Fiscal Rules with Discretion for an Economic Union

Guillaume Sublet

Université de Montréal and CIREQ †

March 11, 2022

The design of a fiscal rule involves a trade-off between committing governments to a fiscally responsible budget and giving governments the discretion to respond to shocks. What is the optimal degree of discretion for deficit-biased governments that are facing shocks to their fiscal needs? The tail of the distribution of shocks determines the optimal degree of discretion. If the tail is thin, an optimal rule features a cap on public spending enforced by off-equilibrium sanctions. If the tail is thick, an optimal rule grants more discretion than a cap could achieve at the cost of on-equilibrium sanctions. An optimal rule featuring on-equilibrium sanctions also features a threshold below which public spending is exempt from sanctions. The optimal exemption balances a loss of discipline on low levels of spending with an economy of sanctions on high levels of spending. These findings suggest avenues to reform the Stability and Growth Pact.

Keywords: Fiscal Rule, Discretion, Economic Union, Sanctions, Exemptions, Delegation, Mechanism Design without Transfers.

JEL classification: D02, D82, E02, E6, F36, F45, F55, H1, H62, P16

† Email: guillaume.sublet@umontreal.ca.

I thank Manuel Amador and Pierre Yared for comments and suggestions that greatly improved the paper. I thank Julien Bengui, Rui Castro, Louphou Coulibaly, Job Boerma, Barış Kaymak, Thomas Phelan, Michel Poitevin, Georgios Stefanidis, Immo Schott, Smriti Upadhyay, and the audiences of academic seminars and conferences for helpful suggestions and comments. I thank Joan Gieseke for expert copyediting. I am responsible for any shortcomings of this paper.

Introduction

Societies delegate the conduct of fiscal policies to their government. A fiscal rule sets the terms of the delegation contract. The design of a fiscal rule involves a fundamental trade-off between the need for discretion to respond to shocks and the need for discipline to act in the best interest of the citizenry. A main insight from the literature is that, for a single country, the optimal fiscal rule balances the need for both discretion and discipline in a stark way. Under some conditions, the optimal rule grants discretion below a threshold and imposes discipline with maximally enforced sanctions above the threshold.[2] In this paper, I study the design of a fiscal rule for an economic union. While the main insight from the literature prevails under more stringent conditions, fiscal rules featuring mild sanctions are optimal under an alternative set of conditions.

Fiscal rules are an important component of the constitution of economic unions. Unlike a fiscal union, an economic union's lack of fiscal integration limits transfers across countries. As a result, the key trade-off remains between discretion and discipline. Unlike a single country, an economic union can use financial sanctions to discipline its government members and use any revenue from the sanctions as contributions to its budget. For instance, the Excessive Deficit Procedure of the Stability and Growth Pact specifies non-interest-bearing deposits and fines starting at 0.2% of GDP for deficits above 3% of GDP. The revenues from fines are earmarked for the European Stability Mechanism. I argue that the lower cost of financial sanctions relative to non-financial sanctions matters for the optimal structure and stringency of a fiscal rule. How then should an economic union design its fiscal rule?

To answer this question, I solve the problem of a central authority designing a fiscal rule for a union of small open economies. The central authority uses financial sanctions, and each small open economy is subject to the usual tensions studied in the literature. As such, if financial sanctions confer no advantage over non-financial sanctions, the model nests a standard model from the literature (Amador, Werning, and Angeletos (2006)). Each government makes spending and borrowing decisions in response to shocks to its citizenry's need for public spending. The objective of each government is present biased in the sense that it discounts the future at a higher rate than the citizenry. The present bias captures in a reduced form the incentive to overborrow

[2] See the seminal paper of Amador, Werning, and Angeletos (2006) and the recent contribution of Halac and Yared (2022a).

on the part of government members of an economic union, whether the overborrowing is due to the common pool problem caused by a common monetary authority that cannot commit or is simply a result of the shortsightedness caused by political turnover. The combination of shocks and a present-biased objective calls for both discretion and discipline.

The design of a fiscal rule is subject to two constraints. The first constraint captures a limitation caused by the lack of fiscal integration. The rule can only levy sanctions; it cannot use transfers. In the context of Europe, Brunnermeier, James, and Landau (2016) attribute the limits on transfers between European countries to two main concerns: the risk of moral hazard and the potential loss of sovereignty over the national budget. The second constraint gives rise to a trade-off between discretion and discipline. The prescription of the rule is not contingent on the shock realization. For instance, the prescription of the European fiscal rule could hardly have been contingent on the severity of the shock to public spending needs resulting from the banking crisis in Ireland.

The last defining characteristic of a rule is its enforcement mechanism. A single country necessarily relies on non-financial sanctions, such as costly fiscal adjustments, for the enforcement of its fiscal rule. An economic union can use financial sanctions as an additional means of enforcement. Irrespectively of the instrument used, sanctions inevitably penalize both the citizenry and the government. Financial sanctions, however, differ from non-financial sanctions in the relative penalty they impose on the citizenry and the government because the economic union can use the revenues from financial sanctions to partially compensate the citizenry. To avoid undoing the discipline imposed on the government, the use of the revenues must compensate the citizenry of the economic union more than it benefits the government. I show that the asymmetry in the compensation depends on whether the revenues from sanctions expand the budget of the economic union—as the European Union does—or reduce the members' contribution to the budget of the union. To focus on the key trade-off between discipline and discretion rather than the choice of instrument, I summarize the advantage of financial sanctions by a welfare weight on the cost of meting out sanctions.

The analysis covers the entire range of relevant welfare weights. At one end of the range, the economic union perfectly recycles the revenues from sanctions. Since meting out sanctions would then entail no loss of welfare, the optimal rule implements the first-best allocation with a marginal sanction schedule that corrects for the deficit bias. At the other end of the range, revenues from sanctions are wasted, and financial sanctions have no advantage over non-financial sanctions. The model then coincides with the model of delegation in Amador, Werning, and Angeletos (2006). The relevant case for an economic union lies in between the two polar cases.

The analysis delivers two main findings. The first main finding resolves the key trade-off between discipline and discretion for an economic union. The optimal sanction schedule balances the need for discipline, governed by the degree of present bias, and the need for discretion, governed by the thickness of the tail of the distribution of shocks.

A main insight from the literature carries over to the context of an economic union. Under some conditions, a cap on public spending is optimal. The lower cost associated with financial sanctions implies that the conditions are more stringent than they are for non-financial sanctions. I interpret the conditions in terms of two characteristics of the economic union. First, granting discretion below the cap is optimal for economic unions whose members have a low degree of present bias. Second, imposing off-equilibrium sanctions above the threshold is optimal if the distribution of shocks has a thin tail. If the tail is thin, large fiscal needs are relatively unlikely, and the optimal rule features prohibitively large (i.e., off-equilibrium) sanctions.

In contrast, if the tail of the distribution of shocks is thick, large fiscal needs are relatively likely, and the optimal rule is either no rule or one that provides discipline with mild (i.e., on-equilibrium) sanctions. On-equilibrium sanctions achieve a finer balance between discipline and discretion than a cap could achieve. I draw on insights from the first-order approach to determine a candidate marginal sanction schedule. It is only a candidate because it solves a doubly relaxed problem that ignores the standard monotonicity constraint and the constraint on transfers due to the lack of fiscal integration.

The second main finding shows how to best address the limitations due to the lack of fiscal integration. To comply with the limit on transfers, it is optimal to truncate the candidate marginal sanction schedule. The structure of the sanction schedule at and above the truncation threshold depends on the severity of the degree of present bias. For a low degree of present bias, the candidate marginal sanction schedule takes negative values below a threshold. The optimal truncation discards the negative part of the candidate marginal sanction schedule. The resulting sanction schedule grants discretion below the truncation threshold.

If the degree of present bias is not low, however, the limitations due to the lack of fiscal integration are more severe. As a partial substitute for transfers, an optimal rule features a threshold below which public spending is *exempt* from sanctions. The exemption truncates a positive part of the candidate marginal sanction schedule. The benefit of the exemption is that it lowers the level of the sanction schedule above the truncation threshold. It does so while preserving the same marginal sanction schedule, and hence the same discipline, above the threshold. Below the threshold, however, there is a loss of discipline. The exemption causes a kink (i.e., a change in slope) in the sanction schedule. In contrast, the current design of the Stability and Growth Pact features a notch (i.e., a change in level) from 0 to 0.2% of GDP. The concluding section contains suggestions to reform the Stability and Growth Pact.

This paper offers a methodological contribution. The optimal design of a rule maps to a mechanism design problem with limited transfers. The constraint on transfers calls for Lagrangian techniques (Luenberger (1969), Amador, Werning, and Angeletos (2006), and Amador and Bagwell (2013, 2020)). I use insights from the standard first-order approach to shed light on the optimality conditions of the powerful, yet less commonly used, global Lagrangian methods.

Related literature

This paper relates to the literature on the political economy of fiscal rules in economic unions with limited commitment (Beetsma and Uhlig (1999), Cooper and Kempf (2004), Chari and Kehoe (2007), Chari and Kehoe (2008), Aguiar, Amador, Farhi, and Gopinath (2015), Dovis and Kirpalani (2020), and Dovis and Kirpalani (2021)). Yared (2019) surveys the literature. Some of this literature provides the micro-foundations for modeling the deficit bias on the part of members of an economic union with a present bias and focuses on the optimal stringency of a cap on deficit. The analysis to follow focuses on the structure of an optimal fiscal rule.

This paper relates to the literature on optimal delegation and mechanism design with limited transfers (Holmström (1977), Melumad and Shibano (1991), Martimort and Semenov (2006), Alonso and Matoushek (2008), Kováč and Mylovanov (2009), Ambrus and Egorov (2017)). Methodologically, this paper is closest to Amador, Werning, and Angeletos (2006), Amador and Bagwell (2013), Amador, Bagwell, and Frankel (2018), and Amador and Bagwell (2020). Throughout the analysis, I highlight the connection between the standard first-order approach and the powerful—and increasingly popular—global Lagrangian method. This paper provides easy-to-check conditions on the distribution of types (i.e., fiscal needs) and the degree of bias between the principal and the agents (i.e., the present bias of the governments) to determine whether the optimal rule features on-equilibrium sanctions (also known as "money burning"). This first main finding mirrors a finding in Diamond (1998), who sheds light on the role of the thickness of the tail of the distribution of types for the optimal income tax schedule. Non-linear taxes aim to redistribute with transfers. In contrast, a rule uses sanctions that jointly penalize the principal and the agents to correct a bias. Besides, a rule cannot reward with transfers.

This paper contributes to the literature on the design of rules to discipline a policy-making authority to act in the best interest of the citizenry (Athey, Atkeson, and Kehoe (2005), Amador, Werning, and Angeletos (2006), Ambrus and Egorov (2013), Amador and Bagwell (2013), Halac and Yared (2014), Halac and Yared (2020a), Halac and Yared (2020b), Halac and Yared (2022a), Halac and Yared (2022b), and Halac and Yared (2022c)). This paper builds on the seminal contribution of Amador, Werning, and Angeletos (2006) and the more recent contribution of Halac and Yared (2022a) to study the design of a fiscal rule for an economic union.

This paper relates to the literature on fiscal rules when the government can default. Felli, Piguillem, and Shi (2021) show that the risk of default makes it optimal to introduce a default rule and to condition the stringency of the fiscal rule on the level of debt. A related literature quantitatively evaluates the benefit of fiscal rules (see Hatchondo, Martinez, and Roch (2020), Alfaro and Kanczuk (2019), and Aguiar, Amador, and Fourakis (2020)). This paper calls for the quantitative evaluation of financial sanctions to discipline the members of an economic union.

1 Model

The economic union comprises a continuum of small open economies, each with their own government. Each government decides how much to spend and borrow in response to shocks to the fiscal needs of its citizenry. Each government is present biased in the sense that it discounts the future at a higher rate than its citizenry. The combination of shocks to fiscal needs and a present-biased objective creates a need for both discretion and discipline.

Formally, each small open economy is subject to idiosyncratic shocks to its *fiscal needs*, denoted θ, which is private information to the government.[3] The shocks follow a distribution F whose

[3] Kocherlakota (2016) introduces the broader concept of *nonrulable* information which aptly captures the economic content of this assumption: the prescription of the rule is not contingent on the realization of the shock.

support is an interval Θ with lower bound $\underline{\theta} > 0$ and supremum $\bar{\theta} > \underline{\theta}$.[4] Although the supremum can be infinite, the first moment is assumed to be finite. The distribution F is twice continuously differentiable with density f. The *tail* of the distribution refers to $1 - F$.

The preferences of each government over the allocation of public spending over time are represented by the objective function

$$\theta U(g) + \beta W(x), \tag{1}$$

where the utility index $\theta U(\cdot)$ denotes the utility from public spending $g \geq 0$, and the continuation value $W(\cdot)$ is a function of future assets x.[5] Both utility indexes are twice continuously differentiable and strictly increasing. The index U is strictly concave and satisfies Inada conditions and the index W is concave. The degree of *present bias* of the government is $1 - \beta \in (0, 1]$ and, to simplify the notation, the discount factor of the citizenry is subsumed into the continuation value. The present bias captures the overborrowing tendency on the part of members of an economic union that share a common monetary authority lacking the ability to commit to a policy plan. Other political economy frictions, such as political turnover or household heterogeneity in discount rates, also induce a present-biased government (for a survey, see Yared (2019)).

Each government's budget constraint is

$$g + x + \tau_f(g) = T, \tag{2}$$

where $T > 0$ denotes the government revenues, and τ_f denotes the financial sanction schedule, $\tau_f(g) \geq 0$ for $g \geq 0$. The non-negativity constraint on sanctions models the limit on transfers due to the lack of fiscal integration.[6] For simplicity, the gross interest rate is exogenous and normalized to one. The timing of financial sanctions is irrelevant for the government because it can borrow or save to allocate the burden of financial sanctions over time.

[4] Following Section 5.4 of Amador, Werning, and Angeletos (2006), Appendix OB illustrates how shocks to fiscal needs can be interpreted as shocks to government revenues with a CARA utility index.

[5] Amador, Werning, and Angeletos (2006) show that the results apply to a multi-period environment with iid shocks. Halac and Yared (2014) study an infinite horizon environment with persistent shocks and Halac and Yared (2022c) study the design of self-enforced fiscal rules.

[6] Atkeson and Lucas (1992) study the case with transfers but without present bias in an infinite horizon version of this economy. Amador, Werning, and Angeletos (2004) and Galperti (2015) study the case with transfers and present bias.

As far as the objective of a government is concerned, financial sanctions and non-financial sanctions are equivalent. Substituting the budget constraint (2) into the objective of the government (1) gives

$$\theta U(g) + \beta W(T - g) - \beta \tau(g),$$

where the mapping of the financial sanction schedule $\tau_f(\cdot)$ into a sanction schedule $\tau(\cdot)$ follows from $\tau(g) = W(T - g) - W(T - g - \tau_f(g))$. The objective is to design a *fiscal rule*, which is a sanction schedule $\tau(\cdot)$ that satisfies the *no-transfer constraint* $\tau(g) \geq 0$ for $g \geq 0$.

The welfare of the economic union is the aggregation of the welfare of the citizenry of each member. Assuming the law of large numbers, the aggregation over members of the economic union and the expectation over shocks reduces to the single integral in the following objective:

$$\int_{\Theta} \left[\theta U(g(\theta)) + W(T - g(\theta)) - \rho \tau(g(\theta))\right] dF(\theta), \tag{3}$$

where ρ denotes the *welfare weight* on the cost of sanctions for the economic union.[7] An economic union with $\rho = 0$ *perfectly recycles* the revenues from sanctions since meting out sanctions on the governments has no effect on the welfare of the union. An economic union with $\rho \in (0, 1)$ recycles sanctions, albeit with a loss, hence financial sanctions are *partially wasteful*. For $\rho = 1$, sanctions are wasteful, which corresponds to the case of *non-financial sanctions*.[8] This paper covers the full range of the welfare weights, $\rho \in [0, 1]$.

The welfare weight ρ summarizes the extent to which forming an economic union matters for the design of a fiscal rule. It depends on the way in which the economic union recycles the revenues from sanctions. Appendix A elaborates two models of the budget of the economic union to provide micro-foundations for the welfare weight ρ. If the economic union uses the revenues from sanctions to expand its budget, then ρ measures the degree of home bias in the citizenry's preferences for public spending. Instead, if it uses the revenues from sanctions to reduce the contribution of its members to its budget, then ρ measures the degree to which the budget "leaks".

[7] Unless the continuation value is linear, aggregating welfare *after* mapping financial sanctions into sanctions is not without loss. It implicitly assumes that only the first moment of the distribution of financial sanctions matters for welfare (for more details, see Appendix A). On the upside, the analysis remains tractable.

[8] An economic union with $\rho = 1$ reduces to the economic environment in Amador, Werning, and Angeletos (2006). See Halac and Yared (2022a) for an environment with $\rho = 1$ and limited enforcement.

Some useful definitions. An *allocation* is a distribution of public spending $g(\cdot)$ with $g(\theta) \geq 0$ for $\theta \in \Theta$. A fiscal rule $\tau(\cdot)$ *implements* an allocation $g(\cdot)$ if for $\theta \in \Theta$:

$$g(\theta) \in \arg\max_{g \geq 0} \; \theta U(g) + \beta W(T-g) - \beta\tau(g). \tag{4}$$

Consider a rule $\tau(\cdot)$ and the allocation it implements $g(\cdot)$. A strictly positive sanction on g is *on-equilibrium* if there exists $\theta \in \Theta$ such that $g(\theta) = g$, i.e., a government with fiscal needs θ finds it worthwhile to incur the sanction. Otherwise, the sanction is *off-equilibrium*. Let g_d denote the *discretionary allocation*, which is the allocation that solves the government problem (4) in the absence of a fiscal rule. Define the *wedge* Δ in the Euler equation of the government as follows: $(1-\Delta(g,\theta))\theta U'(g) = \beta W'(T-g)$. The wedge evaluated at the discretionary allocation is $\Delta(g_d(\theta), \theta) = 0$. A rule that constrains a government to spend less, i.e., $g(\theta) \leq g_d(\theta)$, induces a positive wedge. The *bias* ν denotes the difference, before sanctions, between the objective of the government and the welfare of the citizenry. It reads $\nu(g) = (\beta - 1)W(T-g)$.

1.1 Designing a rule

The optimal design of a rule consists of finding the sanction schedule that maximizes the welfare at the allocation implemented by the rule. This subsection use the revelation principle to map the design of a rule into a mechanism design problem without transfers. The composition of the sanction schedule and the allocation gives the *money-burning* schedule $t(\theta) = \tau(g(\theta))$ for $\theta \in \Theta$. Note that on-equilibrium sanctions impose discipline at the cost of burning money (i.e., it uses resources) whereas off-equilibrium sanctions do not burn money.

Incentive compatibility constraints guarantee the implementability of the allocation as in (4). An allocation $g(\cdot)$ is *incentive compatible* given a money-burning schedule $t(\cdot)$ if

$$\theta U(g(\theta)) + \beta W(T - g(\theta)) - \beta t(\theta) \geq \theta U(g(\hat{\theta})) + \beta W(T - g(\hat{\theta})) - \beta t(\hat{\theta}), \quad \text{for } \theta, \hat{\theta} \in \Theta. \tag{IC}$$

A fiscal rule is *optimal* if the allocation it implements $g(\cdot)$ and the associated schedule $t(\cdot)$ solve

$$\max_{g(\cdot),\; t(\cdot)} \left\{ \int_{\Theta} \big[\theta U(g(\theta)) + W(T-g(\theta)) - \rho t(\theta)\big] dF(\theta) \;\Big|\; \text{(IC) and } t(\theta) \geq 0 \text{ for } \theta \in \Theta \right\}. \tag{5}$$

The intercept of the schedule, if left implicit, is $t(\underline{\theta}) = 0$ and $\tau(g) = 0$ for $g \leq g(\underline{\theta})$. The no-transfer constraint sets program (5) apart from the design of a mechanism with transfers. The solution method exploits powerful Lagrangian techniques to allow for the no-transfer constraint (for a description of the solution method, see Section 6).

1.2 Optimal fiscal rule with perfectly recycled sanctions

For an economic union that perfectly recycles sanctions, i.e., $\rho = 0$, on-equilibrium sanctions can implement the first-best allocation at no cost. The first-best allocation $g^*(\cdot)$ solves the citizenry's Euler equation $\theta U'(g^*(\theta)) = W'(T - g^*(\theta))$ for $\theta \in \Theta$. It suffices to verify that the first-best allocation is compatible with incentives and that the associated sanction schedule satisfies the no-transfer constraint. The first-best allocation is increasing and hence compatible with incentives. Rewriting the Euler equation as $\Delta(g^*(\theta), \theta)\theta U'(g^*(\theta)) = \nu'(g^*(\theta))$ shows that the sanction schedule induces a wedge that is commensurate with the marginal bias. The marginal sanction is set equal to the marginal bias, which is positive, for $g \geq g^*(\underline{\theta})$. For $g < g^*(\underline{\theta})$, $\tau(g) = 0$.

