint is slack for all $\theta \in [\theta^L, \theta^H]$, the perturbation is feasible. The change in social welfare from the perturbation is equal to

$$\int_{\theta^M}^{\theta^H} \left(\frac{dg(\theta^H)}{d\theta^H} \left(\theta U'(g(\theta^H)) - W'(x(\theta^H)) \right) + \delta \frac{dV(\theta^H)}{d\theta^H} \right) f(\theta) d\theta$$
$$+ \frac{d\theta^M}{d\theta^H} \left(\theta^M U(g(\theta^L)) + W(x(\theta^L)) + \delta V(\theta^L) - \theta^M U(g(\theta^H)) - W(x(\theta^H)) - \delta V(\theta^H) \right) f(\theta^M).$$

Note that by the private information constraint for θ^H,

$$\frac{dg(\theta^H)}{d\theta^H} \left(\theta^H U'(g(\theta^H)) - \beta W'(x(\theta^H)) \right) + \beta \delta \frac{dV(\theta^H)}{d\theta^H} = 0, \tag{34}$$

and by indifference of θ^M,

$$\theta^M U(g(\theta^L)) + \beta W(x(\theta^L)) + \beta \delta V(\theta^L) = \theta^M U(g(\theta^H)) + \beta W(x(\theta^H)) + \beta \delta V(\theta^H). \tag{35}$$

Substituting with these expressions, the change in social welfare is equal to

$$\frac{dg(\theta^H)}{d\theta^H} U'(g(\theta^H)) \int_{\theta^M}^{\theta^H} \left(\theta - \frac{\theta^H}{\beta} \right) f(\theta) d\theta + \frac{d\theta^M}{d\theta^H} \theta^M \left(U(g(\theta^L)) - U(g(\theta^H)) \right) \left(1 - \frac{1}{\beta} \right) f(\theta^M). \tag{36}$$

Differentiating (35) with respect to θ^H and substituting with (34) yields

$$\frac{d\theta^M}{d\theta^H} = \frac{dg(\theta^H)}{d\theta^H} U'(g(\theta^H)) \frac{(\theta^H - \theta^M)}{U(g(\theta^H)) - U(g(\theta^L))}.$$

Substituting back into (36) and dividing by $\frac{1}{\beta} \frac{dg(\theta^H)}{d\theta^H} U'(g(\theta^H)) > 0$, we find that the change

in social welfare takes the same sign as

$$S^H(\theta^H, \theta^M) = \int_{\theta^M}^{\theta^H} \left(\beta\theta - \theta^H\right) f(\theta) d\theta + (1-\beta)\,\theta^M f(\theta^M)(\theta^H - \theta^M).$$

Since $Q'(\theta) > 0$ for all $\theta \in [\theta^M, \theta^H]$, Lemma 4 implies $S^H(\theta^H, \theta^M) > 0$, and thus the perturbation strictly increases social welfare. Hence, in any solution to (11)-(15), if $V(\theta)$ is continuous and $Q'(\theta) > 0$ over a given interval, then $V'(\theta) = 0$ over the interval.

It follows from Case 1 and Case 2 that $V(\theta)$ is a step function in any solution to (11)-(15). We now show that in any such solution, $g(\theta)$ and $V(\theta)$ must be right- or left-continuous at each $\theta \in (\underline{\theta}, \overline{\theta}]$. Suppose by contradiction that this is not true at some $\theta \in (\underline{\theta}, \overline{\theta}]$. Denote the left limit by $\{g(\theta^-), V(\theta^-)\} = \lim_{\theta' \uparrow \theta} \{g(\theta'), V(\theta')\}$. By Lemma 2,

$$0 < \theta \left(U(g(\theta)) - U(g(\theta^-))\right) = \beta \left(W(x(\theta^-)) + \delta V(\theta^-) - W(x(\theta)) - \delta V(\theta)\right).$$

Given $\beta \in (0,1)$, this implies

$$\theta \left(U(g(\theta)) - U(g(\theta^-))\right) < W(x(\theta^-)) + \delta V(\theta^-) - W(x(\theta)) - \delta V(\theta).$$

It follows that a perturbation that assigns $\{g(\theta^-), V(\theta^-)\}$ to type θ is incentive feasible, strictly increases social welfare from type θ, and does not affect social welfare from types other than θ. Hence, $g(\theta)$ and, by Lemma 2, $V(\theta)$ must be left- or right-continuous at each $\theta \in (\underline{\theta}, \overline{\theta}]$.

Step 2. We show that in any solution to (11)-(15), if $\frac{dg(\theta)}{d\theta} > 0$, then $V(\theta) \in \{\underline{V}, \overline{V}\}$.

If $\frac{dg(\theta)}{d\theta} > 0$ at some $\theta \in \Theta$, then $g(\theta)$ is continuous and strictly increasing over some interval, call it $[\theta^L, \theta^H]$, which contains θ. By Step 1, $V(\theta)$ must be constant over the interval at some value $V \in [\underline{V}, \overline{V}]$, which implies $g(\theta) = g^f(\theta)$ for all $\theta \in [\theta^L, \theta^H]$. We establish that $V \in \{\underline{V}, \overline{V}\}$. Suppose by contradiction that $\underline{V} < V < \overline{V}$. As in Step 1, it suffices to consider the following two cases:

Case 1: Suppose $Q'(\theta) < 0$ for all $\theta \in [\theta^L, \theta^H]$. Then given $V \in (\underline{V}, \overline{V})$, the arguments in Case 1 of Step 1 above apply, implying that there exists an incentive feasible flattening perturbation that rotates the spending schedule clockwise over $[\theta^L, \theta^H]$ and strictly increases social welfare. Therefore, we must have $V(\theta) \in \{\underline{V}, \overline{V}\}$ for all $\theta \in [\theta^L, \theta^H]$.

Case 2: Suppose $Q'(\theta) > 0$ for all $\theta \in [\theta^L, \theta^H]$. Then given $V \in (\underline{V}, \overline{V})$, the arguments in Case 2 of Step 1 above apply, implying that there exists an incentive feasible steepening

perturbation that drills a hole in the spending schedule within $[\theta^L, \theta^H]$ and strictly increases social welfare. Therefore, we must have $V(\theta) \in \{\underline{V}, \overline{V}\}$ for all $\theta \in [\theta^L, \theta^H]$.

Step 3. We show that in any solution to (11)-(15), if $\frac{dg(\theta)}{d\theta} = 0$, then $V(\theta) \in \{\underline{V}, \overline{V}\}$.

Suppose by contradiction that $\frac{dg(\theta)}{d\theta} = 0$ and $V(\theta) \in (\underline{V}, \overline{V})$ for some type $\theta \in \Theta$. By Step 1 and Step 2, this type belongs to a stand-alone segment, call it $[\theta^L, \theta^M]$, such that $g(\theta) = g$ and $V(\theta) = V$ for all $\theta \in [\theta^L, \theta^M]$, for some $g > 0$ and $V \in (\underline{V}, \overline{V})$, and $g(\theta)$ jumps at each boundary unless $\theta^L = \underline{\theta}$ or $\theta^M = \overline{\theta}$. Observe that the self-enforcement constraint must be slack for all $\theta \in (\theta^L, \theta^M)$. This follows from the fact that the spending rate and the continuation value are constant over $[\theta^L, \theta^M]$. Hence, if (13) were to hold as an equality at some $\theta' \in (\theta^L, \theta^M)$, this constraint would be violated for either $\theta \in [\theta^L, \theta')$ or $\theta \in (\theta', \theta^M]$.

We show that there exists an incentive feasible perturbation that strictly increases social welfare. We consider segment-shifting steepening and flattening perturbations that marginally change the constant spending rate by $dg \lesseqgtr 0$ and change V so as to keep unchanged the government welfare of either type θ^L or type θ^M. As we describe next, which perturbation within this class we perform depends on the shape of the function $Q(\theta)$ over $[\theta^L, \theta^M]$.

