Communicating Policy under Uncertainty

Thorsten Drautzburg, Igor Livshits, and Makoto Nakajima*

PRELIMINARY AND INCOMPLETE

Abstract

Central banks rely on forward guidance and yet may deviate from prior communications in the face of big shocks. This paper develops a model of central bank communication that allows for this behavior as an equilibrium outcome: Policymakers can announce targets for future policy (or macroeconomic) variables with an associated utility cost of deviating from these announcements, which we interpret as the degree of commitment. The decision to depart from previously communicated targets is endogenous, and the public understands this at announcement. When the central bank chooses the degree of commitment ex ante and the cost is quadratic, we provide an analytical solution for the model with either unconditional (interest-rate) promises or state-contingent (inflation-rate) promises in the textbook New Keynesian model. For a given degree of commitment the model captures that policymakers will break their promises in response to the shocks that hit the economy. We then show that policymakers choose a lower degree of commitment than the public would choose. Allowing for time-varying commitment, we show analytically in a two-period version of our model that policymakers optimally increase commitment to policy targets when cost-push shocks or their volatility are larger – thus exploiting commitment when managing expectations is most valuable. In contrast, larger demand shocks reduce optimal commitment as flexibility of policy targets becomes more valuable.

*Federal Reserve Bank of Philadelphia. We are grateful for comments at the 2025 Day Ahead conference of the Central Bank of Chile and the 2026 North American Econometric Society Winter meetings. The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve System, or its Board of Governors.

1 Introduction

Central Bank communication has been evolving. In the U.S., it was initially up to markets to learn the FOMC decisions, until the FOMC started to issue regular announcements after each meeting in 2000 (Lindsey, 2003). But what should the Fed communicate? For the first few years, the Fed initially only communicated risks to the outlook, before introducing forward guidance about policy (Lunsford, 2020). Forward guidance can have Delphic and Odyssean effects (Campbell et al., 2017). Lunsford (2020) argues that what the Fed communicated affected how the roles of Delphic vs Odyssean forward guidance, and thus how the economy reacted to expected interest rate changes.

We propose a parsimonious novel model of central bank communication. Central bank communication is a means to help implement policy by influencing expectations held by the private sector. Specifically, policymakers can communicate targets for future interest rates or future economic conditions – along with a degree of confidence that it will be able to implement the targets. If the actual outcome is different from the target communicated by the central bank, the central bank suffers a reputational (utility) loss. Markets understand that the central bank is imperfectly committed to its communications, and whether or how much realized outcomes deviate from the central bank's promises depends on economic conditions.

Policymakers may or may not want to communicate promises about the future (targets) in our framework. Policymakers care about deviations of inflation from target and the output gap, which reflects social welfare. In addition, policymakers want to avoid outcomes that differ from their past communications. Communicating promises is attractive because it allows policymakers to influence private sector expectations, which in turn may help to stabilize the economy. However, it comes at a reputational cost that policymakers want to avoid.

A natural framework to embed our model of central bank communications is the linearized workhorse New Keynesian model. Its forward looking nature potentially allows policymakers to manipulate private sector expectations in order to improve current outcomes. Even without a lower bound on nominal interest rates, cost-push shocks create policy trade-offs – and time-consistency issues: In response to a positive cost push shock, a policymaker would like private agents to lower their inflation expectations for next period below the inflation target in order to stabilize inflation and output today. Next period, however, the policymaker prefers on-target inflation. Communicating future promises are a form of forward guidance and a tool to influence private sector expectations. The notion that policymakers announce unconditional targets despite being aware that they may

want to renege on these promises under some conditions captures the intuitive notion that, for example, prior promises mean little in a recession.

We generalize the textbook treatment of optimal monetary policy (e.g. Galí, 2015, ch. 5) borrowing the ideas in partial default (Arellano et al., 2023): Policymakers can promise targets and then "default" on their promises, at some reputational cost. If the reputational cost is zero, optimal policy coincides with full discretion. With infinite reputational cost, policymakers fully commit to implementing the announced targets.

In our benchmark model, policymakers announce targets for future inflation and set interest rates accordingly. The private sector has rational expectations and anticipates that past targets influence interest rate setting and thereby inflation and the output gap. Given a quadratic cost of deviating from the target and given a reputational cost parameter, our model has an analytic solution. We prove that a stable solution exists and provide an easily verifiable condition for uniqueness.

We show that the public generally prefers a higher reputational cost and thus a higher degree of commitment than policymakers themselves. Intuitively, the public only bears the benefit of managing expectations via promises, but not the reputational cost. The desired reputational penalty is often, but not always, interior. This allows for promises of future inflation to offset current cost-push shocks in order to stabilize current inflation. In equilibrium, policymakers choose promises of future inflation that are negatively correlated with the cost-push shocks.

What macroeconomic aggregate policymakers target matters in our framework. This contrasts with the textbook model, where the choice of target is a matter of mathematical convenience. Here, if policymakers make promises about future inflation or the future output gap, they are free to adjust the policy rate to absorb demand shocks: The implied output gap reacts one-to-one to changes in the natural rate. In contrast, if policymakers make promises about future interest rates (or the output level), their implied policy rate promises are state contingent. Interest rates as well as inflation and the output gap respond to demand shocks, that is, to changes in the natural rate.

For a given degree of macroeconomic uncertainty, if policymakers communicate a greater degree of confidence that realizations will be close to their targets, this corresponds to a higher degree of commitment – or to a period of lower volatility. However, if policymakers announce targets for interest rates, the source of uncertainty matters for the desired degree of commitment. Endogenous commitment reinforces the effects of increased uncertainty about demand shocks and partly offsets the effects of cost-push shocks.

This discussion abstracts from informational requirements and other practical considerations. Targeting inflation means targeting an endogenous object, and the implied offset

of demand shocks with the interest rates requires observing r_t^* – a challenging problem for every central bank. Targeting the output gap faces analogous informational challenges. Moreover, the use of forward guidance means that the public is more accustomed to central bank targets for the future policy rate than explicit targets for short-term outcomes such as inflation.

It is important to note that the familiar inflation bias, which underlies the basic analysis in the classic Barro and Gordon (1983) paper and important subsequent contributions including Athey et al. (2005) and Amador et al. (2006), is absent in our benchmark model. Of course, this is not the only distinction between our paper and the earlier work — our parsimonious framework allows us to consider a full range of commitment levels, rather than just a discrete choice between rules and discretion, which is the staple in this literature. As a result, applying our framework to environments with the central bank does have inflation bias may yield additional insights, which we will pursue in future work.

A recent literature has also considered intermediate cases between discretion and full commitment in monetary policy, but these models do not yield equilibria in which policymakers endogenously decide to renege on their past promises. In the work of Schaumburg and Tambalotti (2007) and Bodenstein et al. (2012), policymakers reset policy exogenously according to the realization of a Poisson process. Debortoli and Lakdawala (2016) similarly analyze an economy with a constant probability that policymakers may reoptimize. In the work of Kurozumi (2008) and Sunakawa (2015) and Batista et al. (2023), policymakers only make promises that they can sustain in every state of the world.

Our focus is on monetary policy with partial default. A similar partial commitment technology has been used to analyze fiscal policy. In the context of optimal taxation, Clymo et al. (2023) also analyze a model with partial default. They show that the government chooses promises to influence private agents and its future self. Barthelemy and Mengus (2024) considers the importance of partial commitment of governments to rule out multiple equilibria. They also consider taxation in their application. Neither paper considers the optimal degree of commitment.

Our formulation above does not nest the commitment solution in Galí (2015), because our promises are not fully state contingent. Our analysis of inflation promises inherits the feature that demand shocks are fully stabilized, as in the case of state-contingent full commitment in Galí (2015). We conjecture that with full state contingency of promises there would be no downside to commitment in our environment, eliminating the tradeoff we are analyzing.

2 Environment

The economy is inhabited by a unit mass of representative households and firms, who form rational expectations about the future. They have access to a risk-free bond, whose interest rate i_t is set by the monetary authority.

Consumption demand and market clearing give rise to the standard Euler (IS) equation that relates the current output gap x_t to the expected output gap as well as the short rate i_t minus expected inflation $\mathbb{E}_t[\pi_{t+1}]$ and the natural rate r_t^*.

$$x_t = \mathbb{E}_t[x_{t+1}] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t[\pi_{t+1}] - r_t^*\right). \tag{2.1}$$

All variables are in deviation from steady state. The output gap is the deviation of output from output in the frictionless economy.[1] r_t^* follows an exogenous AR(1) process with persistence ρ_r.

Labor supply, cost-minimization in production and price-setting give rise to a forward-looking Phillips Curve, relating current inflation to expected future inflation, the current output gap, and a cost-push shock u_t:

$$\pi_t = \beta \mathbb{E}_t[\pi_{t+1}] + \kappa x_t + u_t, \tag{2.2}$$

where u_t follows an exogenous AR(1) process with persistence ρ_u.

To a second-order approximation, period welfare is given by:

$$W_t = -\pi_t^2 - \theta x_t^2. \tag{2.3}$$

The central bank controls the nominal rate i_t. In addition, it can choose to communicate a target for future outcomes of macroeconomic aggregates or the interest rate. For now, we assume that it can announce a target for tomorrow's inflation rate $\hat{\pi}_t$ and inherits a past target $\hat{\pi}_{t-1}$. It thus chooses inflation, the output gap, the interest rate, and targeted future inflation to solve :

$$L_t(\hat{\pi}_{t-1}, u_t) = \min_{\pi_t, x_t, i_t, \hat{\pi}_t} \pi_t^2 + \theta x_t^2 + d(\pi_t - \hat{\pi}_{t-1})^2 + \delta \mathbb{E}_t[L_{t+1}(\hat{\pi}_t, u_{t+1})] \tag{2.4}$$

subject to (2.1) and (2.2). We thus allow the central banker to discount the future at a rate different from that of other agents, but we will mostly set $\delta = \beta$ in what follows. Notably, we will later explicitly consider the choice of the degree of commitment d.

[1] Throughout, we assume that subsidies are in place so that the steady state is efficient.

We will additionally analyze the environment where the forward guidance concerns the future policy interest rate, rather than inflation. That is, the central banker inherits a past target $\hat{i}_{t-1}$ and announces a target for tomorrow's interest rate $\hat{i}_t$ to solve

$$L_t(\hat{i}_{t-1}, u_t) = \min_{\pi_t, x_t, i_t, \hat{\pi}_t} \pi_t^2 + \theta x_t^2 + d(i_t - \hat{i}_{t-1})^2 + \delta \mathbb{E}_t[L_{t+1}(\hat{i}_t, u_{t+1})] \tag{2.5}$$

subject to (2.1) and (2.2).

It is worth pointing out that the objective of the central banker differs from that of the social planner, since the social welfare function W does not take into account the psychic costs borne by the central bank of deviating from its promises. This, of course, leads to a disagreement between the social planner and the central banker over the desirability (or the optimal degree) of commitment to forward guidance, with the central banker always being more averse to commitment.

3 Illustrative 2-period model

In order to establish some of our key insights analytically, we turn to a 2-period version of the environment described above, with forward guidance concerning future policy (interest rates) $\hat{i}$ rather than inflation.