2 Off-equilibrium sanctions

In this section, I show that a main insight from the literature carries over to the context of an economic union. Under more stringent conditions than those found in the literature (i.e., for $\rho = 1$), a cap on public spending is optimal for an economic union with $\rho < 1$.

A cap on public spending is a simple fiscal rule. It grants discretion below a threshold and imposes off-equilibrium sanctions above the threshold. To separate the cost of limiting discretion from the benefit of addressing the government's bias, I decompose the social welfare function into the sum of the objective of the government, the bias, and the cost of meting out sanctions,

$$\int_{\Theta} [\theta U(g(\theta)) + \beta W(T - g(\theta)) - \nu(g(\theta)) - \rho t(\theta)] \, dF(\theta). \tag{6}$$

I define the allocation implemented by the cap in two steps. A first step parameterizes the cap by the fiscal needs fulfilled at the cap.

Definition. *The* discretion and off-equilibrium sanctions *allocation, denoted $g_d^p(\cdot)$, is defined as follows. For $\theta \in \Theta$,*

$$g_d^p(\theta) = \begin{cases} g_d(\theta_p) & \text{for } \theta > \theta_p \\ g_d(\theta) & \text{for } \theta \leq \theta_p. \end{cases}$$

A second step sets the threshold θ_p—where the subscript stands for the prohibitive nature of off-equilibrium sanctions—to satisfy a first-order condition of the Lagrangian method.

I postpone describing the solution method to focus on the economics. I rephrase the first-order condition in terms of derivatives of the components of the social welfare (6) as follows: $\theta_p = \inf \left\{ \tilde{\theta} \in \Theta \mid \text{Inequality (7) holds for } \hat{\theta} \geq \tilde{\theta} \right\}$, and

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\theta \Delta(g_d(\tilde{\theta}), \theta)\right] \mathrm{d}F(\theta) \leq \left(\frac{\nu'(g_d(\tilde{\theta}))}{U'(g_d(\tilde{\theta}))} + \rho \, \frac{\hat{\theta}\Delta(g_d(\tilde{\theta}), \hat{\theta})}{\beta} \right) (1 - F(\hat{\theta})). \tag{7}$$

The wedge on the left-hand side comes from the objective of the government, the first two terms in (6). It determines the cost of limiting discretion. On the right-hand side, the marginal bias determines the benefit of discipline. The last term is the marginal sanction, which is null for $\hat{\theta} = \tilde{\theta}$ and positive for $\hat{\theta} > \tilde{\theta}$.

Inequality (7) is a function of two thresholds: $\tilde{\theta}$ identifies the threshold at which public spending bunches, and $\hat{\theta}$ determines the range over which the bunching is evaluated. If inequality (7) is satisfied for $\tilde{\theta} = \underline{\theta}$ and $\hat{\theta} \geq \underline{\theta}$, then $\theta_p = \underline{\theta}$. Otherwise, the definition of θ_p encapsulates two requirements. First, by continuity, inequality (7) holds with equality for $\hat{\theta} = \tilde{\theta} = \theta_p > \underline{\theta}$. Second, inequality (7) must hold for all $\hat{\theta} \geq \tilde{\theta} = \theta_p$. While the first requirement is well understood, this paper contributes to our understanding of the second requirement.

The first requirement, that is inequality (7) holding with equality for $\hat{\theta} = \tilde{\theta} = \theta_p > \underline{\theta}$, determines the stringency of the cap. It sets θ_p to equate the marginal cost to the marginal benefit of lowering the cap. On the left-hand side, the marginal cost is the loss of discretion for the governments that are constrained by the cap (by the envelope theorem, the cost of marginally expanding the range of fiscal needs that are constrained is null). On the right-hand side, the marginal benefit is the marginal correction of the bias. The wedge is null at the discretionary allocation for $\hat{\theta} = \tilde{\theta}$. The first requirement simplifies to a familiar condition in the literature on the optimal stringency of a cap: $\beta \, \mathbb{E}[\theta | \theta \geq \theta_p] = \theta_p$, for an interior θ_p (see Proposition 1 in Halac and Yared (2018)). The condition does not depend on the welfare weight ρ precisely because the sanctions are off-equilibrium. Without present bias (i.e., $\beta = 1$), there is no benefit to a cap and the cap does not bind (i.e., $\theta_p = \bar{\theta}$). In contrast, if the degree of present bias is severe enough, in the sense that $\beta \mathbb{E}[\theta] \leq \underline{\theta}$, then the cap binds for all fiscal needs (i.e., $\theta_p = \underline{\theta}$) if the second requirement is also satisfied.

The second requirement, that is inequality (7) must hold for $\hat{\theta} \geq \tilde{\theta} = \theta_p$, determines the structure of the fiscal rule above $g_d(\theta_p)$. It ensures that off-equilibrium sanctions above $g_d(\theta_p)$ dominate on-equilibrium sanctions. At any point $\hat{\theta} \geq \theta_p$, the mechanism designer has the choice to continue enforcing the cap, or to offer an alternative while preserving the bunching of public spending between θ_p and $\hat{\theta}$. On the left-hand side of inequality (7), the marginal cost of enforcing the cap at $\hat{\theta}$ is the loss of discretion. On the right-hand side, the marginal bias measures the marginal benefit of enforcing the cap at $\hat{\theta}$. The difference gives the marginal cost, net of the marginal benefit, of enforcing the cap at $\hat{\theta}$. Alternatively, consider a kink in the sanction schedule that preserves the bunching of public spending at $g_d(\theta_p)$ between θ_p and $\hat{\theta}$, but not beyond $\hat{\theta}$. The kink is caused by jump in the marginal sanction schedule from zero to $\hat{\theta}\Delta(g_d(\theta_p), \hat{\theta})U'(g_d(\theta_p))$. A mass $1 - F(\hat{\theta})$ of governments would choose to incur the marginal sanction, which gives the second term in the parenthesis on the right-hand side of inequality (7). In sum, for $\hat{\theta} \geq \tilde{\theta} = \theta_p$, inequality (7) verifies that the marginal cost net of the marginal benefit of enforcing the cap at $\hat{\theta}$ is lower than the marginal cost of resorting to on-equilibrium sanctions.

The second requirement is equivalent to a simple upper bound on the thickness of the tail of the distribution of shocks.

Lemma 1. *Suppose that inequality* (7) *holds with equality for some* $\tilde{\theta} < \bar{\theta}$ *and* $\hat{\theta} = \tilde{\theta}$*. Then, for* $\hat{\theta} \in [\tilde{\theta}, \bar{\theta})$*, inequality* (7) *holds if and only if* $\beta E[\theta|\theta \geq \hat{\theta}] - \rho\hat{\theta} \leq \beta E[\theta|\theta \geq \tilde{\theta}] - \rho\tilde{\theta}$.

The proof is in Appendix OA.1. Intuitively, the thickness of the tail of the distribution of shocks governs the need for discretion. For a sufficiently thin tail, the need for discretion is sufficiently low that granting no discretion above a threshold with off-equilibrium sanctions dominates on-equilibrium sanctions. For instance, for $\rho = \beta$ and a log-concave tail, the first requirement is sufficient for the second requirement. In contrast, for $\rho = \beta$ and a log-convex tail, the second requirement is not satisfied for an interior threshold and θ_p is either $\underline{\theta}$ or $\bar{\theta}$. Although ρ does not alter the stringency of an interior cap, the lower is ρ the more stringent are the conditions under which off-equilibrium sanctions dominate on-equilibrium sanctions.

2.1 Low degree of present bias

The previous subsection focuses on the optimality of off-equilibrium sanctions above a threshold. This subsection focuses on the optimality of granting discretion below the threshold.

Since Amador, Werning, and Angeletos (2006), a lower bound on the elasticity of the density of the distribution of shocks has been understood to imply that granting discretion below a threshold is optimal for a rule enforced by non-financial sanctions. For $\rho < 1$, a more stringent lower bound is needed to exclude financial sanctions below a threshold.

Assumption L. $\frac{\theta f'(\theta)}{f(\theta)} \geq -\frac{1+\rho-\beta}{1-\beta}$.

Assumption L relates the elasticity of the density of the distribution of shocks to the degree of present bias. The following lemma shows that, under some conditions, Assumption L puts an upper bound on the degree of present bias as a function of the inverse of the elasticity of $1 - F$. Appendix OA.2 contains the proof.

Lemma 2. *If $\theta_p \in (\underline{\theta}, \bar{\theta})$ and Assumption L holds for $\theta \leq \theta_p$, then $1 - \beta \leq \rho \frac{1-F(\theta)}{\theta f(\theta)}$ for $\theta \leq \theta_p$.*

The lemma suggests the following definition: the degree of present bias is *low* if there exists $\theta^* > \underline{\theta}$ such that $1 - \beta \leq \rho \frac{1-F(\theta)}{\theta f(\theta)}$ for $\theta \leq \theta^*$. Lemma 2 implies that if the tail of the distribution of shocks does not decrease too fast—in the sense of Assumption L—up to a point after which it decreases fast enough—in the sense of Lemma 1—then the degree of present bias is low.

Proposition 1 (Optimal fiscal rule for a low degree of present bias and a thin tail)**.** *Suppose $\underline{\theta} < \theta_p$ and Assumption L holds for $\theta \leq \theta_p$. A fiscal rule that implements the* discretion and off-equilibrium sanctions *allocation is optimal.*

The proof is in Appendix B.1. The result extends a main insight from the literature to the context of an economic union. For non-financial sanctions, i.e., $\rho = 1$, the proposition nests Proposition 3 in Amador, Werning, and Angeletos (2006).

Proposition 1 relies on two conditions. The first one, Assumption L, implies that the result holds for an economic union with a low degree of present bias up to θ_p. The second condition is embedded in the definition of the threshold θ_p. Lemma 1 shows that the cap binds if the tail is sufficiently thin above θ_p. The conditions in Proposition 1 are more stringent for $\rho < 1$ than they are for non-financial sanctions $\rho = 1$. The rest of the paper characterizes optimal fiscal rules for economic unions that do not satisfy the conditions in Proposition 1.

3 On-equilibrium sanctions

This section contains the main findings. First, under some conditions, the need for discretion associated with a thick tail of the distribution of shocks justifies resorting to on-equilibrium sanctions above a threshold. Second, granting discretion below a threshold addresses the limitations that the no-transfer constraint places on the design of a fiscal rule. To start, I use insights from the first-order approach to characterize a candidate schedule of on-equilibrium sanctions.

The following standard result of the first-order approach exploits the incentive compatibility constraints to characterize, for a given intercept $t(\underline{\theta})$, the money-burning schedule associated with a non-decreasing allocation (Myerson (1981)). The proof is in Appendix OA.3.

Lemma 3 (Incentive compatible allocations)**.** *An allocation $g(\cdot)$ is incentive compatible given a money-burning schedule $t(\cdot)$ if and only if $g(\cdot)$ is non-decreasing and*

$$t(\theta) = t(\underline{\theta}) + \frac{\theta}{\beta}U(g(\theta)) + W(T - g(\theta)) - \frac{\underline{\theta}}{\beta}U(g(\underline{\theta})) - W(T - g(\underline{\theta})) - \frac{1}{\beta}\int_{\underline{\theta}}^{\theta} U(g(\tilde{\theta}))\, d\tilde{\theta}. \quad (8)$$

Lemma 3 is useful in substituting the money-burning schedule with a function of the allocation in the objective of program (5). The resulting objective functional reads as follows:

$$\int_{\Theta} \left[-\nu(g(\theta)) + (1 - \tfrac{\rho}{\beta})\left(\theta U(g(\theta)) + \beta W(T - g(\theta))\right) + \tfrac{\rho}{\beta}\tfrac{1-F(\theta)}{f(\theta)}U(g(\theta))\right] dF(\theta)$$
$$- \tfrac{\rho}{\beta}\left(t(\underline{\theta}) - \underline{\theta}\, U(g(\underline{\theta})) - \beta W(T - g(\underline{\theta}))\right).$$

Maximizing the objective point-wise for $\theta > \underline{\theta}$ implicitly defines a candidate marginal sanction schedule. It is only a candidate for the optimal marginal sanction schedule because the allocation it implements solves a relaxed problem ignoring two constraints: the usual monotonicity condition and the no-transfer constraint.

Definition (Public spending implemented by the candidate marginal sanction)**.** *Let $\rho \in [0, 1)$. For θ such that $\rho - \beta < \rho\frac{1-F(\theta)}{\theta f(\theta)}$, define $g_n(\theta)$ as follows:*

$$\frac{\nu'(g_n(\theta))}{U'(g_n(\theta))} = \theta\Delta(g_n(\theta), \theta) + \frac{\rho}{\beta}\left(\frac{1 - F(\theta)}{f(\theta)} - \theta\Delta(g_n(\theta), \theta)\right). \quad (9)$$

For $\rho = 1$, the candidate marginal sanction is not defined because equation (9) is independent of g_n.[9] It is without loss since Proposition 1 in Halac and Yared (2022a) shows that, for $\rho = 1$, an optimal rule necessarily relies on extreme—bang-bang—sanctions. For $\rho \in [\beta, 1)$, the bound

[9]The subscript stands for the non-prohibitive nature of on-equilibrium sanctions. Combining the terms in (9) that depend on $g_n(\theta)$ gives $\frac{\nu'(g_n(\theta))}{U'(g_n(\theta))} - (1 - \frac{\rho}{\beta})\theta\Delta(g_n(\theta), \theta) = \frac{1-\rho}{1-\beta}\frac{\nu'(g_n(\theta))}{U'(g_n(\theta))} - (1 - \frac{\rho}{\beta})\theta$, which is independent of $g_n(\theta)$ for $\rho = 1$.

on the elasticity of $1 - F$ at θ guarantees that $g_n(\theta)$ is strictly positive. For $\rho < \beta$, g_n is defined for $\theta \in \Theta$.

Equation (9) equates the marginal benefit of on-equilibrium sanctions on the left-hand side to their marginal cost on the right-hand side. For $\rho = 0$, equation (9) reduces to the equation characterizing the first-best allocation (see Subsection 1.2 on optimal rules with perfectly recycled sanctions). For $\rho > 0$, the marginal cost has an additional component. To be compatible with incentives, a marginal sanction on a given spending $g_n(\theta)$ raises the level of the sanction schedule on spending above $g_n(\theta)$, which affects a mass $1 - F(\theta)$ of governments. I refer to the term $\rho\frac{1-F(\theta)}{f(\theta)}$ as the *incentive cost* of a marginal sanction. The wedge measures the extent to which the marginal sanction discourages spending at $g_n(\theta)$.

The weight $1/\beta > 1$ on the incentive cost highlights a difficulty in disciplining a present-biased government. A present-biased government tilts the allocation of the burden of the financial sanction toward the future. As a result, on-equilibrium sanctions tend to impose a disproportionate cost on the citizenry. For $\rho = \beta$, the asymmetry in the welfare weight on the cost of sanctions exactly offsets the additional burden of sanctions caused by the government's present bias.

The characterization of the candidate marginal sanction in (9) ignores two constraints. First, to be compatible with incentives, the allocation must be non-decreasing.

Lemma 4 (Monotonicity of the allocation implemented by the candidate marginal sanction)**.** *Suppose that $g_n(\theta)$ is well-defined for $\theta \in (\theta_*, \theta^*)$. Then g_n is non-decreasing at $\theta \in (\theta_*, \theta^*)$ if and only if the derivative of $\rho\frac{1-F}{f}$ at θ is not smaller than $\rho - \beta$.*

Appendix OA.4 contains the proof. The lemma gives a sufficient condition for the candidate marginal sanction at $g_n(\theta)$ to be on-equilibrium. A marginal sanction at $g_n(\theta)$ is on-equilibrium if the slope of g_n, which is governed by the slope of the inverse hazard rate, is positive at θ.

Second, the sanction schedule must satisfy the no-transfer constraint. A sufficient condition for a non-negative sanction schedule is a non-negative marginal sanction schedule starting from a non-negative intercept. Lemma 5 shows that the marginal sanction is non-negative for a sufficiently low incentive cost.

Lemma 5 (Non-negative candidate marginal sanction)**.** *Suppose that $g_n(\theta)$ is well-defined at $\theta \in \Theta$. Then $g_n(\theta) \le g_d(\theta)$ if and only if $\rho\frac{1-F(\theta)}{\theta f(\theta)} \le 1 - \beta$.*

Appendix OA.5 contains the proof. The proof relies on the observation that the wedge at $g_n(\theta)$ is $\Delta(g_n(\theta),\theta) = -\frac{1}{1-\rho}\big(\rho\frac{1-F(\theta)}{\theta f(\theta)} - (1-\beta)\big)$.

Lemma 4 and Lemma 5 reveal the distinct roles played by the slope and the level of the inverse hazard rate for the design of a rule. The level of the inverse hazard rate governs the incentive cost of a marginal sanction, which is a determinant of the level of the candidate marginal sanction. The slope of the inverse hazard rate determines the need for discretion corrected for the associated incentive cost. I refer to $\frac{d}{d\theta}(\rho\frac{1-F(\theta)}{f(\theta)} - (\rho-\beta)\theta)$ as the *virtual need for discretion*.[10] Intuitively, the slope of the inverse hazard rate governs the thickness of the tail of the distribution of shocks. A thick tail indicates a high virtual need for discretion. To accommodate a high virtual need for discretion, the candidate marginal sanction schedule induces a graduated schedule of sanctions.

To summarize, two conditions guarantee the validity of $g_n(\theta)$ as building block of a solution. The virtual need for discretion, governed by the slope of the inverse hazard rate, is not too low at θ (see Lemma 4). The incentive cost satisfies the following assumption (see the definition of g_n and Lemma 5).

Assumption I. $\rho - \beta < \rho\frac{1-F(\theta)}{\theta f(\theta)} \leq 1-\beta$.

For $\rho \leq \beta$, the left inequality is invariably satisfied. Assumption I suggests two definitions to complement the definition of a low degree of present bias. The degree of present bias is *intermediate* if there exists $\theta^* > \underline{\theta}$ such that Assumption I holds for $\theta \leq \theta^*$. The degree of present bias is *high* if the following assumption holds instead.

Assumption H. $\rho\frac{1-F(\theta)}{\theta f(\theta)} \leq \rho - \beta$ *for* $\theta \in \Theta$.

For the rest of this section, assume $\rho < 1$ so that if Assumption I holds at θ, then $g_n(\theta)$ is defined. The next candidate solution features on-equilibrium sanctions above a threshold and grants discretion below the threshold.

Definition. *The* discretion and on-equilibrium sanctions *allocation, denoted* $g_d^n(\cdot)$*, is defined, for* $\theta \in \Theta$*, as follows:*

$$g_d^n(\theta) = \begin{cases} g_n(\theta) & \text{for } \theta > \theta_n \\ g_d(\theta) & \text{for } \theta \leq \theta_n, \end{cases}$$

where $\theta_n = \inf\left\{\tilde{\theta} \in \Theta \,\middle|\, \text{Assumption I holds for } \theta \geq \tilde{\theta}\right\}$.

[10] Following Myerson (1981), the term virtual qualifies a concept augmented by the incentive cost. I interpret the following measure of the slope of the first-best allocation g^* as the need for discretion: $\beta = \frac{d}{d\theta}\theta(1-\Delta(g^*(\theta),\theta))$.

The threshold is set at the lowest fiscal need such that g_n is well-defined and associated with a non-negative marginal sanction. If $0 < g_n(\theta) \leq g_d(\theta)$ for $\theta \in \Theta$, then Assumption I holds for $\theta \in \Theta$ and $\theta_n = \underline{\theta}$. If $\theta_n \in (\underline{\theta}, \bar{\theta})$, the continuity of both F and f implies that the threshold satisfies $g_n(\theta_n) = g_d(\theta_n)$ if the upper bound in Assumption I binds. If the lower bound in Assumption I is constraining instead, then $\lim_{\theta \to \theta_n} g_n(\theta) = 0$. In contrast, if the set in the definition of θ_n is empty, then set $\theta_n = \bar{\theta}$. For instance, if shocks are Pareto distributed, $1 - F(\theta) = \theta^{-\gamma}$, and the elasticity of the tail $\gamma \in (1, \frac{\rho}{1-\beta})$, then the need for discretion is such that the candidate solution does not feature sanctions since $\theta_n = \bar{\theta}$. The order of θ_n and θ_p depends on the distribution of shocks. For $\rho = \beta$, $\theta_n \leq \theta_p$ for a strictly log-convex $1 - F$ and $\theta_p \leq \theta_n$ for a log-concave $1 - F$.

