Macroeconomic Regimes as Institutional Architectures[*]

Endogenous Fiscal Credibility and State-Dependent Monetary Policy

Federico Pablo Vacalebre

ORCID: 0009-0009-4946-3464

Pontificia Universidad Católica Argentina (UCA), Buenos Aires, Argentina

Corresponding author: Federico Pablo Vacalebre
Address: Av. Alicia Moreau de Justo 1300 (C1107AAZ), Buenos Aires, Argentina
Telephone: +54 9 11 5178 0798
Fax: Not available
E-mail: fvacalebre@uca.edu.ar

Abstract

This paper studies how beliefs about fiscal backing shape monetary–fiscal regimes and how fiscal choices influence those beliefs. Credibility is updated from observed compliance using discounted Beta learning. Conditional on credibility, a threshold separates fiscal from monetary dominance, governs the response of the price level to interest-rate-rule disturbances, and changes the inflation-variance-minimizing monetary response. A debt-dependent fiscal limit yields a credibility frontier under a frozen effective-backing approximation; this approximation matches the local fiscal response at steady-state debt but does not establish a fully endogenous-feasibility equilibrium. The paper's distinctive contribution is an optimizing fiscal authority. In a deterministic benchmark, optimal policy reaches the monetary-dominance threshold from outside a credibility trap, while sustained dominance requires an additional maintenance condition. That condition holds in the computed rebuilding cases, not for all parameters. Relative to an independent exogenous-backing benchmark, fiscal choices make regime transitions depend on political incentives and inherited credibility. Numerical results with debt and private political-cost shocks show persistent low credibility and preventive fiscal backing at higher obligations. Argentina serves as an illustration rather than an empirical test.

Keywords: monetary–fiscal interactions, fiscal credibility, fiscal dominance, regime switching, discounted Beta learning, fiscal limits, credibility traps, optimal monetary policy.

JEL Classification: D83, E31, E52, E58, E62, E63, H63.

[*] The author thanks the editor and an anonymous referee for constructive comments. The usual disclaimer applies.

1. Introduction

The interaction between monetary and fiscal policy is usually organized around the taxonomy of Leeper (1991). A monetary authority is active when it raises the nominal interest rate more than one-for-one with inflation, and a fiscal authority is passive when it adjusts primary surpluses to stabilize debt. When monetary policy is active and fiscal policy passive, the interest-rate rule anchors the price level and the fiscal authority guarantees solvency. When the roles are reversed, the price level adjusts to make the real value of debt consistent with the path of surpluses. The taxonomy is precise, but it treats the fiscal stance as a known parameter. In practice what agents observe is not a parameter but a record: a history of episodes in which the fiscal authority did or did not deliver the adjustment that monetary policy required. What matters for the pricing of nominal debt is the inference agents draw from that record about whether fiscal backing will be forthcoming.

This paper places that inference at the center of the monetary–fiscal framework. Credibility is defined as the probability agents assign to the fiscal authority's willingness to back public debt when backing is feasible, and it is modeled as a belief that evolves endogenously with observed fiscal compliance. Under discounted Beta learning, in which older observations receive geometrically declining weight, the posterior mean probability of backing is a geometrically weighted stock of past compliance, so that the two interpretations of credibility that coexist in the literature, credibility as an accumulated reputation and credibility as a subjective probability of regime survival, coincide in a single state variable with an explicit law of motion. The paper asks two questions: what changes, for equilibrium determination and for policy design, once the regime matrix is conditioned on that state; and what additional economic insights follow when the evolution of the state is determined by the fiscal authority's choices rather than by an exogenous process.

The answers are organized in three layers, which differ in what they require of credibility. The first layer consists of results that are conditional on the credibility state and hold under any process that generates it. The active–passive classification generalizes to a credibility-weighted matrix: fiscal policy is effectively passive only when the probability-weighted response of surpluses to debt exceeds the real interest rate, which defines a threshold below which the economy is fiscally dominant even in periods in which the fiscal authority does adjust (Proposition 1). Below the threshold, a positive disturbance to the interest-rate rule lowers inflation on impact by an amount that rises with credibility but raises the price level

permanently, a formal version of the mechanism that Sims (2011) described as stepping on a rake; above it, the same disturbance lowers the price level (Proposition 2). Under a loss that counts only the inflation variance generated by surplus disturbances, the variance-minimizing response coefficient is passive and close to a peg below the threshold, while above it any active coefficient minimizes the loss (Proposition 3). With a stochastic fiscal limit in the manner of Bi (2012), the threshold becomes a credibility frontier that rises with debt (Proposition 4); this step rests on a frozen effective-backing approximation, which is exact when the limit is constant and matches the local fiscal response at steady-state debt without establishing full dynamic equivalence; Section 8 states and evaluates it. In first moments, these results are those of an economy with a known partial-backing fraction equal to current credibility, which places them in the tradition of Aiyagari and Gertler (1985); uncertainty about backing adds only second-moment terms. The second layer concerns belief dynamics. Because credibility is a stock of past compliance, the boundaries of the regime matrix move with the fiscal record, a default lowers credibility multiplicatively, and the probability of regime breakdown follows from the learning rule and the fiscal limit (Proposition 4). The third layer concerns the fiscal authority's incentive to invest in credibility, and it is the only one that requires credibility to respond to fiscal choices. In a deterministic benchmark, the set of credibility levels from which the authority never backs is an interval starting at zero, and from every level outside it the optimal plan reaches the threshold; a closed form approximates the boundary of that set (Proposition 5). In that benchmark, entry into monetary dominance depends on initial credibility and a one-time loss of credibility that enters the trap set is permanent, whereas with exogenous backing that is independent over time the long-run behavior of credibility does not depend on its initial level (Proposition 6). Monetary dominance is sustained after entry under a stated condition on the maintenance policy, which holds in every computed case with rebuilding but not for all parameter values. Solving the decision problem with debt and private political-cost shocks numerically shows that higher obligations raise the frontier but also raise the cost of remaining in fiscal dominance, that low credibility becomes a highly persistent region rather than an absorbing state, and that governments with high credibility back preventively when debt rises (Section 9.4).

The contribution is not the idea that credibility is endogenous, which has a long lineage discussed in Section 3, nor the idea that fiscal backing matters for the effects of monetary policy, which goes back at least to Aiyagari and Gertler (1985) and Leeper (1991), nor the

idea that monetary–fiscal regimes switch, which is the subject of a well-developed literature. It is to establish which implications of a belief state about fiscal backing require that the state respond to fiscal choices, and to characterize those choices. The model is deliberately simple: prices are flexible, the real interest rate is constant, and debt is one-period and nominal. That simplicity is what delivers closed-form thresholds, and the limits it imposes are discussed where they bind. Expectations are formed under anticipated utility, so that the equilibrium concept is a learning equilibrium rather than a fully consistent Bayesian equilibrium; Section 9.5 states which links of the feedback between fiscal incentives, beliefs, and outcomes are solved within the model and which are assumed.

The remainder of the paper proceeds as follows. Section 2 summarizes the monetary–fiscal framework and the theory of fiscal limits on which the analysis builds. Section 3 positions the contribution in the literature. Section 4 presents the model and the law of motion of credibility. Sections 5 to 7 derive the equilibrium, the state-dependent effects of disturbances to the interest-rate rule, and the variance-minimizing monetary response. Section 8 introduces fiscal limits and derives the breakdown probability. Section 9 sets out the fiscal authority's decision problem, compares endogenous with exogenous credibility, and solves for the optimal backing policy as a function of credibility and debt. Section 10 uses Argentina as a brief illustration, and Section 11 concludes. Proofs and numerical methods are collected in the Appendix.

2. Background: Monetary–Fiscal Regimes and Fiscal Limits

The analysis builds on four established elements, which are stated here only to fix notation. A policy regime, in the sense used since Lucas (1976), is the set of rules that agents expect to govern policy, as distinct from the instruments those rules deploy. The time-inconsistency argument of Kydland and Prescott (1977) and the reputational extension of Barro and Gordon (1983) explain why the credibility of those rules, rather than their announcement, is what disciplines expectations, and North and Weingast (1989) showed that such credibility rests on institutional arrangements that constrain the future exercise of authority.

The first element is the government budget constraint. With debt and the primary surplus measured relative to output, the debt ratio evolves as

$$b_t = \frac{1 + r_t}{1 + g_t}\, b_{t-1} - s_t \tag{1}$$

where r_t is the real rate on public debt and g_t the real growth rate. Solvency is an intertemporal condition on the expected path of surpluses rather than on the current balance (Blanchard & Fischer, 1989). The second element is the fiscal reaction function. Bohn (1998) showed that a positive response of the primary surplus to debt is sufficient for sustainability, and the local stability condition for debt requires that the response exceed the interest–growth differential,

$$s_t = \bar{s} + \gamma\,(b_{t-1} - b^{*}) + \varepsilon_t, \qquad \gamma > r - g \tag{2}$$

The third element is the interest-rate rule, $i_t = \bar{\imath} + \varphi(\pi_t - \pi^{*}) + \varphi_y y_t$, which satisfies the Taylor principle when $\varphi > 1$ (Taylor, 1993). Clarida et al. (2000) documented that the move to an active rule coincided with the transition to low inflation in the United States.

The fourth element is the joint determination of the price level. Leeper (1991) showed that a unique bounded equilibrium requires one authority to be active and the other passive. When fiscal policy is active, the price level is determined by the valuation equation of the fiscal theory of the price level (Sims, 1994; Woodford, 2001; Cochrane, 2023),

$$\frac{B_{t-1}}{P_t} = E_t \sum_{j=0}^{\infty} \frac{S_{t+j}}{(1 + r)^{j}} \tag{3}$$

which equates the real value of nominal debt to the present value of real primary surpluses. The unpleasant arithmetic of Sargent and Wallace (1981) is the monetarist counterpart of the same constraint, and Canzoneri et al. (2001) examined whether the data are consistent with fiscal determination of the price level. Table 1 collects the four configurations.

Table 1. *The monetary–fiscal regime matrix (Leeper, 1991)*

Fiscal policy	Monetary policy	Configuration	Price level determined by
Passive	Active	Monetary dominance	Interest-rate rule
Active	Passive	Fiscal dominance	Valuation equation (3)
Active	Active	Conflict	No bounded equilibrium
Passive	Passive	Weak anchor	Indeterminate

The theory of fiscal limits qualifies the assumption that surpluses can rise without bound. Bi (2012) and Bi et al. (2013) model the maximum politically sustainable surplus as a stochastic

object, so that the present value of maximal surpluses defines a stochastic debt limit $b_t^{\max}$ and the probability that debt exceeds it,

$$\theta_t = \Pr_t(b_t > b_t^{\max}) = \Psi_t(b_t), \qquad \partial \Psi_t / \partial b_t \geq 0 \tag{4}$$

becomes a state-dependent probability of default or monetization. Reinhart and Rogoff (2009) document how frequently that probability has materialized historically. The matrix and the fiscal limit are the background. What they leave open is how agents come to believe that the fiscal authority is passive, and what follows when that belief is itself a state variable.

3. Related Literature and the Contribution

Three bodies of work come close to the mechanism developed here, and the contribution is best stated by its difference from each.

The first models credibility as an endogenous state. In Isard et al. (2001), the credibility of an inflation-targeting central bank and the inflation expectations it anchors evolve with the policy record, and endogenous credibility strengthens the case for forward-looking rules. Hommes and Lustenhouwer (2019) study an inflation-targeting central bank whose credibility depends on its past success in meeting the target, in a New Keynesian model with heterogeneous expectations, and show that credibility conditions both the admissible policy region and the risk of deflationary spirals at the zero lower bound. Earlier, Barro and Gordon (1983) and Cukierman (1992) treated reputation as an asset accumulated slowly and lost quickly. Closer to the fiscal side, Demirel (2012) studies imperfect fiscal credibility in a monetary–fiscal policy game and shows that the value of monetary commitment depends on the degree of fiscal commitment. With the exception of Demirel (2012), the credibility at stake in these contributions is that of the monetary authority with respect to its inflation target. The present paper models the credibility of fiscal backing as a belief that evolves with the compliance record and determines which side of the Leeper matrix the economy occupies.

The second body of work studies switching between monetary–fiscal regimes. Davig and Leeper (2007) show that when monetary policy switches between regimes, determinacy depends on the distribution of current and future regimes rather than on the current coefficient, which generalizes the Taylor principle. Chung et al. (2007) let both monetary and fiscal policy switch between active and passive configurations and show that an active

monetary rule does not insulate the economy from fiscal disturbances when agents attach positive probability to a future regime in which fiscal policy does not stabilize debt. Davig and Leeper (2008) make monetary regime changes endogenous by letting them occur when observed variables cross thresholds. Bianchi (2012) estimates an evolving monetary–fiscal mix, and Bianchi and Melosi (2017) analyze how uncertainty about the future policy mix shapes macroeconomic dynamics. Farmer et al. (2009) provide general conditions for determinacy in Markov-switching rational expectations models. Bianchi and Ilut (2017) are especially close in emphasizing agents' beliefs about fiscal backing: they show that shifts in beliefs about the policy mix can generate inflation without a contemporaneous change in fundamentals and can account for failed and successful disinflations. In this literature, the transition probabilities are either fixed parameters or functions of observed macroeconomic states. Here, the belief about fiscal backing is an endogenous state updated from observed compliance, and that state moves the active–passive regime boundary itself. That difference has economic content, because it makes the boundaries of the regime matrix move with the history of fiscal behavior and gives the fiscal authority an instrument, compliance, with which to move them.

The third is the analysis of how fiscal backing conditions the effects of policy. Aiyagari and Gertler (1985) showed that the consequences of conventional monetary and fiscal policy depend on how government bonds are backed, by future taxes or by future money creation, an early statement of the mechanism that the fiscal theory of the price level later developed. Sims (2011) argued that monetary tightening without fiscal backing can lower inflation temporarily and raise it later, and Cochrane (2023) develops the fiscal-theory account of interest-rate policy. The present paper derives a credibility threshold at which the long-run effect of a rule disturbance on the price level changes sign. Relative to Aiyagari and Gertler (1985), the paper does not replace the configuration of backing with beliefs as an end in itself. Conditional on the belief state, the probability of backing plays the role of a partial-backing parameter, and Section 5 shows that the first-moment results coincide with those of an economy with known partial backing. What modeling backing as an uncertain and evolving belief adds is threefold: the configuration of backing becomes a state with a law of motion, so that the regime boundaries move with the fiscal record; uncertainty about whether backing materializes adds a second-moment channel that shifts the mean-square stability boundaries; and the fiscal authority acquires an asset, its credibility, in which it can choose to invest or not (Section 9).

The contribution is therefore the connection of elements that the literature has largely kept apart: the active–passive taxonomy of Leeper (1991), regime uncertainty in the tradition of Davig and Leeper (2007) and Chung et al. (2007), and a belief state about fiscal backing that evolves with observed policy and that the fiscal authority can influence. The results fall into three layers: (A) results conditional on the credibility state, namely Propositions 1 to 3 and parts (i) and (ii) of Proposition 4, the latter under the frozen effective-backing approximation of Section 8, which do not depend on how the state is generated; (B) results that depend on the learning dynamics, namely the breakdown probability of Proposition 4(iii) and the effect of defaults through (12); and (C) results that depend on the fiscal authority's choice of backing, namely Propositions 5 and 6 and the numerical results of Section 9.4. Only layer (C) distinguishes the model from a benchmark in which backing is exogenous and independent over time.