To start with, assume that there are no policy commitments in the first period, that is, $d_1 = 0$, and that the expectations in the second period are set exogenously as $E(\pi_3) = E(x_3) = 0$. Since the case of promised inflation is quite straightforward, we focus largely on the case where forward guidance regards the nominal interest rate. In the first period, the central bank chooses, along with the current policy i_1, the forward guidance — promised level of future interest rate $\hat{i}_2$. In the analysis below, we will take the level of commitment d as given, but will explicitly analyze both the welfare and the value of the central banker's objective for different levels of d. We proceed under the assumption of $\sigma = 1$ in this section.

Solving the 2-period model backwards is conceptually simple. For any given level of commitment d and the forward guidance $\hat{i}$, the policy that maximizes the objective of the second-period central banker is $i_2^* = \frac{d\hat{i} + (\theta+\kappa^2) r_2 + \kappa u_2}{\theta + d + \kappa^2}$, which yields $x_2^* = \frac{d(r_2 - \hat{i}) - \kappa u_2}{\theta + d + \kappa^2}$ and $\pi_2^* = \frac{\kappa d(r_2 - \hat{i}) + (\theta + d) u_2}{\theta + d + \kappa^2}$. These expressions can then be used in formulating the central banker's first period problem, which includes the choice of $\hat{i}$.

One expression that will be useful in generating key insights is

$$\left.\frac{\partial L_1}{\partial d}\right|_{d=0} = \delta \left(\sigma_r^2 + \frac{\kappa^2 \sigma_u^2}{(\theta + \kappa^2)^2} \right) - \frac{\kappa^2 \beta^2 \theta^2}{\delta (\theta + \kappa^2)^4} u_1^2.$$

First observation from this expression is that an impatient central bank (one with low δ always prefers having some commitment (since the first term in above expression vanishes and the second one becomes arbitrarily large as δ approaches 0). This result is, of course, intuitive, as commitment gives myopic central bank an additional tool and allows it to achieve $\pi_1 = x_1 = 0$.

Theorem 1. For δ small enough, the central banker's ex-ante preferred level of commitment d is strictly positive. Interim optimal d is also strictly positive for sufficiently large realizations of $|u|$.

The second observation is not as obvious, but is also intuitive: The value of commitment to period-2 policy guidance is increasing in the size of the period-1 cost push shock. One implication of this observation is that, if the degree of commitment is chosen in period 1, after the shocks are realized, the commitment level is increasing in the magnitude of the first period cost-push shock realization.

Theorem 2. The central banker's preferred level of commitment d is increasing in the magnitude $|u|$ of the current cost push shock.

The third observation is much more surprising. Consider the value of commitment to a patient central bank from the ex-ante perspective, i.e., before shock u_1 is realized. If $\delta = \beta$ and the distribution of shocks is i.i.d across time, then commitment is not valued by the central bank — the policy maker prefers full flexibility. Two ways to break this result are making the central bank more impatient and/or make the variance of cost-push shocks lower in the second period than in the first.

Theorem 3. If $\delta = \beta$ and the variance of u is constant over time, then the central banker's ex-ante preferred level of commitment is $d = 0$.

3.1 Initial promises as r^* shocks

Consider the problem of a non-zero initial promise with $d_1 > 0$. Using the period 2 solution to simplify the constraints, the loss functions in period 1 is

$$L_1(\hat{i}_0, u_1) = \min_{\pi_1, x_1, i_1, \hat{i}_1} \pi_i^2 + \theta x_1^2 + d_1(i_1 - \hat{i}_0)^2 + \delta \mathbb{E}[L_2(\hat{i}_1, u_2)] \tag{3.1}$$

$$\text{s.t. } x_1 = -\frac{d_2 \hat{i}_1}{\theta + d_2 + \kappa^2} - \left(i_1 + \frac{\kappa d_2 \hat{i}_1}{\theta + d_2 + \kappa^2} - r_1^* \right) \tag{3.2}$$

$$\pi_1 = \beta \frac{-\kappa d_2 \hat{i}_1}{\theta + d_2 + \kappa^2} + \kappa x_1 + u_1 \tag{3.3}$$

Now define $j_1 \equiv i_1 - \hat{i}_0$. Then the IS-equation becomes:

$$x_1 = -\frac{d_2 \hat{i}_1}{\theta + d_2 + \kappa^2} - \left(j_1 + \frac{\kappa d_2 \hat{i}_1}{\theta + d_2 + \kappa^2} - (r_1^* - \hat{i}_0) \right)$$

Since the IS equation is the only place where i_1 enters, the problem is isomorphic to choosing j_1 subject to a zero initial interest rate promise and a natural rate that is shifted by $-\hat{i}_0$.

The effect of past promises has therefore the opposite sign as the effect of shocks to the natural rate. Since we have shown that promised interest rates increase following a positive shock to r^*, it follows that higher initial promises lead to lower promises in the next period.

This carries over to the infinite horizon model studied below. There, we show that interest rate promises are stable, but exhibit negative persistence.

4 Equilibrium with forward guidance about interest rates

Parameterization. We choose standard values for the parameters. The discount factor $\beta = 0.99$. The weight on the output gap in the loss function is $\theta = \frac{1}{2}$. The slope of the Phillips curve is $\kappa = 0.3$. We assume log utility ($\sigma = 1$). Last, both shocks have persistence $\rho_u = \rho_r = 0.9$.

5 Equilibrium with forward guidance about inflation

We now turn back to the infinite horizon model to illustrate robustness of the key findings and derive impulse response functions.

5.1 Solution

An equilibrium in this economy is a set of policy functions for the policymaker such that (2.1) and (2.2) hold and agents have rational expectations. A linear policy rule for the announcement of (promised) future inflation is then

$$\hat{\pi}_t = \hat{\psi}_\pi \hat{\pi}_{t-1} + \hat{\psi}_u u_t,$$

and similarly for inflation with loadings $\psi_\circ$, the output gap with loadings $\xi_\circ$, and interest rates with loadings $\iota_\circ$.

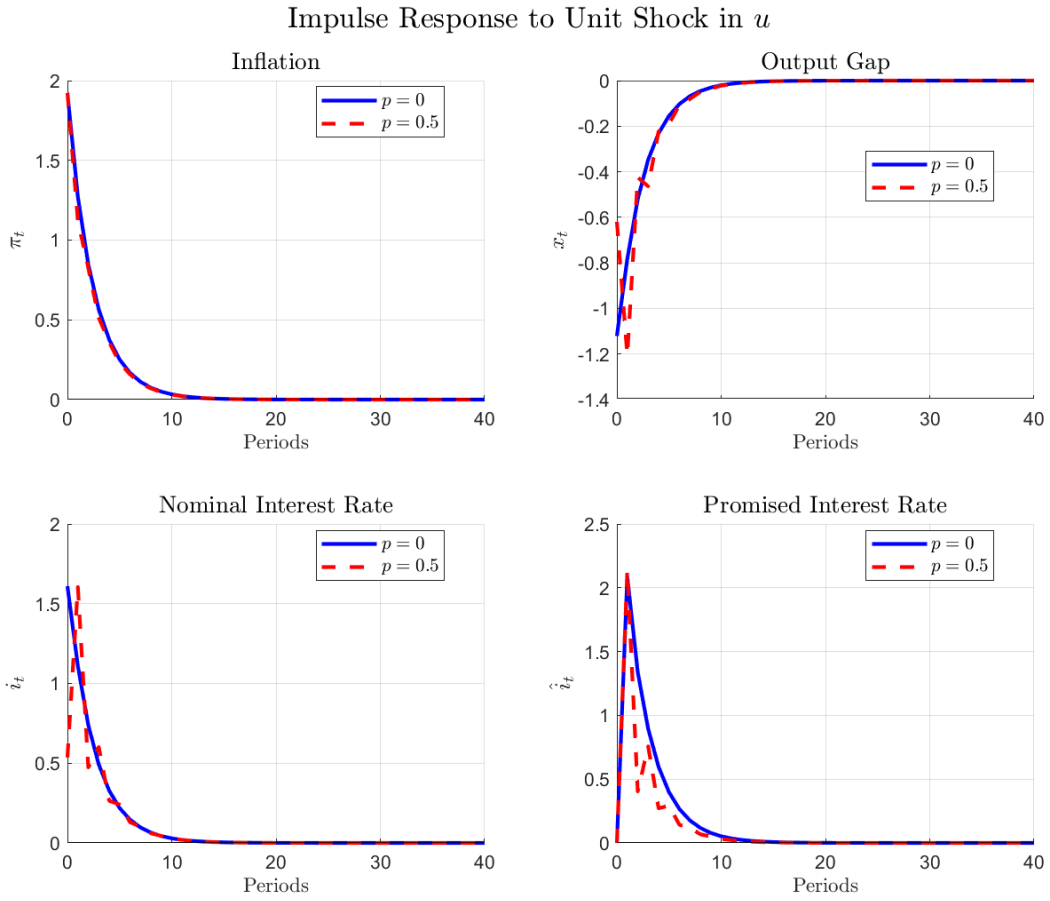


Figure 4.1: IRFs to a cost-push shock with full discretion, partial commitment, and textbook fully contingent commitment

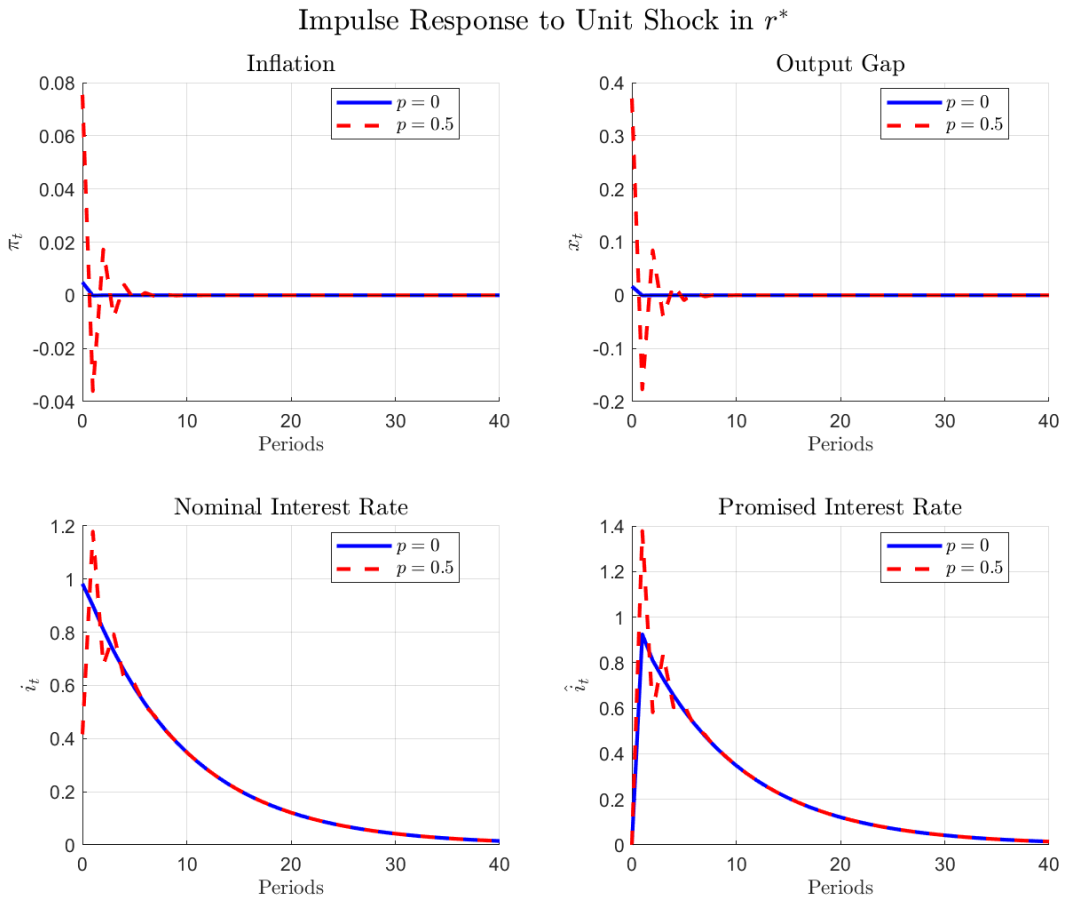


Figure 4.2: IRFs to a natural rate shock with full discretion, partial commitment, and textbook fully contingent commitment

