The Optimal Allocation of Policy-Making Between Executive Agencies and the Legislature *

Alessandro Dovis
University of Pennsylvania
and NBER
adovis@upenn.edu

Rishabh Kirpalani
University of Wisconsin-Madison
rishabh.kirpalani@wisc.edu

Guillaume Sublet
University of Montreal
guillaume.sublet@umontreal.ca

July 2026

Abstract

When should society allocate policy-making to legislatures rather than executive agencies? We develop a model in which biased and polarized policymakers choose policy in response to shocks, and society designs both the mandate of an executive agency and the outside options that shape legislative bargaining. We show that the optimal set of legislative outside options contains no intervals when the bias is small, whereas the optimal delegation set for executive agencies is an interval under standard conditions. We then show that legislative bargaining is optimal when the average bias is small relative to polarization, whereas executive delegation is optimal when the bias is large. Legislative bargaining allows policy to respond to large shocks, but only when both factions agree, thereby limiting unilateral action by the faction in power. Delegation to executive agencies, by contrast, is valuable when the bias is sufficiently severe because a narrow mandate can sharply restrict policy choices. Delegation is thus valuable for constraining choices, not for expanding flexibility as is usually emphasized. Finally, we analyze credibility and show that legislative bargaining is always credible, while narrow executive mandates are credible when the bias arises from time inconsistency but not when it is exogenous.

First version: February 2024. This paper previously circulated as "On the Optimal Allocation of Policy-Making." We thank Mark Aguiar, Manuel Amador, Renee Bowen, Alberto Bisin, Hülya Eraslan, Marina Halac, Navin Kartik, Facundo Piguillem, Ali Shourideh, and Pierre Yared for useful comments.

1 Introduction

The allocation of policy-making authority is a fundamental issue in the design of political institutions. Institutions such as the legislature and executive agencies provide mechanisms, including supermajority requirements and mandates, that help align policymakers' incentives with those of society. The goal of this paper is to understand how policy-making should be allocated between the legislature and executive agencies. The conventional view is that delegation to an agency such as a central bank or an environmental agency like the EPA allows for fast and responsive policy-making; in contrast, legislative bargaining in the presence of political polarization can be slow and lead to political gridlock, even on desirable policies. However, a major downside of delegation is the lack of direct accountability and representation.[1]

Building on this institutional context, this paper studies the optimal allocation of decision-making authority in a setting with three key features: policymaker *bias* relative to societal preferences, shocks that make *flexibility* valuable, and political *polarization* between different societal factions that disagree on the optimal policy. Our main result is that delegating to the legislature is preferred if policymaker bias is small relative to the degree of polarization, while delegating to an executive agency is preferred if policymaker bias is large.

In our model, the two institutional settings differ in whether ex post bargaining is available. When the bias is small, this possibility allows policy to respond to severe shocks, and because it requires the agreement of both factions, it prevents the faction in power from implementing policies too distant from society's preferences. In contrast, when the bias is large, society would like to constrain the choices of policymakers, which can be achieved by delegating to an agency with a narrow mandate that prevents ex post renegotiation. This contrasts with the aforementioned conventional view that society should delegate policy-making to executive agencies to increase flexibility. We also consider whether these institutional arrangements are credible, in the sense that the legislature cannot agree to change them. We show that in our baseline model, where the bias is exogenous, these narrow mandates are not credible, while they are credible when the bias arises from a time-inconsistency problem.

We consider a model where society must decide on a policy as a function of an aggregate shock. The population is divided into two factions who differ in their preferred policies, capturing political polarization. Each faction is represented by a policymaker whose preferences match those of the faction, except for a common constant bias term. This bias can capture either differences in preferences between factions and their representatives or time-inconsistency issues. Factions are in power with an exogenous probability. Society's

[1] See, for example, Tucker (2019).

welfare is a weighted average of the factions' utilities. We compare two institutions, that is two protocols, that specify how policy is chosen. The first, *executive delegation*, grants the policymaker representing the faction in power (the executive) the authority to unilaterally choose a policy from a pre-specified set, which we term the delegation set. The second, *legislative bargaining*, allows the policymaker in power to unilaterally choose a policy from a given delegation set, but the policymaker can also implement a policy outside this set if the other faction agrees.[2] This ability to renegotiate ex post in legislative bargaining is the critical feature that distinguishes the two institutions.

Our analysis proceeds in two parts that answer two different questions. The first part is normative and asks which of the two institutions is optimal from society's perspective. We start by characterizing the optimal outcomes under both institutions. For executive delegation, we provide a set of necessary and sufficient conditions under which the optimal delegation set is an interval that results in discretion for moderate shocks but imposes a cap and a floor for extreme shocks. Within this set, policy varies both because of the underlying state and because of the identity of the faction in power. The latter, which we term *political risk*, is undesirable from society's perspective. The size of the delegation set is chosen to optimally trade off this desire for flexibility against the costs associated with policymaker bias and political risk. For example, if there is no bias and no political risk, it is optimal to grant full flexibility to the policymaker. By contrast, if either policymaker bias or political risk gets large enough, it is optimal to delegate only a single point to the executive.

Under legislative bargaining, the optimal set of outside options contains no intervals when the policymakers' bias is small.[3] Thus, the faction in charge has almost no unilateral discretion. Limiting unilateral discretion under legislative bargaining is beneficial because it mitigates political risk while still allowing policy to respond to large shocks when policymakers can reach an agreement. We show that this discreteness result is independent of the distribution of shocks.

Our main result provides conditions under which each of these two institutional settings is preferred from society's perspective. Fixing a level of polarization, we find that legislative bargaining is preferred if policymaker bias is small enough, while executive delegation is preferred if the bias is large enough. To understand this result, first note that for a fixed delegation set, legislative bargaining gives policymakers more discretion to respond to shocks. Clearly, factions will agree only to a policy that lies between the ideal policies of the two factions. If the policymaker bias is small, then this extra discretion through renegotiation is valuable to society. Thus, legislative bargaining is preferred when

[2] We can think of the policymaker in power as an executive who makes a policy proposal that must go through legislative approval.

[3] This set can capture state-contingent levels of mandatory spending as in Bowen et al. (2017).

the bias is small. However, when the bias is large, delegation is preferred to legislative bargaining. The reason for this is that while bargaining increases policy flexibility, it also limits society's ability to correct the policymaker bias. If the bias is large, this increased flexibility is detrimental to society and thus delegation to an executive is preferred.

Another way to limit policy flexibility is to restrict the set of policies over which factions can bargain. We show that, when the bias is large, a hybrid institution that restricts the ability of factions to renegotiate under legislative bargaining attains the same welfare as executive delegation. For a large bias, the key force is not whether policy-making authority is formally allocated to the executive or the legislature, but whether society can limit ex post renegotiation.

The second part of our analysis is positive: we ask whether the institutions chosen by society are credible. We say an arrangement is credible if the legislature cannot agree to change it. This matters because in practice the mandates of executive agencies are set by legislatures, not fixed at a constitutional stage, so a legislature that imposes a narrow mandate can also agree to relax it. When the bias is large and exogenous, narrow mandates are not credible: the gains from a binding mandate accrue only to society, which plays no role in the bargaining process. Thus, both factions renegotiate ex post toward their own preferences. The result changes when the bias arises from time inconsistency, which we model in a Barro–Gordon (Barro and Gordon, 1983) setting: there, policymakers value the constraint themselves, so a narrow mandate is credible when the problem is severe. Legislative bargaining, by contrast, is always credible.

The key takeaway from our analysis is that allocating policy-making to executive agencies is desirable when the policymaker bias is large not because of these agencies' ability to nimbly respond to shocks, but rather because of society's ability to impose judicially enforced mandates on these agencies. As we discuss in the paper, this finding can help interpret the recent discussions of the Chevron doctrine (U.S. Supreme Court, 1984) and its overturning. In contrast, delegating policy-making to the legislature is desirable when polarization is large relative to the bias, because it allows society to respond to large shocks while simultaneously limiting political risk. In this sense, inaction by the legislature can be a feature and not a problem. This logic provides a rationale for institutional arrangements that require broad agreement before policy can change, including supermajority rules, practices such as the filibuster in the U.S. Senate, and electoral systems that do not grant a majority premium.

Our results can help inform decisions about which types of policies should be allocated to an executive agency versus assigned to a legislature. In our model, policies that suffer from severe time-inconsistency problems relative to the degree of polarization should be delegated to an executive agency with a narrow mandate. This prediction is consistent with the delegation of monetary policy, arguably subject to time-inconsistency problems,

to central banks with inflation targeting. Our model also suggests that policies for which polarization is significant relative to the policymaker bias should be delegated to the legislature. This prediction is consistent with two features of redistributive policy, which is arguably highly polarized: it is chosen within the legislature, and it responds to large shocks but not to small ones. For example, transfer programs do not vary much for small business cycle shocks, but there is often bipartisan agreement to expand these programs after large shocks such as the 2008 recession and the 2020 Covid pandemic.

Related Literature

Our paper builds on two literatures: delegation to biased policymakers and legislative bargaining. Its main contribution is to compare these institutional settings and identify when each is preferred.

Our modeling of executive delegation builds on a large literature studying the trade-off between discipline and discretion in the delegation of policy-making authority, including applications to monetary policy (Athey et al., 2005) and fiscal policy (Halac and Yared, 2014, 2020, 2022a; Sublet, 2023). This literature provides conditions under which the optimal delegation set is an interval (Amador et al., 2006; Alonso and Matouschek, 2008; Amador and Bagwell, 2013; Amador et al., 2026). Sublet (2023) studies when caps are optimal in an environment with money burning. We extend this logic to executive delegation with policymaker heterogeneity, which introduces political risk: policy depends not only on the state but also on the identity of the faction in power.[4]

Our modeling of legislative bargaining builds on Baron and Ferejohn (1989) and subsequent work. This literature highlights the role of the status quo, or outside option, in shaping bargaining outcomes.[5] We contribute by studying the optimal design of the set of outside options. In this respect, our analysis is close to Piguillem and Riboni (2021), who study how fiscal rules shape equilibrium policy in a legislative bargaining environment. Harstad (2010) also studies policymaker bias in legislative bargaining and the role of supermajority requirements; our focus is instead on the optimal design of policy sets and the allocation of authority between executive delegation and legislative bargaining.

Our result that the optimal set of outside options contains no intervals follows from a logic different from that in the delegation literature. The set contains no intervals

[4]Our model with heterogeneity among policymakers (two agents) maps to a multidimensional screening problem without transfers (see, for example, Laffont et al. (1987), Rochet and Choné (1998), Moser and Olea de Souza e Silva (2019), and Boerma et al. (2022) for problems with transfers and Amador et al. (2006) for a problem without transfers).

[5]Papers in this literature study bargaining with an endogenous status quo, where the outcome of one negotiation becomes the default for future negotiations; see, for example, Baron (1996), Kalandrakis (2004), Diermeier and Fong (2011), Bowen et al. (2014), Dziuda and Loeper (2016), Piguillem and Riboni (2015), and Ali et al. (2023). Bowen et al. (2017) allow for a separation between current policy choices and future outside options.

because this limits the bargaining power of the faction in charge and moderates bargaining outcomes. This distinguishes our result from the examples of discrete delegation sets in single-agent problems in Alonso and Matouschek (2008) and Amador et al. (2026). It also differs from veto-based delegation models such as Mylovanov (2008) and Kartik et al. (2021). In Kartik et al. (2021), veto bargaining can be represented as a delegation problem in which a principal offers a set of actions and an informed agent can either choose from that set or take an exogenous outside option. In our model, society instead designs the outside-option set itself in order to shape bargaining between two biased agents. The default set is therefore not a menu for screening one informed agent, but an institutional device for affecting bargaining power and moderating policy outcomes.

Our paper also shares some conclusions with the literature on monetary delegation and time inconsistency. Rogoff (1985) argues that delegating monetary policy to a central banker who places greater weight on inflation stabilization can alleviate the time-inconsistency problem. In that case, delegation is valuable because the policymaker has preferences that generate better outcomes from society's perspective. By contrast, we assume that policymaker preferences are the same under the legislature and the executive, and show how mandates themselves can be designed to alleviate time-inconsistency problems.

Our paper is also related to other studies of the optimal design of institutions. We share our motivation with Alesina and Tabellini (2007, 2008), who highlight normative criteria for delegating policy-making to politicians versus bureaucrats. Unlike these authors, we assume that policymakers have identical preferences within the legislature and executive but that the institutions differ in how policy is chosen. Maskin and Tirole (2004) and Persson et al. (1997) also compare institutions for disciplining policymakers whose preferences diverge from those of voters, in both cases through electoral accountability. We abstract from electoral accountability and compare two institutions whose biased policymakers are disciplined through the optimal design of the executive agency's mandate or the set of outside options. Additionally, our paper has some similarities with Aghion et al. (2004), who study a constitutional design problem to understand how much unilateral power should be granted to factions in power, modeled as the size of the supermajority needed to pass a policy reform. Their problem is similar to our legislative bargaining problem in which society chooses the set of outside options for the faction in power. However, we compare policy outcomes under institutions that involve joint decision-making with those under an institution in which a single executive chooses policy.

The problem of separation of powers between the executive and legislature has been widely studied in the political science and legal studies literatures. Epstein and O'Halloran (1999) provide an overview of the literature on the incentives of the legislature to delegate

policy-making.[6] Callander and Krehbiel (2014) show that under supermajoritarianism, the legislature may delegate to an executive agency run by another biased agent to break political gridlock. We model political risk in executive agencies and study the optimal institution from society's perspective.

The rest of the paper is organized as follows. Section 2 presents the main model. Sections 3 and 4 study the optimal outcome under executive delegation and legislative bargaining, respectively. Section 5 delivers the paper's main result and establishes conditions under which each institutional setting is preferable. Section 6 discusses the implications of our main results. Section 7 shows that our main results extend to a Barro–Gordon model. Section 8 studies the credibility of the institutional setting chosen at the constitutional stage. Finally, Section 9 concludes the paper.

2 Model

Consider a static model in which society must decide on a policy $\pi \in \mathbb{R}$ as a function of an aggregate shock $z \in Z \equiv [\underline{z}, \overline{z}]$. The distribution of shocks is F with density f. The population is divided into two *factions*, or parties, indexed by a preference parameter $\theta \in \{\theta_H, \theta_L\}$. We refer to $\Delta = \theta_H - \theta_L$ as the *polarization* between factions. The population share of faction θ_i is α_i for $i \in \{H, L\}$. Citizens of faction θ have preferences[7]

$$w(\pi, z, \theta) = (z + \theta)\pi - \frac{1}{2}\pi^2.$$

We evaluate societal welfare according to a utilitarian criterion:

$$v(\pi, z) = \sum_i \alpha_i w(\pi, z, \theta_i). \tag{1}$$

Faction θ is represented by a policymaker with preferences

$$u(\pi, z, \theta) = w(\pi, z, \theta) + \nu\pi.$$

The bias ν can capture the difference in preferences between the factions and their representative politicians. For example, politicians may be biased relative to society because they are prone to being captured by special interest groups or to having empire-building motives. In Section 7 we show that the bias can capture, in reduced form, time-inconsistency problems that arise when a policymaker lacks the ability to commit to a policy plan.[8]

[6] See also Aranson et al. (1982), McCubbins et al. (1987), Rao (2015), and Volden (2002).

[7] Note that these preferences can be alternatively represented as $-\frac{1}{2}(\pi - \theta - z)^2$ that only differs from w by a constant term. The extra constant term does not matter for choices.

[8] While we focus on a constant bias in this section, our results extend to environments with a non-constant

Notice that policymaker preferences for θ_H can be written as $u(\pi, z, \theta_H) = v(\pi, z) + (\nu + \alpha_L \Delta)\pi$ which implies that the true or effective bias of the policymaker also depends on the degree of polarization, Δ. Similarly for θ_L, $u(\pi, z, \theta_L) = v(\pi, z) + (\nu - \alpha_H \Delta)\pi$. In other words, our economy is equivalent to one in which there are two types of policymakers with different degrees of bias relative to society.

Type θ policymaker's preferred policy in response to a shock z is denoted by $\pi_f(z, \theta) = z + \theta + \nu$. The preferred policy for society is $\pi^*(z) = z + \bar{\theta}$ with $\bar{\theta} = \alpha_L \theta_L + \alpha_H \theta_H$. Both policymakers agree that the policy must be increasing in the aggregate state z. However, policymaker θ_H prefers a higher policy than does policymaker θ_L for all z. The distance between these preferred policies is the degree of polarization Δ. If $\nu \in (-\alpha_L \Delta, \alpha_H \Delta)$, society's preferred policy is a convex combination of the policymakers' preferred policies. Henceforth, we refer to a bias in this range as *small*. In contrast, society's preferred policy is higher than $\pi_f(z, \theta_H)$ if the bias is sufficiently negative and lower than $\pi_f(z, \theta_L)$ if it is sufficiently positive.

It is useful to define as a benchmark the *best constant policy* that does not vary with the shock z.[9] This policy solves

$$\bar{\pi}^* = \arg\max_{\pi} \int v(\pi, z) f(z) dz,$$

and it satisfies $\bar{\pi}^* = \pi^*(E[z]) = E[z] + \bar{\theta}$.

Two institutional settings We next consider the problem of how society should delegate and structure the policy-making process. We consider two institutional settings that differ in how policies are determined.

First, we consider *delegating to executive agencies (DE)*. In DE, society delegates policy-making authority to an executive representing the faction in power. We assume that a faction is in power with probability equal to its population share α_i.[10] The executive chooses any policy within a delegation set D. Without polarization, $\Delta = 0$, our model nests the standard model studied in the literature on delegation without money burning.

Second, we consider *delegating to a legislature (LB)*. In this case, the two policymakers can *bargain* over the chosen policy, with the outside option being a policy chosen from a set D by the policymaker in charge. In other words, the policymaker in power can unilaterally choose a policy from a prescribed set D or choose another policy outside the set D if the other policymaker agrees. This institutional setup is a simple version of legislative

bias, as in Amador and Bagwell (2013).

[9] Athey et al. (2005) refer to this policy as the expected Ramsey policy.

[10] The assumption that the probability a faction is in power equals its share in the population is made for simplicity.

bargaining that has been studied in the political economy literature. Typically in this literature the executive (or agenda setter) can unilaterally choose one policy only—D is a singleton. Here we study the optimal design of the set D.

While there are, of course, other possible institutional designs, we argue that these two capture essential trade-offs inherent in the allocation of policy-making power. Delegation to an executive represents the transfer of decision-making authority to a single agency, with the aim of promoting flexibility and responsiveness to changing circumstances. The head of the agency is appointed by the faction in power at the time and can choose any policy within the set given to the agency, its mandate. For example, this institutional setting captures the case of an executive agency, like the EPA or a central bank, that has freedom to choose a policy within its mandate; the faction in power appoints an agency head who shares its preferences.[11]

Legislative bargaining, on the other hand, keeps authority in the hands of elected representatives, ensuring a degree of democratic accountability and control. The faction in charge has some flexibility to choose among a given set of options (e.g., mandatory spending) but can choose outside the set if the opposition agrees (e.g., discretionary spending). In Section 5, we also study a hybrid institution that nests both DE and LB.

3 Delegation to executive agencies

We characterize the optimal delegation to executive agencies. We provide necessary and sufficient conditions for the optimal delegation set to be an interval with a cap and a floor. In contrast to the standard delegation environment, giving policymakers the flexibility to respond to shocks generates political risk, in that the chosen policy depends on the policymaker in charge, which is undesirable for society. Consequently, the optimal choice of the cap or floor takes this political risk into account. In fact, we show that for high levels of bias or high levels of polarization it is optimal to leave no discretion to policymakers, and the best outcome is the best constant policy defined in the previous section.

Formally, we can write society's problem of setting limits on executive discretion as a mechanism design problem without transfers, where society chooses a set of feasible policies D and the policy chosen by the executive of faction θ in state z must satisfy the incentive compatibility constraint

$$\pi(z,\theta) \in \arg\max_{\pi \in D} u(\pi, z, \theta). \tag{2}$$

[11] Central banks differ in the degree of influence that their governor has in determining monetary policy. Large boards and mandates for board members that span different administrations are ways to reduce the link between the executive and central banks.

This constraint highlights that under this setting the policymaker in power unilaterally chooses the policy independently of the other faction. Society chooses a set D to solve

$$\max_{D\subseteq\mathbb{R},\pi(z,\theta)} \int_{\underline{z}}^{\bar{z}} \sum_i \alpha_i v\left(\pi\left(z,\theta_i\right),z\right) f(z)dz \tag{3}$$

subject to (2).[12]

The problem (3) can be equivalently written as

$$\max_{\pi(z,\theta)} \int_{\underline{z}}^{\bar{z}} \sum_i \alpha_i v\left(\pi\left(z,\theta_i\right),z\right) f(z)dz \tag{4}$$

subject to

$$u(\pi(z,\theta),z,\theta) \geq u(\pi(\hat{z},\hat{\theta}),z,\theta) \quad \text{for } z,\hat{z}\in Z,\, \theta,\hat{\theta}\in\{\theta_L,\theta_H\}. \tag{5}$$

The main proposition in this section provides necessary and sufficient conditions on the distribution of z such that the optimal delegation set is an interval with a potentially binding cap *and* floor. To show that such an interval is optimal, we first consider a more constrained Ramsey problem in which society is restricted to choosing an interval $[\pi_l,\pi_h]$. This problem is

$$\max_{\pi_l,\pi_h} \int_{z} \sum_i \alpha_i v\left(\pi\left(z,\theta_i\right),z\right) f(z)dz \tag{6}$$

subject to

$$\pi(z,\theta)\in\arg\max_{\pi\in[\pi_l,\pi_h]} u(\pi,z,\theta), \quad \forall\theta.$$

The solution to this problem can be parametrized by two thresholds, z_l and z_h, such that $\pi_l=\pi_f\left(z_l,\theta_L\right)$ and $\pi_h=\pi_f\left(z_h,\theta_L\right)$. If the solution to (6) is the best constant policy, then $z_l=z_h$ and $\pi_f\left(z_l,\theta_L\right)=E[z]+\sum_i\alpha_i\theta_i$.[13] If $z_l<z_h$ instead, then either the cap is not binding (i.e., $z_h=\bar{z}+\Delta$), or the cap is binding and the threshold z_h is characterized in the next result.

Lemma 1 (Ramsey: optimal cap)**.** *Define the distribution $K(z)\equiv\alpha_L F(z)+\alpha_H F(z-\Delta)$, with $\kappa(z)$ being the associated density with support $[\underline{z},\overline{z}+\Delta]$, and E_κ the expectation with respect to this distribution. Let $d(z)=F(z)-F(z-\Delta)$. If $\Delta\alpha_L+\nu>0$ and $\alpha_H>0$, then the optimal*

[12]The restriction to deterministic mechanisms is without loss of generality for the results on delegation to executives. Given the assumption on preferences, allowing for stochastic mechanisms is equivalent to allowing for money burning. Sublet (2023) shows that money burning is suboptimal under Assumption 1. See Kartik et al. (2021) for a more general treatment of stochastic mechanisms.

[13]Note that the implied z_l and z_h may be outside the support of the interval $[\underline{z},\overline{z}+\Delta]$. For instance, $z_l\leq\underline{z}$ if the bias is large in the sense that $\pi_f\left(\underline{z},\theta_L\right)>E[z]+\sum_i\alpha_i\theta_i$. This is inconsequential for the solution of the Ramsey problem, which is defined over the policy space π. This alternative formulation in the shock space z allows us to state Assumption 1 in a manner similar to that in the delegation literature.

cap is binding. Moreover, an interior binding cap $z_h \in (z_l, \bar{z} + \Delta)$ *must satisfy*

$$\nu(1 - K(z_h)) + \Delta\alpha_L\alpha_H d(z_h) = (E_\kappa[z|z \geq z_h] - z_h)\,(1 - K(z_h))\,. \tag{7}$$

A similar result, stated in Appendix A, characterizes the optimal floor z_l.

Lemma 1 makes clear that the optimal cap trades off flexibility against the social cost of discretion. The term $(E_\kappa[z|z \geq z_h] - z_h)\,(1 - K(z_h))$ on the right-hand side captures the marginal cost of tightening the cap: a lower cap prevents the policymaker from adjusting policy efficiently in response to high realizations of the aggregate shock, and this cost is larger when the expected distance of the state above the cap is greater. The left-hand side captures the marginal benefit of tightening the cap. The first component, ν, is the policymaker-bias term: when $\nu > 0$, a tighter cap limits the extent to which the policymaker can exploit discretion to move policy away from the social optimum. The second component, $\Delta\alpha_L\alpha_H d(z_h)$, reflects political risk. Because the high-θ policymaker prefers systematically higher policies, a relaxation of the cap affects policy for lower realizations of z when that faction is in power. As a result, a looser cap not only increases flexibility, but also makes policy outcomes more sensitive to the identity of the policymaker. This political-risk cost is increasing in polarization, Δ, and is largest when both factions are likely to be in power, i.e., when $\alpha_L\alpha_H$ is large.[14] Thus, the optimal cap equates the marginal cost of limiting flexibility in high states with the marginal benefit of reducing both policymaker bias and political risk.

Next, we show that the solution to this constrained Ramsey problem is also a solution to the more general problem (4). To do so, we apply the tools developed by Amador et al. (2006) and Amador and Bagwell (2013) to derive sufficient conditions under which an interval is optimal, and we show that these conditions are also necessary.. These conditions are stated below.

Assumption 1. *Let* $K(z), d(z), z_l, z_h$ *be defined as before. We assume that*

1. $(\nu - \alpha_H\Delta)\,\alpha_L f(z) + (\nu + \alpha_L\Delta)\,\alpha_H f(z - \Delta) + K(z)$ *is non-decreasing for all* $z \in (z_l, z_h)$.
2. $\Delta\alpha_L\alpha_H d(\hat{z}) + \nu(1 - K(\hat{z})) \geq (E_\kappa[z|z \geq \hat{z}] - \hat{z})(1 - K(\hat{z}))$ *for all* $\hat{z} \in [z_h, \bar{z} + \Delta)$.
3. $\Delta\alpha_L\alpha_H d(\hat{z}) - \nu K(\hat{z}) \geq (\hat{z} - E_\kappa[z|z \leq \hat{z}])K(\hat{z})$ *for all* $\hat{z} \in (\underline{z}, z_l]$.

For example, Assumption 1 is satisfied if the distribution of z is uniform, as we show in the Appendix. We also verify numerically that this holds for the truncated normal distribution under the parametrization we use in our numerical example below.

Proposition 1. *The solution to* (3) *is an interval delegation set* $D = [\pi_l, \pi_h]$ *if and only if the cutoffs* π_l *and* π_h *are the solution to the Ramsey problem* (6) *and Assumption 1 is satisfied.*

[14] Note that this second term may incentivize a binding cap even if the common bias term $\nu < 0$.

In contrast to the standard delegation problem, the proof of this proposition is complicated by the fact that there is uncertainty about both the state z and the type of the policymaker θ. We show that the solution to our problem can be characterized as the solution to a simpler problem with only one dimension of uncertainty, $\theta + z$. This allows us to apply the tools developed by the delegation literature.

First, we show that the conditions in Assumption 1 are necessary. In particular, Condition 1 guarantees that granting the executive discretion between the Ramsey bounds is indeed optimal. If we were not to grant discretion in this interval and were instead to consider a small perturbation in which an infinitesimal subset of flexible actions is removed, this would imply that some types bunch above their preferred policy while others bunch below. The change in society's welfare associated with this perturbation is positive if Condition 1 is not satisfied.

Conditions 2 and 3 in Assumption 1 guarantee that it is suboptimal to add points outside the Ramsey bounds, which would result in a disconnected delegation set. In particular, Condition 2 rules out adding a point to the delegation set above the Ramsey cap. This condition has the same marginal costs and benefits interpretation provided for the optimality of the Ramsey cap. The term $\left(E_\kappa[z \mid z \geq \hat{z}] - \hat{z}\right)\left(1 - K\left(\hat{z}\right)\right)$ captures the expected flexibility gain from allowing policy to adjust in states above $\hat{z}$: this gain is larger when shocks in the relevant tail are, on average, far above the threshold. The term $\Delta\alpha_L\alpha_H d(\hat{z}) + \nu(1 - K(\hat{z}))$ captures the benefit of preserving bunching.