4. A Model of Endogenous Fiscal Credibility

4.1 Environment

The economy is the simplest one in which the price level can be determined either by monetary or by fiscal policy. Prices are flexible, output is exogenous, and the real interest rate is constant at $r = \beta^{-1} - 1$, where β is the discount factor. Write $R \equiv \beta^{-1}$. The government issues one-period nominal debt. All variables are expressed as deviations from a zero-inflation steady state in which real debt equals $\bar{b} > 0$. The Fisher relation and the interest-rate rule are

$$\hat{\imath}_t = E_t \pi_{t+1} \tag{5}$$

$$\hat{\imath}_t = \varphi\, \pi_t + u_t, \qquad u_t = \rho\, u_{t-1} + e_t \tag{6}$$

where u_t is a persistent deviation from the rule, interpreted as a contractionary shift in the rule when positive (Section 6 derives its effect on the equilibrium nominal rate), with $|\rho| < 1$. Let $w_t \equiv b_{t-1} + \bar{b}\, \hat{\imath}_{t-1}$ denote the nominal value of maturing debt inherited at date t, which is predetermined. The linearized government budget constraint (Appendix A.1) is

$$b_t = R\left(w_t - \bar{b}\, \pi_t\right) - s_t \tag{7}$$

so that surprise inflation reduces the real value of maturing debt in proportion to $\bar{b}$, which is the revaluation channel of the fiscal theory.

4.2 Fiscal backing and the credibility state

Fiscal policy is described by an indicator of backing. In each period the fiscal authority either adjusts the surplus to the inherited debt ($\xi_t = 1$) or does not ($\xi_t = 0$),

$$s_t = \xi_t \gamma w_t + \varepsilon_t, \qquad \xi_t \in \{0,1\} \tag{8}$$

where $\gamma > r$, so that a backed period is passive in the sense of (2), and ε_t is a surplus disturbance with variance σ^2. The innovations e_t and ε_t are mean zero and mutually independent, and under anticipated utility the backing draws are conditionally independent across periods given C_t; Section 8 adds the feasibility event. The indicator is the fiscal analogue of compliance with a rule: in a backed period the government delivers the adjustment that stabilizes debt, and in an unbacked period it does not. Substituting (6) to (8) gives the law of motion of maturing debt,

$$w_{t+1} = \lambda(\xi_t)\, w_t + \bar{b}(\varphi - R)\, \pi_t + \bar{b}\, u_t - \varepsilon_t, \qquad \lambda(\xi_t) \equiv R - \xi_t \gamma \tag{9}$$

Agents do not know whether future periods will be backed. Credibility C_t is defined as the probability they assign to backing in any future period in which backing is feasible,

$$C_t \equiv \Pr\big(\xi_{t+j} = 1 \mid \mathcal{I}_t^0, \text{ backing feasible at } t+j\big), \qquad j \geq 0 \tag{10}$$

which is the schematic object $C_t = \Pr_t(\text{fiscal backing})$ that organizes the rest of the analysis. It measures willingness, not capacity. In Sections 4 to 7 backing is always feasible, so C_t is also the effective probability of backing; Section 8 introduces the fiscal limit and the distinction between the two. Expectations are formed under anticipated utility in the sense formalized by Cogley and Sargent (2008): agents update C_t with the discounted Beta filter of Section 4.3 and form expectations treating the current belief as the probability that will govern future periods. The equilibrium at each date is therefore the rational expectations equilibrium of an economy in which backing occurs independently each period with probability C_t, which is a special case of the Markov-switching environment of Davig and Leeper (2007) and Chung et al. (2007) with transition probabilities equal across states. It is convenient to define the credibility-weighted dynamic coefficient of debt, $\bar{\lambda}(C_t) \equiv R - C_t \gamma$.

Timing is as follows. At the beginning of period t, w_t and the prior belief C_t are predetermined, and $\mathcal{I}_t^0$ denotes the beginning-of-period information set. Conditional on C_t, the backing state ξ_t is realized and determines current fiscal behavior. The within-period equilibrium is solved under the anticipated-utility convention that C_t is the transition

probability governing future backing. At the end of the period, the observed backing outcome updates the prior to C_{t+1} through (11). Thus C_t is the beginning-of-period belief, and learning is applied between periods. Until Section 9, backing outcomes are taken as given: the results of Sections 5 to 8 hold for any process that generates them.

4.3 Credibility as a stock and as a belief

Agents use a discounted Beta filter: they hold a Beta prior over the probability of backing and discount older observations by a factor $1 - g \in (0,1)$, so that the effective sample has length $1/g$ in the stationary state. Older observations receive geometrically declining weight, and the posterior mean obeys the constant-gain update (Appendix A.2)

$$C_{t+1} = (1 - g)\, C_t + g\, \xi_t \tag{11}$$

The posterior probability of backing is an exponentially weighted average of the compliance record. This is the sense in which credibility is simultaneously a stock and a belief. It is a stock because each episode of compliance or non-compliance is carried forward with geometrically declining weight, so that a regime inherits the credibility accumulated or destroyed by its predecessors. It is a belief because the same object is the subjective probability that prices nominal debt. The gain g measures how quickly the record is revised: a small gain is the formal counterpart of credibility that is accumulated slowly. The filter is an operational learning rule. It is not the Bayesian posterior for a fixed Bernoulli probability, which would weight all observations equally, nor the posterior of agents who know the fiscal authority's decision problem (Section 9.5).

Events that are more informative than a single period of non-compliance can be incorporated by weighting them as multiple observations. If a default $D_t = 1$ counts as $\kappa_D \geq 1$ additional observations of non-backing, the update becomes

$$C_{t+1} = \frac{(1 - g)\, C_t + g\, \xi_t}{1 + g\, \kappa_D D_t} \tag{12}$$

where, for tractability, the effective sample size is renormalized to its stationary value $1/g$ after an exceptional default observation. Under this normalization, (12) is the scalar law of motion for credibility, and a default scales credibility down multiplicatively. This is the

asymmetry described by Cukierman (1992): credibility built over many periods of compliance can be lost in a single episode.[1]

5. Equilibrium under Endogenous Credibility

Equilibrium is characterized within the class of minimum-state-variable solutions linear in $(w_t, u_t, \varepsilon_t)$ with coefficients that may depend on the current backing realization, and boundedness is defined in conditional expectation, as in Davig and Leeper (2007). Appendix A.3 states the mean-square refinement of Farmer et al. (2009), which shifts the boundaries slightly without changing the results. The central object is the credibility threshold

$$C^* \equiv \frac{r}{\gamma} \tag{13}$$

at which the credibility-weighted coefficient $\bar{\lambda}(C)$ equals one.

Proposition 1 (credibility-weighted regime matrix). *Let $\gamma > r$. (i) If $\varphi > 1$ and $C_t > C^*$, there is a unique bounded equilibrium, in which inflation does not respond to debt or to surplus disturbances, $\pi_t = -u_t/(\varphi - \rho)$. (ii) If $\varphi < 1$ and $C_t < C^*$, there is a unique bounded equilibrium,*

$$\pi_t = \psi(\xi_t)\, w_t + \theta\, u_t + \kappa\, \varepsilon_t, \qquad \bar{\psi} = \frac{\bar{\lambda} - \varphi}{\bar{b}\,(R - \varphi)}, \quad \psi(\xi_t) = \bar{\psi}\,\frac{\lambda(\xi_t)}{\bar{\lambda}}, \quad \kappa = -\frac{\bar{\psi}}{\bar{\lambda}} \tag{14}$$

with $\bar{\psi} > 0$, in which inflation responds to inherited debt in backed and unbacked periods alike. (iii) If $\varphi > 1$ and $C_t < C^*$, no bounded equilibrium exists. (iv) If $\varphi < 1$ and $C_t > C^*$, the equilibrium is not unique.

The proposition replaces the fiscal coefficient of the Leeper matrix with its credibility-weighted counterpart. Fiscal policy is effectively passive when $C_t\gamma > r$, that is, when the expected response of surpluses to debt exceeds the real interest rate. Two implications deserve emphasis. First, the regime is determined by the record rather than by the current action. In region (ii), inflation responds to debt even in a period in which the fiscal authority does adjust, because $\psi(1) > 0$: a single backed period does not buy monetary dominance

[1] The reduced-form law of motion in the earlier version of this framework, in which credibility rose with compliance and fell with inflation, default, and discretion, is nested in (12) when those events enter as observations of non-backing with different weights.

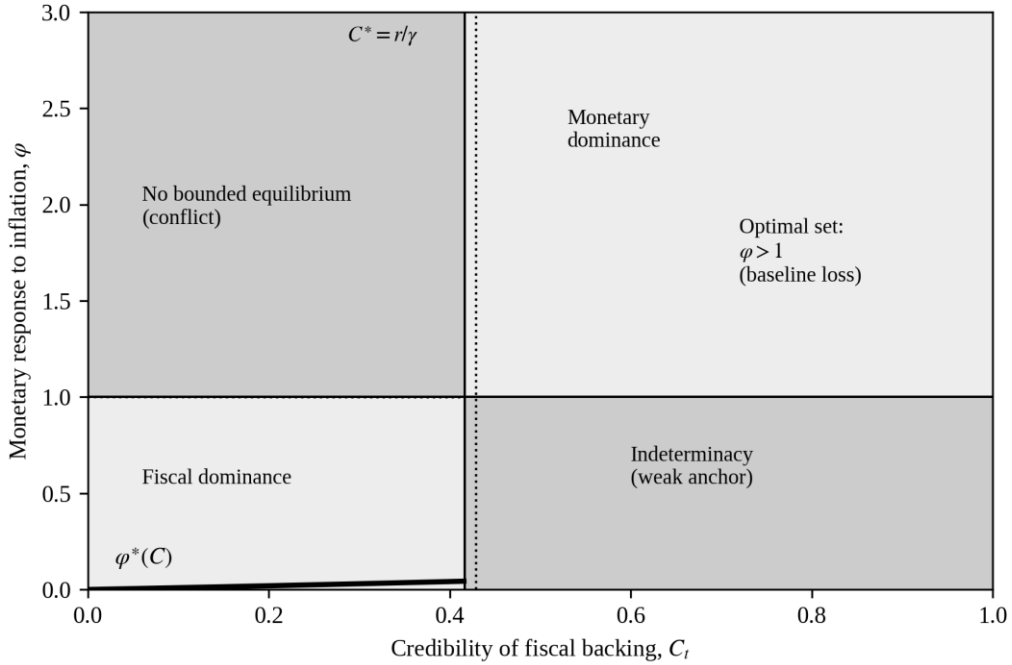
when the record leaves C_t below the threshold. This is the belief-based counterpart of the result of Chung et al. (2007) that an active monetary rule does not insulate the economy from fiscal disturbances when unbacked fiscal regimes are expected. Second, the boundaries of the matrix are not fixed. Because C_t moves with compliance according to (11), an economy can cross from region (i) to region (iii) without any change in either policy rule, simply because a sequence of unbacked periods erodes credibility. Table 2 restates Table 1 in these terms, and Figure 1 maps the regions in the (C, φ) plane.

Proposition 1 is a result about the credibility state, not about how the state is generated. Two observations make this precise. First, it applies date by date to any process for C_t, including an exogenous one, because under anticipated utility the within-period equilibrium depends only on the current belief. Second, in first moments the probability of backing acts as a partial-backing parameter. In an economy in which the fiscal authority adjusts the surplus by a known fraction δ of the backed response, $s_t = \delta\gamma w_t + \varepsilon_t$, the dynamic coefficient of debt is $R - \delta\gamma$, and with $\delta = C_t$ the coefficients $\bar{\psi}$, θ, and κ of (14) and the first-moment determinacy conditions are the same (Appendix A.3). The configuration of backing therefore determines the regime in the sense of Aiyagari and Gertler (1985) and Leeper (1991). Uncertainty about whether backing materializes adds only the second-moment term $C_t(1 - C_t)\gamma^2$, which enters the mean-square boundaries of Appendix A.3 and the variance in Proposition 3. What the belief interpretation adds is that C_t has a law of motion driven by fiscal behavior, which Sections 8 and 9 exploit.

Table 2. *The credibility-weighted regime matrix*

Credibility	Monetary response	Configuration	Response of inflation to inherited debt
$C_t > C^*$	$\varphi > 1$	Monetary dominance	None
$C_t < C^*$	$\varphi < 1$	Fiscal dominance	Positive, in backed and unbacked periods
$C_t < C^*$	$\varphi > 1$	Conflict	No bounded equilibrium
$C_t > C^*$	$\varphi < 1$	Weak anchor	Indeterminate

Figure 1. *Regime map and variance-minimizing monetary response in the credibility–policy plane*



Note. Solid boundaries are the first-moment conditions of Proposition 1; dotted lines are the mean-square refinements of Appendix A.3. The thick line is the variance-minimizing response $\varphi^*(C)$ in the fiscal-dominance region (Proposition 3). Calibration: $\beta = 0.96$, $\gamma = 0.10$ (so $C^* = 0.42$), $\bar{b} = 0.6$.

6. Policy-Rule Disturbances and the Credibility Threshold

The credibility state matters economically because it changes the sign of the effect of monetary policy on the price level. Consider a positive innovation in u_t, an upward shift in the intercept of the interest-rate rule (6) that would raise the policy rate at an unchanged inflation rate. Because the real rate is constant and the nominal rate satisfies the Fisher relation (5), the equilibrium response of the nominal rate is itself endogenous and need not be positive. The analysis therefore distinguishes the rule disturbance u_t, the equilibrium nominal rate $\hat{\imath}_t$, and the responses of inflation and of the price level, and refers to $u_t > 0$ as a contractionary rule disturbance rather than as a tightening.

Proposition 2 (state-dependent transmission). *(i) If* $\varphi > 1$ *and* $C_t > C^*$*, a unit contractionary rule disturbance lowers inflation on impact by* $1/(\varphi - \rho)$ *and lowers the*

price level permanently by $1/[(1-\rho)(\varphi-\rho)]$. *(ii) If* $\varphi < 1$ *and* $C_t < C^*$, *the impact effect on inflation is*

$$\frac{\partial \pi_t}{\partial u_t} = \theta = -\frac{C_t \gamma}{(R-\varphi)\left(\bar{\lambda}-\rho\right)} \leq 0 \tag{15}$$

which is zero when $C_t = 0$ *and increasing in absolute value with* C_t, *while the long-run effect on the price level is*

$$\sum_{k=0}^{\infty} \frac{\partial E_t \pi_{t+k}}{\partial u_t} = \frac{\left(\bar{\lambda}-\varphi\right)(R-\rho)/(1-\varphi) - C_t \gamma}{(1-\rho)(R-\varphi)\left(\bar{\lambda}-\rho\right)} > 0 \tag{16}$$

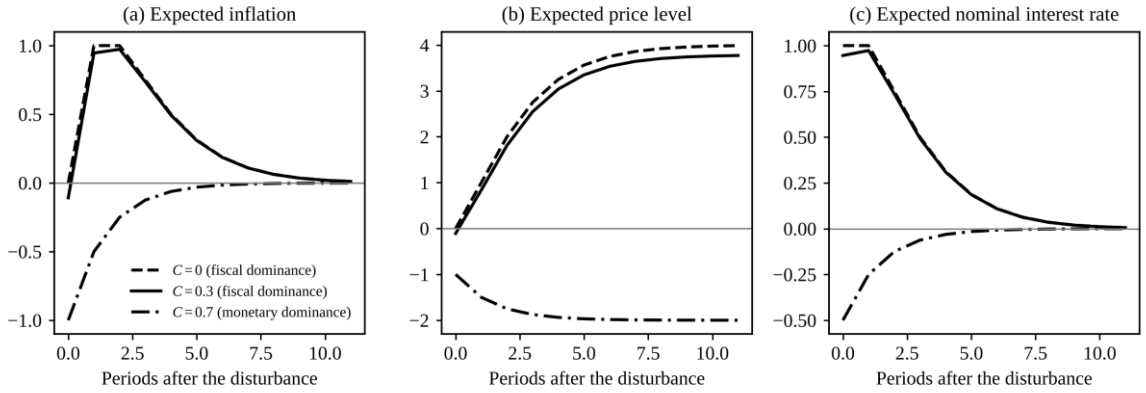
The long-run effect of the disturbance on the price level therefore changes sign at the threshold C^*. *(iii) The impact effect on the equilibrium nominal rate is* $-\rho/(\varphi-\rho)$ *in case (i), negative for* $\rho > 0$, *and* $1+\varphi\theta = [(R-\varphi)(R-\rho) - RC_t\gamma]/\left[(R-\varphi)\left(\bar{\lambda}-\rho\right)\right]$ *in case (ii), positive if and only if* $(R-\varphi)(R-\rho) > RC_t\gamma$.