The *discretion and on-equilibrium* sanctions allocation need not be increasing. Lemma 4 shows that g_n is increasing if and only if the derivative of $\rho\frac{1-F}{f}$ at θ is not smaller than $\rho - \beta$. If $\bar{\theta} < \infty$, the condition cannot be satisfied close to $\bar{\theta}$. For a compact Θ, the need for discipline overwhelms the need for discretion at the top.

Definition. *The* discretion, on-equilibrium, and off-equilibrium sanctions *allocation, denoted $g_d^{np}(\cdot)$, is defined for $\theta \in \Theta$, as follows:*

$$g_d^{np}(\theta) = \begin{cases} g_n(\theta_{np}) & \text{for } \theta \geq \theta_{np} \\ g_n(\theta) & \text{for } \theta_{n'} \leq \theta \leq \theta_{np} \\ g_d(\theta) & \text{for } \theta \leq \theta_{n'}, \end{cases}$$

in which $\theta_{n'} = \inf\left\{\tilde{\theta} \in \Theta \;\middle|\; \text{Assumption I holds for } \theta \in [\tilde{\theta}, \theta_{np}]\right\}$, and
$\theta_{np} = \inf\left\{\tilde{\theta} \in \Theta \mid \text{Inequality (10) holds for } \hat{\theta} \geq \tilde{\theta}\right\}$,

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\theta \Delta(g_n(\tilde{\theta}), \theta)\right] dF(\theta) \leq \left(\frac{\nu'(g_n(\tilde{\theta}))}{U'(g_n(\tilde{\theta}))} + \rho \frac{\hat{\theta}\Delta(g_n(\tilde{\theta}), \hat{\theta})}{\beta}\right)(1 - F(\hat{\theta})). \tag{10}$$

The definition of $\theta_{n'}$ adapts the definition of θ_n for fiscal rules imposing off-equilibrium sanctions above a threshold. The analysis of the definition of θ_{np} follows the analysis of the definition of θ_p with one exception regarding the right-hand side of (10). For a fiscal rule featuring on-equilibrium sanctions, the marginal benefit of a cap is the marginal discipline and, additionally, the economy of the marginal sanction. A strictly positive wedge at $g_n(\theta_{np}) < g_d(\theta_{np})$ captures an additional benefit of off-equilibrium sanctions for a fiscal rule featuring on-equilibrium sanctions. Off-equilibrium sanctions economize on the loss associated with on-equilibrium sanctions. The similarity between the following lemma and Lemma 1 confirms that the economic content of the definitions of θ_p and θ_{np} is otherwise the same. Appendix OA.6 contains the proof.

Lemma 6. *Suppose that inequality* (10) *holds with equality for some* $\tilde{\theta} < \bar{\theta}$ *and* $\hat{\theta} = \tilde{\theta}$. *Then, for* $\hat{\theta} \in [\tilde{\theta}, \bar{\theta})$, *inequality* (10) *holds if and only if* $\beta E[\theta|\theta \geq \hat{\theta}] - \rho\hat{\theta} \leq \beta E[\theta|\theta \geq \tilde{\theta}] - \rho\tilde{\theta}$.

3.1 Low degree of present bias

This subsection contains two findings. First, for a thick tail—defined by a lower bound on the slope of the inverse hazard rate—an optimal fiscal rule imposes graduated on-equilibrium sanctions. Second, to address the limitations that the no-transfer constraint places on the design of a fiscal rule, it is optimal to truncate the candidate marginal sanction schedule below a threshold.

Proposition 2 (Optimal fiscal rule for a low degree of present bias and a thick tail)**.** *Suppose* $\underline{\theta} < \theta_{n'}$ *and Assumption L holds for* $\theta \leq \theta_{n'}$. *If the derivative of* $\rho\frac{1-F}{f}$ *is not smaller than* $\rho - \beta$ *for* $\theta \in [\theta_{n'}, \theta_{np}]$, *then a fiscal rule that implements the* discretion, on-equilibrium, and off-equilibrium sanctions *allocation is optimal.*

The proof is in Appendix B.2. The optimal fiscal rule has three parts. A first part grants discretion below $g_d(\theta_{n'})$. The optimal sanction is null below $g_d(\theta_{n'})$ because, as the next lemma shows, Assumption L implies a low degree of present bias relative to the incentive cost.

Lemma 7 (Implications of Assumption L)**.** *1) If* $\theta_{n'} \in (\underline{\theta}, \bar{\theta})$ *and Assumption L holds for* $\theta \leq \theta_{n'}$, *then* $1 - \beta \leq \rho\frac{1-F(\theta)}{\theta f(\theta)}$ *for* $\theta \leq \theta_{n'}$. *2) Assumption L is equivalent to* $\frac{d}{d\theta}\theta\Delta(g_n(\theta), \theta) \geq \frac{1+\rho-\beta}{1-\beta}\Delta(g_n(\theta), \theta)$.

The proof is in Appendix OA.7. The first statement extends the insight of Lemma 2 to fiscal rules with on-equilibrium sanctions. The second statement shows that Assumption L is equivalent to a bound on the elasticity of the wedge evaluated at g_n. It implies an upper bound on the virtual need for discretion because the wedge evaluated at $g_n(\theta)$ is negatively related to the inverse hazard rate. Lemma 7 shows that the optimal rule grants discretion below $g_d(\theta_{n'})$ because discipline is too costly, not because of a high virtual need for discretion. The optimal fiscal rule grants discretion below a threshold to comply with the no-transfer constraint because the candidate marginal sanction schedule is negative below $g_d(\theta_{n'})$ and zero at $g_d(\theta_{n'})$ (by definition of $\theta_{n'}$). Proposition 2 confirms that it is optimal to discard the negative part of the candidate marginal sanction schedule.

A second part of the fiscal rule imposes graduated on-equilibrium sanctions on spending between $g_d(\theta_{n'})$ and $g_n(\theta_{np})$. The conditions on the derivative of the incentive cost between $\theta_{n'}$ and θ_{np} guarantee that the virtual need for discretion is sufficiently high to justify imposing graduated on-equilibrium sanctions. If g_n is decreasing over a subinterval of $[\theta_{n'}, \theta_{np}]$, the solution would be to "iron" g_n as in Myerson (1981). Ironing the allocation requires a jump in the marginal sanction instead of a continuous schedule of marginal sanctions. The marginal sanction schedule jumps at $g_n(\theta_{np})$ precisely for this reason. A third part of the fiscal rule imposes off-equilibrium sanctions above $g_n(\theta_{np})$. The tail of the distribution of shocks is sufficiently thin above θ_{np} that the need for discipline overwhelms the need for discretion above θ_{np}.

3.2 Intermediate degree of present bias

This previous subsection showed that for a low degree of present bias, it is optimal to truncate the negative part of the candidate marginal sanction schedule. For an intermediate degree of present bias, the limitations due to the lack of fiscal integration are more severe. As a result, it is optimal to truncate a positive part of the candidate marginal sanction schedule.

To gain insights into how the no-transfer constraint matters for the design of the optimal fiscal rule, it is useful to decompose the sanction schedule in two parts: the intercept and the marginal sanction schedule. For an intermediate degree of present bias, the candidate marginal sanction is positive. The no-transfer constraint forces the intercept to be non-negative. Truncating the candidate marginal sanction schedule below a threshold is a partial substitute for a negative intercept. The truncation shifts the sanction schedule downward, as a negative intercept would do. Unlike a negative intercept, however, the truncation entails a loss of discipline because it grants an *exemption* from the candidate marginal sanction below a threshold.

The exemption causes a jump in the marginal sanction schedule and the resulting kink induces the spending of some governments to bunch at the exemption threshold. For the following definition, suppose that Assumption I holds for $\theta \in \Theta$.

Definition. *The* exemption and on-equilibrium sanctions *allocation, denoted $g_x^n(\cdot)$, is defined for $\theta \in \Theta$ as follows:*

$$g_x^n(\theta) = \begin{cases} g_n(\theta) & \text{for } \theta > \theta_x \\ g_n(\theta_x) & \text{for } \theta \leq \theta_x, \end{cases}$$

where $\theta_x = \sup \left\{ \tilde{\theta} \in \Theta \mid \textit{Inequality } (11) \textit{ holds for } \hat{\theta} \leq \tilde{\theta} \right\}$, *and*

$$\frac{\nu'(g_n(\tilde{\theta}))}{U'(g_n(\tilde{\theta}))} F(\hat{\theta}) - \int_{\underline{\theta}}^{\hat{\theta}} \left[\theta \Delta(g_n(\tilde{\theta}), \theta) \right] dF(\theta) \leq \rho \frac{\hat{\theta} \Delta(g_n(\tilde{\theta}), \hat{\theta})}{\beta} (1 - F(\hat{\theta})). \tag{11}$$

The first-order conditions from the Lagrangian methods suggest the definition of the exemption threshold θ_x. Inequality (11) is a function of two thresholds: $\tilde{\theta}$ determines the level at which public spending bunches, and $\hat{\theta}$ determines the range of the bunching. If inequality (11) is satisfied for $\tilde{\theta} = \bar{\theta}$ and $\hat{\theta} \leq \bar{\theta}$, then $\theta_x = \bar{\theta}$. Otherwise, θ_x encapsulates two requirements. First, by continuity, inequality (11) holds with equality at $\hat{\theta} = \tilde{\theta} = \theta_x$. Second, inequality (11) holds for $\hat{\theta} \leq \tilde{\theta} = \theta_x$.

The first requirement determines the leniency of the exemption. It sets θ_x to equate the marginal cost to the marginal benefit of the exemption. On the left-hand side of (11), the marginal cost amounts to the loss of discipline net of the gain in discretion for countries whose fiscal need is below the exemption threshold. On the right-hand side, the marginal benefit is the marginal economy of sanctions. For $\hat{\theta} = \tilde{\theta} = \theta_x$, the exemption from the marginal sanction $U'(g_n(\theta_x))\theta_x \Delta(g_n(\theta_x), \theta_x)$ benefits a mass $1 - F(\theta_x)$ of governments with fiscal needs above θ_x.

The second requirement determines the structure of the fiscal rule below the threshold. It checks that for any $g \leq g_n(\theta_x)$, the exemption dominates any alternative. As a result of the exemption, the marginal sanction schedule jumps from 0 to $U'(g_n(\theta_x))\theta_x \Delta(g_n(\theta_x), \theta_x)$ at $g_n(\theta_x)$. Consider $\hat{\theta} \leq \theta_x$. The mechanism designer can impose a marginal sanction that a government with fiscal need $\hat{\theta}$ would incur, while preserving the bunching of public spending between $\hat{\theta}$ and θ_x. The marginal cost of such a switch to on-equilibrium sanctions is the marginal sanction $U'(g_n(\theta_x))\hat{\theta}\Delta(g_n(\theta_x), \hat{\theta})$. The marginal benefit is the discipline from the marginal sanction net of the loss of discretion for governments with fiscal needs below $\hat{\theta}$. Inequality (11) implies that the marginal benefit of switching from the exemption to on-equilibrium sanctions lies below the marginal cost at any point below the exemption threshold.

The second requirement is equivalent to a sufficiently thick tail of the distribution of shocks.

Lemma 8. *Suppose that inequality* (11) *holds with equality for some* $\tilde{\theta} \in \Theta$ *and* $\hat{\theta} = \tilde{\theta}$*. Then, for* $\hat{\theta} \leq \tilde{\theta}$*, inequality* (11) *holds if and only if*

$$\int_{\hat{\theta}}^{\tilde{\theta}} \left[\rho \tfrac{1-F(\theta)}{f(\theta)} - (\rho - \beta)\theta \right] dF(\theta) \leq \int_{\hat{\theta}}^{\tilde{\theta}} \left[\rho \tfrac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - (\rho - \beta)\tilde{\theta} \right] dF(\theta). \tag{12}$$

The intuition for this result relates the thickness of the distribution of shocks to the marginal benefits and costs of an exemption. The inverse hazard rate governs the incentive cost that a marginal sanction on public spending g imposes on all public spending above g. An exemption benefits the citizenry in two ways. It economizes on the incentive cost of the marginal sanction and it grants discretion below the exemption. The thicker is the tail of the distribution of shocks between $\hat{\theta}$ and θ_x for $\hat{\theta} \leq \theta_x$, the greater are the two benefits. For instance, for $\rho = \beta$, a log-convex tail is sufficiently thick for the first requirement to imply the second requirement.

Although the inequality determining the exemption threshold resembles the inequalities determining the thresholds for off-equilibrium sanctions, their economic content differ. The resemblance stems from the shared origin of these characterizations. Both come from the first-order conditions of the Lagrangian methods. The economic content of the exemption pertains to the limitations due to the lack of fiscal integration. Again, the exemption is an imperfect substitute for a transfer from the economic union to its members. In contrast, the economic content of the threshold for off-equilibrium sanctions pertains to the trade-off between the need for discretion and the need for discipline. The next allocation is defined if Assumption I holds below θ_{np}.

Definition (Exemption, on-equilibrium, and off-equilibrium sanctions)**.** *The* exemption, on-equilibrium, and off-equilibrium sanctions *allocation, denoted $g_x^{np}(\cdot)$, is defined for $\theta \in \Theta$ as follows:*

$$g_x^{np}(\theta) = \begin{cases} g_n(\theta_{np}) & \text{for } \theta > \theta_{np} \\ g_n(\theta) & \text{for } \theta_x \leq \theta \leq \theta_{np} \\ g_n(\theta_x) & \text{for } \theta \leq \theta_x, \end{cases}$$

where the thresholds θ_x and θ_{np} are defined in (11) *and* (10)*, respectively.*

The last candidate allocation sets a cap on public spending to fulfill the average fiscal need.

Definition (Tight cap)**.** *The* tight cap *allocation, denoted $g_c(\cdot)$, is $g_c(\theta) = g_c$ for $\theta \in \Theta$, where g_c solves the following Euler equation: $W'(T - g_c) = \mathbb{E}[\theta]U'(g_c)$.*

The next proposition is the second main result of this paper. For an intermediate degree of present bias, if an optimal fiscal rule features on-equilibrium sanctions, then it features a kink in the sanction schedule caused by an exemption from sanctions below a threshold.

Proposition 3 (Optimal fiscal rules for an intermediate degree of present bias)**.**

1) Suppose that Assumption I holds for $\underline{\theta}$. If the derivative of $\rho\frac{1-F}{f}$ is smaller than $\rho - \beta$ for

$\theta \in \Theta$, *then a fiscal rule that implements the* tight cap *allocation is optimal.*

2) Suppose that Assumption I holds for $\theta \leq \theta_{np}$. *If the derivative of* $\rho\frac{1-F}{f}$ *is not smaller than* $\rho - \beta$ *for* $\theta \in [\underline{\theta}, \theta_{np}]$, *then a fiscal rule that implements the* exemption, on-equilibrium, and off-equilibrium sanctions *allocation is optimal.*

Appendix B.3 contains the proof. The insight regarding the balance between the needs for discretion and discipline from Propositions 1 and 2 carries over to an economic union with an intermediate degree of present bias. The additional insight—the optimality of granting an exemption below a threshold—matters for economic unions with an intermediate degree of present bias. The exemption balances the loss of discipline below θ_x with the gain from a lower level of sanctions meted out above θ_x.

4 High degree of present bias

This subsection shows that for a high degree of present bias, the need for discipline outweigh the need for discretion.

The tight cap is set such that it fulfills the expected fiscal need. Assumption H implies that the *tight cap allocation* constrains governments irrespectively of the shock to their fiscal needs, i.e., $g_c \leq g_d(\underline{\theta})$ (see Appendix OA.10 for the formal statement and its proof). The proof of the following proposition is in Appendix B.4.

Proposition 4 (Optimal fiscal rule for a high degree of present bias)**.** *Suppose that Assumption H holds. A fiscal rule that implements the* tight cap *allocation is optimal.*

5 Examples

This section illustrates the main findings with computed examples.[11] With CRRA utility indices $U(g) = g^{1-\eta}/(1-\eta)$ and $W(x) = (\omega + x)^{1-\eta}/(1-\eta)$, the discretionary allocation is $g_d(\theta) = (\omega + T)\big(1 + (\beta/\theta)^{\frac{1}{\eta}}\big)^{-1}$. The grey line in the panels of Figures 1 and 2 depicts the discretionary allocation. For an economic union with $\rho = \beta$, the allocation implemented by the candidate

[11] The code is available on the GitHub page guillaumesublet/Fiscal_Rules_with_Discretion_Economic_Union.

marginal sanction schedule is

$$g_n(\theta) = (\omega + T)\left(1 + (\beta/\theta)^{1/\eta}\left(\frac{\beta}{1-\beta}\frac{1-F(\theta)}{\theta f(\theta)}\right)^{-1/\eta}\right)^{-1}. \tag{13}$$

The dashed line in the panels of Figure 2 depicts g_n. The allocations g_d and g_n are two of the three building blocks of the candidate solutions. The third building block is the bunching of public spending.

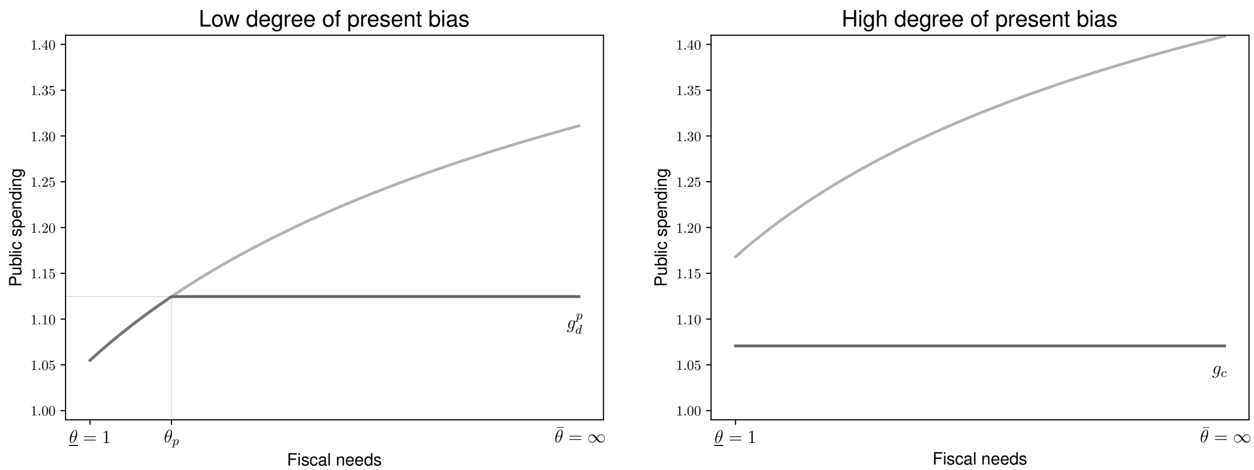

Figure 1: Optimal fiscal rule with non-financial sanctions, i.e., $\rho = 1$.

The grey line depicts the discretionary allocation. The black line depicts the allocation implemented by an optimal fiscal rule. Fiscal needs are exponentially distributed with parameter $\lambda = 3$. The only difference between the parameters for the two panels is the degree of present bias: $1 - \beta = 0.2$ for the left panel and $1 - \beta = 0.6$ for the right panel.

The parameters that are common across the examples are the fiscal revenues T normalized to 1, and the coefficient of relative risk aversion $\eta = 2$. The parameter $\omega = 1$ sets the average deficit as a percentage of fiscal revenues for economies with a low degree of present bias to 12% in the absence of a fiscal rule.[12]

The parameters that differ across the examples are the ones determining the optimal structure of a fiscal rule. They are the welfare weight on the cost of sanctions ρ, the degree of present bias $1 - \beta$, and the distribution of fiscal needs F. They are set to best illustrate the findings

[12] The choice is based on data for the euro area. The average deficit for the three years prior to the implementation of the Stability and Growth Pact was 4.9% of GDP (Source: OECD (2021), General government deficit). Fiscal revenues averaged 40.8% of GDP (Source: Eurostat (gov_10a_taxag)). Combining these two moments gives a target average deficit as a percentage of fiscal revenues of 4.9%/40.8% = 12%.