Using condition (ii) in Proposition 1, suppose first that $\int_{\theta^L}^{\theta^M} Q(\theta^M)d\theta < \int_{\theta^L}^{\theta^M} Q(\theta)d\theta$. Then we consider a perturbation that marginally changes the spending rate by $dg < 0$ and keeps type θ^L equally well off. This means that $\frac{dV}{dg}$ is given by

$$-\left(\theta^L U'(g) - \beta W'(x)\right) + \beta\delta\frac{dV}{dg} = 0, \tag{37}$$

where $x = 1 - g$. Note that this perturbation reduces θ^M, in the sense that the highest types in $[\theta^L, \theta^M]$, arbitrarily close to θ^M, will now jump up to a higher allocation. This higher allocation is the allocation over which type θ^M was initially indifferent if $\theta^M < \overline{\theta}$ and such an allocation satisfies self-enforcement constraints; otherwise, we let the perturbation introduce these allocations (which will satisfy private information constraints, and will involve flexible spending and maximal punishment if the self-enforcement constraint of θ^M is binding in the original allocation). Calling the allocation to which type θ^M jumps up the allocation of type θ^H, the indifference condition of θ^M is

$$\theta^M U(g) + \beta W(x) + \beta\delta V = \theta^M U(g(\theta^H)) + \beta W(x(\theta^H)) + \beta\delta V(\theta^H).$$

To verify feasibility, note that the self-enforcement constraint is slack for all $\theta \in (\theta^L, \theta^M)$, $V \in (\underline{V}, \overline{V})$, and the government welfare of types θ^L and θ^M remains unchanged with the

perturbation. Hence, the perturbation is feasible for dg arbitrarily close to zero.

The change in social welfare due to the perturbation is equal to

$$\int_{\theta^L}^{\theta^M} \left(-\left(\theta U'(g) - W'(x)\right) + \delta \frac{dV}{dg} \right) f(\theta)\, d\theta$$
$$+ \frac{d\theta^M}{dg} \left(\theta^M U(g) + W(x) + \delta V - \theta^M U(g(\theta^H)) - W(x(\theta^H)) - \delta V(\theta^H) \right) f(\theta^M).$$

Substituting with (37) and the indifference condition of type θ^M yields that the change in social welfare is equal to

$$-U'(g) \int_{\theta^L}^{\theta^M} \left(\theta - \frac{\theta^L}{\beta} \right) f(\theta) d\theta - \frac{d\theta^M}{dg} \theta^M \left(U(g(\theta^H)) - U(g) \right) \left(1 - \frac{1}{\beta} \right) f(\theta^M). \tag{38}$$

Differentiating the indifference condition of type θ^M and substituting with (37) yields

$$\frac{d\theta^M}{dg} = -U'(g) \frac{\left(\theta^M - \theta^L\right)}{U\left(g\left(\theta^H\right)\right) - U(g)}.$$

Substituting back into (38) and dividing by $\frac{1}{\beta} U'(g) > 0$, we find that the change in social welfare takes the same sign as

$$-S^L(\theta^L, \theta^M) = -\left[\int_{\theta^L}^{\theta^M} \left(\beta\theta - \theta^L \right) f(\theta) d\theta + (1-\beta)\, \theta^M f(\theta^M)(\theta^M - \theta^L) \right],$$

which can be rewritten as

$$-S^L(\theta^L, \theta^M) \;=\; -\int_{\theta^L}^{\theta^M} \int_{\theta}^{\theta^M} Q'(\widetilde{\theta}) d\widetilde{\theta} d\theta = -\int_{\theta^L}^{\theta^M} (Q(\theta^M) - Q(\theta)) d\theta = 0. \tag{39}$$

By the assumption that $\int_{\theta^L}^{\theta^M} Q(\theta^M) d\theta < \int_{\theta^L}^{\theta^M} Q(\theta) d\theta$, the above expression is strictly positive, implying that the perturbation strictly increases social welfare.

We can perform analogous perturbations in the other instances of condition (ii) in Proposition 1. If $\int_{\theta^L}^{\theta^M} Q(\theta^M) d\theta > \int_{\theta^L}^{\theta^M} Q(\theta) d\theta$, we consider a perturbation that marginally changes the spending rate by $dg > 0$ and keeps type θ^L equally well off. Arguments analogous to those above imply that the perturbation is incentive feasible. Moreover, one can verify that the change in social welfare from the perturbation is given by $S^L(\theta^L, \theta^M)$, which is strictly positive in this case. If $\int_{\theta^L}^{\theta^M} Q(\theta^M) d\theta = \int_{\theta^L}^{\theta^M} Q(\theta) d\theta$, then by condition (ii) we must have $\int_{\theta^L}^{\theta^M} Q(\theta^L) d\theta \neq \int_{\theta^L}^{\theta^M} Q(\theta) d\theta$. We then consider perturbations that marginally

change the spending rate by $dg \lesseqgtr 0$ and keep type θ^M equally well off. Arguments analogous to those above imply that these perturbation are incentive feasible. Moreover, one can verify that the implied change in social welfare is given by either $S^H(\theta^M, \theta^L)$ or $-S^H(\theta^M, \theta^L)$, depending on the sign of dg, where

$$S^H(\theta^M, \theta^L) = \int_{\theta^L}^{\theta^M} \int_{\theta^L}^{\theta} Q'(\widetilde{\theta}) d\widetilde{\theta} d\theta = \int_{\theta^L}^{\theta^M} (Q(\theta) - Q(\theta^L)) d\theta. \tag{40}$$

Hence, given $\int_{\theta^L}^{\theta^M} Q(\theta^L) d\theta \neq \int_{\theta^L}^{\theta^M} Q(\theta) d\theta$, one of these perturbations strictly increases social welfare in this case.

A.3 Proof of Lemma 3

Step 1. We show that in any solution to (11)-(15), if $V(\theta^{**}) = \underline{V}$, then $\theta^{**} \geq \widehat{\theta}$.

By Proposition 1, if $V(\theta^{**}) = \underline{V}$ for some $\theta^{**} \in \Theta$, then $V(\theta) = \underline{V}$ over an interval $[\theta^L, \theta^H]$ that contains θ^{**}. We establish that $\theta^L \geq \widehat{\theta}$. Suppose by contradiction that $\theta^L < \widehat{\theta}$. Note that the self-enforcement constraint (13) requires $g(\theta) = g^f(\theta)$ for all $\theta \in [\theta^L, \theta^H]$. Then we can perform a flattening perturbation that rotates the spending schedule clockwise over a subset of $[\theta^L, \theta^H]$ that contains type θ^L and is below $\widehat{\theta}$, analogous to the perturbation used in Step 2 in the proof of Proposition 1. By the arguments in that step, this perturbation is incentive feasible. Moreover, since $Q'(\theta) < 0$ for all types θ in the subset (by the subset being below $\widehat{\theta}$ and Assumption 2), the perturbation strictly increases social welfare. It follows that $\theta^L \geq \widehat{\theta}$, proving the claim in this step.

Step 2. We show that in any solution to (11)-(15), if $V(\theta^{**}) = \underline{V}$, then $V(\theta) = \underline{V}$ for all $\theta \geq \theta^{**}$.

Suppose by contradiction that $V(\theta^{**}) = \underline{V}$ for some $\theta^{**} \in \Theta$ and $V(\theta) > \underline{V}$ for some $\theta > \theta^{**}$. By Step 1, $\theta^{**} \geq \widehat{\theta}$. We first show that there exist $\theta^H > \theta^L > \theta^{**}$ such that $V(\theta) = \overline{V}$ for all $\theta \in [\theta^L, \theta^H]$. Suppose this is not true. Then by the contradiction assumption and Proposition 1, $V(\theta)$ must jump up at $\overline{\theta}$. However, this contradicts the claim in Step 1 in the proof of Proposition 1 that $V(\theta)$ is left-continuous at $\overline{\theta}$. Therefore, given Proposition 1 and the contradiction assumption, we must have $\theta^H > \theta^L > \theta^{**}$ such that $V(\theta) = \overline{V}$ for all $\theta \in [\theta^L, \theta^H]$.