Below the threshold the mechanism is the one Sims (2011) described. When the condition in part (iii) holds, as in the calibration of Figure 2, the disturbance raises the equilibrium nominal rate and therefore the nominal value of maturing debt. A fraction C_t of that increase is expected to be backed by future surpluses, and to that extent the disturbance is disinflationary on impact. The remaining fraction is not backed, and must be reconciled with the valuation equation through a higher price level in subsequent periods. The impact disinflation is increasing in credibility and the subsequent reflation is decreasing in it, but within the fiscally dominant region the second always dominates in the long run (Appendix A.4). Above the threshold, the disturbance lowers inflation and, for $\rho > 0$, also the equilibrium nominal rate: with flexible prices and a constant real rate, a persistent rule disturbance lowers expected inflation and hence, by (5), the nominal rate, so that a contractionary rule disturbance and a rise in the observed policy rate do not coincide. Figure 2 shows the three patterns: a credible economy disinflates monotonically with a lower nominal rate, whereas an economy below the threshold, in which the nominal rate rises, disinflates briefly, if at all, and then inflates.

The proposition gives a precise answer, within the model, to the question of whether there is a level of credibility below which a contractionary monetary disturbance becomes counterproductive for the price level in the absence of fiscal adjustment. There is, and it is $C^* = r/\gamma$. The threshold is lower when the backed fiscal response is stronger and when the

real interest rate is lower, which formalizes the observation that the same monetary action can stabilize one economy and destabilize another with an identical central bank. Proposition 2 compares transmission across credibility states and policy configurations. It holds for any process that generates C_t and does not by itself identify an incremental effect of making credibility endogenous, which is the subject of Section 9.

Figure 2. *Expected responses to a contractionary rule disturbance under low and high credibility*



Note. Calibration as in Figure 1, with $\rho = 0.5$; $\varphi = 0.5$ in the fiscally dominant cases and $\varphi = 1.5$ in the monetary-dominance case. Unit innovation in u_t. Panel (c) plots the equilibrium nominal rate, $\hat{\imath}_t = \varphi\pi_t + u_t = E_t\pi_{t+1}$, which on impact rises by 1.00 and 0.95 for $C = 0$ and $C = 0.3$ and falls by 0.50 under monetary dominance (Proposition 2(iii)).

7. Optimal Monetary Policy Conditional on Credibility

If credibility shapes how monetary policy is transmitted, the monetary rule should condition on it. To study that dependence under a deliberately narrow criterion, consider a monetary authority that chooses the response coefficient to minimize the unconditional variance of inflation induced by surplus disturbances, taking C_t as given and treating it as constant, in line with anticipated utility. The criterion excludes output, the variance of debt, the rule disturbance u_t, and real-rate disturbances. It is not a welfare function, and optimal in this section means optimal under this loss. The admissible set consists of the coefficients $\varphi \geq 0$ for which inflation is mean-square bounded.

Proposition 3 (state-dependent variance-minimizing response). *In the monetary-dominance region, inflation is insulated from surplus disturbances for every* $\varphi > 1$. *In the*

fiscal-dominance region, mean-square stability requires $\varphi^2 m(C_t) < 1$*, with* $m(C) \equiv 1 + C(1-C)\gamma^2/\bar{\lambda}^2 \geq 1$*, and the variance of inflation is*

$$V(\varphi; C_t) = \frac{\sigma^2}{\bar{b}^2} \frac{\left(\bar{\lambda} - \varphi\right)^2}{\bar{\lambda}^2 (R - \varphi)^2 \left[1 - m(C_t)\, \varphi^2\right]} \tag{17}$$

(i) The variance at an interest-rate peg, $V(0; C) = \sigma^2 \beta^2 / \bar{b}^2$*, is independent of credibility, and its slope at the peg is* $-2C_t\gamma/\left(\bar{\lambda}R\right)$ *times its level, so that the optimal response satisfies* $\varphi^*(0) = 0$ *and* $\varphi^*(C_t) > 0$ *for* $C_t \in (0, C^*)$*. (ii)* V *has a local minimum close to the peg, given to a first-order approximation by*

$$\varphi^p(C_t) \approx \frac{\bar{\lambda}^{-1} - R^{-1}}{m(C_t) - \bar{\lambda}^{-2} + R^{-2}} \tag{18}$$

(iii) As C_t *approaches* C^**,* V *acquires a second local minimum close to* $\varphi = 1/m(C_t)$*, at the edge of the mean-square stability region, where the variance of debt becomes large. To leading order in* r*, this second minimum lies below the near-peg minimum in a neighborhood of the threshold when* $C^* > 1/2$*; thus* $1/2$ *is an analytical approximation to the switching boundary rather than an exact cutoff, and the exact boundary, which depends weakly on* β*, is determined numerically. For thresholds below the boundary, the global optimum is the near-peg branch* $\varphi^p(C_t)$ *throughout the fiscal-dominance region; for sufficiently high thresholds, it switches to a passive response just below unity for credibility in an interval* $\left(\hat{C}, C^*\right)$ *below the threshold. In both cases* $\varphi^*(C_t) < 1$ *below the threshold, and above it the optimal set under the baseline loss is the active region* $\varphi > 1$*.*

The economics of the near-peg branch is the following. With one-period nominal debt, a positive response of the policy rate to inflation raises debt service and therefore the inflation needed to reconcile the valuation equation, so that the monetary authority minimizes inflation volatility by letting the revaluation occur rather than resisting it. The optimal response nevertheless rises with credibility, because a fraction C_t of the debt-service cost of leaning against inflation is backed. In the baseline calibration ($C^* = 0.42 < 1/2$), the near-peg branch is the global optimum throughout the fiscal-dominance region, φ^* rises monotonically from zero to about 0.04 at the threshold, an order of magnitude close to the real interest rate, and the error of the approximation (18) is below 0.0002 (Appendix A.5). The thick line in Figure 1 plots this optimum against the regime boundaries of Proposition 1.

Part (iii) qualifies the near-peg result when the threshold is high, with $1/2$ as the leading-order approximation to the boundary (Appendix A.5 reports the exact values), that is, when the backed fiscal response γ is weak relative to the real interest rate. Close to such a threshold, $\bar{\lambda}$ is close to one, and a response just below unity nearly eliminates the response of inflation to debt, at the cost of pushing debt dynamics to the edge of mean-square stability. Because the loss function counts only inflation volatility, that trade is worthwhile near the threshold. For example, with $\beta = 0.96$ and $\gamma = 0.05$ ($C^* = 0.83$), the global optimum switches from the near-peg branch to $\varphi^* \approx 0.999$ at $\hat{C} \approx 0.82$ (Appendix A.5). The result that survives in all cases is the one that matters for the policy mix: below the threshold the optimal response is passive, and aggressiveness has value only once credibility crosses the threshold, where the Taylor principle becomes feasible. In the monetary-dominance region the baseline loss is flat with respect to surplus disturbances for all $\varphi > 1$, so the model selects the active region rather than a unique coefficient.[2]

Within the model and under this loss, a deterioration in fiscal credibility that moves the economy below the threshold therefore calls for a passive response, and discontinuously so, because what is required is not a different monetary coefficient but a different policy mix. The conclusion is conditional on the environment: with nominal rigidities, long-term debt, or a loss that also counts output or debt volatility, the variance-minimizing coefficient below the threshold need not be close to a peg. The policy that restores the effectiveness of monetary aggressiveness is fiscal, and it operates through C_t, which is the subject of Section 9.

8. Fiscal Limits, the Credibility Frontier, and Endogenous Breakdown

The threshold C^* assumes that a willing fiscal authority is always able to back. The theory of fiscal limits implies that it is not. Throughout the paper, C_t measures beliefs about willingness: the probability that the fiscal authority backs in a period in which backing is feasible. Feasibility is a separate object. To map the fiscal-limit concept in (4) into the linearized model, let $\Psi(w_t)$ denote the probability that the inherited maturing-obligation state w_t exceeds fiscal capacity, so that backing is infeasible. This is the local counterpart of

[2] An arbitrarily small real-rate disturbance provides a tie-breaker in favor of a more aggressive feasible response: with such disturbances added to (5), inflation volatility is decreasing in φ throughout the monetary-dominance region (Appendix A.5). This disturbance is not part of the baseline loss.

$\Psi_t(b_t)$ in (4); exogenous time variation in the fiscal-limit distribution is suppressed, and Ψ is non-decreasing. The effective probability of backing is then

$$\tilde{C}_t \equiv C_t[1 - \Psi(w_t)] \tag{19}$$

which equals C_t in Sections 4 to 7, where the limit does not bind. Willingness and capacity enter once each. A period in which the limit binds is observed as such and carries no information about willingness, so (11) updates C_t only in periods in which backing is feasible.[3] The timing within period t is as follows. At the beginning of the period agents know w_t and hold the belief C_t. Feasibility is then drawn, with probability $1 - \Psi(w_t)$, and observed. If backing is feasible, the fiscal authority backs with probability q_t, its actual strategy, which agents estimate by C_t. Inflation and w_{t+1} follow from (9), and C_{t+1} follows from (11) if backing was feasible and equals C_t otherwise. Thus C_t is the subjective probability of willingness, $1 - \Psi(w_t)$ the probability of capacity, $\tilde{C}_t$ the subjective probability that an adjustment actually occurs, and q_t the authority's true probability of backing when feasible; incapacity enters only through Ψ. Two expectations must be kept apart. Before feasibility is realized, the probability that period t carries an adjustment is $\tilde{C}_t$; once feasibility is observed, it is C_t if backing is feasible and zero otherwise. The current period therefore enters the equilibrium through the realized effective indicator $\xi_t F_t$, and $\tilde{C}_t$ governs only the ex-ante expectation, formed with beginning-of-period information, of effective backing at future dates. Incapacity is thus counted once in each period. In Section 9, where ξ_t is set to zero when backing is infeasible, ξ_t denotes this realized effective indicator.

Replacing C_t by $\tilde{C}_t$ in the equilibrium conditions of Proposition 1 requires an assumption in addition to anticipated utility. Anticipated utility freezes C_t, a belief about a parameter that agents do not know. It does not freeze $1 - \Psi(w_{t+j})$, which is a known function of future obligations, themselves determined by future inflation through (9). With exact expectations, the expected backed surplus at $t + j$ is $C_t \gamma E_t\{[1 - \Psi(w_{t+j})]w_{t+j}\}$, a nonlinear functional of the equilibrium distribution of w_{t+j}, and the linear characterization of Proposition 1 does not apply. Proposition 4 therefore rests on a frozen effective-backing approximation: agents

[3] A default enters (12) as evidence about willingness only when it is read as strategic non-backing. A default forced by the fiscal limit is the infeasibility event itself and does not enter (12), which avoids counting capacity twice.

evaluate feasibility at all future dates at its current value, replacing $E_t\{[1-\Psi(w_{t+j})]w_{t+j}\}$ by $[1-\Psi(w_t)]E_t w_{t+j}$, and treat $\tilde{C}_t$, like C_t, as constant. The freeze uses only beginning-of-period information, (C_t, w_t), applies at every horizon $j \geq 1$, and is re-evaluated each period as C_t and w_t change. It differs from anticipated utility in its object: anticipated utility treats as constant a parameter that agents do not know, whereas the freeze neglects a known dependence of future feasibility on future debt. Under this approximation the within-period equilibrium is that of Proposition 1 with $\tilde{C}_t$ in place of C_t, which yields a frontier in the credibility–debt plane.

The approximation is exact in one case and locally accurate in another (Appendix A.6). If Ψ is constant over the relevant range of obligations, the effective backing indicator is a Bernoulli variable with probability $C_t(1-\Psi)$ independent of w, and Proposition 1 holds with that probability. At the steady state $w = 0$, the expected backed surplus $C_t\gamma[1-\Psi(w)]w$ has slope $C_t\gamma[1-\Psi(0)]$, which is the frozen slope, so the approximation matches the local fiscal response at steady-state debt. This is a coincidence of first-order fiscal responses at one point, not a dynamic equivalence, because obligations move away from zero along equilibrium paths. Away from $w = 0$, and for $w \geq 0$ with Ψ differentiable and non-decreasing, the frozen slope $1-\Psi(w)$ exceeds the marginal slope $h'(w) = 1-\Psi(w)-\Psi'(w)w$ by $\Psi'(w)w \geq 0$. Where $h'(w) > C^*$, so that the marginal frontier $C^*/h'(w)$ is well defined, a local characterization based on the marginal slope requires more credibility for monetary dominance: the frontier (20) is a lower bound for it. Where $h'(w) \leq C^*$, no credibility satisfies the marginal condition, and wherever (20) exceeds one, so does the alternative.

In the calibration of Section 9.4, the two frontiers coincide at $w = 0$, equal 0.419 and 0.429 at $w = 0.10$ and 0.430 and 0.496 at $w = 0.15$, and exceed one at $w \approx 0.32$ and $w \approx 0.20$, respectively. Three implications follow. The threshold C^* and every result without a binding limit, including Propositions 5 and 6, which are derived for $\Psi = 0$, are unaffected. The qualitative content of Proposition 4, that a binding limit raises the credibility required for monetary dominance and that beyond some debt level no credibility suffices, also holds under the alternative characterization; but the location of the frontier at positive obligations depends on the approximation, and (20) understates the credibility required as debt approaches the limit. Finally, the numerical results of Section 9.4, which use (20) to determine the regime, are conditional on the frozen effective-backing approximation.