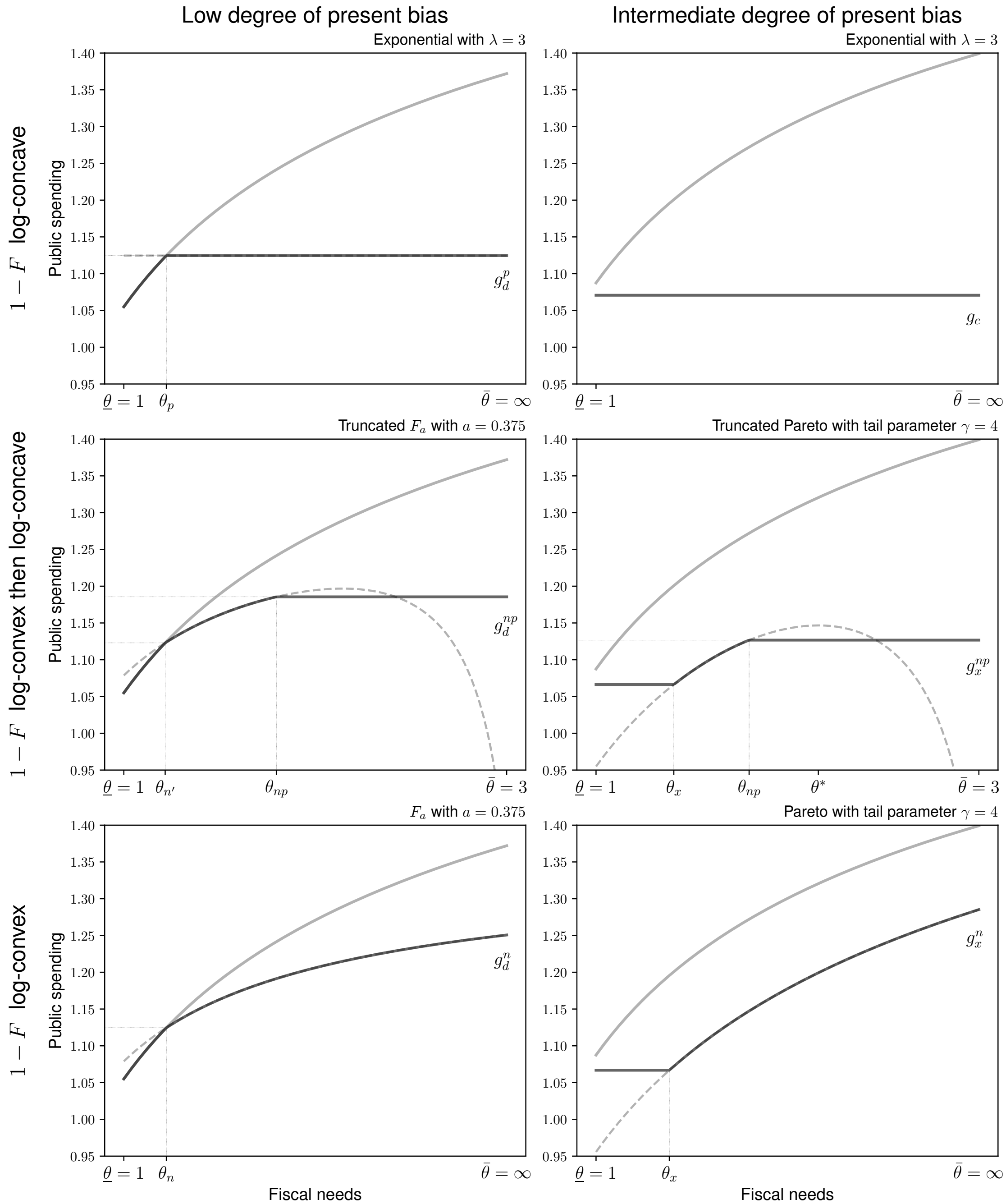

Figure 2: Optimal fiscal rule with financial sanctions, $\rho = \beta$.

Panels on the left depict allocations for an economic union with a low degree of present bias (i.e., $1 - \beta = 0.2$). Panels on the right depict allocations for an economic union with an intermediate degree of present bias (i.e., $1 - \beta = 0.3$). The grey line depicts the discretionary allocation. The dashed line depicts the allocation implemented by the candidate marginal sanction schedule. The black line depicts the allocation implemented by the optimal fiscal rule. The distribution of fiscal needs is displayed at the top right of each panel.

of this paper.[13] The captions of Figures 1 and 2 contain the specifications that differ across the examples.

The contrast between Figures 1 and 2 highlights the importance of the enforcement mechanism for the design of a fiscal rule. The contrast between the panels on the left and on the right in Figures 1 and 2 depicts how the need for discipline, governed by $1-\beta$, matters for the design of an optimal fiscal rule. Figure 1 depicts allocations implemented by optimal fiscal rules enforced by non-financial sanctions (i.e., Propositions 1 and 4 for $\rho = 1$). Figure 2 depicts allocations implemented by optimal fiscal rules enforced by financial sanctions (Propositions 1, 2, and 3 for $\rho = \beta < 1$). For each panel, I verify that the economic union satisfies the conditions of the relevant Proposition in Appendix OC.

Comparing the optimal fiscal rules for different distributions of shocks emphasizes the role of the virtual need for discretion in shaping the sanction schedule. I consider three distributions that differ in the virtual need for discretion that they imply. The constant inverse hazard rate $1/\lambda$ of the exponential distribution implies that the virtual need for discretion is null for $\rho = \beta$ because $0 = \frac{d}{d\theta}(\rho\frac{1}{\lambda} - (\rho - \beta)\theta)$. The linear inverse hazard rate of the Pareto distribution implies a constant virtual need for discretion $\beta - \rho(1 - 1/\gamma)$, where $\gamma > 1$ denotes the tail parameter. The hazard rate of the third distribution, denoted F_a, is a convex combination of the hazard rates of the exponential distribution and the Pareto distribution, as follows:[14] $h_a(\theta) = a\lambda + (1-a)\gamma/\theta$, and $a \in (0, 1)$. The inverse hazard rate of the distribution F_a implies a decreasing virtual need for discretion. The increasing severity of the marginal sanction schedule that implements the discretion and on-equilibrium sanctions allocation in the bottom left panel of Figure 2 reflects the decreasing virtual need for discretion implied by F_a. The parameters for the exponential distribution, the Pareto distribution, and the coefficient a for the distribution F_a are all set so that the three distributions have the same mean. The distributions for the panels in the middle row in Figure 2 are truncated versions of F_a and the Pareto distribution. For ease of comparison, the parameters of the truncated distributions are kept the same as the ones for their non-truncated counterparts.[15]

[13] A quantitative evaluation of the Stability and Growth Pact is left for future research. The model would need to be extended to feature heterogeneity across members of the economic union and an endogenous interest rate.

[14] The hazard rate $h_a(\cdot)$ uniquely characterizes the distribution $F_a(\theta) = 1 - \exp\left(-\int_{\underline{\theta}}^{\theta} h_a(x)dx\right)$ for $\theta \in \Theta$.

[15] Truncating the Pareto and F_a distributions at $\bar{\theta} = 3$ does not significantly alter the mean.

6 Solution method for the design of rules

The no-transfer constraint sets program (5) apart from mechanism design problems in which transfers are possible. Unlike the incentive compatibility constraints, the no-transfer constraint cannot be easily summarized in the objective function to be maximized point-wise without resorting to Lagrangian methods. This section outlines how I use the first-order conditions of the Lagrangian method to identify a candidate solution and to find conditions for global optimality.

The first step uses a Lagrange multiplier function to combine the no-transfer constraint and the objective in a Lagrangian. Let $\Lambda : \Theta \mapsto [0,1]$ be a non-decreasing function such that $\lim_{\theta \to \bar{\theta}} \Lambda(\theta) = 1$ and $1 - \Lambda$ is integrable. A valid Lagrange multiplier function is non-decreasing, which is the infinite dimensional analog of a non-negative Lagrange multiplier for the Kuhn-Tucker theorem with finitely many inequality constraints. Define the Lagrangian, with Lagrange multiplier function Λ, as a functional on $\Phi \equiv \{(u, \underline{t}) \mid u : \Theta \mapsto \mathbb{R}_+$ is non-decreasing, and $\underline{t} \in \mathbb{R}_+\}$ as follows:

$$\mathcal{L}(u, \underline{t}|\Lambda) \equiv \int_{\Theta} \left[\theta u(\theta) + W(T - U^{-1}(u(\theta))) - \rho t(\theta, u, \underline{t})\right] dF(\theta) + \int_{\Theta} t(\theta, u, \underline{t}) d\Lambda(\theta), \tag{14}$$

where $t(\theta, u, \underline{t})$ is the schedule associated with allocation $U^{-1}(u(\cdot))$ in Lemma 3 and $t(\underline{\theta}) = \underline{t}$.

The Gateaux derivative in the direction $(h, h_t) \in \Phi$ is defined as follows:[16]

$$\partial\mathcal{L}(u, \underline{t}, h, h_t|\Lambda) \equiv \lim_{\alpha \downarrow 0} \frac{1}{\alpha} \left[\mathcal{L}(u + \alpha h, \underline{t} + \alpha h_t|\Lambda) - \mathcal{L}(u, \underline{t}|\Lambda)\right]. \tag{15}$$

The next lemma gives optimality conditions in terms of the Gateaux derivative evaluated at the solution. The optimality conditions are that the Gateaux derivative evaluated at the solution is null in the direction of the solution and non-positive in any non-decreasing direction.

Lemma 9 (Lemma of optimality). *If there exists a non-decreasing $u^* \equiv U(g^*)$ and $\underline{t}^*$ in the convex cone Φ and a non-decreasing function $\Lambda^* : \Theta \mapsto [0,1]$ such that $\lim_{\theta \to \bar{\theta}} \Lambda^*(\theta) = 1$ and $1 - \Lambda^*$ is integrable, and if*

$$\partial\mathcal{L}(u^*, \underline{t}^*, u^*, \underline{t}^*|\Lambda^*) = 0, \quad \textit{and} \quad \partial\mathcal{L}(u^*, \underline{t}^*, h, h_t|\Lambda^*) \leq 0 \quad \textit{for all } (h, h_t) \in \Phi,$$

then $g^* \equiv U^{-1}(u^*)$ and the associated money-burning schedule t^* characterized by (8) with $t^*(\underline{\theta}) = \underline{t}^*$ solve the mechanism design problem (5).

[16] The existence of the Gateaux differential follows from Lemma A.1 p. 390 of Amador, Werning, and Angeletos (2006).

The proof in Appendix OA.9 is an application of the global theory of constrained optimization (Chapter 8 in Luenberger (1969), Lemma 1 p.227, and Theorem 1 p.220), as used in Lemma A.2 in Amador, Werning, and Angeletos (2006) and Theorem 1 in Amador and Bagwell (2013). A part of the proof shows that the degree of concavity of the Lagrangian depends on ρ. The Lagrangian is strictly concave for $\rho \in [0, 1)$ and a non-decreasing Lagrange multiplier function (see the proof of Lemma 9). The Lagrangian is linear for $\rho = 1$.

The solution method appears to ask the designer to guess the solution and verify that it satisfies the optimality conditions in Lemma 9. Guessing the solution amounts to the arrangement of three building blocks. The first building block obtains from ignoring the monotonicity and the no-transfer constraint to determine the spending g_n for $\rho < 1$. The second building block is the discretionary allocation. It is a natural candidate because of the no-transfer constraint. Third, the allocation may be constant over subintervals of Θ.

I use the optimality conditions of Lemma 9 to determine the arrangement of the three building blocks. The first optimality condition sets the Gateaux derivative of the Lagrangian to zero. It is the first requirement in the definition of the thresholds between the different building blocks $\theta_p, \theta_{np}, \theta_x$. It also determines the Lagrange multiplier function. In turn, Assumption L is precisely the condition needed for the Lagrange multiplier function to be non-decreasing. The second optimality condition verifies that the Gateaux derivative of the Lagrangian in any non-decreasing direction is negative. It is the second requirement in the definition of the thresholds $\theta_p, \theta_{np}, \theta_x$.

7 Concluding remarks

To conclude, I use the findings from this paper to evaluate the Excessive Deficit Procedure (EDP) of the Stability and Growth Pact and propose avenues for reforms. For some context, the following quote reflects the current financial sanction schedule of the EDP:

> A non-interest-bearing deposit of 0.2% of GDP may be requested from a euro area country that is placed in EDP. [...] In case of non-compliance with the initial recommendation for corrective action, this non-interest-bearing deposit will be converted into a fine. *EU Economic governance "Six Pack" - State of Play, Memo/11/647.*

First, there is a threshold below which discretion prevails and above which the country is placed in EDP. This paper lends support to this feature of the Stability and Growth Pact.

Second, the EDP features a jump in the level of financial sanctions—a notch point—from 0 to a non-interest-bearing deposit of 0.2% of GDP. Initially, the sanction is the forgone interest on the deposit. This notch could be optimal if it is an off-equilibrium sanction. Off-equilibrium sanctions are part of the optimal design of a fiscal rule if the distribution of shocks has a sufficiently thin tail. Otherwise, the optimal design features a kink point at the exemption threshold. Expressing the sanction as a percentage of the deficit above the threshold, instead of a percentage of GDP, would turn the notch into a kink.

Third, the EDP could convert the deposit into a fine. The sanction would then be the deposit instead of the forgone interest on this deposit. This sharp increase in the sanction is not the relevant difference in the choice of instrument to impose discipline. Indeed, the designer can set a fine that matches the forgone interest on a deposit. The relevant difference between a non-interest-bearing deposit and a fine is the illiquid nature of the fine. If the economic union is better able to recycle illiquid sanctions, by say investing the revenues in union-specific public goods, then a fine is a better instrument (i.e., lower ρ) to enforce a rule with on-equilibrium sanctions. Appendix A contains a discussion of the choice of instrument.

The quantitative evaluation of the potential for a reform of the Stability and Growth Pact has been left for future research. Two considerations stand out. First, the E.U. is a union of heterogeneous countries. This paper shows that the cross-country heterogeneity in the degree of present bias and in the thickness of the tail of the distribution of shocks determines the heterogeneity in the competing needs for discipline and discretion across members of the economic union. Second, a union of countries may be large enough for its fiscal rule to affect the interest rate. Halac and Yared (2018) solve for the optimal stringency of a cap on spending in an environment with an endogenous interest rate.

The global Lagrangian method used in this paper is widely applicable to solve mechanism design problems with limited transfers. Werning (2007) study Pareto efficient income taxation. The requirement that a reform be Pareto-improving limits transfers. The method also applies to design the cost of verifying the state to determine whether escape clauses apply (see Halac and Yared (2020b) for the design of a rule with costly state verification). Another promising application is the study of the optimal illiquidity of retirement savings accounts for households who under-save for their retirement (see Laibson et al. (1998), and Beshears et al. (2020)).

References

Aguiar, M., Amador, M., Farhi, E. and Gopinath, G. (2015). Coordination and Crisis in Monetary Unions, Quarterly Journal of Economics, 130(4): 1727-1780

Aguiar, M., Amador, M., and Fourakis, S. (2020). On the Welfare Losses from External Sovereign Borrowing. IMF Economic Review, 68(1): 163-194.

Alfaro, L. and Kanczuk, F. (2019). Fiscal Rules and Sovereign Default. Working paper.

Alonso, R., and Matouschek, N. (2008). Optimal Delegation. Review of Economic Studies, 75: 259-293

Amador, M., and Bagwell, K. (2021). Regulating a Monopolist With Uncertain Costs Without Transfers. Working paper.

Amador, M., and Bagwell, K. (2020). Money Burning in the Theory of Delegation. Games and Economic Behavior, 112: 382-412

Amador, M., Bagwell, K., and Frankel, A. (2018). A Note on Interval Delegation. Economic Theory Bulletin, 6(2): 239-249

Amador, M. and Bagwell, K. (2013). The Theory of Optimal Delegation With an Application to Tariff Caps. Econometrica, 81: 1541-1599

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

Amador, M., Werning, I. and Angeletos, G.-M. (2004). Redistribution and Corrective Taxation. Working Paper

Ambrus, A., and Egorov, G. (2017). Delegation and Nonmonetary Incentives, 171: 101-135

Ambrus, A., and Egorov, G. (2013). Comment on “Commitment vs. Flexibility”, Econometrica, 81(5): 2113-2124

Athey, S., Atkeson, A., and Kehoe, P. (2005). The Optimal Degree of Discretion in Monetary Policy, Econometrica, 73(5): 1431-1475

Atkeson, A. and Lucas, R. E. (1992). On Efficient Distribution with Private Information. Review of Economic Studies, 59(3): 427-453

Azzimonti, M., Battaglini, M., and Coate, S. (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

Bagnoli, M. and Bergstrom, T. (2005). Log-Concave Probability and Its Applications. Economic Theory, 26: 445-469

Beetsma, R. and Uhlig, H. (1999). An Analysis of the Stability and Growth Pact, 109: 546-571

Beshears, J., Choi, J. J., Clayton, C., Harris, C., Laibson, D., and Madrian, B. C. (2020). Optimal Illiquidity. Working paper.

Brunnermeier, M. K., James, H., and Landau, J.-P. (2016) *The Euro and the Battle of Ideas*, Princeton University Press

Chari, V. V., and Kehoe, P. (2007). On the need for fiscal constraints in a monetary union. Journal of Monetary Economics, 54: 2399-2408

Chari, V. V., and Kehoe, P. (2008). Time Inconsistency and Free-Riding in a Monetary Union. Journal of Money, Credit and Banking, 40: 1329-1355

Cooper, R. and Kempf, H. (2004). Overturning Mundell: Fiscal Policy in a Monetary Union. Review of Economic Studies, 71: 371-396

Diamond, P. (1998). Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates. American Economic Review, 88: 83-95

Dovis, A. and Kirpalani, R. (2021). Rules without Commitment: Reputation and Incentives. *forthcoming* Review of Economic Studies

Dovis, A., and Kirpalani, R. (2020). Fiscal Rules, Bailouts, and Reputation in Federal Governments. American Economic Review *forthcoming*

European Commission (2019). Vade Mecum on the Stability and Growth Pact. European Economy Institutional Paper 101.

Felli, C., Piguillem, F., and Shi, L. (2021). Fiscal Rules and Discretion with Risk of Default. Working paper.

Galperti, S. (2015). Commitment, Flexibility, and Optimal Screening of Time Inconsistency. Econometrica, 83: 1425-1465

Halac, M., and Yared, P. (2020a). Inflation Targeting under Political Pressure. in Independence, Credibility, and Communication of Central Banking, edited by Ernesto Pastén and Ricardo Reis, Santiago, Chile: Central Bank of Chile, 7-27

Halac, M., and Yared, P. (2020b). Commitment vs. Flexibility with Costly Verification. Journal of Political Economy, 128: 4523-4573

Halac, M., and Yared, P. (2022a). Fiscal Rules and Discretion under Limited Enforcement. Econometrica, *forthcoming*

Halac, M., and Yared, P. (2022b). Instrument vs. Target-Based Rules. Review of Economic Studies, 89: 312-345

Halac, M., and Yared, P. (2022c). A Theory of Fiscal Responsibility and Irresponsibility. *working paper*

Halac, M., and Yared, P. (2018). Fiscal Rules and Discretion in a World Economy. American Economic Review, 108: 2305-2334

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

Hatchondo, J.-C., Martinez, L., and Roch, F. (2020). Fiscal Rules and the Sovereign Default Premium, *forthcoming* American Economic Journal: Macroeconomics

Holmström, B. (1977). On Incentives and Control in Organizations. Ph.D. thesis, Stanford Graduate School of Business

Kocherlakota, N. (2016). Rules versus Discretion: A Reconsideration. Brookings Papers on Economic Activity, 1-36

Kováč, E. and Mylovanov, T. (2009). Stochastic mechanisms in settings without monetary transfers: The regular case. Journal of Economic Theory, 144: 1373-1395

Laibson, D., I. Repetto, A., Tobacman, J., Hall, R., Gale, W. and Akerlof, G., A. (1998). Self-Control and Saving for Retirement. Brookings Papers on Economic Activity 91-196

Lledó, V., Yoon, S., Fang, X., Mbaye, S., and Kim, Y. (2017). Fiscal Rules at a Glance. IMF background paper.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods.* Series in Decision and Control. New York: Wiley

Martimort, D. and Semenov, A. (2006). Continuity in Mechanism Design Without Transfers. Economic Letters, 93: 182-189

Melumad, N. D. and Shibano, T. (1991). Communication in Settings with No Transfers. The Rand Journal of Economics, 22: 173-198

Myerson, R. B. (1981). Optimal Auction Design. Mathematics of Operations Research, 6: 58-73

Werning, I. (2007). Pareto Efficient Income Taxation. working paper.

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

Appendix

A Determinants of the welfare cost of sanctions

This section shows that the way in which the economic union recycles the revenues from sanctions matters for the welfare cost of sanctions. It contains two models of the budget of the economic union. The two models differ in the way in which the economic union uses the revenues from sanctions. In the first model, inspired by the Excessive Deficit Procedure of the Stability and Growth Pact, the revenues from sanctions expand the budget of the economic union. In the second model, the revenues from sanctions reduce the contributions of each members to the budget of the union.

For this section, I assume a money-metric continuation value. As a result, the social welfare function depends only on the first moment of the distribution of financial sanctions meted out. This is because if $W(x) = x$, then $\tau(g) = \tau_f(g)$.