We next establish that $g(\theta) = g$ for all $\theta \in [\theta^L, \theta^H]$ and some $g > 0$. Suppose by contradiction that $\frac{dg(\theta)}{d\theta} > 0$ for some type $\theta \in [\theta^L, \theta^H]$. Note that the private information constraint (12) implies $g(\theta) = g^f(\theta)$ in the neighborhood of such a type θ. Then we can

perform an incentive feasible steepening perturbation in this neighborhood as that described in Step 1 in the proof of Proposition 1, which drills a hole in the spending schedule $g(\theta)$. By the arguments in that step, this perturbation strictly increases social welfare. It follows that $\frac{dg(\theta)}{d\theta} = 0$ for any $\theta > \theta^{**}$ such that $V(\theta) = \overline{V}$.

Finally, we show that a segment $[\theta^L, \theta^H]$ with $V(\theta) = \overline{V}$ and $g(\theta) = g > 0$ for all $\theta \in [\theta^L, \theta^H]$ and $\theta^L > \theta^{**}$ cannot exist. Consider the closest such segment to θ^{**}, so that θ^L is the lowest point above θ^{**} at which the continuation value jumps from $\underline{V}$ to $\overline{V}$. Take θ^H to be the lowest point above θ^L at which $g(\theta)$ jumps, or take $\theta^H = \overline{\theta}$ if $g(\theta)$ does not jump above θ^L. Then $[\theta^L, \theta^H]$ is a stand-alone segment with constant spending g and continuation value $V = \overline{V}$. Moreover, note that $g > g^f(\theta^L)$; this follows from Lemma 2 given the jump in continuation value at θ^L. There are two cases to consider.

Case 1: Suppose $g < g^f(\theta^H)$. Note that by $\theta^L > \theta^{**} \geq \widehat{\theta}$ and Assumption 2, $\int_{\theta^L}^{\theta^H} Q(\theta)d\theta > \int_{\theta^L}^{\theta^H} Q(\theta^L)d\theta$. Moreover, given $g^f(\theta^H) > g > g^f(\theta^L)$, we can perform an incentive feasible segment-shifting steepening perturbation analogous to that in Step 3 in the proof of Proposition 1; this perturbation marginally reduces g and lowers V so as to keep the government welfare of type θ^L unchanged while letting type θ^H jump to a higher spending rate. By the arguments in that step, this perturbation strictly increases social welfare, contradicting the assumption that $g < g^f(\theta^H)$.

Case 2: Suppose $g \geq g^f(\theta^H)$. Then Lemma 2 requires $\theta^H = \overline{\theta}$, as the private information constraint would be violated if $g(\theta) \geq g^f(\theta^H)$ jumps up and $V(\theta)$ weakly decreases at θ^H. Note that since the self-enforcement constraint (13) is satisfied for $\theta = \theta^L$ and $g \geq g^f(\theta^H)$, this constraint must hold as a strict inequality for all $\theta \in (\theta^L, \theta^H]$. Then we can perform an incentive feasible segment-shifting steepening perturbation that marginally reduces g and lowers V so as to keep the government welfare of type θ^L unchanged, like that in Case 1 above, except that in this case θ^H needs not jump to a higher spending rate. Using the representation in (18), the change in social welfare from the perturbation is equal to

$$-U'(g)\int_{\theta^L}^{\theta^H} Q(\theta)d\theta,$$

which is strictly positive given $\theta^L \geq \widehat{\theta}$ and Assumption 2. This contradicts the assumption that $g \geq g^f(\theta^H)$.

Case 1 and Case 2 imply that a segment $[\theta^L, \theta^H]$ with $V(\theta) = \overline{V}$ and $g(\theta) = g > 0$ for all $\theta \in [\theta^L, \theta^H]$ and $\theta^L > \theta^{**}$ cannot exist, completing the proof of this step.

Step 3. We show that in any solution to (11)-(15), $V(\theta)$ is right-continuous at $\underline{\theta}$ with

$V(\underline{\theta}) = \overline{V}$.

By the previous steps and Proposition 1, if $V(\theta) = \underline{V}$ for some $\theta \in \Theta$, then there exists $\theta^{**} \in \Theta$ such that $V(\theta) = \underline{V}$ for all $\theta > \theta^{**}$ and, if $\theta^{**} > \underline{\theta}$, $V(\theta) = \overline{V}$ for all $\theta \in (\underline{\theta}, \theta^{**})$. We begin by establishing that indeed $\theta^{**} > \underline{\theta}$. Suppose by contradiction that $\theta^{**} = \underline{\theta}$ and thus $V(\theta) = \underline{V}$ for all $\theta \in (\underline{\theta}, \overline{\theta}]$. Then the self-enforcement constraint (13) implies $g(\theta) = g^f(\theta)$ for all $\theta \in (\underline{\theta}, \overline{\theta}]$. Consider an incentive feasible global perturbation that assigns $V(\theta) = \overline{V}$ to all $\theta \in [\underline{\theta}, \overline{\theta}]$ and assigns type $\underline{\theta}$ the limiting allocation to its right. This perturbation is incentive feasible and strictly increases social welfare, contradicting the assumption that $\theta^{**} = \underline{\theta}$.

Given $\theta^{**} > \underline{\theta}$, we complete the proof by showing that $V(\underline{\theta}) = \overline{V}$. Suppose by contradiction that $V(\underline{\theta}) < \overline{V}$. By the claims above, this implies that the continuation value jumps up from $V(\underline{\theta})$ to $\overline{V}$ at $\underline{\theta}$. By Lemma 2, the private information constraint at $\underline{\theta}$ requires $\lim_{\theta \downarrow \underline{\theta}} g(\theta) > g^f(\underline{\theta})$. Moreover, the private information constraint for higher types requires that there exist $\theta^L > \underline{\theta}$ such that $g(\theta) = g^f(\theta^L)$ and $V(\theta) = \overline{V}$ for all $\theta \in (\underline{\theta}, \theta^L]$. Then we can perform an incentive feasible global perturbation that assigns types $\theta \in [\underline{\theta}, \theta^L]$ the allocation $\{g(\theta), V(\theta)\} = \{g^f(\theta), \overline{V}\}$. Since $g^f(\theta) > g^{fb}(\theta)$ and $\theta U(g(\theta)) + W(x(\theta))$ is strictly concave, this perturbation strictly increases social welfare. It follows that we must have $V(\underline{\theta}) = \overline{V}$ in any solution to (11)-(15).

A.4 Proof of Proposition 2

By Proposition 1, Lemma 3, and the self-enforcement constraint (13), in any solution to (11)-(15), there exists $\theta^{**} > 0$ such that $\{g(\theta), V(\theta)\} = \{g^f(\theta), \underline{V}\}$ for all $\theta > \theta^{**}$ and $V(\theta) = \overline{V}$ for all $\theta < \theta^{**}$ (where it is possible that $\theta^{**} > \overline{\theta}$). Moreover, by the same logic as in the last paragraph of Step 1 in the proof of Proposition 1, $V(\theta^{**}) = \overline{V}$ (with $g(\theta^{**}) < g^f(\theta^{**})$) in any such solution. Since the self-enforcement constraint holds with equality at θ^{**}, it follows that

$$\theta^{**} U(g(\theta^{**})) + \beta W(x(\theta^{**})) + \beta\delta\overline{V} = \theta^{**} U(g^f(\theta^{**})) + \beta W(x^f(\theta^{**})) + \beta\delta\underline{V}. \tag{41}$$

This characterizes the allocation for $\theta \geq \theta^{**}$. To characterize the allocation for $\theta < \theta^{**}$, we first establish that $g(\theta)$ is continuous over $[\underline{\theta}, \theta^{**}]$. Recall from the proof of Lemma 3 that $\theta^{**} \geq \widehat{\theta}$. There are two cases to consider:

Case 1: Suppose by contradiction that $g(\theta)$ has a point of discontinuity below $\widehat{\theta}$: there is a type $\theta^M < \widehat{\theta}$ which is indifferent between choosing $\lim_{\theta \uparrow \theta^M} g(\theta)$ and $\lim_{\theta \downarrow \theta^M} g(\theta) > \lim_{\theta \uparrow \theta^M} g(\theta)$.