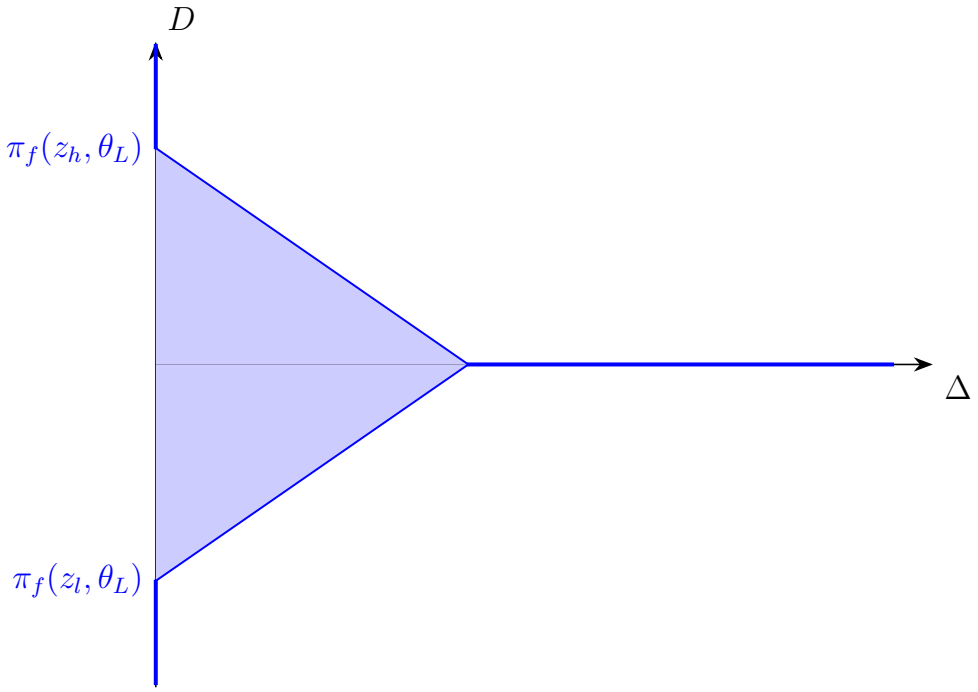
Finally, we use Lagrangian methods as in Amador and Bagwell (2013) to prove that the conditions are sufficient. First, we show that under Assumption 1 the Lagrangian is concave by appropriately constructing the function for the Lagrange multiplier on the incentive constraint associated with the Ramsey solution. Then, we show that Assumption 1 is sufficient for the solution to the Ramsey problem to maximize the Lagrangian.

Note that when $\Delta = 0$ (no polarization), the conditions in Assumption 1 are equivalent to those in Amador and Bagwell (2013) with a constant bias.[15]

The choice of the optimal interval highlights a trade-off between *political risk* and *flexibility* that is absent in the case in which there is certainty about the policymaker type. By political risk, we mean that the chosen policy varies depending on the type of the policymaker in charge. To understand this trade-off, suppose that $\nu = 0$. If $\Delta = 0$, then it is optimal for society to grant policymakers complete discretion in how they choose policy, since preferences are completely aligned. Suppose instead that $\Delta > 0$. Then, if society continues to grant complete discretion to policymakers, the chosen policy will on average align with the one preferred by society, but the presence of political polarization will lead to undesirable movements in policy as a function of θ. Consequently, it is no

[15]Note that if Assumption 1 is violated, the optimal solution can be a disconnected set, as shown in Alonso and Matouschek (2008).

Figure 1: Optimal delegation sets as a function of polarization, Δ



longer optimal to grant complete discretion to policymakers. To see this, note that a sufficient condition for a binding cap is that the θ_H policymaker's effective bias is positive, $\Delta\alpha_L + \nu > 0$. This is more likely to be satisfied for a larger degree of polarization Δ because it increases the effective bias of the θ_H policymaker. Similarly, for any level of Δ, a larger absolute value of ν implies greater misalignment in preferences between society and policymakers, and consequently it is optimal to grant policymakers less discretion.

The next lemma shows that a sufficiently large polarization Δ or bias ν collapses the optimal delegation set to a single point.

Lemma 2. *For any fixed $\Delta \geq 0$, there is a threshold $\bar{\nu}(\Delta)$ such that for $|\nu| \geq \bar{\nu}(\Delta)$, the delegation set is a single point. Similarly, for any fixed ν, there is a threshold $\Delta(\nu)$ such that for $\Delta \geq \Delta(\nu)$, the delegation set is a single point.*

Figure 1 illustrates the lemma for $\nu = 0$: for a sufficiently large level of polarization, there is no discretion and the set D contains only the best constant policy.[16] We will refer to this situation as a *narrow mandate.*

[16] In the figure, the size of the delegation set is also monotonically decreasing in Δ, a feature that holds in our numerical example.

4 Legislative bargaining

We now turn to study the best outcome under legislative bargaining. Relative to the previous case, the policymaker in charge has the option to unilaterally choose a policy from the set D, but it may also choose $\pi \notin D$ if the other policymaker agrees. In this section we characterize the optimal set D. We show that, contrary to the case of executive delegation, the set D contains no intervals. The typical outcome does not respond to small realizations of the shock z but allows some flexibility for large realizations, in sharp contrast to the outcome under executive delegation.

To set up the problem, we define $\pi_D(z, \theta)$ as the policy that would be unilaterally chosen by the policymaker in charge, θ, from the set D. This policy choice is the outside option in bargaining and must satisfy (2). Given a shock realization z, if type θ is in power, it can implement any policy in the set $\mathcal{R}(z, \theta, D)$, where

$$\mathcal{R}(z, \theta, D) = \{\pi | u(\pi, z, \hat{\theta}) \geq u(\pi_D(z, \theta), z, \hat{\theta}),\ \hat{\theta} \neq \theta\}.$$

In other words, the policymaker of type θ can implement any policy that the other type prefers to the outside option.

We can then write the problem for society when it delegates policy-making to a legislature as

$$\max_{D \subseteq \mathbb{R}, \pi(z,\theta)} \int_{\underline{z}}^{\bar{z}} \sum_i \alpha_i v\left(\pi\left(z, \theta_i\right), z\right) f(z) dz \tag{8}$$

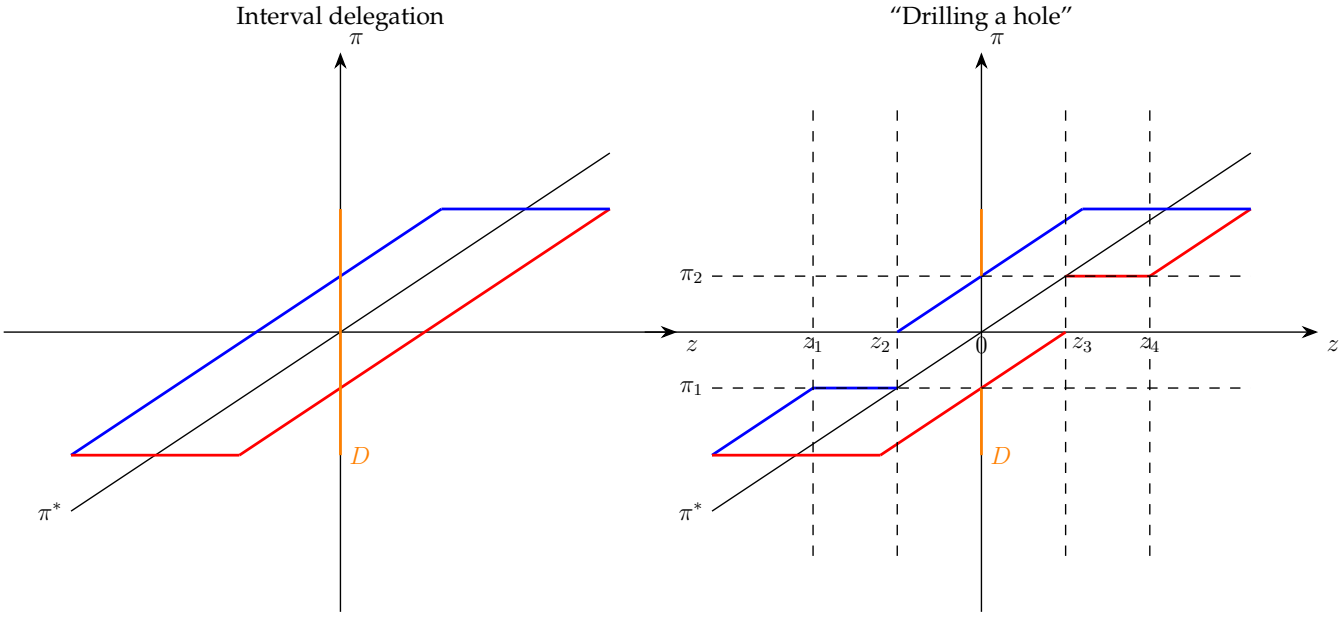
subject to

$$\pi(z, \theta) \in \arg \max_{\pi \in D \cup \mathcal{R}(z,\theta,D)} u(\pi, z, \theta). \tag{9}$$

As described in the previous section, the presence of political risk creates undesirable movements in policy. We showed that under executive delegation, this risk is mitigated by limiting the discretion granted to executive agencies through the bounds of an interval. In contrast, in the presence of bargaining, political risk can be mitigated by limiting the ability of the policymaker in charge to unilaterally choose policies. The way to do this is to restrict the set of unilateral policies available to the policymaker in charge, D. In particular, we show that when the policymaker's bias is small, it is optimal to grant discretion only over a discrete set of points. This contrasts sharply with the finding of the previous section.

We say that a delegation set D contains an interval if there is an interval $[z_a, z_b]$, with $z_a < z_b$, such that $\pi_f([z_a, z_b], \theta_L) \subseteq D$ or $\pi_f([z_a, z_b], \theta_H) \subseteq D$.

Figure 2: Sub-optimality of interval delegation in LB



Proposition 2. *Suppose that the bias is small, i.e.,* $\nu \in (-\alpha_L\Delta, \alpha_H\Delta)$. *Then, the set* D *that solves* (8) *does not contain intervals.*

Note that the proposition requires only that societal preferences can be written as a convex combination of policymaker preferences and does not rely on any distributional assumptions. In the Appendix, we characterize the solution when $\nu \notin (-\alpha_L\Delta, \alpha_H\Delta)$.

To see how this differs from the executive delegation case in Proposition 1, note that when society delegates policy-making to a legislature, the choice of D affects not only the policy implemented when the policymakers of the two factions fail to agree—just as in DE—but also the bargaining power of the policymaker in charge. Restricting discretion in the set of outside options weakens the faction in power and increases the other faction's ability to discipline policy. Unlike in DE, drilling a hole in the delegation set does not create inefficiencies, because policymakers can renegotiate. In DE, such holes force bunching at inefficient policies; under LB, those inefficiencies are renegotiated away.

To illustrate the mechanism, suppose that $\nu \in (-\alpha_L\Delta, \alpha_H\Delta)$ and that the set D contains an interval, as shown in the first panel of Figure 2. Suppose further that faction θ_L is in power. Then, whenever the low type's preferred policy lies in that interval, the low type can implement it unilaterally, so bargaining exerts no discipline on policy choice.

Now consider drilling a hole in the set of outside options D by removing a small interval (π_1, π_2), as illustrated in the second panel of Figure 2. This prevents the low type from unilaterally implementing its preferred policy for $z \in [0, z_4]$. For $z \in [0, z_3]$, the relevant outside option is π_1, and the low type still implements its preferred policy because the high type prefers $\pi_f(z, \theta_L)$ to π_1. By contrast, for $z \in [z_3, z_4]$, the relevant

outside option is π_2, and no agreement is possible: the low type prefers policies below π_2, while the high type prefers policies above it. As a result, the implemented policy is π_2, which is socially preferred to the low type's ideal policy. A symmetric argument applies when the high type is in power, for $z \in [z_1, 0]$. Hence, removing the small interval (π_1, π_2) improves social welfare.

Intuitively, removing the interval (π_1, π_2) results in lower bargaining power for the policymaker in charge and induces greater moderation in the policy choice. That is, the policy does not vary too much in the direction preferred by the θ_L type. Since $\nu \in (-\alpha_L\Delta, \alpha_H\Delta)$ and society's preferences are a convex combination of policymakers' preferences, this moderation is welfare-improving.

Note that the set of outside options D may contain multiple such points if the range of z is large enough. However, if the range of z is small enough, it may be optimal to have a unique outside option. This is particularly appealing because we can interpret this case as one where there is a given level of mandatory spending (the singleton D) that can be changed through discretionary spending only if there is agreement in the legislature.

It is instructive to compare the effect of removing an interval in our setting with its effect in the delegation literature. In Kartik et al. (2021), removing an interval from the menu offered to a single informed agent has two opposing effects. It raises the expected action toward the proposer's ideal, since types in the removed interval bunch at its endpoints and, when the type density is increasing, more move up than down. But it also raises the dispersion of the induced action, concentrating mass at the two endpoints, which a risk-averse proposer dislikes. Removing an interval is therefore optimal only where the density increases quickly enough to outweigh this dispersion cost, so the optimal set is an interval for some type densities and disconnected for others. The lack of intervals in Alonso and Matouschek (2008) and Amador et al. (2026) likewise requires conditions on the distribution of types.

In our model, removing an interval from the set of outside options has a different effect, because the set governs bargaining between two factions rather than screening one agent. It lowers the outside option of the faction in power in the affected states, weakening its bargaining power and moderating the implemented policy toward society's preference. When the bias is small, it carries no offsetting cost: whenever both factions would prefer a policy in the removed interval, they reach it by agreement, so no flexibility is lost. Thus, removing intervals when the bias is small is always weakly optimal, and the optimal set of outside options is discrete for every distribution in this case

Finally, we show that if the degree of polarization in society is sufficiently large, the best that society can do is to set the outside option to the best constant policy, $D = \{\bar{\pi}^*\}$; given this, the policymakers never agree to depart from it in the bargaining phase, and the implemented policy is $\bar{\pi}^*$ in every state.

Lemma 3. *For any ν, there exists a $\bar{\Delta}(\nu) > 0$ such that for all $\Delta > \bar{\Delta}(\nu)$, the optimal set is $D = \{\bar{\pi}^*\}$, where $\bar{\pi}^*$ is the best constant policy, and the allocation implemented by LB is $\pi_{LB}(z, \theta) = \bar{\pi}^*$.*

5 How should society allocate policy-making?

In this section, we analyze the conditions under which delegation to an executive agency (DE) or legislative bargaining (LB) is optimal for societal welfare. To this end, we compare the solutions to (2) and (9) for any choice of D and show that this comparison alone provides conditions under which one of the two institutional settings dominates in terms of welfare, independently of the choice of the delegation set. Our main result is that the preferred institution depends on the size of the policymaker bias relative to political polarization. When the bias is small, legislative bargaining is preferred because bargaining moderates political risk while preserving flexibility for sufficiently large shocks. When the bias is large, executive delegation is preferred because society wants to restrict the discretion of policymakers and prevent ex post agreements that move policy too far from the social optimum. Finally, when polarization is sufficiently large, both institutions implement the best constant policy, and society is indifferent between them.

Proposition 3.

1. *Suppose $\Delta > 0$. If the bias is small, i.e., $\nu \in (-\alpha_L\Delta, \alpha_H\Delta)$, then society weakly prefers LB to DE. The preference is strict if either $\nu > -(\bar{z} - E[z]) + \alpha_H\Delta$ or $\nu < (E[z] - \underline{z}) - \alpha_L\Delta$.*

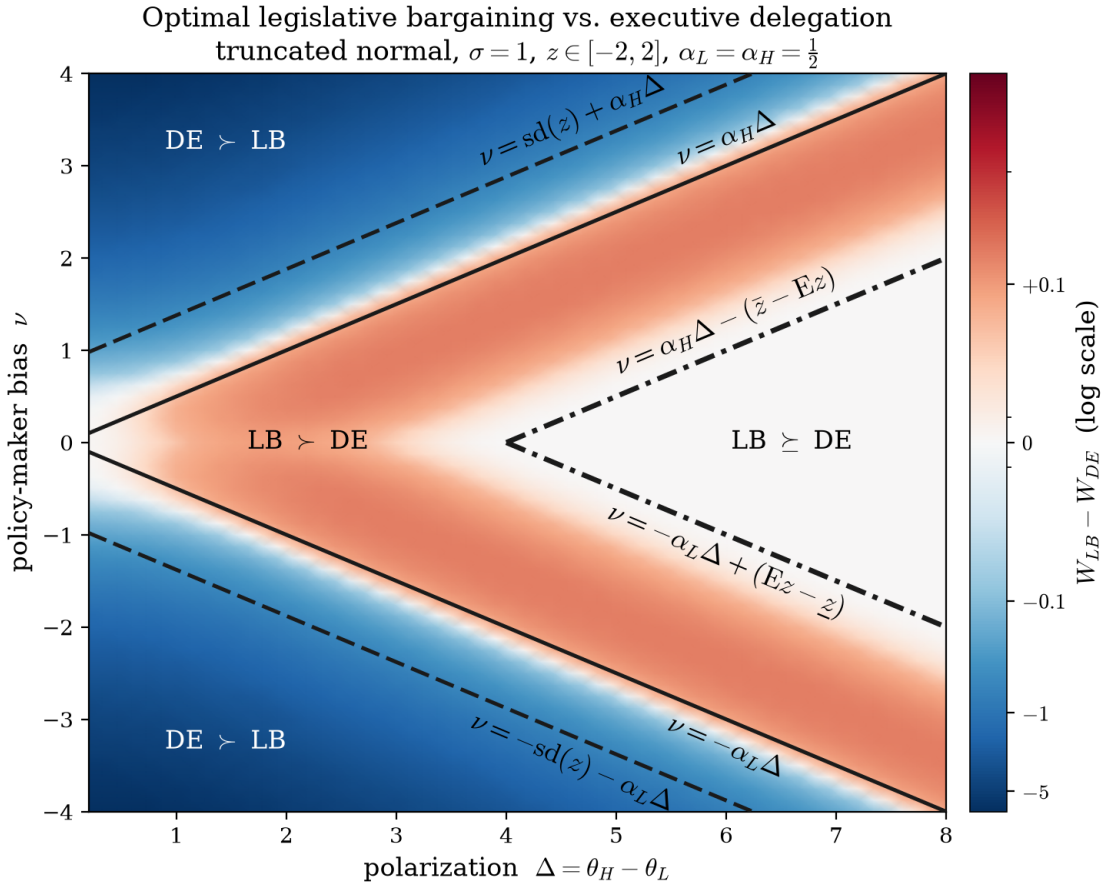
2. *If the bias is sufficiently large, i.e., either $\nu > sd(z) + \alpha_H\Delta$ or $\nu < -sd(z) - \alpha_L\Delta$, where $sd(z)$ denotes the standard deviation of z, then society strictly prefers DE to LB.*

3. *If $\Delta = 0$, society prefers DE to LB, and strictly so if $\nu \neq 0$.*

The proposition is illustrated in Figure 3. Fix the level of political polarization $\Delta > 0$. If the bias ν is small enough, then society prefers LB to DE (at least weakly). This is true irrespective of the distribution of shocks z, because for a small bias, for any delegation set, the policy implemented by a legislature dominates the policy implemented by the executive in terms of welfare for any realization of the shock.

To understand this result, suppose that the set of outside options in LB is identical to the delegation set in DE. Under LB, bargaining gives policymakers more discretion to respond to shocks. In fact, through bargaining, factions can agree to implement policies beyond those that are at the discretion of the executive. Clearly, factions will agree only to policies that lie between the ideal policies of the two factions. This is beneficial if the bias ν is small (in absolute value) because $\pi_f(z, \theta_L) \leq \pi_{LB}(z, \theta; D) \leq \pi_f(z, \theta_H)$. Thus, the

Figure 3: Preferred institutional setting as a function of polarization, Δ, and policymaker bias, ν



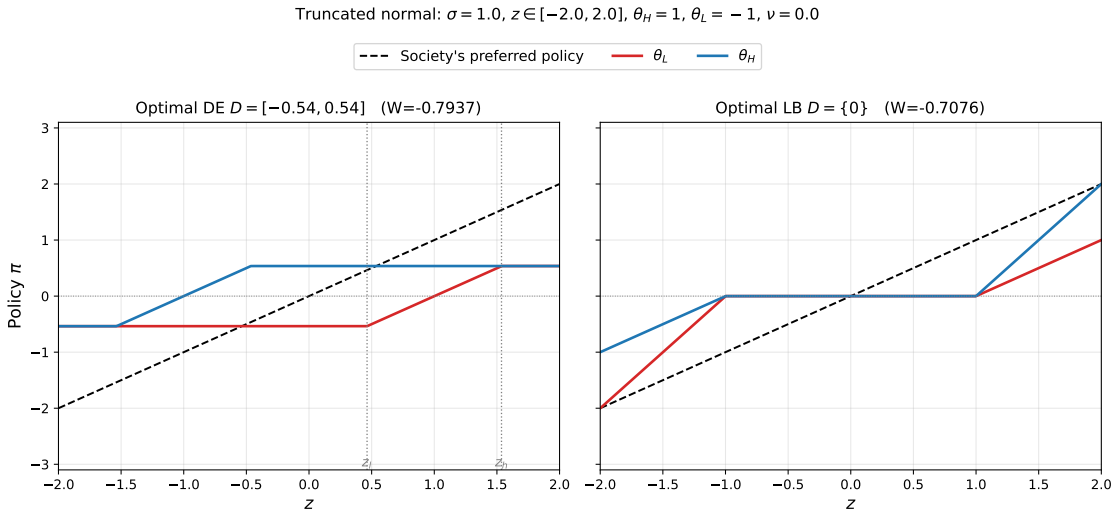
renegotiated policy is closer to society's preferences than what can be unilaterally chosen by each agent.

Intuitively, when $|\nu|$ is small, the flexible policy chosen by the average type in response to z is close to society's optimal policy; therefore, granting flexibility to the average type is desirable. However, under DE, allowing for more discretion with respect to z also implies greater political risk, since π will also respond to θ. Thus, there is a trade-off between flexibility and political risk. In fact, as Lemma 2 shows, as Δ gets large, it becomes optimal to eliminate this political risk even at the cost of eliminating any flexibility with respect to z. In the case of LB, however, bargaining between factions moderates this political risk and generates outcomes close to what the average type would have chosen, which is valuable to society.

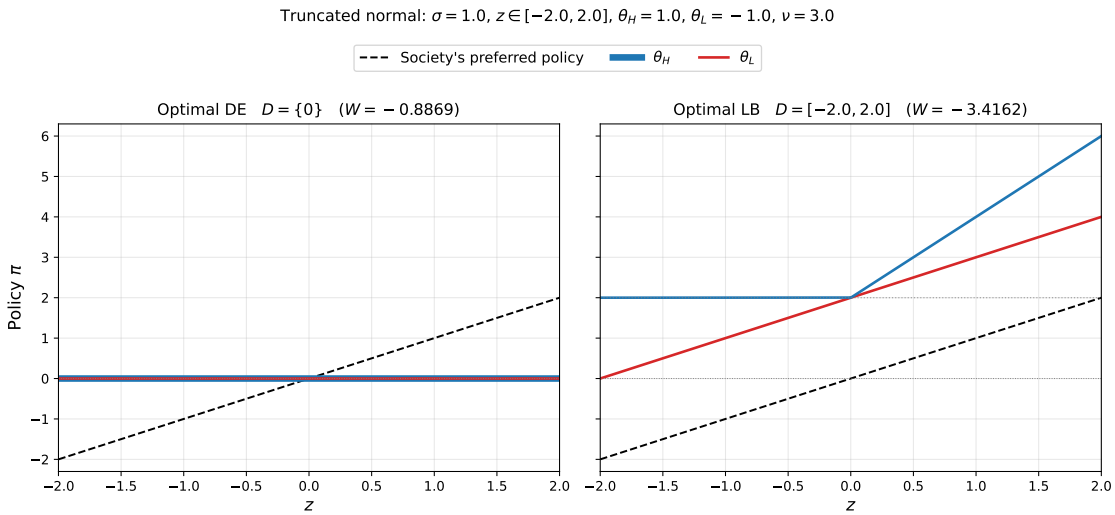
This is illustrated in the top panel of Figure 4, which plots the optimal outcomes under DE and LB for a numerical example where the shock z follows a truncated normal distribution, there is no bias, $\nu = 0$, and $\Delta = 2$. For this parameterization, under LB the optimal set D is simply $\{0\}$. For intermediate realizations of z, the outcome does not

Figure 4: Outcomes for a numerical example

1. Small bias



2. Large bias



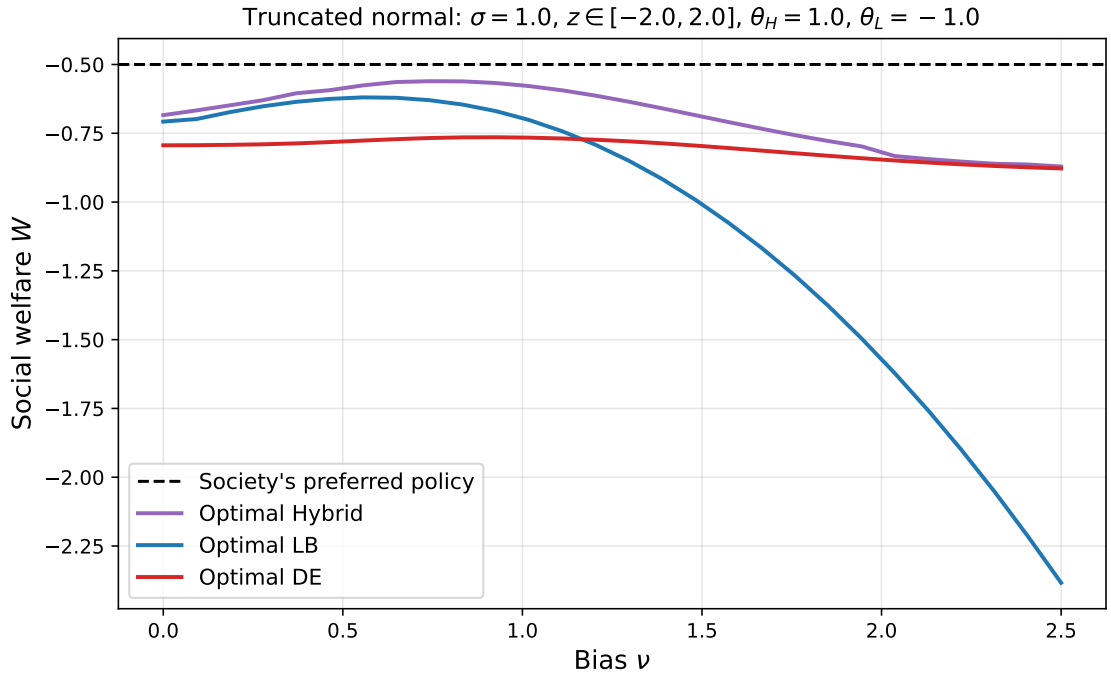
respond to shocks because factions cannot agree to a policy different from $\pi = 0$. In this sense, there is political gridlock. However, for large realizations of z, both factions prefer to renegotiate and the policy responds to shocks. Note that in this case, gridlock is a *feature* of the optimal institutional choice because it avoids costly political risk. Moreover, the ability to respond to these large shocks generates benefits relative to the case of DE, which does not allow such renegotiation and limits the response to large shocks with a cap and a floor.

These gains are strictly positive when either $\nu > -(\bar{z} - E[z]) + \alpha_H \Delta$ or $\nu < (E[z] - \underline{z}) - \alpha_L \Delta$. These conditions identify when this extra flexibility has positive value. If the optimal DE allocation already coincides with the best constant policy and no socially valuable renegotiation is possible, then LB only replicates DE. By contrast, LB strictly improves on DE whenever shocks are sufficiently important relative to polarization, so that at least one tail of the shock distribution creates gains from moving away from the best constant policy. With the quadratic specification, this occurs when either the high-state policy desired by the low type lies above the best constant policy, or the low-state policy desired by the high type lies below the best constant policy.

Eventually, as the level of polarization increases, the outcomes from the two institutions coincide with the best constant policy. This follows as a corollary of Lemma 2 and Lemma 3. Under LB, large polarization makes agreement between factions increasingly difficult. The set of policies acceptable to both factions shrinks, and eventually no useful renegotiation occurs. At the same time, under DE, large polarization increases political risk: the policy chosen by the executive depends strongly on which faction is in power. To eliminate this political risk, society optimally chooses a narrow mandate. Hence both institutions implement the same allocation, namely the best constant policy. Combining this with Proposition 3, the welfare advantage of LB over DE is non-monotone in Δ. At $\Delta = 0$ it is weakly negative, and strictly negative whenever $\nu \neq 0$. It becomes strictly positive over an intermediate range of polarization, and it returns to zero for Δ large enough that both institutions implement the best constant policy.