Proposition 4 (credibility frontier and breakdown). *Under the frozen effective-backing approximation: (i) monetary dominance requires* $\tilde{C}_t > C^*$*, that is,* $C_t > C^F(w_t)$*, where*

$$C^F(w_t) \equiv \frac{C^*}{1 - \Psi(w_t)} \tag{20}$$

is non-decreasing in debt, increasing where Ψ *is, and exceeds one, so that monetary dominance is infeasible at any level of credibility, once* $\Psi(w_t) \geq 1 - C^*$*. (ii) If* $C_t > C^F(w_t)$*, the optimal set under the baseline loss is the active region* $\varphi > 1$*; otherwise, the optimal response* $\varphi^*(C_t, w_t)$ *is given by Proposition 3 evaluated at* $\tilde{C}_t$*. Parts (i) and (ii) are characterizations conditional on the frozen approximation, which is exact when* Ψ *is constant and matches the local fiscal response at* $w_t = 0$ *without implying full dynamic equivalence; they are not theorems about equilibrium determination in a full Markov model with endogenous feasibility. (iii) Let* q_t *denote the probability with which the fiscal authority actually backs at* t *when backing is feasible, its strategy, of which* C_t *is the agents' estimate. If* $C_t > C^F(w_t)$ *and backing is feasible at* t*, the probability that the economy leaves this region at* $t + 1$*, that is, that* $C_{t+1} < C^F(w_{t+1})$*, is*

$$\begin{aligned} \Pr_t(\text{breakdown}_{t+1}) \quad &= q_t \Pr_t[(1-g)C_t + g < C^F(w_{t+1}) \mid \xi_t = 1] \\ &\quad + (1 - q_t) \Pr_t[(1-g)C_t < C^F(w_{t+1}) \mid \xi_t = 0] \end{aligned} \tag{21}$$

where the conditional distribution of w_{t+1} *follows from (9): for given shocks, non-backing raises* w_{t+1} *by* γw_t *relative to backing. If* $w_t \geq 0$*,* Ψ *is non-decreasing, and the shocks* (u_t, ε_t) *are independent of* (C_t, w_t, ξ_t)*, the probability (21) is non-increasing in* C_t *and in* q_t*, and non-decreasing in* w_t*. The probability perceived by agents is (21) with* C_t *in place of* q_t*. Part (iii) is exact conditional on the region: the approximation enters only through the definition of the region by (20).*

The breakdown probability is derived from the learning dynamics (11) and the fiscal-limit block (4) rather than introduced as a separate reduced-form function, and the conditioning on ξ_t in (21) makes explicit that non-backing harms the regime twice: it erodes credibility by gC_t, and it leaves a higher debt state that raises the frontier. Rising debt alone also raises the frontier, so that an unchanged level of credibility becomes insufficient, which is the credibility analogue of the result in Bi (2012) that risk premia and inflation remain contained until debt approaches the limit and then rise abruptly. A default treated as strategic non-backing shifts credibility down multiplicatively by (12), which can move the economy

several regions at once. The joint dependence on C_t and w_t is the formal content of the statement that the optimal aggressiveness of monetary policy varies with credibility and with fiscal space. It also implies that low current inflation is weak evidence of a durable regime, since an economy can sit just above the frontier with a high probability of breakdown. Proposition 4(iii) takes the backing strategy q_t as given; Section 9 derives it from the fiscal authority's problem.

9. Fiscal Incentives to Invest in Credibility

The analysis so far has taken the backing decision as given, and Propositions 1 to 4 hold for any process that generates it. This section asks what follows when the fiscal authority chooses backing knowing that compliance today raises credibility tomorrow. Section 9.1 states the decision problem, Section 9.2 characterizes rebuilding in a benchmark without a fiscal limit, Section 9.3 compares endogenous with exogenous credibility, Section 9.4 solves the problem with debt and a fiscal limit, and Section 9.5 identifies which links of the feedback are modeled and which are assumed.

9.1 The fiscal decision problem

The state at the beginning of period t is (C_t, w_t): inherited credibility and inherited maturing obligations. With one-period debt, $w_t \equiv b_{t-1} + \bar{b}\,\hat{\imath}_{t-1}$ is the predetermined debt variable of the model, so that a policy rule $a_t^*(C_t, b_t)$ stated in terms of debt corresponds here to $\xi_t^*(C_t, w_t)$, with w_t including the interest due on maturing debt. The action is the backing decision $a_t \equiv \xi_t \in \{0,1\}$, the binary choice already embedded in (8). A continuous adjustment intensity would require a separate theory of how partial adjustment is read by agents, and is not pursued. Within period t the timing is as follows. Feasibility is drawn first: with probability $\Psi(w_t)$ backing is infeasible, $\xi_t = 0$, and, as in Section 8, the period carries no information about willingness, so that $C_{t+1} = C_t$. If backing is feasible, the authority chooses ξ_t. The regime is determined by $\tilde{C}_t$ (Proposition 4, under the frozen effective-backing approximation of Section 8), which, as in Section 8, governs the ex-ante expectation of future effective backing evaluated at (C_t, w_t), while the current period enters only through the realized ξ_t: the monetary authority sets $\varphi = \varphi^p(\tilde{C}_t)$ from (18) below the frontier and an active coefficient above it, in line with the sequencing corollary below. The surplus

disturbance is then realized, w_{t+1} follows (9) at the equilibrium inflation rate of Proposition 4, and C_{t+1} follows (11). The authority solves

$$V(C,w) = \Psi(w)\, Q_0^{\emptyset}(C,w) + [1-\Psi(w)]\min_{\xi\in\{0,1\}} Q_\xi(C,w) \tag{22}$$

$$Q_\xi(C,w) = \xi\, \chi(w) + L\big(\tilde{C},w,\xi\big) + \beta_g\, E\, V\left((1-g)\, C + g\, \xi,\ w'_\xi\right) \tag{23}$$

where w'_ξ is the value of w_{t+1} implied by (9) when $\xi_t = \xi$, and $Q_0^{\emptyset}$ is Q_0 with credibility left unchanged, C in place of $(1-g)C$. The political cost of a backed period, $\chi(w) = \chi_0 + \chi_1(\gamma w/\sigma)^2$, rises with the size of the required adjustment γw, and an unbacked period costs nothing; in the notation $\kappa(a_t)$ of a generic adjustment cost, $\kappa(\xi_t) = \xi_t \chi(w_t)$. The macroeconomic loss is the expected squared inflation rate implied by the equilibrium of Proposition 4, conditional on the state and on the authority's decision,

$$L\big(\tilde{C},w,\xi\big) = L_F\ \mathbb{1}\big[\tilde{C} < C^*\big]\frac{\psi(\xi)^2 w^2 + \kappa^2\sigma^2}{\beta^2\sigma^2/\bar{b}^2} \tag{24}$$

where $\psi(\xi)$ and κ are the coefficients of (14) evaluated at $\tilde{C}$ and $\varphi^p\big(\tilde{C}\big)$. Because $\pi_t = \psi(\xi_t)w_t + \kappa\varepsilon_t$ with $E\varepsilon_t = 0$, conditional on (w_t,ξ_t) inflation has mean $\psi(\xi_t)w_t$ and variance $\kappa^2\sigma^2$, and (24) is its second moment, $E[\pi_t^2 \mid w_t,\xi_t] = \psi(\xi_t)^2 w_t^2 + \kappa^2\sigma^2$, not its variance. The second moment is the appropriate criterion here: the term $\psi(\xi)^2 w^2$ is the inflation that revalues inherited obligations, which is predictable given the state but is precisely the cost of fiscal dominance that backing reduces, and dropping it would remove the debt-dependent incentive to back. The normalization makes L_F the loss per period spent in fiscal dominance with obligations at their steady-state value and an interest-rate peg, since $\kappa^2\sigma^2 = \beta^2\sigma^2/\bar{b}^2$ at $\varphi = 0$ for every $\tilde{C}$; under $\varphi^p\big(\tilde{C}\big)$ the loss at $w = 0$ lies between $0.996L_F$ and L_F for every $\tilde{C} < C^*$. In monetary dominance, inflation responds neither to w_t nor to ε_t, and the loss is zero. The criterion of Section 7 is related but distinct. Proposition 3 minimizes the unconditional variance (17), which, because unconditional mean inflation is zero, equals the unconditional second moment of inflation. Holding credibility fixed in the fiscal-dominance region, with w_t at its stationary distribution and ξ_t drawn with probability C, the expectation of (24) equals L_F/s_0^2 times (17), where $s_0^2 \equiv \beta^2\sigma^2/\bar{b}^2$ is the normalizing constant in (24); under these conditions the two criteria are proportional, not identical (Appendix A.9). Thus (24) conditions on the current state and on the authority's own decision, whereas (17) averages over both.

Two features of (22) to (24) carry the economics. Backing is an investment, because it raises C_{t+1} through (11), and in the fiscal-dominance region it also has a current return, because $\psi(1) < \psi(0)$: a backed period reduces the inflation required to revalue inherited obligations. Debt enters twice, raising the frontier $C^F(w)$ through capacity and raising the cost of remaining in fiscal dominance through w^2. For the probability of backing to be a genuine probability from the agents' point of view, as the interpretation of C_t and q_t requires, the authority's decision cannot be a known deterministic function of a public state. The minimal source of private information is a political-cost shock: each option carries an additional cost $\nu_t(\xi) = \varsigma[\gamma_E - e_t(\xi)]$, where the $e_t(\xi)$ are independent standard type-I extreme value variables and γ_E is Euler's constant, so that the shocks have mean zero and scale ς. The information sets are as follows. The authority observes (C_t, w_t), the feasibility draw, and ν_t, and knows its own policy. Private agents observe w_t, whether backing was feasible, and the outcome ξ_t; they do not observe ν_t and do not know the authority's policy, and they update C_t by (11). The minimum in (22) is replaced by its expectation over ν_t, which, with mean-zero shocks, is exactly $-\varsigma\ln[\exp(-Q_0/\varsigma) + \exp(-Q_1/\varsigma)]$, and the probability of backing when feasible is

$$q^*(C, w) = \frac{1}{1 + \exp\{[Q_1(C, w) - Q_0(C, w)]/\varsigma\}} \tag{25}$$

which is the strategy q_t of Proposition 4(iii). As $\varsigma \to 0$ the policy becomes deterministic. Because both options have finite values, the probability of backing lies strictly between zero and one whenever the cost shocks have positive scale. Appendix A.9 describes the numerical solution.

9.2 Rebuilding and the credibility trap

The investment motive can be isolated in a benchmark without a binding fiscal limit, $\Psi = 0$, with inherited obligations at their steady-state value, $w_t = 0$, and without private cost shocks. The authority then bears a cost $\chi = \chi(0) > 0$ in each period in which it backs and a loss $L_F > 0$ in each period spent below the threshold, normalized to zero at or above it in line with Proposition 3, and (22) reduces to $V(C) = \min_{\xi \in \{0,1\}}\{\chi\xi + L_F \mathbb{1}[C < C^*] + \beta_g V\big((1-g)C + g\xi\big)\}$, with the convention that $C = C^*$ counts as monetary dominance. The authority discounts the future at $\beta_g \leq \beta$. Following the political-economy literature, a low β_g captures political turnover and fragmented decision-making, which bias governments

toward present expenditure (Alesina & Tabellini, 1990; Persson & Tabellini, 2000). Proposition 5 states first what holds for the exact solution of this problem, and then a closed form obtained by comparing two strategies under an approximate maintenance cost: rebuilding by backing in consecutive periods until the threshold is reached and then maintaining credibility at or above it, against never backing, with the cost of maintenance approximated by χC^* per period, the cost of backing at the long-run frequency C^* spread uniformly over time.[4]

Starting from $C_0 < C^*$ and in the absence of default, (11) implies that the number of consecutive backed periods required to reach the threshold is

$$n(C_0) = \left\lceil \frac{\ln[(1-C^*)/(1-C_0)]}{\ln(1-g)} \right\rceil \tag{26}$$

which is decreasing in initial credibility and in the gain.

Proposition 5 (rebuilding and the credibility trap). *Consider the benchmark without a fiscal limit. (i) Exact solution. The value function V is non-increasing in C. The trap set $T \equiv \{C < C^*: V(C) = L_F/(1-\beta_g)\}$ is either empty or an interval with lower end zero, and it does not shrink as χ/L_F rises. From $C_0 \in T$, never backing is optimal and, with ties broken in favour of not backing, $C_t \in T$ for all t, $C_t \to 0$, and the economy is in fiscal dominance in every period. From $C_0 < C^*$ outside T, every optimal plan reaches $[C^*, 1]$ in finite time; part (i) does not state that credibility remains in $[C^*, 1]$ thereafter. (ii) Closed-form approximation. If the cost of remaining at or above C^* is approximated by χC^* per period, then, between rebuilding in consecutive periods and never backing, rebuilding is preferred if and only if $\beta_g^{n(C_0)}[L_F + \chi(1-C^*)] \geq \chi$, that is, if and only if $C_0 \geq C^{trap}$, where*

$$C^{\text{trap}} = 1 - (1-C^*)(1-g)^{-\lfloor \bar{n} \rfloor}, \qquad \bar{n} = \frac{\ln\{\chi/[L_F + \chi(1-C^*)]\}}{\ln\beta_g} \tag{27}$$

Within this approximate problem, C^{trap} is non-increasing in β_g, L_F, and g, and non-decreasing in χ; if $C^{trap} \leq 0$, rebuilding is preferred from every initial level, and if $L_F \leq \chi C^*$, it is preferred from none.

[4] In the deterministic benchmark the authority controls ξ_t, so maintenance can hold pathwise: backing whenever $(1-g)C_t < C^*$ keeps $C_t \geq C^*$ in every period. The long-run frequency of this rule exceeds C^* because each backed period overshoots the threshold; the closed form uses the limiting cost χC^*, and Appendix A.8 computes the exact cost. With a fiscal limit or private cost shocks, credibility is stochastic, and maintenance can keep it above the frontier only with high probability (Section 9.4).

Part (i) holds for all parameter values (Appendix A.8). It does not determine whether the authority backs in every period between the boundary of T and C^*, nor how it behaves above C^*; these features, and the accuracy of part (ii), were computed by solving the benchmark Bellman equation exactly for 36 combinations of $\beta_g \in \{0.80, 0.90, 0.95\}$, $g \in \{0.05, 0.10, 0.20\}$, and $\chi/L_F \in \{0.25, 0.5, 1, 2\}$, with the calibration of Figure 1 (Appendix A.8). In every combination the exact policy never backs below an exact threshold C_e^{trap}, which coincides with the upper end of T up to grid precision, and backs in every period between C_e^{trap} and C^*. Above C^*, the policy follows the maintenance rule 'back if and only if $C_t < C^*/(1-g)$' in 28 combinations, never backs in 5, and backs only on an upper sub-interval of $[C^*, C^*/(1-g))$ in 3; all combinations without full maintenance have $\chi = 2L_F$. The maintenance rule keeps credibility at or above C^* along every path, but it overshoots the threshold, so that its long-run backing frequency, 0.443, 0.470, and 0.500 for $g = 0.05$, 0.10, and 0.20, exceeds $C^* = 0.417$, and its present cost exceeds $\chi C^*/(1-\beta_g)$. As a result, the closed form (27) never overstates the exact threshold. It coincides with it in many combinations; the mean absolute difference is 0.018, and the largest difference, 0.182, arises for $\beta_g = 0.80$, $g = 0.20$, and $\chi = L_F$, where the exact threshold is 0.271 and the closed form gives 0.089. In two combinations, $\beta_g = 0.95$ and $\chi = 2L_F$ with $g = 0.10$ and $g = 0.20$, the closed form recommends rebuilding from 0.352 and 0.271 while the exact solution never backs below C^*. There, the present cost of the maintenance rule from C^*, $19.74L_F$ and $20.51L_F$, is close to or above the value of remaining in fiscal dominance forever, $20L_F$, whereas the approximation puts it at $16.67L_F$. On the grid, the exact threshold is non-increasing in β_g and non-decreasing in χ; the latter also follows from part (i). It is not monotone in g: for $\beta_g = 0.95$ and $\chi = 2L_F$ it rises from 0.409 at $g = 0.05$ to C^* at $g = 0.10$, because faster learning shortens rebuilding but enlarges the overshoot and hence the cost of maintenance. These comparative statics are numerical findings on the grid, not general results. The same computations show that reaching C^* does not by itself imply remaining above it. In the 8 combinations without full maintenance, credibility falls below C^* from every computed initial level above it, and, since in those combinations $T = [0, C^*)$, the economy then remains in fiscal dominance; Proposition 6 returns to this point.

