Reassessing Central Bank Reputation:

Beyond Long-Run Expectations†

Tomás E. Caravello   Alex Carrasco   Pedro Martínez-Bruera

March 26, 2026

Click here for the latest version.

We study the implications of uncertainty about a central bank's preference for inflation stability for monetary policy design. A hawkish reputation, defined as a reputation for placing large weight to inflation stabilization, reduces stabilization costs by dampening the pass-through of shocks to short-run inflation expectations. Because the private sector learns from the bank's policy actions, the optimal response to cost-push shocks is more aggressive than that of a myopic policymaker: the bank internalizes the value of its reputation. Furthermore, the bank treats reputation as an asset: it invests in it when perceived as dovish, and spends it when perceived as hawkish. Using cross-sectional variation in U.S. private forecasts, we find evidence consistent with these mechanisms. Quantitatively, welfare gains from the optimal policy are larger during the Great Moderation than during the Great Inflation, and delegating policy to a conservative but myopic central banker, as in Rogoff (1985), closely approximates the optimal policy.

†Martínez-Bruera: MIT, pmbruera@mit.edu. I thank my advisors, Iván Werning, Christian K. Wolf, and Ricardo Caballero, for guidance throughout this project. Caravello: MIT, tomasec@mit.edu. Carrasco: MIT, alexcm@mit.edu Comments and suggestions from Olivier Blanchard, Javier García-Cicco, Emilio Espino, Gonzalo Huertas, Vicente Jimenez G., Jackson Mejia, Valeria Morales, Andy Neumeyer, Mikkel Plagborg-Møller, Jim Poterba, Francisco Roldán, and David Romer, as well as seminar participants at Banca d'Italia, CREI, EPGE-FGV Rio, MIT Macro Lunch, MIT Macroeconomics Seminar, The University of Chicago, Universidad de Buenos Aires, Universidad de San Andrés, Universidad Torcuato di Tella, and University of Notre Name are greatly appreciated.

1 Introduction

Central bank reputation matters for the effectiveness of monetary policy: when the public believes the central bank is committed to low and stable inflation, disinflation can be achieved at lower output cost. A long tradition in monetary economics emphasizes that time-inconsistency and fiscal dominance can undermine this reputation (Barro and Gordon 1983; Rogoff 1985; Kydland and Prescott 1977). The solutions to these concerns focused on achieving a low inflation target through institutional design (Rogoff 1985) or commitment mechanisms (Barro and Gordon 1983), and central banks have succeeded on this dimension: in advanced economies, *long-run* inflation expectations—an empirical measure of target credibility—have remained anchored and stable during periods of high volatility.

Yet, debates about central bank reputation persist, now focused on a different and theoretically underexplored dimension: uncertainty about the aggressiveness with which the central bank responds to cost-push shocks—that is, how it trades off inflation stabilization against output costs when the two objectives conflict in the *short run*. This short-run reputation shapes perceptions of whether central banks are "hawkish" or "dovish" and, more broadly, the private sector's beliefs about the bank's reaction function (Hamilton et al. 2011; Aikman et al. 2024; Bauer et al. 2024; Bocola et al. 2024). We build a framework to study the implications of this form of reputation, test its main predictions using US data, and quantify the welfare gains from following the optimal policy.

We consider a tractable New Keynesian model, where the central bank has a dual mandate over inflation and output gap stability. The bank's reaction to shocks is shaped by the relative priority given to inflation stability: a stronger priority leads to more aggressive rate increases in response to inflationary pressures. The private sector does not know this relative priority, and learns about it from observing the bank's actions.

We follow the misspecified learning literature (Esponda and Pouzo 2016; Sargent 1999) and assume the private sector believes the central bank is myopic: they incorrectly believe the bank optimizes period-by-period without internalizing how its actions affect beliefs. Given this, the private sector forms rational forecasts using Bayesian updating. This assumption enables us to summarize beliefs over the central bank's preferences, an infinite-dimensional object, into a sufficient statistic-reputation-that we can measure. A shift in beliefs toward a more hawkish central bank improves reputation.[1] Our model delivers two implications for

[1]We later extend the analysis to the case where the private sector correctly anticipates the bank's incen-

reputation.

First, it shapes the propagation of shocks: a hawkish reputation anchors short-run inflation expectations, reducing the pass-through of cost-push shocks and lowering the cost of stabilization.[2] Second, it evolves endogenously: since the private sector learns from the central bank's actions, reputation responds to monetary policy surprises. Although a hawkish reputation is always beneficial, maintaining it is costly: it requires strong reactions that signal a high commitment to inflation stability.

This tension between current stabilization costs and future reputation gains shapes the central bank's optimal policy. Reputation has two implications for optimal policy. First, optimal policy reacts more strongly to shocks than the myopic benchmark; it *overreacts* because it internalizes the reputation effects of its actions.[3] In response to a cost-push shock, the magnitude of the output gap response is larger and the magnitude of the inflation response is smaller than the myopic policy.

The second implication is that whether the central bank improves its reputation or not depends on its current reputation. Overreaction does not necessarily imply it will always surprise the private sector. The intuition is straightforward: while reputation is beneficial, a very hawkish reputation creates expectations of very aggressive reactions, which are costly for the central bank. We prove that when the bank is perceived as relatively dovish, optimal policy calls for improving it: in response to a positive cost-push shock, the bank raises rates by more than the private sector expects. Conversely, when the bank is perceived as relatively hawkish, the central bank spends its reputation: it raises rates by less than expected.

Our theoretical insights extend in two directions. First, overreaction holds under correctly specified beliefs and regardless of functional form assumptions on learning, requiring only that inflation moves in the direction of the shock signals weaker commitment. Second, the logic extends beyond cost-push shocks: when the divine coincidence fails, it is optimal to overreact to demand disturbances as well. Overreaction emerges as a general principle of optimal monetary policy.

---

tives and show that most results carry over, at the cost of losing the tractable measurement framework.

[2] In this sense, we provide a short- and medium-run analog of anchored expectations in Bernanke (2007): we focus on how a central bank's reputation shapes the propagation of shocks through short- and medium-run expectations, and thus the trade-offs it faces in real time.

[3] We use the term "overreaction" to emphasize that policy reacts more than what a myopic central bank with the same preferences would choose, not that the response is excessive relative to the social optimum or what the public expected.

Empirically, testing the two theoretical predictions of our model requires measuring reputation, which is unobservable from time-series data alone: realized outcomes conflate changes in reputation with changes in the state of the economy. Our theory is also informative about how to do so: when forecasters share the same beliefs about the central bank but disagree about future economic conditions, their forecasts co-move in a way that is informative about reputation. We prove that a simple cross-sectional regression of output gap forecasts on inflation forecasts recovers reputation at any point in time: a smaller slope signals that the private sector believes the central bank places greater priority on inflation stabilization

We estimate the Fed's reputation using the Blue Chip Survey of Financial Forecasts (BCFF), which provides individual forecasts from professional forecasters at major financial and economic institutions for the US economy. Using these estimates, we find evidence consistent with both predictions: a more hawkish reputation reduces the pass-through of oil price news shocks to inflation expectations, and an unexpected monetary tightening improves reputation. Furthermore, our measure captures a dimension of reputation distinct from traditional metrics: it has a near-zero correlation with long-run inflation expectations, the standard empirical proxy for target credibility. Finally, we show that the relationship between a central bank's perceived hawkishness and its perceived reduced-form Taylor rule coefficient on inflation is not monotone: a central bank with a high reputation has the same perceived coefficient as a central bank with a low one.

To quantify the importance of the reputation channel, we fit the model on quarterly U.S. data from two benchmark episodes: the Great Inflation (1973:Q1–1982:Q4) and the Great Moderation (1992:Q1–2019:Q4). We calibrate the model under the assumption that the Fed is myopic, and then calculate the counterfactual moments when the Fed follows the optimal policy.

The quantitative results support the central mechanism of the paper. In both episodes, the optimal central bank responds more aggressively to cost-push shocks than the myopic benchmark, reducing inflation volatility at the cost of higher output-gap volatility and thereby improving welfare through better future reputation. The gains are present in both samples, but they are larger in the Great Moderation, where the reputation channel has more time to compound, and more modest during the Great Inflation, where shocks are larger and the sample is shorter.

Even though the welfare gains from the optimal policy are quantitatively relevant, implementing it is demanding: it requires tracking reputation in real time, and understanding how the private sector interprets the central bank's actions.

This challenge echoes a classic problem in the literature on central bank reputation. In his seminal paper, Rogoff (1985) showed that a simple institutional solution, delegating monetary policy to a conservative central banker, could offset the inflationary bias arising from time inconsistency, lowering the long-run inflation target. Can Rogoff's approach approximate the optimal policy in our context?

To answer this question, we evaluate the performance of the optimal policy and the hawkish myopic delegate. We calibrate the latter to match the optimal policy's long-run reputation. While the optimal and hawkish myopic policies coincide for small shocks when reputation fluctuates around its stochastic steady state, they diverge with large shocks. Since large shocks are relatively infrequent, both policies deliver similar results in terms of welfare, while the myopic policy is substantially simpler to implement. Thus, appointing a hawkish myopic delegate is a close substitute for the optimal policy.

Related Literature This perspective connects to a broad literature on credibility, commitment, and expectations formation, while shifting the focus to dynamic short- and medium-run interactions. Our work mainly contributes to three literatures.

First, we shift the perspective of the central bank's reputation from perceived inflation bias to perceived commitment to stabilizing inflation. The classic literature emphasizes the strategic interaction that produces inflation bias (Barro and Gordon 1983; Kydland and Prescott 1977). There can be incentives to build reputation over time when there is uncertainty about the central bank's inflation bias (Backus and Driffill 1985, Barro 1986, Canzoneri 1985, Vickers 1986). However, private information may complicate inference and create perverse incentives (Cukierman and Meltzer 1986) and central banks may mimic types to avoid detection (King et al. 2008; Lu et al. 2016; Kostadinov and Roldán 2025). We reframe this logic: by overreacting, the bank mimics greater hawkishness than its true preferences, not to hide inflation bias, but to anchor short-run expectations.