In contrast, when $|\nu|$ is large, the preferences of society and those of the average policymaker type are severely misaligned. In this case, society would like to limit the discretion granted to the average type. As already discussed, limiting discretion under LB for large shocks is difficult since policymakers can always choose to renegotiate policies outside D. However, through DE society can limit policymakers' discretion by granting a narrow mandate. In fact, under the bound provided in the proposition, delegating a single point to the executive achieves higher welfare than any allocation that can be implemented via LB. This is because factions can ex post agree to a policy that is, for any z, farther from society's preference than the policy with no discretion. This scenario is illustrated in the second panel of Figure 4.

Figure 5: Values for a numerical example



Proposition 3 provides sufficient conditions for either institution to be preferred by society. As illustrated in Figure 3, there is an intermediate range of values of the bias, $\nu \in (\alpha_H \Delta, \text{sd}(z) + \alpha_H \Delta)$, where these sufficient conditions are not satisfied. We now discuss the optimal choice of institution in this range using a numerical example. We consider a truncated normal distribution of shocks and a symmetric economy. We show numerically that there exists a threshold in the intermediate region such that DE is strictly preferred to LB if and only if $\nu \geq \nu^*(\Delta)$, as shown in the heatmap in Figure 3 and in Figure 5 for a given Δ. Thus, DE is preferred to LB only when the bias is sufficiently large, in which case society imposes a relatively narrow mandate on the executive agency.

Hybrid Institutional Setting

Thus far, we have focused on two institutional settings meant to capture legislatures and executive agencies. However, other institutional arrangements may deliver higher welfare. In particular, in the LB environment considered above, there are no restrictions on the policies that can be chosen whenever both policymakers agree. Placing limits on this bargaining set may improve welfare by preserving the benefits of LB while constraining the policymakers' bias.

We therefore extend the LB framework to allow for restrictions on the set of policies available in the bargaining game. This more general institutional environment nests both LB and DE as special cases and is therefore always weakly welfare-improving. We show

that the gains from this hybrid arrangement are especially large for intermediate values of the bias: in this region, combining bargaining with limits on the bargaining set delivers substantial welfare improvements. By contrast, when the bias is large, the optimal hybrid institution coincides with DE, since the main objective is to eliminate biased policy by restricting bargaining. When the bias is small, additional restrictions on the bargaining set have limited value, and LB captures most of the attainable gains.

Formally, we can write the problem as choosing two sets, D and $\bar{D}$. The set D contains the policies that the policymaker in power can unilaterally choose. The set $\bar{D}$ is the set of policies that can be chosen if there is agreement between the two policymakers. Unlike in the baseline LB, society can now choose a cap and/or a floor to restrict the set. The set $\bar{D}$ takes the form $[\bar{\pi}_{\min}, \bar{\pi}_{\max}]$. Thus, given a shock realization z, if the policymaker of type θ is in power, the set of policies that the other type will agree upon is

$$\mathcal{R}(z, \theta, D, \bar{D}) = \{\pi \in \bar{D} | u(\pi, z, \hat{\theta}) \geq u(\pi_D(z, \theta), z, \hat{\theta}),\ \hat{\theta} \neq \theta\}$$

where, recall, $\pi_D(z, \theta)$ is the best policy that type θ can choose unilaterally from the set D. Note that this institutional setting nests LB and DE as special cases. If $\bar{D}$ is the entire real line, it coincides with LB. If instead $\bar{D} = D$, it coincides with DE because policymakers have unilateral discretion within D and, with $\bar{D} = D$, renegotiation does not expand the set of policies that can be implemented.

We can then write society's problem as

$$W_{\bar{LB}} = \max_{D, \bar{D}, \pi(\theta, z)} \int \sum_i \alpha_i v\left(\pi\left(\theta_i, z\right), z\right) f\left(z\right) dz$$

subject to

$$\pi\left(\theta_i, z\right) = \arg \max_{\pi \in D \cup \mathcal{R}(z, \theta_i, D, \bar{D})} u\left(\pi, z, \theta_i\right)$$

and $D \subseteq \bar{D} \subseteq \mathbb{R}$.

We show that whenever the bias is *not* small, $\nu \notin [-\alpha_L \Delta, \alpha_H \Delta]$, it is optimal to restrict $\bar{D}$ to be a strict subset of the range of the socially optimal flexible policy, $\pi^*(z)$. The reason is that, once policymakers' preferences are sufficiently biased relative to society's, unrestricted bargaining can lead to agreements over policies that are too extreme for social welfare. A cap or floor in $\bar{D}$ prevents such outcomes while preserving the possibility of renegotiation when valuable. For sufficiently large values of the bias, it is optimal to have $\bar{D} = D = \{\bar{\pi}^*\}$, effectively mimicking the narrow mandate case in DE.

Proposition 4. *Suppose that $\nu \notin [-\alpha_L \Delta, \alpha_H \Delta]$. Then $\bar{D}$ is a strict subset of range $(\pi^*(z))$. In particular, if ν is sufficiently large, then $\bar{D} = D = \{\bar{\pi}^*\}$ where $\bar{\pi}^*$ is the best constant policy.*

Figure 5 illustrates the implications of the proposition in our numerical example. For

sufficiently large values of the bias, the hybrid institution coincides with DE. In this region, preference misalignment is so severe that society prefers to eliminate discretion altogether. The optimal policy is therefore the best constant policy, and allowing bargaining does not improve welfare.

For intermediate values of the bias, the hybrid institution strictly improves on both baseline DE and LB. Relative to DE, the hybrid institution uses bargaining between factions to allow for a flexible response to shocks with dampened political risk. Relative to LB, the cap or floor prevents policymakers from agreeing on excessively distorted policies that reflect their bias. These gains are largest precisely when the bias is large enough that unrestricted LB performs poorly, but not so large that all discretion should be eliminated.

When the bias is small, the gains from adding caps and floors to LB are limited. Bargaining between the two factions already prevents extreme outcomes, so the baseline LB institution captures most of the welfare gains available from flexibility and compromise.

Overall, these results reinforce the conclusions of our analysis of LB and DE. For small bias, pure LB is approximately the best institution because it allows for flexibility and dampens political risk. For large bias, the best that society can do is to restrict policy and disallow ex post renegotiation. This can be implemented either by delegating policy to an executive agency with a narrow mandate or by restricting the set of policies that legislatures can choose (as in the hybrid system with $D = \bar{D}$).

6 Implications

We now discuss some takeaways from our analysis. First, our results emphasize that delegating policy-making to executive agencies is desirable when the policymaker bias is large not because of the executive agencies' ability to nimbly respond to shocks but rather because of society's ability to impose mandates on these agencies that are enforced by the judicial branch. In contrast, delegating to the legislature is desirable when polarization is large relative to the policymaker bias, because it allows society to respond to large shocks while simultaneously limiting political risk. In this sense, Congress's inaction can be a feature and not a problem.

Justice Gorsuch's recent concurrence in U.S. Supreme Court, 2025 articulates a similar logic:

> [L]egislating can be hard and take time. And, yes, it can be tempting to bypass Congress when some pressing problem arises. But the deliberative nature of the legislative process was the whole point of its design. Through that process, the Nation can tap the combined wisdom of the people's elected representatives, not just that of one faction or man. There, deliberation tempers

impulse, and compromise hammers disagreements into workable solutions.

The case of environmental regulation is a good illustration of the contrast between the conventional view of policy-making and the one argued for in this paper. As noted by Klyza and Sousa (2013), the increased role for the executive in environmental policy was seen as a solution to the political gridlock on environmental issues in Congress; however, this increase in executive discretion has led to significant political risk in the enacted policies. For example, the authors highlight work that shows that fines and violation notices to polluters dropped significantly during the George W. Bush administration as compared with the Clinton administration. The recent rescission of the EPA's greenhouse gas endangerment finding further illustrates this point. The 2009 finding effectively granted EPA the authority to regulate greenhouse gas emissions across multiple sectors under the Clean Air Act. Consistent with our model's predictions, this broad delegation generated significant political risk: the Trump administration rescinded the finding entirely in February 2026. This policy reversal is precisely the type of costly political risk that our analysis suggests could be mitigated by instead allocating such a polarized policy issue to the legislature. Our paper argues that political gridlock in Congress is a feature of the system that can limit this political risk.

In summary, our results caution strongly against allocating policy-making to an agency because of political gridlock within the legislature. As emphasized previously, this gridlock can be part of the optimal mechanism and is used to limit undesirable political risk in policy outcomes. For example, broadening central bank mandates to include objectives such as climate change or inequality to get around Congress's inaction might lead to unwanted political risk and excessive volatility in these outcomes.[17]

Second, our results imply that when it is optimal to delegate policy-making to an executive agency, a relatively narrow mandate should be imposed since delegation is valuable precisely because of the ability of the judicial system to enforce these mandates. Our results can help interpret the recent discussions of the Chevron doctrine and its overturning. The Chevron doctrine (U.S. Supreme Court, 1984) held that courts should defer to an agency's interpretation of an ambiguous statute if the interpretation was reasonable. This principle granted executive agencies substantial discretion within the framework provided by Congress and led to political risk in enacted policies, as in the case of the aforementioned environmental regulation. However, the recent *Loper Bright Enterprises v. Raimondo* (U.S. Supreme Court, 2024) decision marks a significant shift in that courts now take a more active role in interpreting statutes, reducing the deference previously given to agencies. Our results imply that narrow mandates under delegation

[17]This argument is related to the critique in Bessent (2025) of the expansion of the goals of central banks – what he terms "Gain-of-Function" in monetary policy – because of the idea that "central banks are the only game in town."

are optimal, and this finding aligns with the recent ruling, which can be interpreted as a tightening of mandates. The *Loper Bright* decision reflects a judicial preference for more precise legislative guidelines and reduced agency discretion, which is consistent with our result that when policy bias is significant and it is optimal to delegate, tighter control over executive discretion is beneficial. Of course, this argument relies on judges who are immune from bias and political risk; otherwise, policies delegated to the executive can continue to suffer from undesirable political risk.

A classic example of a policy handled by the legislature is income tax policy, while an example of policy that is delegated to an executive agency is monetary policy. Through the lens of our model, allowing Congress to decide on personal income tax policy is optimal if the degree of polarization on this issue is large relative to the bias. Arguably, there is significant disagreement over income tax policy between the different factions in Congress. Our results also provide predictions on how tax policy should respond to shocks under LB. Indeed, consistent with observation, taxes and transfers are typically not responsive to small shocks (for example, business cycle shocks) but are responsive to large shocks (for example, the Covid crisis), for which all factions agree on how to change these policies.

In contrast, in our model, delegating to a central bank is optimal if monetary policy suffers from a large bias relative to polarization. This bias can arise from time inconsistency, a link that we make more precise in Section 7. Observationally, the fact that we see relatively high-frequency movements in monetary policy might seem at odds with a narrow mandate. However, the movements in instruments (for example, interest rates) might still be consistent with a narrow mandate on outcomes, as is indeed the case for most central banks that have inflation-targeting mandates. Within the inflation-targeting framework, the central bank has full flexibility in how to achieve this mandate, but it has no flexibility in choosing its target. While our baseline model is silent on the discussion of instruments versus targets (see, for example, Atkeson and Kehoe (2001); Atkeson et al. (2007); Halac and Yared (2022b)), we could easily extend our model to incorporate this distinction. In particular, suppose that π corresponds to inflation, and $\pi = g(i, z)$, where i is the nominal interest rate controlled by the central bank. All our results would go through in this framework and would call for a narrow mandate on π but full flexibility on the nominal interest rate.

7 Time inconsistency and source of the bias

So far, we have assumed that the policymaker bias, ν, is an exogenous preference parameter. For example, politicians may be biased relative to society because they are prone

to being captured by special interest groups. In this section, we show that bias can also arise from time-inconsistency problems. We make this point within the context of a Barro–Gordon model and show that our results extend to this setting.

We follow Athey et al. (2005) and let x denote nominal wage inflation, u_n denote the unemployment rate, and z denote the state of the economy. Suppose that z is distributed with density f over the support $Z = [\underline{z}, \bar{z}]$ and with mean $E[z] = 0$. The shock affects the desired level of inflation, and there is *heterogeneity* in this desired level. The payoff of type θ when the state of the economy is z is $-\frac{1}{2}\left[u_n^2 + (\pi - z - \theta)^2\right]$, where π denotes the growth rate of money (which also pins down price inflation). Unemployment is determined by a static Phillips curve $u_n = U + x - \pi$, where U denotes the natural rate of unemployment. Substituting the Phillips curve in the payoff of type θ gives

$$R(x, \pi, z, \theta) = -\frac{1}{2}\left[(U + x - \pi)^2 + (\pi - z - \theta)^2\right].$$

A *monetary authority* of type θ sets the growth rate of money $\pi(z, \theta)$ as a function of the shock. Private agents have rational expectations and set wage growth equal to the expected money growth rate

$$x = \sum_i \alpha_i \int \pi(z, \theta_i) f(z) dz, \tag{10}$$

where α_i denotes the probability that the monetary authority is of type θ_i. We refer to (10) as the implementability constraint. The timing of events is as follows: private agents set nominal wage growth x, the state z is realized, and with probability α_i the monetary authority θ_i sets the money growth rate.

Under commitment, society's preferred allocation solves

$$\max_{\pi(\cdot)} \sum_i \alpha_i \int R(x, \pi(z), z, \theta_i) f(z) dz$$

subject to the implementability constraint (10). An alternative interpretation of the aforementioned preferences is that they can arise from a Nash bargaining game between the different factions with bargaining weights α_i. The solution, which we call the full-information Ramsey allocation, is $\pi^*(z) = \bar{\theta} + \frac{z}{2}$, and private agents rationally expect $x = \bar{\theta}$, where $\bar{\theta} = \sum_i \alpha_i \theta_i$. The best constant policy is $\bar{\pi}^* = \bar{\theta}$.

Without commitment, the monetary authority solves $\max_\pi R(x, \pi, z, \theta)$ taking x as given. The best response is then $\pi_f(z, \theta; x) = \frac{U + x + \theta + z}{2}$. If private agents believe that the monetary authority is acting with full discretion, then, using (10), the expected nominal wage inflation is $x = U + \bar{\theta}$ and $\pi(z, \theta) = \bar{\theta} + \frac{z}{2} + U + \frac{\theta - \bar{\theta}}{2}$. The static Nash outcome

differs from society's preferred allocation in two respects: an inflation bias U, which stems from the monetary authority's attempt to exploit the Phillips curve, and political risk, captured by the term $(\theta - \bar{\theta})/2$. These differences motivate the institutional choice problem described in the previous sections. In particular, we characterize when it is optimal to use an executive agency (DE) versus the legislature (LB) to conduct monetary policy.

We can then write the problem for DE as

$$\max_{D,\pi(z,\theta),x} \sum_i \sum_j \alpha_i \alpha_j \int R(x, \pi(z, \theta_j), z, \theta_i) f(z) dz \tag{11}$$

subject to the incentive compatibility constraint for all θ

$$\pi(z, \theta) = \arg\max_{\pi \in D} R(x, \pi, z, \theta)$$

and the implementability constraint (10).

The problem for LB can be written as

$$\max_{D \subseteq \mathbb{R}, \pi(z,\theta), x} \sum_i \sum_j \alpha_i \alpha_j \int R(x, \pi(z, \theta_j), z, \theta_i) f(z) dz \tag{12}$$

subject to

$$\pi(z, \theta) \in \arg \max_{\pi \in D \cup \mathcal{R}(z,\theta,D)} R(x, \pi, z, \theta) \tag{13}$$

and the implementability constraint (10), where, analogously to Section 4, $\pi_D(z, \theta; x) \equiv \arg\max_{\pi' \in D} R(x, \pi', z, \theta)$ is the policy that the faction in power would implement unilaterally under disagreement, and

$$\mathcal{R}(z, \theta, D; x) \equiv \left\{ \pi : R\left(x, \pi, z, \hat{\theta}\right) \geq R\left(x, \pi_D(z, \theta; x), z, \hat{\theta}\right), \ \hat{\theta} \neq \theta \right\}$$

is the set of policies that the opposition accepts.

To see the connection with our baseline model, consider the Lagrangian for the two problems above and let λ be the Lagrange multiplier on the implementability constraint; it is then as if society had biased preferences

$$R\left(x, \pi(z, \theta), z, \bar{\theta}\right) - \lambda \pi(z, \theta).$$

Thus, the term $\lambda \pi(z, \theta)$ captures the bias term $\nu \pi$ in the case analyzed so far. The main difference is that now the bias is endogenous to the allocation and dependent on the severity of the time-inconsistency problem as captured by the term U.

The next proposition shows that the conclusion of Proposition 3 extends to this econ-

omy, where the severity of the time-inconsistency problem U plays the role of the exogenous bias ν:

Proposition 5. *For any $\Delta > 0$, define $U_L(\Delta) \equiv \frac{1}{2}\alpha_H \Delta$ and $U_H(\Delta) \equiv \alpha_H \Delta + \sqrt{\frac{\alpha_H^2 \Delta^2 + \text{var}(z)}{2}}$. If $U < U_L(\Delta)$, then society weakly prefers LB to DE. If $U > U_H(\Delta)$, then society strictly prefers DE to LB. If $\Delta = 0$, society prefers DE to LB.*

8 Credibility

As shown in the previous sections, executive delegation is valuable if mandates can be enforced and factions can be prevented from renegotiating ex post. However, in practice, since the decision to delegate is typically passed by the legislature, the legislature can either change the mandate to allow greater discretion or set policy itself within the legislature. For example, the legislature has the authority to change the mandate of the central bank or decide monetary policy itself. Therefore, it is important to understand whether the institutional arrangements described earlier are *credible*. In this section, we show that when DE is optimal, it is not credible when the bias is exogenous (as in our baseline economy), but it is credible when the bias arises from a time-inconsistency problem.

We say that an institutional arrangement is credible if the legislature cannot agree to change it. Formally, let $V_{\hat{\theta}}^i(D, \theta)$ denote the expected value to policymaker of faction $\hat{\theta}$ of having faction θ in power, with delegation set D, in institutional setting $i \in \{DE, LB\}$. An institutional setting consisting of $i \in \{DE, LB\}$ and a delegation set D is *credible* if, for any faction θ in power, that faction's policymaker prefers (i, D) to any alternative institution $(j, \hat{D})$ that the other faction's policymaker would agree to. That is,

$$V_\theta^i(D, \theta) \geq \max_{j, \hat{D}} \left\{ V_\theta^j(\hat{D}, \theta) \;\middle|\; V_{\hat{\theta}}^j(\hat{D}, \theta) \geq V_{\hat{\theta}}^i(D, \theta) \right\} \quad \text{for all } \theta.$$

We first study whether the institutional arrangements are credible in our baseline economy with exogenous bias.

Proposition 6. *If the bias is not small (i.e., $\nu > \alpha_H \Delta$ or $\nu < -\alpha_L \Delta$), then the optimal DE is not credible. Thus, if DE is optimal then it is not credible.*

With a large exogenous bias, the gains from preventing ex post renegotiation are realized only by society. Therefore, narrow mandates are not credible, since both factions find it optimal to renegotiate, thus bringing the chosen policy closer to their own preferences and away from society's.[18]

[18] In this analysis, we abstract from the potential disciplining role of elections, which may deter policymakers from supporting reforms that would allow them to choose policies closer to their bias.

In contrast, when the bias results from a strong degree of time inconsistency, a well-designed narrow mandate delegated to the executive is credible. Note that here we assume that the mandate is chosen before private agents act and set x.

Proposition 7. *For all Δ, there exists $\bar{U}_H(\Delta)$ such that for all $U \geq \bar{U}_H(\Delta)$, DE is preferred to LB and the optimal DE outcome is a credible narrow mandate at the best constant policy.*

In the Barro–Gordon model, each policymaker would like to commit ex ante not to best respond ex post to x, to avoid the cost of expected inflation. Therefore, policymakers value the commitment to not renegotiate ex post, which makes the narrow mandate attractive if the time-inconsistency problem is severe. One way of attaining this benefit is by delegating policy-making to an agency that follows the best constant policy. However, this is costly for two reasons. First, policy can no longer respond to the shock z, and, second, the inflation target differs from the one preferred by any individual type θ (since it corresponds to the average type $\bar{\theta}$). We show that the benefits outweigh the costs for all types if the degree of time inconsistency, U, is large relative to the volatility of z and the degree of polarization.

Note that it is critical that the mandate be chosen before x is chosen by the private agents, since commitment is valuable only if x can be manipulated. If instead the mandate can be chosen after x is chosen, the narrow mandate would no longer be credible, since it would always be optimal to best respond to x. However, in a dynamic environment, the narrow mandate can continue to be credible by standard reputational arguments (Abreu (1988)). Interestingly, political polarization makes it easier to support good outcomes. This is because the faction in charge cannot unilaterally choose its best response, as it must obtain the approval from the opposition. In turn, this limitation reduces the static gains from deviating and makes deviations less attractive. This mechanism is similar to the one in Chari et al. (2020) for a monetary union.

In contrast to Proposition 6, LB is always credible, as we show next. To fix ideas, suppose that the set D in LB contains a single point π_o.

Proposition 8. *Suppose we are in the environment with exogenous bias and that $D = \{\pi_o\}$. Then LB is always credible.*

This result suggests that Congress will never choose to delegate policy-making so long as the faction in power needs approval from the opposing faction. This interpretation seems in contrast to the observed expansion of delegation to executive agencies. We view this overdelegation as arising from the executive branch, where it may be possible for the faction in power to use executive agencies to implement that faction's desired policies without much approval from the opposition. In the context of the model, this scenario would correspond to the case in which the faction in power were no longer subject to the participation constraint; thus, it would prefer to delegate.

9 Conclusion

This paper studies the optimal delegation of decision-making authority in an economy with three key features: policymaker *bias* relative to societal preferences, shocks that make *flexibility* valuable, and political *polarization* between different factions in society that disagree on the optimal policy. We show that allocating policy-making to the legislature is preferred when polarization is large relative to the bias, while allocating policy-making to an executive agency is preferred in the opposite case. We argue that, in contrast to conventional wisdom, the legislature is the institution through which flexibility is granted, and not the executive agency, because when it is optimal to delegate to the latter, it is optimal to do so with a relatively narrow mandate. We also study the credibility of these institutions and show that while legislative bargaining is always credible, executive delegation is credible only when the bias arises from time-inconsistency problems.

Our work provides a normative theory of these institutions. One interesting avenue for future work is to develop a positive theory for the prevalence of these institutions and, in particular, account for the rise of executive agencies in policy-making. Our model abstracts from many features that would make executive delegation attractive for Congress. For example, as argued by Epstein and O'Halloran (1999), politicians might choose to delegate policies that could hurt their re-election chances, such as the closure of military bases in the 1980s.

Finally, given the narrow mandates that are optimal in the context of executive delegation, it would be interesting to study the design of these mandates in a dynamic economy in the presence of bias and political risk while accounting for their credibility. See, for example, Dovis and Kirpalani (2021).

References

Abreu, D. (1988): "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 56, 383–396. 28

Aghion, P., A. Alesina, and F. Trebbi (2004): "Endogenous Political Institutions," *The Quarterly Journal of Economics*, 119, 565–611. 5

Alesina, A. and G. Tabellini (2007): "Bureaucrats or Politicians? Part I: A Single Policy Task," *American Economic Review*, 97, 169–179. 5

——— (2008): "Bureaucrats or Politicians? Part II: Multiple Policy Tasks," *Journal of Public Economics*, 92, 426–447. 5

Ali, S. N., B. D. Bernheim, A. W. Bloedel, and S. C. Battilana (2023): "Who Controls the Agenda Controls the Legislature," *American Economic Review*, 113, 3090–3128. 4

Alonso, R. and N. Matouschek (2008): "Optimal Delegation," *The Review of Economic Studies*, 75, 259–293. 4, 5, 11, 15

Amador, M. and K. Bagwell (2013): "The Theory of Optimal Delegation with an Application to Tariff Caps," *Econometrica*, 81, 1541–1599. 4, 7, 10, 11, 38

Amador, M., K. Bagwell, and L. Carpizo (2026): "Generalized Delegation," Working paper. 4, 5, 15

Amador, M., I. Werning, and G.-M. Angeletos (2006): "Commitment vs. Flexibility," *Econometrica*, 74, 365–396. 4, 10, 39

Aranson, P. H., E. Gellhorn, and G. O. Robinson (1982): "Theory of Legislative Delegation," *Cornell L. Rev.*, 68, 1. 6

Athey, S., A. Atkeson, and P. J. Kehoe (2005): "The Optimal Degree of Discretion in Monetary Policy," *Econometrica*, 73, 1431–1475. 4, 7, 25

Atkeson, A., V. V. Chari, and P. J. Kehoe (2007): "On the Optimal Choice of a Monetary Policy Instrument," Staff Report 394, Federal Reserve Bank of Minneapolis. 24

Atkeson, A. and P. J. Kehoe (2001): "The Advantage of Transparent Instruments of Monetary Policy," Tech. rep., National Bureau of Economic Research. 24

Baron, D. P. (1996): "A Dynamic Theory of Collective Goods Programs," *American Political Science Review*, 90, 316–330. 4

Baron, D. P. and J. A. Ferejohn (1989): "Bargaining in Legislatures," *American Political Science Review*, 83, 1181–1206. 4

Barro, R. J. and D. B. Gordon (1983): "Rules, Discretion and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics*, 12, 101–121. 3

Bessent, S. (2025): "The Fed's New Gain-of-Function Monetary Policy," *The International Economy*, 39, 16–68. 23

Boerma, J., A. Tsyvinski, and A. P. Zimin (2022): "Bunching and Taxing Multidimensional Skills," Tech. rep., National Bureau of Economic Research. 4

Bowen, T. R., Y. Chen, and H. Eraslan (2014): "Mandatory versus Discretionary Spending: The Status Quo Effect," *American Economic Review*, 104, 2941–2974. 4

Bowen, T. R., Y. Chen, H. Eraslan, and J. Zápal (2017): "Efficiency of Flexible Budgetary Institutions," *Journal of Economic Theory*, 167, 148–176. 2, 4

Callander, S. and K. Krehbiel (2014): "Gridlock and Delegation in a Changing World," *American Journal of Political Science*, 58, 819–834. 6

Chari, V., A. Dovis, and P. J. Kehoe (2020): "Rethinking Optimal Currency Areas," *Journal of Monetary Economics*, 111, 80–94. 28

Diermeier, D. and P. Fong (2011): "Legislative Bargaining with Reconsideration," *The Quarterly Journal of Economics*, 126, 947–985. 4

Dovis, A. and R. Kirpalani (2021): "Rules without Commitment: Reputation and Incentives," *The Review of Economic Studies*, 88, 2833–2856. 29

Dziuda, W. and A. Loeper (2016): "Dynamic Collective Choice with Endogenous Status Quo," *Journal of Political Economy*, 124, 1148–1186. 4

Epstein, D. and S. O'Halloran (1999): *A Transaction Cost Politics Approach to Policy Making under Separate Powers*, Cambridge: Cambridge university press. 5, 29

Halac, M. and P. Yared (2014): "Fiscal Rules and Discretion under Persistent Shocks," *Econometrica*, 82, 1557–1614. 4

——— (2020): "Commitment versus Flexibility with Costly Verification," *Journal of Political Economy*, 128, 4523–4573. 4

——— (2022a): "Fiscal Rules and Discretion under Limited Enforcement," *Econometrica*, 90, 2093–2127. 4

——— (2022b): "Instrument-based versus Target-based Rules," *The Review of Economic Studies*, 89, 312–345. 24

Harstad, B. (2010): "Strategic Delegation and Voting Rules," *Journal of Public Economics*, 94, 102–113. 4

Kalandrakis, A. (2004): "A Three-player Dynamic Majoritarian Bargaining Game," *Journal of Economic Theory*, 116, 294–14. 4

Kartik, N., A. Kleiner, and R. Van Weelden (2021): "Delegation in Veto Bargaining," *American Economic Review*, 111, 4046–4087. 5, 9, 15

Klyza, C. M. and D. J. Sousa (2013): *American Environmental Policy: Beyond Gridlock, Updated and Expanded Edition*, Cambridge, MA: MIT Press. 23

Laffont, J.-J., E. Maskin, and J.-C. Rochet (1987): "Optimal Nonlinear Pricing with Two-dimensional Characteristics," *Information, Incentives and Economic Mechanisms*, 256–266. 4, 36

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*, Series in Decision and Control, Wiley. 38

Maskin, E. and J. Tirole (2004): "The Politician and the Judge: Accountability in Government," *American Economic Review*, 94, 1034–1054. 5

McCubbins, M. D., R. G. Noll, and B. R. Weingast (1987): "Administrative Procedures as Instruments of Political Control," *The Journal of Law, Economics, and Organization*, 3, 243–277. 6

Moser, C. and P. Olea de Souza e Silva (2019): "Optimal Paternalistic Savings Policies," *Columbia Business School Research Paper*. 4

Mylovanov, T. (2008): "Veto-based Delegation," *Journal of Economic Theory*, 138, 297–307. 5

Persson, T., G. Roland, and G. Tabellini (1997): "Separation of Powers and Political Accountability," *The Quarterly Journal of Economics*, 112, 1163–1202. 5

Piguillem, F. and A. Riboni (2015): "Spending-biased Legislators: Discipline Through Disagreement," *The Quarterly Journal of Economics*, 130, 901–949. 4

——— (2021): "Fiscal Rules as Bargaining Chips," *The Review of Economic Studies*, 88, 2439–2478. 4

Rao, N. (2015): "Administrative Collusion: How Delegation Diminishes the Collective Congress," *NYUL Rev.*, 90, 1463. 6

Rochet, J.-C. and P. Choné (1998): "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 783–826. 4

Rogoff, K. (1985): "The Optimal Degree of Commitment to an Intermediate Monetary Target," *The Quarterly Journal of Economics*, 100, 1169–1189. 5

Sublet, G. (2023): "The Optimal Degree of Discretion in Fiscal Policy," Working paper. 4, 9

Tucker, P. (2019): *Unelected Power: The Quest for Legitimacy in Central Banking and the Regulatory State*, Princeton University Press. 1

U.S. Supreme Court (1984): "Chevron U.S.A., Inc. v. Natural Resources Defense Council, Inc." Supreme Court of the United States, 468 U.S. 837. 3, 23

——— (2024): "Loper Bright Enterprises v. Raimondo," Supreme Court of the United States, no. 22-451. 23

——— (2025): "Learning Resources, Inc. v. Trump," Supreme Court of the United States, no. 24-1287. 22

Volden, C. (2002): "A Formal Model of the Politics of Delegation in a Separation of Powers System," *American Journal of Political Science*, 111–133. 6

Appendix

A Proofs of results in Section 3

A.1 Proof of Proposition 1

The Ramsey step

As described in the main text, we prove this proposition in two steps. First, we consider the Ramsey problem (6) in which the society is restricted to delegating an interval and chooses the bounds of this interval. Second, we show that no other incentive feasible allocation dominates this.