In the deterministic benchmark, the credibility trap is the dynamic counterpart of the multiplicity that a static reading of the matrix only suggests. Two economies with identical institutions and identical policy rules can settle in different regimes because they inherit

different credibility, and the low-credibility economy remains there not because rebuilding is impossible but because, from where it starts, the horizon over which the costs must be borne exceeds the horizon of the decision-maker. The trap is deeper where political horizons are short, which links the model to the observation of Besley and Persson (2011) that fiscal capacity is an accumulated institutional stock that regimes reliant on inflation tend to underinvest in. The proposition also identifies the role of the gain. Faster learning shortens rebuilding, by Proposition 5, but makes breakdown faster, by Proposition 4, and in the exact solution it also raises the cost of maintenance: an environment in which the record is revised quickly is one in which regimes are both easier to restore and easier to lose.

With a fiscal limit and debt held at a constant level w with $C^F(w) < 1$, the same argument applies with $C^F(w)$ in place of C^* in (26) and (27): rebuilding requires reaching the frontier rather than the benchmark threshold, and maintenance requires backing with frequency of at least $C^F(w)$. In the closed form, C^{trap} is increasing in the threshold it is computed from, and C^F is increasing in debt, so that the trap is deeper when debt is higher, provided the loss from fiscal dominance does not itself depend on debt. When $\Psi(w) \geq 1 - C^*$, no level of credibility suffices, and rebuilding requires reducing debt as well as accumulating a compliance record. Section 9.4 shows that, once the loss reflects the inflation needed to revalue inherited obligations, as in (24), debt also raises the cost of remaining in fiscal dominance, and that this second effect can reverse the conclusion.

9.3 Exogenous versus endogenous credibility

What does endogeneity add? Consider two economies that share the monetary block, the learning rule (11), the fiscal limit, and the equilibrium of Propositions 1 to 4. In Model E, backing when feasible is drawn independently each period with a fixed probability $\bar{q}$, unrelated to (C_t, w_t) and to political costs. In Model EN, it is chosen by the fiscal authority as in (22). Model E is a benchmark, not a stand-in for every exogenous process: an exogenous process with state-dependent backing probabilities can display other properties, and one that mechanically reproduced the policy of Model EN would reproduce its outcomes. The comparison isolates what follows from deriving the dependence of backing on the state from political costs, horizons, and losses.

Proposition 6 (endogenous versus i.i.d. exogenous backing). *(i) In both models, the equilibrium at date t is given by Propositions 1 to 4 evaluated at C_t and w_t, with Proposition*

4 understood under the frozen effective-backing approximation when the fiscal limit binds. Any single realized path of credibility generated by Model EN is also generated by some exogenous sequence of backing outcomes; this equivalence is pathwise and does not extend to the distribution of paths, to incentives, or to counterfactuals. (ii) In Model E with $\Psi = 0$*, two credibility paths driven by the same backing outcomes from* C_0 *and* C_0' *satisfy* $|C_t - C_t'| = (1-g)^t|C_0 - C_0'|$*. Hence* C_t *converges in distribution to a unique invariant distribution with mean* $\bar{q}$*, which does not depend on* C_0*,* β_g*,* χ*, or* L_F*, and a one-time change in credibility changes* C_{t+k} *by* $(1-g)^k$ *times its size. (iii) In the benchmark of Proposition 5, suppose that* T *and its complement in* $[0, C^*)$ *are both nonempty. Then entry into monetary dominance under the optimal policy of Model EN depends on initial credibility: from* $C_0 \in T$*, with ties broken in favour of not backing, the economy is in fiscal dominance in every period, while from* $C_0 < C^*$ *outside* T *credibility reaches* $[C^*, 1]$ *in finite time. A one-time fall in credibility that moves* C_t *into* T *has permanent effects, unlike in Model E. If, in addition, the optimal policy backs at every* $C \geq C^*$ *with* $(1-g)C < C^*$*, then* $[C^*, 1]$ *is invariant, so credibility never falls below* C^* *after entry, and monetary dominance in the strict sense of Proposition 1 holds whenever* $C_t > C^*$*. (iv) In both models, over any N periods in which backing is feasible, the averages of* C_t *and of* ξ_t *taken over those periods differ by at most* $1/(gN)$.

Part (i) is a pathwise observational-equivalence result. A single history of credibility cannot distinguish the two models, and Propositions 1 to 4 are not, by themselves, evidence that credibility is endogenous. The models differ in the mapping from primitives to the distribution of histories and in their counterfactuals. Under Model E, initial credibility and political parameters are irrelevant for the long run; under Model EN, in the benchmark, entry into monetary dominance depends on initial credibility and on β_g, χ, L_F, and g, backing depends on the credibility state, and a credibility loss that enters T is permanent. These are history-dependence results for this benchmark; they do not imply that history dependence is impossible under every exogenous process. Part (iii) separates three levels of evidence. Its first two sentences are general: they follow from Proposition 5(i) for all parameter values. Its last sentence is a conditional consequence: the invariance of $[C^*, 1]$ follows from the stated maintenance condition, which is a property of the optimal policy that the analysis does not derive from primitives. Whether the condition holds is a numerical finding. In the 36 combinations of Section 9.2, T and its complement are both nonempty in 11, T is empty in

17, and the complement is empty in 8. In each of the 28 combinations in which some level below C^* lies outside T, the condition holds, and from each of 25 initial levels between the boundary of T and C^* and 60 levels above C^* the optimal path reaches or stays in $[C^*, 1]$ and never leaves it within 600 periods. In the remaining 8, all with $\chi = 2L_F$, $T = [0, C^*)$, the authority never backs below C^*, and the condition fails: in 5 of them it never backs at all, and in 3 it backs only on an upper part of the band $\{C \geq C^* : (1-g)C < C^*\}$. From every computed level above C^* credibility then falls below C^* and remains in fiscal dominance. These are cases without rebuilding from below, not of rebuilding followed by collapse. Sustained monetary dominance after entry is therefore a property of the computed maintenance cases, not a general result. Because the benchmark counts $C = C^*$ as monetary dominance by convention, whereas Proposition 1 requires $C > C^*$, reaching C^* means entering $[C^*, 1]$; along the computed paths credibility stays strictly above C^* after entry, by at least 2.1×10^{-5}, so the convention does not affect these statements. Throughout, "reaching the threshold" means entering $[C^*, 1]$, and strict monetary dominance refers to $C_t > C^*$. In the baseline with $\chi = L_F$, $\beta_g = 0.90$, and $g = 0.10$, the exact solution gives $T = [0, 0.178)$ up to grid precision, and the optimal path from every computed level outside T reaches C^* and remains above it (Appendix A.8). The long-run backing frequency of Model EN from above the trap is 0.470. Model E with $\bar{q}$ equal to that frequency spends between 0.67 and 0.68 of periods 201 to 400 at or above C^* from every initial level considered, against none or all of them in Model EN. The independence from initial conditions does not depend on that choice of $\bar{q}$: with $\bar{q} = 0.30$, C^*, and 0.60 the shares are about 0.14, 0.49, and 0.95, again whatever the initial level, and the same holds over 1,200 periods. Part (iv) says only that the average belief equals the average backing frequency. It does not imply that beliefs are correct conditional on the state: in the benchmark, backing is a deterministic function of C_t while agents treat it as a Bernoulli draw with probability C_t. Nor does it establish a self-confirming equilibrium in the sense of Fudenberg and Levine (1993) or a Bayesian equilibrium.

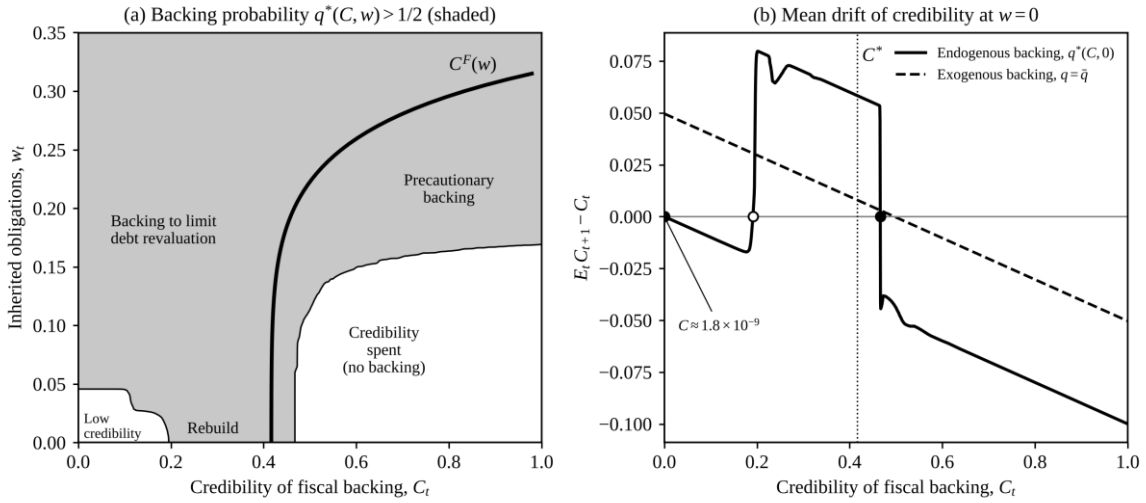
9.4 Credibility and debt: the optimal backing rule

Appendix A.9 solves (22) to (25) numerically on a grid for (C, w), with the calibration of Figure 1, $\sigma = 0.02$, a normal fiscal limit $\Psi(w) = \Phi\big((w - 0.30)/0.08\big)$, $\chi_0 = L_F$, $\chi_1 = 0.25L_F$, $\beta_g = 0.90$, $g = 0.10$, and $\varsigma = 0.05L_F$. The parameter values are illustrative, and

the results below are properties of the calibrated model. Figure 3(a) plots the states in which the probability of backing when feasible exceeds one half, together with the frontier $C^F(w)$. Four features stand out. The regime and the equilibrium inflation rate that enter (24) are computed with the frozen effective-backing approximation of Section 8, so the location of the frontier and of the regions of Figure 3(a) at positive obligations is conditional on the frozen effective-backing approximation.

First, low credibility is self-perpetuating only at low inherited obligations. In the fiscal-dominance region inflation revalues maturing debt each period, so that obligations remain close to zero, and the only cost of staying there is the loss from surplus disturbances. At $w = 0$, the probability of backing is below one half for credibility below about 0.195, close to the exact benchmark value of Section 9.3, and it falls to 1.8×10^{-9} at $C = 0$: from there, the sequence of costly backed periods needed to reach the frontier is not worth undertaking. Second, at higher obligations the authority backs with high probability at every level of credibility below the frontier, at least 0.99 for $w = 0.05$, because backing reduces the inflation required to revalue current obligations and because the loss from fiscal dominance rises with w^2. Debt therefore has two opposing effects on the incentive to rebuild: it raises the frontier, which lengthens rebuilding, and it raises the cost of remaining in fiscal dominance. With a loss that does not depend on debt, as in Section 9.2, only the first is present and the trap deepens with debt; with the loss (24), the second dominates in the calibration. Third, in the monetary-dominance region the authority does not back when obligations are low and credibility is comfortably above the frontier, so that it spends credibility, but it backs preventively once obligations rise: at $C = 0.70$ the probability of backing is 0.0001 at $w = 0.10$, 0.06 at $w = 0.15$, and 1.00 at $w = 0.20$. Backing then keeps debt away from the fiscal limit, where the frontier rises steeply and monetary dominance becomes infeasible. This precautionary motive is the joint management of willingness and capacity, and it does not arise when backing is exogenous and independent of the state. Fourth, with private cost shocks the backing decision is stochastic from the agents' point of view, but its probability depends on the state.

Figure 3. *Optimal backing policy and the mean drift of credibility*



Note. Panel (a): shaded regions are states in which the probability of backing when feasible, $q^*(C, w)$ from (25), exceeds one half; the thick line is the credibility frontier $C^F(w)$ of (20). Panel (b): expected change in credibility from (28) with w held at zero, under endogenous backing (q^* evaluated off the grid) and under exogenous backing with $\bar{q} = 0.496$, the long-run backing frequency of the endogenous model from high initial credibility. Markers indicate zeros of the drift; filled markers satisfy $dq^*/dC < 1$. The lowest zero lies at $C \approx 1.8 \times 10^{-9}$, not at zero. Calibration as in Figure 1, with $\sigma = 0.02$, $\Psi(w) = \Phi\big((w - 0.30)/0.08\big)$, $\chi_0 = L_F$, $\chi_1 = 0.25L_F$, $\beta_g = 0.90$, $g = 0.10$, and $\varsigma = 0.05L_F$ (Appendix A.9).

By (11), expected credibility evolves according to

$$E_t C_{t+1} - C_t = g[1 - \Psi(w_t)][q^*(C_t, w_t) - C_t] \tag{28}$$

Because $q^* > 0$ whenever $\varsigma > 0$, the drift at $C = 0$ is $g[1 - \Psi(0)]q^*(0,0) > 0$, so that $C = 0$ is not a rest point of (28). Holding w at zero, (28) has three zeros in the calibration, all computed off the grid (Appendix A.9). The lowest is a near-zero low-credibility rest point at $C \approx 1.8 \times 10^{-9}$, where $q^*(C, 0) = C$ to machine precision; the slope of the map $C \mapsto C + g[1 - \Psi(0)][q^*(C, 0) - C]$ there is 0.90, so it is locally stable under the stated updating map. The second, at about 0.191, is unstable, since $dq^*/dC \approx 34$, and separates the two regions. The third, at about 0.466, lies just above the threshold, where q^* falls steeply: it is stable under the continuous-time version of (28), since $dq^*/dC < 1$, but the slope of the discrete map with $g = 0.10$ is about -136, so credibility fluctuates around it, as under the

maintenance rule, rather than converging to it. These are fixed points of the mean drift with w held constant, not invariant distributions of the joint process for (C_t, w_t); they are related to, but weaker than, the expectational-stability concept of Evans and Honkapohja (2001). Under Model E the drift has a single zero, at $\bar{q}$. Table 4 reports finite-horizon simulations of the joint process. Starting from $w_0 = 0$, Model EN spends no time at or above the frontier in periods 301 to 600 from $C_0 \leq 0.15$ and at least 99.97 percent of it from $C_0 \geq 0.25$, with brief exits only, and the same holds over 1,200 periods. Model E spends about the same share from every initial level, whichever of three values of $\bar{q}$ is used. Because q^* is positive everywhere, the low region can be left in any period: none of 400 runs started at $C_0 = 0.05$ left it within 1,200 periods, and the most direct exit, 3 consecutive backed periods from the low rest point, has probability of about 10^{-18}. The low-credibility region is therefore highly persistent rather than absorbing, and the simulations do not establish permanence at an infinite horizon.