Second, our empirical measure of reputation connects directly to recent work showing that perceptions of monetary policy vary over the business cycle (Hamilton et al. 2011, Bauer et al. 2024, Bocola et al. 2024). This literature interprets shifts in perceived Taylor-rule coefficients as changes in hawkishness. Given the skepticism about whether central banks actually follow such a rule (Svensson 2003; Nakamura et al. 2025), we provide a model where our empirical measure maps directly to perceptions of an optimizing central bank's preferences. Further, we show that there is no one-to-one mapping from the perceived Taylor-rule coefficient to reputation: a central bank with a stronger reputation may not need to react as aggressively because expectations are already well anchored.

Third, we contribute to the literature on robust solutions to credibility problems. Rogoff (1985) proposed delegating policy to a hawkish central banker as a way to reduce inflation bias. We show numerically that this prescription is not only a suitable approximation of the optimal policy, but also robust.

Finally, closest to our framework is the contemporaneous work by Bocola et al. (2025), who also study monetary policy when the public is uncertain about the central bank's relative weight on inflation stability. Empirically, they examine how reputation affects the anchoring of long-run inflation expectations. We instead focus on short-run expectations and use cross-sectional forecast data for our empirical analysis.[4] We view our contributions as complementary: reputation for hawkishness matters not only for the anchoring of long-run inflation expectations, but also shapes the short-run costs of stabilization

Roadmap The rest of the paper is organized as follows. Section 2 develops the main mechanism in a three-period economy. Section 3 describes the economy. Section 4 derives the optimal policy. Section 5 presents empirical evidence from U.S. forecast data. **??** provides a quantitative exercise evaluating the relative performance of the optimal policy. **??** provides evaluates the performance of delegating to a hawkish but myopic central banker.

2 The Reputation Channel in a Three-Period Model

To build intuition, this section introduces a simple model of reputation as perceived hawkishness. Private-sector learning makes it optimal to trade a deeper recession today in exchange for a better inflation-output trade-off tomorrow. To illustrate the mechanism, we use a three-period New Keynesian model. The Phillips curve is given by

$$\pi_t = \kappa y_t + \beta \mathbb{E}_t^P \left[\pi_{t+1}\right] + \varepsilon_t \qquad t = 0, 1, 2,$$

where y_t is the output gap, π_t is inflation, and ε_t is a cost-push shock. $\mathbb{E}_t^P[\cdot]$ denotes private-sector expectations, and $\mathbb{E}_t^P[\pi_3] = 0$. We assume the expectations operator, $\mathbb{E}_t^P[\cdot]$ is fixed

[4] There are also modeling differences. Bocola et al. (2025) adopts a two-type environment and assumes the private sector internalizes that the central bank follows its optimal policy. We allow for a continuum of types and, in our baseline, assume agents believe the central bank is myopic. However, as shown in Section 4, this distinction is not essential for our results.

at time t, but it varies over time. The central bank has a dual mandate over the output gap and inflation,

$$\mathcal{W}_0 = -\frac{1}{2}\mathbb{E}_0^{CB}\left[\sum_{t=0}^{2}\beta^t\left(y_t^2 + \lambda\pi_t^2\right)\right]$$

where $\mathbb{E}_t^{CB}[\cdot]$ denotes the central bank's expectations, and λ its relative weight on inflation stability. The private sector knows the structure of the economy but not the value of λ. The true preference is fixed, but what evolves is the private sector's perception of it: the bank's reputation, which shapes how the private sector's expectations—$\mathbb{E}_t^P[\cdot]$—are formed. The central bank maximizes $\mathcal{W}_0$ subject to the Phillips curve. This toy model previews the trade-off at the heart of our results for optimal policy. While the full framework generalizes these insights, the simplified setting captures the main logic.

Solving by backward induction yields, at $t = 2$,

$$y_2 = -\underbrace{\frac{\kappa\lambda}{1+\kappa^2\lambda}}_{\psi^y}\varepsilon_2 \qquad \pi_2 = \underbrace{\frac{1}{1+\kappa^2\lambda}}_{\psi^\pi}\varepsilon_2.$$

At $t = 1$, and given y_2 and π_2

$$y_1 = -\psi^y\left(\varepsilon_1 + \beta\mathbb{E}_1^P[\pi_2]\right) \qquad \pi_1 = \psi^\pi\left(\varepsilon_1 + \beta\mathbb{E}_1^P[\pi_2]\right).$$

Finally, at $t = 0$, noticing that even though π_2 is fixed for each realization, but $\mathbb{E}_1^P[\cdot]$ is not

$$y_0 = \overbrace{-\psi^y\left(\varepsilon_0 + \beta\mathbb{E}_0^P[\pi_1]\right)}^{\text{Myopic Stabilization}} \overbrace{+\psi^\pi\beta\mathbb{E}_t^{CB}\left[\beta\frac{\partial\mathbb{E}_1^P[\pi_2]}{\partial\pi_0}y_1\right]}^{\text{Intertemporal Smoothing}}$$

$$\pi_0 = \overbrace{\psi^\pi\left(\varepsilon_0 + \beta\mathbb{E}_0^P[\pi_1]\right)}^{\text{Myopic Stabilization}} \overbrace{+\kappa\psi^\pi\beta\mathbb{E}_t^{CB}\left[\beta\frac{\partial\mathbb{E}_1^P[\pi_2]}{\partial\pi_0}y_1\right]}^{\text{Intertemporal Smoothing}}.$$

Two terms characterize the optimal policy. The first, *myopic stabilization*, is the policy of a central bank without commitment, stabilizing the output gap and inflation each period. The second, *intertemporal smoothing*, captures the ability of current policy to influence future inflation expectations. Formally, the effect of the current output gap on future expected inflation is given by

$$\frac{\partial\mathbb{E}_1^P[\pi_2]}{\partial\pi_0} = \frac{\partial\mathbb{E}_1^P[\psi^\pi]}{\partial\pi_0}\mathbb{E}_1^P[\varepsilon_2].$$

Because ψ^π decreases with λ, the pass-through of shocks to expectations depends on the central bank's perceived hawkishness. Intuitively, if the central bank is perceived as more hawkish, agents expect shocks to pass through less strongly to expectations. This dependence of pass-through to beliefs is precisely where the reputation channel operates: a hawkish stance shifts beliefs toward higher values of λ, dampening the effect of shocks on expectations.

We refer to this as the *reputation channel*. When the private sector perceives the central bank as highly committed to stable inflation, it has a *hawkish* reputation; otherwise, it has a *dovish* one. Reputation shapes the sensitivity of inflation expectations to shocks. Having isolated the mechanism in a three-period model, we generalize it to an infinite-horizon New Keynesian economy.

3 The Economy

The economy consists in a textbook New Keynesian model with incomplete information. This section lays out the main equations that define the equilibrium in our economy. We provide a detailed description and derivation in Appendix A

Private Sector Block The competitive equilibrium in this economy is summarized by a dynamic IS equation and a New Keynesian Phillips Curve (NKPC):

$$y_t = \mathbb{E}_t^P [y_{t+1}] - \frac{1}{\sigma} \left(i_t - \mathbb{E}_t^P [\pi_{t+1}] - r_t^n \right) \tag{1}$$

$$\pi_t = \kappa y_t + \beta \mathbb{E}_t^P [\pi_{t+1}] + \varepsilon_t \tag{2}$$

where y_t is the output gap, π_t is inflation, i_t is the risk-free nominal interest rate, r_t^n is the real natural rate, and ε_t is a markup (cost-push) shock. Following Woodford (2003a), demand and productivity shocks enter through the natural rate, r_t^n:

$$r_t^n \equiv \rho + \upsilon \mathbb{E}_t^P [\Delta a_{t+1}] - \mathbb{E}_t^P [\Delta z_{t+1}]$$

where a_t is (log) TFP, and z_t (log) embodies shocks to the household's discount factor. A positive demand shock increases the natural rate, while a positive productivity shock lowers it. The expectations operator $\mathbb{E}^P [\cdot]$ denotes private-sector beliefs, which satisfy the Law of Iterated Expectations. We assume both households and firms have the same information set. Equations (1) and (2) fix notation and make clear where private-sector expectations enter the equilibrium.

Central Bank and Monetary Policy Regime The central bank's dual mandate over the output gap and inflation is represented by the welfare loss function

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t^{CB}\left[\sum_{k=0}^{\infty}\beta^k\left(y_{t+k}^2 + \lambda\pi_{t+k}^2\right)\right] \tag{3}$$

where $\mathbb{E}_t^{CB}[\cdot]$ denotes the central bank's expectations. The parameter λ measures the weight placed on inflation stabilization relative to output-gap stabilization, and the objective will be optimized subject to the private-sector block, with expectations playing the key role in transmitting reputation.

This specification, while stylized, is standard: it arises from a second-order approximation to household welfare in this economy and is widely used in applied policy analysis (e.g., Federal Reserve Board 2016, Barnichon and Mesters 2023).[5] In this approximation, the parameter λ depends on deep parameters of the economy. Here, we assume it is exogenous and dictated by the policymaker's preferences.

To avoid full revelation of the central bank's preferences from its actions, we include the following information friction: At the start of period t, the central bank announces the nominal interest rate, i_t, before the current demand and productivity shocks are realized, that is, before the current value of the natural rate of interest is realized. The central bank sets policy based on its private forecast of the natural rate $\mathbb{E}_t^{CB}[r_t^n]]$ which differs from the eventual realization by the forecast error ν_t:

$$r_t^n = \mathbb{E}_t^{CB}[r_t^n] + \nu_t$$

where $\mathbb{E}_t^P[\nu_t] = 0$. The private sector does not observe the central bank's forecast, only the realized natural rate. Since $\mathbb{E}_t^P[\nu_t] = 0$, the private sector believes the central bank is not biased on average

$$\mathbb{E}_t^P\left[\mathbb{E}_t^{CB}[r_t^n]\right] = r_t^n.$$

Because the central bank's forecast is unobservable, the noise term ν_t prevents the private sector from perfectly inferring the central bank's preferences from observed policy choices.[6]

[5] We derive the second-order approximation of welfare in Appendix A.