A.1 Home-biased preferences for public spending

Suppose the revenues from sanctions are used to expand the budget of the economic union. The budget of the economic union finances union-specific public goods. For instance, in the context of the Stability and Growth Pact, the fine contributes to a union-specific public good, the European Stability Mechanism (Section 2.4.2 of the Vade Mecum on the Stability and Growth Pact, European Commission 2019).

The citizenry puts a weight of $(1-\rho) \in (0,1)$ on a contribution to the budget of the economic union relative to a contribution to the country's future assets. Hence ρ measures the degree of *home bias* in the citizenry's preferences for public good spending

$$\int_\Theta [\theta U(g(\theta)) + x(\theta) + (1-\rho)b]\, dF(\theta). \tag{16}$$

The revenues from sanctions contribute b to the budget of the economic union:

$$b = \int_\Theta \tau_f(g(\theta)) dF(\theta). \tag{17}$$

Substituting the budget constraint of each country and the budget constraint of the economic union (17) in (16) gives the same social welfare function as the one used in the main text,

$$\int_\Theta [\theta U(g(\theta)) + W(T - g(\theta)) - \rho\tau(g(\theta))]\, dF(\theta).$$

Ultimately, the economic union chooses how to use the revenues from sanctions. The next subsection considers an economic union that uses the revenues from sanctions to reduce the contribution of its members instead of expanding its budget.

A.2 The budget of the economic union as a "leaky bucket"

Suppose the economic union needs to finance an exogenous expenditure G. To do so, each member contributes c to finance the expenditures not covered by the revenues from sanctions. In the spirit of the "leaky bucket" model of public finance, a fraction $1-\rho \in (0,1)$ of the revenues from sanctions accrue to the budget and a fraction ρ "leaks" (see Okun (1975)).[17] The budget constraint of the economic union is

$$c = G - (1-\rho)\int_{\Theta} \tau_f(g(\theta))dF(\theta). \tag{18}$$

The reduction in the contribution amounts to a uniform transfer to the governments. This shows that the level of the limit in the no-transfer constraint is inessential.

The budget constraint of each government depends on the revenues from taxation net of the contribution to the union, that is,

$$g + x + \tau_f(g) = T - c. \tag{19}$$

Substituting (19) in the government's objective gives: $\theta U(g) + \beta(T - g - c) - \beta\tau_f(g)$, where each government takes c as given and independent of the sanction it pays since it has a negligible impact on the budget of the economic union.

The welfare of the economic union is $\int_{\Theta} [\theta U(g(\theta)) + W(x(\theta))]\, dF(\theta)$. Substituting the budget constraints (18) and (19) in the welfare of the economic union gives the same social welfare function, up to an irrelevant constant, as the one used in the main text:

$$\int_{\Theta} [\theta U(g(\theta)) + W(T - g(\theta)) - \rho\tau(g(\theta))]\, dF(\theta) - G.$$

B Proofs of main results

The Lagrangian (14), after rescaling the objective by β and the Lagrange multiplier by $\rho\beta$, reads

$$\begin{aligned}
\mathcal{L}(u,\underline{t}|\Lambda) = &\int_{\Theta}\left[-\beta\nu(U^{-1}(u(\theta))) + \rho\tfrac{1-F(\theta)}{f(\theta)}u(\theta)\right] dF(\theta)\\
&+ (\beta-\rho)\int_{\Theta}\left[\theta u(\theta) + \beta W(T - U^{-1}(u(\theta)))\right] dF(\theta)\\
&+ \rho\left(\underline{\theta}u(\underline{\theta}) + \beta W(T - U^{-1}(u(\underline{\theta}))) - \underline{t}\right)\Lambda(\underline{\theta})\\
&+ \rho\int_{\Theta}\left[\theta u(\theta) + \beta W(T - U^{-1}(u(\theta)))\right] \mathrm{d}\Lambda(\theta) - \rho\int_{\Theta}[u(\theta)(1-\Lambda(\theta))]\, d\theta.
\end{aligned}$$

[17] To put a lower bound on the extent of the leak, suppose that the only source of leakage is the administrative cost of the fiscal rule. Based the budget of the European Union, if the administrative cost of the fiscal rule is commensurate with the administrative cost of the economic union, then the leak amounts to 6% of the revenues.

The Gateaux derivative in the direction (h, h_t), defined in (15), reads as follows:[18]

$$\partial\mathcal{L}(u,\underline{t},h,h_t|\Lambda) = \int_\Theta \Big[\Big(-\beta \tfrac{\nu'(U^{-1}(u(\theta)))}{U'(U^{-1}(u(\theta)))} + \rho \tfrac{1-F(\theta)}{f(\theta)}\Big) h(\theta)\Big]\, dF(\theta) \tag{20}$$
$$+ (\beta-\rho)\int_\Theta \big[\theta\Delta(U^{-1}(u(\theta)),\theta)h(\theta)\big]\, dF(\theta) + \rho\left(\underline{\theta}\Delta(U^{-1}(u(\underline{\theta})),\underline{\theta})h(\underline{\theta}) - h_t\right)\Lambda(\underline{\theta})$$
$$+ \rho\int_\Theta \big[\theta\Delta(U^{-1}(u(\theta)),\theta)h(\theta)\big]\, \mathrm{d}\Lambda(\theta) - \rho\int_\Theta [h(\theta)(1-\Lambda(\theta))]\, d\theta.$$

B.1 Proof of Proposition 1

Proof of Proposition 1. The proof consists of applying Lemma 9. Let $u^*(\theta) = U(g_d^p(\theta))$ for $\theta \in \Theta$ and $\underline{t}^* = 0$. Since g_d^p is non-decreasing, $(u^*, \underline{t}^*) \in \Phi$. The Lagrange multiplier function is

$$\rho\Lambda^*(\theta) = \begin{cases} \rho & \text{for } \theta \geq \theta_p \\ \rho F(\theta) + (1-\beta)\theta f(\theta) & \text{for } \theta \in (\underline{\theta}, \theta_p) \\ 0 & \text{for } \theta = \underline{\theta}. \end{cases}$$

A valid Lagrange multiplier function is non-decreasing. The lower bound on the elasticity of the density in Assumption L holding for $\theta \leq \theta_p$ is equivalent to the Lagrange multiplier Λ^* being non-decreasing on $(\underline{\theta}, \theta_p)$. The jumps at $\underline{\theta}$ and θ_p must be non-decreasing. The jump at $\underline{\theta}$ is non-negative since f is non-negative, $0 < \beta \leq 1$, and $\rho > 0$. The jump at θ_p is non-negative, as shown in Lemma 2. Lemma 2 shows that Assumption L holding for $\theta \leq \theta_p$ and $\beta E[\theta|\theta \geq \hat{\theta}] - \rho\hat{\theta} \leq \beta E[\theta|\theta \geq \theta_p] - \rho\theta_p$ holding for $\hat{\theta} \geq \theta_p$ implies $\rho\frac{1-F(\theta)}{\theta f(\theta)} \geq 1 - \beta$ for $\theta \leq \theta_p$, which is a non-negative jump of Λ^* at θ_p. Note also that $1 - \Lambda^*$ is integrable because $1 - F$ is integrable and the expectation exists.

The rest of the proof checks that the conditions in terms of Gateaux derivatives in Lemma 9 are satisfied. The Gateaux derivative (15) in the direction of h, with Lagrange multiplier function Λ^* reads

$$\partial\mathcal{L}(u,\underline{t},h,h_t|\Lambda^*) = \int_\Theta \Big[\Big(-\beta \tfrac{\nu'(U^{-1}(u(\theta)))}{U'(U^{-1}(u(\theta)))} + \rho \tfrac{1-F(\theta)}{f(\theta)}\Big) h(\theta)\Big]\, dF(\theta)$$
$$+ (\beta-\rho)\int_\Theta \big[\theta\Delta(U^{-1}(u(\theta)),\theta)h(\theta)\big]\, dF(\theta)$$
$$+ \int_{\underline{\theta}}^{\theta_p} \big[\theta\Delta(U^{-1}(u(\theta)),\theta)h(\theta)\big]\, \rho d\Lambda^*(\theta) - \rho\int_{\underline{\theta}}^{\theta_p} [(1-\Lambda^*(\theta))\; h(\theta)]\, d\theta.$$

[18]The existence of the Gateaux differential follows from Lemma A.1 p. 390 of Amador, Werning, and Angeletos (2006).

Rewriting the Euler equations characterizing g_d gives

$$\beta\frac{\nu'(U^{-1}(u^*(\theta)))}{U'(U^{-1}(u^*(\theta)))} = \begin{cases} (1-\beta)\theta_p & \text{for } \theta > \theta_p \\ (1-\beta)\theta & \text{for } \theta \leq \theta_p \end{cases} \quad \text{and,} \quad \theta\Delta(U^{-1}(u^*(\theta)),\theta) = \begin{cases} \theta-\theta_p & \text{for } \theta > \theta_p \\ 0 & \text{for } \theta \leq \theta_p. \end{cases}$$

After substitution of these two expressions, the Gateaux derivative evaluated at u^* simplifies to

$$\partial\mathcal{L}(u^*,\underline{t}^*,h,h_t|\Lambda^*) = \int_{\underline{\theta}}^{\theta_p}\left[\left(-(1-\beta)\theta + \rho\tfrac{1-F(\theta)}{f(\theta)} - \rho\tfrac{1-\Lambda^*(\theta)}{f(\theta)}\right)f(\theta)h(\theta)\right]d\theta \tag{21}$$

$$+\int_{\theta_p}^{\bar\theta}\left[\left(-(1-\beta)\theta_p + \rho\tfrac{1-F(\theta)}{f(\theta)} + (\beta-\rho)\left(\theta-\theta_p\right)\right)f(\theta)h(\theta)\right]d\theta. \tag{22}$$

The Lagrange multiplier Λ^* over $(\underline{\theta},\theta_p]$ is defined so that the integral (21) is null, for $(h,h_t)\in\Phi$. Suppose that $\bar\theta<\infty$ so that for any $(h,h_t)\in\Phi$, h is bounded (the case $\bar\theta=\infty$ is addressed below). For h bounded, the following term is null:

$$\lim_{\hat\theta\to\bar\theta}\int_{\hat\theta}^{\bar\theta}\left[\left(-(1-\beta)\theta_p + \rho\tfrac{1-F(\theta)}{f(\theta)} + (\beta-\rho)(\theta-\theta_p)\right)f(\theta)\right]d\theta\ h(\hat\theta) = 0,$$

hence, integrating (22) by parts gives

$$\int_{\theta_p}^{\bar\theta}\left[\left(-(1-\beta)\theta_p + \rho\tfrac{1-F(\theta)}{f(\theta)} + (\beta-\rho)(\theta-\theta_p)\right)f(\theta)\right]d\theta\ h(\theta_p) \tag{22.1}$$

$$+\int_{\theta_p}^{\bar\theta}\left[\int_{\hat\theta}^{\bar\theta}\left[\left(-(1-\beta)\theta_p + \rho\tfrac{1-F(\theta)}{f(\theta)} + (\beta-\rho)(\theta-\theta_p)\right)f(\theta)\right]d\theta\right]dh(\hat\theta). \tag{22.2}$$

As the next claim shows, θ_p is defined so that the inner integral in (22.2) is negative for $\hat\theta\geq\theta_p$ and null for $\hat\theta=\theta_p$ (which also implies that (22.1) is null).

Claim 1. *Inequality* (7) *is equivalent to*

$$\int_{\hat\theta}^{\bar\theta}\left[\left(-(1-\beta)\tilde\theta + \rho\tfrac{1-F(\theta)}{f(\theta)} + (\beta-\rho)(\theta-\tilde\theta)\right)f(\theta)\right]d\theta \leq 0. \tag{23}$$

Proof of Claim 1. After substitution of the wedge evaluated at the discretionary allocation, i.e., $\hat\theta\Delta(g_d(\tilde\theta),\hat\theta)=\hat\theta-\tilde\theta$, inequality (7) reads as follows:

$$\int_{\hat\theta}^{\bar\theta}\left[\theta-\tilde\theta\right]dF(\theta) \leq \frac{\nu'(g_d(\tilde\theta))}{U'(g_d(\tilde\theta))}(1-F(\hat\theta)) + \rho\frac{\hat\theta-\tilde\theta}{\beta}(1-F(\hat\theta)).$$

After multiplying both sides by β and grouping terms, the inequality reads

$$\int_{\hat\theta}^{\bar\theta}\left[\rho(\tilde\theta-\hat\theta)+\beta(\theta-\tilde\theta)\right]dF(\theta) \leq \beta\frac{\nu'(g_d(\tilde\theta))}{U'(g_d(\tilde\theta))}(1-F(\hat\theta)).$$

Substituting the definition of the bias $\nu'(g_d(\tilde{\theta})) = (1-\beta)W'(T - g_d(\tilde{\theta}))$ on the right-hand side and using the Euler equation defining the discretionary allocation gives

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\tilde{\theta} - \hat{\theta}) + \beta(\theta - \tilde{\theta})\right] dF(\theta) \leq (1-\beta)\tilde{\theta}(1 - F(\hat{\theta})). \tag{24}$$

Adding and subtracting $\rho(\theta - \tilde{\theta})$ to the integrand on the left-hand side gives

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\theta - \hat{\theta}) + (\beta - \rho)(\theta - \tilde{\theta})\right] dF(\theta) \leq (1-\beta)\tilde{\theta}(1 - F(\hat{\theta})).$$

Integrating the left-hand side by parts gives inequality (23). □

Since $\theta_p > \underline{\theta}$, inequality (7) holds with equality for $\hat{\theta} = \tilde{\theta} = \theta_p$, hence (23) holds with equality for $\hat{\theta} = \tilde{\theta} = \theta$. By definition of $\theta_p > \underline{\theta}$, (22.1) is null and the inner integral of (22.2) is negative for $\hat{\theta} \geq \theta_p$.

Consider $(h, h_t) = (u^*, \underline{t}^*)$. The Gateaux derivative $\partial\mathcal{L}(u^*, \underline{t}^*, u^*, \underline{t}^*|\Lambda^*)$ is null since (22.1) and (22.2) are both null. Line (22.2) is null for $h = u^*$ because $dh(\theta) = du^*(\theta) = 0$ for $\theta \geq \theta_p$.

Consider any $(h, h_t) \in \Phi$. The Gateaux derivative $\partial\mathcal{L}(u^*, \underline{t}^*, h, h_t|\Lambda^*)$ is negative since (22.1) is null and (22.2) is negative. Line (22.2) is negative because $dh \geq 0$ since $(h, h_t) \in \Phi$ and (23) is negative.

For $\bar{\theta} = \infty$, consider a sequence of environments, indexed by $m \in \mathbb{N}$, with $\Theta = [\underline{\theta}, \bar{\theta}_m]$, $\bar{\theta}_m < \infty$, and $\lim_{m\to\infty} \bar{\theta}_m = \infty$. Denote by F_m the truncation of F on $[\underline{\theta}, \bar{\theta}_m]$, defined as follows: $F_m(\theta) = \frac{F_\theta}{F(\theta_m)}$ for $\theta \in [\underline{\theta}, \bar{\theta}_m]$ and $F_m(\theta) = 1$ for $\theta \geq \bar{\theta}_m$. Note that F_m converges weakly to F. Also, since F is twice continuously differentiable, f_m is continuous and it converges point-wise to f. For each $m \in \mathbb{N}$, denote the solution of the environment with the truncated distribution F_m by $g_d^p(\cdot; m)$ and the threshold at which the cap binds by $\theta_p^{(m)}$.

First, note that $g_d^p(\cdot; m)$ converges point-wise to $g_d^p(\cdot)$ if $\lim_{m\to\infty} \theta_p^{(m)} = \theta_p$. By assumption $\theta_p > \underline{\theta}$. For $\theta_p < \bar{\theta}$, it is the lowest fiscal need that solves $\beta\, \mathbb{E}[\theta|\theta \geq \theta_p] = \theta_p$, and satisfies (see Lemma 1), $\beta E[\theta|\theta \geq \hat{\theta}] - \rho\hat{\theta} \leq \beta E[\theta|\theta \geq \theta_p] - \rho\theta_p$ for $\hat{\theta} \geq \theta_p$. The threshold $\theta_p^{(m)}$ is characterized analogously. For $\theta_p^{(m)} < \bar{\theta}$, it is the lowest fiscal need that solves $\beta\, \mathbb{E}_m[\theta|\theta \geq \theta_p^{(m)}] = \theta_p^{(m)}$, and, by Lemma 1, $\beta E_m[\theta|\theta \geq \hat{\theta}] - \rho\hat{\theta} \leq \beta E_m[\theta|\theta \geq \theta_p^{(m)}] - \rho\theta_p^{(m)}$ for $\hat{\theta} \geq \theta_p^{(m)}$. For any $\hat{\theta} > \underline{\theta}$, because F_m converges weakly to F, $\lim_{m\to\infty} \mathbb{E}_m\left[\frac{\theta}{\hat{\theta}}|\theta \geq \hat{\theta}\right] = \mathbb{E}\left[\frac{\theta}{\hat{\theta}}|\theta \geq \hat{\theta}\right]$. Because taking the limit preserves weak inequalities, $\lim_{m\to\infty} \theta_p^{(m)} \geq \theta_p$. Since F_m is a right-truncation of F, $\mathbb{E}_m\left[\frac{\theta}{\hat{\theta}}|\theta \geq \hat{\theta}\right] \leq \mathbb{E}\left[\frac{\theta}{\hat{\theta}}|\theta \geq \hat{\theta}\right]$. Hence $\theta_p^{(m)} \leq \theta_p$ for every m. It follows that $\lim_{m\to\infty} \theta_p^{(m)} \leq \theta_p$. Combining the two inequalities, $\lim_{m\to\infty} \theta_p^{(m)} = \theta_p$, and $g_d^p(\cdot, m)$ converges pointwise to $g_d^p(\cdot)$.

Second, note that since $(g_d^p(\cdot, m))_{m\in\mathbb{N}}$ is a uniformly bounded sequence and $(f_m)_{m\in\mathbb{N}}$ is bounded by an integrable density, the dominated convergence theorem implies that the sequence of social welfare (with the incentive compatible $t(\cdot; m)$ substituted in) resulting from the sequence of truncated economies converges to the social welfare of the non-truncated economy. Hence a fiscal rule with a null intercept that implements $g_d^p(\cdot)$ is optimal for the non-truncated economy. □

B.2 Proof of Proposition 2

Proof of Proposition 2. The proof consists of applying Lemma 9. A valid allocation is non-decreasing. The allocation g_d^{np} is continuous since $g_d(\theta_n) = g_n(\theta_n)$ for $\theta_n \geq \underline{\theta}$. The discretionary allocation g_d is strictly increasing. The public spending $g_n(\theta)$ is defined for $\theta \in (\theta_n, \theta_{np})$ because Assumption I holds for $\theta \in (\theta_n, \theta_{np})$. Lemma 4 implies that g_n is non-decreasing over (θ_n, θ_{np}) because the derivative of $\rho\frac{1-F(\theta)}{f(\theta)}$ is not smaller than $\rho - \beta$. The utility index U is strictly increasing so the utility profile from the *discretion, on-equilibrium, and off-equilibrium sanctions* allocation $u^*(\theta) = U(g_d^{np}(\theta))$ for $\theta \in \Theta$ and $\underline{t}^* = 0$ satisfies $(u^*, \underline{t}^*) \in \Phi$.

The Lagrange multiplier function is

$$\rho\Lambda^*(\theta) = \begin{cases} \rho & \text{for } \theta \geq \theta_n \\ \rho F(\theta) + (1-\beta)\theta f(\theta) & \text{for } \theta \in (\underline{\theta}, \theta_n) \\ 0 & \text{for } \theta = \underline{\theta}. \end{cases}$$

A valid Lagrange multiplier function is non-decreasing. The lower bound on the elasticity of the density in Assumption L holding for $\theta \leq \theta_n$ is equivalent to the Lagrange multiplier Λ^* being non-decreasing on $(\underline{\theta}, \theta_n)$. The jumps at $\underline{\theta}$ and θ_n must be non-negative. The jump at $\underline{\theta}$ is non-negative since f is non-negative, $0 < \beta \leq 1$, and $\rho > 0$. By Lemma 5, the jump at θ_n is non-negative since either $\theta_n < \bar{\theta}$ in which case $g_n(\theta_n) = g_d(\theta_n)$ and the Lagrange multiplier is continuous at θ_n, or $\theta_n = \bar{\theta}$ in which case $g_n(\theta_n) \leq g_d(\theta_n)$ and the jump is non-negative. Note also that $1 - \Lambda^*$ is integrable because $1 - F$ is integrable and the expectation exists.