Note that given $V(\theta) = \overline{V}$ 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 $g^f(\theta^L)$ and types $\theta \in (\theta^M, \theta^H]$ bunched at $g^f(\theta^H)$, for some $\theta^L < \theta^M < \theta^H$. Now consider perturbing the rule 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 clearly incentive feasible. The change in social welfare from the perturbation is equal to

$$\frac{dg^f(\theta^L)}{d\theta^L} \int_{\theta^L}^{\theta^M} \left(\theta U'(g^f(\theta^L)) - W'(x^f(\theta^L))\right) f(\theta)\, d\theta$$
$$+\frac{d\theta^M}{d\theta^L} \left[\theta^M \left(U(g^f(\theta^L)) - U(g^f(\theta^H))\right) + W(x^f(\theta^L)) - W(x^f(\theta^H))\right] f(\theta^M).$$

By the definition of $g^f(\theta^L)$, we have $\theta^L U'(g^f(\theta^L)) = \beta W'(x^f(\theta^L))$, and by indifference of type θ^M (with $V(\theta)$ constant over $[\underline{\theta}, \widehat{\theta})$), we have

$$\theta^M U(g^f(\theta^L)) + \beta W(x^f(\theta^L)) = \theta^M U(g^f(\theta^H)) + \beta W(x^f(\theta^H)). \tag{42}$$

Substituting these into the expression above yields that the change in social welfare due to the perturbation is equal to

$$\frac{dg^f(\theta^L)}{d\theta^L} U'(g^f(\theta^L)) \int_{\theta^L}^{\theta^M} \left(\theta - \frac{\theta^L}{\beta}\right) f(\theta)\, d\theta + \frac{d\theta^M}{d\theta^L} \left(\frac{1}{\beta} - 1\right) \theta^M f(\theta^M) \left(U(g^f(\theta^H)) - U(g^f(\theta^L))\right). \tag{43}$$

Note that differentiating the indifference condition (42) with respect to θ^L (and substituting again with $\theta^L U'(g^f(\theta^L)) = \beta W'(x^f(\theta^L))$) yields

$$\frac{d\theta^M}{d\theta^L} = \frac{dg^f(\theta^L)}{d\theta^L} U'(g^f(\theta^L)) \frac{(\theta^M - \theta^L)}{U(g^f(\theta^H)) - U(g^f(\theta^L))}.$$

Substituting this back into (43) and dividing by $\frac{1}{\beta} \frac{dg^f(\theta^L)}{d\theta^L} U'(g^f(\theta^L)) > 0$, we find that the change in social welfare takes the same sign as

$$S^L(\theta^L, \theta^M) = \int_{\theta^L}^{\theta^M} \left(\beta\theta - \theta^L\right) f(\theta)\, d\theta + (1 - \beta)\, \theta^M f(\theta^M)(\theta^M - \theta^L).$$

By $\theta^M < \widehat{\theta}$, Assumption 2, and Lemma 4, $S^L(\theta^L, \theta^M) > 0$. Thus, the perturbation strictly increases social welfare, showing that $g(\theta)$ cannot jump at a point below $\widehat{\theta}$.

Case 2: Suppose by contradiction that $g(\theta)$ is discontinuous at a point $\theta \in [\widehat{\theta}, \theta^{**}]$. Note that since $V(\theta) = \overline{V}$ for all $\theta \in [\widehat{\theta}, \theta^{**}]$, we can apply the same logic as in Step 2 in the

proof of Lemma 3 to show that $\frac{dg(\theta)}{d\theta} = 0$ over any continuous interval in $[\widehat{\theta}, \theta^{**}]$. Now using the arguments in Case 1 above, this means that if $g(\theta)$ jumps at a point in $[\widehat{\theta}, \theta^{**}]$, then there exists a stand-alone segment with constant spending in $[\widehat{\theta}, \theta^{**}]$. However, using again the arguments in Step 2 in the proof of Lemma 3, we can then show that there exists an incentive feasible segment-shifting steepening perturbation that strictly increases social welfare. Thus, $g(\theta)$ cannot jump at a point $\theta \in [\widehat{\theta}, \theta^{**}]$.

Case 1 and Case 2 imply that $g(\theta)$ is continuous over $[\underline{\theta}, \theta^{**}]$. It follows that the allocation over this range must be bounded discretion with either a minimum spending rate or a maximum spending rate or both. The arguments in Step 3 in the proof of Lemma 3 imply that a minimum spending rate is suboptimal, and Lemma 2 implies $g(\theta^{**}) = g^f(\theta^*)$ for $g^f(\theta^*) < g^f(\theta^{**})$ which satisfies (41).

To complete the proof, we now show that $\theta^* < \overline{\theta}$, so the maximum spending rate binds for types $\theta \in (\theta^*, \overline{\theta}]$. Suppose by contradiction that this is not the case, meaning that $\{g(\theta), V(\theta)\} = \{g^f(\theta), \overline{V}\}$ for all $\theta \in \Theta$. Consider an incentive feasible perturbation which assigns $\{g(\theta), V(\theta)\} = \{g^f(\overline{\theta} - \varepsilon), \overline{V}\}$ for all $\theta \in [\overline{\theta} - \varepsilon, \overline{\theta}]$, where $\varepsilon > 0$ is chosen to be small enough as to continue to satisfy the self-enforcement constraint (13) for all types. Using the representation in (18), the change in social welfare from this perturbation is equal to

$$\int_{\overline{\theta}-\varepsilon}^{\overline{\theta}} (U(g^f(\overline{\theta} - \varepsilon)) - U(g^f(\theta)))Q(\theta).$$

Since $g^f(\overline{\theta} - \varepsilon) < g^f(\theta)$ and $Q(\theta) < 0$ for all $\theta \in (\overline{\theta} - \varepsilon, \overline{\theta})$ given $\varepsilon > 0$ arbitrarily small, the perturbation strictly increases social welfare.

B Supplementary Appendix for Online Publication

B.1 Proof of Lemma 1

We begin by proving necessity. Conditional on a public history h^{t-1}, type θ_t can choose to follow the equilibrium strategy associated with (h^{t-1}, θ_t') from t onward for any $\theta_t' \neq \theta_t$, instead of that associated with (h^{t-1}, θ_t). This provides immediate spending rate $g_t(\theta^{t-1}, \theta_t')$ and continuation value $V_{t+1}(\theta^{t-1}, \theta_t')$. Condition (9) is necessary for any such unobservable deviation to be weakly dominated. In addition, conditional on a public history h^{t-1}, type θ_t can choose spending rate $g^f(\theta_t)$. In equilibrium, this leads to a continuation value V' associated with public history $\{h^{t-1}, g^f(\theta_t)\}$. By definition of $\underline{V}$, the continuation value V' weakly exceeds $\underline{V}$. Thus, condition (10) is necessary for a deviation to $g^f(\theta_t)$ to be weakly dominated.

To establish sufficiency, take a spending sequence that satisfies (9) and (10). We construct equilibrium strategies as follows. Consider a public history along the equilibrium path in which h^{t-1} corresponds to a positive probability realization of the spending sequence. Given θ_t, the government chooses the level of spending associated with a history of shocks θ^t under the spending sequence. If instead h^{t-2} corresponds to a positive probability realization of the spending sequence, but $\{h^{t-2}, g_{t-1}\}$ does not, then the government follows the continuation strategy associated with continuation value $\underline{V}$. Condition (9) implies that any unobservable deviation is weakly dominated, and condition (10) implies that any observable deviation is weakly dominated. Therefore, the spending sequence is supported by these equilibrium strategies.

B.2 Proof of Proposition 3

Assume that (24) does not hold. We begin with some preliminaries.

For a given threshold θ', denote by $b(\theta')$ the type exceeding θ' at which (20) holds:

$$b(\theta')U(g^f(\theta')) + \beta W(x^f(\theta')) + \beta\delta\overline{V} = b(\theta')U(g^f(b(\theta'))) + \beta W(x^f(b(\theta'))) + \beta\delta\underline{V}. \quad (44)$$

Note that for any θ', $b(\theta') > \theta'$ is uniquely defined, with $b'(\theta') > 0$. By the definition of a maximally enforced deficit limit in Definition 1, $b(\theta^*) = \theta^{**}$, and by the definition of θ_b in equation (25), $b(\theta_b) = \overline{\theta}$.