We now consider the Ramsey problem in which society chooses an interval $D = [\pi_l, \pi_h]$. Given this interval, and our assumption on policymaker preferences, the choice of a policymaker of type θ_L is

$$\pi_D(z, \theta_L) = \min\{\max\{\pi_f(z, \theta_L), \pi_l\}, \pi_h\}$$

and

$$\pi_D(z, \theta_H) = \pi_D(z + \Delta, \theta_L).$$

The latter equality implies that we can write the policy choice of θ_H in terms of θ_L. To do so we need to extend the range of states to $[\underline{z}, \overline{z} + \Delta]$.

We can parametrize the bounds of the interval by threshold state realizations z_l, z_h so that $D = [\pi_l, \pi_h] = [\pi_f(z_l, \theta_L), \pi_f(z_h, \theta_L)]$.

Next, we prove Lemma 1, which characterizes the optimal threshold z_h if it is interior.

Proof. A sufficient condition for the cap to be binding is that the effective bias of type θ_H is strictly positive, i.e., $\Delta\alpha_L + \nu > 0$ and $\alpha_H > 0$. Suppose by way of contradiction that the optimal cap is not binding. Without a binding cap the high type's preferred policy is strictly larger than the highest policy that society could want to implement $\pi_f(z, \theta_H) > \pi^*(\bar{z})$ for $z \in [\bar{z} - \Delta\alpha_L - \nu, \bar{z}]$. Because $\alpha_H > 0$, it follows that a cap at $\pi^*(\bar{z})$ strictly improves welfare, which contradicts that a non-binding cap is optimal.

If the solution to the Ramsey problem is an interior cap, i.e., $z_h \in (z_l, \bar{z} + \Delta)$, then z_h necessarily solves

$$\begin{aligned}\max_{z_h} &\int_{z_l}^{z_h} \left[\alpha_L v(\pi_f(z, \theta_L), z) + \alpha_H v(\pi_f(z, \theta_L), z - \Delta)\frac{f(z - \Delta)}{f(z)}\right] f(z)dz \\ &+ \int_{z_h}^{\bar{z}+\Delta} \left[\alpha_L v(\pi_f(z_h, \theta_L), z) + \alpha_H v(\pi_f(z_h, \theta_L), z - \Delta)\frac{f(z - \Delta)}{f(z)}\right] f(z)dz.\end{aligned}$$

The first-order condition with respect to z_h gives

$$0 = \int_{z_h}^{\bar{z}+\Delta} \left[\alpha_L v_1(\pi_f(z_h, \theta_L), z) + \alpha_H v_1(\pi_f(z_h, \theta_L), z - \Delta) \frac{f(z-\Delta)}{f(z)} \right] f(z)\, dz. \tag{14}$$

Using

$$v_1(\pi_f(z_h, \theta_L), z) = z - z_h + \alpha_H \Delta - \nu,$$

and

$$v_1(\pi_f(z_h, \theta_L), z - \Delta) = z - z_h - \alpha_L \Delta - \nu,$$

in the first-order condition (14), we obtain

$$\begin{aligned}
0 &= \int_{z_h}^{\bar{z}+\Delta} \left[\alpha_L (z - z_h + \alpha_H \Delta - \nu) + \alpha_H (z - z_h - \alpha_L \Delta - \nu) \frac{f(z-\Delta)}{f(z)} \right] f(z)\, dz \\
&= \int_{z_h}^{\bar{z}+\Delta} (z - z_h - \nu) (\alpha_L f(z) + \alpha_H f(z-\Delta))\, dz + \int_{z_h}^{\bar{z}+\Delta} [\alpha_L \alpha_H \Delta f(z) - \alpha_H \alpha_L \Delta f(z-\Delta)]\, dz \\
&= \int_{z_h}^{\bar{z}+\Delta} (z - z_h - \nu) \kappa(z)\, dz + \alpha_L \alpha_H \Delta \int_{z_h}^{\bar{z}+\Delta} [f(z) - f(z-\Delta)]\, dz.
\end{aligned}$$

□

Without heterogeneity in θ, i.e., $\Delta = 0$, an optimal cap, if it is interior, equates the bias on the left-hand side to the loss of discretion on the right-hand side, $\nu = E_\kappa[z|z \geq z_h] - z_h$. Thus, a cap is binding only if the policymaker's bias is positive.

With heterogeneity in θ, i.e., $\Delta > 0$, the distribution K captures the higher loss of discretion for θ_H than that for θ_L. Heterogeneity adds a second component to the benefit of a cap and d captures the net disagreement over the cap. It aggregates the net disagreement for all policies greater than the cap. For example, suppose that the distribution is uniform. Then, absent the bias, societal preferences are a convex combination of those of the low and high type. If $\Delta > 0$ the cost to society of θ_H choosing its flexible policy depends on the weights of this convex combination.

Note that the cap binds if $\Delta\alpha_L + \nu > 0$ because, for shocks distributed on a compact support, the mean residual life is zero at the upper bound $\bar{z}$. The analysis for an optimal floor mirrors the analysis for the cap.

Lemma 4 (Ramsey: optimal floor)**.** *If $\nu - \Delta\alpha_H < 0$ and $\alpha_L > 0$, then the optimal floor is binding. Moreover, an interior binding floor $z_l \in (\underline{z}, z_h)$ must satisfy*

$$\Delta\alpha_L\alpha_H d(z_l) - \nu K(z_l) = (z_l - E_\kappa[z|z \leq z_l])K(z_l). \tag{15}$$

The proof mirrors that of Lemma 1 and is omitted.

In the standard delegation setting with one-dimensional private information, the optimal delegation set is an interval with either a binding floor or cap. In this case both the cap and floor can be binding depending on which faction is in charge.

The verification step

With one dimension of private information, say the shocks z, the global incentive compatibility constraints are equivalent to a local incentive compatibility constraint—the envelope condition—and a monotonicity condition. In our setting, incentives must also be compatible across the other dimension of private information. The next lemma shows that under our assumptions about preferences, global incentive compatibility is equivalent to the two conditions previously described as well as a simple condition that connects the allocations of the two types. Following Laffont et al. (1987), we call the latter the *integrability condition*.

Lemma 5. *An allocation $\pi(z, \theta)$ satisfies (5) if and only if,*

$$u(\pi(z, \theta_L), z, \theta_L) = u(\pi(\underline{z}, \theta_L), \underline{z}, \theta_L) + \int_{\underline{z}}^{z} \pi(x, \theta_L) dx, \text{ for } z \in [\underline{z}, \bar{z} + \Delta], \tag{16}$$

$$\pi(z, \theta_L) \text{ is non-decreasing in } z, \tag{17}$$

$$\pi(z, \theta_H) = \pi(z + \Delta, \theta_L). \tag{18}$$

Proof. Let $U(z, \theta) = u(\pi(z, \theta), z, \theta)$. For $z > z'$,

$$\frac{U(z, \theta) - U(z', \theta)}{z - z'} \geq \pi(z', \theta),$$

and

$$\pi(z, \theta) \geq \frac{U(z, \theta) - U(z', \theta)}{z - z'}.$$

Combining the two inequalities,

$$\pi(z, \theta) \geq \frac{U(z, \theta) - U(z', \theta)}{z - z'} \geq \pi(z', \theta).$$

Hence, $\pi(z, \theta)$ is non-decreasing in z. A monotonic function is almost everywhere differentiable so it is almost everywhere continuous. Taking the limit $z \to z'$ gives (16).

Because $u(\pi, z, \theta) = u(\pi, z + \theta - \hat{\theta}, \hat{\theta})$, it follows that $\pi(z, \theta_H) = \pi(z + \theta_H - \theta_L, \theta_L)$, which gives (18). Incentive compatibility for θ_H follows from the integrability condition

and incentive compatibility for θ_L with respect to z; formally,

$$\begin{aligned}
u\left(\pi\left(z,\theta_H\right),z,\theta_H\right) &= u\left(\pi\left(z+\Delta,\theta_L\right),z,\theta_H\right)\\
&= u\left(\pi\left(z+\Delta,\theta_L\right),z+\Delta,\theta_L\right)\\
&= u\left(\pi\left(\underline{z},\theta_L\right),\underline{z},\theta_L\right)+\int_{\underline{z}}^{z+\Delta}\pi\left(x,\theta_L\right)dx\\
&= u\left(\pi\left(\underline{z},\theta_L\right),\underline{z},\theta_L\right)+\int_{\underline{z}}^{\underline{z}+\Delta}\pi\left(x,\theta_L\right)dx+\int_{\underline{z}+\Delta}^{z+\Delta}\pi\left(x,\theta_L\right)dx\\
&= u\left(\pi\left(\underline{z}+\Delta,\theta_L\right),\underline{z}+\Delta,\theta_L\right)+\int_{\underline{z}+\Delta}^{z+\Delta}\pi\left(x,\theta_L\right)dx\\
&= u\left(\pi\left(\underline{z},\theta_H\right),\underline{z},\theta_H\right)+\int_{\underline{z}}^{z}\pi\left(x,\theta_H\right)dx.
\end{aligned}$$

□

Conditions (16) and (17) are the standard envelope and monotonicity conditions for incentive compatibility for one dimension of private information. The integrability condition (18) ensures incentive compatibility across the other dimension of private information. The intuition for the integrability condition follows from the linearity of preferences in both dimensions of private information, which implies $u(\pi, z+\Delta, \theta_L) = u(\pi, z, \theta_H)$. Note that incentive compatibility for type θ_H obtains from the integrability condition and incentive compatibility for type θ_L over an extended range of shocks $[\underline{z}, \bar{z}+\Delta]$ to capture incentives over policies above θ_L's preferred policy.

The proof for the verification step has three parts. First, we map the two-dimensional delegation problem (3) into a standard one-dimensional delegation problem. Second, we use global Lagrangian methods to account for the incentive compatibility constraints and derive a set of sufficient optimality conditions. Third, we guess a Lagrange multiplier function and verify that the solution satisfies the optimality conditions.

Part I. The upshot of Lemma 5 is that the two-dimensional delegation problem (3) maps to a one-dimensional delegation problem. Note that the domain of definition of π is extended by Δ so that the integrability condition characterizes $\pi(z, \theta_H)$ for $z \in [\underline{z}, \bar{z}]$. Defining

$$\begin{aligned}
V\left(\pi_L,\pi_H,z\right) &= \alpha_L v\left(\pi_L,z\right)+\alpha_H v\left(\pi_H,z\right)\\
&= \alpha_L\sum_i \alpha_i w\left(\pi_L,\theta_i,z\right)+\alpha_H\sum_i \alpha_i w\left(\pi_H,\theta_i,z\right),
\end{aligned}$$

we can then express problem (4) as:

$$\max_{\pi(\cdot)} \int_{\underline{z}}^{\bar{z}} V\left(\pi\left(z\right), \pi\left(z+\Delta\right), z\right) f(z) dz \tag{19}$$

subject to the envelope condition (16) and the monotonicity condition (17).

Part II. We use global Lagrangian methods to account for the constraint that there are no transfers in (19). We start by defining the Lagrangian. Denote the Lagrange multiplier function on the continuum of equality constraints (16) by $\Lambda : [\underline{z}, \bar{z}+\Delta] \mapsto \mathbb{R}$ with $1-\Lambda$ integrable. The Lagrangian is a functional on the set of allocations that satisfy the monotonicity condition,

$$\Phi = \{\pi \mid \pi : [\underline{z}, \bar{z}+\Delta] \mapsto \mathbb{R} \text{ is non-decreasing}\}$$

defined as follows: $\mathcal{L} : \Phi \to \mathbb{R}$,

$$\begin{aligned}\mathcal{L}(\pi) = &\int_{\underline{z}}^{\bar{z}} \left[V\left(\pi(z), \pi(z+\Delta), z\right)\right] f(z) dz \\ &+ \int_{\underline{z}}^{\bar{z}+\Delta} \left[u(\pi(z), z, \theta_L) - u(\pi(\underline{z}), \underline{z}, \theta_L) - \int_{\underline{z}}^{z} \pi(x) dx\right] d\Lambda(z).\end{aligned} \tag{20}$$

Integrating the nested integrals in (20) by parts, the Lagrangian reads,

$$\begin{aligned}\mathcal{L}(\pi) = &\int_{\underline{z}}^{\bar{z}} V\left(\pi(z), \pi(z+\Delta), z\right) f(z) dz \\ &+ \int_{\underline{z}}^{\bar{z}+\Delta} u(\pi(z), z, \theta_L) d\Lambda(z) - u(\pi(\underline{z}), \underline{z}, \theta_L)(1-\Lambda(\underline{z})) - \int_{\underline{z}}^{\bar{z}+\Delta} \pi(z)(1-\Lambda(z)) dz.\end{aligned}$$

The global theory of constrained optimization outlined in Luenberger (1969) Chapter 8 shows that the maximizer of a Lagrangian is the solution to the constrained optimization problem of interest. In particular, Theorem 1 in Section 8.3 in Luenberger (1969) (see also Appendix B in Amador and Bagwell (2013)) implies that a maximizer of the Lagrangian (20) is a solution to (19). Denote by $\partial\mathcal{L}(\pi, h) = \lim_{\alpha \downarrow 0} \frac{1}{\alpha}[\mathcal{L}(\pi + \alpha h) - \mathcal{L}(\pi)]$ the Gateaux derivative of the Lagrangian in the direction $h : [\underline{z}, \bar{z}+\Delta] \to \mathbb{R}$. Lemma 1 in Section 8.7 in Luenberger (1969) implies that if there exists π in the convex cone Φ and an integrable function Λ such that $\mathcal{L}$ is concave,

$$\partial\mathcal{L}(\pi, \pi) = 0, \tag{21}$$

and

$$\partial\mathcal{L}(\pi, h) \leq 0 \text{ for } h \in \Phi, \tag{22}$$

then π maximizes the Lagrangian, and hence is a solution to (19).

The Gateaux derivative of the Lagrangian is given by[19]

$$\begin{aligned}\partial\mathcal{L}(\pi, h) = &\int_{\underline{z}}^{\bar{z}} \left[V_1\left(\pi(z), \pi(z+\Delta), z\right) h(z) + V_2\left(\pi(z), \pi(z+\Delta), z\right) h(z+\Delta)\right] f(z)dz \\ &+ \int_{\underline{z}}^{\bar{z}+\Delta} u_1(\pi(z), z, \theta_L)h(z)d\Lambda(z) - u_1(\pi(\underline{z}), \underline{z}, \theta_L)h(\underline{z})(1-\Lambda(\underline{z})) \\ &- \int_{\underline{z}}^{\bar{z}+\Delta} h(z)(1-\Lambda(z))dz.\end{aligned}$$

Part III. We guess that the solution to the delegation problem (19) coincides with the solution to the Ramsey problem parametrized by z_l and z_h, that is

$$\pi(z) = \begin{cases} \pi_f(z_l, \theta_L) & \text{for } z \in [\underline{z}, z_l] \\ \pi_f(z, \theta_L) & \text{for } z \in (z_l, z_h) \\ \pi_f(z_h, \theta_L) & \text{for } z \in [z_h, \bar{z}+\Delta]. \end{cases}$$

Note that this encompasses the case in which π is constant. In that case, $z_l = z_h$. If $z_l < \underline{z}$, we adopt the convention that the set $[\underline{z}, z_l]$ is empty. Similarly, if $z_h > \bar{z} + \Delta$, we adopt the convention that the set $[z_h, \bar{z}+\Delta]$ is empty.

We also guess the following Lagrange multiplier function

$$\Lambda(z) = \begin{cases} 1 - K(z) & \text{for } z \in [\underline{z}, z_l] \\ 1 + (\nu - \alpha_H\Delta)\,\alpha_L f(z) + (\nu + \alpha_L\Delta)\,\alpha_H f(z-\Delta) & \text{for } z \in (z_l, z_h) \\ 2 - K(z) & \text{for } z \in [z_h, \bar{z}+\Delta], \end{cases} \tag{23}$$

and verify the three optimality conditions: concavity of $\mathcal{L}(\pi)$, (21), and (22).

Concavity of the Lagrangian. We can rewrite the Lagrangian as follows:

$$\begin{aligned}\mathcal{L}(\pi) = &\int_{\underline{z}}^{\bar{z}} \left[V(\pi(z), \pi(z+\Delta), z)\right] f(z)dz - \int_{\underline{z}}^{\bar{z}+\Delta} u(\pi(z), z, \theta_L)\kappa\left(z\right) dz \\ &+ \int_{\underline{z}}^{\bar{z}+\Delta} u(\pi(z), z, \theta_L)\left[\frac{d\Lambda(z)}{dz} + \kappa\left(z\right)\right] dz - \int_{\underline{z}}^{\bar{z}+\Delta} \pi(z)(1-\Lambda(z))dz.\end{aligned}$$

where we used that $\Lambda\left(\underline{z}\right) = 1$. The first line is linear in π by definition of V and κ. The second line is concave if $\Lambda(z) + K(z)$ is non-decreasing because u is concave and the second

[19] Amador et al. (2006) show that the Gateaux differential exists in Lemma A.1 p. 390.

term is linear in π. Hence, it suffices that $\Lambda(z) + K(z)$ is non-decreasing for $\mathcal{L}(\pi)$ to be concave. By construction, $\Lambda(z) + K(z)$ is constant for $z \leq z_l$ and for $z \geq z_h$. Assumption 1 part 1 implies that $\Lambda(z) + K(z)$ is non-decreasing for $z \in (z_l, z_h)$. Lastly, Assumption 1 parts 2 and 3 and the Ramsey optimality conditions (7) and (15) imply that $\Lambda(z) + K(z)$ is non-decreasing at z_l and z_h.

Condition (21). Substituting the marginal utility at the allocation implemented by a cap and a floor (recall $u(\pi, z, \theta_i) = (z + \theta)\pi - \frac{1}{2}\pi^2 + \nu\pi$)

$$u_1(\pi(z), z, \theta_L) = \begin{cases} z - z_l & \text{for } z \in [\underline{z}, z_l] \\ 0 & \text{for } z \in (z_l, z_h) \\ z - z_h & \text{for } z \in [z_h, \bar{z} + \Delta], \end{cases}$$

in the Gateaux derivative gives

$$\partial\mathcal{L}(\pi, h) = \int_{\underline{z}}^{z_l} \left[\alpha_L(z - z_l + \alpha_H\Delta - \nu)f(z) + \alpha_H(z - z_l - \alpha_L\Delta - \nu)f(z - \Delta)\right] h(z)dz \quad (24)$$

$$+ \int_{\underline{z}}^{z_l} (z - z_l)h(z)d\Lambda(z) - \int_{\underline{z}}^{z_l} h(z)(1 - \Lambda(z))dz \quad (25)$$

$$+ \int_{z_l}^{z_h} \left[(\alpha_H\Delta - \nu)\alpha_L f(z) - (\nu + \alpha_L\Delta)\alpha_H f(z - \Delta) - (1 - \Lambda(z))\right] h(z)dz \quad (26)$$

$$+ \int_{z_h}^{\bar{z}+\Delta} \left[\alpha_L(z - z_h + \alpha_H\Delta - \nu)f(z) + \alpha_H(z - z_h - \alpha_L\Delta - \nu)f(z - \Delta)\right] h(z)dz \quad (27)$$

$$+ \int_{z_h}^{\bar{z}+\Delta} (z - z_h)h(z)d\Lambda(z) - \int_{z_h}^{\bar{z}+\Delta} h(z)(1 - \Lambda(z))dz \quad (28)$$

where we used that $\Lambda(\underline{z}) = 1$. The following argument shows that for z_l and z_h that parametrize the solution to the Ramsey problem (6), and given the Lagrange multiplier function (23), condition (21) is satisfied. First, note that line (26) is zero by definition of the Lagrange multiplier function (23). To show that the lines (24) and (25) sum up to zero when evaluated in the direction $h(z) = \pi(z)$, substituting the Lagrange multiplier

function (23) and rearranging terms gives

$$\begin{aligned}
&\int_{\underline{z}}^{z_l} [\alpha_L(z - z_l + \alpha_H\Delta - \nu)f(z) + \alpha_H(z - z_l - \alpha_L\Delta - \nu)f(z - \Delta)]\, h(z)dz \\
&- \int_{\underline{z}}^{z_l} (z - z_l)h(z)\kappa(z)dz - \int_{\underline{z}}^{z_l} h(z)K(z)dz \\
=&\int_{\underline{z}}^{z_l} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z - \Delta)]\, h(z)dz \\
&+ \int_{\underline{z}}^{z_l} [(z - z_l)(\alpha_L f(z) + \alpha_H f(z - \Delta))]\, h(z)dz - \int_{\underline{z}}^{z_l} (z - z_l)h(z)\kappa(z)dz - \int_{\underline{z}}^{z_l} h(z)K(z)dz \\
=&\int_{\underline{z}}^{z_l} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z - \Delta)]\, h(z)dz - \int_{\underline{z}}^{z_l} h(z)K(z)dz \\
=&\int_{\underline{z}}^{z_l} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z - \Delta) - K(z)]\, h(z)dz. \qquad (29)
\end{aligned}$$

Note that $\pi(z)$ is constant and equal to $\pi(z_l)$ for $z \in [\underline{z}, z_l]$. As a result, the constant $h(z) = \pi(z_l)$ can be taken out of the integral in (29).

The optimality condition for z_l in Lemma 4 implies that (29) equals zero. A similar argument shows that the lines (27) and (28) sum to zero when evaluated in the direction of the constant $h(z) = \pi(z_h)$ for $z \in [z_h, \bar{z} + \Delta]$.

Condition (22). The following argument shows that, given the Lagrange multiplier (23), the conditions in Assumption 1 parts 2 and 3 imply that the optimality condition (22) is satisfied. Again, line (26) is zero by definition of the Lagrange multiplier function (23) for $z \in [z_l, z_h]$. Integrating (29) by parts isolates the contribution of dh to the Gateaux derivative

$$\begin{aligned}
&\int_{\underline{z}}^{z_l} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z - \Delta) - K(z)]\, h(z)dz \\
=&-\int_{\underline{z}}^{z_l}\int_{\underline{z}}^{\hat{z}} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z - \Delta) - K(z)]\, dzdh(\hat{z}) \\
&+ \int_{\underline{z}}^{z} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z - \Delta) - K(z)]\, h(z)|_{\underline{z}}^{z_l} \\
=&-\int_{\underline{z}}^{z_l}\int_{\underline{z}}^{\hat{z}} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z - \Delta) - K(z)]\, dzdh(\hat{z}) \\
&+ \int_{\underline{z}}^{z_l} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z - \Delta) - K(z)]\, dzh(z_l).
\end{aligned}$$

The optimality condition for z_l from Lemma 4 implies that

$$\int_{\underline{z}}^{z_l} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z - \Delta) - K(z)]\, dz = 0.$$

Noting that h is non decreasing, it suffices that

$$\int_{\underline{z}}^{\hat{z}} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z-\Delta) - K(z)]\, dz \geq 0$$

for $\hat{z} \in [\underline{z}, z_l]$ for lines (24) and (25) to contribute a negative term to the Gateaux derivative. The condition in part 3 of Assumption 1 implies that the term is non-negative. To see this, rewrite the left-hand side of the inequality as follows

$$\begin{aligned}
&\int_{\underline{z}}^{\hat{z}} [\alpha_L(\alpha_H\Delta - \nu)f(z) + \alpha_H(-\alpha_L\Delta - \nu)f(z-\Delta) - K(z)]\, dz \\
&\qquad = \alpha_L\alpha_H\Delta \int_{\underline{z}}^{\hat{z}} [f(z) - f(z-\Delta)]\, dz - \nu K(\hat{z}) - \int_{\underline{z}}^{\hat{z}} K(z)dz \\
&\qquad = \alpha_L\alpha_H\Delta \int_{\underline{z}}^{\hat{z}} [f(z) - f(z-\Delta)]\, dz - \nu K(\hat{z}) - \hat{z}K(\hat{z}) + \int_{\underline{z}}^{\hat{z}} z\kappa(z)dz \\
&\qquad = \alpha_L\alpha_H\Delta \left[F(\hat{z}) - F(\hat{z}-\Delta)\right] - \nu K(\hat{z}) - \hat{z}K(\hat{z}) + \int_{\underline{z}}^{\hat{z}} z\kappa(z)dz.
\end{aligned}$$

A specular logic applies for lines (27) and (28). $Q.E.D.$

Next we show the necessary part:

Lemma 6. *If Condition 1 in Assumption 1 is not satisfied, then drilling a hole in the interval* $[\pi_f(z_l, \theta_L), \pi_f(z_h, \theta_L)]$ *improves welfare.*

Proof. Suppose that the delegation set contains the interval $[\pi_f(z_l, \theta_L), \pi_f(z_h, \theta_L)]$ but Condition 1 in Assumption 1 fails at some $z_a \in (z_l, z_h)$.

Consider an alternative set $\hat{D}$ obtained by removing the interval $(\pi_f(z_a, \theta_L), \pi_f(z_b, \theta_L))$ for some z_b such that $z_b - z_a < \Delta$ and $[z_a, z_b] \subseteq [z_l, z_h]$. Formally,

$$\hat{D} \equiv D \setminus (\pi_f(z_a, \theta_L), \pi_f(z_b, \theta_L)).$$

We prove this lemma in three steps. First, we characterize the policy under $\hat{D}$. Second, we compute the welfare effect of drilling the hole. Finally, in step 3, we show that societal welfare is larger under $\hat{D}$ relative to D.

Step 1: Characterizing the policy under $\hat{D}(z_b, z_a)$.

For $z \in [z_a, z_b]$, policymaker θ_L must choose between $\pi_f(z_a, \theta_L)$ and $\pi_f(z_b, \theta_L)$. Hence,

there exists $z^* \in (z_a, z_b)$ such that

$$\pi_{\hat{D}}(z, \theta_L) = \begin{cases} \pi_D(z, \theta_L) & \text{for } z \leq z_a, \\ \pi_f(z_a, \theta_L) & \text{for } z \in [z_a, z^*], \\ \pi_f(z_b, \theta_L) & \text{for } z \in (z^*, z_b], \\ \pi_D(z, \theta_L) & \text{for } z \geq z_b. \end{cases}$$

The threshold z^* is such that θ_L is indifferent between $\pi_f(z_b, \theta_L)$ and $\pi_f(z_a, \theta_L)$, that is

$$u\left(\pi_f(z_a, \theta_L), z^*, \theta_L\right) = u\left(\pi_f(z_b, \theta_L), z^*, \theta_L\right). \tag{30}$$

Since u is quadratic, $z^* = \frac{z_a + z_b}{2}$. By the integrability condition $\pi(z, \theta_H) = \pi(z + \Delta, \theta_L)$, an analogous characterization holds for type θ_H over $z \in [z_a - \Delta, z_b - \Delta]$, with indifference threshold $z^* - \Delta$.