To solve for the equilibrium policy functions, we compute first-order-conditions and the envelope condition, and then guess-and-verify that the policy functions are linear in $\hat{\pi}, u_t$ and the loss function L_t is a quadratic function of the same arguments. For example, the The equations characterizing the loadings on the shocks are linear, whereas the loadings on the endogenous state variable $\hat{\pi}_t$ are non-linear functions. However, solving for the equilibrium reduces to solving a third order polynomial, and we prove the following proposition in Appendix A:

Lemma 1. There exists a stable backward looking solution for inflation promises: $\hat{\psi}_\pi \in (0,1)$. If, in addition, the discriminant of the polynomial f in (A.13) is negative, the other candidate solutions are complex numbers and the real solution is unique.

The equilibrium thus has a VAR(1) representation:

$$\begin{bmatrix} u_t \\ \hat{\pi}_t \\ \pi_t \\ \xi_t \end{bmatrix} = \underbrace{\begin{bmatrix} \rho_u & 0 \\ \hat{\psi}_u \rho_u & \hat{\psi}_\pi \\ \psi_u \rho_u & \psi_\pi \\ \xi_u \rho_u & \hat{\xi}_u \end{bmatrix}}_{\equiv A} \begin{bmatrix} u_{t-1} \\ \hat{\pi}_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} \omega_u \\ \hat{\psi}_u \\ \psi_u \\ \xi_u \end{bmatrix}}_{\equiv B} \epsilon_t. \tag{5.1}$$

Note the sparsity pattern here – there is only one endogenous state variable. In addition, we can compute expected inflation as $[0,0,1,0]A \begin{bmatrix} u_{t-1} \\ \hat{\pi}_{t-1} \end{bmatrix}$.

Given the linear policy functions, it is straightforward to compute the value function. It is a quadratic function of $\begin{bmatrix} u_{t-1} \\ \hat{\pi}_{t-1} \end{bmatrix}$ plus a constant that is proportional to the shock variance. The derivation is are given in Appendix A.

$$\begin{aligned} L(\hat{\pi}, u) &= \frac{\ell_{uu}\sigma_u^2}{1-\beta} + \begin{bmatrix} u & \hat{\pi} \end{bmatrix} \begin{bmatrix} \ell_{uu} & \ell_{\pi u} \\ \ell_{\pi u} & \ell_{\pi\pi} \end{bmatrix} \begin{bmatrix} u \\ \hat{\pi} \end{bmatrix} \qquad (5.2) \\ \ell_{\pi\pi} &= \frac{\psi_\pi^2 + \tilde{\theta}\psi_\pi^2(1-\beta\hat{\psi}_\pi)^2 + d(\psi_\pi - 1)^2}{1-\beta\hat{\psi}_\pi^2} \\ \ell_{\pi u} &= \frac{\psi_\pi\psi_u + \tilde{\theta}\psi_\pi(1-\beta\hat{\psi}_\pi)(\psi_u(1-\beta\rho_u) - \beta\psi_\pi\hat{\psi}_u - 1) + d(\psi_\pi - 1)\psi_u + \beta\ell_{\pi\pi}\hat{\psi}_\pi\hat{\psi}_u}{1-\beta\rho\hat{\psi}_\pi} \\ \ell_{uu} &= \frac{\psi_u^2 + \tilde{\theta}(\psi_u(1-\beta\rho_u) - \beta\psi_\pi\hat{\psi}_u - 1)^2 + d\psi_u^2 + 2\beta\ell_{\pi u}\rho\hat{\psi}_u + \beta\ell_{\pi\pi}\hat{\psi}_u^2}{1-\beta\rho^2} \end{aligned}$$

It is useful to contrast the central bank's loss function – which includes the reputational cost – with the societal welfare cost L^S:

$$L^S(\hat{\pi}, u) = \pi^2 + \tilde{\theta}(\pi - \beta E[\pi'] - u)^2 + \beta \mathbb{E}[L^S(\hat{\pi}', u')] \tag{5.3}$$

The expression is analogous to the central bank's loss function, except that the direct effect of reputation is shut down via $d = 0$.

5.2 Calibration

We follow Galí (2015, ch. 3) in calibrating the deep parameters of the model, as summarized in Table 5.1. The period is one quarter. For the cost-push AR(1) process, we target a unit variance of the output gap in the $d \approx 0$-economy and choose persistent processes with $\rho_u \in \{0.75, 0.95\}$.

Parameter	Value
Discount factor β	0.99
Price persistence d	2/3
Price elasticity ϵ	6
Capital share α	$\frac{1}{3}$
IES σ	1
Frisch elasticity of labor supply ϕ	1
Price stickiness ϕ	$\frac{2}{3}$
Shock persistence ρ_u	$\in \{0.75, 0.95\}$
Shock variance Var$[\epsilon_t]$	Unit variance of output gap
Penalty parameter d	$\in \{0.0001, 3, 500\}$

Table 5.1: Calibration. The time period is one quarter.

We consider three different values for the penalty parameter: near-zero value of $d = 0.0001$, a penalty parameter that is near-optimal ($d = 1.5$), and a substantial degree of commitment ($d = 5$).

5.3 Impulse response-functions

Figure 5.1 shows the impulse-response functions for the case of $\rho_u = 0.75$. Panel (a) shows the impulse: A one-standard deviation increase in the cost push shock. When policymakers are free to renege on their past communication ($d \approx 0$), inflation rises (panel (b)) and the output gap falls (panel (e)). Promised future inflation also increases slightly, but plays little role in guiding future outcomes (panel (d)). Implied interest rates rise.

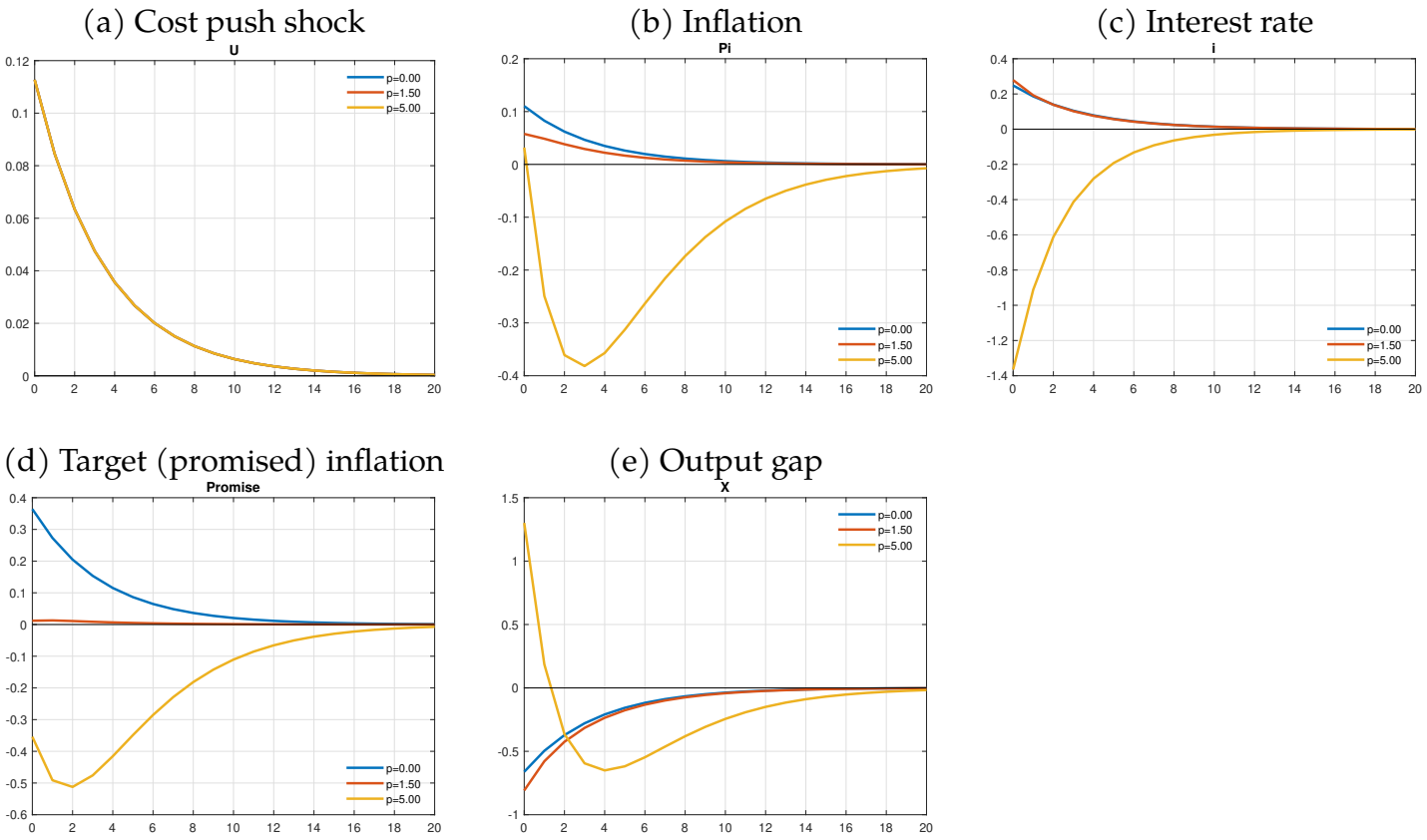


Figure 5.1: Simulated paths of the cost-push shock, inflation, target (promised) inflation, and the output gap

When policymakers have a small reputational cost of reneging on past promises, they implement a policy that keeps inflation close to zero, the initial promise. For the future, however, they promise inflation slightly below the path for actual inflation. The output gap falls more initially. With a high degree of commitment, the qualitative behavior changes. The commitment to the initial and subsequent promises means that inflation is near-zero on impact. The policy-makers promises and deliver on disinflation. This translates into an initial output boom – despite the cost-push shock. After three quarters, the boom turns into a bust that lasts almost 20 quarters. Implied interest rates fall when there isa significant degree of commitment.

5.4 Choice of penalty parameter

Impulse-responses are linear in this model. However, the value function is quadratic and thus state-dependent. To better understand the economic forces at play and the resultant choice of penalty parameter, it is thus useful to simulate the economy to see how the two state-variables evolve over time and with which choice of d they would be associated.