[6] Another interpretation of ν_t is that it represents pure monetary policy shocks.

Private Sector Expectations' Formation Process The private sector knows the full structure of the economy, but not the central bank's relative weight on inflation, λ, nor their forecast for the natural rate $\mathbb{E}_t^{CB}[r_t^n]$. They believe the central bank is myopic, and share a common prior, μ_t, over possible values of $\lambda \in (0, \infty)$. They assume a central bank of type $\lambda = \tilde{\lambda}$ maximizes the dual mandate (3) subject to the ex-ante equilibrium conditions, (1) and (2)

$$\tilde{y}_t = \mathbb{E}_t^{CB}\left[\mathbb{E}_t^P[y_{t+1}]\right] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t^{CB}\left[\mathbb{E}_t^P[\pi_{t+1}]\right] - \mathbb{E}_t^{CB}[\tilde{r}_t^n]\right) \tag{4}$$

$$\tilde{\pi}_t = \kappa \tilde{y}_t + \beta \mathbb{E}_t^{CB}\left[\mathbb{E}_t^P[\pi_{t+1}]\right] + \varepsilon_t \tag{5}$$

where $\tilde{y}_t := \mathbb{E}_t^{CB}[y_t]$ and $\tilde{\pi}_t := \mathbb{E}_t^{CB}[\pi_t]$ denote the allocation the central bank seeks to implement.[7]

We impose no anticipated learning (Marcet and Sargent 1989; Eusepi and Preston 2018; Kreps 1998; Evans and Honkapohja 2001): agents forecast the future using today's beliefs, without accounting for the fact that they will update them in the future. This assumption avoids the infinite regress of beliefs about future beliefs and keeps the problem tractable.

The private sector's belief about λ determines how shocks today feed into expectations. Let $\psi^y := \frac{\kappa\lambda}{1+\kappa^2\lambda}$ and $\psi^\pi := \frac{1}{1+\lambda\kappa^2}$. From the private sector's perspective, the following result describes how a myopic central bank of type $\lambda = \tilde{\lambda}$ reacts to shocks.

Lemma 1. *Under no anticipated learning, a myopic central bank with $\lambda = \tilde{\lambda}$ implements:*

$$\tilde{y}_t^M\left(\tilde{\lambda}\right) = -\psi^y\left(\tilde{\lambda}\right)\sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^P[\psi^\pi]\right)^s \mathbb{E}_t[\varepsilon_{t+s}] = -\psi^y\left(\tilde{\lambda}\right)X_t \tag{6}$$

$$\tilde{\pi}_t^M\left(\tilde{\lambda}\right) = \psi^\pi\left(\tilde{\lambda}\right)\sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^P[\psi^\pi]\right)^s \mathbb{E}_t[\varepsilon_{t+s}] = \psi^\pi\left(\tilde{\lambda}\right)X_t \tag{7}$$

where $\psi^y\left(\tilde{\lambda}\right) = \frac{\kappa\tilde{\lambda}}{1+\kappa^2\tilde{\lambda}}$, $\psi^\pi\left(\tilde{\lambda}\right) = \frac{1}{1+\kappa^2\tilde{\lambda}}$, *and* $X_t = \sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^P[\psi^\pi]\right)^s \mathbb{E}_t[\varepsilon_{t+s}]$.

Proof. See Appendix A ■

Given these perceived allocations, the object of interest is how agents form expectations.

[7] We impose this assumption for tractability; in Appendix B we show that all conclusions continue to hold once it is relaxed.

The k-period-ahead forecasts of output gap and inflation are:

$$\mathbb{E}_t^P[y_{t+k}] = -\mathbb{E}_t^P[\psi^y]\,\mathbb{E}_t^P[X_{t+k}] + \mathbb{E}_t^P[\eta_{t+k}] \tag{8}$$

$$\mathbb{E}_t^P[\pi_{t+k}] = \mathbb{E}_t^P[\psi^\pi]\,\mathbb{E}_t^P[X_{t+k}] + \kappa\mathbb{E}_t^P[\eta_{t+k}] \tag{9}$$

where $\eta_t := y_t - \tilde{y}_t = -\frac{1}{\sigma}\nu_t$ denotes the central bank's forecast error for the output gap, and $\mathbb{E}_t^P[X_{t+k}] = \sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^P[\psi^\pi]\right)^s \mathbb{E}_t[\varepsilon_{t+k+s}]$. Under our modeling assumptions, $\mathbb{E}_t^P[\eta_{t+k}] = 0$, but it need not be the case in the data.[8] These forecasts depend on the expected values of ψ^y and ψ^π, the parameters through which reputation shapes the transmission of shocks.

Demand shocks are always fully stabilized in the baseline model, which is why they do not affect inflation expectations. In Appendix B we extend our insights to a case where demand shocks have a sizeable effect on expectations, and show that all of our conclusions hold.

The Central Bank's reputation Private-sector expectations for inflation and the output gap depend on beliefs about λ only through their effect on $\mathbb{E}_t^P[\psi^y]$ and $\mathbb{E}_t^P[\psi^\pi]$. Since $\psi^y = \kappa^{-1}(1-\psi^\pi)$, the two move in opposite directions: a higher $\mathbb{E}_t^P[\psi^\pi]$ means a lower $\mathbb{E}_t^P[\psi^y]$, and vice versa: stabilizing inflation comes at the cost of volatility of the output gap. We refer to $\mathbb{E}_t^P[\psi^\pi]$ as the ***reputation*** of the central bank, and say reputation *improves* when $\mathbb{E}_t^P[\psi^\pi]$ falls, that is, when the central bank is perceived as more hawkish.

Shifts in reputation correspond to movements in the entire belief distribution, μ_t, rather than just a change in the odds of any single type.[9] However, given our assumptions, reputation is a *sufficient statistic* for private-sector beliefs. That is, the equilibrium allocation and optimal policy depend on beliefs only through their effect on reputation. Hence, the infinite-dimensional state variable μ_t can be replaced by the one-dimensional object $\mathbb{E}_t^P[\psi^\pi]$.

From (8)-(9), a lower $\mathbb{E}_t^P[\psi^\pi]$ shapes expectations through two channels:

1. **Direct:** holding $\mathbb{E}_t^P[X_{t+k}]$ fixed, the private sector expects less inflation and a deeper recession in equilibrium in response to a positive cost-push shock.

2. **Indirect:** a stronger reputation makes the private sector discount future shocks more

[8] We demonstrate in Appendix C that the case with evolving beliefs over η_{t+k} is isomorphic to a setting where the private sector learns about the central bank's inflation bias or long-run inflation target.

[9] In Appendix A we derive comparative statics from the belief distribution to reputation.

heavily: if they are confident that the central bank will stabilize inflation in the future, those shocks matter less for today's expectations.

From the central bank's perspective, a lower value of $\mathbb{E}_t^P[\psi^\pi]$ is unambiguously desirable: if the private sector is confident that the central bank will stabilize inflation, expectations barely respond to shocks.[10] Achieving and maintaining such a reputation is costly, and the optimal policy trades off these costs against the benefits.

Timing and Belief Updating At the start of each period, the central bank privately forecasts future shocks. Based on these forecasts, it sets the nominal interest rate for that period. The private sector observes the policy choice but not the bank's internal forecast, takes its consumption and production decisions, and only then updates its beliefs once outcomes for period t are realized. In other words, consumption and production decisions come first, and belief updating about the bank's preferences comes afterward. Figure 1 summarizes the sequence.

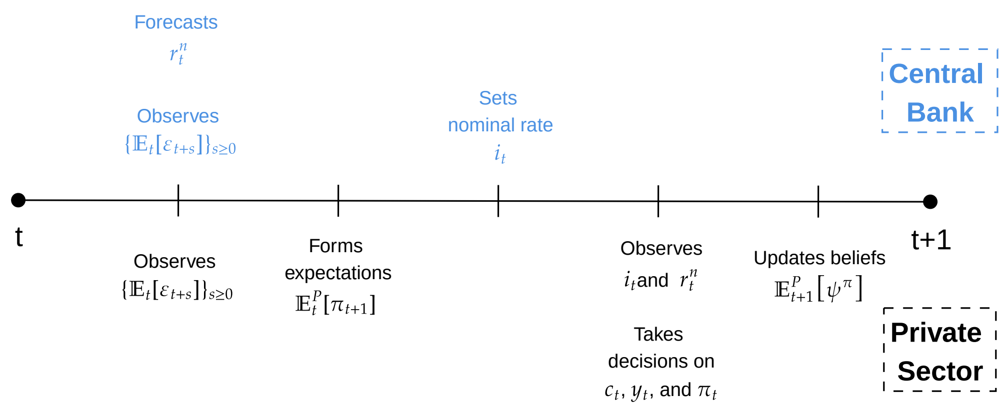


Figure 1: Timing within period

Given beliefs μ_t, households and firms update beliefs each period as follows. At the end of period t, after all variables have been realized, the public performs the following inference to update beliefs from μ_t to μ_{t+1}:

1. The cost-push shocks are realized $\{\mathbb{E}_t[\varepsilon_{t+s}]\}_{s\geqslant0}$.

[10]This may no longer be true at the Zero Lower Bound. A more dovish reputation can serve as a way to get out of the ZLB. A larger pass-through of shocks on inflation expectations can offset the effect of a negative natural interest rate. We explore this situation in Appendix C.