Table 4. *Share of periods in monetary dominance over finite horizons: endogenous versus exogenous backing*

Initial credibility, C_0	**0.05**	**0.15**	**0.25**	**0.35**	**0.50**	**0.70**
Periods 301–600						
Model EN	0.00	0.00	1.00	1.00	1.00	1.00
Model E, $\bar{q} = 0.496$	0.75	0.75	0.74	0.76	0.74	0.74
Model E, $\bar{q} = 0.398$	0.42	0.44	0.43	0.43	0.43	0.43
Model E, $\bar{q} = 0.30$	0.13	0.14	0.13	0.14	0.14	0.14
Periods 601–1,200						
Model EN	0.00	0.00	1.00	1.00	1.00	1.00
Model E, $\bar{q} = 0.496$	0.75	0.76	0.75	0.75	0.75	0.74
Model E, $\bar{q} = 0.398$	0.44	0.43	0.43	0.43	0.43	0.42
Model E, $\bar{q} = 0.30$	0.14	0.14	0.13	0.14	0.14	0.14

Note. Share of periods with $\tilde{C}_t \geq C^*$, averaged over 200 replications, from $w_0 = 0$. Model E draws backing, when feasible, independently with probability $\bar{q}$: 0.496 is the backing frequency of Model EN in feasible periods 301 to 600 of 200 runs from $C_0 = 0.60$; 0.398 is the same frequency pooled over 420 runs from 21 initial levels between 0 and 1; and 0.30 is an arbitrary lower value. Both models share the monetary block, the learning rule, the fiscal limit, and the shocks; finite-horizon shares do not establish long-run permanence. Calibration as in Figure 3.

Table 5 reports how the low-credibility region responds to political horizons, the speed of learning, fiscal capacity, and the cost of adjustment. Within the calibrated model, the backing threshold at $w = 0$ rises from 0.195 to 0.281 when β_g falls to 0.85 and to 0.311 when g falls to 0.05, and it vanishes for $\beta_g = 0.95$ or $g = 0.20$, in which case backing is more likely than not even at zero credibility. Fiscal capacity and the debt-dependent part of the political cost leave the threshold at $w = 0$ essentially unchanged, because they matter at higher obligations, where they shift the frontier and the region of precautionary backing in Figure 3(a). These are comparisons at a few parameter values, not monotonicity results.

Table 5. *Backing threshold, backing probability at zero credibility, and the frontier under alternative parameters*

Parameter change (others at baseline)	**Backing threshold at** $w = 0$	$q^*(0,0)$	**Frontier** $C^F(0.15)$
Baseline	0.195	1.8×10^{-9}	0.430
$\beta_g = 0.85$	0.281	1.9×10^{-9}	0.430
$\beta_g = 0.95$	0.000	0.86	0.430
$g = 0.05$	0.311	1.9×10^{-9}	0.430
$g = 0.20$	0.000	0.73	0.430
$\chi_1 = 0$	0.192	1.8×10^{-9}	0.430
$\chi_1 = 0.5L_F$	0.197	1.8×10^{-9}	0.430
Lower capacity, $\Psi(w) = \Phi\big((w - 0.25)/0.08\big)$	0.197	1.8×10^{-9}	0.466
Higher capacity, $\Psi(w) = \Phi\big((w - 0.35)/0.08\big)$	0.195	1.8×10^{-9}	0.419
$\varsigma = 0.02L_F$	0.195	1.4×10^{-22}	0.430

Note. The backing threshold is the lowest credibility level below the frontier at which $q^*(C, 0) = 1/2$, computed off the grid; a value of zero means that backing is more likely than not from every level. Baseline as in Figure 3 ($\beta_g = 0.90$, $g = 0.10$, $\chi_1 = 0.25L_F$, $\Psi(w) = \Phi\big((w - 0.30)/0.08\big)$, $\varsigma = 0.05L_F$).

Combining Propositions 1, 3, 4, and 5 yields a sequencing result, which is also the monetary rule assumed in the fiscal problem (22).

Corollary (sequencing). *In the environment of Propositions 1, 3, and 4, during rebuilding the monetary response that avoids the conflict region and minimizes the baseline loss is*

passive, $\varphi^*(\tilde{C}_t) < 1$. *Adopting* $\varphi > 1$ *before* C_t *crosses the frontier* $C^F(w_t)$*, which equals* C^* *in the benchmark, places the economy in the region without a bounded equilibrium. Within this environment, the order that sustains a transition is fiscal compliance first and monetary aggressiveness second.*

The corollary is an implication of the specified environment rather than a general prescription. It does not imply that monetary institutions are irrelevant during a stabilization. It implies that their contribution operates through what they allow the fiscal authority to do, rather than through the coefficient of the interest-rate rule, until the compliance record has moved credibility past the frontier.

9.5 What is endogenous and what is assumed

The feedback emphasized in this section runs from fiscal incentives to fiscal adjustment, from adjustment to credibility, from credibility to macroeconomic outcomes, and from outcomes back to incentives. Within the model, the authority's choice responds to incentives through (22) and (25); outcomes respond to credibility through the equilibrium of Propositions 1 to 4 under anticipated utility, which enters the authority's loss (24) and the law of motion of obligations; and outcomes feed back into incentives because credibility and inherited obligations are the state of (22). The link from adjustment to beliefs is assumed. The discounted Beta filter (11) is an operational learning rule, not the posterior of agents who know the authority's problem: agents do not use the fact that backing depends on (C_t, w_t), nor do they anticipate how their own beliefs will evolve. Under anticipated utility they treat C_t as the constant probability governing future backing, and the monetary block is solved under the same convention. The authority, by contrast, anticipates the agents' updating rule and the monetary response. These pieces are not jointly consistent in the sense of a general equilibrium with rational expectations: the authority's anticipation, the private expectations, the monetary rule, and the transition of debt are coherent only under the stated learning convention. When the fiscal limit binds, the regime and the within-period equilibrium are further conditional on the frozen effective-backing approximation of Section 8, under which future feasibility is evaluated at its current value.

Credibility is therefore an equilibrium object only in a limited sense. The private cost shocks make backing genuinely uncertain for agents. At the zeros of (28), the agents' belief equals the probability of backing that the authority chooses at that state, so that beliefs are not

revised on average; and by Proposition 6(iv), average beliefs equal average backing along any path. Neither property establishes that beliefs are correct conditional on the state, a self-confirming equilibrium, or expectational stability in the technical sense; the paper uses the more limited terms 'fixed point of the mean drift' and 'locally stable under the stated updating map'. A fully consistent equilibrium would require agents to learn about the authority's type or cost process with knowledge of its strategy, and would require Propositions 1 to 4 to be recomputed with beliefs that anticipate future backing, in which case the within-period equilibrium would no longer depend on C_t alone. The results of Sections 5 to 8 should accordingly be read as conditional on the anticipated-utility convention, and the fiscal extension as a partial closing of the loop, in which fiscal incentives are optimizing and beliefs are adaptive.

10. Argentina: A Brief Illustration

Argentina is used here as an illustration of the mechanisms, not as a test of them. Table 3 maps three periods into the regions of the model.

Table 3. *Argentina's recent regime sequence read through the model*

Period	Monetary stance	Fiscal credibility	Model region	Model-consistent pattern
1975–1991	Accommodating	Well below the frontier	Fiscal dominance	Inflation absorbs fiscal disturbances; tightening episodes reversed (Proposition 2)
1991–2001	Rule renouncing accommodation (convertibility)	Below a rising frontier	Conflict	Breakdown through default (Proposition 4)
2002–2023	Accommodating, with intermittent tightening	Low after default	Fiscal dominance, trap	Persistent inflation; short-lived disinflations (Propositions 2 and 5)

The chronic-inflation period corresponds to region (ii) of Proposition 1. With a fiscal record that left credibility far below the threshold, Proposition 2 predicts that episodes of monetary tightening unaccompanied by a durable change in fiscal behavior would at best produce transitory disinflation followed by a return of inflation, and Proposition 3 implies that an active Taylor response is not the variance-minimizing policy before fiscal credibility is restored.

The convertibility period is the most informative case for the framework, because it combined a monetary arrangement that renounced the accommodation margin with a fiscal

record that had not crossed the frontier. In the logic of the model, this configuration is analogous to the conflict region of Proposition 1, rather than a literal parameterization of the Taylor rule, and the model predicts no bounded equilibrium in that region under unchanged policies. Proposition 4 adds the mechanism through which the configuration resolved: rising debt raised the credibility frontier while the record of fiscal behavior did not lift credibility above it, so that the probability of breakdown rose until the regime ended in default. To the extent that the default was read as a decision rather than as pure inability, (12) implies that it lowered credibility multiplicatively, which is the channel through which the subsequent period inherited a low starting point.

The period after 2002 is consistent with the trap of the benchmark in Propositions 5 and 6. Starting from credibility depressed by default, the model predicts that a fiscal authority with a short political horizon would not undertake the sustained sequence of backed periods required to reach the threshold, that inflation would remain responsive to fiscal disturbances, and that monetary tightening would yield only transitory disinflation. Section 9.4 adds that such a low-credibility region is compatible with moderate measured obligations, because inflation itself keeps real obligations low and thereby weakens the incentive to adjust.

The illustration leaves the central quantities unmeasured, and the natural next step is empirical. The compliance indicator can be constructed from the record of fiscal targets and from primary-balance responses to debt estimated in the manner of Bohn (1998), and the credibility state can be filtered from that record using (11) or recovered from sovereign spreads. Proposition 2 implies a testable prediction: the cumulative response of the price level to a contractionary rule disturbance, identified as a shift in the rule rather than as a change in the observed policy rate and estimated by state-dependent local projections, should switch from negative to positive as the filtered credibility state falls below its threshold, whereas the impact response of inflation may be negative on both sides of the threshold. Proposition 6 adds predictions that distinguish endogenous from exogenous credibility: compliance should depend on the filtered credibility state, being absent at low credibility and concentrated near the frontier, and large losses of credibility should have persistent effects. That exercise is left for separate work.

11. Conclusion

This paper has modeled the credibility of fiscal backing as a belief that evolves with observed compliance, has embedded it in the active–passive framework of monetary–fiscal interaction, and has asked which of its implications require the fiscal authority to influence it. Under discounted Beta learning, the reputation stock and the subjective probability of backing are the same object, and the regime matrix becomes a function of it. Conditional on that state, fiscal policy is effectively passive only when credibility exceeds the threshold $C^* = r/\gamma$; a contractionary rule disturbance lowers the price level above the threshold and raises it below, although it can lower inflation on impact in both cases and lowers the equilibrium nominal rate above it; the coefficient that minimizes the inflation variance generated by surplus disturbances is passive below the threshold, while any active coefficient minimizes it above; and a stochastic fiscal limit converts the threshold into a frontier that rises with debt, conditional on a frozen effective-backing approximation that matches the local fiscal response at steady-state debt, does not establish a fully endogenous-feasibility equilibrium, and understates the credibility required as debt approaches the limit. These results hold for any process generating credibility and coincide, in first moments, with those of an economy with known partial backing. What endogeneity adds is an account of why governments build, maintain, or neglect credibility. In a deterministic benchmark in which the fiscal authority internalizes its effect on credibility, there is a trap set from which it never backs, entry into monetary dominance depends on history, and a credibility loss that enters the trap set is permanent, none of which occurs when backing is exogenous and independent over time; monetary dominance is sustained after entry in the computed cases with rebuilding, but not for all parameter values. In a stochastic extension with debt and private political-cost shocks, computed numerically, low credibility is highly persistent rather than absorbing, and governments with high credibility back preventively when debt rises.

The simplicity of the environment is deliberate, and it marks the natural extensions. Long-term debt and nominal rigidities would add the intertemporal reallocation of inflation emphasized by Sims (2011) and Cochrane (2023), would allow the threshold to depend on debt maturity, and would allow a contractionary rule disturbance to raise the real rate. A second credibility state for the monetary authority would allow the two reputations to interact. A fully consistent belief process, in which agents learn about the fiscal authority's type with knowledge of its strategy rather than under the anticipated-utility approximation

of Cogley and Sargent (2008), would close the remaining link of the feedback identified in Section 9.5, and making the fiscal backing decision depend on observed thresholds would connect the framework to the endogenous regime changes of Davig and Leeper (2008). And the empirical program outlined in Section 10 would discipline the threshold and the political parameters with data. The central message should survive these extensions: whether a monetary action stabilizes or destabilizes depends on a state variable, the credibility of fiscal backing, that is built by fiscal behavior, that fiscal authorities have incentives to build only under some conditions, and that monetary policy alone cannot supply.

Appendix

A.1 The linearized budget constraint

With one-period nominal debt B_t and price level P_t, the real flow constraint is $B_t/P_t = (1 + i_{t-1})B_{t-1}/P_t - S_t$. Linearizing around a steady state with zero inflation, $1 + i = R$ and real debt $\bar{b}$, and writing b_t and s_t for deviations of real debt and the real surplus, gives

$$b_t = R\big[b_{t-1} + \bar{b}\,(\hat{\imath}_{t-1} - \pi_t)\big] - s_t \tag{A.1}$$

which is (7) with $w_t \equiv b_{t-1} + \bar{b}\,\hat{\imath}_{t-1}$. Using $w_{t+1} = b_t + \bar{b}\,\hat{\imath}_t$, the rule (6), and the fiscal rule (8) yields (9).

A.2 Discounted Beta learning

Let the prior over the backing probability be Beta with pseudo-counts (α_t, δ_t) and $n_t = \alpha_t + \delta_t$. With forgetting factor $1 - g$, the update is $\alpha_{t+1} = (1 - g)\alpha_t + \xi_t$ and $\delta_{t+1} = (1 - g)\delta_t + (1 - \xi_t) + \kappa_D D_t$. The effective sample converges to $n = 1/g$ in the absence of default. This discounted (constant-gain) filter is not the standard posterior for a fixed Bernoulli probability, since it deliberately downweights older observations. Evaluating the posterior mean $C_{t+1} = \alpha_{t+1}/n_{t+1}$ at $n_t = 1/g$,

$$C_{t+1} = \frac{(1 - g)\,C_t/g + \xi_t}{1/g + \kappa_D D_t} = \frac{(1 - g)\,C_t + g\,\xi_t}{1 + g\,\kappa_D D_t} \tag{A.2}$$

which is (12), and (11) when $D_t = 0$. After a default, $n_{t+1} = 1/g + \kappa_D$ exceeds its stationary value; (12) treats it as renormalized to $1/g$ before the next update, so that C_t remains a sufficient scalar state.

A.3 Proof of Proposition 1 and the mean-square refinement

Under the timing of Section 4.2, C_t is fixed at the beginning of the period, ξ_t is realized within it, and expectations about ξ_{t+1} are formed with C_t. Guess $\pi_t = \psi(\xi_t)w_t + \theta u_t + \kappa\varepsilon_t$. Since ξ_{t+1} is independent of w_{t+1} under the agents' beliefs, $E_t\pi_{t+1} = \bar{\psi}\,w_{t+1} + \theta\rho\,u_t$ with $\bar{\psi} = C_t\psi(1) + (1 - C_t)\psi(0)$. Substituting (6) and (9) into (5),

$$\big[\varphi - \bar{\psi}\,\bar{b}(\varphi - R)\big]\pi_t = \bar{\psi}\,\lambda(\xi_t)\,w_t + \big(\bar{\psi}\,\bar{b} + \theta\rho - 1\big)\,u_t - \bar{\psi}\,\varepsilon_t \tag{A.3}$$

Matching coefficients on w_t and averaging over ξ_t gives $\bar{\psi}\big[\varphi - \bar{\psi}\,\bar{b}(\varphi - R) - \bar{\lambda}\big] = 0$. Either $\bar{\psi} = 0$ or $\bar{\psi} = \big(\bar{\lambda} - \varphi\big)/\big[\bar{b}(R - \varphi)\big]$.