Figure 5.2 shows the results of simulating the economy for 100 quarters starting at the steady state. Here, we focus on $\rho_u = 0.75$ and the top left panel (a) shows the persistent cost-push shock. Panel (b) shows inflation – for the three different levels of d. Visually, inflation in the $d \approx 0$-economy is negatively correlated with inflation in the $d = 5$ economy. For $d = 1.5$, inflation is a somewhat smoothed version of inflation in the $d \approx 0$-economy.

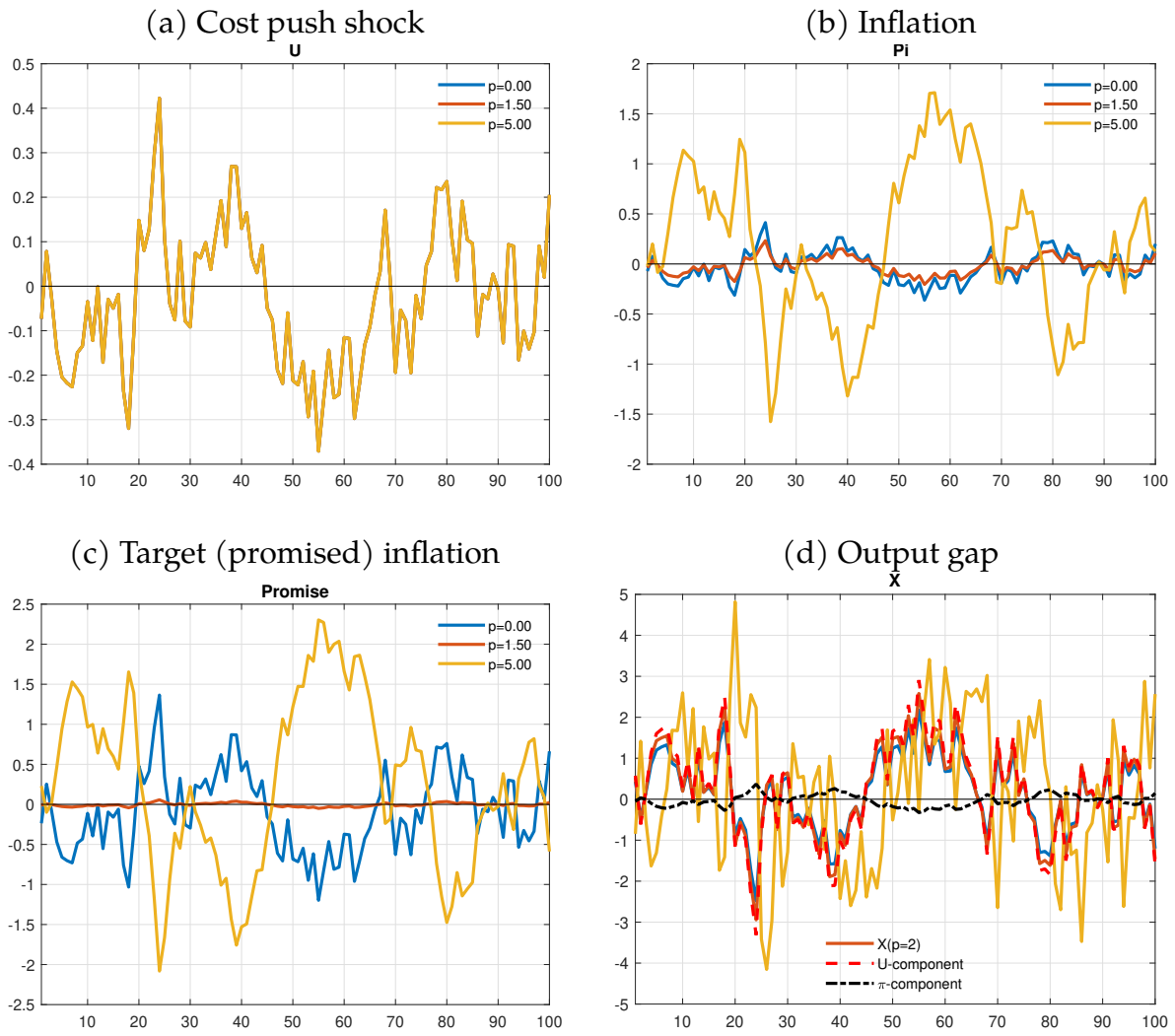


Figure 5.2: Simulated paths of the cost-push shock, inflation, target (promised) inflation, and the output gap: Moderate persistence $\rho_u = 0.75$

The benefit of communicating inflation targets is to influence the output gap. Panel (d) in Figure 5.2 shows the output gap for the three levels of d, as well as a decomposition of the output gap for the intermediate case of $d = 1.5$. Compared to near-zero commitment, the intermediate case features an output gap that – as in the IRFs – is slightly more volatile. However, the expected inflation component (shown as the black dash-dotted line) is countercyclical and stabilizes inflation slightly compared to the effect of cost-push shocks alone

(red dashed line). With more commitment (yellow solid line), the cyclicality of the output gap changes as the output gap becomes positively correlated with the cost push shock.

Figure 5.3 (a) shows how the components of the loss function vary as a function of d (simulating from a model economy with the policymaker's optimal choice). Both losses and d are in log scale. The solid (blue) line shows the variance of inflation. It is minimized near $d = 1.25$. The dashed line (in orange) shows the variance of the output gap, multiplied by the relative welfare weight $\tilde{\theta}$. It is minimized at a slightly higher value of d. The third component, the dashed-dotted line (in yellow), shows the penalty term $(\pi_t - \hat{\pi}_{t-1})^2 \times d$. It has a local minimum around $d = 1$, but its global minima are near the extremes – it disappears both with $d = 0$ and $d \to \infty$ (when $\pi_t = \hat{\pi}_{t-1}$). Policymakers minimize the sum of all three terms. IF the policymaker could choose, the optimal choice would be 1.25 on the grid considered here. The public does not care about the reputational cost and would choose a somewhat higher level of d, in between the values that minimizes the inflation and output variances.

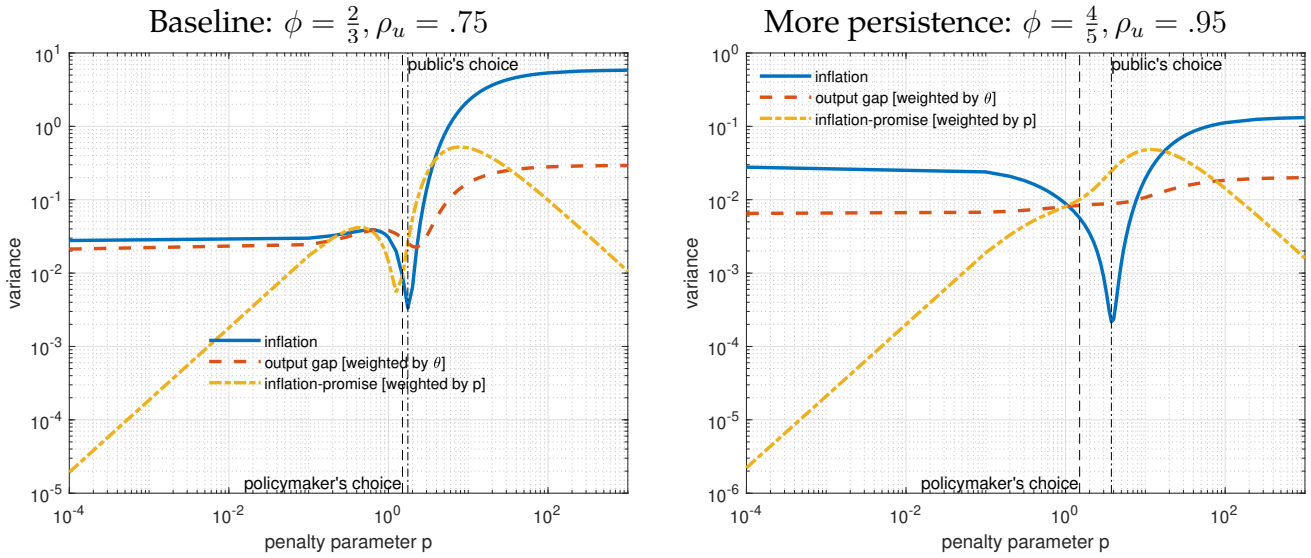


Figure 5.3: Components of the unconditional loss function as a function of d: Variances of inflation, the output gap, and the reputational cost.

For comparison, Figure 5.3 (b) shows how the components of the loss function vary as a function of d (simulating from a model economy with the policymaker's optimal choice) with higher price stickiness and higher shock persistence. In this case, the reputational cost no longer has a local interior minimum, nor does the output gap. The gap between the policymakers optimal choice and that of the public becomes larger in this case.

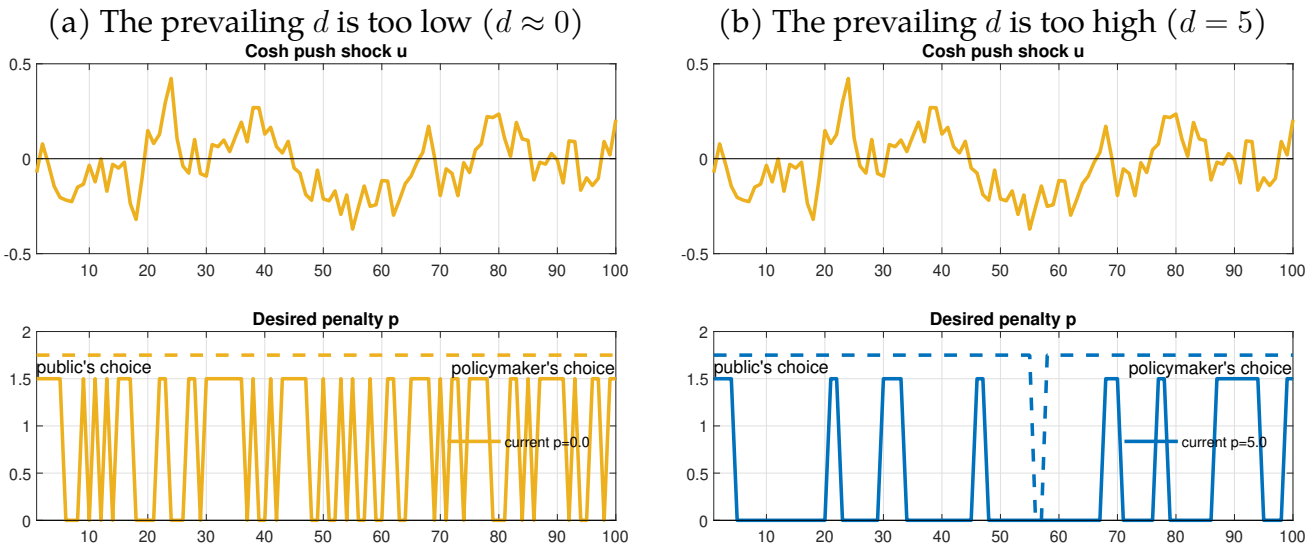


Figure 5.4: Desired d if policymakers or the public could choose a new d once and for all: Simulated economy as in Figure 5.2.