Consider a maximally enforced deficit limit with threshold θ' and associated $b(\theta')$. Under such a limit, types $\theta \in [\theta', b(\theta')]$ are bunched at the flexible spending rate of θ' and receive the maximal reward $\overline{V}$, whereas types $\theta > b(\theta')$ spend at their flexible rate and receive

the maximal punishment $\underline{V}$. We will consider a perturbation where we marginally reduce the threshold θ' and let type $b(\theta')$, whose self-enforcement constraint initially holds with equality, jump up to a spending rate $g^f(b(\theta'))$ with maximal punishment $\underline{V}$. The change in social welfare from this perturbation is equal to:

$$\begin{aligned} M(\theta') &= -\frac{dg^f(\theta')}{d\theta'}\int_{\theta'}^{b(\theta')}\left(\theta U'(g^f(\theta')) - W'(x^f(\theta'))\right) f(\theta)d\theta \\ &\quad -b'(\theta')\begin{pmatrix} b(\theta')U(g^f(\theta')) + W(x^f(\theta')) + \delta\overline{V} \\ -b(\theta')U(g^f(b(\theta'))) - W(x^f(b(\theta'))) - \delta\underline{V} \end{pmatrix} f(b(\theta')). \end{aligned}$$

By the definition of $g^f(\theta')$, we have $\theta' U'(g^f(\theta')) = \beta W'(x^f(\theta'))$. Substituting with this and type $b(\theta')$'s binding self-enforcement constraint (44) yields

$$\begin{aligned} M(\theta') &= -\frac{dg^f(\theta')}{d\theta'}U'(g^f(\theta'))\int_{\theta'}^{b(\theta')}\left(\theta - \frac{\theta'}{\beta}\right) f(\theta)\, d\theta \\ &\quad -b'(\theta')\left(\frac{1}{\beta} - 1\right) b(\theta') f(b(\theta'))\left(U(g^f(b(\theta'))) - U(g^f(\theta'))\right). \end{aligned} \tag{45}$$

Note that differentiating the indifference condition (44) with respect to θ' (and substituting again with $\theta' U'(g^f(\theta')) = \beta W'(x^f(\theta'))$ yields

$$b'(\theta') = \frac{dg^f(\theta')}{d\theta'}U'(g^f(\theta'))\frac{(b(\theta') - \theta')}{U(g^f(b(\theta'))) - U(g^f(\theta'))}.$$

Substituting this back into (45), we obtain

$$M(\theta') = \frac{1}{\beta}\frac{dg^f(\theta')}{d\theta'}U'(g^f(\theta'))\left[-\int_{\theta'}^{b(\theta')}(\beta\theta - \theta')\, f(\theta)\, d\theta - (1-\beta)\, b(\theta') f(b(\theta'))(b(\theta') - \theta')\right]. \tag{46}$$

Note that using integration by parts, we have

$$\int_{\theta'}^{b(\theta')}(1 - F(\theta))\, d\theta = (1 - F(b(\theta')))b(\theta') - (1 - F(\theta'))\theta' + \int_{\theta'}^{b(\theta')}\theta f(\theta)\, d\theta,$$

which yields

$$\begin{aligned} \int_{\theta'}^{b(\theta')} Q(\theta)d\theta &= (1 - F(b(\theta')))b(\theta') - (1 - F(\theta'))\theta' + \int_{\theta'}^{b(\theta')}\beta\theta f(\theta)\, d\theta \\ &= (1 - F(b(\theta')))b(\theta') - (1 - F(b(\theta')))\theta' + \int_{\theta'}^{b(\theta')}(\beta\theta - \theta')\, f(\theta)\, d\theta. \end{aligned} \tag{47}$$

Rearranging terms yields

$$-\int_{\theta'}^{b(\theta')} (\beta\theta - \theta')\, f(\theta)\, d\theta = -\int_{\theta'}^{b(\theta')} Q(\theta)d\theta + (1 - F(b(\theta')))(b(\theta') - \theta').$$

Substituting back into (46), we obtain that the change in social welfare from the perturbation is equal to

$$M(\theta') = \frac{1}{\beta}\frac{dg^f(\theta')}{d\theta'}U'(g^f(\theta'))\left[\int_{\theta'}^{b(\theta')} (Q(b(\theta')) - Q(\theta))\, d\theta\right]. \tag{48}$$

Since $\frac{1}{\beta}\frac{dg^f(\theta')}{d\theta'}U'(g^f(\theta')) > 0$, $M(\theta')$ takes the same sign as the term in square brackets. The derivative of $M(\theta')$ is equal to

$$\begin{aligned} M'(\theta') \quad = \quad & \frac{1}{\beta}\frac{d\left(\frac{dg^f(\theta')}{d\theta'}U'(g^f(\theta'))\right)}{d\theta'}\left[\int_{\theta'}^{b(\theta')} (Q(b(\theta')) - Q(\theta))\, d\theta\right] \\ & + \frac{1}{\beta}\frac{dg^f(\theta')}{d\theta'}U'(g^f(\theta'))\left[Q(\theta') - Q(b(\theta')) + Q'(b(\theta'))b'(\theta')(b(\theta') - \theta')\right]. \end{aligned} \tag{49}$$

The following auxiliary lemmas will be useful to prove the proposition.

Lemma 5. *Suppose (24) does not hold. Any solution to (11)-(15) is a maximally enforced deficit limit with $\theta^* \leq \theta_b$.*

Proof. By Proposition 2, all we need to prove is that any optimal deficit limit sets a threshold $\theta^* \leq \theta_b$. Suppose by contradiction that an optimal deficit limit sets $\theta^* > \theta_b$. Consider a perturbation where we reduce this threshold marginally. Note that the self-enforcement constraint is satisfied for all $\theta \in \Theta$ given a limit $\theta^* > \theta_b$; that is, $b(\theta^*) > \overline{\theta}$. Thus, substituting with $f(\theta) = 0$ for all $\theta > \overline{\theta}$ in (45), the change in social welfare from this perturbation is equal to

$$-\frac{dg^f(\theta^*)}{d\theta^*}U'(g^f(\theta^*))\int_{\theta^*}^{\overline{\theta}}\left(\theta - \frac{\theta^*}{\beta}\right) f(\theta)d\theta. \tag{50}$$

Analogous to the derivation in (47), note that

$$\int_{\theta^*}^{\overline{\theta}} Q(\theta)d\theta = \int_{\theta^*}^{\overline{\theta}} (\beta\theta - \theta^*)\, f(\theta)\, d\theta,$$

and thus $\int_{\theta^*}^{\overline{\theta}}\left(\theta - \frac{\theta^*}{\beta}\right) f(\theta)d\theta$ takes the same sign as $\int_{\theta^*}^{\overline{\theta}} Q(\theta)d\theta$. Recall that by (23), $\int_{\theta_e}^{\overline{\theta}} Q(\theta)d\theta = 0$. Since (24) does not hold, $\theta_b > \theta_e$, and by the contradiction assump-

tion, $\theta^* > \theta_b > \theta_e$. It follows then from Assumption 2 that $\int_{\theta^*}^{\overline{\theta}} Q(\theta)d\theta < 0$ and thus $\int_{\theta^*}^{\overline{\theta}} \left(\theta - \frac{\theta^*}{\beta}\right) f(\theta)\, d\theta < 0$. Since $\frac{dg^f(\theta^*)}{d\theta^*} U'(g^f(\theta^*)) > 0$, we obtain from (50) that the perturbation strictly increases social welfare. The claim follows. □

Lemma 6. *Suppose (24) does not hold and consider a maximally enforced deficit limit with threshold $\theta' \leq \theta_b$ and associated $b(\theta')$. If $M(\theta') = 0$ and $b(\theta') \geq \widehat{\theta}$, then $M'(\theta') > 0$.*

Proof. Using (48) and the fact that $\frac{1}{\beta}\frac{dg^f(\theta')}{d\theta'} U'(g^f(\theta')) > 0$, the condition $M(\theta') = 0$ implies

$$\int_{\theta'}^{b(\theta')} (Q(b(\theta')) - Q(\theta))\, d\theta = 0.$$

By $b(\theta') \geq \widehat{\theta}$, $Q(b(\theta')) < 0$. Given Assumption 2, the above equality therefore requires $\int_{\theta'}^{b(\theta')} Q(\theta)d\theta < 0$ and $Q(\theta') > Q(b(\theta'))$. Moreover, substituting the above equality in (49), we obtain

$$M'(\theta') = \frac{1}{\beta}\frac{dg^f(\theta')}{d\theta'} U'(g^f(\theta')) \left[Q(\theta') - Q(b(\theta')) + Q'(b(\theta'))b'(\theta')(b(\theta') - \theta')\right].$$