In the first solution inflation is independent of debt, the Fisher block gives $\theta = -1/(\varphi - \rho)$, and $E_t w_{t+1} = \bar{\lambda}\, w_t + \cdots$, which is bounded in expectation if and only if $\bar{\lambda} < 1$, that is, $C_t > C^*$. The Fisher block has a unique bounded solution if and only if $\varphi > 1$. In the second solution, substituting $\psi(\xi_t)$ into (9) gives $w_{t+1} = (\varphi/\bar{\lambda})\lambda(\xi_t) w_t + \cdots$, whose conditional mean coefficient is φ, so that the solution is bounded if and only if $\varphi < 1$; $\bar{\psi} > 0$ requires $\bar{\lambda} > \varphi$, which holds when $C_t < C^*$ and $\varphi < 1$. Matching the remaining coefficients gives $\kappa = -\bar{\psi}/\bar{\lambda}$ and $\theta = (\bar{\psi}\,\bar{b} - 1)/(\bar{\lambda} - \rho) = -C_t\gamma/[(R - \varphi)(\bar{\lambda} - \rho)]$. Region (i) admits only the first bounded solution, region (ii) only the second, region (iii) neither, and region (iv) both, which establishes the proposition. The same steps apply to an economy with deterministic partial backing, $s_t = \delta\gamma w_t + \varepsilon_t$: $\lambda(\xi_t)$ is replaced by $R - \delta\gamma$ in (9) and (A.3), and with $\delta = C_t$ the coefficients $\bar{\psi}$, θ, and κ and the first-moment conditions are unchanged. Only the mean-square conditions below, which depend on $m(C)$, differ, since $m = 1$ under deterministic backing.

Under mean-square stability (Farmer et al., 2009), the first solution requires $E[\lambda(\xi)^2] < 1$, or $C_t > C^*_{ms} \equiv (R^2 - 1)/[R^2 - (R - \gamma)^2]$, and the second requires $\varphi^2 m(C_t) < 1$, where $m(C) = E[\lambda(\xi)^2]/\bar{\lambda}^2 = 1 + C(1 - C)\gamma^2/\bar{\lambda}^2$. In the calibration, $C^* = 0.417$ and $C^*_{ms} = 0.429$, and $1/\sqrt{m(C)}$ exceeds 0.998 throughout the fiscal-dominance region.

A.4 Proof of Proposition 2

Part (i) follows from the first solution in A.3. For part (ii), taking expectations of (9) under the second solution, $E_t w_{t+k+1} = \varphi\, E_t w_{t+k} + \bar{b}\,(R - \rho)\rho^k u_t/(\bar{\lambda} - \rho)$, using $(\varphi - R)\theta + 1 = (R - \rho)/(\bar{\lambda} - \rho)$. Summing over k and using $E_t \pi_{t+k} = \bar{\psi}\, E_t w_{t+k} + \theta\rho^k u_t$,

$$\sum_{k=0}^{\infty} E_t\, \pi_{t+k} = \frac{1}{1 - \rho}\left[\frac{\bar{\psi}\,\bar{b}\,(R - \rho)}{(1 - \varphi)(\bar{\lambda} - \rho)} + \theta\right] u_t \tag{A.4}$$

which simplifies to (16). Since $\bar{\lambda} > 1 > \varphi$ in the region, $(\bar{\lambda} - \varphi)/(1 - \varphi) > 1$, so the numerator of (16) exceeds $R - \rho - C_t\gamma = \bar{\lambda} - \rho > 0$, and the denominator is positive. For part (iii), the rule (6) gives the impact effect on the nominal rate, $\varphi\ \partial\pi_t/\partial u_t + 1$. In case (i) it equals $1 - \varphi/(\varphi - \rho) = -\rho/(\varphi - \rho)$, which also equals $\rho\ \partial\pi_t/\partial u_t$, as (5) requires. In case (ii) it equals $1 + \varphi\theta$, and $(R - \varphi)(\bar{\lambda} - \rho) - \varphi C_t\gamma = (R - \varphi)(R - \rho) - RC_t\gamma$ gives the expression in the text. The sign condition holds in the calibration of Figure 2, with impact effects 1.00 and 0.95 for $C = 0$ and $C = 0.3$, but can fail when φ and ρ are both close to

one. All expressions in Propositions 1 and 2 and in this appendix were verified by symbolic substitution.

A.5 Proof of Proposition 3 and numerical verification

Under the second solution with surplus disturbances only, $\pi_t = \psi(\xi_t)w_t + \kappa\varepsilon_t$ and $w_{t+1} = a(\xi_t)w_t - \left(\varphi/\bar{\lambda}\right)\varepsilon_t$ with $a(\xi) = \varphi\lambda(\xi)/\bar{\lambda}$. Since $E[a^2] = \varphi^2 m(C)$, the stationary variance of w is $\varphi^2\sigma^2/\left[\bar{\lambda}^2(1-\varphi^2 m)\right]$ when $\varphi^2 m < 1$. Because w_t is independent of (ξ_t, ε_t) and $E[\psi(\xi)^2] = \bar{\psi}^2 m$,

$$V = \bar{\psi}^2 m\,\frac{\varphi^2\sigma^2}{\bar{\lambda}^2(1-\varphi^2 m)} + \frac{\bar{\psi}^2\sigma^2}{\bar{\lambda}^2} = \frac{\bar{\psi}^2\sigma^2}{\bar{\lambda}^2(1-\varphi^2 m)} \tag{A.5}$$

which gives (17) after substituting $\bar{\psi}$. The logarithmic derivative is $-2/\left(\bar{\lambda}-\varphi\right) + 2/(R-\varphi) + 2m\varphi/(1-m\varphi^2)$, which equals $-2C_t\gamma/\left(\bar{\lambda}R\right)$ at $\varphi = 0$; expanding it to first order in φ and setting it to zero gives (18), which establishes parts (i) and (ii).

For part (iii), write $m = 1+\mu$ and $\bar{\lambda} = 1+\epsilon$, with $\epsilon = (C^* - C_t)\gamma$. At the threshold ($\epsilon = 0$), the factor $(1-\varphi)^2/(1-m\varphi^2)$ is minimized at $\varphi = 1/m$, where it equals $\mu/(1+\mu)$, so that

$$V(1/m; C^*) = \frac{\sigma^2}{\bar{b}^2}\,\frac{\mu}{(1+\mu)(R-1/m)^2}, \qquad \mu = C^*(1-C^*)\gamma^2 = r^2\,\frac{1-C^*}{C^*} \tag{A.6}$$

To leading order in r, this equals $\left(\sigma^2/\bar{b}^2\right)(1-C^*)/C^*$, while the near-peg branch attains a value close to $V(0) = \sigma^2\beta^2/\bar{b}^2$. To leading order, the second minimum is therefore lower at the threshold, and by continuity in a neighborhood below it, when $(1-C^*)/C^* < \beta^2 \approx 1$, that is, $C^* > 1/2$; the exact boundary is reported below. At the second minimum, $1 - m\varphi^2 = \mu/(1+\mu)$ is small, so that the variance of w derived above is large.

The analytical characterization was checked by global numerical minimization of (17). For each C, V was evaluated on a dense grid covering the whole admissible interval $[0, 1/\sqrt{m(C)})$, with additional logarithmically spaced points near its upper bound, and the grid minimum was refined by bounded minimization. The closed form (17) was also verified by simulation. The numerical verification is restricted to parameter pairs satisfying the maintained assumption $\gamma > r = \beta^{-1} - 1$, equivalently $C^* < 1$. Of the pairs in the grid $\beta \in \{0.90, 0.96, 0.99\}$ and $\gamma \in \{0.05, 0.10, 0.30\}$, seven are admissible: $\beta = 0.90$ with $\gamma = 0.30$,

and $\beta \in \{0.96,0.99\}$ with all three values of γ. Over these seven pairs, three results hold. First, the near-peg branch is increasing in C in every case, and the error of (18) on that branch is below 0.004 (below 0.0002 for $\beta \geq 0.96$). Second, the global minimum coincides with the near-peg branch throughout the fiscal-dominance region in the six pairs with $C^* < 1/2$. Third, the single pair with $C^* > 1/2$, namely $\beta = 0.96$ and $\gamma = 0.05$ ($C^* = 0.833$), displays the switch of part (iii): the global optimum moves to $\varphi^* \approx 0.999$ for $C \geq \hat{C} \approx 0.820$, while (18) gives about 0.043 there. A bisection over γ, within the admissible range, locates the exact switching boundary at $C^* = 0.497$, 0.500, and 0.500 for $\beta = 0.90$, 0.96, and 0.99, close to the leading-order value $1/2$ but weakly dependent on β. The baseline calibration of Figures 1 and 2 ($C^* = 0.42$) lies on the near-peg side. In the monetary-dominance region, adding a real-rate disturbance $z_t = \rho z_{t-1} + e_t^z$ to (5) gives $\pi_t = (z_t - u_t)/(\varphi - \rho)$, whose variance is decreasing in φ.

A.6 Proof of Proposition 4

Parts (i) and (ii). Under the frozen effective-backing approximation, agents' expected surplus response at every future date is $\tilde{C}_t \gamma w_{t+j}$, and the effective backing indicator $\xi_t F_t$, where F_t indicates feasibility, is a Bernoulli variable with probability $\tilde{C}_t$ given w_t. The proof of Proposition 1 then applies with $\tilde{C}_t$ in place of C_t, including the mean-square refinement with $m(\tilde{C}_t)$, which gives part (i); Proposition 3 with $\tilde{C}_t$ gives part (ii). The monotonicity of (20) follows from Ψ being non-decreasing. *Exactness and local coincidence.* (a) If $\Psi(w) = p$ for all relevant w, then $E[\xi_t F_t \mid w_t] = C_t(1-p)$ at every date, the frozen approximation is an identity, and Proposition 1 holds with $C_t(1-p)$: substituting the solution (14) with $C_t(1-p)$ into (5), (6), and (9) leaves a zero residual for both realizations of the effective backing indicator. (b) Let $h(w) = [1 - \Psi(w)]w$, so that the expected backed surplus is $C_t \gamma h(w)$. Then $h'(w) = 1 - \Psi(w) - \Psi'(w)w$ and $h'(0) = 1 - \Psi(0)$, the frozen slope at $w = 0$, so the first-order fiscal response around $w = 0$ coincides with that of the frozen approximation; this is a local coincidence at steady-state debt, not a dynamic equivalence. (c) Assume $w \geq 0$ and Ψ differentiable and non-decreasing, $\Psi' \geq 0$. Then $h(w)/w - h'(w) = \Psi'(w)w \geq 0$, so the frozen slope weakly exceeds the marginal slope, and, where the denominator is positive with $h'(w) > C^*$, the frontier $C^*/h'(w)$ weakly exceeds (20); where $h'(w) \leq C^*$, no credibility satisfies the marginal condition. These expressions were checked by symbolic computation. *Part (iii).* By (11), $C_{t+1} = (1-g)C_t + g\,\xi_t$ in a period

in which backing is feasible. By (9), $w_{t+1} = \lambda(\xi_t)w_t + \bar{b}(\varphi - R)\pi_t + \bar{b}u_t - \varepsilon_t$, and in monetary dominance $\pi_t = -u_t/(\varphi - \rho)$ does not depend on ξ_t or w_t, so that, for given shocks, w_{t+1} is larger by γw_t after non-backing. The economy leaves the region at $t+1$ when $C_{t+1} < C^F(w_{t+1})$. Conditioning on ξ_t, with $\Pr(\xi_t = 1) = q_t$, gives (21) by the law of total probability, exactly for the region defined by (20). Because the shocks are independent of (C_t, w_t, ξ_t) and $C^F(w_{t+1})$ does not depend on C_t, both conditional events shrink as C_t rises. For $w_t \geq 0$, the event after non-backing contains the event after backing, because credibility is lower and debt is higher, so (21) is non-increasing in q_t. Since $\lambda(\xi_t) > 0$ and C^F is non-decreasing, both events expand as w_t rises. Agents, who do not observe q_t, evaluate the same expression with their estimate C_t.

A.7 Proof of Proposition 5

Rebuilding yields $-(\chi + L_F)\left(1 - \beta_g^n\right)/\left(1 - \beta_g\right) - \beta_g^n \chi C^*/\left(1 - \beta_g\right)$, and not rebuilding yields $-L_F/\left(1 - \beta_g\right)$, since without backing credibility decays toward zero and remains below the threshold. Rebuilding is preferred if and only if

$$\beta_g^n[L_F + \chi(1 - C^*)] \geq \chi \tag{A.7}$$

that is, $n(C_0) \leq \lfloor \bar{n} \rfloor$. By (26) this holds if and only if $(1 - C^*)/(1 - C_0) \geq (1 - g)^{\lfloor \bar{n} \rfloor}$, which gives (27). The comparative statics follow because $\bar{n}$ is increasing in β_g and L_F and decreasing in χ, and $(1 - g)^{-\lfloor \bar{n} \rfloor}$ is increasing in g. Replacing C^* by a higher threshold lowers $1 - C^*$ directly and lowers $\bar{n}$, so that C^{trap} is increasing in the threshold, which gives the statement for the frontier $C^F(w)$ in the text.

The derivation is exact given the maintenance cost $\chi C^*/\left(1 - \beta_g\right)$. This proves Proposition 5(ii); part (i) is proved in Appendix A.8, which also computes the exact cost.

A.8 Proofs of Propositions 5(i) and 6, and exact solution of the benchmark

Numerical solution. The benchmark Bellman equation of Section 9.2 was solved by value function iteration on a grid of 40,001 equally spaced points on [0,1], evaluating the continuation value at the deterministic transitions $(1 - g)C$ and $(1 - g)C + g$ by linear interpolation, with a sup-norm tolerance of 10^{-12}. For the baseline ($\chi = L_F, \beta_g = 0.90, g = 0.10$), the upper end of T is 0.1778, 0.1777, and 0.1777 on grids of 10,001, 40,001, and 160,001 points, and $Q_1 - Q_0$ changes sign between adjacent grid points at the boundary. For

each of the 36 parameter combinations of Section 9.2 (solved on 20,001 points), the policy was classified below and above C^*, the maintenance rule was simulated for 3,000 periods to obtain its long-run backing frequency and to verify that credibility never falls below C^*, the present cost of that rule from C^* was computed, and the exact threshold was compared with (27). For the baseline, the computed policy backs from every level at or above the boundary of T below C^*, and simulation from 82 initial levels between that boundary and 0.99 confirms that credibility stays at or above C^* once it has reached it; this last property is numerical. For each of the 36 combinations, the same procedure was applied from 25 initial levels between the boundary of T and C^*, when that interval is nonempty, and from 60 levels between C^* and 0.999, over 600 periods, recording first entry into $[C^*, 1]$, any later exit, and entry into T; the counts are reported in Section 9.3.