The Gateaux derivative (20), with a Lagrange multiplier function equal to 1 for $\theta \geq \theta_n$, reads

$$\partial\mathcal{L}(u,\underline{t},h,h_t|\Lambda^*) = \int_{\underline{\theta}}^{\theta_n} \rho\left[\theta\Delta(U^{-1}(u(\theta)),\theta)h(\theta)\right]d\Lambda^*(\theta) \tag{25}$$

$$+\int_{\underline{\theta}}^{\theta_n}\left[\left(-\beta\tfrac{\nu'(U^{-1}(u(\theta)))}{U'(U^{-1}(u(\theta)))}+\rho\tfrac{1-F(\theta)}{f(\theta)}+(\beta-\rho)\theta\Delta(U^{-1}(u(\theta)),\theta)-\rho(1-\Lambda^*(\theta))\right)h(\theta)\right]dF(\theta) \tag{26}$$

$$+\int_{\theta_n}^{\theta_{np}}\left[\left(-\beta\tfrac{\nu'(U^{-1}(u(\theta)))}{U'(U^{-1}(u(\theta)))}+\rho\tfrac{1-F(\theta)}{f(\theta)}+(\beta-\rho)\theta\Delta(U^{-1}(u(\theta)),\theta)\right)h(\theta)\right]\ dF(\theta) \tag{27}$$

$$+\int_{\theta_{np}}^{\bar{\theta}}\left[\left(-\beta\tfrac{\nu'(U^{-1}(u(\theta)))}{U'(U^{-1}(u(\theta)))}+\rho\tfrac{1-F(\theta)}{f(\theta)}+(\beta-\rho)\theta\Delta(U^{-1}(u(\theta)),\theta)\right)h(\theta)\right]\ dF(\theta). \tag{28}$$

The last step shows that the conditions in terms of Gateaux derivatives in Lemma 9 are met. The term (25) evaluated at u^* is null irrespectively of the direction of the Gateaux derivative (h,h_t) because $g_d^{np}(\theta)=g_d(\theta)$ for $\theta\leq\theta_n$ implies $\Delta(U^{-1}(u^*(\theta)),\theta)=0$ for $\theta\leq\theta_n$. The choice of Lagrange multiplier over $(\underline{\theta},\theta_n)$ is precisely the condition needed for the term (26) to be null irrespectively of the direction (h,h_t) of the Gateaux derivative. The definition of g_n in (9) implies that the term (27) evaluated at u^* is null irrespectively of the direction (h,h_t) of the Gateaux derivative.

Using the definition of the bias to get $\theta\Delta(U^{-1}(u(\theta)),\theta)=\theta-\frac{\beta}{1-\beta}\frac{\nu'(U^{-1}(u(\theta)))}{U'(U^{-1}(u(\theta)))}$, and the following characterization of u^* above θ_{np}: $\beta\frac{1-\rho}{1-\beta}\ \frac{\nu'(U^{-1}(u^*(\theta)))}{U'(U^{-1}(u^*(\theta)))}=\rho\frac{1-F(\theta_{np})}{f(\theta_{np})}+(\beta-\rho)\theta_{np}$, for $\theta\geq\theta_{np}$, line (28) reads

$$\int_{\theta_{np}}^{\bar{\theta}}\left[\left(\rho\tfrac{1-F(\theta)}{f(\theta)}-\rho\tfrac{1-F(\theta_{np})}{f(\theta_{np})}+(\beta-\rho)(\theta-\theta_{np})\right)h(\theta)\right]\ dF(\theta). \tag{29}$$

Integrating (29) by parts gives

$$\int_{\theta_{np}}^{\bar{\theta}}\left[\rho\tfrac{1-F(\theta)}{f(\theta)}-\rho\tfrac{1-F(\theta_{np})}{f(\theta_{np})}+(\beta-\rho)\left(\theta-\theta_{np}\right)\right]\ dF(\theta)\ h(\theta_{np}) \tag{29.1}$$

$$+\int_{\theta_{np}}^{\bar{\theta}}\left[\int_{\hat{\theta}}^{\bar{\theta}}\left[\rho\tfrac{1-F(\theta)}{f(\theta)}-\rho\tfrac{1-F(\theta_{np})}{f(\theta_{np})}+(\beta-\rho)\left(\theta-\theta_{np}\right)\right]\ dF(\theta)\right]dh(\hat{\theta}), \tag{29.2}$$

where I used that $\lim_{\hat{\theta}\to\bar{\theta}}\int_{\hat{\theta}}^{\bar{\theta}}\left[\rho\frac{1-F(\theta)}{f(\theta)}-\rho\frac{1-F(\theta_{np})}{f(\theta_{np})}+(\beta-\rho)\left(\theta-\theta_{np}\right)\right]dF(\theta)h(\hat{\theta})$ is zero since h is bounded if $\bar{\theta}<\infty$. If $\bar{\theta}=\infty$, the result follows from taking the limit of a sequence of environments with compact support as in the proof of Proposition 1 in Appendix B.1.

The definition of θ_{np} is precisely so that the inner integral in (29.2) is negative for $\hat{\theta}\geq\theta_{np}$ and null for $\hat{\theta}=\theta_{np}$ (which also implies that (29.1) is null).

Claim 2. *Inequality* (10) *is equivalent to*

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho\frac{1-F(\theta)}{f(\theta)} + (\beta-\rho)(\theta-\tilde{\theta})\right] dF(\theta) \le \rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})}(1-F(\hat{\theta})). \tag{30}$$

Proof of Claim 2. The definition of the wedge implies the following identity:

$$\hat{\theta}\Delta(g,\hat{\theta}) = \hat{\theta} - \frac{\beta W'(T-g)}{U'(g)} = \hat{\theta} - \tilde{\theta} + \tilde{\theta}\Delta(g,\tilde{\theta}).$$

Multiplying both sides of inequality (10) by β and substituting the identity for the wedge gives

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\beta(\theta-\tilde{\theta}) + \beta\tilde{\theta}\Delta(g_n(\tilde{\theta}),\tilde{\theta})\right]\, dF(\theta) \le \left(\beta\frac{\nu'(g_n(\tilde{\theta}))}{U'(g_n(\tilde{\theta}))} + \rho(\hat{\theta}-\tilde{\theta}) + \rho\tilde{\theta}\Delta(g_n(\tilde{\theta}),\tilde{\theta})\right)(1-F(\hat{\theta})).$$

Rearranging terms gives

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\tilde{\theta}-\hat{\theta}) + \beta(\theta-\tilde{\theta})\right]\, dF(\theta) \le \left(\beta\frac{\nu'(g_n(\tilde{\theta}))}{U'(g_n(\tilde{\theta}))} + (\rho-\beta)\tilde{\theta}\Delta(g_n(\tilde{\theta}),\tilde{\theta})\right)(1-F(\hat{\theta})).$$

Using the definition of g_n, i.e., equation (9), on the right-hand side gives

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\tilde{\theta}-\hat{\theta}) + \beta(\theta-\tilde{\theta})\right]\, dF(\theta) \le \rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})}(1-F(\hat{\theta})). \tag{31}$$

Add and subtract $\rho\theta$ to the integrand on the left-hand side and rearrange terms to get

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\theta-\hat{\theta}) + (\beta-\rho)(\theta-\tilde{\theta})\right]\, dF(\theta) \le \rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})}(1-F(\hat{\theta})).$$

Integrating the left-hand side by parts gives inequality (30). □

For $(h, h_t) \in \Phi$, h is non-decreasing and because the inner integral from (29.2) is non-positive, the integral (29.2) is non-positive for $(h, h_t) \in \Phi$. The Gateaux derivative of the Lagrangian at the candidate solution $(u^*, \underline{t}^*)$ is null in the direction $(h, h_t) = (u^*, \underline{t}^*)$ and negative in all directions $(h, h_t) \in \Phi$. □

B.3 Proof of Proposition 3

Proof of Proposition 3 part 1). The proof consists of applying Lemma 9. Let $u^*(\theta) = U(g_c)$ for $\theta \in \Theta$ and $\underline{t}^* = 0$, hence $(u^*, \underline{t}^*) \in \Phi$. The Lagrange multiplier function $\Lambda^*(\theta) = 1$ for $\theta \in \Theta$ is valid since it is non-decreasing and $1 - \Lambda^*$ is integrable.

The marginal bias and the wedge evaluated at g_c are $\frac{\nu'(g_c)}{U'(g_c)} = (1-\beta)\int_\Theta \tilde{\theta} dF(\tilde{\theta})$, and $\theta\Delta(g_c,\theta)) = \theta - \beta\int_\Theta \tilde{\theta} dF(\tilde{\theta})$. The Gateaux derivative (20), evaluated at $(u^*, \underline{t}^*)$ and in the direction (h, h_t), given the constant Lagrange multiplier Λ^*, reads

$$\partial\mathcal{L}(u^*, \underline{t}^*, h, h_t|\Lambda^*) = \int_\Theta \left[\left(\rho\tfrac{1-F(\theta)}{f(\theta)} - \beta(1-\rho)\int_\Theta \tilde{\theta}\, dF(\tilde{\theta}) + (\beta-\rho)\theta\right) h(\theta)\right] dF(\theta) \tag{32}$$

$$+ \left(\left(\underline{\theta} - \beta\int_\Theta \tilde{\theta}\, dF(\tilde{\theta})\right) h(\underline{\theta}) - h_t\right)\rho. \tag{33}$$

Integrating the inverse hazard rate by parts gives the following: $\int_{\underline{\theta}}^{\hat{\theta}} \frac{1-F(\theta)}{f(\theta)}\, dF(\theta) = \hat{\theta}(1-F(\hat{\theta})) - \underline{\theta} + \int_{\underline{\theta}}^{\hat{\theta}} \theta dF(\theta)$. Since the expectation of θ exists, $\hat{\theta}(1-F(\hat{\theta})) \to 0$ as $\hat{\theta} \to \bar{\theta}$. Substitute $\underline{\theta}$ in (33),

$$\left(\underline{\theta} - \beta\int_\Theta \tilde{\theta}\, dF(\tilde{\theta})\right) h(\underline{\theta}) - h_t = -\left(\int_\Theta \left[\tfrac{1-F(\theta)}{f(\theta)} - (1-\beta)\theta\right]\, dF(\theta)\right) h(\underline{\theta}) - h_t.$$

First, I show that the Gateaux derivative in the direction of the candidate solution is null when evaluated at the candidate solution. Consider the direction $h(\theta) = u^*(\theta)$ for $\theta \in \Theta$ and $h_t = \underline{t}^* = 0$. Since h is constant, $h(\theta) = h(\underline{\theta})$ for $\theta \in \Theta$ and $h(\underline{\theta})$ can be taken out of the expectation in (32). Since $-\beta(1-\rho) + \beta - \rho = -\rho(1-\beta)$, it follows that $\partial\mathcal{L}(u^*, \underline{t}^*, u^*, \underline{t}^*|\Lambda^*) = 0$ as claimed.

It remains to show that the Gateaux derivative evaluated at $(u^*, \underline{t}^*)$ in any direction $(h, h_t) \in \Phi$ is non-positive. Integrating (32) by parts gives

$$\int_\Theta \left[\rho\tfrac{1-F(\theta)}{f(\theta)} - \beta(1-\rho)\int_\Theta \tilde{\theta} dF(\tilde{\theta}) + (\beta-\rho)\theta\right] dF(\theta) h(\underline{\theta}) \tag{32.1}$$

$$+ \int_\Theta \left[\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho\tfrac{1-F(\theta)}{f(\theta)} - \beta(1-\rho)\int_\Theta \tilde{\theta}\, dF(\tilde{\theta}) + (\beta-\rho)\theta\right] dF(\theta)\right] dh(\hat{\theta}). \tag{32.2}$$

Claim 3. *Suppose Assumption I holds for* $\underline{\theta}$ *and the derivative of* $\rho\frac{1-F(\theta)}{f(\theta)}$ *is smaller than* $\rho - \beta$ *for* $\theta \in \Theta$, *then*

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho\tfrac{1-F(\theta)}{f(\theta)} - \beta(1-\rho)\int_\Theta \tilde{\theta}\, dF(\tilde{\theta}) + (\beta-\rho)\theta\right] dF(\theta) \le 0 \tag{32.2.i}$$

for all $\hat{\theta} \in \Theta$ *and* $\int_\Theta \left[\frac{1-F(\theta)}{f(\theta)} - (1-\beta)\theta\right]\, dF(\theta) \le 0$.

Proof of Claim 3. The proof consists of first showing that the inequality (32.2.i) holds for $\hat{\theta} = \underline{\theta}$. The condition on the slope of the inverse hazard rate implies $\rho\frac{1-F(\theta)}{f(\theta)} - (\rho-\beta)\theta \le \rho\frac{1-F(\underline{\theta})}{f(\underline{\theta})} - (\rho-\beta)\underline{\theta}$ for $\theta \in \Theta$. Assumption I for $\underline{\theta}$ implies $\rho\frac{1-F(\underline{\theta})}{f(\underline{\theta})} \le (1-\beta)\underline{\theta}$. Combining the two inequalities gives $\rho\frac{1-F(\theta)}{f(\theta)} - (\rho-\beta)\theta \le (1-\rho)\underline{\theta}$. Taking expectations on both sides gives

$$\rho\int_\Theta \tfrac{1-F(\theta)}{f(\theta)} dF(\theta) - (\rho-\beta)\int_\Theta \theta dF(\theta) \le (1-\rho)\underline{\theta}. \tag{34}$$

Integrating the inverse hazard rate by parts to substitute $\underline{\theta}$ in (34) gives

$$\rho \int_{\Theta} \tfrac{1-F(\theta)}{f(\theta)} dF(\theta) - (\rho - \beta) \int_{\Theta} \theta dF(\theta) \leq (1-\rho) \left(\int_{\Theta} \theta dF(\theta) - \int_{\Theta} \tfrac{1-F(\theta)}{f(\theta)} \, dF(\theta) \right),$$

which simplifies to $\int_{\Theta} \frac{1-F(\theta)}{f(\theta)} dF(\theta) \leq (1-\beta) \int_{\Theta} \theta dF(\theta)$. Note that for $\hat{\theta} = \underline{\theta}$, inequality (32.2.i) reduces to $\int_{\Theta} \left[\frac{1-F(\theta)}{f(\theta)} - (1-\beta)\theta \right] dF(\theta)$.

Second, I show that given that inequality (32.2.i) holds for $\hat{\theta} = \underline{\theta}$, then the condition on the derivative of the inverse hazard rate implies that inequality (32.2.i) holds for $\hat{\theta} \in \Theta$. Rewrite inequality (32.2.i) as follows: $E\left[\rho \frac{1-F(\theta)}{f(\theta)} - (\rho - \beta)\theta \;\middle|\; \theta \geq \hat{\theta} \right] \leq \beta(1-\rho) \int_{\Theta} \theta \, dF(\theta)$. The condition on the slope of the inverse hazard rate implies that $\rho \frac{1-F(\theta)}{f(\theta)} - (\rho - \beta)\theta$ is decreasing. It follows that the conditional expectation $E\left[\rho \frac{1-F(\theta)}{f(\theta)} - (\rho - \beta)\theta \;\middle|\; \theta \geq \hat{\theta} \right]$ is a decreasing function of $\hat{\theta}$. Since the inequality holds for $\hat{\theta} = \underline{\theta}$, it follows that it holds for $\hat{\theta} \in \Theta$. □

For any $(h, h_t) \in \Phi$, h is non-decreasing so inequality (32.2.i) implies that line (32.2) is non-positive. Claim 3 also shows that line (32.1) is non-positive. □

Proof of Proposition 3 part 2). The proof consists of applying Lemma 9. Let $u^*(\theta) = U(g_x^{np}(\theta))$ for $\theta \in \Theta$ and $\underline{t}^* = 0$. The lower bound on the derivative of the inverse hazard rate implies that $g_x^{np}(\theta)$ is non-decreasing, hence $(u^*, \underline{t}^*) \in \Phi$. The Lagrange multiplier function is $\Lambda^*(\theta) = 1$ for $\theta \in \Theta$ is valid since it is non-decreasing and $1 - \Lambda$ is integrable.

Substituting the wedge $\theta\Delta(g, \theta) = \theta - \frac{\beta}{1-\beta} \frac{\nu'(g)}{U'(g)}$ and the constant Lagrange multiplier function Λ^* in the Gateaux derivative (20) gives:

$$\begin{aligned} \partial\mathcal{L}(u, \underline{t}, h, h_t | \Lambda^*) = &\int_{\Theta} \left[\left(\rho \tfrac{1-F(\theta)}{f(\theta)} - \beta \left(\tfrac{1-\rho}{1-\beta} \right) \tfrac{\nu'(U^{-1}(u(\theta)))}{U'(U^{-1}(u(\theta)))} + (\beta - \rho)\theta \right) h(\theta) \right] dF(\theta) \\ &+ \left(\left(\underline{\theta} - \tfrac{\beta}{1-\beta} \tfrac{\nu'(U^{-1}(u(\underline{\theta})))}{U'(U^{-1}(u(\underline{\theta})))} \right) h(\underline{\theta}) - h_t \right) \rho. \end{aligned}$$

The definition of g_x^{np} implies the following profile of utility:

$$\beta \, \frac{1-\rho}{1-\beta} \, \frac{\nu'(U^{-1}(u^*(\theta)))}{U'(U^{-1}(u^*(\theta)))} = \begin{cases} \rho \frac{1-F(\theta_{np})}{f(\theta_{np})} + (\beta - \rho)\theta_{np} & \text{for } \theta \geq \theta_{np} \\ \rho \frac{1-F(\theta)}{f(\theta)} + (\beta - \rho)\theta & \text{for } \theta_x \leq \theta \leq \theta_{np} \\ \rho \frac{1-F(\theta_x)}{f(\theta_x)} + (\beta - \rho)\theta_x & \text{for } \theta \leq \theta_x \, . \end{cases}$$

The Lagrangian, evaluated at the allocation u^*, reduces to

$$\partial\mathcal{L}(u^*, \underline{t}^*, h, h_t|\Lambda^*) = \int_{\theta_{np}}^{\bar{\theta}} \left[\left(\rho\tfrac{1-F(\theta)}{f(\theta)} - \rho\tfrac{1-F(\theta_{np})}{f(\theta_{np})} + (\beta-\rho)(\theta-\theta_{np})\right) h(\theta)\right] dF(\theta) \tag{35}$$

$$+ \int_{\underline{\theta}}^{\theta_x} \left[\left(\rho\tfrac{1-F(\theta)}{f(\theta)} - \rho\tfrac{1-F(\theta_x)}{f(\theta_x)} + (\beta-\rho)(\theta-\theta_x)\right) h(\theta)\right] dF(\theta) \tag{36}$$

$$+ \left(\left(\underline{\theta} - \tfrac{\rho}{1-\rho}\tfrac{1-F(\theta_x)}{f(\theta_x)} - \tfrac{\beta-\rho}{1-\rho}\theta_x\right) h(\underline{\theta}) - h_t\right)\rho. \tag{37}$$

Integrating (35) and (36) by parts and grouping terms,

$$\partial\mathcal{L}(u^*, \underline{t}^*, h, h_t|\Lambda^*) = -\rho h_t \tag{38}$$

$$+ \int_{\theta_{np}}^{\bar{\theta}} \left[\rho\tfrac{1-F(\theta)}{f(\theta)} - \rho\tfrac{1-F(\theta_{np})}{f(\theta_{np})} + (\beta-\rho)(\theta-\theta_{np})\right] dF(\theta)\ h(\theta_{np}) \tag{39}$$

$$+ \int_{\theta_{np}}^{\bar{\theta}} \left[\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho\tfrac{1-F(\theta)}{f(\theta)} - \rho\tfrac{1-F(\theta_{np})}{f(\theta_{np})} + (\beta-\rho)(\theta-\theta_{np})\right] dF(\theta)\right] dh(\hat{\theta}) \tag{40}$$

$$+ \left(\int_{\underline{\theta}}^{\theta_x} \left[\rho\tfrac{1-F(\theta)}{f(\theta)} - \rho\tfrac{1-F(\theta_x)}{f(\theta_x)} + (\beta-\rho)(\theta-\theta_x)\right] dF(\theta) + \left(\underline{\theta} - \tfrac{\rho}{1-\rho}\tfrac{1-F(\theta_x)}{f(\theta_x)} - \tfrac{\beta-\rho}{1-\rho}\theta_x\right)\rho\right) h(\theta_x) \tag{41}$$

$$- \int_{\underline{\theta}}^{\theta_x}\left[\int_{\underline{\theta}}^{\hat{\theta}}\left[\rho\tfrac{1-F(\theta)}{f(\theta)} - \rho\tfrac{1-F(\theta_x)}{f(\theta_x)} + (\beta-\rho)(\theta-\theta_x)\right]dF(\theta) + \left(\underline{\theta} - \tfrac{\rho}{1-\rho}\tfrac{1-F(\theta_x)}{f(\theta_x)} + \tfrac{\beta-\rho}{1-\rho}\theta_x\right)\rho\right] dh(\hat{\theta}), \tag{42}$$

using that $\lim_{\hat{\theta}\to\bar{\theta}} \int_{\hat{\theta}}^{\bar{\theta}} \left[\rho\frac{1-F(\theta)}{f(\theta)} - \rho\frac{1-F(\theta_{np})}{f(\theta_{np})} + (\beta-\rho)(\theta-\theta_{np})\right] dF(\theta) h(\hat{\theta})$ is zero since h is bounded for $\bar{\theta} < \infty$. If $\bar{\theta} = \infty$, the result follows from taking the limit of a sequence of environments with compact support as in the proof of Proposition 1 in Appendix B.1.