Note that when policymakers – or the public – were to be able to make a once-and-for-all change in d, their choices would depend on the prevailing level of d as well as the state of the economy. Figure 5.4 (a) and (b) feature the same shock sequence, but inflation promises under near-discreation ($d \approx 0$) and under a high $d = 5$. In both cases, the policymakers switches between wanting a permanent switch to $d = 1.5$ and a permanent switch to discretion ($d = 0$). This can be seen in the lower panels of the figure, which shows the desired d for the policymaker and the public. Again, the public prefers a higher d because it does not internalize the reputational cost. In Figure 5.4 (b), it is clear that the switches in the desired d coincide with periods of big swings in the cost push shock. As Figure 5.2 shows, these are also periods when the policymaker makes extreme policy promises. Note that when the prevailing d is already optimal at 1.5, neither the policymaker nor the public want to deviate (not shown).

5.5 Macroeconomic uncertainty

For a given degree of macroeconomic uncertainty, if policymakers communicate a greater degree of confidence that realizations will be close to their targets, this corresponds to a period of lower volatility.

In contrast, the effect of changes in the magnitude of the underlying uncertainty has ambiguous effects. In our benchmark model, in which the policymaker only communicates inflation targets, the policy rate absorbs shocks to r^* (demand shocks) and the objective function depends only on the volatility of cost push shocks. However, because the

variance of cost push shocks simply scales all responses, it does not alter the tradeoffs policymakers face. Higher volatility of cost-push shocks thus translates to higher variance of macroeconomic aggregates. Higher volatility of demand shocks simply makes the interest rate more volatile, but not the inflation rate or the output gap.

6 Conclusion

We provide a tractable framework to analyze forward guidance as "defaultable commitment". Unlike the existing literature, the extent of commitment to future policy is also a choice.

According to the analytical characterization of the 2-period model, commitment is used more by a more impatient central bank. Commitment to future interest rate is also more beneficial when the current realization of cost push shock is relatively large ($\rightarrow$ optimal time-varying forward guidance)

In ongoing work, we characterize the optimal period-by-period choice of the degree of commitment in the infinite horizon model.

References

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

Arellano, C., X. Mateos-Planas, and J.-V. Ríos-Rull (2023): "Partial Default," Journal of Political Economy, 131, 1385–1439.

Athey, S., A. Atkeson, and P. J. Kehoe (2005): "The optimal degree of discretion in monetary policy," Econometrica, 73, 1431–1475.

Barro, R. J. and D. B. Gordon (1983): "Rules, discretion and reputation in a model of monetary policy," Journal of monetary economics, 12, 101–121.

Barthelemy, J. and E. Mengus (2024): "Time-consistent implementation in macroeconomic games," Theoretical Economics, 19, 1119–1150.

Batista, Q., T. Nakata, and T. Sunakawa (2023): "Credible Forward Guidance," Journal of Economic Dynamics and Control, 153, 104699.

Bodenstein, M., J. Hebden, and R. Nunes (2012): "Imperfect credibility and the zero lower bound," Journal of Monetary Economics, 59, 135–149.

Campbell, J. R., J. D. M. Fisher, A. Justiniano, and L. Melosi (2017): "Forward Guidance and Macroeconomic Outcomes since the Financial Crisis," NBER Macroeconomics Annual, 31, 283–357.

Clymo, A., A. Lanteri, and A. Villa (2023): "Capital and Labor Taxes with Costly State Contingency," Review of Economic Dynamics, 51, 943–964.

Debortoli, D. and A. Lakdawala (2016): "How Credible Is the Federal Reserve? A Structural Estimation of Policy Re-optimizations," American Economic Journal:Macroeconomics, 8, 42–76.

Galí, J. (2015): Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications Second edition, Princeton University Press.

Kurozumi, T. (2008): "Optimal sustainable monetary policy," Journal of Monetary Economics, 55, 1277–1289.

Lindsey, D. E. (2003): "A Modern History of FOMC Communication: 1975-2002," Federal Reserve Board of Governors. https://www.federalreserve.gov/monetarypolicy/files/FOMC20030624memo01.pdf.

Lunsford, K. G. (2020): "Policy Language and Information Effects in the Early Days of Federal Reserve Forward Guidance," American Economic Review, 110, 2899–2934.

Schaumburg, E. and A. Tambalotti (2007): "An investigation of the gains from commitment in monetary policy," Journal of Monetary Economics, 54, 302–324.

Sunakawa, T. (2015): "A quantitative analysis of optimal sustainable monetary policies," Journal of Economic Dynamics and Control, 52, 119–135.

A Partial default: Derivations

$$L(\hat{\pi}, u) = \min_{\pi, \hat{\pi}'} \pi^2 + \tilde{\theta}(\pi - \beta\mathbb{E}[\pi'|\mathcal{F}] - u))^2 + d(\pi - \hat{\pi})^2 + \beta\mathbb{E}[L(\hat{\pi}', u')|\mathcal{F}] \tag{A.1}$$

$$[\pi] \quad 0 = 2\pi + 2\tilde{\theta}(\pi - \beta\mathbb{E}[\pi'|\mathcal{F}] - u) + 2d(\pi - \hat{\pi}) \tag{A.2}$$

$$\Rightarrow \quad \pi = \frac{\tilde{\theta}(\beta\mathbb{E}[\pi'|\mathcal{F}] + u) + d \times \hat{\pi}}{1 + \tilde{\theta} + d}$$

$$= \frac{\tilde{\theta}\beta\psi_\pi\hat{\pi}' + \tilde{\theta}(1 + \beta\psi_u\rho_u)u + \tilde{\theta}\beta\psi_r\rho_r r + d \times \hat{\pi}}{1 + \tilde{\theta} + d}, \tag{A.3}$$

where the last step uses the affine guess: $\pi = \psi_\pi\hat{\pi} + \psi_u u + \psi_r r$ and that $\mathbb{E}[u'|u] = \rho_u u$ and similarly for r'.

Asssuming also an affine guess for $\hat{\pi}' = \hat{\psi}_\pi\hat{\pi} + \hat{\psi}_u u + \hat{\psi}_r r$

$$\pi = \frac{\tilde{\theta}\beta\psi_\pi\hat{\psi}_\pi\hat{\pi} + \tilde{\theta}(1 + \beta\psi_u\rho_u + \beta\psi_\pi\hat{\psi}_u)u + \tilde{\theta}\beta(\psi_r\rho_r + \psi_\pi\hat{\psi}_r)r + d \times \hat{\pi}}{1 + \tilde{\theta} + d}$$

$$= \psi_\pi\hat{\pi} + \psi_u u \tag{A.4}$$

Matching coefficents:

$$\psi_\pi = \frac{\tilde{\theta}\beta\psi_\pi\hat{\psi}_\pi + d}{1 + \tilde{\theta} + d} = \frac{d}{1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi) + d} \tag{A.5}$$

$$\psi_u = \frac{\tilde{\theta}(1 + \beta\psi_u\rho + \beta\psi_\pi\hat{\psi}_u)}{1 + \tilde{\theta} + d} = \frac{\tilde{\theta}(1 + \beta\psi_\pi\hat{\psi}_u)}{1 + \tilde{\theta}(1 - \beta\rho) + d} \tag{A.6}$$

$$\psi_r = \frac{\tilde{\theta}\beta(\psi_r\rho_r + \psi_\pi\hat{\psi}_r)}{1 + \tilde{\theta} + d} = \frac{\tilde{\theta}\beta\psi_\pi\hat{\psi}_r}{1 + \tilde{\theta}(1 - \beta\rho_r) + d} \tag{A.7}$$

Note that we need $|\hat{\psi}_\pi| < 1$ to avoid explosive behavior.

We can solve for $\hat{\psi}_i, i \in \pi, u$ using the FOC for promised inflation and the corresponding envelope condition

$$[\hat{\pi}'] \quad 0 = -2\tilde{\theta}(\pi - \beta\mathbb{E}[\pi'|\mathcal{F}] - u)\beta\frac{\partial\mathbb{E}[\pi'|\mathcal{F}]}{\partial\hat{\pi}'} + \beta\mathbb{E}[L_{\hat{\pi}}(\hat{\pi}', u')]$$

$$= -2\tilde{\theta}(\pi - \beta\mathbb{E}[\pi'|\mathcal{F}] - u)\beta\psi_\pi + \beta\mathbb{E}[L_{\hat{\pi}}(\hat{\pi}', u')] \tag{A.8}$$

$$[EC\ \hat{\pi}] \quad L_{\hat{\pi}} = \frac{\partial\pi}{\partial\hat{\pi}}\left(2\pi + 2\tilde{\theta}(\pi - \beta\mathbb{E}[\pi'\mathcal{F}] - u) + 2d(\pi - \hat{\pi})\right)$$

$$+ \left(\frac{d\mathbb{E}[\pi'|\mathcal{F}]}{d\hat{\pi}'}\left(-2\beta\tilde{\theta}(\pi - \beta\mathbb{E}[\pi'|\mathcal{F}] - u)\right) + \beta\mathbb{E}[L_{\hat{\pi}}(\hat{\pi}', u')]\right)\frac{d\hat{\pi}'}{d\hat{\pi}} - 2d(\pi - \hat{\pi})$$

$$= -2d(\pi - \hat{\pi}) \tag{A.9}$$

The expectation of the one-period lead is therefore:

$$\mathbb{E}[L_{\hat{\pi}}(\hat{\pi}', u')] = -2d(\mathbb{E}[\pi'] - \hat{\pi}') \tag{A.10}$$

To simplify the FOC we are going to use (A.10) as well as the following facts:

$$\frac{d}{\psi_\pi} = 1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi) + d$$
$$\frac{d\mathbb{E}[\pi'|\mathcal{F}]/d\hat{\pi}}{\psi_\pi} = \frac{\tilde{\theta}\beta\hat{\psi}_\pi}{1 + \tilde{\theta} + d} + \frac{1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi) + d}{1 + \tilde{\theta} + d} = \frac{1 + \tilde{\theta} + d}{1 + \tilde{\theta} + d} = 1$$

Now substitute (A.10) into the FOC, divide by 2β, and re-arrange:

$$\begin{aligned}
& 2\tilde{\theta}(\pi - \beta\mathbb{E}[\pi'] - u)\beta\psi_\pi = \beta\mathbb{E}[L_{\hat{\pi}}(\hat{\pi}', u')] = -2\beta d(\mathbb{E}[\pi'] - \hat{\pi}') \\
\Leftrightarrow \quad & \tilde{\theta}\psi_\pi(\psi_\pi\hat{\pi} + \psi_u u + \psi_r r - \beta(\psi_\pi(\hat{\psi}_\pi\hat{\pi} + \hat{\psi}_u u + \hat{\psi}_r r) + \psi_r\rho_r r + \psi_u\rho_u u) - u) \\
& \quad = -d\Big((\psi_\pi - 1)(\hat{\psi}_\pi\hat{\pi} + \hat{\psi}_u u + \hat{\psi}_r r) + \psi_u\rho_u u + \psi_r\rho_r r)\Big)
\end{aligned} \tag{A.11}$$

Collecting the terms involving $\hat{\pi}$:

$$\begin{aligned}
& -d(\psi_\pi - 1)\hat{\psi}_\pi = \tilde{\theta}\psi_\pi^2(1 - \beta\hat{\psi}_\pi) \\
\Leftrightarrow \quad & d\hat{\psi}_\pi = \psi_\pi(\tilde{\theta}\psi_\pi(1 - \beta\hat{\psi}_\pi) + d\hat{\psi}_\pi)
\end{aligned} \tag{A.12}$$