Step 2: Computing the welfare effect $\delta(z_b, z_a)$.

Notice that the welfare implications of delegating $\hat{D}$ instead of D depend only on the policy for $z \in [z_a, z_b]$ for type θ_L and for $z \in [z_a - \Delta, z_b - \Delta]$ for type θ_H. Then, the previous steps imply that the welfare effect of "drilling a hole" in D is

$$\begin{aligned} \delta(z_b, z_a) = &\int_{z_a}^{z^*} \left[v(\pi_f(z_a, \theta_L), z) - v(\pi_f(z, \theta_L), z)\right] \alpha_L f(z)\, dz \\ &+ \int_{z_a}^{z^*} \left[v(\pi_f(z_a, \theta_L), z - \Delta) - v(\pi_f(z, \theta_L), z - \Delta)\right] \alpha_H f(z - \Delta)\, dz \\ &+ \int_{z^*}^{z_b} \left[v(\pi_f(z_b, \theta_L), z) - v(\pi_f(z, \theta_L), z)\right] \alpha_L f(z)\, dz \\ &+ \int_{z^*}^{z_b} \left[v(\pi_f(z_b, \theta_L), z - \Delta) - v(\pi_f(z, \theta_L), z - \Delta)\right] \alpha_H f(z - \Delta)\, dz. \end{aligned} \tag{31}$$

Since, $v(\pi, z) = (z + \bar{\theta})\pi - \frac{1}{2}\pi^2$, for any $\hat{z}$, using $\pi_f(\hat{z}, \theta_L) - \pi_f(z, \theta_L) = \hat{z} - z$, $\bar{\theta} - \theta_L = \alpha_H \Delta$, and $\bar{\theta} - \Delta - \theta_L = -\alpha_L \Delta$, we have

$$\begin{aligned} v(\pi_f(\hat{z}, \theta_L), z) - v(\pi_f(z, \theta_L), z) &= -(z - \hat{z}) \left[\frac{z - \hat{z}}{2} + (\alpha_H \Delta - \nu)\right], \\ v(\pi_f(\hat{z}, \theta_L), z - \Delta) - v(\pi_f(z, \theta_L), z - \Delta) &= -(z - \hat{z}) \left[\frac{z - \hat{z}}{2} - (\alpha_L \Delta + \nu)\right]. \end{aligned}$$

Letting

$$\Psi(z) \equiv (\nu - \alpha_H \Delta)\alpha_L f(z) + (\nu + \alpha_L \Delta)\alpha_H f(z - \Delta),$$

and substituting into (31),

$$\delta(z_b, z_a) = \int_{z_a}^{z^*} (z - z_a)\Psi(z)\, dz - \int_{z_a}^{z^*} \frac{(z - z_a)^2}{2}\kappa(z)\, dz$$
$$- \int_{z^*}^{z_b} (z_b - z)\Psi(z)\, dz - \int_{z^*}^{z_b} \frac{(z_b - z)^2}{2}\kappa(z)\, dz. \tag{32}$$

Step 3: Showing that societal welfare is larger under $\hat{D}$ relative to D for $z_b > z_a$.

Fix z_a and consider the effect of varying z_b. We will show that $\delta(z_a, z_a) = 0$ is a strict local minimum of $\delta(z_b, z_a)$ at $z_b = z_a$. This implies that for z_b close enough to z_a, the welfare gain of $\hat{D}(z_b, z_a)$ relative to D is positive.

As a first step, let's take the derivative of (32) with respect to z_b and substitute $z^* = (z_a + z_b)/2$ and $\partial z^*/\partial z_b = 1/2$ to obtain

$$\frac{\partial \delta(z_b, z_a)}{\partial z_b} = S(z_b, z_a)$$

where

$$S(z_b, z_a) \equiv \frac{z_b - z_a}{2}\Psi(z^*) - \int_{z^*}^{z_b} \Psi(z)\, dz - \int_{z^*}^{z_b} (z_b - z)\kappa(z)\, dz. \tag{33}$$

Taking the limit as z_b tends to z_a yields $\lim_{z_b \to z_a} S(z_b, z_a) = 0$.

Next, we differentiate (33) with respect to z_b to obtain

$$\frac{\partial S(z_b, z_a)}{\partial z_b} = \Psi(z^*) - \Psi(z_b) + \frac{z_b - z_a}{4}\left[\Psi'(z^*) + \kappa(z^*)\right] - \int_{z^*}^{z_b} \kappa(z)\, dz. \tag{34}$$

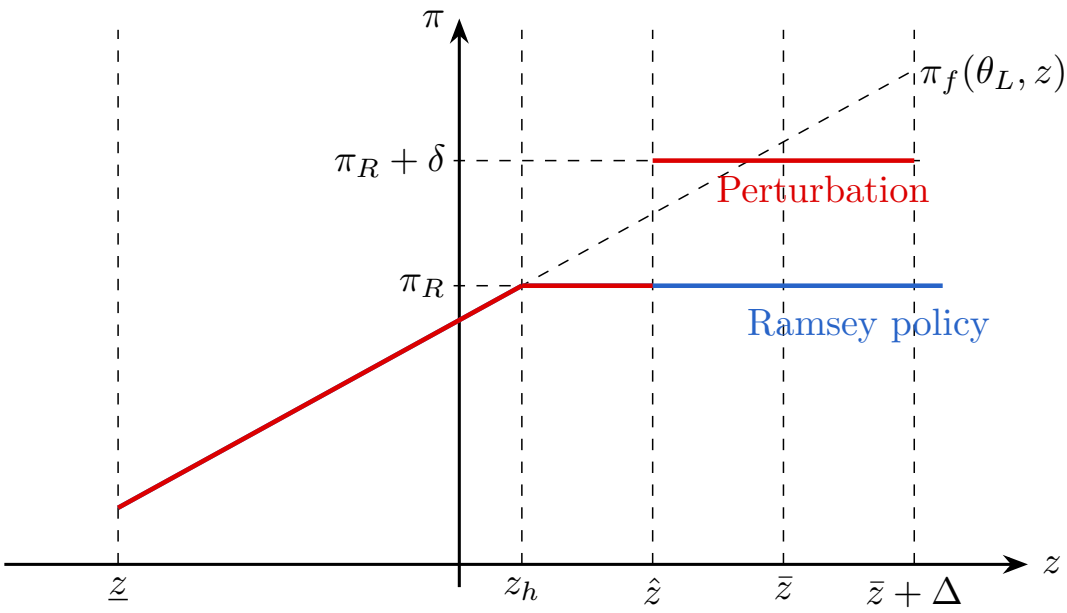
Taking the limit as z_b tends to z_a yields $\lim_{z_b \to z_a} \frac{\partial S(z_b, z_a)}{\partial z_b} = 0$.

Finally, we show that $\delta(z_b, z_a)$ is strictly convex at $z_b = z_a$. Differentiating (34) with respect to z_b and evaluating at $z_b = z_a$ (where $z^* = z_a$),

$$\begin{aligned}
\left.\frac{\partial^2 S(z_b, z_a)}{\partial z_b^2}\right|_{z_b = z_a} &= \tfrac{1}{2}\Psi'(z_a) + \tfrac{1}{4}\left[\Psi'(z_a) + \kappa(z_a)\right] - \Psi'(z_a) - \kappa(z_a) + \tfrac{1}{2}\kappa(z_a) \\
&= \left(\tfrac{1}{2} + \tfrac{1}{4} - 1\right)\Psi'(z_a) + \left(\tfrac{1}{4} - 1 + \tfrac{1}{2}\right)\kappa(z_a) \\
&= -\tfrac{1}{4}\left[\Psi'(z_a) + \kappa(z_a)\right] \\
&= -\tfrac{1}{4}\frac{d}{dz}\left[\Psi(z) + K(z)\right]\Big|_{z = z_a},
\end{aligned}$$

which is positive given that Condition 1 fails at $z_0 = z_a$. Since $\frac{\partial S(z_a, z_a)}{\partial z_b} = 0$ and $\left.\frac{\partial^2 S(z_b, z_a)}{\partial z_b^2}\right|_{z_b = z_a} > 0$, we have $\frac{\partial S(z_b, z_a)}{\partial z_b} > 0$ for z_b close to z_a. In turn, $S(z_a, z_a) = 0$ and $\partial S/\partial z_b > 0$ imply $S(z_b, z_a) > 0$, and hence $\delta(z_b, z_a) > 0$ for z_b close to z_a. $Q.E.D.$

Figure 6: Perturbation that adds one point above Ramsey cap



Lemma 7. *It is not optimal to add one point above the Ramsey cap if and only if condition 2 in Assumption 1 is satisfied.*

Proof. Suppose the delegation set is the interval with upper cap

$$\pi_R \equiv \pi^f(z_h, \theta_L) = z_h + \theta_L + \nu.$$

For any $\hat{z} > z_h$, consider a perturbation that adds the point

$$\pi_R + \delta(\hat{z})$$

to the delegation set, where $\delta(\hat{z})$ is defined by the indifference condition

$$u\left(\pi_R, \hat{z}, \theta_L\right) = u\left(\pi_R + \delta(\hat{z}), \hat{z}, \theta_L\right). \tag{35}$$

Note that if $z_h < \underline{z}$ because the solution to the Ramsey problem is a constant policy, then we are also considering perturbations that increase the optimal policy by a constant for every possible realization of the shock z. This is because the jump would happen at $\hat{z}$ outside of the support of F. The perturbation is depicted in Figure 6.

The policy associated with this perturbation coincides with the Ramsey policy for $z \leq \hat{z}$ and it is equal to $\pi_R + \delta\left(\hat{z}\right)$ for $z > \hat{z}$. Thus, the change in social welfare induced by

this perturbation is

$$\Delta V(\hat{z}) = \alpha_L \int_{\hat{z}}^{\bar{z}} \left[v(\pi_R + \delta(\hat{z}), z) - v(\pi_R, z)\right] f(z)dz + \alpha_H \int_{\hat{z}-\Delta}^{\bar{z}} \left[v(\pi_R + \delta(\hat{z}), z) - v(\pi_R, z)\right] f(z)dz. \tag{36}$$

First, note that, using our functional form, the indifference condition (35) can be written as

$$\pi_R \left(\hat{z} + \theta_L + \nu\right) - \frac{1}{2}\pi_R^2 = \left(\pi_R + \delta\right)\left(\hat{z} + \theta_L + \nu\right) - \frac{1}{2}\left(\pi_R + \delta\right)^2$$

which implies that

$$0 = \delta\left[\left(\hat{z} + \theta_L + \nu\right) - \pi_R - \frac{1}{2}\delta\right] = \delta\left[\left(\hat{z} + \theta_L + \nu\right) - \left(z_h + \theta_L + \nu\right) - \frac{1}{2}\delta\right]$$

where the second equality uses $\pi_R = z_h + \theta_L + \nu$. Thus,

$$\delta\left(\hat{z}\right) = 2\left(\hat{z} - z_h\right) > 0 \quad \forall \hat{z} > z_h \tag{37}$$

Next, using

$$v(\pi, z) = (z + \bar{\theta})\pi - \frac{1}{2}\pi^2,$$

we have

$$v(\pi_R + \delta, z) - v(\pi_R, z) = \delta(z + \bar{\theta} - \pi_R) - \frac{1}{2}\delta^2.$$

Let $b_L \equiv \nu - \alpha_H \Delta$ be the effective bias of θ_L. Since $\bar{\theta} + b_L = \theta_L + \nu$, it follows that

$$\pi_R = z_h + \theta_L + \nu = z_h + \bar{\theta} + b_L.$$

Hence

$$\begin{aligned} v(\pi_R + \delta, z) - v(\pi_R, z) &= \delta(z - z_h - b_L) - \frac{1}{2}\delta^2 \\ &= 2(\hat{z} - z_h)(z - z_h - b_L) - 2(\hat{z} - z_h)^2 \\ &= 2(\hat{z} - z_h)(z - \hat{z} - b_L). \end{aligned} \tag{38}$$

where the second equality follows from using (37) to substitute for δ, and the third equality simply rearranges terms. Using (38) into (36), we have

$$\Delta V(\hat{z}) = 2(\hat{z} - z_h)\left[\alpha_L \int_{\hat{z}}^{\bar{z}} (z - \hat{z} - b_L) f(z)\, dz + \alpha_H \int_{\hat{z}-\Delta}^{\bar{z}} (z - \hat{z} - b_L) f(z)\, dz\right]. \tag{39}$$

We now rewrite the term in brackets. Using the change of variables $x = z + \Delta$ in the

second integral, we have

$$
\begin{aligned}
&\alpha_L \int_{\hat{z}}^{\bar{z}} (z-\hat{z}-b_L) f(z)\, dz + \alpha_H \int_{\hat{z}-\Delta}^{\bar{z}} (z-\hat{z}-b_L) f(z)\, dz \\
&= \alpha_L \int_{\hat{z}}^{\bar{z}} (z-\hat{z}-b_L) f(z)\, dz + \alpha_H \int_{\hat{z}}^{\bar{z}+\Delta} (x-\hat{z}-b_L-\Delta) f(x-\Delta)\, dx \\
&= \alpha_L \int_{\hat{z}}^{\bar{z}} (z-\hat{z}-(\nu-\alpha_H\Delta)) f(z)\, dz + \alpha_H \int_{\hat{z}}^{\bar{z}+\Delta} (x-\hat{z}-\nu-\alpha_L\Delta) f(x-\Delta)\, dx \\
&= \left[\alpha_L \int_{\hat{z}}^{\bar{z}} (z-\hat{z}-\nu) f(z)\, dz + \alpha_H \int_{\hat{z}}^{\bar{z}+\Delta} (x-\hat{z}-\nu) f(x-\Delta)\, dx\right] \\
&+ \left[\alpha_L\alpha_H\Delta \int_{\hat{z}}^{\bar{z}} f(z)\, dz - \alpha_H\alpha_L\Delta \int_{\hat{z}}^{\bar{z}+\Delta} f(x-\Delta)\, dx\right] \\
&= \int_{\hat{z}}^{\bar{z}+\Delta} (x-\hat{z}-\nu)\kappa(x)\, dx - \alpha_L\alpha_H\Delta \left[F(\hat{z}) - F(\hat{z}-\Delta)\right] \\
&= \big(1-K(\hat{z})\big)\Big(E_\kappa[x-\hat{z} \mid x \geq \hat{z}] - \nu\Big) - \Delta\alpha_L\alpha_H \left[F(\hat{z}) - F(\hat{z}-\Delta)\right]
\end{aligned}
$$

where the second equality follows from $b_L = \nu - \alpha_H\Delta$, the third is just splitting terms, and the fourth one uses the definition of $\kappa(z)$.

Substituting this expression into (39) gives

$$
\Delta V(\hat{z}) = 2(\hat{z}-z_h)\big(1-K(\hat{z})\big)\left(E_\kappa[x-\hat{z} \mid x \geq \hat{z}] - \nu - \Delta\alpha_L\alpha_H \frac{d(\hat{z})}{1-K(\hat{z})}\right)
$$

where, recall, $d(\hat{z}) = F(\hat{z}) - F(\hat{z}-\Delta)$. Finally, since $\hat{z} > z_h$ and $1-K(\hat{z}) > 0$,

$$
\Delta V(\hat{z}) \leq 0 \iff E_\kappa[x-\hat{z} \mid x \geq \hat{z}] \leq \nu + \Delta\alpha_L\alpha_H \frac{d(\hat{z})}{1-K(\hat{z})}.
$$

$Q.E.D.$

A symmetric argument gives that it is not optimal to add one point below the Ramsey floor if and only if condition 3 in Assumption 1 is satisfied.

A.2 Assumption 1 is satisfied for uniformly distributed shocks

With $f = 1/(\bar{z}-\underline{z})$, the density of κ is

$$
\kappa(z) = \begin{cases} \dfrac{\alpha_L}{\bar{z}-\underline{z}} & z \in [\underline{z},\ \underline{z}+\Delta), \\ \dfrac{1}{\bar{z}-\underline{z}} & z \in [\underline{z}+\Delta,\ \bar{z}], \\ \dfrac{\alpha_H}{\bar{z}-\underline{z}} & z \in (\bar{z},\ \bar{z}+\Delta]. \end{cases}
$$

For $z \in [\underline{z}+\Delta, \bar{z}]$, straightforward calculations give $d(z) = \Delta/(\bar{z}-\underline{z})$ and $1-K(z) = (\bar{z}-z+\alpha_H\Delta)/(\bar{z}-\underline{z})$, so

$$\frac{d(z)}{1-K(z)} = \frac{\Delta}{\bar{z}-z+\alpha_H\Delta}, \tag{40}$$

and for $\hat{z} \in [\underline{z}+\Delta, \bar{z}]$,

$$E_\kappa[z \mid z \geq \hat{z}] - \hat{z} = \frac{(\bar{z}-\hat{z})^2/2 + \alpha_H\Delta\big(\bar{z}-\hat{z}+\Delta/2\big)}{\bar{z}-\hat{z}+\alpha_H\Delta}. \tag{41}$$

For $\hat{z} \in (\bar{z}, \bar{z}+\Delta]$ one can verify $d(\hat{z})/(1-K(\hat{z})) = 1/\alpha_H$ and $E_\kappa[z \mid z \geq \hat{z}] - \hat{z} = (\bar{z}+\Delta-\hat{z})/2$.

Condition 2. Define $G(\hat{z}) \equiv \Delta\alpha_L\alpha_H\, d(\hat{z}) + \nu\big(1-K(\hat{z})\big) - \big(E_\kappa[z \mid z \geq \hat{z}] - \hat{z}\big)\big(1-K(\hat{z})\big)$. Recall that this is the marginal cost of increasing the upper bound in the Ramsey problem. Optimality requires that $G(z_h) \geq 0$ (the first-order condition (7) gives $G(z_h) = 0$ if the upper bound is interior). Since $1-K(\hat{z}) > 0$ on $[z_h, \bar{z}+\Delta)$, we can write $G(\hat{z}) = \big(1-K(\hat{z})\big)g(\hat{z})$ where $g(\hat{z}) \equiv \Delta\alpha_L\alpha_H \frac{d(\hat{z})}{1-K(\hat{z})} + \nu - \big(E_\kappa[z \mid z \geq \hat{z}] - \hat{z}\big)$, so $g(z_h) \geq 0$. Thus, it suffices to show $g'(\hat{z}) > 0$ for all $\hat{z} \geq z_h$, which then gives $G(\hat{z}) = \big(1-K(\hat{z})\big)g(\hat{z}) \geq 0$.

Case 1: $\hat{z} \in [\underline{z}, \underline{z}+\Delta)$. Let $D(\hat{z}) \equiv \bar{z} - \underline{z} - \alpha_L(\hat{z}-\underline{z})$. Since $F(\hat{z}-\Delta) = 0$, one has $d(\hat{z}) = F(\hat{z}) = (\hat{z}-\underline{z})/(\bar{z}-\underline{z})$ and $1-K(\hat{z}) = D(\hat{z})/(\bar{z}-\underline{z})$, hence $d(\hat{z})/(1-K(\hat{z})) = (\hat{z}-\underline{z})/D(\hat{z})$. Write $E_\kappa[z \mid z \geq \hat{z}] - \hat{z} = N(\hat{z})/D(\hat{z})$ where

$$N(\hat{z}) = \frac{\alpha_L(\underline{z}+\Delta-\hat{z})^2}{2} + \frac{(\bar{z}-\hat{z})^2 - (\underline{z}+\Delta-\hat{z})^2}{2} + \alpha_H\left[(\bar{z}-\hat{z})\Delta + \frac{\Delta^2}{2}\right].$$

Differentiating gives $N'(\hat{z}) = \alpha_L(\hat{z}-\underline{z}) - (\bar{z}-\underline{z}) = -D(\hat{z})$. Applying the quotient rule and using $D(\hat{z}) + \alpha_L(\hat{z}-\underline{z}) = \bar{z}-\underline{z}$,

$$g'(\hat{z}) = \frac{D(\hat{z})^2 + \alpha_H(\bar{z}-\underline{z})^2}{2\,D(\hat{z})^2} > 0.$$

Case 2: $\hat{z} \in [\underline{z}+\Delta, \bar{z}]$. Using (40)–(41) and differentiating with respect to $\hat{z}$,

$$g'(\hat{z}) = \frac{(\bar{z}-\hat{z})^2/2 + \alpha_H\Delta\big(\bar{z}-\hat{z}+\Delta/2\big)}{(\bar{z}-\hat{z}+\alpha_H\Delta)^2}.$$

The numerator equals $\frac{1}{2}\big[(\bar{z}-\hat{z})^2 + 2\alpha_H\Delta(\bar{z}-\hat{z}) + \alpha_H\Delta^2\big]$, which is always positive: its discriminant as a quadratic in $(\bar{z}-\hat{z})$ is $(2\alpha_H\Delta)^2 - 4\alpha_H\Delta^2 = -4\alpha_H\alpha_L\Delta^2 < 0$.

Case 3: $\hat{z} \in (\bar{z}, \bar{z}+\Delta]$. Here $d(\hat{z})/(1-K(\hat{z})) = 1/\alpha_H$, so $g(\hat{z}) = \alpha_L\Delta + \nu - (\bar{z}+\Delta-\hat{z})/2$ and $g'(\hat{z}) = 1/2 > 0$.

In all three cases g is strictly increasing, so $g(\hat{z}) > g(z_h) \geq 0$ for all $\hat{z} > z_h$. Since $1-K(\hat{z}) \geq 0$, we conclude $G(\hat{z}) = \big(1-K(\hat{z})\big)g(\hat{z}) \geq 0$ for all $\hat{z} \in [z_h, \bar{z}+\Delta]$.

Condition 3. The argument is symmetric to Condition 2.

Condition 1. Note that $f(z)$ and $f(z-\Delta)$ are step functions that are either 0 or $1/(\bar{z}-\underline{z})$. Thus,

$$h(z) \;=\; (\nu-\alpha_H\Delta)\alpha_L f(z)+(\nu+\alpha_L\Delta)\alpha_H f(z-\Delta)+K(z)$$

is strictly increasing wherever it is continuous because $K(z)$ is strictly increasing. It remains to check that the jumps at $z=\underline{z}+\Delta$ and $z=\bar{z}$ are non-negative whenever these points lie in (z_l,z_h).

Jump at $\bar{z}$: When $f(z)$ switches from $1/(\bar{z}-\underline{z})$ to 0, the jump in h equals $\alpha_L(\alpha_H\Delta-\nu)/(\bar{z}-\underline{z})$ and it is negative only if $\nu-\alpha_H\Delta>0$, or the effective bias of the θ_L policymaker is positive. Next, note that $\bar{z}$ lies in (z_l,z_h) only when $z_h>\bar{z}$ and so the cap is not binding for the θ_L type. Evaluating the left-hand side minus the right-hand side of the first-order condition (7) at $\hat{z}=\bar{z}$ gives

$$\begin{aligned}\left(1-K\left(\bar{z}\right)\right)\left(\alpha_L\Delta+\nu-\Delta/2\right)=G\left(\bar{z}\right)&=\left(1-K\left(\bar{z}\right)\right)g\left(\bar{z}\right)\\&<\left(1-K\left(\bar{z}\right)\right)g\left(z_h\right)=0\end{aligned}$$

where we used that $g'(\hat{z})>0$ for $z_h>\bar{z}$ as we showed above, $K(\bar{z})<1$, and $g(z_h)=0$. Thus,

$$\nu<\Delta/2-\alpha_L\Delta<\Delta-\alpha_L\Delta=\alpha_H\Delta$$

So, if $z_h>\bar{z}$ then $\nu<\alpha_H\Delta$. Hence the jump is non-negative.

Jump at $\underline{z}+\Delta$: When $f(z-\Delta)$ switches from 0 to $1/(\bar{z}-\underline{z})$, the jump equals $(\nu+\alpha_L\Delta)\alpha_H/(\bar{z}-\underline{z})$. This point lies in (z_l,z_h) only when $z_l<\underline{z}+\Delta$; by the symmetric argument for Condition 3, this requires $\nu\geq-\alpha_L\Delta$, so the jump is non-negative. $Q.E.D.$

A.3 Proof of Lemma 2

For a given Δ, the upper bound of the delegation set $\bar{\pi}$ depends on the bias ν, as well as on the mean-residual-life $E_\kappa[\hat{z}|\hat{z}\geq z]-z$ and the function $d(z)\equiv F(z)-F(z-\Delta)$. Lemma 1 shows that if there is $z_h\in(z_l,\bar{z}+\Delta)$ such that $\pi_f(z_h,\theta_L)=\bar{\pi}$, then it solves equation (7). If, instead, the bias is negative and large in the sense that $\nu<-\Delta\alpha_L$, then (7) does not have a solution in $[z_l,\bar{z}+\Delta]$ and the upper bound is not constraining because $z_h=\bar{z}+\Delta$. If the bias is positive and large in the sense that $\Delta\alpha_L\alpha_H d(z)+\nu(1-K(z))>(E_\kappa[\hat{z}|\hat{z}\geq z]-z)(1-K(z))$ for $z\in(z_l,\bar{z}+\Delta)$, then $\bar{\pi}=\underline{\pi}=\bar{\pi}^*$.

For a given ν, there exists a $\Delta(\nu)$ large enough such that the optimal delegation set is a single point. To see this, let $\Delta(\nu)$ be the smallest value of Δ such that $\bar{z}+\theta_L+\nu<\underline{z}+\bar{\theta}$ and $\underline{z}+\theta_H+\nu>\bar{z}+\bar{\theta}$ so that the preferred range of policies for type θ_H (θ_L) is always above (below) that for society. In this case, granting any flexibility will result in type θ_H

choosing the largest possible policy and type θ_L choosing the lowest possible one. Society strictly prefers the best constant policy over this. $Q.E.D.$

B Proofs of results in Section 4

B.1 Proof of Proposition 2

The proof of Proposition 2 follows as a corollary to the following Lemma.

Lemma 8. *Given some $z_a < z_b$, suppose that for all $z \in [z_a, z_b]$ we have*

$$(-\nu + \alpha_H \Delta)\, \alpha_L f(z) + (\alpha_L \Delta + \nu)\, \alpha_H f(z - \Delta) > 0. \tag{42}$$

Then, the delegation set does not contain intervals in $[\pi_f(z_a, \theta_L), \pi_f(z_b, \theta_L)]$.

Proposition 2 follows from this Lemma because if the bias is small, $\nu \in (-\alpha_L \Delta, \alpha_H \Delta)$, then (42) holds for all z. We next use this Lemma to characterize the case when the bias is not small.

Proof. Suppose, by way of contradiction, that the delegation set contains an interval that includes $[\pi_f(z_a, \theta_L), \pi_f(z_b, \theta_L)]$. Without loss, suppose $z_b - z_a < \Delta$.

Consider an alternative set $\hat{D}$ obtained by removing the interval $(\pi_f(z_a, \theta_L), \pi_f(z_b, \theta_L))$ from D, i.e., "drilling a hole" in D. Formally,

$$\hat{D}(z_b, z_a) \equiv D \setminus (\pi_f(z_a, \theta_L), \pi_f(z_b, \theta_L)).$$

We prove this lemma in three steps. First, we characterize the relevant outside option associated with $\hat{D}$. Second, we characterize the chosen policies under LB given the outside options from step 1. Finally, in step 3, we show that societal welfare is larger under $\hat{D}$ relative to D.

Step 1: Characterizing the relevant outside option associated with $\hat{D}$.

The set of available policies through bargaining $\mathcal{R}(z, \theta, \hat{D})$ depends on the outside option the executive θ has in $\hat{D}$. For $z \in [z_a, z_b]$ and θ_L, the relevant set of policies to choose from is the discrete set $\{\pi_f(z_a, \theta_L), \pi_f(z_b, \theta_L)\}$. Hence, there exists $z^* \in (z_a, z_b)$,

$$\pi_{\hat{D}}(z, \theta_L) = \begin{cases} \pi_D(z, \theta_L) & \text{for } z \leq z_a, \\ \pi_f(z_a, \theta_L) & \text{for } z \in [z_a, z^*], \\ \pi_f(z_b, \theta_L) & \text{for } z \in (z^*, z_b], \\ \pi_D(z, \theta_L) & \text{for } z \geq z_b. \end{cases}$$

The threshold z^* is such that θ_L is indifferent between $\pi_f(z_b, \theta_L)$ and $\pi_f(z_a, \theta_L)$, that is

$$z^* = \frac{z_a + z_b}{2}. \tag{43}$$

Step 2: Characterizing the chosen policies under LB.