Since $\frac{1}{\beta}\frac{dg^f(\theta')}{d\theta'} U'(g^f(\theta')) > 0$, $Q(\theta') > Q(b(\theta'))$ (as just established), and $Q'(b(\theta')) > 0$ (by $b(\theta') \geq \widehat{\theta}$), it follows that $M'(\theta') > 0$. □

We now proceed to prove the claims in the proposition. Suppose first that condition (26) holds. Suppose by contradiction that an optimal deficit limit sets $\theta^* < \theta_b$ and thus $\theta^{**} = b(\theta^*) < \overline{\theta}$. Since this is an interior optimum, the first-order condition $M(\theta^*) = 0$ must be satisfied, and by the proof of Lemma 3, $\theta^{**} \geq \widehat{\theta}$. Hence, it follows from Lemma 6 that $M'(\theta^*) > 0$. Now note that by condition (26), $M(\theta_b) \leq 0$, that is, a perturbation that marginally decreases the threshold from θ_b weakly reduces social welfare. Since $M(\theta^*) = 0$, $M'(\theta^*) > 0$, and $M(\theta)$ is continuous, it follows that there must exist a point θ_0, with $\theta^* < \theta_0 \leq \theta_b$, such that $M(\theta_0) = 0$ and $M'(\theta_0) \leq 0$. However, given $M(\theta_0) = 0$ and $b(\theta_0) > \theta^{**} \geq \widehat{\theta}$, Lemma 6 implies $M'(\theta_0) > 0$, a contradiction. It follows that a deficit limit with $\theta^* < \theta_b$ cannot be optimal. Moreover, by Lemma 5, a deficit limit with $\theta^* > \theta_b$ cannot be optimal either, and thus the unique optimal rule is a deficit limit with threshold $\theta^* = \theta_b$.

Suppose next that condition (26) does not hold. Then $M(\theta_b) > 0$, which means that a perturbation that marginally decreases the threshold from θ_b strictly increases social welfare. It follows that an optimal deficit limit cannot set $\theta^* = \theta_b$, and by Lemma 5, an optimal deficit limit must set $\theta^* < \theta_b$. Moreover, by Lemma 6, such an optimum must

be unique. The reason is that any interior optimum must satisfy the first-order condition $M(\theta^*) = 0$. Given $M(\theta)$ continuous, the existence of multiple optima would then require the existence of a point θ' with $M(\theta') = 0$ and $M'(\theta') \leq 0$, contradicting Lemma 6.

Finally, we show that if (26) does not hold, the unique solution $\theta^* < \theta_b$ must satisfy $\theta^* > \theta_e$. Note that θ_e is defined by equation (23), which implies $Q(\theta_e) > 0$ and, for any $\theta' \in (\theta_e, \overline{\theta})$, $\int_{\theta_e}^{\theta'} Q(\theta)d\theta > 0$. It follows that if $\theta^* < \theta_e$, then $\int_{\theta^*}^{b(\theta^*)} Q(\theta)d\theta > 0$. Now if $\theta^* < \theta_e < \theta_b$, the first-order condition $M(\theta^*) = 0$ must be satisfied, and by the proof of Lemma 3, $b(\theta^*) \geq \widehat{\theta}$. However, as shown in the proof of Lemma 6 above, these conditions require $\int_{\theta^*}^{b(\theta^*)} Q(\theta)d\theta < 0$, yielding a contradiction. We conclude that $\theta^* < \theta_e$ cannot hold.

B.3 Proof of Proposition 4

Let $\theta^L, \theta^H \in \Theta$ and $\Delta > 0$ be defined as in Definition 2. We prove the proposition by proving the following three claims.

Claim 1. If Assumption 2 is strictly violated and a maximally enforced deficit limit $\{\theta^*, \theta^{**}\}$ is a solution to (11)-(15) for given values $\{\underline{V}, \overline{V}\}$, then $\theta^* \leq \theta^L$ and $\theta^{**} \geq \theta^H$.

Proof of Claim 1. Suppose Assumption 2 is strictly violated. Suppose by contradiction that a maximally enforced deficit limit with $\theta^* > \theta^L$ is a solution to (11)-(15). Then analogous to Step 1 in the proof of Proposition 1, consider a perturbation that drills a hole in the spending schedule in the range $[\theta^L, \theta^L + \varepsilon]$ for arbitrarily small $\varepsilon > 0$ satisfying $\theta^L + \varepsilon < \min\{\theta^*, \theta^L + \Delta\}$. This perturbation is incentive feasible. Moreover, since $Q(\theta)$ is strictly increasing in this range, the arguments in Step 1 in the proof of Proposition 1 imply that this perturbation strictly increases social welfare. It follows that $\theta^* > \theta^L$ cannot hold and we must have $\theta^* \leq \theta^L$.

Next, suppose by contradiction that a maximally enforced deficit limit with $\theta^{**} < \theta^H$ is a solution to (11)-(15). Then consider a perturbation of the allocation for types $\theta \in [\theta^H - \varepsilon, \theta^H]$ for arbitrarily small $\varepsilon > 0$ satisfying $\theta^H - \varepsilon > \max\{\theta^{**}, \theta^H - \Delta\}$. The perturbation assigns these types a constant spending rate $\widehat{g}$ and continuation value $\widehat{V}$ so as to leave the government welfare of types $\theta^H - \varepsilon$ and θ^H unchanged. This perturbation is incentive feasible, and we can show that it strictly increases social welfare. Since the self-enforcement constraint (13) binds for types $\theta^H - \varepsilon$ and θ^H under the original and thus the perturbed allocation, the representation of government welfare in (17) implies

$$\int_{\theta^H - \varepsilon}^{\theta^H} (U(\widehat{g}) - U(g^f(\theta)))d\theta = 0. \tag{51}$$

Using the representation in (18), the change in social welfare from the perturbation is equal to

$$\frac{1}{\beta}\int_{\theta^H-\varepsilon}^{\theta^H}(U(\widehat{g})-U(g^f(\theta)))Q(\theta)d\theta. \tag{52}$$

This is an integral over the product of two terms. Note that $U(\widehat{g})-U(g^f(\theta))$ is strictly decreasing in θ and $Q(\theta)$ is also strictly decreasing in θ over the range $[\theta^H-\varepsilon,\theta^H]$. Hence, the two terms are positively correlated with one another, which implies that the change in social welfare is strictly greater than

$$\int_{\theta^H-\varepsilon}^{\theta^H}(U(\widehat{g})-U(g^f(\theta)))d\theta\int_{\theta^H-\varepsilon}^{\theta^H}Q(\theta)d\theta, \tag{53}$$

which equals zero by (51). It follows that the perturbation strictly increases social welfare; therefore, $\theta^{**}<\theta^H$ cannot hold and we must have $\theta^{**}\geq\theta^H$.

Claim 2. Suppose Assumption 2 is strictly violated. There exist continuation values $\{\underline{V},\overline{V}\}$ for which no solution to (11)-(15) is a maximally enforced deficit limit.

Proof of Claim 2. Suppose Assumption 2 is strictly violated. By Claim 1, if a maximally enforced deficit limit $\{\theta^*,\theta^{**}\}$ solves (11)-(15), then $\theta^*\leq\theta^L$ and $\theta^{**}\geq\theta^H$. Consider the indifference condition (20) which defines, for any given θ^*, a unique value of $\theta^{**}>\theta^*$. This condition shows that $(\theta^{**}-\theta^*)$ is continuous in $(\overline{V}-\underline{V})$ and approaches 0 as $(\overline{V}-\underline{V})$ goes to 0. It follows that there exists $\Lambda>0$ such that if $(\overline{V}-\underline{V})<\Lambda$, then $(\theta^{**}-\theta^*)$ is small enough that $\theta^*\leq\theta^L<\theta^H\leq\theta^{**}$ cannot hold. Thus, if the continuation values $\{\underline{V},\overline{V}\}$ satisfy $(\overline{V}-\underline{V})<\Lambda$, a maximally enforced deficit limit is not a solution to (11)-(15).