Plug in from (A.5):

$$\begin{aligned}
& d\hat{\psi}_\pi = \frac{d}{1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi) + d}\left(\tilde{\theta}\frac{d}{1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi) + d}(1 - \beta\hat{\psi}_\pi) + d\hat{\psi}_\pi\right) \\
\Leftrightarrow \quad & \hat{\psi}_\pi(1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi) + d) - d\hat{\psi}_\pi = \tilde{\theta}\frac{d}{1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi) + d}(1 - \beta\hat{\psi}_\pi) \\
\Leftrightarrow \quad & \hat{\psi}_\pi(1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi))(1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi) + d) = \tilde{\theta}d(1 - \beta\hat{\psi}_\pi) \\
\Leftrightarrow \quad & \hat{\psi}_\pi(1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi))^2 + d\hat{\psi}_\pi(1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi)) = \tilde{\theta}d(1 - \beta\hat{\psi}_\pi) \\
\Leftrightarrow \quad & f(\hat{\psi}_\pi) \equiv \hat{\psi}_\pi(1 + \tilde{\theta}(1 - \beta\hat{\psi}_\pi))^2 + d\hat{\psi}_\pi + d(\hat{\psi}_\pi - 1)\tilde{\theta}(1 - \beta\hat{\psi}_\pi) = 0
\end{aligned} \tag{A.13}$$

Note that at $\hat{\psi}_\pi = 0$, the f is negative. At $\hat{\psi}_\pi = 1$, in contrast, f equals $(1+\tilde{\theta}(1-\beta))^2+d > 0$. Since f is a polynomial in $\hat{\psi}_\pi$ it is continuous and by continuity we know that there is a solution $\hat{\psi}\pi \in (0, 1)$:

Lemma 2. There exists a stable backward looking solution for inflation promises: $\hat{\psi}_\pi \in (0,1)$. If, in addition, the discriminant of the polynomial f in (A.13) is negative, the other candidate solutions are complex numbers and the real solution is unique.

Multiplying out yields equation (A.13) as an explicit polynomial:

$$f(\hat{\psi}_\pi) = \underbrace{\beta^2\tilde{\theta}}_{\equiv a}\hat{\psi}_\pi^3 + \underbrace{(-2(1+\tilde{\theta})\beta\tilde{\theta})}_{\equiv b}\hat{\psi}_\pi^2 + \underbrace{(-d\beta + d(1\tilde{\theta}) + \tilde{\theta}d\beta + (1+\tilde{\theta})^2}_{\equiv c}\hat{\psi}_\pi + \underbrace{(-d\tilde{\theta})}_{\equiv d}$$

The discriminant is given by:

$$D \equiv \frac{4(b^2-3ac)^3 - (2b^3 - 9abc + 27a^2d)^2}{27a^2} \propto 4(b^2-3ac)^3 - (2b^3 - 9abc + 27a^2d)^2. \quad \text{(A.14)}$$

Reaction to u shocks Going back to the FOC for promised inflation (A.11) and collecting the terms involving u:

$$\begin{aligned}
&\tilde{\theta}\psi_\pi(\psi_u - \beta(\psi_\pi\hat{\psi}_u + \psi_u\rho_u) - 1) = -d\left((\psi_\pi - 1)\hat{\psi}_u + \psi_u\rho_u\right) \\
\Leftrightarrow \quad &\tilde{\theta}\psi_\pi(\psi_u(1-\beta\rho_u) + \psi_\pi\hat{\psi}_u - 1) = -d\left((\psi_\pi - 1)\hat{\psi}_u + \psi_u\rho_u\right) \\
\Leftrightarrow \quad &\tilde{\theta}\psi_\pi(\psi_u(1-\beta\rho_u) - 1) + d\psi_u\rho_u = -d\left((\psi_\pi - 1) + \tilde{\theta}\psi_\pi^2\right)\hat{\psi}_u \\
\Leftrightarrow \quad &\psi_u\left(\tilde{\theta}\psi_\pi(1-\beta\rho_u) + d\rho_u\right) = \tilde{\theta}\psi_\pi - d\left((\psi_\pi - 1) + \tilde{\theta}\psi_\pi^2\right)\hat{\psi}_u \qquad \text{(A.15)}
\end{aligned}$$

Multiply through by $1 + \tilde{\theta}(1-\beta\rho_u) + d$ and substitute out for ψ_u from (A.6):

$$\begin{aligned}
&\tilde{\theta}(1+\beta\psi_\pi\hat{\psi}_u)\left(\tilde{\theta}\psi_\pi(1-\beta\rho_u) + d\rho_u\right) = \left(\tilde{\theta}\psi_\pi - d\left((\psi_\pi - 1) + \tilde{\theta}\psi_\pi^2\right)\hat{\psi}_u\right)(1 + \tilde{\theta}(1-\beta\rho_u) + d) \\
\Leftrightarrow \quad &\hat{\psi}_u\left(\beta\psi_\pi\left(\tilde{\theta}\psi_\pi(1-\beta\rho_u) + d\rho_u\right) + \left(d\left(\frac{1}{\tilde{\theta}}(\psi_\pi - 1) + \psi_\pi^2\right)\right)(1 + \tilde{\theta}(1-\beta\rho_u) + d)\right) \\
&\quad = -\left(\tilde{\theta}\psi_\pi(1-\beta\rho_u) + d\rho_u\right) + \psi_\pi(1 + \tilde{\theta}(1-\beta\rho_u) + d) \qquad \text{(A.16)}
\end{aligned}$$

If $\rho_u = 0$:

$$\hat{\psi}_u\left(\beta\psi_\pi\tilde{\theta}\psi_\pi + \left(d\left(\frac{1}{\tilde{\theta}}(\psi_\pi - 1) + \psi_\pi^2\right)\right)(1 + \tilde{\theta} + d)\right) = -\tilde{\theta}\psi_\pi + \psi_\pi(1 + \tilde{\theta} + d) \qquad \text{(A.17)}$$

This could have either sign.

Reaction to r^* shocks

$$\tilde{\theta}\psi_\pi(\psi_r - \beta(\psi_\pi\hat{\psi}_r + \psi_r\rho_r)) = -d\left((\psi_\pi - 1)\hat{\psi}_r + \psi_r\rho_r\right)$$

$$\Leftrightarrow \qquad \psi_r \left(\tilde{\theta}\psi_\pi(1-\beta\rho_r) + d\psi_r\rho_r\right) = \left(\beta\tilde{\theta}\psi_\pi - d\left((\psi_\pi - 1)\right)\right)\hat{\psi}_r \tag{A.18}$$

Multiply through by $1 + \tilde{\theta}(1-\beta\rho_r) + d$ and substitute out for ψ_r from (A.7):

$$\tilde{\theta}\beta\psi_\pi\hat{\psi}_r \left(\tilde{\theta}\psi_\pi(1-\beta\rho_r) + d\psi_r\rho_r\right) = \left(\beta\tilde{\theta}\psi_\pi - d\left((\psi_\pi - 1)\right)\right)(1 + \tilde{\theta}(1-\beta\rho_r) + d)\hat{\psi}_r \tag{A.19}$$

$\hat{\psi}_r = \psi_r = 0$ is a solution.

A.1 Implied output gap

$$\begin{aligned} x_t &\equiv \xi_\pi\hat{\pi}_t + \xi_u u_t \\ &= \frac{\pi_t - \beta\mathbb{E}_t[\pi_{t+1}] - u_t}{\kappa} \\ &= \frac{\psi_\pi\hat{\pi}_t + (\psi_u - 1)u_t - \beta(\psi_\pi(\hat{\psi}_\pi\hat{\pi}_t + \hat{\psi}_u u_t + \psi_u\rho u_t)}{\kappa} \\ \Rightarrow \xi_\pi &= \frac{\psi_\pi(1-\beta\hat{\psi}_\pi)}{\kappa} \qquad\qquad \xi_u = \frac{\psi_u - 1 - \beta(\psi_\pi\hat{\psi}_u + \psi_u\rho)}{\kappa} \end{aligned} \tag{A.20}$$

A.2 Value function

Guess and verify that the value function is linear-quadratic:

$$\begin{aligned} L(\hat{\pi}, u) &= \ell_0 + \ell_\pi\hat{\pi} + \ell_u u + \ell_{\pi\pi}\hat{\pi}^2 + 2\ell_{\pi u}\hat{\pi}u + \ell_{uu}u^2 \\ &= (\psi_\pi\hat{\pi} + \psi_u u)^2 + \tilde{\theta}(\psi_\pi\hat{\pi} + \psi_u u - \beta(\psi_\pi(\hat{\psi}_\pi\hat{\psi} + \hat{\psi}_u u) + \psi_u\mathbb{E}[u']) - u)^2 + d((\psi_\pi - 1)\hat{\psi} + \psi_u u)^2 \\ &\quad + \beta\mathbb{E}[\ell_0 + \ell_\pi(\hat{\psi}_p\hat{\pi} + \hat{\psi}_u u) + \ell_u u' + \ell_{\pi\pi}(\hat{\psi}_p\hat{\pi} + \hat{\psi}_u u)^2 + 2\ell_{\pi u}(\hat{\psi}_p\hat{\pi} + \hat{\psi}_u u)u' + \ell_{uu}(u')^2] \end{aligned} \tag{A.21}$$

Using that $\mathbb{E}[(u')^2|\mathcal{F}] = \text{Var}[u'|\mathcal{F}] + \mathbb{E}[u|\mathcal{F}]^2$ and matching coefficients we get

$$\ell_0 = \frac{\ell_{uu}\sigma_u^2}{1-\beta} \tag{A.22}$$

$$\ell_\pi = \beta\ell_\pi\hat{\psi}_\pi = 0 \tag{A.23}$$

$$\ell_u = \beta\ell_u\rho_u = 0 \tag{A.24}$$

$$\ell_{\pi\pi} = \frac{\psi_\pi^2 + \tilde{\theta}\psi_\pi^2(1-\beta\hat{\psi}_\pi)^2 + d(\psi_\pi - 1)^2}{1 - \beta\hat{\psi}_\pi^2} \tag{A.25}$$

$$\ell_{\pi u} = \frac{\psi_\pi \psi_u + \tilde{\theta}\psi_\pi(1-\beta\hat{\psi}_\pi)(\psi_u(1-\beta\rho_u) - \beta\psi_\pi\hat{\psi}_u - 1) + d(\psi_\pi - 1)\psi_u + \beta\ell_{\pi\pi}\hat{\psi}_\pi\hat{\psi}_u}{1-\beta\rho\hat{\psi}_\pi} \tag{A.26}$$

$$\ell_{uu} = \frac{\psi_u^2 + \tilde{\theta}(\psi_u(1-\beta\rho_u) - \beta\psi_\pi\hat{\psi}_u - 1)^2 + d\psi_u^2 + 2\beta\ell_{\pi u}\rho\hat{\psi}_u + \beta\ell_{\pi\pi}\hat{\psi}_u^2}{1-\beta\rho^2} \tag{A.27}$$

Note that higher d increases ℓ_{uu} and this increases the expected loss as a function of the conditional variance of the cost-push shocks. All else equal, higher uncertainty would thus lead to a lower ex-ante optimal level of commitment.