The terms (39) and (40) with $\hat{\theta} = \theta_{np}$ are null and (40) is non-positive for $\hat{\theta} \geq \theta_{np}$ (see Claim 2 in the proof of Proposition 2 in Appendix B.2).

Claim 4. *For $\hat{\theta} \leq \theta_x$,*

$$\int_{\underline{\theta}}^{\hat{\theta}} \left[\rho\tfrac{1-F(\theta)}{f(\theta)} - \rho\tfrac{1-F(\theta_x)}{f(\theta_x)} - (\rho-\beta)(\theta-\theta_x)\right]\ dF(\theta) \geq -\rho\underline{\theta} + \rho\left(\tfrac{\rho}{1-\rho}\tfrac{1-F(\theta_x)}{f(\theta_x)} - \tfrac{\rho-\beta}{1-\rho}\theta_x\right), \tag{43}$$

with equality at $\hat{\theta} = \theta_x$.

Proof of Claim 4. Inequality (43) holds with equality at $\hat{\theta} = \theta_x$ precisely because of the definition of an interior θ_x in (11).

Since the derivative of $\rho\frac{1-F(\theta)}{f(\theta)}$ is not smaller than $\rho - \beta$ over $\theta \in [\underline{\theta}, \theta_{np}]$ and $\theta_x \leq \theta_{np}$, integrating from θ to θ_x gives $\rho\frac{1-F(\theta)}{f(\theta)} - \rho\frac{1-F(\theta_x)}{f(\theta_x)} \leq (\rho-\beta)(\theta-\theta_x)$. The left-hand side of inequality (43) is hence decreasing as a function of $\hat{\theta}$, and the inequality holds with equality at $\hat{\theta} = \theta_x$. It follows that inequalities (43) hold for $\hat{\theta} \leq \theta_x$ as claimed. □

The last step of the proof shows that the conditions in terms of Gateaux derivatives in Lemma 9 are satisfied. First, consider the different terms (38)-(42) of the Gateaux derivative in the direction of the solution $(h, h_t) = (u^*, \underline{t}^*)$. Since $\underline{t}^* = 0$, the term on line (38) is zero. Lines (39) and (41) are zero by Claim 2 and Claim 4. Lines (40) and (42) are zero because u^* is constant above θ_{np} and below θ_x. Hence, $\partial\mathcal{L}(u^*, \underline{t}^*, u^*, \underline{t}^*|\Lambda^*) = 0$ as desired.

Second, consider the Gateaux derivative (38)-(42) in any direction $(h, h_t) \in \Phi$. Since $h_t \geq 0$, the term on line (38) is negative. Claim 2 and Claim 4 imply that lines (40) and (42) are negative since h is non-decreasing and lines (39) and (41) are zero. Hence, $\partial\mathcal{L}(u^*, \underline{t}^*, h, h_t|\Lambda^*) \leq 0$ for $(h, h_t) \in \Phi$ as desired. □

B.4 Proof of Proposition 4

Proof of Proposition 4. The proof is identical to the proof of Proposition 3 part 1) up to Claim 3. The rest of the proof uses implications from the definition of a high degree of present bias to imply the same conclusion as the one in Claim 3. Note that

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho\frac{1-F(\theta)}{f(\theta)} - \beta(1-\rho)\int_{\Theta} \theta \, dF(\theta) - (\rho-\beta)\theta\right] dF(\theta) \leq -\beta(1-\rho)\int_{\Theta} \theta \, dF(\theta) \leq 0,$$

where the first inequality uses Assumption H, $\rho\frac{1-F(\theta)}{f(\theta)} \leq (\rho-\beta)\theta$, for $\theta \in \Theta$, and the second inequality follows from $\rho \leq 1$. A high degree of present bias also implies $\frac{1-F(\theta)}{\theta f(\theta)} \leq 1-\beta$ for $\theta \in \Theta$ because $\frac{\rho-\beta}{\rho} \leq 1-\beta$ for $0 < \rho \leq 1$. Hence $\int_{\hat{\theta}}^{\bar{\theta}} \left[\frac{1-F(\theta)}{f(\theta)} - (1-\beta)\theta\right] dF(\theta) \leq 0$ for $\hat{\theta} \in \Theta$. For $(h, h_t) \in \Phi$, h is non-decreasing and $h_t \geq 0$. Hence $\partial\mathcal{L}(u^*, \underline{t}^*, h, h_t|\Lambda^*) \leq 0$ for $(h, h_t) \in \Phi$. □

Online Appendix

OA Proofs

OA.1 Lemma 1 on the second requirement in the definition of θ_p

Proof of Lemma 1. A first step consists of rewriting inequality (7) as one of the first-order conditions of the Lagrangian method. The first step is Claim 1 in the proof of Proposition 1 in Appendix B.1. For convenience, I repeat the claim here.

Claim 1. *Inequality* (7) *is equivalent to*

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\left(-(1-\beta)\tilde{\theta} + \rho \frac{1-F(\theta)}{f(\theta)} + (\beta-\rho)(\theta-\tilde{\theta}) \right) f(\theta) \right] d\theta \leq 0. \tag{23}$$

The proof of Claim 1 is in the proof of Proposition 1 in Appendix B.1. It also shows that inequalities (23) and inequality (24) are equivalent.

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\tilde{\theta}-\hat{\theta}) + \beta(\theta-\tilde{\theta}) \right] dF(\theta) \leq (1-\beta)\tilde{\theta}(1-F(\hat{\theta})). \tag{24}$$

The second step uses the equivalence between inequalities (7) and (24) and the assumption that inequality (7) holds with equality for some $\tilde{\theta} < \bar{\theta}$ and $\hat{\theta} = \tilde{\theta}$ to get:

$$\int_{\tilde{\theta}}^{\bar{\theta}} \left[\beta(\theta-\tilde{\theta}) \right] dF(\theta) = (1-\beta)\tilde{\theta}(1-F(\tilde{\theta})). \tag{44}$$

Multiplying both sides of (24) by $(1-F(\tilde{\theta}))$ and using (44) for the right-hand side gives

$$(1-F(\tilde{\theta})) \int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\tilde{\theta}-\hat{\theta}) + \beta(\theta-\tilde{\theta}) \right] dF(\theta) \leq (1-F(\hat{\theta})) \int_{\tilde{\theta}}^{\bar{\theta}} \left[\beta(\theta-\tilde{\theta}) \right] dF(\theta)$$

which simplifies to

$$(1-F(\tilde{\theta})) \int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\tilde{\theta}-\hat{\theta}) + \beta\theta \right] dF(\theta) \leq (1-F(\hat{\theta})) \int_{\tilde{\theta}}^{\bar{\theta}} \left[\beta\theta \right] dF(\theta).$$

Rearranging terms give

$$(1-F(\tilde{\theta}))\beta \int_{\hat{\theta}}^{\bar{\theta}} \theta \mathrm{d}F(\theta) - (1-F(\tilde{\theta}))(1-F(\hat{\theta}))\rho\hat{\theta} \leq (1-F(\hat{\theta}))\beta \int_{\tilde{\theta}}^{\bar{\theta}} \theta \mathrm{d}F(\theta) - (1-F(\hat{\theta}))(1-F(\tilde{\theta}))\rho\tilde{\theta}.$$

Since $\tilde{\theta} \leq \hat{\theta} < \bar{\theta}$, both sides can be divided by $(1-F(\tilde{\theta}))(1-F(\hat{\theta}))$ to give the result. □

OA.2 Lemma 2 on an implication of Assumption L

Proof of Lemma 2. The proof proceeds in three steps. The first step shows that the definition of θ_p implies $\rho\frac{1-F(\theta_p)}{\theta_p f(\theta_p)} \geq 1-\beta$ if $\theta_p \in (\underline{\theta}, \bar{\theta})$. The second requirement in the definition of θ_p is the following inequality on the conditional tail expectation (see Lemma 1): $\beta E[\theta|\theta \geq \hat{\theta}] - \rho\hat{\theta} \leq \beta E[\theta|\theta \geq \theta_p] - \rho\theta_p$ for $\hat{\theta} \geq \theta_p$. The derivative of the left-hand side with respect to $\hat{\theta}$ reads $-\beta\frac{\hat{\theta}f(\hat{\theta})}{1-F(\hat{\theta})} + \beta\mathbb{E}[\theta|\theta \geq \hat{\theta}]\frac{f(\hat{\theta})}{1-F(\hat{\theta})} - \rho$. The derivative must be negative at $\hat{\theta} = \theta_p$ because the inequality holds with equality at θ_p. Using that $\beta\mathbb{E}[\theta|\theta \geq \theta_p] = \theta_p$ for an interior θ_p gives $-\beta\frac{\theta_p f(\theta_p)}{1-F(\theta_p)} + \theta_p\frac{f(\theta_p)}{1-F(\theta_p)} - \rho \leq 0$, which completes the first step.

The second step proves the following claim.

Claim 5. *Suppose that Assumption L holds for $\theta \leq \theta^*$ and there exists $\theta_* \leq \theta^*$ such that $\rho\frac{1-F(\theta_*)}{\theta_* f(\theta_*)} < 1-\beta$. Then $\rho\frac{1-F(\theta)}{\theta f(\theta)} < 1-\beta$ for $\theta \in [\theta_*, \theta^*]$.*

Proof of Claim 5 . For any $\theta \leq \theta^*$,

$$\begin{aligned}\frac{d}{d\theta}\left(\rho\frac{1-F(\theta)}{f(\theta)} - \theta(1-\beta)\right) &= -\rho - \rho\frac{1-F(\theta)}{\theta f(\theta)}\frac{\theta f'(\theta)}{f(\theta)} - (1-\beta)\\ &\leq \rho\frac{1-F(\theta)}{\theta f(\theta)}\frac{1-\beta+\rho}{1-\beta} - (1-\beta+\rho)\\ &= \frac{1-\beta+\rho}{1-\beta}\left(\rho\frac{1-F(\theta)}{\theta f(\theta)} - (1-\beta)\right),\end{aligned}$$

in which the inequality follows from Assumption L. By assumption, $\rho\frac{1-F(\theta_*)}{\theta_* f(\theta_*)} - 1 - \beta < 0$ for $\theta_* \leq \theta^*$. Given that $\frac{1-\beta+\rho}{1-\beta} \geq 0$, combining the two inequalities implies $\frac{d}{d\theta}\left(\rho\frac{1-F(\theta)}{f(\theta)} - \theta(1-\beta)\right) < 0$ for $\theta \in [\theta_*, \theta^*]$. □

The last step shows that $\rho\frac{1-F(\theta)}{\theta f(\theta)} \geq 1-\beta$ for $\theta \leq \theta_p$ by contradiction. Suppose not, so there exists $\theta_* < \theta_p$ such that $\rho\frac{1-F(\theta_*)}{\theta_* f(\theta_*)} < 1-\beta$. Claim 5 implies that $\rho\frac{1-F(\theta_*)}{\theta_* f(\theta_*)} < 1-\beta$ for $\theta \in [\theta_*, \theta_p]$, which contradicts $\rho\frac{1-F(\theta_p)}{\theta_p f(\theta_p)} \geq 1-\beta$. □

OA.3 Lemma 3 on incentive compatible money-burning schedules

Proof of Lemma 3. The proof follows the argument in Myerson (1981). Suppose that $g(\cdot)$ is incentive compatible given a money-burning schedule $t(\cdot)$. Define $V(\theta) = \theta U(g(\theta)) + \beta W(T - g(\theta)) - \beta t(\theta)$ and $u(\theta) = U(g(\theta))$. Consider $\theta > \hat{\theta}$, incentive compatibility implies,

$$V(\theta) \geq V(\hat{\theta}) + (\theta - \hat{\theta})u(\hat{\theta}), \quad \text{and} \quad V(\hat{\theta}) \geq V(\theta) + (\hat{\theta} - \theta)u(\theta).$$

The inequalities combined imply that $u(\cdot)$ is non-decreasing,

$$u(\theta) \geq \frac{V(\theta) - V(\hat{\theta})}{\theta - \hat{\theta}} \geq u(\hat{\theta}).$$

Since U is strictly increasing and $u(\cdot)$ is non-decreasing, g is also non-decreasing and $V(\cdot)$ is continuous and differentiable almost everywhere. Taking the limit, $V'(\theta) = u(\theta)$. Integrating from $\underline{\theta}$ to θ gives $V(\theta) = V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} u(\theta)$. Replacing V and u by their definitions gives (8).

Suppose instead that $g(\cdot)$ is non-decreasing and, for a given $t(\underline{\theta})$, define $t(\cdot)$ according to (8). Using the definitions of V and u, rewrite (8) as follows: $V(\theta) = V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} u(\theta) d\theta$. For $\theta \geq \hat{\theta}$,

$$V(\theta) - V(\hat{\theta}) = \int_{\hat{\theta}}^{\theta} u(\theta) d\theta \geq \int_{\hat{\theta}}^{\theta} u(\hat{\theta}) d\theta = (\theta - \hat{\theta}) u(\hat{\theta}).$$

The inequality holds because a non-decreasing $g(\cdot)$ implies that $u(\cdot)$ is also non-decreasing. Substituting the definitions of V and u gives the incentive compatibility constraints. □

OA.4 Lemma 4 on the monotonicity of g_n

Proof of Lemma 4. Using the observation from footnote 9, rewrite equation (9) as follows:

$$\frac{1-\rho}{1-\beta} \frac{\nu'(g_n(\theta))}{U'(g_n(\theta))} = \frac{\rho}{\beta} \frac{1-F(\theta)}{f(\theta)} + \frac{\beta - \rho}{\beta} \theta.$$

For $\rho < 1$, we have $0 < \frac{1-\rho}{1-\beta}$. The bias ν is convex and the utility index U is strictly concave, so the ratio $\frac{\nu'(g_n)}{U'(g_n)}$ is unambiguously increasing in the argument g_n. The right-hand side is non-decreasing in θ if and only if the derivative of $\rho \frac{1-F(\theta)}{f(\theta)}$ is not smaller than $\rho - \beta$. □

OA.5 Lemma 5 on non-negative marginal sanctions

Proof of Lemma 5. The Euler equation characterizing g_d is $\theta U'(g_d(\theta)) = \beta W'(T - g_d(\theta))$. For ease of comparison, I rewrite equation (9) defining g_n as follows:

$$\left(\frac{\rho}{1-\rho} \frac{1-F(\theta)}{\theta f(\theta)} + \frac{\beta - \rho}{1-\rho} \right) \theta U'(g_n(\theta)) = \beta W'(T - g_n(\theta)).$$

With U strictly concave and W concave, $g_n(\theta) \leq g_d(\theta)$ if and only if the term in the parenthesis is smaller than one. Hence $g_n(\theta) \leq g_d(\theta)$ if and only if $\rho \frac{1-F(\theta)}{\theta f(\theta)} \leq 1 - \beta$. □

OA.6 Lemma 6 on the second requirement in the definition of θ_{np}

Proof of Lemma 6. A first step consists of rewriting inequality (10) as one of the first-order conditions of the Lagrangian method. The first step is Claim 2 in the proof of Proposition 2 in Appendix B.2. I repeat the claim here for convenience.

Claim 2. *Inequality* (10) *is equivalent to*

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho\frac{1-F(\theta)}{f(\theta)} + (\beta-\rho)(\theta-\tilde{\theta})\right] dF(\theta) \leq \rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})}(1-F(\hat{\theta})). \tag{30}$$

The proof of Claim 2 is in the proof of Proposition 2 in Appendix B.2. It also shows that inequality (30) and inequality (31) are equivalent.

$$\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\tilde{\theta}-\hat{\theta}) + \beta(\theta-\tilde{\theta})\right]\, dF(\theta) \leq \rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})}(1-F(\hat{\theta})). \tag{31}$$

The second step uses the equivalence between inequalities (10) and (31) and the assumption that inequality (10) holds with equality for some $\tilde{\theta} < \bar{\theta}$ and $\hat{\theta} = \tilde{\theta}$ to get:

$$\int_{\tilde{\theta}}^{\bar{\theta}} \left[\beta(\theta-\tilde{\theta})\right]\, dF(\theta) = \rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})}(1-F(\tilde{\theta})). \tag{45}$$

Multiplying both sides of (31) by $(1-F(\tilde{\theta}))$ and using (45) for the right-hand side gives

$$(1-F(\tilde{\theta}))\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\tilde{\theta}-\hat{\theta}) + \beta(\theta-\tilde{\theta})\right] dF(\theta) \leq (1-F(\hat{\theta}))\int_{\tilde{\theta}}^{\bar{\theta}} \left[\beta(\theta-\tilde{\theta})\right] dF(\theta),$$

which simplifies to

$$(1-F(\tilde{\theta}))\int_{\hat{\theta}}^{\bar{\theta}} \left[\rho(\tilde{\theta}-\hat{\theta}) + \beta\theta\right] dF(\theta) \leq (1-F(\hat{\theta}))\int_{\tilde{\theta}}^{\bar{\theta}} \left[\beta\theta\right] dF(\theta).$$

Rearranging terms give:

$$(1-F(\tilde{\theta}))\beta\int_{\hat{\theta}}^{\bar{\theta}} \theta \mathrm{d}F(\theta) - (1-F(\tilde{\theta}))(1-F(\hat{\theta}))\rho\hat{\theta} \leq (1-F(\hat{\theta}))\beta\int_{\tilde{\theta}}^{\bar{\theta}} \theta \mathrm{d}F(\theta) - (1-F(\hat{\theta}))(1-F(\tilde{\theta}))\rho\tilde{\theta}.$$

Since $\tilde{\theta} \leq \hat{\theta} < \bar{\theta}$, both sides can be divided by $(1-F(\tilde{\theta}))(1-F(\hat{\theta}))$ to give the result. □

OA.7 Lemma 7 on the implications of Assumption L

Proof of Lemma 7. For the first part, note that $\theta_n \in (\underline{\theta}, \bar{\theta})$ implies that $\rho\frac{1-F(\theta_n)}{\theta_n f(\theta_n)} = 1-\beta$. The argument is by contradiction. Suppose that there exists $\theta_* \leq \theta_n$ such that $\rho\frac{1-F(\theta_*)}{\theta_* f(\theta_*)} < 1-\beta$.

Claim 5, implies that $\rho\frac{1-F(\theta)}{\theta f(\theta)} < 1-\beta$ for $\theta \in [\theta_*, \theta_n]$, which contradicts $\rho\frac{1-F(\theta_n)}{\theta_n f(\theta_n)} = 1-\beta$ (for Claim 5, see Proof of Lemma 2 in Appendix OA.2).

The second part uses $\Delta(g_n(\theta), \theta) = -\frac{1}{1-\rho}\big(\rho\frac{1-F(\theta)}{\theta f(\theta)} - (1-\beta)\big)$.

$$\begin{aligned}
\frac{d}{d\theta}\theta\Delta(g_n(\theta),\theta) &= \frac{1}{1-\rho}\frac{d}{d\theta}\left(\theta(1-\beta) - \rho\tfrac{1-F(\theta)}{f(\theta)}\right) \\
&= \frac{1}{1-\rho}\frac{1+\rho-\beta}{1-\beta}\left(1-\beta+\rho\tfrac{1-F(\theta)}{\theta f(\theta)}\tfrac{\theta f'(\theta)}{f(\theta)}\tfrac{1-\beta}{1+\rho-\beta}\right) \\
&\geq \frac{1+\rho-\beta}{1-\beta}\frac{1}{1-\rho}\left(1-\beta-\rho\tfrac{1-F(\theta)}{\theta f(\theta)}\right) \\
&= \frac{1+\rho-\beta}{1-\beta}\Delta(g_n(\theta),\theta),
\end{aligned}$$

where the inequality is equivalent to Assumption L. □

OA.8 Lemma 8 on the second requirement in the definition of θ_x

Proof of Lemma 8. A first step consists of rewriting inequality (11) as the first-order condition of the Lagrangian method. I record this step in the following claim.