Proof of Proposition 5(i). For $C' > C$, applying from C' the action sequence that is optimal from C yields a credibility path that is at least as high in every period, because (11) is increasing in C_t; losses are non-increasing in credibility and costs are identical, so $V(C') \leq V(C)$. Never backing from $C < C^*$ yields exactly $L_F/(1-\beta_g)$, because $(1-g)^t C < C^*$ for all t; never backing from any C yields at most this value, so $V \leq L_F/(1-\beta_g)$ everywhere, and T is the set of $C < C^*$ at which the bound is attained. If $C \in T$ and $C' < C$, then $L_F/(1-\beta_g) \geq V(C') \geq V(C) = L_F/(1-\beta_g)$, so T is an interval with lower end zero. From $C \in T$ never backing attains $V(C)$, and the path $(1-g)^t C$ remains in T; with ties broken in favour of not backing, this is the selected plan. Optimal plans that back from $C \in T$ can exist only when backing and not backing are exactly indifferent. From $C < C^*$ outside T, $V(C) < L_F/(1-\beta_g)$; if an optimal plan never reached C^*, every period would cost at least L_F and $V(C) \geq L_F/(1-\beta_g)$, a contradiction. Finally, $V(C; \chi, L_F) = L_F \tilde{V}(C; \chi/L_F)$, where $\tilde{V}$ is the value with unit loss, which is non-decreasing in χ/L_F and bounded by $1/(1-\beta_g)$, so T does not shrink as χ/L_F rises. *Proof of Proposition 6.* Part (i): under anticipated utility the within-period equilibrium depends only on (C_t, w_t) and the realized ξ_t, so Propositions 1 to 4 apply date by date; setting the exogenous backing outcomes equal to those realized in Model EN reproduces its path of credibility, but not the probabilities with which those outcomes occur. Part (ii): with $\Psi = 0$ every period is feasible and (11) gives $C_{t+1} - C'_{t+1} = (1-g)(C_t - C'_t)$ for common ξ_t, which yields the contraction. Iterating, $C_t = (1-g)^t C_0 + g \sum_{s<t} (1-g)^{t-1-s} \xi_s$; the distribution of the second term does not

depend on C_0 or on the fiscal parameters, it converges because the weights are summable, its limit is unique by the contraction, and its mean converges to $\bar{q}$. Part (iii): its first two statements follow from Proposition 5(i). Paths from T remain in T under the stated tie-breaking, and paths from outside T reach $[C^*, 1]$, so two initial levels on either side of the boundary differ in whether monetary dominance is ever reached; a fall into T is permanent, whereas in Model E its effect decays at rate $1 - g$ by part (ii). For the last statement, let $0 < g \leq 1$, $0 < C^* < 1$, and $C \geq C^*$. If $(1-g)C \geq C^*$, then $C_{t+1} \geq (1-g)C \geq C^*$ whatever the action; after non-backing, $C_{t+1} = C^*$ only if $(1-g)C = C^*$. If $(1-g)C < C^*$, the condition requires backing, and $C_{t+1} = (1-g)C + g \geq (1-g)C^* + g = C^* + g(1 - C^*) > C^*$, strictly, since $g(1-C^*) > 0$. Hence $[C^*, 1]$ is invariant under any policy satisfying the condition. Proposition 5(i) does not imply the condition, and the computations of Section 9.3 show that it fails in some parameter combinations. Part (iv): in a feasible period, (11) gives $C_{t+1} - C_t = g(\xi_t - C_t)$, and in an infeasible period $C_{t+1} = C_t$. Summing over the N feasible periods in a window, $\sum(\xi_s - C_s) = (C_{\text{end}} - C_{\text{start}})/g$, so the two averages over those periods differ by at most $1/(gN)$. The identity concerns averages only and carries no information about conditional correctness of beliefs.

A.9 Numerical solution of the problem with debt

Loss identities. With $\pi_t = \psi(\xi_t)w_t + \kappa\varepsilon_t$, $E[\pi_t \mid w_t, \xi_t] = \psi(\xi_t)w_t$, $\text{Var}[\pi_t \mid w_t, \xi_t] = \kappa^2\sigma^2$, and $E[\pi_t^2 \mid w_t, \xi_t] = \psi(\xi_t)^2 w_t^2 + \kappa^2\sigma^2$, which is (24) up to normalization. From the agents' perspective, with ξ_t equal to one with probability C, $E[\psi(\xi_t)] = \bar{\psi}$ and $E[\psi(\xi_t)^2] = \bar{\psi}^2 m(C)$. In fiscal dominance, $w_{t+1} = (\varphi/\bar{\lambda})\lambda(\xi_t)w_t - (\varphi/\bar{\lambda})\varepsilon_t$, so the stationary second moment is $E[w^2] = \varphi^2\sigma^2/[\bar{\lambda}^2(1 - m\varphi^2)]$, and, because w_t is independent of the current ξ_t and ε_t, $E[\pi_t^2] = \bar{\psi}^2 m\, E[w^2] + \kappa^2\sigma^2 = \bar{\psi}^2\sigma^2/[\bar{\lambda}^2(1 - m\varphi^2)]$, which equals (17). Hence, in the fiscal-dominance region with C fixed and under the stationarity conditions of Section 7, the expected value of the loss (24) is $E[L] = (L_F/s_0^2)E[\pi_t^2] = (L_F/s_0^2)\text{Var}(\pi_t)$, that is, (17) multiplied by L_F/s_0^2, with $s_0^2 = \beta^2\sigma^2/\bar{b}^2$; the identity does not extend to paths along which credibility varies. These identities were checked by symbolic computation and by simulation. *Algorithm.* Problem (22) to (25) was solved by value function iteration on a grid for (C, w), with bilinear interpolation, a Gauss–Hermite rule for ε_t, and the exact expectation of the minimum over the mean-zero extreme-value cost shocks. In infeasible periods there is no cost shock, credibility is not updated, and $Q_0^{\emptyset}$ uses C in place of $(1-g)C$.

Values of w_{t+1} outside the grid are truncated to its bounds; in the baseline 2.0 percent of the next-period quadrature nodes fall outside, and widening the bounds to $[-0.30,0.70]$ leaves the reported results unchanged (Table A1). The rule disturbance and real-rate disturbances are set to zero, so that inflation is zero in monetary dominance; below the frontier the monetary coefficient is $\varphi^p(\tilde{C}_t)$ from (18), and above it $\varphi = 1.5$. In the baseline, iteration stops after 198 iterations, and the Bellman residual is $\| TV - V \|_\infty = 8.7 \times 10^{-10}$, where V is the final iterate. The discretized operator T, which combines bilinear interpolation and quadrature with nonnegative weights, the truncation of w, and the expected minimum over the cost shocks, is a contraction with modulus β_g, so the distance between V and the fixed point V^* of that operator satisfies $\| V - V^* \|_\infty \leq \| TV - V \|_\infty / (1 - \beta_g) \approx 8.7 \times 10^{-9}$. This bounds the error relative to the discretized problem, not the approximation error relative to the continuous model, which is assessed by the sensitivity analysis below. The probability q^*, the backing threshold of Table 5, and the zeros of (28) are computed off the grid, by evaluating Q_0 and Q_1 at arbitrary (C, w) with the converged V as continuation and computing $\ln q^* = -\ln[1 + \exp((Q_1 - Q_0)/\varsigma)]$ stably; the near-zero zero of (28) is found by root-finding in $\ln C$. Simulations use bilinear interpolation of $\ln q^*$ on the grid, 200 replications per initial level, a common seed for both models, and horizons of 600 and 1,200 periods with the first half discarded. *Sensitivity.* Table A1 reports the main numerical objects under alternative numerical choices. The backing threshold and the unstable zero vary by less than 0.02 across the linear-interpolation variants. Cubic-spline interpolation lowers the threshold and produces additional zeros near the frontier, because the spline oscillates around the kinks of V at the frontier; linear interpolation, which preserves monotonicity, is used for the results in the text. The near-zero zero equals $q^*(0,0)$ to within its own precision, and its magnitude depends strongly on ς, while its existence and positivity do not. Replication files in Python, with a README, a parameter file, and regression tests that fail if a number reported in the text differs from the reproduced output, are available from the author on request.

Table A1. *Numerical sensitivity of the problem with debt*

Variant	Threshold at $w = 0$	$q^*(0,0)$	Unstable zero	Upper zero	Bellman residual
Baseline	0.195	1.8×10^{-9}	0.191	0.466	8.7×10^{-10}
401×281 grid	0.197	1.9×10^{-9}	0.183	0.464	8.7×10^{-10}

Variant	**Threshold at $w = 0$**	$q^*(0,0)$	**Unstable zero**	**Upper zero**	**Bellman residual**
801×561 grid	0.199	2.0×10^{-9}	0.183	0.464	8.7×10^{-10}
$w \in [-0.30, 0.70]$	0.195	1.8×10^{-9}	0.191	0.466	8.7×10^{-10}
5 Gauss–Hermite nodes	0.195	1.8×10^{-9}	0.191	0.466	8.7×10^{-10}
15 Gauss–Hermite nodes	0.195	1.8×10^{-9}	0.191	0.466	8.7×10^{-10}
Tolerance 10^{-11}	0.195	1.8×10^{-9}	0.191	0.466	8.6×10^{-12}
Cubic-spline interpolation	0.176	2.1×10^{-9}	0.168	0.468 (+4)	8.4×10^{-10}
$\varsigma = 0.02 L_F$	0.195	1.4×10^{-22}	0.195	0.466	8.7×10^{-10}
$\varsigma = 0.10 L_F$	0.194	4.2×10^{-5}	0.182	0.466	8.7×10^{-10}

Note. Baseline: 201×141 grid on $[0,1] \times [-0.20, 0.50]$, 9 Gauss–Hermite nodes, tolerance 10^{-9}, bilinear interpolation, $\varsigma = 0.05 L_F$. Each row changes one element. Zeros of (28) at $w = 0$ other than the near-zero one, which equals $q^*(0,0)$ up to its precision; '(+k)' indicates k additional zeros.

References

Aiyagari, S. R., & Gertler, M. (1985). The backing of government bonds and monetarism. *Journal of Monetary Economics, 16*(1), 19–44. https://doi.org/10.1016/0304-3932(85)90004-2

Alesina, A., & Tabellini, G. (1990). A positive theory of fiscal deficits and government debt. *The Review of Economic Studies, 57*(3), 403–414. https://doi.org/10.2307/2298021

Barro, R. J., & Gordon, D. B. (1983). Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics, 12*(1), 101–121. https://doi.org/10.1016/0304-3932(83)90051-X

Besley, T., & Persson, T. (2011). *Pillars of prosperity: The political economics of development clusters*. Princeton University Press. https://doi.org/10.1515/9781400840526

Bi, H. (2012). Sovereign default risk premia, fiscal limits, and fiscal policy. *European Economic Review, 56*(3), 389–410. https://doi.org/10.1016/j.euroecorev.2011.11.001

Bi, H., Leeper, E. M., & Leith, C. (2013). Uncertain fiscal consolidations. *The Economic Journal, 123*(566), F31–F63. https://doi.org/10.1111/ecoj.12011

Bianchi, F. (2012). Evolving monetary/fiscal policy mix in the United States. *American Economic Review, 102*(3), 167–172. https://doi.org/10.1257/aer.102.3.167

Bianchi, F., & Ilut, C. (2017). Monetary/fiscal policy mix and agents' beliefs. *Review of Economic Dynamics, 26*, 113–139. https://doi.org/10.1016/j.red.2017.02.011

Bianchi, F., & Melosi, L. (2017). Escaping the Great Recession. *American Economic Review, 107*(4), 1030–1058. https://doi.org/10.1257/aer.20160186

Blanchard, O. J., & Fischer, S. (1989). *Lectures on macroeconomics*. MIT Press.

Bohn, H. (1998). The behavior of U.S. public debt and deficits. *The Quarterly Journal of Economics, 113*(3), 949–963. https://doi.org/10.1162/003355398555793

Canzoneri, M. B., Cumby, R. E., & Diba, B. T. (2001). Is the price level determined by the needs of fiscal solvency? *American Economic Review, 91*(5), 1221–1238. https://doi.org/10.1257/aer.91.5.1221

Chung, H., Davig, T., & Leeper, E. M. (2007). Monetary and fiscal policy switching. *Journal of Money, Credit and Banking, 39*(4), 809–842. https://doi.org/10.1111/j.1538-4616.2007.00047.x

Clarida, R., Galí, J., & Gertler, M. (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory. *The Quarterly Journal of Economics, 115*(1), 147–180. https://doi.org/10.1162/003355300554692

Cochrane, J. H. (2023). *The fiscal theory of the price level*. Princeton University Press. https://doi.org/10.1515/9780691243245

Cogley, T., & Sargent, T. J. (2008). Anticipated utility and rational expectations as approximations of Bayesian decision making. *International Economic Review, 49*(1), 185–221. https://doi.org/10.1111/j.1468-2354.2008.00477.x

Cukierman, A. (1992). *Central bank strategy, credibility, and independence: Theory and evidence*. MIT Press.

Davig, T., & Leeper, E. M. (2007). Generalizing the Taylor principle. *American Economic Review, 97*(3), 607–635. https://doi.org/10.1257/aer.97.3.607

Davig, T., & Leeper, E. M. (2008). Endogenous monetary policy regime change. In L. Reichlin & K. West (Eds.), *NBER International Seminar on Macroeconomics 2006* (pp. 345–391). University of Chicago Press. https://doi.org/10.1086/653984

Demirel, U. D. (2012). The value of monetary policy commitment under imperfect fiscal credibility. *Journal of Economic Dynamics and Control, 36*(6), 813–829. https://doi.org/10.1016/j.jedc.2012.02.006

Evans, G. W., & Honkapohja, S. (2001). *Learning and expectations in macroeconomics*. Princeton University Press. https://doi.org/10.1515/9781400824267

Farmer, R. E. A., Waggoner, D. F., & Zha, T. (2009). Understanding Markov-switching rational expectations models. *Journal of Economic Theory, 144*(5), 1849–1867. https://doi.org/10.1016/j.jet.2009.05.004

Fudenberg, D., & Levine, D. K. (1993). Self-confirming equilibrium. *Econometrica, 61*(3), 523–545. https://doi.org/10.2307/2951716

Hommes, C., & Lustenhouwer, J. (2019). Inflation targeting and liquidity traps under endogenous credibility. *Journal of Monetary Economics, 107*, 48–62. https://doi.org/10.1016/j.jmoneco.2019.01.027

Isard, P., Laxton, D., & Eliasson, A.-C. (2001). Inflation targeting with NAIRU uncertainty and endogenous policy credibility. *Journal of Economic Dynamics and Control, 25*(1–2), 115–148. https://doi.org/10.1016/S0165-1889(99)00071-8

Kydland, F. E., & Prescott, E. C. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy, 85*(3), 473–491. https://doi.org/10.1086/260580

Leeper, E. M. (1991). Equilibria under active and passive monetary and fiscal policies. *Journal of Monetary Economics, 27*(1), 129–147. https://doi.org/10.1016/0304-3932(91)90007-B

Lucas, R. E., Jr. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy, 1*, 19–46. https://doi.org/10.1016/S0167-2231(76)80003-6

North, D. C., & Weingast, B. R. (1989). Constitutions and commitment: The evolution of institutions governing public choice in seventeenth-century England. *The Journal of Economic History, 49*(4), 803–832. https://doi.org/10.1017/S0022050700009451

Persson, T., & Tabellini, G. (2000). *Political economics: Explaining economic policy*. MIT Press.

Reinhart, C. M., & Rogoff, K. S. (2009). *This time is different: Eight centuries of financial folly*. Princeton University Press. https://doi.org/10.1515/9781400831722

Sargent, T. J., & Wallace, N. (1981). Some unpleasant monetarist arithmetic. *Federal Reserve Bank of Minneapolis Quarterly Review, 5*(3), 1–17. https://doi.org/10.21034/qr.531

Sims, C. A. (1994). A simple model for study of the determination of the price level and the interaction of monetary and fiscal policy. *Economic Theory, 4*(3), 381–399. https://doi.org/10.1007/BF01215378

Sims, C. A. (2011). Stepping on a rake: The role of fiscal policy in the inflation of the 1970s. *European Economic Review, 55*(1), 48–56. https://doi.org/10.1016/j.euroecorev.2010.11.010

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy, 39*, 195–214. https://doi.org/10.1016/0167-2231(93)90009-L

Woodford, M. (2001). Fiscal requirements for price stability. *Journal of Money, Credit and Banking, 33*(3), 669–728. https://doi.org/10.2307/2673890