Claim 3. Suppose Assumption 2 is weakly violated. There exist continuation values $\{\underline{V},\overline{V}\}$ for which not every solution to (11)-(15) is a maximally enforced deficit limit.

Proof of Claim 3. Suppose Assumption 2 is weakly violated and a maximally enforced deficit limit $\{\theta^*,\theta^{**}\}$ is a solution to (11)-(15). Then $\{\theta^*,\theta^{**}\}$ satisfy condition (20) and analogous arguments as in the proof of Claim 2 above imply that there exist continuation values $\{\underline{V},\overline{V}\}$ such that $\theta^*\leq\theta^L<\theta^H\leq\theta^{**}$ cannot hold. This means that given such continuation values, any maximally enforced deficit limit $\{\theta^*,\theta^{**}\}$ solving (11)-(15) must have either $\theta^*>\theta^L$ or $\theta^{**}<\theta^H$ (or both). Suppose first that $\theta^*>\theta^L$. Then consider a perturbation as in the proof of Claim 1 above which drills a hole in the range $[\theta^L,\theta^L+\varepsilon]$ for arbitrarily small $\varepsilon>0$ satisfying $\theta^L+\varepsilon<\min\left\{\theta^*,\theta^L+\Delta\right\}$. The same arguments as in the proof of Claim 1, given $Q'(\theta)\geq 0$ for $\theta\in[\theta^L,\theta^L+\varepsilon]$, imply that this perturbation weakly increases social welfare. The resulting allocation is therefore a solution to (11)-(15),

and it is not a maximally enforced deficit limit.

Suppose next that $\theta^{**} < \theta^H$. Then consider a perturbation as in the proof of Claim 1 above which modifies the allocation for types $\theta \in [\theta^H - \varepsilon, \theta^H]$ for arbitrarily small $\varepsilon > 0$ satisfying $\theta^H - \varepsilon > \max\{\theta^{**}, \theta^H - \Delta\}$. The perturbation assigns these types a constant spending rate $\widehat{g}$ and continuation value $\widehat{V}$ so as to leave the government welfare of types $\theta^H - \varepsilon$ and θ^H unchanged. The same arguments as in the proof of Claim 1, given $Q'(\theta) \leq 0$ for $\theta \in [\theta^H - \varepsilon, \theta^H]$, imply that this perturbation weakly increases social welfare. The resulting allocation is therefore a solution to (11)-(15), and it is not a maximally enforced deficit limit.

B.4 Proof of Proposition 5

The proof of this proposition is analogous to the proof of Proposition 2, with the uniqueness claim following from the same logic as in the proof of Proposition 3. We therefore describe this proof only briefly here, focusing on the steps that are different from those in the previous proofs.

Observe first that Proposition 1 applies when solving the program in (16), by the same arguments as when solving the program in (11). The reason is that perturbations that apply whenever $Q'(\theta) > 0$ in the maximization of social welfare will now apply whenever $Q'(\theta) < 0$ in the minimization of social welfare, and vice versa.

The analog of Lemma 3 also holds: in any solution to (16), $V(\theta)$ is weakly increasing, right-continuous at $\theta = \underline{\theta}$, and left-continuous at $\theta = \overline{\theta}$ with $V(\overline{\theta}) = \overline{V}$. Steps 1-3 of the proof of Lemma 3 are isomorphic in the sense that the arguments applying to types $\theta < \widehat{\theta}$ now apply to types $\theta > \widehat{\theta}$, and vice versa. However, there are two notable modifications to the proof of this lemma; we describe them next.

The first modification involves Case 2 in Step 2 in the proof of Lemma 3. In that case, we take a segment $[\theta^L, \theta^H]$ with $\theta^H = \overline{\theta}$ and perform a segment-shifting perturbation that marginally reduces g and lowers V so as to keep the government welfare of type θ^L unchanged. In contrast, we would now take a segment $[\theta^L, \theta^H]$ with $\theta^L = \underline{\theta}$ and perform a segment-shifting perturbation that marginally increases g and lowers V so as to keep the government welfare of type θ^H unchanged. Note that using the logic of the representation in (18) but with reference type $\overline{\theta}$ instead of $\underline{\theta}$, we can write social welfare (normalized by debt) as

$$\frac{1}{\beta}\overline{\theta}U(g(\overline{\theta})) + W(x(\overline{\theta})) + \delta V(\overline{\theta}) + \frac{1}{\beta}\int_{\underline{\theta}}^{\overline{\theta}} U(g(\theta))Q_p(\theta)d\theta, \tag{54}$$

where

$$Q_p(\theta) = Q(\theta) - 2(1 - F(\theta)). \tag{55}$$

Using this representation, the change in social welfare from the perturbation is equal to

$$U'(g) \int_{\theta^L}^{\theta^H} Q_p(\theta) d\theta.$$

Since $Q_p(\theta) < 0$ for all $\theta \in \Theta$, it follows that the perturbation strictly reduces social welfare. This contradicts the assumption that the original allocation would solve (16).

The second modification concerns an argument in Step 3 in the proof of Lemma 3. In that argument, we rule out the possibility that $V(\theta) = \underline{V}$ for all $\theta \in \Theta$ in a solution to (11). We can show that this possibility cannot arise in a solution to (16) either, although the argument is different. To show this here, suppose by contradiction that $V(\theta) = \underline{V}$ for all $\theta \in \Theta$ in a solution to (16). The self-enforcement constraint (13) then implies $g(\theta) = g^f(\theta)$ for all $\theta \in \Theta$. Moreover, given the representation in (18), social welfare equals

$$\frac{1}{\beta}\underline{\theta} U(g^f(\underline{\theta})) + W(x^f(\underline{\theta})) + \delta \underline{V} \; + \frac{1}{\beta} \int_{\underline{\theta}}^{\overline{\theta}} U(g^f(\theta)) Q(\theta) d\theta. \tag{56}$$

Consider a global perturbation in which all types $\theta \in \Theta$ are assigned the allocation corresponding to a maximally enforced surplus limit $\{\theta_p^*, \theta_p^{**}\}$ with $\theta_p^{**} > \underline{\theta}$ and $Q(\theta_p^{**}) < 0$ (and θ_p^{**}'s self-enforcement constraint binding given the flexible allocation of θ_p^*, as defined in Proposition 5). Note that $Q(\theta_p^{**}) < 0$ is feasible since $Q(\overline{\theta}) < 0$. Under this construction, $Q(\theta) < 0$ for all $\theta \geq \theta_p^{**}$ by Assumption 2. Using an analogous representation to (56) and taking into account that the perturbation keeps type $\underline{\theta}$'s allocation unchanged, we find that the change in social welfare from the perturbation is equal to

$$\int_{\theta_p^{**}}^{\theta_p^*} (U(g^f(\theta_p^*)) - U(g^f(\theta)) Q(\theta) d\theta. \tag{57}$$

Since $g^f(\theta_p^*) > g^f(\theta)$ and $Q(\theta) < 0$ for all $\theta \in [\theta_p^{**}, \theta_p^*]$, it follows that the perturbation strictly reduces social welfare. Therefore, the original allocation with $V(\theta) = \underline{V}$ for all $\theta \in \Theta$ is not a solution to (16).

Given Proposition 1 and the analog of Lemma 3 as described above, the next step is to show that $g(\theta)$ is continuous for $\theta \geq \theta_p^{**}$. Here analogous arguments to those in the proof of Proposition 2 apply. The optimality of a binding minimum spending limit then also follows from analogous arguments as in that proof, with the exception that we now appeal to the

representation of social welfare in (54). Finally, uniqueness of the optimal limit follows from analogous logic as in the proof of Proposition 3.