Claim 6. *Inequality* (11) *is equivalent to*

$$\int_{\underline{\theta}}^{\hat{\theta}} \left[\rho\tfrac{1-F(\theta)}{f(\theta)} - (\rho-\beta)(\theta-\tilde{\theta})\right]\, dF(\theta) \geq \rho\tfrac{1-F(\tilde{\theta})}{f(\tilde{\theta})}F(\hat{\theta}) - \rho\underline{\theta} + \rho\left(\tfrac{\rho}{1-\rho}\tfrac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - \tfrac{\rho-\beta}{1-\rho}\tilde{\theta}\right). \quad (46)$$

Proof of Claim 6. The definition of the wedge implies the following identity:

$$\hat{\theta}\Delta(g,\hat{\theta}) = \hat{\theta} - \frac{\beta W'(T-g)}{U'(g)} = \hat{\theta} - \tilde{\theta} + \tilde{\theta}\Delta(g,\tilde{\theta}).$$

Multiplying both sides of inequality (11) by β and substituting the identity for the wedge gives

$$\beta\frac{\nu'(g_n(\tilde{\theta}))}{U'(g_n(\tilde{\theta}))}F(\hat{\theta}) - \beta\int_{\underline{\theta}}^{\hat{\theta}}\left[\theta-\tilde{\theta}+\tilde{\theta}\Delta(g_n(\tilde{\theta}),\tilde{\theta})\right]dF(\theta) \leq \rho(\hat{\theta}-\tilde{\theta}+\tilde{\theta}\Delta(g_n(\tilde{\theta}),\tilde{\theta}))(1-F(\hat{\theta})).$$

Grouping terms gives:

$$\left(\beta\frac{\nu'(g_n(\tilde{\theta}))}{U'(g_n(\tilde{\theta}))} + (\rho-\beta)\tilde{\theta}\Delta(g_n(\tilde{\theta}),\tilde{\theta})\right)F(\hat{\theta}) \leq \int_{\underline{\theta}}^{\hat{\theta}}[\beta(\theta-\tilde{\theta})-\rho(\hat{\theta}-\tilde{\theta})]dF(\theta)+\rho(\hat{\theta}-\tilde{\theta}+\tilde{\theta}\Delta(g_n(\tilde{\theta}),\tilde{\theta})).$$

Using the definition of g_n, i.e., equation (9), on the left-hand side gives

$$\rho\tfrac{1-F(\tilde{\theta})}{f(\tilde{\theta})}F(\hat{\theta}) \leq \int_{\underline{\theta}}^{\hat{\theta}}[\beta(\theta-\tilde{\theta})-\rho(\hat{\theta}-\tilde{\theta})]dF(\theta) + \rho(\hat{\theta}-\tilde{\theta}) + \rho\tilde{\theta}\Delta(g_n(\tilde{\theta}),\tilde{\theta}).$$

Again, using the definition of g_n, i.e., equation (9), expressed as a function of the wedge reads: $\tilde{\theta}\Delta(g_n(\tilde{\theta}),\tilde{\theta}) = \tilde{\theta} - \left(\frac{\rho}{1-\rho}\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - \frac{\rho-\beta}{1-\rho}\tilde{\theta}\right)$, which, upon substitution in the previous inequality gives

$$\rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})}F(\hat{\theta}) \leq \int_{\underline{\theta}}^{\hat{\theta}} \left[\beta(\theta-\tilde{\theta}) - \rho(\hat{\theta}-\tilde{\theta})\right]dF(\theta) + \rho(\hat{\theta}-\tilde{\theta}) + \rho\tilde{\theta} - \rho\left(\frac{\rho}{1-\rho}\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - \frac{\rho-\beta}{1-\rho}\tilde{\theta}\right).$$

Upon subtracting $\rho\underline{\theta}$ on both sides and rearranging terms, the inequality reads

$$\rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})}F(\hat{\theta}) - \rho\underline{\theta} + \rho\left(\frac{\rho}{1-\rho}\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - \frac{\rho-\beta}{1-\rho}\tilde{\theta}\right) \leq \int_{\underline{\theta}}^{\hat{\theta}} \left[\beta(\theta-\tilde{\theta}) - \rho(\hat{\theta}-\tilde{\theta})\right]dF(\theta) + \rho\hat{\theta} - \rho\underline{\theta}.$$

Adding and subtracting $\rho\theta$ to the integrand on the right-hand side gives

$$\rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})}F(\hat{\theta}) - \rho\underline{\theta} + \rho\left(\frac{\rho}{1-\rho}\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - \frac{\rho-\beta}{1-\rho}\tilde{\theta}\right) \leq \int_{\underline{\theta}}^{\hat{\theta}} \left[\rho(\theta-\hat{\theta}) - (\rho-\beta)(\theta-\tilde{\theta})\right]dF(\theta) + \rho\hat{\theta} - \rho\underline{\theta}.$$

Integration by parts gives the following identity:

$$\rho\int_{\underline{\theta}}^{\hat{\theta}}[\theta-\hat{\theta}]dF(\theta) + \rho\hat{\theta} - \rho\underline{\theta} = \int_{\underline{\theta}}^{\hat{\theta}}\left[\rho\frac{1-F(\theta)}{f(\theta)}\right]dF(\theta),$$

which, after substitution in the right-hand side of the previous inequality, gives (46). □

The second step uses the equivalence in Claim 6 and the assumption that inequality (10) holds with equality for some $\tilde{\theta} < \bar{\theta}$ and $\hat{\theta} = \tilde{\theta}$ to get:

$$\int_{\underline{\theta}}^{\tilde{\theta}}\left[\rho\frac{1-F(\theta)}{f(\theta)} - \rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - (\rho-\beta)(\theta-\tilde{\theta})\right]dF(\theta) = -\rho\underline{\theta} + \rho\left(\frac{\rho}{1-\rho}\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - \frac{\rho-\beta}{1-\rho}\tilde{\theta}\right). \tag{47}$$

Substituting (47) in the right-hand side of (46) gives

$$\int_{\underline{\theta}}^{\hat{\theta}}\left[\rho\frac{1-F(\theta)}{f(\theta)} - \rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - (\rho-\beta)(\theta-\tilde{\theta})\right]dF(\theta) \geq \int_{\underline{\theta}}^{\tilde{\theta}}\left[\rho\frac{1-F(\theta)}{f(\theta)} - \rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - (\rho-\beta)(\theta-\tilde{\theta})\right]dF(\theta).$$

Subtracting the left-hand side on both sides gives

$$0 \geq \int_{\hat{\theta}}^{\tilde{\theta}}\left[\rho\frac{1-F(\theta)}{f(\theta)} - \rho\frac{1-F(\tilde{\theta})}{f(\tilde{\theta})} - (\rho-\beta)(\theta-\tilde{\theta})\right]dF(\theta),$$

which gives (12) after rearranging terms. □

OA.9 Proof of Lemma 9 on the global optimality conditions

Proof of Lemma 9. Lemma A.2 in Amador, Werning, and Angeletos (2006) implies that if the Lagrangian with Lagrange multipliers Λ^* is concave, then the equality and inequality conditions in terms of Gateaux derivatives imply that the Lagrangian is maximized at $u^*, \underline{t}^*$:

$$\mathcal{L}(u^*, \underline{t}^*|\Lambda^*) \geq \mathcal{L}(u, \underline{t}|\Lambda^*) \quad \text{for all } (u,t) \in \Phi.$$

To show the concavity of the Lagrangian with Lagrange multipliers Λ^*, it is convenient to spell out the Lagrangian (14) and factorize the non-linear terms as follows:

$$\begin{aligned}\mathcal{L}(u,\underline{t}|\Lambda^*) \equiv &\int_\Theta \left[u(\theta)\ \left((1-F(\theta)) - \theta\tfrac{\rho-\beta}{\rho}f(\theta)\right)\right] d\theta \\ &-\int_\Theta [u(\theta)(1-\Lambda^*(\theta))]\, d\theta + (\underline{\theta}\ u(\underline{\theta}) - \underline{t})\Lambda^*(\underline{\theta}) + \int_\Theta [\theta\ u(\theta)]\, d\Lambda^*(\theta) \\ &+ \int_\Theta \left[\beta W(T-U^{-1}(u(\theta)))\right] d\left(\tfrac{1-\rho}{\rho}F(\theta) + \Lambda^*(\theta)\right) \\ &+ \beta W(T-U^{-1}(u(\underline{\theta})))\Lambda^*(\underline{\theta}).\end{aligned}$$

The integrands for the integrals in the first two lines are linear in u. For the terms in the remaining two lines, to show that the function $u \mapsto W(T - U^{-1}(u)))$ is concave, note that the utility index U is strictly increasing and concave so its inverse U^{-1} is strictly increasing and convex ($U^{-1'}(U(x)) = 1/U'(x)$ and $U^{-1''}(U(x)) = -U^{-1'}(U(x))U''(x)/U'(x)^2$). Since W is increasing and concave and $-U^{-1}$ is concave, the composition $u \mapsto W(T - U^{-1}(u))$ is concave. A sufficient condition for the Lagrangian to be concave is that the function $\frac{1-\rho}{\rho}F(\theta) + \Lambda^*(\theta)$ be non-decreasing, which is the case since $0 \le \rho \le 1$ and F and Λ^* are both non-decreasing.

It remains to show that the maximizer of a concave Lagrangian at a valid Lagrange multiplier is the solution to the constrained optimization problem of interest. This is precisely what the global theory of constrained optimization does for us.

The following notation maps the environment studied in this paper to Theorem 1 in Amador and Bagwell (2013) p.1575: $X = \{u, \underline{t} \mid u : \Theta \mapsto \mathbb{R},\ \underline{t} \in \mathbb{R}\}$, $Z = \{z \mid z : \Theta \mapsto \mathbb{R}\}$ with norm $||z||= \sup_{\theta\in\Theta}|z(\theta)|$, $\Omega = \{(u, \underline{t}) \in X|\ u \text{ is non-decreasing},\ \underline{t} \ge 0\}$, and $P = \{z \in Z \mid z(\theta) \ge 0 \text{ for } \theta \in \Theta\}$. The objective is a functional $f : \Omega \mapsto \mathbb{R}$ defined as follows:

$$\begin{aligned}f(u,\underline{t}) = &-\int_\Theta \left[u(\theta)\rho\tfrac{1-F(\theta)}{f(\theta)} - \beta\nu(U^{-1}(u(\theta)))\right] dF(\theta) \\ &- (\rho-\beta)\int_\Theta \left[\theta u(\theta) + \beta W(T-U^{-1}(u(\theta)))\right] dF(\theta).\end{aligned}$$

The constraints on limited transfers are defined as follows: $G : \Omega \mapsto Z$,

$$G(u,\underline{t}) = -\left(\underline{t} + \theta\ u(\theta) + \beta W(T-U^{-1}(u(\theta))) - \underline{\theta}\ u(\underline{\theta}) - \beta W(T-U^{-1}(u(\underline{\theta}))) - \int_{\underline{\theta}}^{\theta} u(\tilde{\theta})\ d\tilde{\theta}\right),$$

and their contributions to the Lagrangian are given by $T : Z \mapsto \mathbb{R}$,

$$T(z) = \int_\Theta z(\theta) d\Lambda^*(\theta),$$

which satisfies $T(z) \geq 0$ for all $z \in P$ since Λ^* is non-decreasing. Since $\mathcal{L}(u|\Lambda^*) = -f(u) - T(G(u))$, Theorem 1 from Amador and Bagwell (2013) implies that $(u^*, \underline{t}^*)$ solves

$$\min_{(u,\underline{t})\in\Omega} \{f(u,\underline{t})|-G(u,\underline{t}) \in P\}.$$

Inverting the above mapping from the environment of this paper to Theorem 1 in Amador and Bagwell (2013) and using $t(\cdot)$ defined in (8) as a function of $g(\cdot)$, the allocation $g^* = U^{-1}(u^*)$ and the initial level $\underline{t}^*$ solve the optimization problem:

$$\max_{g\in\Omega,\ \underline{t}\geq 0} \int_\Theta [\theta U(g(\theta)) + \beta W(T - g(\theta)) - \nu(g(\theta)) - \rho t(\theta)]\, dF(\theta)$$

s.t. for all $\theta \in \Theta$:

$$\beta t(\theta) = \beta\underline{t} + \theta U(g(\theta)) + \beta W(T - g(\theta)) - \underline{\theta} U(g(\underline{\theta})) - \beta W(T - g(\underline{\theta})) - \int_{\underline{\theta}}^{\theta} U(g(\tilde{\theta}))d\tilde{\theta}$$

g is non-decreasing

$t(\theta) \geq 0.$

The characterization of incentive compatible allocations in Lemma 3 implies that (g^*, t^*) in which

$$\beta t^*(\theta) \equiv \beta\underline{t}^* + \theta\ U(g^*(\theta)) + \beta W(T - g^*(\theta)) - \underline{\theta}\ U(g^*(\underline{\theta})) - \beta W(T - g^*(\underline{\theta})) - \int_{\underline{\theta}}^{\theta} U(g^*(\tilde{\theta}))\ d\tilde{\theta}$$

solve the mechanism design problem (5). □

OA.10 Proof that Assumption H implies a tight cap

This section contains the formal statement and the proof of an observation made in Section 4.

Claim 7. *Assumption H implies that the tight cap allocation lies below the discretionary allocation; that is, $g_c \leq g_d(\theta)$ for $\theta \in \Theta$.*

Proof of Claim 7. Integrate the inequality in Assumption H to get

$$\int_{\underline{\theta}}^{\bar{\theta}} (1 - F(\theta))d\theta \leq (1 - \beta) \int_{\underline{\theta}}^{\bar{\theta}} \theta f(\theta)d\theta.$$

Integrating $1 - F$ by parts gives

$$\int_{\underline{\theta}}^{\bar{\theta}} \theta f(\theta)d\theta - \underline{\theta} + \lim_{\theta\to\bar{\theta}} \theta(1 - F(\theta)) \leq (1 - \beta) \int_{\underline{\theta}}^{\bar{\theta}} \theta f(\theta)d\theta.$$

Since the expectation of θ is finite and the distribution of θ has a density, $\lim_{\theta\to\bar{\theta}} \theta(1-F(\theta)) = 0$. The inequality reduces to $\beta\ \mathbb{E}[\theta] \leq \underline{\theta}$, which is equivalent to $g_c \leq g_d(\underline{\theta})$. The discretionary allocation is non-decreasing and hence $g_c \leq g_d(\underline{\theta}) \leq g_d(\theta)$. □

OB Shocks to government revenues

This section uses Section 5.4 in Amador, Werning, and Angeletos (2006) to derive shocks to fiscal needs from shocks to government revenues. I then show that the mapping preserves the log-convexity of the tail. Although log-concavity need not be preserved, an example illustrates that it may be preserved. For this section, assume that the utility index exhibits constant absolute risk aversion: $U(g) = e^{-\alpha g}$. The objective of the government does not feature a priori shocks to fiscal needs; it reads $U(g) + \beta W(x)$. The budget constraint is

$$g + x + \tau_f(g) = T + \tilde{T},$$

where $\tilde{T}$ denotes the idiosyncratic shocks to government revenues. Suppose that $\tilde{T} \sim F_{\tilde{T}}$. Substituting the budget constraint in the objective function maps shocks to fiscal revenues into shocks to fiscal needs as follows: $\theta = e^{-\alpha \tilde{T}}$. The distribution of shocks to fiscal needs θ is related to the distribution of shocks to revenues $\tilde{T}$ as follows: $F(x) = 1 - F_{\tilde{T}}\left(-\ln(x)/\alpha\right)$.

The distribution of fiscal needs is a transformed and mirrored version of the distribution of government revenues. Convex (resp. concave) monotonic transformations of a random variable preserve the log-convexity (resp. log-concavity) of the tail. Mirroring preserves the log-concavity/convexity.

Proposition (Log-convex tails)**.** *If $F_{\tilde{T}}$ is log-convex, then $1 - F$ is also log-convex.*

Proof. Let $z(x) = -\frac{1}{\alpha}\ln(x)$. Since z is monotonic and convex, the log-convexity of $F_{\tilde{T}}$ implies that $F_{\tilde{T}} \circ z$ is also log-convex. Since $1 - F = F_{\tilde{T}} \circ z$, the tail of the distribution of θ is also log-convex. □

The analog correspondence for log-concave tails need not hold. The convex function for the change of variable counteracts the concavity of the log of the tail. The following example illustrates that the result holds for some class of distributions.

Example (Log-concave tails)**.** *Fiscal need θ is uniformly distributed on $[0, 1]$ if and only if the shock to fiscal revenue $\tilde{T}$ is exponentially distributed with $F(x) = 1 - \exp\left(-x/\alpha\right)$ on $[0, \infty)$. The uniform and exponential distributions both have a log-concave tail.*

Assumption L pertains to the elasticity of the density. The densities are related as follows: $f(x) = \frac{1}{\alpha x} f_{\tilde{T}}\left(-\ln(x)/\alpha\right)$. A lower bound on the elasticity of f corresponds to an upper bound on the semi-elasticity of $f_{\tilde{T}}$; the correspondence follows from $\frac{xf'(x)}{f(x)} = -\left(1 + \frac{f'_{\tilde{T}}(-\ln(x)/\alpha)}{f_{\tilde{T}}(-\ln(x)/\alpha)}\right)$.

OC Examples

OC.1 Figure 1

For non-financial sanctions (i.e., $\rho = 1$), the optimal fiscal rule is a cap on spending as depicted in the left and right panels of Figure 1. First, consider the left panel. To show that implementing the *discretion and off-equilibrium* sanctions allocation is optimal, it suffices to check that Assumption L holds for $\theta \leq \theta_p$ so that Proposition 1 applies. The elasticity of the density of the exponential distribution is $-\lambda\theta$, which is smaller than $-\frac{2-\beta}{1-\beta}$ for $\theta \leq 2$ since $1 - \beta = 0.2$. The degree of present bias is low and $\underline{\theta} = 1 < \theta_p < 2$. The threshold for the optimal cap is interior because the tail of the exponential distribution is log-concave. In contrast, if shocks to fiscal needs were distributed according to a distribution with a strictly log-convex tail, the cap would either be infinitely loose (i.e., no cap), or it would be tight.

Second, consider the right panel of Figure 1. To show that implementing the *tight cap* allocation is optimal, it suffices to check that Assumption H holds so that Proposition 4 applies. For the exponential distribution, $\frac{1-F(\theta)}{\theta f(\theta)} = \frac{1}{\lambda\theta} \leq \frac{1}{\lambda\underline{\theta}} < 0.6 = 1 - \beta$, so Assumption H holds.

OC.2 Figure 2

For $\rho = \beta < 1$, the dashed line in the panels in Figure 2 depicts the allocation implemented by the candidate marginal sanction schedule.

Consider the top left panel of Figure 2. To show that a fiscal rule that implements the *discretion and off-equilibrium sanctions* allocation is optimal, it suffices to show that Proposition 1 applies. Note that $1 - F$ is log-concave, and Assumption L holds for $\theta \leq \frac{1}{\lambda}\frac{1}{1-\beta}$ and $\theta_p < 1.6\bar{6} = \frac{1}{\lambda}\frac{1}{1-\beta}$. Consider the bottom left panel. Proposition 2 implies that implementing the *discretion and on-equilibrium sanctions* allocation is optimal since $1 - F_a$ is log-convex (so $\theta_{np} = \bar{\theta}$) and Assumption L holds for $\theta \leq \theta_n$. Consider the middle left panel. Proposition 2 applies because the tail of the truncated F_a distribution becomes log-concave after a point that lies above θ_{np} and Assumption L holds for $\theta \leq \theta_{n'}$.

Consider the bottom right panel. To show that implementing the *exemption and on-equilibrium sanctions* allocation is optimal, it suffices to show that part 2) of Proposition 3 applies. Since the tail of the Pareto distribution is log-convex, $\theta_{np} = \bar{\theta}$. To check that Assumption I holds for $\theta \leq \theta_{np}$, note that for the Pareto distribution, $\beta\frac{1-F(\theta)}{\theta f(\theta)} = \frac{\beta}{\gamma} = \frac{0.7}{4} < 0.3 = 1 - \beta$. Consider the middle right panel. The only change from the bottom right panel is that the distribution is truncated. Because a truncated Pareto distribution is log-convex up to a point after which it is log-concave, $\theta_{np} < \bar{\theta}$ in the middle left panel. Last, consider the top right panel. Part 1) of Proposition 3 applies since the exponential distribution has a log-concave tail and Assumption I holds for $\theta = \underline{\theta}$.

Two main lessons emerge from Figure 2. First, the three panels on the left illustrate the role of the tail of the distribution of shocks in determining the optimal severity of sanctions. In all three panels, the sanction schedule imposes increasing discipline. In the middle left panel, for instance, there is no sanction below a threshold, sanctions are on-equilibrium for an intermediate range of spending levels, and sanctions are off-equilibrium above the cap. The bunching caused by off-equilibrium sanctions can be understood as "ironing." In contrast the bunching caused by the exemption cannot be understood as "ironing" because g_n is increasing below the exemption threshold.