- Under μ_t, the public believes that a central bank of type $\lambda = \tilde{\lambda}$ chooses the allocations $\tilde{y}_t^M\left(\tilde{\lambda}\right)$ and $\tilde{\pi}_t^M\left(\tilde{\lambda}\right)$ given by (6) and (7), leading to the ex-post realizations

$$\tilde{y}_t^M\left(\tilde{\lambda}\right) + \eta_t \quad \text{and} \quad \tilde{\pi}_t^M\left(\tilde{\lambda}\right) + \kappa\eta_t$$

- The central bank sets the policy rate i_t to achieve allocations $\tilde{y}_t$ and $\tilde{\pi}_t$

2. The true realizations are

$$y_t = \tilde{y}_t + \eta_t \quad \text{and} \quad \pi_t = \tilde{\pi}_t + \kappa\eta_t$$

3. Upon observing π_t and y_t, the public cannot tell whether the deviations from their expectations are due to the forecast error η_t or a (myopic) central bank's type λ. A myopic central bank of type $\lambda = \tilde{\lambda}$ would choose $\tilde{\pi}^M\left(\tilde{\lambda}\right) = \psi^\pi\left(\tilde{\lambda}\right)X_t$, where $X_t = \sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t\left[\psi^\pi\right]\right)^s\mathbb{E}_t^P\left[\varepsilon_{t+s}\right]$. The private sector infers that such type's forecast error must have been

$$\eta_t\left(\tilde{\lambda}\right) = \kappa^{-1}\left(\pi_t - \psi^\pi\left(\tilde{\lambda}\right)X_t\right) = \kappa^{-1}\left(\tilde{\pi}_t - \psi^\pi\left(\tilde{\lambda}\right)X_t + \kappa\eta_t\right) \tag{10}$$

Beliefs are updated via Bayes' rule:

$$Pr\left(\tilde{\lambda}\middle|\,\pi_t, y_t\right) = \frac{Pr\left(\pi_t, y_t|\,\tilde{\lambda}\right)}{Pr\left(\pi_t, y_t\right)}Pr\left(\tilde{\lambda}\right) \Longleftrightarrow \mu_{t+1}\left(\tilde{\lambda}\right) = \frac{f_\eta\left(\eta_t\left(\tilde{\lambda}\right)\right)}{\int f_\eta\left(\eta_t\left(a\right)\right)\mu_t(a)da}\mu_t\left(\tilde{\lambda}\right) \tag{11}$$

where $f_\eta\left(\cdot\right)$ is the time-invariant probability distribution of η_t.[11]

Because ψ^π is a sufficient statistic for how λ affects inflation expectations (Lemma 1), it is equivalent, and more transparent, to describe beliefs over ψ^π rather than λ.

We assume $\eta_t \sim \mathcal{N}\left(0, \tau_\eta^{-1}\right)$ and that the prior over ψ^π is a truncated normal on $[0, 1]$:

$$\tilde{\psi}^\pi \sim \Psi_t\Big|_{\Psi_t\in[0,1]}, \quad \Psi_t \sim \mathcal{N}\left(\overline{\psi}_t, \tau_t^{-1}\right)$$

[11] Our conclusions are unchanged if forecast errors are expressed in deviations from expected inflation, $\tilde{\pi}_t - \tilde{\pi}_t^M$, expected output gap $\tilde{y}_t - \tilde{y}_t^M$, or the first-order condition $\tilde{y}_t^M + \lambda\kappa\tilde{\pi}_t^M$. We use inflation deviations for simplicity

Lemma 2. *Under these assumptions, the prior is conjugate, and*

$$\mathbb{E}_t^P\left[\psi^\pi\right] = \overline{\psi}_t - \frac{1}{\sqrt{\tau_t}} \frac{\phi\left(\sqrt{\tau_t}\left(1-\overline{\psi}_t\right)\right) - \phi\left(-\sqrt{\tau_t}\overline{\psi}_t\right)}{\Phi\left(\sqrt{\tau_t}\left(1-\overline{\psi}_t\right)\right) - \Phi\left(-\sqrt{\tau_t}\overline{\psi}_t\right)} \tag{12}$$

where ϕ and Φ are the standard normal PDF and CDF, and

$$\frac{\mathbb{E}_t^P\left[\psi^\pi\right]}{\partial\overline{\psi}_t} = \tau_t \mathbb{V}_t^P\left[\psi^\pi\right]$$

moreover, $\mathbb{V}_t^P\left[\psi^\pi\right]$ is hump-shaped, and peaks when $\overline{\psi}_t = \frac{1}{2}$.

Proof. This is a direct consequence of the truncated normal functional form assumption. ■

Lemma 2 establishes that the private sector's Bayesian learning has a conjugate prior, whose properties we discuss in detail in Appendix B.

A dovish central bank accommodates most cost-push shocks, so ψ^π is near one; a hawkish central bank prioritizes inflation stability, so ψ^π is close to zero, unlike two-type models (e.g. Backus and Driffill 1985, Bocola et al. 2025), we allow a continuum of types. Hence, reputation varies along an extensive margin: beliefs shift the entire distribution over ψ^π, not just the probability of a given type. Reputation is, then, the perception of how hawkish the central bank is, rather than the probability that the central bank is the hawkish one.

Under the above structure, $\Psi_{t+1} \sim \mathcal{N}\left(\overline{\psi}_{t+1}, \tau_{t+1}^{-1}\right)$ with

$$\overline{\psi}_{t+1} = \omega_t X_t^{-1} \overbrace{\left(\tilde{\pi}_t + \kappa\eta_t\right)}^{=\pi_t} + \left(1-\omega_t\right)\overline{\psi}_t \tag{13}$$

where

$$\omega_t = \frac{\left(\kappa^{-1}X_t\right)^2 \tau_\eta}{\tau_t + \left(\kappa^{-1}X_t\right)^2 \tau_\eta} \quad \text{and} \quad \tau_{t+1} = \tau_t + \left(\kappa^{-1}X_t\right)^2 \tau_\eta$$

ω_t is the Kalman gain: it rises when shocks are large relative to noise, but the *effective* updating weight $\omega_t X_t^{-1}$ is hump-shaped in X_t because large shocks dilute the signal about its type.

Policy affects beliefs through $\pi = \tilde{\pi}_t + \kappa\eta_t$: in the presence of inflationary shocks, $X_t > 0$, lower inflation shifts beliefs towards more hawkish types. Then, policy choices affect reputation, which will be central to the optimal policy problem we study next.

Central Bank's Problem given the private sectors beliefs, μ_t, the central bank maximizes

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t^{CB}\left[\sum_{s=0}^{\infty}\beta^s\left(\tilde{y}_t^2 + \lambda\tilde{\pi}_t^2\right)\right] \tag{PP}$$

subject to

(i) The New Keynesian Phillips Curve (NKPC)

$$\tilde{\pi}_t = \kappa\tilde{y}_t + \beta\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right] + \varepsilon_t \qquad \forall t$$

which links today's inflation to the output gap, expected inflation in $t+1$, and cost-push shocks.

(ii) The private sector's learning structure,

$$\begin{aligned}
\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right] =& \mathbb{E}_t^P\left[\psi^\pi\right]\mathbb{E}_t^P\left[X_{t+1}\right] \\
\mathbb{E}_t^P\left[\psi^\pi\right] =& \overline{\psi}_t - \frac{1}{\sqrt{\tau_t}}\frac{\phi\left(\tau_t\left(1-\overline{\psi}_t\right)\right) - \phi\left(-\tau_t\overline{\psi}_t\right)}{\Phi\left(\tau_t\left(1-\overline{\psi}_t\right)\right) - \Phi\left(-\tau_t\overline{\psi}_t\right)} \\
\overline{\psi}_{t+1} =& \overline{\psi}_t + \omega_t\left[\left(\mathbb{E}_t^P\left[\psi^\pi\right] - \overline{\psi}_t\right) + \kappa\frac{\eta_t}{X_t} + \left(\frac{\tilde{\pi}_t}{X_t} - \mathbb{E}_t^P\left[\psi^\pi\right]\right)\right]
\end{aligned}$$

where $\omega_t = \frac{\left(\kappa^{-1}X_t\right)^2\tau_\eta}{\tau_t + \left(\kappa^{-1}X_t\right)^2\tau_\eta}$ and $\tau_{t+1} = \tau_t + \left(\kappa^{-1}X_t\right)^2\tau_\eta$

We do not need to write the dynamic IS equation (4) explicitly. Given any desired path for (ex-ante) inflation and the output gap, the central bank can choose the interest rate to implement it.[12] More importantly, Forward Guidance is irrelevant. At time t, the private sector's inflation expectations depend only on the central bank's current reputation and cost-push shocks. Promises about the future path of variables have no effect, making the policy time-consistent.

4 Optimal Policy

We now turn to the central bank's optimal policy when its reputation evolves endogenously. At each date t, the central bank chooses the (ex-ante) output gap, $\tilde{y}_t$, and inflation, $\tilde{\pi}_t$,

[12]We assume the Zero Lower Bound never binds.

balancing the conventional stabilization trade-off with the benefits from influencing how it is perceived.

4.1 Main Results

Our theoretical analysis yields two main results: overreaction and reputation management. Consider a positive cost-push shock. Relative to a myopic central bank, the optimal policy overreacts: it raises rates more. Relative to the private sector's expectation, the optimal policy manages its reputation: improve it when it is low, and spend it when it is high. We now explain the mechanisms driving each result.

Overreaction Our first result is that the optimal policy reacts more aggressively than the myopic benchmark. Consider a positive cost-push shock. The myopic problem is the familiar trade-off between inflation and recession. However, it misses that the central bank's actions also reveal information about its preferences: conditional on a forecast error η_t, higher inflation signals a lower taste for inflation stabilization. To see how this incentive shapes the optimal policy, notice that the optimal output gap satisfies

$$\tilde{y}_t\left(\tilde{\psi^\pi}\right) = \overbrace{\tilde{y}_t^M\left(\tilde{\psi^\pi}\right)}^{\text{Myopic Stabilization}} + \overbrace{\tilde{\psi^\pi}\sum_{s=1}^{\infty}\beta^s\mathbb{E}_t^{CB}\left[\beta\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+1+s}\right]}{\partial\tilde{\pi}_t}\tilde{y}_{t+s}\left(\tilde{\psi^\pi}\right)\right]}^{\text{Intertemporal Smoothing}} \tag{14}$$

Intertemporal smoothing means that, if the central bank strengthens its reputation, future shocks will have a smaller impact on inflation expectations. This creates an incentive to overreact, relative to a myopic central bank. If the current shock is persistent, overreaction reduces its direct future damage. If it is *iid*, it insures against future shocks.