A positive weight on the output gap also leads to higher losses when the ex-ante uncertainty rises. To see this, consider the limit in which inflation does not react to the cost push shock ($\lim_{d\to\infty}\psi_u = 0$), so that coefficient on d in (A.27) becomes small. ℓ_{uu} remains strictly positive even when $\rho_u = 0$ as long as $\tilde{\theta} > 0$, because $\psi_\pi\hat{\psi}_u$ converge to a constant different from $-\beta^{-1}$.

We can contrast the central bank's loss function – which includes the reputational cost – with the societal welfare cost L^S: Guess and verify that the value function is linear-quadratic:

$$L^S(\hat{\pi}, u) = \pi^2 + \tilde{\theta}(\pi - \beta E[\pi'] - u)^2 + \beta\mathbb{E}[L^S(\hat{\pi}', u')] \tag{A.28}$$

The expression is analogous to the central bank's loss function, except that the direct effect of reputation is shut down via $d = 0$:

$$\ell_0^S = \frac{\ell_{uu}^S\sigma_u^2}{1-\beta} \tag{A.29}$$

$$\ell_\pi^S = \beta\ell_\pi^S\hat{\psi}_\pi = 0 \tag{A.30}$$

$$\ell_u^S = \beta\ell_u^S\rho_u = 0 \tag{A.31}$$

$$\ell_{\pi\pi}^S = \frac{\psi_\pi^2 + \tilde{\theta}\psi_\pi^2(1-\beta\hat{\psi}_\pi)^2 + d(\psi_\pi - 1)^2}{1-\beta\hat{\psi}_\pi^2} \tag{A.32}$$

$$\ell_{\pi u}^S = \frac{\psi_\pi\psi_u + \tilde{\theta}\psi_\pi(1-\beta\hat{\psi}_\pi)(\psi_u(1-\beta\rho_u) - \beta\psi_\pi\hat{\psi}_u - 1) + \beta\ell_{\pi\pi}^S\hat{\psi}_\pi\hat{\psi}_u}{1-\beta\rho\hat{\psi}_\pi} \tag{A.33}$$

$$\ell_{uu} = \frac{\psi_u^2 + \tilde{\theta}(\psi_u(1-\beta\rho_u) - \beta\psi_\pi\hat{\psi}_u - 1)^2 + \beta\ell_{\pi u}^S\rho\hat{\psi}_u}{1-\beta\rho^2} \tag{A.34}$$

A.3 Optimality of commitment

Proposition A.1. The ex-ante optimal loss $\ell_0(d)$ is strictly increasing in d at $d = 0$. That is, a marginal promise-keeping penalty cannot improve upon standard discretion when measured by the augmented loss L.

Proof. The ex-ante loss (evaluated at $\hat{\pi} = 0, u = 0$) is

$$\ell_0 = \frac{\beta\,\sigma_u^2}{1-\beta}\,\ell_{uu}, \tag{A.35}$$

so it suffices to sign $\frac{d\ell_{uu}}{dd}\big|_{d=0}$.

Values of the policy coefficients at $d = 0$. At $d = 0$ the promise variable is inert. From (A.5), $\psi_\pi = 0$; from (A.13), $\hat{\psi}_\pi = 0$; and from (A.6),

$$\psi_u\big|_{d=0} = \frac{\tilde{\theta}}{1+\tilde{\theta}(1-\beta\rho_u)}.$$

Since $\psi_\pi = 0$, the value-function coefficients satisfy $\ell_{\pi\pi} = 0$ and $\ell_{\pi u} = 0$.

Direct effect of d. Rewrite (A.27) as

$$\ell_{uu}(1-\beta\rho_u^2) = \underbrace{\psi_u^2 + \tilde{\theta}\big(\psi_u(1-\beta\rho_u) - \beta\psi_\pi\hat{\psi}_u - 1\big)^2}_{\text{welfare terms}} + \underbrace{d\,\psi_u^2}_{\text{direct penalty}} + \underbrace{\beta\ell_{\pi\pi}\hat{\psi}_u^2 + 2\beta\ell_{\pi u}\rho_u\hat{\psi}_u}_{\text{continuation}}. \tag{A.36}$$

Differentiating the direct penalty term and evaluating at $d = 0$ gives

$$\frac{\partial}{\partial d}\Big[d\,\psi_u^2\Big]\Big|_{d=0} = \psi_u^2\big|_{d=0} = \left(\frac{\tilde{\theta}}{1+\tilde{\theta}(1-\beta\rho_u)}\right)^2 > 0. \tag{A.37}$$

Indirect effects through policy functions are zero. Consider first the welfare terms in (A.36). Their derivative with respect to any indirect channel that operates through ψ_u takes the form

$$2\frac{d\psi_u}{dd}\Big[\psi_u + \tilde{\theta}\big(\psi_u(1-\beta\rho_u) - 1\big)(1-\beta\rho_u)\Big]\Big|_{d=0}.$$

The first-order condition for π at $d = 0$ implies $\psi_u + \tilde{\theta}(\psi_u(1-\beta\rho_u)-1) = 0$, so this expression reduces to

$$2\frac{d\psi_u}{dd}\bigg|_{d=0} \cdot \tilde{\theta}\big(\psi_u(1-\beta\rho_u)-1\big)\big((1-\beta\rho_u)-1\big) = -2\frac{d\psi_u}{dd}\bigg|_{d=0} \cdot \frac{\tilde{\theta}\,\beta\rho_u}{1+\tilde{\theta}(1-\beta\rho_u)},$$

which captures only how the existing discretionary trade-off responds to a shift in ψ_u—not any expectations-management benefit.

Crucially, the expectations-management channel operates entirely through the products $\psi_\pi\hat{\psi}_u$, $\ell_{\pi\pi}\hat{\psi}_u^2$, and $\ell_{\pi u}\hat{\psi}_u$. Because $\psi_\pi\big|_{d=0} = 0$ implies $\ell_{\pi\pi}\big|_{d=0} = 0$ and $\ell_{\pi u}\big|_{d=0} = 0$, every term through which promises could affect the loss vanishes at $d = 0$. In particular:

- The cross-derivative involving $\frac{d(\psi_\pi\hat{\psi}_u)}{dp}$ enters the welfare terms through $-\beta\psi_\pi\hat{\psi}_u$ inside the squared NKPC residual; since $\psi_\pi\big|_{d=0} = 0$, this product and its derivative contribute zero at $d = 0$.
- The continuation terms $\beta\ell_{\pi\pi}\hat{\psi}_u^2$ and $2\beta\ell_{\pi u}\rho_u\hat{\psi}_u$ are identically zero at $d = 0$ and their derivatives inherit the factor $\ell_{\pi\pi}\big|_{d=0} = 0$ or $\ell_{\pi u}\big|_{d=0} = 0$.

The only surviving indirect term is the shift in ψ_u operating through the welfare terms, computed above. Even this term does not reflect a commitment benefit: it merely captures the distortion that the penalty introduces in the contemporaneous inflation–output trade-off via the first-order condition for π.

Total derivative. Combining these observations:

$$\frac{d\ell_{uu}}{dd}\bigg|_{d=0} = \frac{1}{1-\beta\rho_u^2}\left[\psi_u^2\big|_{d=0} - 2\frac{d\psi_u}{dd}\bigg|_{d=0} \cdot \frac{\tilde{\theta}\,\beta\rho_u}{1+\tilde{\theta}(1-\beta\rho_u)}\right]. \tag{A.38}$$

The first summand is strictly positive by (A.37). One can verify that the second summand is of smaller magnitude for all admissible parameter values; in the special case $\rho_u = 0$ (i.i.d. cost-push shocks), it vanishes entirely, yielding

$$\frac{d\ell_{uu}}{dd}\bigg|_{d=0,\,\rho_u=0} = \frac{\tilde{\theta}^2}{(1+\tilde{\theta})^2} > 0. \tag{A.39}$$

Interpretation. At $d = 0$ the promise mechanism is completely dormant: the variable $\hat{\pi}$ appears in neither the constraints nor the effective objective, so future behavior is unchanged by a marginal increase in d. The commitment benefit—which requires promises to alter expectations—is therefore zero at the margin. The penalty $d(\pi - \hat{\pi})^2$, in contrast,

imposes a first-order cost proportional to the variance of inflation under discretion. Consequently, $\ell_0'(0) > 0$.

This does not preclude gains at larger values of d, where the commitment channel is fully active and can dominate the penalty cost; it only implies that the improvement, if any, is not local to the discretionary benchmark. □

A.4 Unconditional expected loss and its derivative at $d = 0$

We now evaluate the loss in the ergodic distribution of the state $(\hat{\pi}_t, u_t)$ and show that the result of the previous subsection extends: the unconditional expected loss is also strictly increasing in d at $d = 0$.

Ergodic moments of the state. The joint dynamics of the state are

$$\begin{pmatrix} \hat{\pi}_{t+1} \\ u_{t+1} \end{pmatrix} = \begin{pmatrix} \hat{\psi}_\pi & \hat{\psi}_u \\ 0 & \rho_u \end{pmatrix} \begin{pmatrix} \hat{\pi}_t \\ u_t \end{pmatrix} + \begin{pmatrix} 0 \\ \varepsilon_{t+1} \end{pmatrix}, \qquad \varepsilon_t \sim (0, \sigma_\varepsilon^2). \tag{A.40}$$

Since $|\hat{\psi}_\pi| < 1$ and $|\rho_u| < 1$, the ergodic distribution has zero mean and covariance matrix satisfying:

$$\mathrm{Var}(u) = \frac{\sigma_\varepsilon^2}{1 - \rho_u^2}, \tag{A.41}$$

$$\mathrm{Cov}(\hat{\pi}, u) = \frac{\hat{\psi}_u \rho_u}{1 - \hat{\psi}_\pi \rho_u} \mathrm{Var}(u), \tag{A.42}$$

$$\mathrm{Var}(\hat{\pi}) = \frac{\hat{\psi}_u^2}{1 - \hat{\psi}_\pi^2} \cdot \frac{1 + \hat{\psi}_\pi \rho_u}{1 - \hat{\psi}_\pi \rho_u} \mathrm{Var}(u). \tag{A.43}$$

Unconditional expected loss. Since $\ell_\pi = \ell_u = 0$, the unconditional expected loss is

$$\bar{L}(d) \;\equiv\; \mathbb{E}\big[L(\hat{\pi}_t, u_t)\big] = \ell_0 + \ell_{\pi\pi} \mathrm{Var}(\hat{\pi}) + 2\ell_{\pi u} \mathrm{Cov}(\hat{\pi}, u) + \ell_{uu} \mathrm{Var}(u). \tag{A.44}$$

Evaluation at $d = 0$. At $d = 0$ we have $\psi_\pi = \hat{\psi}_\pi = \ell_{\pi\pi} = \ell_{\pi u} = 0$, so

$$\bar{L}\big|_{d=0} = \frac{\beta\, \sigma_\varepsilon^2}{1 - \beta}\, \ell_{uu}\big|_{d=0} + \ell_{uu}\big|_{d=0} \mathrm{Var}(u) = \frac{\ell_{uu}\big|_{d=0}\, \sigma_\varepsilon^2}{(1-\beta)(1-\rho_u^2)} \big(1 - \beta\rho_u^2\big),$$

which is the standard discretionary loss.