We first characterize the policy chosen when θ_L is in power. We will show that for $z \in (z_a, z_b)$ the solution to the bargaining problem (9) is

$$\begin{cases} \pi_f(z, \theta_L) & \text{for } z \in [z_a, z^*], \\ \pi_f(z_b, \theta_L) & \text{for } z \in [z^*, z_b]. \end{cases}$$

To see this, first notice that for $z \in (z_a, z^*)$ the relevant outside option for θ_L is

$$\pi_{\hat{D}}(z, \theta_L) = \pi_f(z_a, \theta_L) < \pi_f(z, \theta_L).$$

Thus, the θ_H policymaker prefers $\pi_f(z, \theta_L)$ to the outside option $\pi_f(z_a, \theta_L)$. Hence, the participation constraint in (9) is slack and the solution is $\pi_f(z, \theta_L)$ for $z \in (z_a, z^*)$.

Next consider $z \in (z^*, z_b)$. In this case, the relevant outside option is

$$\pi_{\hat{D}}(z, \theta_L) = \pi_f(z_b, \theta_L) = \pi_f(z_b - \Delta, \theta_H) < \pi_f(z_a, \theta_H) < \pi_f(z, \theta_H)$$

where the first inequality follows from considering the case $z_b - z_a < \Delta$. Hence, for $z \in (z^*, z_b)$ and $\pi < \pi_{\hat{D}}(z, \theta_L)$,

$$u(\pi, z, \theta_H) < u(\pi_{\hat{D}}(z, \theta_L), z, \theta_H).$$

Therefore, $\pi \notin \mathcal{R}(z, \theta_L, \hat{D})$ and the solution in (9) for $z \in (z^*, z_b)$ is $\pi_{\hat{D}}(z, \theta_L)$.

Next, we characterize the policy chosen when θ_H is in power. We will show that for $z \in (z_a - \Delta, z_b - \Delta)$ the solution to the bargaining problem (9) is

$$\begin{cases} \pi_f(z_a - \Delta, \theta_H) & \text{for } z \in [z_a - \Delta, z^* - \Delta], \\ \pi_f(z, \theta_H) & \text{for } z \in [z^* - \Delta, z_b - \Delta]. \end{cases}$$

To see this, first notice that for $z \in (z^* - \Delta, z_b - \Delta)$,

$$\pi_{\hat{D}}(z, \theta_H) = \pi_f(z_b - \Delta, \theta_H) > \pi_f(z, \theta_H),$$

and hence, because θ_L accepts θ_H's offer, the solution to the bargaining problem (9) for $z \in (z^* - \Delta, z_b - \Delta)$ is $\pi_f(z, \theta_H)$.

Next, for $z \in (z_a - \Delta, z^* - \Delta)$,

$$\pi_{\hat{D}}(z, \theta_H) = \pi_f(z_a - \Delta, \theta_H) = \pi_f(z_a, \theta_L) > \pi_f(z_b - \Delta, \theta_L) \geq \pi_f(z - \Delta, \theta_L),$$

where the first inequality follows from considering the case $z_b - z_a < \Delta$. Hence, for $z \in (z_a - \Delta, z^* - \Delta)$ and $\pi > \pi_{\hat{D}}(z, \theta_H)$,

$$u(\pi, z, \theta_L) < u(\pi_{\hat{D}}(z, \theta_H), z, \theta_L)$$

and $\pi \notin \mathcal{R}(z, \theta_H, \hat{D})$. The solution to (9) for $z \in (z_a - \Delta, z^* - \Delta)$ is $\pi_f(z_a - \Delta, \theta_H)$.

Step 3: Showing that societal welfare is larger under $\hat{D}$ relative to D for $z_b > z_a$.

Notice that the welfare implications of delegating $\hat{D}$ instead of D depend only on the solution to the bargaining problem (9) for $z \in [z_a, z_b]$ for θ_L and for $z \in [z_a - \Delta, z_b - \Delta]$ for θ_H. Then, the previous steps imply that the welfare effect of "drilling a hole" in D is

$$\begin{aligned}
\delta(z_b, z_a) = &\int_{z_a}^{z^*} \left[\frac{1}{2}\pi_f(z - \Delta, \theta_H)^2 - \frac{1}{2}\pi_f(z_a - \Delta, \theta_H)^2\right] \alpha_H f(z - \Delta) dz \\
&+ \int_{z_a}^{z^*} \left[\left(z - \Delta + \sum_i \alpha_i \theta_i\right) (\pi_f(z_a - \Delta, \theta_H) - \pi_f(z - \Delta, \theta_H))\right] \alpha_H f(z - \Delta) dz \\
&+ \int_{z^*}^{z_b} \left[\frac{1}{2}\pi_f(z, \theta_L)^2 - \frac{1}{2}\pi_f(z_b, \theta_L)^2\right] \alpha_L f(z) dz \\
&+ \int_{z^*}^{z_b} \left[\left(z + \sum_i \alpha_i \theta_i\right) [\pi_f(z_b, \theta_L) - \pi_f(z, \theta_L)]\right] \alpha_L f(z) dz. \qquad (44)
\end{aligned}$$

We fix z_a and consider the effect of varying z_b in $[z_a, z_a + \varepsilon]$. Thus, z_b indexes the size of the hole, $\hat{D}(z_b, z_a)$. We will show that $\delta(z_b, z_a)$ has a local minimum at $\delta(z_a, z_a) = 0$. This implies for z_b close enough to z_a, the welfare gain of $\hat{D}(z_b, z_a)$ relative to D is positive.

As a first step, let's take the derivative of (44) with respect to z_b and substitute $\pi_f(z_b, \theta_L) = z_b + \theta_L + \nu$, and $\pi_f(z - \Delta, \theta_H) = \pi_f(z, \theta_L)$ to obtain

$$\begin{aligned}
\frac{\partial \delta(z_b, z_a)}{\partial z_b} = &\alpha_H \frac{\partial z^*(z_b, z_a)}{\partial z_b} \left[\frac{1}{2}\pi_f(z^*, \theta_L)^2 - \frac{1}{2}\pi_f(z_a, \theta_L)^2\right] f(z^* - \Delta) \qquad (45) \\
&+ \alpha_H \frac{\partial z^*(z_b, z_a)}{\partial z_b} \left[(z^* + \theta_L - \alpha_L \Delta)(\pi_f(z_a, \theta_L) - \pi_f(z^*, \theta_L))\right] f(z^* - \Delta) \\
&- \alpha_L \frac{\partial z^*(z_b, z_a)}{\partial z_b} \left[\frac{1}{2}\pi_f(z^*, \theta_L)^2 - \frac{1}{2}\pi_f(z_b, \theta_L)^2\right] f(z^*) \\
&- \alpha_L \frac{\partial z^*(z_b, z_a)}{\partial z_b} \left[(z^* + \theta_L + \alpha_H \Delta)(\pi_f(z_b, \theta_L) - \pi_f(z^*, \theta_L))\right] f(z^*) \\
&+ \alpha_L \frac{\partial \pi_f(z_b, \theta_L)}{\partial z} \int_{z^*}^{z_b} [z - z_b - \nu + \alpha_H \Delta] f(z) dz.
\end{aligned}$$

Since u is quadratic we also have that $\frac{\partial \pi_f(z_b,\theta_L)}{\partial z}=1$. The implicit function theorem for the threshold z^* defined in (43) implies that

$$\frac{\partial z^*(z_b,z_a)}{\partial z_b}=\frac{\partial \pi_f(z_b,\theta_L)}{\partial z}\frac{z_b-z^*}{\pi_f(z_b,\theta_L)-\pi_f(z_a,\theta_L)}.$$

Therefore, we can write

$$\frac{\partial \delta(z_b,z_a)}{\partial z_b}=S(z_b,z_a)$$

where

$$S(z_b,z_a)\equiv \alpha_L\int_{z^*}^{z_b}\left[z-z_b-\nu+\alpha_H\Delta\right]f(z)dz+\frac{z_b-z^*(z_b,z_a)}{\pi_f(z_b,\theta_L)-\pi_f(z_a,\theta_L)}A\left(z_b,z_a\right)$$

where

$$\begin{aligned}A\left(z_b,z_a\right)\equiv&\alpha_H\left[\frac{1}{2}\pi_f(z^*,\theta_L)^2-\frac{1}{2}\pi_f(z_a,\theta_L)^2\right]f(z^*-\Delta)\\&+\alpha_H\left[\left(z^*+\theta_L-\alpha_L\Delta\right)\left(\pi_f(z_a,\theta_L)-\pi_f(z^*,\theta_L)\right)\right]f(z^*-\Delta)\\&-\alpha_L\left[\frac{1}{2}\pi_f(z^*,\theta_L)^2-\frac{1}{2}\pi_f(z_b,\theta_L)^2\right]f(z^*)\\&-\alpha_L\left[\left(z^*+\theta_L+\alpha_H\Delta\right)\left(\pi_f(z_b,\theta_L)-\pi_f(z^*,\theta_L)\right)\right]f(z^*).\end{aligned}$$

Taking the limit as z_b tends to z_a, and using L'Hôpital's rule for the term $\frac{z_b-z^*}{\pi_f(z_b,\theta_L)-\pi_f(z_a,\theta_L)}$, yields $\lim_{z_b\to z_a}\partial\delta(z_b,z_a)/\partial z_b=0$. Finally, we show that $\delta\left(z,z_a\right)$ is strictly convex at $z=z_a$. We have

$$\frac{\partial S(z_b,z_a)}{\partial z_b}|_{(z_a,z_a)}=\alpha_L\frac{\partial}{\partial z_b}\left[\int_{z^*}^{z_b}\left[z-z_b-\nu+\alpha_H\Delta\right]f(z)dz\right]+\frac{\partial}{\partial z_b}\left[\frac{z_b-z^*}{\pi_f(z_b,\theta_L)-\pi_f(z_a,\theta_L)}A\left(z_b,z_a\right)\right].$$

Examining the first term, we have

$$\left.\frac{\partial\int_{z^*}^{z_b}\left[z-z_b-\nu+\alpha_H\Delta\right]f(z)dz}{\partial z_b}\right|_{z_b=z_a}=\frac{1}{2}(\alpha_H\Delta-\nu)f(z_a).$$

For the second term we have

$$\left.\frac{\partial\frac{z_b-z^*}{\pi_f(z_b,\theta_L)-\pi_f(z_a,\theta_L)}A\left(z_b,z_a\right)}{\partial z_b}\right|_{z_b=z_a}=\left.\frac{z_b-z^*}{\pi_f(z_b,\theta_L)-\pi_f(z_a,\theta_L)}A_1(z_b,z_a)\right|_{z_b=z_a}=\frac{1}{2\frac{\partial\pi_f(z_a,\theta_L)}{\partial z}}A_1(z_a,z_a),$$

where we used $\lim_{z_b \to z_a} \frac{\partial \frac{z_b - z^*}{\pi_f(z_b,\theta_L) - \pi_f(z_a,\theta_L)}}{\partial z_b} A(z_b, z_a) = 0$, because $\frac{\partial \frac{z_b - z^*}{\pi_f(z_b,\theta_L) - \pi_f(z_a,\theta_L)}}{\partial z_b}$ converges as $z_b \to z_a$ and $A(z_a, z_a) = 0$. We have, using $\partial z^*/\partial z_b = 1/2$ and that each bracket vanishes at $z_b = z_a$ (where $z^* = z_a$),

$$\begin{aligned} A_1(z_a, z_a) = &\tfrac{1}{2}\alpha_H \left[\pi_f(z_a, \theta_L) - (z_a + \theta_L - \alpha_L \Delta)\right] f(z_a - \Delta) \\ &- \tfrac{1}{2}\alpha_L \left[(z_a + \theta_L + \alpha_H \Delta) - \pi_f(z_a, \theta_L)\right] f(z_a). \end{aligned}$$

Substituting $\pi_f(z_a, \theta_L) = z_a + \theta_L + \nu$, the common $z_a + \theta_L$ cancels and

$$A_1(z_a, z_a) = \tfrac{1}{2}\alpha_H \left(\nu + \alpha_L \Delta\right) f(z_a - \Delta) - \tfrac{1}{2}\alpha_L \left(\alpha_H \Delta - \nu\right) f(z_a).$$

And so

$$\begin{aligned} \frac{\partial S(z_a, z_a)}{\partial z_b} &= \tfrac{1}{2}\alpha_L(\alpha_H \Delta - \nu) f(z_a) + \tfrac{1}{2} A_1(z_a, z_a) \\ &= \left(\tfrac{1}{2} - \tfrac{1}{4}\right) \alpha_L(\alpha_H \Delta - \nu) f(z_a) + \tfrac{1}{4}\alpha_H(\nu + \alpha_L \Delta) f(z_a - \Delta) \\ &= \tfrac{1}{4}(\alpha_H \Delta - \nu)\alpha_L f(z_a) + \tfrac{1}{4}\left(\nu + \alpha_L \Delta\right)\alpha_H f(z_a - \Delta), \end{aligned}$$

which is positive given condition (42). Q.E.D.

Proposition 2 is a corollary of the above Lemma: if $\nu \in (-\alpha_L \Delta, \alpha_H \Delta)$, then (42) holds for all z, which implies that the delegation set does not contain intervals.

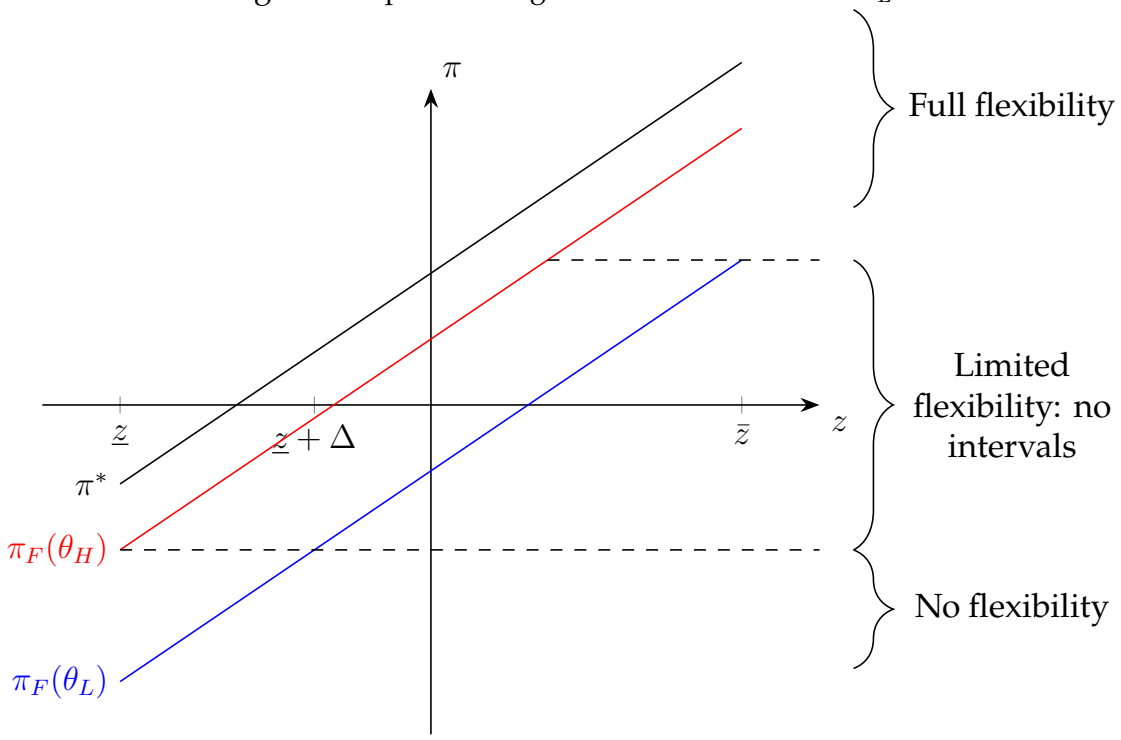
Next, we characterize the optimal set D if the bias is large.

Proposition 9. *If $\nu < -\alpha_L \Delta$ and for $z \in [\underline{z} + \Delta, \bar{z}]$ condition (42) holds, then the set D does not contain intervals in $[\pi_f(\underline{z} + \Delta, \theta_L), \pi_f(\overline{z}, \theta_L)]$ and grants full discretion for $\pi > \pi_f(\tilde{z}_1, \theta_L)$ where $\tilde{z}_1 \in [\overline{z}, \overline{z} + \Delta)$ and no discretion for $\pi < \pi_f(\underline{z} + \Delta, \theta_L)$. Similarly, if $\nu > \alpha_H \Delta$ and for $z \in [\underline{z}, \bar{z} - \Delta]$ condition (42) holds, then the set D does not contain intervals in $[\pi_f(\underline{z} + \Delta, \theta_L), \pi_f(\overline{z}, \theta_L)]$ and grants full discretion for $\pi < \pi_f(\tilde{z}_2, \theta_L)$, where $\tilde{z}_2 \in (\underline{z}, \underline{z} + \Delta]$, and no discretion for $\pi > \pi_f(\overline{z}, \theta_L)$.*

Proof. Suppose that $\nu < -\alpha_L \Delta$. The delegation set D takes the form outlined in Figure 7. If the condition (42) holds for $z \in [\underline{z} + \Delta, \bar{z}]$ then, from the previous lemma the set D does not contain intervals in $[\pi_f(\underline{z} + \Delta, \theta_L), \pi_f(\overline{z}, \theta_L)]$.

Suppose now that the set D contains some point $\tilde{\pi}$ from the interval $[\pi_f(\overline{z}, \theta_L), \pi_f(\overline{z} + \Delta, \theta_L)] = [\pi_f(\overline{z} - \Delta, \theta_H), \pi_f(\overline{z}, \theta_H)]$. We next show that granting full flexibility above $\tilde{\pi}$ is an improvement. Define $\tilde{D}$ to be $D \cup [\tilde{\pi}, \pi_f(\overline{z} + \Delta, \theta_L)]$. We show that societal welfare is higher under $\tilde{D}$. To see why note that if θ_H is in charge and the set is $\tilde{D}$, the chosen policy will correspond to its flexible best response in this new range. Since the equilibrium choices in LB are always in between the flexible policies of θ_H and θ_L, this change is preferable from society's perspective given the level of bias. Now suppose that θ_L is in charge. Since

Figure 7: Optimal delegation set in LB for $\nu < -\alpha_L \Delta$



$\tilde{\pi} > \pi_f(\bar{z}, \theta_L)$, the set $\tilde{D}$ does not affect the relevant outside option for the θ_L policymaker relative to D, so $\mathcal{R}(z, \theta_L, \tilde{D}) = \mathcal{R}(z, \theta_L, D)$ for all z thus leaving the chosen policies the same as if the delegation set is D.

Finally, we show that the optimal set D contains some $\tilde{\pi} \in [\pi_f(\bar{z}, \theta_L), \pi_f(\bar{z}+\Delta, \theta_L)]$. If not, consider adding any point in that interval to D. By the previous argument, we know that it is optimal to grant full flexibility above $\tilde{\pi}$, so consider the set $\tilde{D} = D \cup [\tilde{\pi}, \pi_f(\bar{z}+\Delta, \theta_L)]$. If type θ_H is in charge, the set $\mathcal{R}(z, \theta_H, D) \subset \mathcal{R}(z, \theta_H, \tilde{D})$ which strictly increases the set of policies feasible for the θ_H policymaker and moves the policy choice closer to society. If instead θ_L is in charge, all policies in $\tilde{D}$ above $\tilde{\pi}$ are never going to be the relevant outside option since $\tilde{\pi} \geq \pi_f(\bar{z}, \theta_L)$. Thus, there are two cases to consider. Suppose first that $\pi_{\tilde{D}}(z, \theta_L) \neq \tilde{\pi}$ for all z. Then $\mathcal{R}(z, \theta_L, D) = \mathcal{R}(z, \theta_L, \tilde{D})$ and so the chosen policy is the same. Second, suppose that $\pi_{\tilde{D}}(z, \theta_L) = \tilde{\pi}$ for some z. For such z, it must be that $\pi_D(z, \theta_L) < \pi_f(z, \theta_L)$. In fact, suppose by contradiction that $\pi_D(z, \theta_L) \geq \pi_f(z, \theta_L)$, since $\pi_D(z, \theta_L) < \pi_f(\bar{z}, \theta_L) < \tilde{\pi}$ then θ_L prefers $\pi_D(z, \theta_L)$ to $\tilde{\pi}$, a contradiction. Thus, it must be that $\pi_D(z, \theta_L) < \pi_f(z, \theta_L)$. Then $\pi_f(z, \theta_L) \in \mathcal{R}(z, \theta_L, D)$ and the θ_L policymaker can implement its preferred policy with D. Since the policy under LB is in $[\pi_f(z, \theta_L), \pi_f(z, \theta_H)]$ then the policy is moving (weakly) toward society's preferences for z such that $\pi_{\tilde{D}}(z, \theta_L) = \tilde{\pi}$. Thus, welfare is strictly improving under $\tilde{D}$.

A symmetric logic gives the result for $\nu > \alpha_H \Delta$. $Q.E.D.$

To understand this result, suppose that $\nu < -\alpha_L \Delta$, so that society's preferences are

always closer to those of faction θ_H. Then, the moderation that is desirable when the bias is small (see Proposition 2) is no longer desirable. In this situation, the best case for society is if type θ_H always implemented its flexible policy. However, society cannot just give full bargaining power to the faction in charge, since the low type would also choose its flexible policy when it is in charge. We show that for the range of policies which are chosen by both types, it is optimal to limit bargaining power by not delegating an interval. In contrast, for policies that are large enough so that they would never be chosen by the θ_L type, it is optimal to grant full bargaining power to the faction in charge. A symmetric logic holds for the case in which $\nu > \alpha_H \Delta$.

B.2 Proof of Lemma 3

Throughout, let $\sigma_z \equiv \text{sd}(z)$ denote the standard deviation of z, and define

$$\bar{\Delta}(\nu) = \max\left\{\frac{(\overline{z} - \underline{z}) + \sigma_z - \nu}{\alpha_L}, \frac{(\overline{z} - \underline{z}) + \sigma_z + \nu}{\alpha_H}\right\}.$$

For $\Delta > \bar{\Delta}(\nu)$ each faction's effective bias exceeds the spread of the shocks,

$$\min\{\nu + \alpha_L \Delta,\ \alpha_H \Delta - \nu\} > (\overline{z} - \underline{z}) + \sigma_z.$$

For any given ν, we first show that if Δ is large enough so that $\pi_f(\bar{z}, \theta_L) < \pi^*(\underline{z})$ and $\pi^*(\overline{z}) < \pi_f(\underline{z}, \theta_H)$ then implementing the best constant policy as an outcome is feasible. To see this, suppose that $D = \{\bar{\pi}^*\}$. Then $\mathcal{R}(z, \theta, D) = \{\bar{\pi}^*\}$ for any z, θ. This is because, for any z, $u(\pi, z, \theta_L) < u(\bar{\pi}^*, z, \theta_L)$ for any $\pi > \bar{\pi}^*$ since $\pi_f(z, \theta_L) < \bar{\pi}^*$, and $u(\pi, z, \theta_H) < u(\bar{\pi}^*, z, \theta_H)$ for any $\pi < \bar{\pi}^*$ since $\pi_f(z, \theta_H) > \bar{\pi}^*$. Hence neither faction can obtain the other's agreement to move away from $\bar{\pi}^*$.

We now show that this set D is optimal from society's perspective. We know from Proposition 2 that the set D does not contain intervals. Suppose first that $D \subset [\pi^*(\underline{z}), \pi^*(\overline{z})]$. Suppose by way of contradiction that $D \neq \{\bar{\pi}^*\}$. In particular, suppose that $D = \{\pi_1, \ldots, \pi_N\}$, with $\pi_1 < \ldots < \pi_N$. Then, the outcome under LB is $\pi_{LB}(z, \theta_H) = \pi_N$ and $\pi_{LB}(z, \theta_L) = \pi_1$, for all z. Clearly, this outcome is dominated by the best constant policy from society's perspective.

Second, consider adding one point to D outside the range of π^*. For example, suppose there is a $\pi_{N+1} > \pi^*(\bar{z})$. Adding this point does not change the policy implemented by the θ_L policymaker because their relevant outside options are in $D \backslash \{\pi_{N+1}\}$, but it can increase the bargaining power of the θ_H policymaker and reduce society's welfare.

The only remaining case is $D \cap [\pi^*(\underline{z}), \pi^*(\overline{z})] = \emptyset$, that is, every point of D lies outside the range of π^*. By the argument above that adding an additional outside option only

raises the bargaining power of one faction, it suffices to consider set D with a single point only. Without loss, take $D = \{\pi_o\}$ with $\pi_o > \pi^*(\overline{z})$; the case $\pi_o < \pi^*(\underline{z})$ is symmetric, with the roles of θ_H and θ_L interchanged.

Because D is a singleton, both factions are forced to the outside option π_o, and the outcome depends on how π_o compares to the high type's ideal $\pi_f(z, \theta_H) = z + \theta_H + \nu$.

Case 1: $\pi_o \leq \pi_f(z, \theta_H)$ for all z. Then neither faction can agree to a policy different from π_o. If θ_H is in power it would move up, but any $\pi > \pi_o$ is refused by θ_L, whose ideal policy lies below π_o; and it does not move down, since $\pi_o \leq \pi_f(z, \theta_H)$. If θ_L is in power it would move down, but any $\pi < \pi_o$ is refused by θ_H, since $\pi_o \leq \pi_f(z, \theta_H)$. Hence the implemented allocation is the constant π_o for both factions and all z. Since $\bar{\pi}^*$ is the best constant policy and $\pi_o > \pi^*(\overline{z}) \geq \bar{\pi}^*$, $D = \{\pi_o\}$ is strictly dominated by $D = \{\bar{\pi}^*\}$.

Case 2: $\pi_f(\underline{z}, \theta_H) < \pi_o < \pi_f(\overline{z}, \theta_H)$. Now π_o lies above the high type's ideal for low realizations of z and below it for high realizations. Let $z_o \equiv \pi_o - \theta_H - \nu \in (\underline{z}, \overline{z})$ denote the threshold at which $\pi_f(z_o, \theta_H) = \pi_o$. If θ_H is in power, it implements its ideal where that ideal lies below the outside option and is stuck at π_o otherwise: $\pi(z, \theta_H) = \pi_f(z, \theta_H)$ for $z < z_o$ and $\pi(z, \theta_H) = \pi_o$ for $z \geq z_o$. If θ_L is in power, for $z < z_o$ it can implement a lower policy than π_o that satisfies the high type's participation constraint, and it is stuck at π_o otherwise: $\pi(z, \theta_L) = 2\pi_f(z, \theta_H) - \pi_o$ for $z < z_o$ and $\pi(z, \theta_L) = \pi_o$ for $z \geq z_o$.

Because $\pi_o > \pi_f(\underline{z}, \theta_H) = \underline{z} + \theta_H + \nu$, every one of these policies lies above society's preferred policy $\pi^*(z)$. Letting $b_H \equiv \nu + \alpha_L \Delta = \pi_f(z, \theta_H) - \pi^*(z)$ denote the high type's effective bias, each implemented policy exceeds $\pi^*(z)$ by at least $b_H - (\overline{z} - \underline{z})$: the high type's ideal by exactly b_H; the flat policy π_o by $\pi_o - \pi^*(z) \geq \pi_f(\underline{z}, \theta_H) - \pi^*(\overline{z}) = b_H - (\overline{z} - \underline{z})$; and the renegotiated policy $2\pi_f(z, \theta_H) - \pi_o$ by $z + m$ with $m \equiv 2(\theta_H + \nu) - \bar{\theta} - \pi_o > b_H - \overline{z}$. Hence society's loss $\frac{1}{2}(\pi - \pi^*(z))^2$ is at least $\frac{1}{2}(b_H - (\overline{z} - \underline{z}))^2$ at every z and for either type in power, which exceeds the best-constant loss $\frac{1}{2}\sigma_z^2$ once Δ is large enough that $b_H - (\overline{z} - \underline{z}) > \sigma_z$. So $D = \{\pi_o\}$ is again strictly dominated by $D = \{\bar{\pi}^*\}$, and this holds regardless of the factions' population shares.