The policy trade-off is therefore between a deeper *current* recession in response to inflationary pressures, and a milder *future* one. Therefore, relative to the myopic benchmark, the optimal policy calls for a larger recession. For inflation, the logic is reversed: a larger recession today means, by the NKPC, lower inflation than the myopic benchmark.Proposition 1 formalizes this intuition

Proposition 1. *Under the optimal policy, output gap overreacts, and inflation underreacts. That is,*

$$|\tilde{y}_t| > \left|\tilde{y}_t^M\right| \quad \textit{and} \quad |\tilde{\pi}_t| < \left|\tilde{\pi}_t^M\right|$$

Proof. See Appendix B ■

Figure 2 illustrates the intuition behind Proposition 1 through the lens of price theory. In price theory, consumers maximize their utility subject to their budget constraint. Suppose there is a positive cost-push shock at time t. The NKPC plays the role of the budget constraint: the central bank chooses which point along it to sit at; that is, which combination of recession and inflation maximizes its preferences, dictated by the dual mandate. Since the central bank has a bliss point at zero output gap and inflation, utility increases toward the origin.

(a) Period-t **(b)** Period-$t+1$

Figure 2: Intuition behind Proposition 1, positive cost-push shock

Start with Figure 2a. The black dashed line depicts the NKPC at period t. A myopic central bank chooses the point at which its indifference curve (solid black line) is tangent to the NKPC, trading off recession and inflation without accounting for reputation. This leads to a certain reputation in $t+1$, and consequentially an NKPC at $t+1$, depicted by the dashed black line in Figure 2b. Is policy this optimal?

The red dashed indifference curve depicts the preferences of the optimal central bank, which factors in both the current inflation-output trade-off and private-sector learning.[13] Notice that this hypothetical indifference curve is flatter than that of the myopic central bank at the tangency point: the central bank internalizes that being tougher on inflation today improves its reputation, and therefore chooses a point with a larger recession and lower inflation. The payoff comes in period $t+1$: better reputation reduces the pass-through of shocks onto inflation expectations, shifting the NKPC inward in Figure 2b (dashed blue line), a more favorable budget constraint for the future.

[13]We do not claim this is a true indifference curve, since learning occurs in equilibrium and only points on the NKPC have well-defined learning dynamics. We invoke it purely for illustrative purposes.

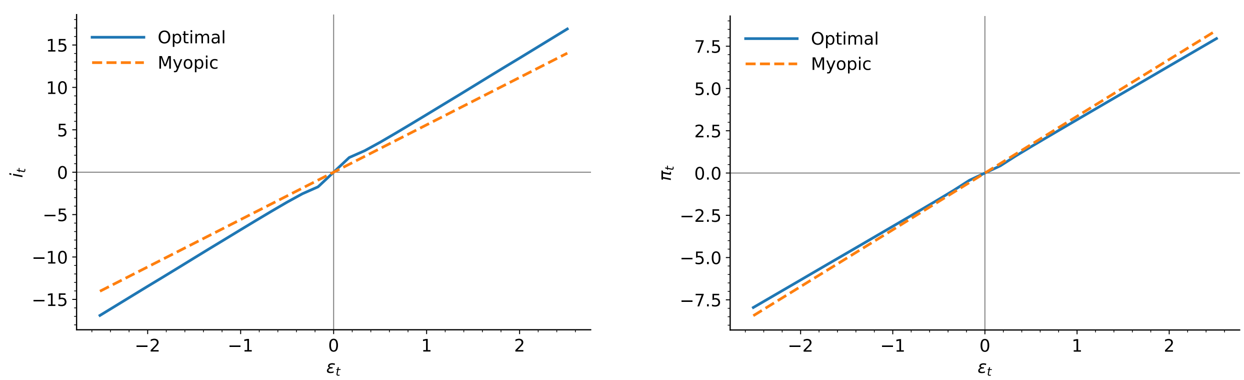
More formally, we need to determine the sign of $\frac{\partial \mathbb{E}^P_{t+s}[\tilde{\pi}_{t+s+1}]}{\partial \tilde{\pi}_t} \tilde{y}_{t+s}$. We can express $\frac{\partial \mathbb{E}^P_{t+s}[\tilde{\pi}_{t+s+1}]}{\partial \tilde{\pi}_t}$ as

$$\frac{\partial \mathbb{E}^P_{t+s}\left[\tilde{\pi}_{t+s+1}\right]}{\partial \tilde{\pi}_t} = \overbrace{\frac{\partial \mathbb{E}^P_{t+s}\left[\tilde{\pi}_{t+s+1}\right]}{\partial \mathbb{E}^P_{t+s}\left[\psi^\pi\right]}}^{-\text{sign}(\tilde{y}_{t+s})} \cdot \overbrace{\frac{\partial \mathbb{E}^P_{t+s}\left[\psi^\pi\right]}{\partial \overline{\psi}_{t+s}}}^{\geqslant 0} \cdot \overbrace{\frac{\partial \overline{\psi}_{t+s}}{\partial \tilde{\pi}_t}}^{\text{sign}(X_t)} \tag{15}$$

The first term captures that better reputation reduces the pass-through of shocks on inflation expectations. It is possible when cost-push shocks are positive, and negative when they are negative. The second term contains the mechanical effect of $\overline{\psi}_t$ on $\mathbb{E}^P_t\left[\psi^\pi\right]$, which is always positive. Finally, the third term reflects how today's inflation affects reputation: with positive cost-push shocks, higher inflation damages reputation, whereas with negative cost-push shocks it improves it.

The product of the first term and $\tilde{y}_{t+s}$ is always negative: worse reputation amplifies inflation expectations in response to shocks, requiring a deeper recession to stabilize inflation. The second term is always positive, reflecting the mechanical effect on $\overline{\psi}_t$ on $\mathbb{E}^P_t\left[\psi^\pi\right]$. Therefore, the overall sign of the expression is determined by the third term: when cost-push shocks are positive, $\frac{\partial \mathbb{E}^P_{t+s}[\tilde{\pi}_{t+s+1}]}{\partial \tilde{\pi}_t} \tilde{y}_{t+s} < 0$, and positive otherwise. The policy implication of Proposition 1 is simple: optimal policy overreacts relative to the myopic benchmark, always pushing rates further in the same direction.[14] The NKPC then converts the output gap into lower inflation.

Figure 3 plots the interest rate and inflation.



(a) Policy function for the interest rate i_t

(b) Policy function for inflation π_t

Figure 3: Policy functions. We assume that the shock ε_t follows an AR(1) process.

[14]We say "further in the same direction" rather than "raises rates more aggressively" because a sufficiently persistent cost-push shock with very low reputation may call for lowering rates — in which case optimal policy lowers them more.

Reputation works in the opposite direction from commitment. Under commitment, the central bank smooths the recession over time, trading a smaller current recession for a larger future one. With reputational concerns, this logic flips: a larger current contraction signals a high λ, resulting in a softer one in the future. This stands in contrast with the traditional literature on reputation, where the private sector is uncertain about whether the central bank has commitment. There, a discretionary central bank may have incentives to shift policy *towards* commitment to improve its reputation, while a commitment one may deviate from its standard solution only to protect it. Here, the optimal policy distorts *away* from commitment with the goal of improving the inflation-output trade-off in the future.

So far, we have established that the optimal policy always overreacts relative to the myopic benchmark — it pushes further in the same direction, regardless of the bank's current reputation. This raises a distinct question: does the bank always want to *improve* its reputation, or can it be optimal to let it *erode*? The answer, as we show next, depends on where the bank currently stands.

Reputation Management Overreaction means responding more aggressively than the myopic benchmark, but this need not improve the central bank's reputation. The dynamics of reputation depends on how hawkish or dovish it is perceived to be. From the central bank's perspective, (13), the better the reputation, lower $\overline{\psi}_t$, the stronger the response needed to maintain it.

By this logic, while a good reputation reduces stabilization costs, sustaining it is itself costly: it requires the central bank to respond aggressively to shocks, potentially at odds with its own preferences. The optimal policy problem can thus be recast as an investment problem: reputation is the capital stock, whose return is lower stabilization costs, and whose investment cost is deviating from the central bank's static optimum. Formally, rewrite (14) as

$$\overbrace{\tilde{y}_t^M\left(\tilde{\psi}^\pi\right)-\tilde{y}_t\left(\tilde{\psi}^\pi\right)}^{\text{Cost: Deviation from Myopic}}=\overbrace{-\tilde{\psi}^\pi\sum_{s=1}^{\infty}\beta^s\mathbb{E}_t^{CB}\left[\beta\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+1+s}\right]}{\partial\tilde{\pi}_t}\tilde{y}_{t+s}\left(\tilde{\psi}^\pi\right)\right]}^{\text{Payoff: Lower sensitivity of expectations}}$$

This equation formalizes the trade-off at the heart of optimal policy: the cost of deviating from what static preferences would dictate, against the payoff of a counterfactually better reputation. When reputation is low—relative to the bank's static preferences—the cost of improving it is small: overreaction is modest, while the payoff from better-anchored short-run expectations is large. The bank therefore invests in reputation. When reputation is high, however, sustaining it requires a large deviation from the static optimum, and it can be optimal for the policymaker to allow it to erode toward a more sustainable level. Proposition 2 formalizes this:

Proposition 2. *There exists a* $\widehat{\psi}$ *such that* $\mathbb{E}_t^P[\psi^\pi] > \mathbb{E}_t^{CB}\left[\mathbb{E}_{t+1}^P[\psi^\pi]\right]$ *when* $\mathbb{E}_t^P[\psi^\pi] > \widehat{\psi}$ *and* $\mathbb{E}_t^P[\psi^\pi] < \mathbb{E}_t^{CB}\left[\mathbb{E}_{t+1}^P[\psi^\pi]\right]$ *otherwise.*

Proof. See Appendix B ■

Notice that Proposition 2 is expressed in terms of the central bank's *expected* reputation on the following period. The central bank controls its reputation only in expectation: realized reputation also depends on forecast errors. A central bank that systematically underestimates the natural rate will lose reputation despite its best efforts, rationalizing why policymakers who "fall behind the curve" can suffer lasting credibility costs.