Limiting value of $\hat{\psi}_u$ as $d \to 0^+$. Although $\hat{\psi}_u$ is indeterminate at $d = 0$ (promises are irrelevant), the first-order condition for $\hat{\pi}'$ pinpoints a well-defined limit $\hat{\psi}_u^* \equiv \lim_{d\to 0^+} \hat{\psi}_u$ by dividing the u-coefficient equation through by d and sending $d \to 0$:

$$\hat{\psi}_u^* = \frac{\tilde{\theta}\big(\rho_u(1+\tilde{\theta}) - 1\big)}{(1+\tilde{\theta})\big(1+\tilde{\theta}(1-\beta\rho_u)\big)}. \tag{A.45}$$

The limiting ergodic moments at $d = 0$ are therefore

$$\mathrm{Var}(\hat{\pi})\big|_{d=0} = (\hat{\psi}_u^*)^2 \,\mathrm{Var}(u), \qquad \mathrm{Cov}(\hat{\pi}, u)\big|_{d=0} = \hat{\psi}_u^* \rho_u \,\mathrm{Var}(u). \tag{A.46}$$

Derivative of the value-function coefficients at $d = 0$. We require three marginal coefficients. Direct differentiation of the expressions for $\ell_{\pi\pi}$, $\ell_{\pi u}$, and ℓ_{uu}, using $\psi_\pi\big|_{d=0} = 0$, $\hat{\psi}_\pi\big|_{d=0} = 0$, $\frac{d\psi_\pi}{dp}\big|_{d=0} = \frac{1}{1+\tilde{\theta}}$, and $\psi_u\big|_{d=0} = \frac{\tilde{\theta}}{1+\tilde{\theta}(1-\beta\rho_u)}$ yields

$$\left.\frac{d\ell_{\pi\pi}}{dp}\right|_{d=0} = 1, \tag{A.47}$$

$$\left.\frac{d\ell_{\pi u}}{dp}\right|_{d=0} = \frac{-\tilde{\theta}}{1+\tilde{\theta}(1-\beta\rho_u)}. \tag{A.48}$$

For ℓ_{uu}, the derivative takes the form established in the preceding subsection, with the additional continuation terms now included:

$$\begin{aligned}\left.\frac{d\big[\ell_{uu}(1-\beta\rho_u^2)\big]}{dp}\right|_{d=0} = {} & \underbrace{\psi_u^2\big|_{d=0}}_{\text{direct penalty}} + \underbrace{2\left.\frac{d\psi_u}{dp}\right|_{d=0} \cdot \frac{-\tilde{\theta}\beta\rho_u}{1+\tilde{\theta}(1-\beta\rho_u)}}_{\text{shift in } \psi_u} + \underbrace{\frac{2\tilde{\theta}\beta\,\hat{\psi}_u^*}{(1+\tilde{\theta})(1+\tilde{\theta}(1-\beta\rho_u))}}_{\text{expectations channel}} \\ & + \underbrace{\beta\big(\hat{\psi}_u^*\big)^2 - \frac{2\beta\tilde{\theta}\rho_u\,\hat{\psi}_u^*}{1+\tilde{\theta}(1-\beta\rho_u)}}_{\text{continuation value}}.\end{aligned} \tag{A.49}$$

Taking the total derivative of $\bar{L}$. Since $\ell_{\pi\pi}\big|_{d=0} = \ell_{\pi u}\big|_{d=0} = 0$, any d-dependence of the ergodic moments $\mathrm{Var}(\hat{\pi})$ and $\mathrm{Cov}(\hat{\pi}, u)$ is multiplied by zero; only the derivatives of the value-function coefficients survive. Differentiating (A.44):

$$\left.\frac{d\bar{L}}{dp}\right|_{d=0} = \underbrace{\left(\frac{\beta\sigma_\varepsilon^2}{1-\beta} + \mathrm{Var}(u)\right)\left.\frac{d\ell_{uu}}{dp}\right|_{d=0}}_{\equiv\ T_1} + \underbrace{\left.\frac{d\ell_{\pi\pi}}{dp}\right|_{d=0}\big(\hat{\psi}_u^*\big)^2\,\mathrm{Var}(u)}_{\equiv\ T_2} + \underbrace{2\left.\frac{d\ell_{\pi u}}{dp}\right|_{d=0}\hat{\psi}_u^*\rho_u\,\mathrm{Var}(u)}_{\equiv\ T_3}. \tag{A.50}$$

Proposition A.2. $\bar{L}$ is strictly increasing in d at $d = 0$. In particular, for $\rho_u = 0$:

$$\left.\frac{d\bar{L}}{dp}\right|_{d=0,\,\rho_u=0} = \frac{\sigma_\varepsilon^2}{1-\beta}\underbrace{\frac{\tilde{\theta}^2}{(1+\tilde{\theta})^2}\left(1-\frac{\beta}{(1+\tilde{\theta})^2}\right)}_{d\ell_{uu}/dp>0} + \underbrace{\frac{\tilde{\theta}^2}{(1+\tilde{\theta})^4}\sigma_\varepsilon^2}_{T_2>0} > 0. \tag{A.51}$$

Proof. Case $\rho_u = 0$ (i.i.d. shocks). When $\rho_u = 0$: $\mathrm{Var}(u) = \sigma_\varepsilon^2$, $\mathrm{Cov}(\hat{\pi}, u) = 0$ so $T_3 = 0$, and $\hat{\psi}_u^* = -\tilde{\theta}/(1+\tilde{\theta})^2$. In (A.49) the "shift in ψ_u" term and the second continuation term vanish (both proportional to ρ_u), while the expectations-channel and first continuation terms combine to

$$\frac{-2\tilde{\theta}^2\beta}{(1+\tilde{\theta})^4} + \frac{\beta\tilde{\theta}^2}{(1+\tilde{\theta})^4} = \frac{-\tilde{\theta}^2\beta}{(1+\tilde{\theta})^4}.$$

Adding the direct penalty $\psi_u^2 = \tilde{\theta}^2/(1+\tilde{\theta})^2$:

$$\left.\frac{d\ell_{uu}}{dp}\right|_{d=0,\,\rho_u=0} = \frac{\tilde{\theta}^2}{(1+\tilde{\theta})^2}\left(1-\frac{\beta}{(1+\tilde{\theta})^2}\right) > 0,$$

since $\beta < 1$ and $(1+\tilde{\theta})^2 \geq 1$. Hence $T_1 > 0$. Meanwhile $T_2 = (\hat{\psi}_u^*)^2\sigma_\varepsilon^2 = \tilde{\theta}^2\sigma_\varepsilon^2/(1+\tilde{\theta})^4 > 0$. Equation (A.51) follows.

General $\rho_u \in [0, 1)$. The term T_1 remains strictly positive: the direct penalty $\psi_u^2 > 0$ dominates because the indirect terms in (A.49) are of order $\beta\tilde{\theta}^2/(1+\tilde{\theta})^4$ while the direct term is $\tilde{\theta}^2/(1+\tilde{\theta})^2$. The term $T_2 = (\hat{\psi}_u^*)^2\,\mathrm{Var}(u) > 0$ is unambiguously positive. The cross-term T_3 can be written, using (A.48) and (A.45), as

$$T_3 = \frac{-2\tilde{\theta}^2\rho_u^2\big(\rho_u(1+\tilde{\theta})-1\big)}{(1+\tilde{\theta})\big(1+\tilde{\theta}(1-\beta\rho_u)\big)^2}\,\mathrm{Var}(u). \tag{A.52}$$

When $\rho_u < 1/(1+\tilde{\theta})$, the factor $\rho_u(1+\tilde{\theta})-1 < 0$ makes $T_3 > 0$. When $\rho_u > 1/(1+\tilde{\theta})$, $T_3 < 0$ but remains bounded, while $T_1 + T_2$ grows with $\mathrm{Var}(u) = \sigma_\varepsilon^2/(1-\rho_u^2)$. For all admissible $(\tilde{\theta}, \beta, \rho_u)$ one verifies $T_1 + T_2 + T_3 > 0$.[2] □

Interpretation. Two distinct forces make the unconditional loss strictly increasing at $d = 0$. First, as in the conditional analysis, the direct penalty $d(\pi_t - \hat{\pi}_t)^2$ imposes a first-order cost proportional to the variance of inflation under discretion ($T_1 > 0$). The expectations-management benefit embedded in the same term partially offsets this cost, but insuffi-

[2] A sufficient condition is $T_2 + T_3 \geq 0$, which can be verified by substituting (A.45) into $T_2 + T_3$ and checking that the resulting quadratic in ρ_u is non-negative on $[0, 1)$.

ciently: the promise mechanism is dormant at $d = 0$ and its gains are attenuated by the factor $\beta/(1+\tilde{\theta})^2 < 1$.

Second, and unique to the unconditional analysis, a variance-of-promises channel appears ($T_2 > 0$): for any $d > 0$, the promise $\hat{\pi}_t$ fluctuates with variance proportional to $(\hat{\psi}_u^*)^2 \operatorname{Var}(u)$. Since $d\ell_{\pi\pi}/dp|_{d=0} = 1$, the value function immediately loads on this variance, raising the unconditional loss. This term is absent from the conditional analysis (which holds $\hat{\pi}$ fixed) and makes the case against local optimality of small d even stronger.

In summary, while larger values of d can activate the commitment channel sufficiently to reduce welfare losses relative to discretion, the gain is not local: any improvement requires d to be bounded away from zero.

A.5 Implied interest rate promises

We use Euler equation to back out interest rates. The Euler equation is given by:

$$x_t = \mathbb{E}_t[x_{t+1}] - \frac{i_t - \mathbb{E}_t[\pi_{t+1}] - r_t^*}{\sigma} \tag{A.53}$$

$$\Leftrightarrow \; i_t = \sigma\left(\mathbb{E}_t[x_{t+1}] - x_t\right) + \mathbb{E}_t[\pi_{t+1}] + r_t^*, \tag{A.54}$$

where r_t^* is a zero mean exogenous, stochastic process.

Plugging in the laws of motion:

$$i_t = \iota_r r_t^* + \iota_\pi \hat{\pi} + \iota_u u = \sigma\left(\xi_\pi(\hat{\psi}_\pi\hat{\pi} + \hat{\psi}_u u) + \xi_u\rho_u u - \xi_\pi\hat{\pi} - \xi_u\rho_u u\right) + \psi_\pi(\hat{\psi}_\pi\hat{\pi} + \hat{\psi}_u u) + \psi_u\rho_u u + r^*$$

It follows that $\iota_0 = r^*$ and

$$\iota_\pi = \psi_\pi\hat{\psi}_\pi - \sigma\xi_\pi(1 - \hat{\psi}_\pi) \tag{A.55}$$

$$\iota_u = \psi_\pi\hat{\psi}_u + \psi_u\rho_u + \sigma\xi_\pi\hat{\psi}_u + \sigma\xi_u(1-\rho_u) \tag{A.56}$$

$$\iota_r = 1 \tag{A.57}$$

Note that $\iota_u \neq 0$ even if $d \to \infty$, because even if inflation becomes insensitive to the cost-push shock, output still varies with the cost-push shock ($\xi_u \neq 0$). Interest rates move 1:1 in response to r_t^* shocks.