Case 3: Now π_o lies above the high type's ideal, $\pi_o > \pi_f(z, \theta_H)$ for all z. If θ_H is in power, the implemented policy is its own ideal policy, which the low type accepts: $\pi(z, \theta_H) = \pi_f(z, \theta_H) = z + \theta_H + \nu$. The deviation from $\pi^*(z) = z + \bar{\theta}$ is the constant effective bias $\theta_H + \nu - \bar{\theta} = \nu + \alpha_L \Delta$, so the welfare loss if θ_H is in power is $\frac{1}{2}(\nu + \alpha_L \Delta)^2$ at every z.

If instead θ_L is in power, the low type would move down toward its ideal policy subject to the high type participation constraint with

$$\pi(z, \theta_L) = 2\pi_f(z, \theta_H) - \pi_o = 2z + 2(\theta_H + \nu) - \pi_o.$$

Writing $m \equiv 2(\theta_H + \nu) - \bar{\theta} - \pi_o$, its deviation from $\pi^*(z)$ is $z + m$, with welfare loss

$\frac{1}{2}E\left[(z+m)^2\right] = \frac{1}{2}\left(\sigma_z^2 + (E[z]+m)^2\right)$.

Thus, the best constant policy is preferred if

$$\frac{1}{2}\sigma_z^2 < \alpha_H \frac{1}{2}(\nu + \alpha_L\Delta)^2 + \alpha_L \frac{1}{2}\left(\sigma_z^2 + (E[z]+m)^2\right)$$

or

$$0 < \alpha_H \frac{1}{2}\left[(\nu + \alpha_L\Delta)^2 - \sigma_z^2\right] + \alpha_L \frac{1}{2}(E[z]+m)^2$$

Our bound on Δ implies that $(\nu + \alpha_L\Delta)^2 > \sigma_z^2$ so the above inequality is satisfied.

In each of these cases $D = \{\pi_o\}$ is strictly dominated by $D = \{\bar{\pi}^*\}$. Together with the earlier cases and Proposition 2, this establishes that $D = \{\bar{\pi}^*\}$ is optimal. $Q.E.D.$

C Proof of Proposition 3

Part 1. Suppose that $\nu \in (-\alpha_L\Delta, \alpha_H\Delta)$. Then, we can write society's utility as a convex combination of the two policymaker types' utilities. In fact, let γ be defined by

$$\gamma\theta_H + (1-\gamma)\theta_L + \nu = \sum_i \alpha_i\theta_i$$

Note that if $\nu \in (-\alpha_L\Delta, \alpha_H\Delta)$, then $\gamma \in [0,1]$. With this definition, we can write

$$\begin{aligned} v(\pi, z) &= \left(z + \sum_i \alpha_i\theta_i\right)\pi - \frac{1}{2}\pi^2 \\ &= (z + \gamma\theta_H + (1-\gamma)\theta_L + \nu)\pi - \frac{1}{2}\pi^2 \\ &= \gamma u(\pi, z, \theta_H) + (1-\gamma)u(\pi, z, \theta_L), \end{aligned} \tag{46}$$

which is indeed a convex combination of policymaker preferences.

Next, consider an arbitrary delegation set D. Under DE, the equilibrium outcome is

$$\pi_D(z, \theta; D) \in \arg\max_{\pi \in D} u(\pi, z, \theta).$$

Under LB, the outcome solves

$$\pi_{LB}(z, \theta; D) \in \arg\max_{\pi} u(\pi, z, \theta)$$

subject to

$$u\left(\pi, z, \hat{\theta}\right) \geq u\left(\pi_D(z, \theta; D), z, \hat{\theta}\right) \quad \hat{\theta} \neq \theta.$$

Thus, it must be that for all z

$$u\left(\pi_{LB}\left(z,\theta;D\right),z,\theta\right) \geq u\left(\pi_{D}\left(z,\theta;D\right),z,\theta\right), \tag{47}$$

$$u\left(\pi_{LB}\left(z,\theta;D\right),z,\hat{\theta}\right) \geq u\left(\pi_{D}\left(z,\theta;D\right),z,\hat{\theta}\right) \quad \hat{\theta}\neq\theta, \tag{48}$$

where the first inequality is a strict inequality whenever there is an agreement to go outside the set D. We can then write the welfare for an arbitrary delegation set D as

$$\begin{aligned}\sum_{i}\alpha_{i}\int v\left(\pi_{D}\left(z,\theta_{i};D\right),z\right)f\left(z\right)dz &\leq \sum_{i}\alpha_{i}\int v\left(\pi_{LB}\left(z,\theta_{i};D\right),z\right)f\left(z\right)dz,\\ &\leq W_{LB}=\max_{D}\sum_{i}\alpha_{i}\int v\left(\pi_{LB}\left(z,\theta_{i};D\right),z\right)f\left(z\right)dz,\end{aligned}$$

where the first inequality follows from the fact that we can write v as a convex combination of policymakers' utility and (47)–(48), and the second inequality follows from optimality. Since the second inequality holds for any D, it must also hold for the delegation set that attains the maximum under DE. Thus, we have $W_{LB}\geq W_{DE}$.

To prove that the inequality is strict if either $\nu>-\left(\bar{z}-E[z]\right)+\alpha_{H}\Delta$ or $\nu<\left(E[z]-\underline{z}\right)-\alpha_{L}\Delta$, it is sufficient to show that the allocation under LB strictly dominates that under DE for any set D. Note that if the delegation set under DE contains an interval, then, from Proposition 2, drilling a hole in the interval and allowing for renegotiation, as we do in LB, strictly improves welfare. Therefore, it must be that $W_{LB}>W_{DE}$. If the delegation set under DE does not contain an interval, allowing for renegotiation strictly improves welfare because, as we show next, there is an agreement to renegotiate outside the set D for a non-negligible set of shocks z. We show this in two steps. First, suppose that D contains only one point which is the best constant policy. Our condition that either $\nu>-\left(\bar{z}-E[z]\right)+\alpha_{H}\Delta$ or $\nu<\left(E[z]-\underline{z}\right)-\alpha_{L}\Delta$ is equivalent to either $\pi_{f}(\bar{z},\theta_{L})>\bar{\pi}^{*}$ or $\pi_{f}(\underline{z},\theta_{H})<\bar{\pi}^{*}$. Thus, there exists a non-negligible set of shocks such that $\pi_{D}(z,\theta_{H})>\pi_{f}(z,\theta_{H})$ (or $\pi_{D}(z,\theta_{L})<\pi_{f}(z,\theta_{L})$), which implies that allowing for renegotiation strictly improves welfare. Second, suppose that D contains at least two points π_{1} and π_{2}. Without loss of generality, consider the case in which type θ_{H} uses at least two points. The switch from one policy to the other happens at the shock $z^{*}<\bar{z}$ that makes type θ_{H} indifferent between the two points, that is $u(\pi_{1},z^{*},\theta_{H})=u(\pi_{2},z^{*},\theta_{H})$. Since u is single-peaked, it must be that $\pi_{1}<\pi_{f}(z^{*},\theta_{H})<\pi_{2}$. Thus, there exists $\epsilon>0$ such that $\pi_{D}(z,\theta_{H})>\pi_{f}(z,\theta_{H})$ for $z\in[z^{*},z^{*}+\epsilon)$. This implies that allowing for renegotiation, as under LB, strictly improves welfare.

Part 2. Next, assume that $\nu>\text{sd}\left(z\right)+\alpha_{H}\Delta$. We will show that delegating a single point to the executive achieves higher welfare than any allocation that can be implemented via

LB. Consider the best constant policy, $\bar{\pi}^* = E[z] + \bar{\theta}$ where $\bar{\theta} = \sum_i \alpha_i \theta_i$. Under DE it is always feasible to implement the best constant policy so

$$W_{DE} \geq \int_{\underline{z}}^{\bar{z}} v(\bar{\pi}^*, z) f(z) dz.$$

Moreover, when $\nu > \alpha_H \Delta$, $v(\pi_f(z, \theta_L), z)$ is an upper bound for social welfare under legislative bargaining, since the equilibrium policy is in $[\pi_f(z, \theta_L), \pi_f(z, \theta_H)]$ and society's preferences are more aligned with those of the θ_L faction. Thus,

$$\int_{\underline{z}}^{\bar{z}} v(\pi_f(z, \theta_L), z) f(z) dz \geq W_{LB}.$$

When $\nu > \text{sd}(z) + \alpha_H \Delta$, direct calculations show that society's value of following the constant policy is higher than that of following type θ_L's preferred policy. To see this, notice that

$$\begin{aligned}
\int_{\underline{z}}^{\bar{z}} v(\pi_f(z, \theta_L), z) f(z) dz &= -\frac{1}{2} \int_{\underline{z}}^{\bar{z}} \sum_i \alpha_i (\pi_f(z, \theta_L) - (z + \theta_i))^2 f(z) dz + \int \sum_i \alpha_i \frac{(z + \theta_i)^2}{2} \\
&= -\frac{1}{2} \int_{\underline{z}}^{\bar{z}} \sum_i \alpha_i (\theta_L - \theta_i + \nu)^2 f(z) dz + \int \sum_i \alpha_i \frac{(z + \theta_i)^2}{2} \\
&= -\frac{1}{2} \left(\nu^2 - 2\alpha_H \Delta \nu + \alpha_H \Delta^2\right) + \int \sum_i \alpha_i \frac{(z + \theta_i)^2}{2}
\end{aligned}$$

and

$$\begin{aligned}
\int_{\underline{z}}^{\bar{z}} v(\bar{\pi}^*, z) f(z) dz &= -\frac{1}{2} \int_{\underline{z}}^{\bar{z}} \sum_i \alpha_i \left(E[z] + \bar{\theta} - (z + \theta_i)\right)^2 f(z) dz + \int \sum_i \alpha_i \frac{(z + \theta_i)^2}{2} \\
&= -\frac{1}{2} \sum_i \alpha_i \left[\sigma_z^2 + \left(\bar{\theta} - \theta_i\right)^2\right] + \int \sum_i \alpha_i \frac{(z + \theta_i)^2}{2} \\
&= -\frac{1}{2} \left[\sigma_z^2 + \alpha_L \alpha_H \Delta^2\right] + \int \sum_i \alpha_i \frac{(z + \theta_i)^2}{2}
\end{aligned}$$

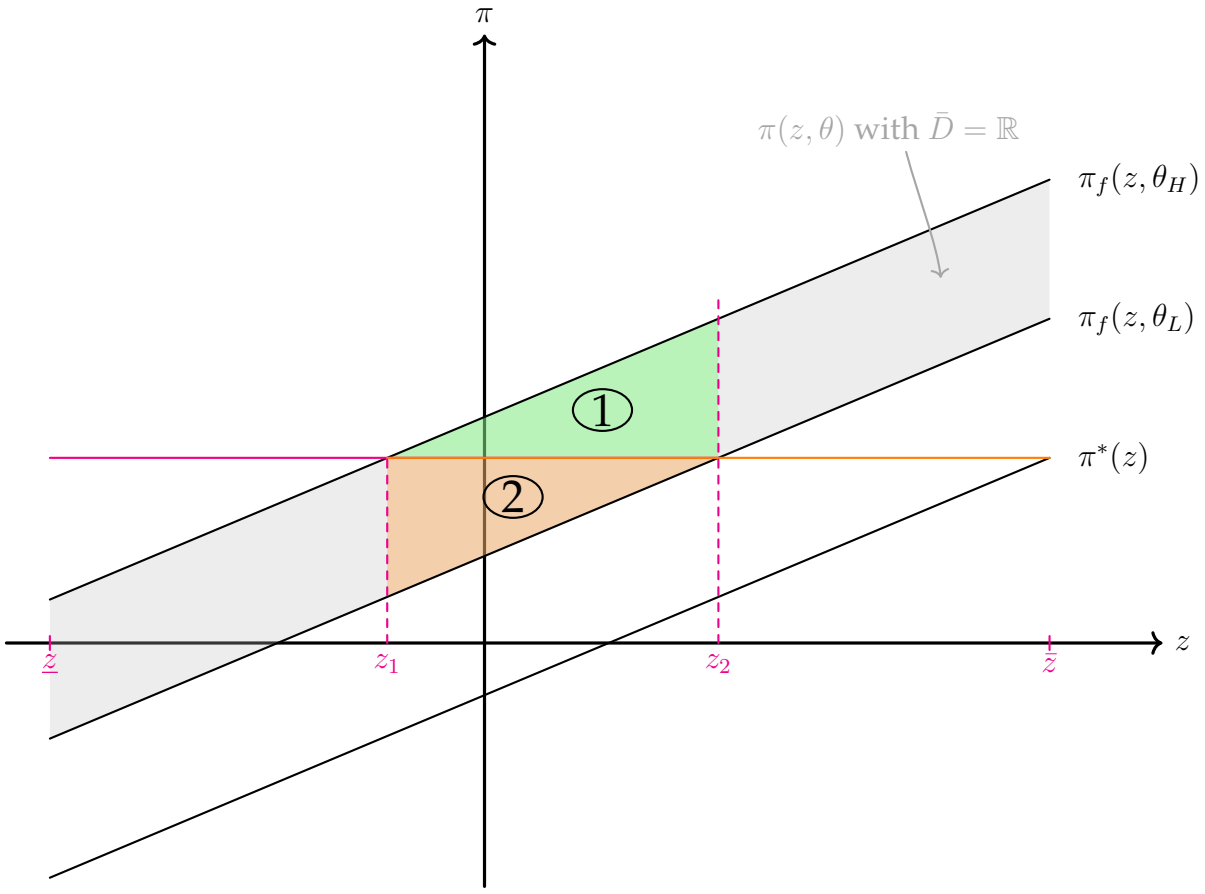
where $\sigma_z^2 = E\left[(E[z] - z)^2\right]$. Thus,

$$\int_{\underline{z}}^{\bar{z}} v(\bar{\pi}^*, z) f(z) dz > \int_{\underline{z}}^{\bar{z}} v(\pi_f(z, \theta_L), z) f(z) dz$$

iff $\sigma_z^2 < (\nu^2 - 2\alpha_H \Delta \nu + \alpha_H^2 \Delta^2) = (\nu - \alpha_H \Delta)^2$, that is, $\nu > \sigma_z + \alpha_H \Delta$. Thus, $W_{DE} > W_{LB}$. A symmetric argument holds for the case when $\nu < -\text{sd}(z) - \alpha_L \Delta$.

Part 3. Finally, suppose that there is no polarization, $\Delta = 0$. Under LB, factions can

Figure 8: Construction of the perturbation in the proof of Proposition 4



always implement their preferred policy, $\pi_{LB}(z,\theta;D)=\pi_f(z,\theta)$. This outcome is always feasible under DE since society can just grant full discretion to the policymaker. Then we have $W_{LB}\leq W_{DE}$.

We now show the inequality is strict whenever $\nu\neq 0$. Suppose $\nu>0$ and consider the DE mandate that caps the policy at $\pi_f(z_h,\theta)$ with $z_h=\bar{z}-\nu$ (or $z_h=\underline{z}$ if $\nu\geq\bar{z}-\underline{z}$), so the implemented policy is $\min\{\pi_f(z,\theta),\pi_f(z_h,\theta)\}$. Relative to full discretion, which attains the LB outcome, welfare changes by

$$W(z_h)-W(\bar{z})=\frac{1}{2}\int_{z_h}^{\bar{z}}\left[\nu^2-(z_h+\nu-z)^2\right]f(z)\,dz.$$

For every $z\in(z_h,\bar{z}]$ we have $0<z-z_h\leq\nu$ and hence $|z_h+\nu-z|<\nu$, so the integrand is strictly positive on the region of integration, and so $W(z_h)>W(\bar{z})$. Thus $W^{DE}>W^{LB}$. A symmetric argument holds for $\nu<0$. $Q.E.D.$

D Proof of Proposition 4

Proof. Suppose by way of contradiction that there are no restrictions on the bargaining set and $\bar{D} = \mathbb{R}$. The equilibrium policy then satisfies $\pi(z, \theta_i) \in [\pi_f(z, \theta_L), \pi_f(z, \theta_H)]$. Consider the following alternative pair $\left(D', \bar{D}'\right)$ where

$$\bar{D}' = \text{range}(\pi^*) = [\pi^*(\underline{z}), \pi^*(\bar{z})]$$
$$D' = D \cap \bar{D}'$$

Define z_1 and z_2 as

$$\pi^*(\bar{z}) = \pi_f(z_1, \theta_H) \quad \pi^*(\bar{z}) = \pi_f(z_2, \theta_L)$$

as in Figure 8.

Consider $z \in [z_2, \bar{z}]$. Facing $\left(D', \bar{D}'\right)$, both factions' ideal policies are higher than both the largest outside option in D' and $\pi^*(\bar{z})$. Thus, they will agree to choose $\pi'(z, \theta_i) = \pi^*(\bar{z})$ for all $z \in [z_2, \bar{z}]$. Since $\pi'(z, \theta_i) = \pi^*(\bar{z}) \leq \pi(z, \theta_L) \leq \pi(z, \theta_i)$, society's utility (weakly) goes up under the alternative sets in this region.

Consider next $z \in [z_1, z_2]$. If the relevant outside option for faction θ_i when facing D is also in $D' \subseteq D$, then $\pi_D(z, \theta_i) = \pi_{D'}(z, \theta_i)$ and

$$\pi^*(z) < \pi_f(z, \theta_i) \leq \pi'(z, \theta_i) = \min\{\pi(z, \theta_i), \pi^*(\bar{z})\} \leq \pi(z, \theta_i)$$

Thus, societal welfare (weakly) goes up because $\pi'(z, \theta_i)$ is closer to society's preferences.

Suppose next that the relevant outside option for faction θ_H when facing D is not in $D' \subseteq D$. Thus, under the original set we have that $\pi_o = \pi_D(z, \theta_H) > \pi^*(\bar{z}) > \pi_f(z, \theta_L)$ for $z \in [z_1, z_2]$. The θ_L policymaker then wants to deviate down. Thus, $\pi(z, \theta_H)$ is either such outside option π_o or $\pi_f(z, \theta_H)$ if also the θ_H policymaker wants to deviate down. Thus, $\pi(z, \theta_H) > \pi^*(\bar{z})$. Since the new policy under $\left(D', \bar{D}'\right)$ must be $\pi'(z, \theta_H) \leq \pi^*(\bar{z})$ and

$$\pi^*(z) < \pi_f(z, \theta_L) \leq \pi'(z, \theta_H) \leq \pi^*(\bar{z}) < \pi(z, \theta_H)$$

Thus, societal welfare goes up because $\pi'(z, \theta_H)$ is closer to society's preferences than $\pi(z, \theta_H)$. (Move from triangle (1) to triangle (2) in Figure 8.)

Suppose next that the relevant outside option for faction θ_L when facing D is not in $D' \subseteq D$ so that $\pi_o = \pi_D(z, \theta_L) > \pi^*(\bar{z})$. This implies that the best outside option in D', $\pi'_o = \pi_{D'}(z, \theta_L)$, is lower than $\pi_f(z, \theta_L)$. To see this, suppose by way of contradiction that

$$\pi_o > \pi^*(\bar{z}) > \pi'_o > \pi_f(z, \theta_L)$$

Then π'_o is closer to θ_L's ideal policy and attains a higher value than π_o, contradicting that

π_o is optimal in D since $\pi_o' \in D$. Thus, it must be that

$$\pi_o > \pi^*(\bar{z}) > \pi_f(z, \theta_L) > \pi_o'$$

Since both types want to increase π from π_o' then $\pi'(z, \theta_L) = \pi_f(z, \theta_L)$ under $\left(D', \bar{D}'\right)$ which is a weak improvement over $\pi(z, \theta_L) \in [\pi_f(z, \theta_L), \pi_f(z, \theta_H)]$.

For $z \leq z_1$, the proposed perturbation has no effect and $\pi'(z, \theta) = \pi(z, \theta)$.

Thus, the proposed variation strictly increases welfare, contradicting that $(D, \mathbb{R})$ is optimal.

Moreover, we know that lowering the cap from $\pi^*(\bar{z})$ to $\pi^*(\bar{z}) - \varepsilon$ is beneficial. We can see this either by replicating the same argument we just did or by picking ε small enough so that the resulting set of outside options $D'' = D' \cap [\pi^*(\underline{z}), \pi^*(\bar{z}) - \varepsilon] = D'$ so this perturbation only affects the bunching for high z and it improves welfare. $Q.E.D.$

E Proofs of results in Section 7

We prove Proposition 5 in two steps. First, we show that there exists $U_L(\Delta)$ such that if $U \leq U_L(\Delta)$, then LB is preferred to DE. Second, we show that if the inflation bias is sufficiently large, then delegating to executives dominates legislative bargaining.

LB preferred to DE if U small

Proposition 10. *If $U \leq U_L(\Delta) = \frac{1}{2}\alpha_H \Delta$, then society prefers LB to DE.*

Proof. Let D denote a delegation set for executives. Let $\pi_D(z, \theta)$ denote the allocation implemented by delegating D to executives, and x_D the inflation expectations associated with $\pi_D(z, \theta)$. Let $\pi_{LB}(z, \theta)$ and x_{LB} denote the allocation and associated inflation expectations implemented by delegating $D + \{\frac{x_{LB} - x_D}{2}\}$ to the legislature. (A fixed-point argument guarantees that such x_{LB} exists because best responses have slope $1/2 < 1$ with respect to x.)

The proof consists of showing that welfare is higher for $\pi_{LB}(z, \theta)$ with x_{LB} than for $\pi_D(z, \theta)$ with x_D. The proof proceeds in three steps. The first step bounds the welfare loss associated with inflation expectations x_{LB} instead of inflation expectations x_D, keeping the discretion granted to executives constant. The second step bounds the welfare gain from renegotiation through legislative bargaining, keeping the inflation expectations constant. The third step shows that the condition $U \leq \frac{1}{2}\alpha_H \Delta$ is sufficient for the welfare gain due to renegotiation to outweigh the welfare loss due to the change in inflation expectations.

Step 1: We evaluate the welfare change due to inflation expectations changing from x_D to x_{LB}. The shift in inflation expectations induces a change in the allocation implemented

by delegating to executives (although inflation expectations x_{LB} are not rational at the allocation considered, this is a useful intermediate step in evaluating the change in welfare in going from DE to LB). Because

$$\pi_f(z,\theta;x_{LB}) = \pi_f(z,\theta;x_D) + \frac{x_{LB}-x_D}{2},$$

consider the delegation set $\hat{D} = D + \{\frac{x_{LB}-x_D}{2}\}$. Intuitively, shifting the delegation set keeps the discretion granted to executives constant at the new inflation expectations. As a result, the allocation implemented by delegating $\hat{D}$ to executives, with inflation expectations x_{LB}, is a vertical shift of the allocation implemented by delegating D with inflation expectations x_D. That is,

$$\pi_{\hat{D},x_{LB}}(z,\theta) = \pi_D(z,\theta) + \frac{x_{LB}-x_D}{2}.$$

We next show that the change in welfare from the allocation $\pi_D(z,\theta)$ with x_D to the allocation $\pi_{\hat{D},x_{LB}}(z,\theta)$ with x_{LB} is

$$-(U + x_D - \bar{\theta})\frac{x_{LB}-x_D}{2} - \left(\frac{x_{LB}-x_D}{2}\right)^2.$$

To see this, let the path from x_D to x_{LB} be parametrized by $t \in [0,1]$ and

$$x(t) = t x_{LB} + (1-t)x_D.$$

Similarly, let the path from $\pi_D(z,\theta)$ to $\pi_{\hat{D},x_{LB}}(z,\theta)$ be parametrized by $t \in [0,1]$ as follows

$$\pi_t(z,\theta) = t\pi_{\hat{D},x_{LB}}(z,\theta) + (1-t)\pi_D(z,\theta).$$

The change in welfare is the integration of marginal changes along the path parametrized by t,

$$\hat{w}(x(1),\pi_{t=1}(\cdot,\cdot)) - \hat{w}(x(0),\pi_{t=0}(\cdot,\cdot)) = \int_0^1 \frac{d\hat{w}(x(t),\pi(\cdot,\cdot))}{dt}dt,$$

where $\hat{w}$ denotes society's expected welfare. Faction $i \in \{H,L\}$ is in power with probability α_i, hence

$$\hat{w}(x,\pi(\cdot,\cdot)) = \alpha_L\hat{\nu}(x,\pi(\cdot,\cdot),\theta_L) + \alpha_H\hat{\nu}(x,\pi(\cdot,\cdot),\theta_H),$$

where

$$\hat{\nu}(x,\pi(\cdot,\cdot),\theta) = \int_{\underline{z}}^{\bar{z}} \left[R\left(x,\pi(z,\theta),z,\bar{\theta}\right)\right] dF(z).$$

The marginal change in welfare is the Gateaux derivative in the direction $h(z,\theta) =$

$\pi_{\hat{D},x_{LB}}(z,\theta) - \pi_D(z,\theta)$ and $h_x = x_{LB} - x_D$,

$$\frac{d\hat{w}(x(t),\pi_t(\cdot,\cdot))}{dt} = \alpha_L \partial\hat{\nu}(x(t),\pi_t(z,\theta),\theta_L; h_x, h(\cdot,\cdot)) + \alpha_H \partial\hat{\nu}(x(t),\pi_t(z,\theta),\theta_H; h_x, h(\cdot,\cdot)),$$

where

$$\begin{aligned}\partial\hat{\nu}(x,\pi(\cdot,\cdot),\theta; h_x, h(\cdot,\cdot)) = &\int_{\underline{z}}^{\bar{z}} \left[\frac{\partial R(x,\pi(z,\theta),z,\bar{\theta})}{\partial x} h_x\right] dF(z) \\ &+ \int_{\underline{z}}^{\bar{z}} \left[\frac{\partial R(x,\pi(z,\theta),z,\bar{\theta})}{\partial \pi} h(z,\theta)\right] dF(z).\end{aligned}$$

Using $\frac{\partial R(x,\pi,z,\bar{\theta})}{\partial x} = -(U + x - \pi)$, and $\frac{\partial R(x,\pi(z,\theta),z,\bar{\theta})}{\partial \pi} = (U + x - \pi) - (\pi - z - \bar{\theta})$, it follows that

$$\partial\hat{\nu}(x(t),\pi_t(z,\theta),\theta; h_x, h(\cdot,\cdot)) = -\left[\frac{U + x_D - \bar{\theta}}{2} + t\frac{x_{LB} - x_D}{2}\right](x_{LB} - x_D),$$

given that $h(z,\theta) = \frac{x_{LB} - x_D}{2}$ and $h_x = x_{LB} - x_D$.

As a result,

$$\hat{w}(x(1),\pi_{t=1}(\cdot,\cdot)) - \hat{w}(x(0),\pi_{t=0}(\cdot,\cdot)) = -\frac{U + x_D - \bar{\theta}}{2}(x_{LB} - x_D) - \left(\frac{x_{LB} - x_D}{2}\right)^2. \quad (49)$$

Step 2: We evaluate the welfare implications of allowing for renegotiation, while keeping inflation expectations fixed at x_{LB}. That is, the welfare change from $\pi_{\hat{D},x_{LB}}(z,\theta)$ to $\pi_{LB}(z,\theta)$, given inflation expectations x_{LB}.

The Gateaux derivative in the direction $h(z,\theta)$, keeping inflation expectations fixed, that is $h_x = 0$, is

$$\partial\hat{\nu}(x_{LB},\pi(\cdot,\cdot),\theta; h_x = 0, h(\cdot,\cdot)) = \int_{\underline{z}}^{\bar{z}} \left[\left(U + x_{LB} + z + \bar{\theta} - 2\pi(z,\theta)\right) h(z,\theta)\right] dF(z).$$

Let the path from $\pi_{\hat{D},x_{LB}}(z,\theta)$ to $\pi_{LB}(z,\theta)$ be parametrized by $t \in [0,1]$ as follows

$$\pi_t^r(z,\theta) = t\pi_{LB}(z,\theta) + (1-t)\pi_{\hat{D},x_{LB}}(z,\theta),$$

and let $h^r(z,\theta) = \pi_{LB}(z,\theta) - \pi_{\hat{D},x_{LB}}(z,\theta)$.