B.5 Sufficient Conditions for $\overline{V} > \underline{V}$

Our analysis in Section 4 takes the set of feasible continuation values $[\underline{V}, \overline{V}]$ as given and assumes $\overline{V} > \underline{V}$. The following lemma provides a sufficient condition under which this inequality is satisfied and thus continuation values can be used as reward and punishment to discipline the government. This condition amounts to the discount factor being sufficiently high, or the government's present bias being sufficiently severe, so the current government places a high enough value on restraining future public spending. We maintain Assumption 2 for this result.

Lemma 7. *Suppose $\delta > \widehat{\delta}$, where $\widehat{\delta} \equiv \frac{1}{1-Q(\widehat{\theta})} \in (0,1)$. Then $\overline{V} > \underline{V}$.*

Proof. Take $\delta > \widehat{\delta}$ and suppose by contradiction that $\overline{V} = \underline{V}$. Then by the self-enforcement constraint (13), in equilibrium the government chooses its flexible spending rate at all dates, that is, $g(\theta) = g^f(\theta)$ and $V(\theta) = \overline{V} = \underline{V}$ for all $\theta \in \Theta$. Using the representation of social welfare in (18), this yields

$$\underline{V} = \frac{1}{\beta}\underline{\theta}U(g^f(\underline{\theta})) + W(x^f(\underline{\theta})) + \delta\underline{V} + \frac{1}{\beta}\left(\int_{\underline{\theta}}^{\overline{\theta}} U(g^f(\theta))Q(\theta)d\theta\right). \tag{58}$$

Given this value of $\underline{V}$, and for a given threshold $\theta^{**} \in (\underline{\theta}, \overline{\theta})$ and $\varepsilon \geq 0$, let us construct an equilibrium with the following allocation:

$$\{g(\theta), V(\theta)\} = \begin{cases} \left\{\min\{g^f(\theta), g^f(\theta^{**}-\varepsilon)\}, \overline{V}(\theta^{**}, \varepsilon)\right\} & \text{if } \theta \leq \theta^{**}, \\ \left\{g^f(\theta), \underline{V}\right\} & \text{if } \theta > \theta^{**}, \end{cases} \tag{59}$$

where

$$\theta^{**}U(g^f(\theta^{**}-\varepsilon)) + \beta W(x^f(\theta^{**}-\varepsilon)) + \beta\delta\overline{V}(\theta^{**},\varepsilon) - \theta^{**}U(g^f(\theta^{**})) - \beta W(x^f(\theta^{**})) - \beta\delta\underline{V} = 0 \tag{60}$$

and $\overline{V}(\theta^{**}, \varepsilon)$ is the corresponding social welfare:

$$\begin{aligned} \overline{V}(\theta^{**},\varepsilon) \;=\; & \frac{1}{\beta}\underline{\theta}U\left(\min\left\{g^f(\underline{\theta}), g^f(\theta^{**}-\varepsilon)\right\}\right) + W\left(\max\left\{x^f(\underline{\theta}), x^f(\theta^{**}-\varepsilon)\right\}\right) + \delta\overline{V}(\theta^{**},\varepsilon) \\ & + \frac{1}{\beta}\left(\begin{array}{c} \int_{\underline{\theta}}^{\theta^{**}-\varepsilon} U\left(g^f(\theta)\right)Q(\theta)\,d\theta + \int_{\theta^{**}-\varepsilon}^{\theta^{**}} U\left(g^f(\theta^{**}-\varepsilon)\right)Q(\theta)\,d\theta \\ + \int_{\theta^{**}}^{\overline{\theta}} U\left(g^f(\theta)\right)Q(\theta)\,d\theta \end{array}\right). \end{aligned} \tag{61}$$

The system (58)-(61) corresponds to an equilibrium under a maximally enforced deficit limit, associated with maximum spending rate $g^f(\theta^{**} - \varepsilon)$. Government types $\theta \leq \theta^{**}$ abide to the limit and receive continuation value $\overline{V}(\theta^{**}, \varepsilon)$ (associated with restarting the equilibrium in the next period), whereas types $\theta > \theta^{**}$ spend at their flexible rate and receive continuation value $\underline{V}$ (associated with choosing the flexible spending rate at all future dates).

For any θ^{**}, the system admits a solution with $\varepsilon = 0$, in which case $\overline{V}(\theta^{**}, \varepsilon) = \underline{V}$. Suppose the system also admitted a solution with $\varepsilon > 0$ for some $\theta^{**} \in (\underline{\theta}, \overline{\theta})$. In this case, equation (60) would imply $\overline{V}(\theta^{**}, \varepsilon) > \underline{V}$. Moreover, since the system (58)-(61) is incentive feasible and satisfies (12)-(15), this would imply $\overline{V} \geq \overline{V}(\theta^{**}, \varepsilon) > \underline{V}$, yielding the contradiction that we are seeking. To prove the claim, we therefore proceed by showing that the system (58)-(61) admits a solution with $\varepsilon > 0$ for some $\theta^{**} \in (\underline{\theta}, \overline{\theta})$. There are two cases to consider.

Case 1: Suppose $\widehat{\theta} > \underline{\theta}$. Choose $\theta^{**} = \widehat{\theta}$. A given $\varepsilon \geq 0$ yields (58), (59), and (61). To establish the existence of a solution with some $\varepsilon > 0$, we must thus verify that the left-hand side of (60) equals 0 for such a value. Note that the left-hand side of (60) equals 0 for $\varepsilon = 0$, is strictly negative for ε sufficiently close to $\widehat{\theta}$, and is continuous in ε. It follows that, if the left-hand side of (60) is positive for $\varepsilon > 0$ arbitrarily close to 0, then there exists $\varepsilon > 0$ for which the left-hand side of (60) equals 0. Consider the derivative of the left-hand side of (60) with respect to ε (where we have substituted with $\theta^{**} = \widehat{\theta}$ and used the fact that $g^f(\theta) = \theta/[\theta + \beta\delta\mathbb{E}[\theta]/(1-\delta)]$):

$$-\frac{\widehat{\theta}}{\widehat{\theta} - \varepsilon} + \frac{\widehat{\theta} + \beta\delta\mathbb{E}[\theta]/(1-\delta)}{\widehat{\theta} - \varepsilon + \beta\delta\mathbb{E}[\theta]/(1-\delta)} - \frac{\delta}{1-\delta}\int_{\widehat{\theta}-\varepsilon}^{\widehat{\theta}} \left(\frac{1}{\widehat{\theta} - \varepsilon} - \frac{1}{\widehat{\theta} - \varepsilon + \beta\delta\mathbb{E}[\theta]/(1-\delta)}\right) Q(\theta) d\theta. \tag{62}$$

This derivative equals 0 at $\varepsilon = 0$. Consider the second derivative of the left-hand side of (60) evaluated at $\varepsilon = 0$:

$$-\frac{1}{\widehat{\theta}} + \frac{1}{\widehat{\theta} + \beta\delta\mathbb{E}[\theta]/(1-\delta)} - \frac{\delta}{1-\delta}\left(\frac{1}{\widehat{\theta}} - \frac{1}{\widehat{\theta} + \beta\delta\mathbb{E}[\theta]/(1-\delta)}\right) Q(\widehat{\theta}).$$

This simplifies to

$$\left(\frac{1}{\widehat{\theta}} - \frac{1}{\widehat{\theta} + \beta\delta\mathbb{E}[\theta]/(1-\delta)}\right)\left(-1 - \frac{\delta}{1-\delta} Q(\widehat{\theta})\right) > 0, \tag{63}$$

where the inequality follows from $\delta > \widehat{\delta}$ and $Q(\widehat{\theta}) \leq Q(\overline{\theta}) = -\overline{\theta} f(\overline{\theta})(1-\beta) < 0$.

By (62) and (63), the left-hand side of (60) achieves a local minimum at $\varepsilon = 0$. This implies that the left-hand side of (60) is positive for $\varepsilon > 0$ arbitrarily close to 0, which completes the argument.

Case 2: Suppose $\widehat{\theta} = \underline{\theta}$. Choose $\theta^{**} = \widehat{\theta} + \upsilon$ for some $\upsilon > 0$ which can be taken to be arbitrarily small. Then the same steps as in Case 1 apply, where the condition analogous to (63) for small enough υ holds given $\delta > \widehat{\delta}$ and the continuity of $Q(\theta)$. The claim follows. □