Figure 4 illustrates the intuition behind Proposition 2 through the lens of price theory. Suppose there is a positive cost-push shock. The dashed black line depicts the NKPC at period t, and the red dot depicts the allocation the central bank must implement to maintain its current reputation into the following period.

(a) Low Reputation **(b)** High Reputation

Figure 4: Dynamics of Reputation: Low vs High Reputation

Consider first Figure 4a, which displays the case of low reputation. Since reputation is low, sustaining it requires only a modest recession paired with relatively high inflation. Yet this is not optimal: at that allocation, the bank's indifference curve (red dashed line) is flatter than the NKPC, signaling a preference for lower inflation and a deeper recession. The bank therefore moves to the tangency of the NKPC and its indifference curve (dashed blue line), implementing a larger recession and lower inflation — and in doing so, improving its reputation.

Figure 4b shows the symmetric case of high reputation. Here, sustaining current reputation demands a very deep recession paired with very low inflation — a costly combination. At that point the indifference curve reveals a preference for more inflation and a shallower recession, so the bank moves away from the sustainability threshold, accepting a deterioration in reputation in exchange for a more favorable allocation.

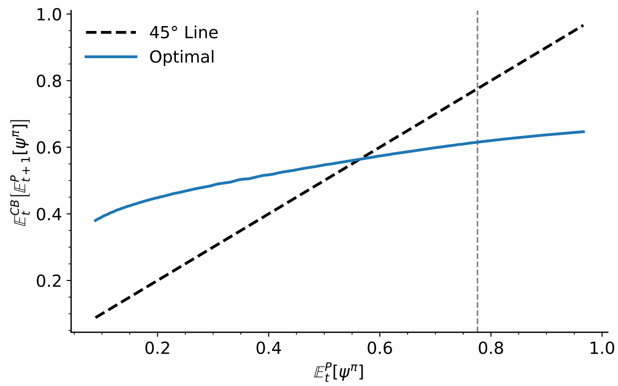


Figure 5: Reputation Dynamics

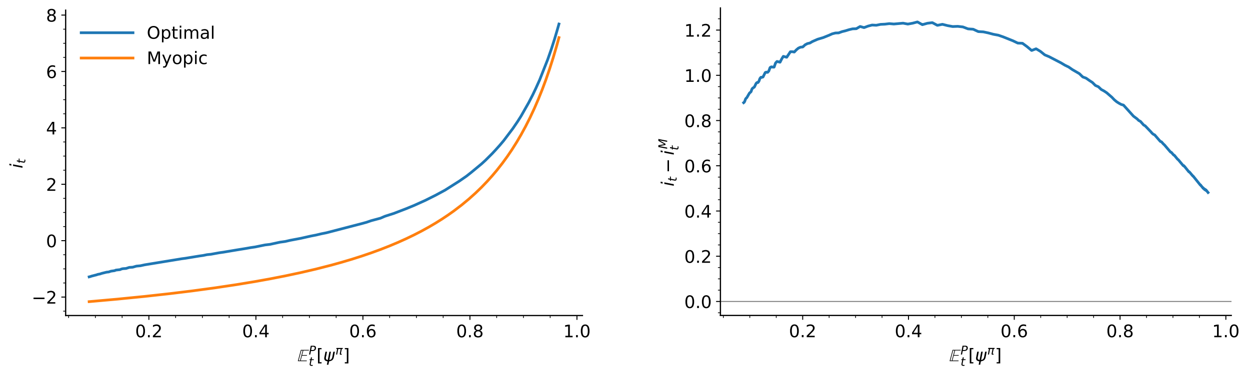
Figure 5 shows that the central bank manages its reputation toward a target, building it when perceived as dovish, letting it decline when perceived as very hawkish. We also observe a direct consequence of our result: there are no multiple equilibria in the long run, as the policy function crosses the 45-degree line only once. This management of reputation raises the question of when the central bank should build or spend its reputation.

4.2 Extensions

Beyond the baseline results, our framework speaks to several policy questions and extends naturally in a number of directions, as we show next.

State-Dependence Should the central bank overreact more when reputation is low or high? The answer, shown in Appendix B, is neither. From (14), overreaction is strongest when beliefs are most sensitive to the central bank's actions. Under our functional form assumptions, this occurs at intermediate levels of reputation: the private sector is uncertain whether the bank is hawkish or dovish. At the extremes, actions are uninformative and beliefs are very rigid. In the middle ground, however, actions are highly informative, and the incentive to overreact peaks. Thus, Overreaction follows an inverse-U shape in reputation.

Figure 6 plots the policy rate overreaction as a function of the central bank's reputation in response to a positive cost-push shock.



(a) Policy function for the interest rate i_t

(b) Interest rate overreaction $i_t - i_t^M$

Figure 6: Policy rate overreaction and the central bank's reputation. Positive shock $\varepsilon_t > 0$

Good Times vs Bad Times A long-debated policy question is when the central bank should build reputation and when it should spend it: is it better to build in good times and deploy it during crises, or are bad times precisely when the bank should "show what it is made of"? Our framework clarifies this.

The size of the shock shapes optimal policy through two channels. First, from (13), the effective weight on the bank's action, $\omega_t X_t^{-1}$, is non-monotone in shock size: it increases when the shock is small—a one percentage point deviation is highly informative about the bank's type—but decreases when the shock is large, as actions become harder to distinguish from the shock itself. Second, a large persistent shock signals a large future recession, raising the incentives to maintain a high reputation.

Taken together, these two channels clarify the policy question. When a shock is large, the bank's actions are less informative, reducing the incentive to overreact. But if the shock is also persistent, spending reputation is unwise: the bank correctly anticipates a large recession in the future, and a good reputation remains valuable. Therefore, the only case in which spending reputation is unambiguously optimal is when the shock is both large and short-lived.

Demand Shocks So far, our discussion has focused on cost-push shocks, which are relatively rare compared to demand-driven fluctuations. Do the same results carry over? To

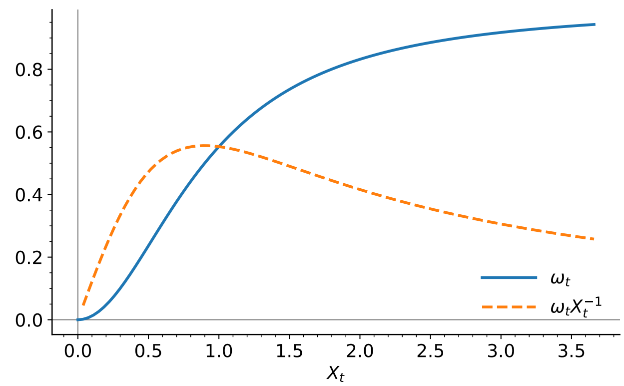


Figure 7: Effective Weight.

address this, we consider a setting where the Divine Coincidence does not hold, modifying the welfare function to penalize large movements in the policy rate:

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t^{CB}\left[\sum_{s=0}^{\infty}\beta^s\left(\tilde{y}_{t+s}^2 + \lambda\tilde{\pi}_{t+s}^2 + \varphi i_{t+s}^2\right)\right] \tag{16}$$

The third term, inspired by Woodford (2003c), captures gradualism in rate-setting. When $\varphi \to 0$, we recover the benchmark where demand shocks are fully stabilized. In Appendix C, we show that if prices are fixed and the private sector is uncertain about φ, the optimal policy problem is isomorphic to the cost-push case: overreaction in the interest rate, underreaction in the output gap, and all conclusions carry over. In this case, reputation is not about the relative priority given to price stability, but rather the central bank's willingness to move rates aggressively to stabilize the economy. These two interpretations of reputation are complementary, and the logic of optimal policy extends naturally to demand-driven fluctuations.

Communication In the baseline model, the central bank cannot communicate its internal forecasts to the private sector. Central bank communication has been widely studied as a tool for managing expectations (Blinder et al., 2008, 2024). Suppose the central bank were able to communicate its own internal forecasts. What would its incentives be?

Consider a positive cost-push shock, which calls for raising rates. To improve its reputation, the central bank can overreact. But it has an additional instrument: it can announce that its estimate of the natural rate is low. By doing so, it leads the private sector to believe that demand conditions alone do not warrant a large rate increase — so if the central bank raises rates aggressively nonetheless, it must be because it is genuinely tough on inflation. Communication thus amplifies the signal embedded in the policy action. More generally, the central bank has an incentive to bias its announcements in the direction *opposite* to the cost-push shock: downward when the shock is positive, and upward when it is negative. Reputation therefore shapes not only the central bank's *actions* but also its *communication strategy*, suggesting that the two should be studied jointly.

Correctly Specified Beliefs So far, we assumed for tractability that the private sector held misspecified beliefs and assumed the central bank to be myopic. In reality, the private sector may be aware that the central bank internalizes the impact of its actions on their beliefs. To capture this, we consider a model in which private beliefs are correctly specified. In Appendix C we prove that, as long as higher inflation signals a lower λ when there are positive shocks (and the opposite when shocks are negative), the central bank continues to overreact relative to the myopic benchmark. The magnitude of overreaction may be smaller than in the misspecified-beliefs case, since private agents correctly anticipate the central bank's incentives, but the qualitative result remains. This result confirms that overreaction is a robust feature of optimal policy. Furthermore, in a three-period economy, all of our results hold.

Reputation about the Long Run Following the traditional literature on reputation and monetary policy, suppose the private sector knows the central bank's relative weight on inflation, λ, but is uncertain about its inflation bias. Uncertainty about the inflation bias is analogous to a situation where the private sector learns about the central bank's inflation target. In Appendix C, we analyze this case and derive the implications for optimal policy. In an effort to stabilize future outcomes, the central bank overreacts to news shocks. This policy can be interpreted as akin to Average Inflation Targeting (see Powell 2020; Eggertsson and Kohn 2023).

5 The Reputation Channel in the Data

Our theory delivers two testable implications. First, the pass-through of cost-push shocks to inflation expectations is smaller when the central bank has a stronger reputation. Second, reputation responds to monetary policy surprises. To test for them, we need to measure reputation, which is unobservable. Our theory is also informative about how to do so. In this section, we show how to retrieve reputation from forecast data and test both predictions using US data.