We are going to show that: If $h^r(z,\theta_L) > 0$, then

$$U + x_{LB} + z + \bar{\theta} - 2\pi_t^r(z,\theta_L) = 2(1-t)h^r(z,\theta_L) + \alpha_H\Delta > 0. \quad (50)$$

If $h^r(z,\theta_L) < 0$, then

$$U + x_{LB} + z + \bar{\theta} - 2\pi_t^r(z,\theta_L) = (1-2t)h^r(z,\theta_L) - \alpha_L\Delta. \tag{51}$$

Similarly, if $h^r(z,\theta_H) < 0$, then

$$U + x_{LB} + z + \bar{\theta} - 2\pi_t^r(z,\theta_H) = 2(1-t)h^r(z,\theta_H) - \alpha_L\Delta < 0. \tag{52}$$

Lastly, if $h^r(z,\theta_H) > 0$, then

$$U + x_{LB} + z + \bar{\theta} - 2\pi_t^r(z,\theta_H) = (1-2t)h^r(z,\theta_H) + \alpha_H\Delta. \tag{53}$$

To see these results, notice that if $h^r(z,\theta_L) > 0$ then, because $\theta_L \leq \theta_H$, the θ_L faction can attain its preferred policy, $\pi_{LB}(z,\theta_L) = \pi_f(z,\theta_L)$. Because R is concave and the marginal payoff for θ_L at $\pi_f(z,\theta_L)$ is zero, it follows that

$$U + x_{LB} - 2\pi_{\hat{D},x_{LB}}(z,\theta_L) + z + \theta_L = 2h^r(z,\theta_L) > 0.$$

More generally, for $t \in [0,1]$, because the marginal utility is linear in the policy, if $h^r(z,\theta_L) > 0$, then

$$U + x_{LB} - 2(\pi_{\hat{D},x_{LB}}(z,\theta_L) + th^r(z,\theta_L)) + z + \theta_L = 2(1-t)h^r(z,\theta_L).$$

Adding $\alpha_H\Delta$ to both sides of the equality gives (50).

Similarly, if $h^r(z,\theta_L) < 0$ instead, then $\pi_{\hat{D},x_{LB}}(z,\theta_L) > \pi_f(z,\theta_H)$. Because R is concave and the marginal payoff for θ_H at $\pi_f(z,\theta_H)$ is zero, it follows that

$$U + x_{LB} - 2\pi_{\hat{D},x_{LB}}(z,\theta_L) + z + \theta_H < 0.$$

Subtracting $\alpha_L\Delta$ from both sides of the inequality gives

$$U + x_{LB} - 2(\pi_{\hat{D},x_{LB}}(z,\theta_L) + th^r(z,\theta_L)) + z + \bar{\theta} = (1-2t)h^r(z,\theta_L) - \alpha_L\Delta$$

which is (51).

The argument for $h^r(z,\theta_H)$ is symmetric.

The change in welfare is the integration of marginal changes along the path parametrized by t,

$$\hat{w}(x_{LB},\pi_{t=1}^r(\cdot,\cdot)) - \hat{w}(x_{LB},\pi_{t=0}^r(\cdot,\cdot)) = \int_0^1 \frac{d\hat{w}(x_{LB},\pi_t^r(\cdot,\cdot))}{dt}dt.$$

Based on (50)–(53), we partition the range of shocks as follows: $Z(+,\theta) \equiv \{z \in Z \,|\, h^r(z,\theta) >$

$0\}$, $Z(-,\theta) \equiv \{z \in Z \mid h^r(z,\theta) < 0\}$, and $Z(=,\theta) \equiv \{z \in Z \mid h^r(z,\theta) = 0\}$, which gives

$$\partial\hat{\nu}(x_{LB}, \pi_t^r(\cdot,\cdot), \theta_L; h_x = 0, h^r(\cdot,\cdot)) = \int_{Z(+,\theta_L)} \left[(2(1-t)h^r(z,\theta_L) + \alpha_H\Delta)\, h^r(z,\theta_L)\right] dF(z) + \int_{Z(-,\theta_L)} \left[((1-2t)h^r(z,\theta_L) - \alpha_L\Delta)\, h^r(z,\theta_L)\right] dF(z).$$

Integrating over the path from $\pi_{\hat{D},x_{LB}}(z,\theta_L)$ to $\pi_{LB}(z,\theta_L)$,

$$\int_0^1 \partial\hat{\nu}(x_{LB}, \pi_t^r(\cdot,\cdot), \theta_L; h_x = 0, h^r(\cdot,\cdot))dt = \int_{Z(+,\theta_L)} \left[h^r(z,\theta_L)^2\right] dF(z) + \alpha_H\Delta \int_{Z(+,\theta_L)} \left[h^r(z,\theta_L)\right] dF(z) - \alpha_L\Delta \int_{Z(-,\theta_L)} \left[h^r(z,\theta_L)\right] dF(z).$$

Similarly for θ_H,

$$\partial\hat{\nu}(x_{LB}, \pi_t^r(\cdot,\cdot), \theta_H; h_x = 0, h^r(\cdot,\cdot)) = \int_{Z(+,\theta_H)} \left[((1-2t)h^r(z,\theta_H) + \alpha_H\Delta)\, h^r(z,\theta_H)\right] dF(z) + \int_{Z(-,\theta_H)} \left[(2(1-t)h^r(z,\theta_H) - \alpha_L\Delta)\, h^r(z,\theta_H)\right] dF(z).$$

And, integrating over the path from $\pi_{\hat{D},x_{LB}}(z,\theta_H)$ to $\pi_{LB}(z,\theta_H)$,

$$\int_0^1 \partial\hat{\nu}(x_{LB}, \pi_t^r(\cdot,\cdot), \theta_H; h_x = 0, h^r(\cdot,\cdot))dt = \int_{Z(-,\theta_H)} \left[h^r(z,\theta_H)^2\right] dF(z) + \alpha_H\Delta \int_{Z(+,\theta_H)} \left[h^r(z,\theta_H)\right] dF(z) - \alpha_L\Delta \int_{Z(-,\theta_H)} \left[h^r(z,\theta_H)\right] dF(z).$$

Step 3: We show that if $U \leq \frac{1}{2}\alpha_H\Delta$, then the benefits outlined in the second step outweigh the cost outlined in the first step and summarized in (49). That is, it suffices to show that

$$\alpha_H\Delta \sum_i \alpha_i \int_{Z(+,\theta_i)} \left[h^r(z,\theta_i)\right] dF(z) - \alpha_L\Delta \sum_i \alpha_i \int_{Z(-,\theta_i)} \left[h^r(z,\theta_i)\right] dF(z) + \sum_i \alpha_i \int_{\underline{z}}^{\bar{z}} \left[h^r(z,\theta_i)\right]^2 dF(z) \geq (U + x_D - \bar{\theta})\frac{x_{LB} - x_D}{2} + \left(\frac{x_{LB} - x_D}{2}\right)^2. \tag{54}$$

Using the implementability condition (10) we can write

$$\begin{aligned} x_{LB} - x_D &= \sum_i \alpha_i \int [\pi_{LB}(z,\theta_i) - \pi_D(z,\theta_i)] f(z) dz \\ &= \sum_i \alpha_i \int [\pi_{LB}(z,\theta_i) - \pi_{\hat{D},x_{LB}}(z,\theta_i) + \pi_{\hat{D},x_{LB}}(z,\theta_i) - \pi_D(z,\theta_i)] f(z) dz \\ &= \sum_i \alpha_i \int [h^r(z,\theta_i) + h(z,\theta_i)] f(z) dz. \end{aligned}$$

where the first equality follows from (10), the second from adding and subtracting $\pi_{\hat{D},x_{LB}}(z,\theta_i)$, and the last one uses the definition of h^r and h. Thus, the change in inflation expectations between the two regimes can be expressed in terms of h^r and h.

Because the set $\hat{D}$ is such that $h(z,\theta_i) = \frac{x_{LB}-x_D}{2}$, it follows that

$$\frac{x_{LB} - x_D}{2} = \sum_i \alpha_i \int [h^r(z,\theta_i)] f(z) dz, \tag{55}$$

and, by Jensen's inequality,

$$\left(\frac{x_{LB} - x_D}{2}\right)^2 \leq \sum_i \alpha_i \int [h^r(z,\theta_i)]^2 f(z) dz.$$

Hence, substituting the above into (54), to prove our result it suffices to show that

$$\begin{aligned} &\alpha_H \Delta \sum_i \alpha_i \int_{Z(+,\theta_i)} [h^r(z,\theta_i)]\, dF(z) - \alpha_L \Delta \sum_i \alpha_i \int_{Z(-,\theta_i)} [h^r(z,\theta_i)]\, dF(z) \\ &\geq (U + x_D - \bar{\theta}) \frac{x_{LB} - x_D}{2}. \end{aligned}$$

The second term on the left-hand side of the inequality subtracts a negative term because $h^r(z,\theta_i) < 0$ for $z \in Z(-,\theta_i)$. Thus, for (54) to hold it suffices to show that

$$\alpha_H \Delta \sum_i \alpha_i \int_{Z(+,\theta_i)} [h^r(z,\theta_i)]\, dF(z) \geq (U + x_D - \bar{\theta}) \frac{x_{LB} - x_D}{2}.$$

Note that

$$\sum_i \alpha_i \int_{Z(+,\theta_i)} [h^r(z,\theta_i)]\, dF(z) \geq \sum_i \alpha_i \int_{\underline{z}}^{\bar{z}} [h^r(z,\theta_i)]\, dF(z) = \frac{x_{LB} - x_D}{2},$$

where the last equality follows from (55). Thus, the following inequality is a sufficient

condition for the result to hold:

$$\alpha_H \Delta > U + x_D - \bar{\theta}.$$

We assumed that $U \leq \frac{1}{2}\alpha_H \Delta$, so for the above inequality to hold it is sufficient to show that $\alpha_H \Delta \geq 2U \geq U + x_D - \bar{\theta}$, or simply that $x_D \leq U + \bar{\theta}$. Suppose by way of contradiction that $x_D > U + \bar{\theta}$. This contradicts that π_D solves the mechanism design problem of delegation to executives: even in the Nash equilibrium (with full flexibility), expected inflation would be $U + \bar{\theta}$, and so the Nash equilibrium outcome would achieve higher welfare than π_D with inflation expectations equal to $U + \bar{\theta}$. But this is not possible since the Nash equilibrium outcome is feasible for the delegation problem. Thus, we have a contradiction; hence $x_D \leq U + \bar{\theta}$, implying that condition (54) holds. $Q.E.D.$

DE preferred to LB if U large

In contrast, if the inflation bias is sufficiently large, then delegating to executives dominates legislative bargaining.

Proposition 11. *If $\Delta = 0$ or $U \geq U_H(\Delta) \equiv \alpha_H \Delta + \sqrt{\frac{\alpha_H^2 \Delta^2 + \text{var}(z)}{2}}$, then society prefers DE to LB.*

Proof. First, suppose that $\Delta = 0$, so that the two factions have identical preferences. Under LB, the faction in power implements its flexible best response in every state. This outcome is feasible under DE: by granting full discretion, $D = \mathbb{R}$, society induces the executive to implement the same allocation as under LB. Thus, $V_{DE} \geq V_{LB}$ and society (weakly) prefers DE to LB for any $U \geq 0$. Next, suppose that $\Delta \neq 0$. We compare a lower bound for the welfare associated with DE to an upper bound for the welfare associated with LB. The welfare associated with DE is at least as high as the welfare resulting from delegating the narrow mandate $D = \{\bar{\pi}^*\}$, where $\bar{\pi}^* = \bar{\theta}$, to the monetary authority. Thus,

$$\begin{aligned} V_{DE} \geq \bar{V}^* &= \int \sum_i \alpha_i R\left(\bar{\pi}^*, \bar{\pi}^*, z, \theta_i\right) f(z)\, dz - \frac{1}{2}\left[U^2 + \sum_i \alpha_i \left(\bar{\theta} - z - \theta_i\right)^2\right] \\ &= -\frac{1}{2}\left[U^2 + \sum_i \alpha_i \left(\bar{\theta} - \theta_i\right)^2 + \text{var}(z)\right] \end{aligned}$$

To find an upper bound for the LB value, note that the LB outcome satisfies

$$\pi_f(z, \theta_L; x_{LB}) \leq \pi_{LB}(z, \theta) \leq \pi_f(z, \theta_H; x_{LB})$$

$$x_{LB} = \sum_i \alpha_i \int \pi_{LB}(z, \theta_i) f(z)\, dz$$

so $x_{LB} \geq x_M(\theta_L) = U + \theta_L$, where $x_M(\theta_L)$ is the expected inflation in the Markov equilibrium in which the θ_L faction always chooses policy. Thus,

$$\begin{aligned}
V_{LB} \leq \bar{V}_{LB} &= \int \sum_i \alpha_i R\left(x_M(\theta_L), \pi_f\left(z, \bar{\theta}; x_M(\theta_L)\right), z, \theta_i\right) f(z)\, dz \\
&= -\frac{1}{2}\int \sum_i \alpha_i \left[\left(U + x_M(\theta_L) - \frac{U + x_M(\theta_L) + z + \bar{\theta}}{2}\right)^2 + \left(\frac{U + x_M(\theta_L) + z + \bar{\theta}}{2} - z - \theta_i\right)^2\right] f(z)\, dz \\
&= -\frac{1}{2}\int \left[\left(U - \frac{\bar{\theta} - \theta_L + z}{2}\right)^2 + \sum_i \alpha_i \left(U + \frac{\bar{\theta} + \theta_L - 2\theta_i - z}{2}\right)^2\right] f(z)\, dz \\
&= -\frac{1}{2}\left[2U^2 + \frac{1}{4}\left(\bar{\theta} - \theta_L\right)^2 + \frac{1}{2}\text{var}(z) + \frac{1}{4}\sum_i \alpha_i \left(\bar{\theta} + \theta_L - 2\theta_i\right)^2 - 2U\left(\bar{\theta} - \theta_L\right)\right]
\end{aligned}$$

Thus, $\bar{V}_{LB} < \bar{V}^*$ if and only if

$$U^2 + \sum_i \alpha_i \left(\bar{\theta} - \theta_i\right)^2 + \text{var}(z) < 2U^2 + \frac{1}{4}\left(\bar{\theta} - \theta_L\right)^2 + \frac{1}{2}\text{var}(z) + \frac{1}{4}\sum_i \alpha_i \left(\bar{\theta} + \theta_L - 2\theta_i\right)^2 - 2U\left(\bar{\theta} - \theta_L\right)$$

$$\sum_i \alpha_i \left(\bar{\theta} - \theta_i\right)^2 + \frac{1}{2}\text{var}(z) < U^2 + \frac{1}{4}\left(\bar{\theta} - \theta_L\right)^2 + \frac{1}{4}\sum_i \alpha_i \left(\bar{\theta} + \theta_L - 2\theta_i\right)^2 - 2U\left(\bar{\theta} - \theta_L\right)$$

Note that, by adding and subtracting $2\bar{\theta}$ inside the square, we have

$$\begin{aligned}
\sum_i \alpha_i \left(\bar{\theta} + \theta_L - 2\theta_i\right)^2 &= \sum_i \alpha_i \left(\theta_L - \bar{\theta} + 2\left(\bar{\theta} - \theta_i\right)\right)^2 \\
&= \sum_i \alpha_i \left(\left(\theta_L - \bar{\theta}\right)^2 + 4\left(\bar{\theta} - \theta_i\right)^2 + 4\left(\theta_L - \bar{\theta}\right)\left(\bar{\theta} - \theta_i\right)\right) \\
&= \left(\theta_L - \bar{\theta}\right)^2 + 4\sum_i \alpha_i \left(\bar{\theta} - \theta_i\right)^2 + 4\left(\theta_L - \bar{\theta}\right)\sum_i \alpha_i \left(\bar{\theta} - \theta_i\right) \\
&= \left(\theta_L - \bar{\theta}\right)^2 + 4\sum_i \alpha_i \left(\bar{\theta} - \theta_i\right)^2
\end{aligned}$$

Thus, $\bar{V}_{LB} < \bar{V}^*$ if and only if

$$\sum_i \alpha_i \left(\bar{\theta} - \theta_i\right)^2 + \frac{1}{2}\text{var}(z) < U^2 + \frac{1}{4}\left(\bar{\theta} - \theta_L\right)^2 + \frac{1}{4}\left[\left(\theta_L - \bar{\theta}\right)^2 + 4\sum_i \alpha_i \left(\bar{\theta} - \theta_i\right)^2\right] - 2U\left(\bar{\theta} - \theta_L\right)$$

$$\frac{1}{2}\text{var}(z) < U^2 + \frac{1}{2}\left(\bar{\theta} - \theta_L\right)^2 - 2U\left(\bar{\theta} - \theta_L\right)$$

equivalently,

$$U^2 > \frac{1}{2}\text{var}(z) - \frac{1}{2}\left(\bar{\theta} - \theta_L\right)^2 + 2U\left(\bar{\theta} - \theta_L\right) = \frac{1}{2}\text{var}(z) - \frac{1}{2}\alpha_H^2 \Delta^2 + 2U\alpha_H \Delta$$

where we used $\bar{\theta} - \theta_L = \alpha_L \theta_L + \alpha_H \theta_H - \theta_L = \alpha_H \Delta$.

Thus, $U_H(\Delta)$ is the larger (positive) root of $2U^2 - 4\alpha_H \Delta U + (\alpha_H^2 \Delta^2 - \text{var}(z)) = 0$,

$$U = \frac{4\alpha_H \Delta \pm \sqrt{16\alpha_H^2 \Delta^2 - 8(\alpha_H^2 \Delta^2 - \text{var}(z))}}{4} = \frac{4\alpha_H \Delta \pm \sqrt{8\alpha_H^2 \Delta^2 + 8\text{var}(z)}}{4}$$
$$= \alpha_H \Delta \pm \sqrt{\frac{\alpha_H^2 \Delta^2 + \text{var}(z)}{2}},$$

and so

$$U_H(\Delta) = \alpha_H \Delta + \sqrt{\frac{\alpha_H^2 \Delta^2 + \text{var}(z)}{2}}.$$

Q.E.D.

F Proofs of results in Section 8

Proof of Proposition 6

Proof. If $\nu > \alpha_H \Delta$, we know that the effective bias of the θ_L policymaker is strictly positive. Thus, the optimal DE will not coincide with the θ_L faction's preferred policy,

$$\int u\left(\pi_{DE}\left(z, \theta_L\right), z, \theta_L\right) f\left(z\right) dz < \int u\left(\pi_f\left(z, \theta_L\right), z, \theta_L\right) f\left(z\right) dz.$$

Assume now that the θ_L faction is in charge and consider a deviation to full flexibility. By the inequality above, the θ_L policymaker strictly gains. Since $\nu > \alpha_H \Delta$, we have that $\bar{\theta} < \theta_L + \nu < \theta_H + \nu$, so the θ_H policymaker prefers $\pi_f\left(z, \theta_L\right)$ to $\pi_{DE}\left(z, \theta_L\right)$ because its preferences are more aligned with the θ_L policymaker's than society's are, that is,

$$\int u\left(\pi_{DE}\left(z, \theta_L\right), z, \theta_H\right) f\left(z\right) dz < \int u\left(\pi_f\left(z, \theta_L\right), z, \theta_H\right) f\left(z\right) dz$$

given our assumption on preferences. Both factions therefore agree to replace the optimal DE with full flexibility, and the optimal DE is not credible. Note that society itself prefers the mandate:

$$\int u\left(\pi_{DE}\left(z, \theta_L\right), z, \bar{\theta} - \nu\right) f\left(z\right) dz > \int u\left(\pi_f\left(z, \theta_L\right), z, \bar{\theta} - \nu\right) f\left(z\right) dz.$$

A symmetric argument follows for $\nu < -\alpha_L \Delta$. For the last part, note that a necessary condition for DE to be optimal is that the bias is large, so it must be that $\nu > \alpha_H \Delta$ or $\nu < -\alpha_L \Delta$, as shown in Proposition 3. Thus, whenever DE is optimal, it is not credible. Q.E.D.

Proof of Proposition 7

Proof. First, we show that if U is large enough, then the best constant policy is credible. Note that the value of the best constant policy for type θ is

$$\bar{V}^*(\theta) = \int R(\bar{\pi}^*, \bar{\pi}^*, z, \theta) f(z) dz = -\frac{1}{2}\left[U^2 + \left(\bar{\theta} - \theta\right)^2 + \text{var}(z)\right].$$

If the alternative regime chosen is LB, the value for type θ is at most $\bar{V}_{LB}(\theta)$ defined as

$$V_{LB}(\theta) \leq \bar{V}_{LB}(\theta) = \int R\left(x_M(\theta_L), \pi_f(z, \theta; x_M(\theta_L)), z, \theta\right) f(z) dz,$$

where $x_M(\theta_L)$ is the Markov equilibrium expected inflation defined in the proof of Proposition 10, and is a lower bound for the equilibrium inflation under LB. Note that

$$\begin{aligned}
\bar{V}_{LB}(\theta) &= -\frac{1}{2}\int\left[\left(U + x_M(\theta_L) - \frac{U + x_M(\theta_L) + z + \theta}{2}\right)^2 + \left(\frac{U + x_M(\theta_L) + z + \theta}{2} - z - \theta\right)^2\right] f(z)dz \\
&= -\frac{1}{2}\int\left[2\left(\frac{U + x_M(\theta_L) - z - \theta}{2}\right)^2\right] f(z)dz = -\frac{1}{2}\int\left[2\left(\frac{2U + \theta_L - z - \theta}{2}\right)^2\right] f(z)dz \\
&= -\left[U^2 + U(\theta_L - \theta) + \frac{1}{4}\int(\theta_L - z - \theta)^2 f(z) dz\right] \\
&= -\left[U^2 + U(\theta_L - \theta) + \frac{1}{4}\text{var}(z) + \frac{1}{4}(\theta_L - \theta)^2\right].
\end{aligned}$$

Thus, $\bar{V}_{LB}(\theta) < \bar{V}^*(\theta)$ for all θ if

$$\begin{aligned}
2\left[U^2 + \frac{1}{4}\text{var}(z)\right] &> U^2 + \left(\bar{\theta} - \theta_L\right)^2 + \text{var}(z) \\
2\left[U^2 - U\Delta + \frac{1}{4}\text{var}(z) + \frac{1}{4}\Delta^2\right] &> U^2 + \left(\bar{\theta} - \theta_H\right)^2 + \text{var}(z).
\end{aligned}$$

Thus,

$$U > \max\left\{\sqrt{\alpha_H^2\Delta^2 + \frac{1}{2}\text{var}(z)}, \Delta + \sqrt{\left(\frac{1}{2} + \alpha_L^2\right)\Delta^2 + \frac{1}{2}\text{var}(z)}\right\} \tag{56}$$

is a sufficient condition for $\bar{V}_{LB}(\theta) < \bar{V}^*(\theta)$ for all θ. This in turn implies that $\bar{V}^*(\theta) > V_{LB}(\theta)$ for all θ and so factions do not want to switch to LB from the best constant policy.

Consider next the possibility that the factions agree to switch to another DE policy. Note that no deviation to a different constant policy can be agreed upon: any constant policy that the faction in power strictly prefers to $\bar{\pi}^*$ is strictly worse than $\bar{\pi}^*$ for the opposition. We next compare the value of the best constant policy to an upper bound for all policies feasible in DE that allow for some discretion. For this class of policies, the

expected inflation must be at least equal to

$$x_{\min} = \theta_L + U + \underline{z}.$$

This minimal amount of expected inflation is higher than that under the best constant policy, $\bar{x}^* = \bar{\theta}$, if U is sufficiently high. Thus, the best constant policy is beneficial in this respect. With this lower bound for inflation, we can write

$$V_{DE}(\theta) \leq \bar{V}_{DE}(\theta) = \int R\left(x_{\min}, \pi_f\left(z, \theta; x_{\min}\right), z, \theta\right) f(z)\, dz,$$
$$V_{DE} \leq \bar{V}_{DE} = \int \sum_i \alpha_i R\left(x_{\min}, \pi_f\left(z, \bar{\theta}; x_{\min}\right), z, \theta_i\right) f(z)\, dz.$$

Note that

$$\begin{aligned}
\bar{V}_{DE}(\theta) &= -\frac{1}{2}\int \left[\left(U + x_{\min} - \frac{U + x_{\min} + z + \theta}{2}\right)^2 + \left(\frac{U + x_{\min} + z + \theta}{2} - z - \theta\right)^2\right] f(z)dz \\
&= -\frac{1}{2}\int \left[2\left(\frac{U + x_{\min} - z - \theta}{2}\right)^2\right] f(z)dz = -\int \left[\left(U + \frac{\theta_L + \underline{z} - z - \theta}{2}\right)^2\right] f(z)dz \\
&= -\left[U^2 + U\left(\theta_L + \underline{z} - \theta\right) + \frac{1}{4}\left(\theta_L - \theta + \underline{z}\right)^2 + \frac{1}{4}\text{var}(z)\right].
\end{aligned}$$

Thus, $\bar{V}_{DE}(\theta) < \bar{V}^*(\theta)$ for all θ if, for all θ,

$$U^2 + 2U\left(\theta_L + \underline{z} - \theta\right) + \frac{1}{2}\left(\theta_L - \theta + \underline{z}\right)^2 > \left(\bar{\theta} - \theta\right)^2 + \frac{1}{2}\text{var}(z),$$

or

$$U > \max\left\{-\underline{z} + \sqrt{\frac{1}{2}\underline{z}^2 + \alpha_H^2\Delta^2 + \frac{1}{2}\text{var}(z)}, (\Delta - \underline{z}) + \sqrt{\frac{1}{2}\left(\Delta - \underline{z}\right)^2 + \alpha_L^2\Delta^2 + \frac{1}{2}\text{var}(z)}\right\}. \tag{57}$$

Then $\bar{V}_{DE}(\theta) < \bar{V}^*(\theta)$ for all θ and so the best constant policy is the preferred outcome under DE for all θ.

Defining $\bar{U}_H$ as the smallest U such that conditions (56) and (57) hold, we have that for all $U \geq \bar{U}_H$ the best constant policy is credible. Furthermore, the best constant policy is optimal for society. In fact, as we showed, for $U \geq \bar{U}_H$ no policy under LB and no policy under DE allowing some discretion can improve on the best constant policy. $Q.E.D.$

Proof of Proposition 8

Proof. Suppose the θ_H policymaker is in charge. Throughout the proof, write $\pi_f(z) \equiv \pi_f(z, \theta_H)$ for the flexible policy of the type in power. First notice that it is not feasible for type θ_H to change D (with the approval of θ_L), since increasing π_o is beneficial for θ_H but detrimental for θ_L. Consider the problem of delegating policy-making to a decision maker with identical preferences, subject to the other faction receiving at least its LB payoff. This problem boils down to choosing an interval $D = [\pi_f(\underline{z}), \pi_f(z^*)]$. Notice that if $\pi_f(z^*) = \pi_o$, then $V_H^{LB}(\pi_o, \theta_H) > V_H^{DE}(D, \theta_H)$, since both have identical payoffs for $z < (\pi_f)^{-1}(\pi_o)$ and LB has strictly higher payoffs for z larger than this threshold owing to the greater flexibility. Therefore, in order for delegation to be strictly preferred, it must be that $\pi_f(z^*) > \pi_o$.

Next, consider the LB problem. Without loss of generality, assume that $\theta_H = -\theta_L = \Delta/2$. Given that we assumed quadratic utility, type θ's ideal policy is $z + \theta + \nu$, and the solution to the LB problem is

$$\pi^{LB}(z, \theta_H) = \begin{cases} z + \theta_H + \nu & \text{if } z \leq \pi_o - \frac{\Delta}{2} - \nu \\ \pi_o & \text{if } z \in \left(\pi_o - \frac{\Delta}{2} - \nu, \pi_o + \frac{\Delta}{2} - \nu\right), \\ \min\{z + \theta_H + \nu, 2(z - \theta_H + \nu) - \pi_o\} & \text{if } z \geq \pi_o + \frac{\Delta}{2} - \nu \end{cases}$$

Now suppose that $\bar{z} \leq 3\theta_H + \pi_o - \nu$, so the participation constraint for the θ_L faction is always binding. Then $V_L^{LB}(\pi_o, \theta_H) = V_L^{DE}(D, \theta_H)$, where D has a cap equal to π_o. Therefore, in this range we have that $V_L^{DE}\left(\tilde{D}, \theta_H\right) < V_L^{LB}(\pi_o, \theta_H)$ for any $\tilde{D}$ with a larger cap. As a result, the best the θ_H faction can do in delegation is to choose $\pi_f(z^*) = \pi_o$, and delegation is not optimal. Now consider the case in which $\bar{z} > 3\theta_H + \pi_o - \nu$. In this case, we have that $V_L^{LB}(\pi_o, \theta_H) > V_L^{DE}(D, \theta_H)$, where D has a cap equal to π_o. This implies that type θ_H can delegate only with a lower cap than π_o. Therefore, the θ_H faction strictly prefers LB. A similar argument holds for the case in which the θ_L policymaker is in charge. Q.E.D.