A key empirical challenge in measuring reputation is that it evolves endogenously over time. Using time-series variation in *realized* output gaps and inflation would conflate changes in the central bank's reputation with changes in the state of the economy. Our theory provides a way to measure reputation, not from time-series variation of realized outcomes, but from cross-sectional variation in *forecasts*.

Suppose at time t, there is a continuum of forecasters indexed by i who share a common model of the economy and identical beliefs about the central bank's preferences. In particular, they all believe the central bank is myopic, have the same prior over the central bank's λ and use the same estimate of the slope of the NKPC, κ.

At the beginning of period t, each forecaster receives a private signal of the period's shocks and forms forecasts accordingly. This induces cross-sectional variation in forecasts that disappears once the shocks are realized and publicly observed. Under this assumption, forecaster i's forecasts for the output gap and inflation are

$$\mathbb{E}_t^i\left[y_{t+k}\right] = -\,\mathbb{E}_t^P\left[\psi^y\right]\mathbb{E}_t^i\left[X_{t+k}\right] + \mathbb{E}_t^i\left[\eta_{t+k}\right] \tag{17}$$

$$\mathbb{E}_t^i\left[\pi_{t+k}\right] = \mathbb{E}_t^P\left[\psi^\pi\right]\mathbb{E}_t^i\left[X_{t+k}\right] + \kappa\mathbb{E}_t^i\left[\eta_{t+k}\right] \tag{18}$$

where

$$\mathbb{E}_t^i\left[X_{t+k}\right] = \sum_{s-0}^{\infty}\left(\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]$$

Suppose we have access to survey data on expectations from these forecasters at period t, and consider the regression

$$\mathbb{E}_t^i\left[y_{t+k}\right] = \gamma_{1t} + \gamma_{2t}\mathbb{E}_t^i\left[\pi_{t+k}\right] + u_{t,k}^i \tag{19}$$

Our first result provides conditions under which we can directly recover reputation.

Proposition 3. *Suppose forecasters only disagree about future cost push shocks, so* $\mathbb{E}_t^i\left[\eta_{t+k}\right] = \mathbb{E}_t^P\left[\eta_{t+k}\right]$. *Then*

$$\gamma_{2t} = -\frac{\mathbb{E}_t^P\left[\psi^y\right]}{\mathbb{E}_t^P\left[\psi^\pi\right]} = -\kappa^{-1}\left(\frac{1}{\mathbb{E}_t^P\left[\psi^\pi\right]} - 1\right) \tag{20}$$

Moreover, a more negative value of γ_{2t} *corresponds to an improvement in the central bank's reputation.*

Proof. See Appendix D ■

Proposition 3 establishes how our theory allows us to retrieve reputation from the data. When forecasters differ only in their expectations about cost-push shocks, we can recover reputation from a simple cross-sectional regression of output-gap forecasts on inflation forecasts. The intuition is straightforward. Consider the optimal policy problem of a myopic central bank. Its first-order condition is

$$y_t = -\kappa\lambda\pi_t$$

A larger λ implies a steeper slope in the regression of the output gap on inflation. Proposition 3 extends this idea to the case where the private sector is uncertain about λ: beliefs over λ induce a correlation structure between forecasts of the output gap and forecasts of inflation, which is informative about the central bank's reputation.

An improvement in reputation, captured by a decline in $\mathbb{E}_t^P[\psi^\pi]$, corresponds to a more negative value of $\gamma_{2,t}$. Furthermore, Proposition 3 is sharp: under those assumptions, $\gamma_{2,t}$ will *only* change in response to changes in reputation. In practice, these assumptions may not hold, resulting in time variation of our estimate independent of changes in reputation. In Section 5.3 we characterize the possible sources of time-variation in $\gamma_{2,t}$ unrelated to reputation and assess their relevance for our tests.

5.1 Data and Estimation

Our primary data source is the Blue Chip Financial Forecasts (BCFF) survey, which provides individual forecasts of interest rates and macroeconomic variables for the U.S. economy. This survey has been used in the literature to measure the private sector's expectations, and more recently, their perceptions about monetary policy (Bauer et al. 2024). Each month, forecasters report projections for their current quarter and four to five quarters ahead. We use the sample from January 1992 to December 2024.[15]

The forecasts of output growth and CPI inflation are reported in quarter-over-quarter annualized terms. Following Bauer et al. (2024), we transform them into year-over-year inflation and output gap forecasts. For inflation, we combine realized CPI with the sur-

[15]Data are available prior to 1992, but earlier surveys report GNP rather than GDP forecasts, so we restrict the sample to avoid potential inconsistencies.

vey responses to build forecasts of year-over-year inflation. For output, we construct GDP forecasts by cumulating quarterly growth projections, using the contemporaneous Archival Federal Reserve Economic Data (ALFRED) vintage for the level of real GDP.[16] Potential output is taken from Congressional Budget Office (CBO) projections, also retrieved in real time from ALFRED. Finally, we calculate the output gap as

$$\mathbb{E}_t^i\left[y_{t+k}\right] = 100 \times \frac{\mathbb{E}_t^i\left[Y_{t+k}\right] - \mathbb{E}_t\left[Y_{t+k}^n\right]}{\mathbb{E}_t\left[Y_{t+k}^n\right]}$$

where Y_{t+k} is real output, and Y_{t+k}^npotential output at horizon $t+k$.

We estimate (19) month-by-month using data from January 1992 to December 2024, including forecaster fixed effects to control for forecaster heterogeneity. Figure 8 plots the time series of $\gamma_{2,t}$. An Augmented Dickey-Fuller test on γ_{2t} rejects the null of a unit root at the 1% level. The absence of a unit root is consistent with our learning model. Reputation fluctuates around a long-run mean, deteriorating when it is above (more hawkish) and improving when it is below (more dovish). This property holds whether the central bank is myopic or follows the optimal policy.

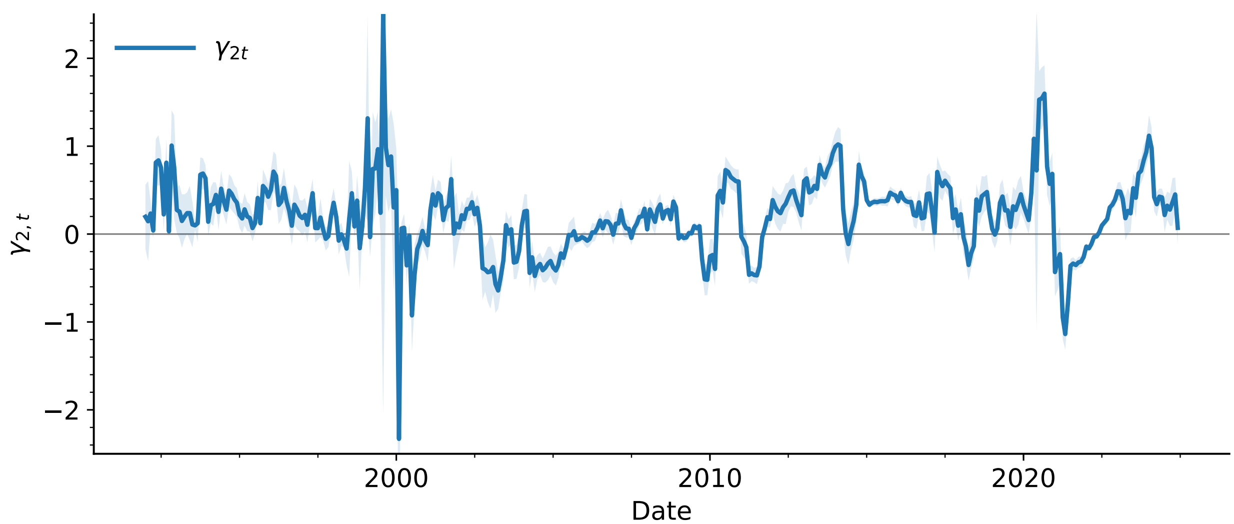


Figure 8: Monthly estimates for $\gamma_{2,t}$

Our measure of perceived hawkishness captures a dimension of reputation distinct from long-run inflation expectations: the correlation between $\gamma_{2,t}$ and the 5-Year, 5-Year forward

[16]ALFRED provides the vintages available to forecasters at each survey date. If the exact date is missing, we use the closest vintage. We assume the survey took place on the first day of each month.

inflation expectation rate from the SPF is only 0.05. That said, the two need not be unrelated: Bocola et al. (2025) find that reputation also affects long-run inflation expectations in Brazil, a setting with greater variation in those expectations.

As a robustness check, we repeat the exercise using quarterly individual forecasts from the Survey of Professional Forecasters (SPF). While BCFF primarily surveys financial market participants and business economists, SPF targets a broader set of professional forecasters, including academics, government agencies, and private sector analysts. Using the same methodology, the correlation coefficient between both estimates of $\gamma_{2,t}$ is 0.60, indicating that both series capture similar underlying variation in reputation. In Appendix D, we replicate the empirical analysis below using SPF data and find qualitatively similar results. We now turn to testing the model's two predictions.

5.2 Testing the Model's Predictions

PREDICTION #1: REPUTATION ANCHORS INFLATION EXPECTATIONS In our model, the payoff of a good reputation is that the pass-through of shocks on inflation expectations is lower. Suppose the cost-push shock follows an AR(1) process, and forecasters disagree on its persistence ρ_i. From (18), the pass-through of a cost push shock of size ε on inflation expectations is given by

$$\mathbb{E}_t^i \left[\pi_{t+k} \middle| \varepsilon_t = \varepsilon \right] - \mathbb{E}_t^i \left[\pi_{t+k} \middle| \varepsilon_t = 0 \right] = \frac{\mathbb{E}_t^P \left[\psi^\pi \right]}{1 - \beta \mathbb{E}_t^P \left[\psi^\pi \right] \rho_i} \rho_i^k$$

Therefore, a better reputation reduces the pass-through. We test this using oil price news shocks from Känzig (2021) and the nonlinear local projection

$$\mathbb{E}_t^i \left[\pi_{t+k} \right] = \beta_0^h + \beta_1^h \varepsilon_t + \beta^2 f\left(\hat{\gamma}_{2,t-1} \right) \varepsilon_t + u_{i,t,h}$$

where $f(\cdot)$ is a positive function that captures how the impulse response may depend on reputation, and $\hat{\gamma}_{2,t-1}$ is our empirical estimate.[17] Oil price news shocks are a natural test: they are cost-push shocks that raise inflation and, because they are anticipated, they affect inflation expectations directly. Under the null, $\beta_2^h = 0$ and reputation does not affect pass-

[17] We use the lag to avoid simultaneity.

through; under the alternative, $\beta_2^h \neq 0$.[18]

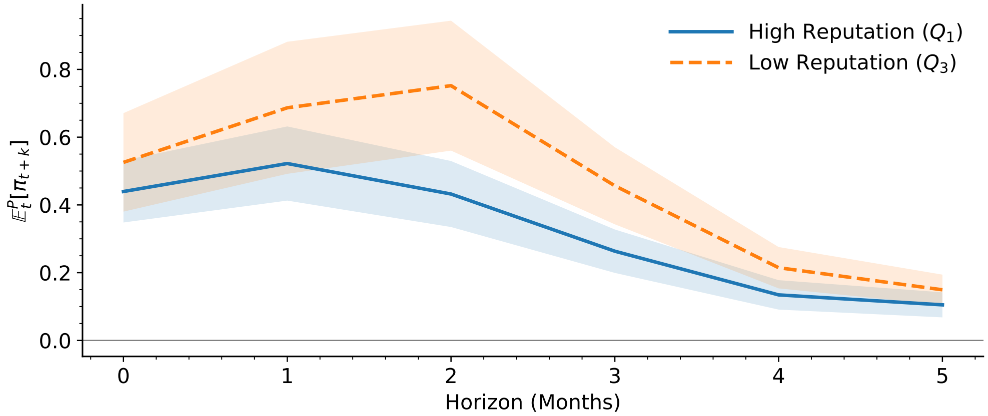


Figure 9: Impulse response of inflation expectations to an oil price news shock for high and low reputation.
Shaded areas denote 95% confidence intervals.

Figure 22 plots the impulse response of inflation expectations to an oil price news shock when reputation is high ($\gamma_{2,t}$ at its first quartile) and when it is low ($\gamma_{2,t}$ at its third quartile), using $f(\hat{\gamma}_{2,t-1}) = exp(-\hat{\gamma}_{2,t-1})$. Pass-through is larger when reputation is dovish and smaller when it is hawkish, consistent with the model's first prediction. The difference vanishes at longer horizons, which may reflect that oil price news shocks are short-lived, or that reputation matters more for short-run than long-run expectations.

Prediction #2: Reputation Responds to Monetary Policy Surprises Our second testable prediction is that reputation responds to unexpected changes in the interest rate. Define $mps_t := i_t - \mathbb{E}_t^P[i_t]$ as the monetary policy surprise. From the private sector's learning model, the impulse response of reputation to a monetary policy surprise is given by[19]

$$\mathbb{E}\left[\overline{\psi}_{t+1}\middle| mps_t = \varepsilon^m\right] - \mathbb{E}\left[\overline{\psi}_{t+1}\middle| mps_t = 0\right] = -\mathbb{E}\left[\omega_t X_t^{-1}\right]\frac{\kappa}{\sigma}\varepsilon^m$$

Crucially, the impact of a monetary policy surprise on reputation depends on the term $\mathbb{E}\left[\omega_t X_t^{-1}\right]$. Intuitively, this term captures the distribution of the shocks. When there are positive shocks, a positive monetary policy surprise improves reputation. However, when

[18]An alternative approach, following Ramey and Zubairy (2018), would be to split the sample into high- and low-reputation regimes. We avoid this since our estimate $\gamma_{2,t}$ can vary for reasons unrelated to reputation, which would introduce misclassification error into the regime assignment.

[19]See Appendix B for details.

there are negative shocks, the opposite occurs. Thus, this term summarizes what occurs on average. If shocks are mean zero and symmetric, then the average effect of a monetary policy surprise is zero.

Using high-frequency shocks from Gertler and Karadi (2015), we estimate

$$\gamma_{2,t+h} = \alpha_h + \beta_h \varepsilon_t^m + \delta_{1,h}\gamma_{1,t-1} + \delta_{2,h}\gamma_{2,t-1} + u_{t,h} \tag{21}$$

where ε_t^m is the FF4 monetary policy surprise from Gertler and Karadi (2015). An unexpected tightening improves reputation: Figure 10 shows that $\gamma_{2,t}$ declines following a monetary policy surprise.

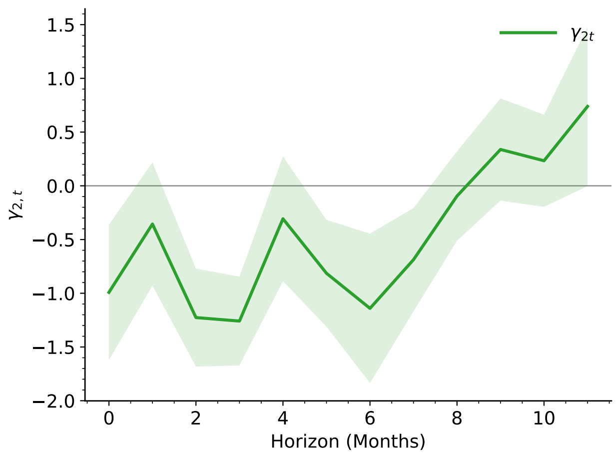


Figure 10: Impulse response to a monetary tightening.
Shaded areas denote 68% confidence intervals.

Our estimates are consistent with $\mathbb{E}\left[\omega_t X_t^{-1}\right] > 0$, suggesting that cost-push shocks are either positive on average or positively skewed.[20] Finally, we show in Appendix D that this finding is robust to controlling for the "Fed Information Effect" (Nakamura and Steinsson 2018, Bauer and Swanson 2022, Bauer and Swanson 2023). In particular, controlling for information in between FOMC announcements does not change our conclusions.

[20]Using (13) and $\mathbb{E}[X_t] = 0$, we have

$$\mathbb{E}\left[\omega_t X_t^{-1}\right] = -\left(\frac{\kappa^{-2}\tau_\eta}{\tau_t}\right)^2 \mathbb{E}\left[X_t^3\right] + \mathbb{E}\left[\mathcal{O}\left(X_t^5\right)\right]$$

If cost-push shocks are positively skewed ($\mathbb{E}\left[X_t^3\right] > 0$), then $\mathbb{E}\left[\omega_t X_t^{-1}\right] > 0$.

5.3 Alternative Sources of Time Variation in $\gamma_{2,t}$

In this section, we discuss how departures from our identifying assumptions affect the interpretation of $\gamma_{2,t}$ and the validity of our empirical tests.

STRUCTURAL BREAKS Our identification assumes a stable slope of the NKPC, κ. Changes in κ affect our estimate through two channels. First, holding beliefs constant, a change in κ directly alters reputation since $\mathbb{E}_t^P[\psi^\pi] = \mathbb{E}_t^P\left[\frac{1}{1+\kappa^2\lambda}\right]$. Second, even holding reputation fixed, $\gamma_{2,t} = -\kappa^{-1}\left(\frac{1}{\mathbb{E}_t^P[\psi^\pi]} - 1\right)$ changes mechanically with κ. For both reasons, estimates across periods with different slopes are not directly comparable. There is broad consensus that the slope of the NKPC changed during and after COVID, so we restrict our sample to the pre-2020 period.

DEMAND SHOCKS Our identification strategy assumes the central bank fully stabilizes (its estimate of) demand shocks. Suppose the first assumption does not hold because the central bank is gradualist and has a triple mandate as in (16). Then, cross-sectional disagreement about demand shocks will induce a positive correlation structure between forecasts of the output gap and forecasts of inflation. In this case $\gamma_{2,t}$ can be positive, and vary because of time-varying cross-sectional disagreement about demand and cost-push shocks. However, as we show in Appendix D, better reputation leads to a smaller value of $\gamma_{2,t}$ in this extended setting.

When we allow for demand shocks to have real effects, a new dimension of reputation appears: reputation about the central bank's gradualism. Therefore, reputation about reaction becomes a two-dimensional object: for cost-push shocks, reputation for hawkishness; for demand shocks, for gradualism. We cannot disentangle both dimensions in our dataset, so our estimate $\gamma_{2,t}$ reflects a combination of both.[21]

While time-varying disagreement is a potential source of variation in $\gamma_{2,t}$, our empirical tests suggest that the reputation channel is still present in the data. Suppose our estimate reflected only time-varying cross-sectional disagreement, with no information about reputation. For our first prediction, our model implies that disagreement does not affect the pass-through of cost-push shocks. Thus, we should expect $\beta_2^h = 0$ for all horizons. We find the opposite. For our second prediction, if monetary policy surprises moved $\gamma_{2,t}$ only by

[21] Separating them would require individual forecasts of demand and cost-push shocks separately, which are not currently available in our data.

changing disagreement, this would be analogous to an information effect: monetary policy surprises would be revealing information about the state of the economy rather than updating beliefs about the central bank's type. Controlling for information revealed between FOMC meetings leaves our results unchanged, arguing against this interpretation.

Belief Heterogeneity We initially assumed the private sector agreed on the central bank's reputation. Suppose instead that beliefs over λ differ across agents. As we show in Appendix D, our estimate $\gamma_{2,t}$ then reflects both the average belief about reputation and the degree of disagreement about λ. As a result, $\gamma_{2,t}$ can vary even if reputation is unchanged. For instance, lower disagreement about the central bank's preferences would reduce $\gamma_{2,t}$ independently of any aggregate change in reputation.

Our model features a representative agent and is therefore not well-suited to directly model belief heterogeneity about the central bank. However, we can offer the following interpretation. Suppose forecasters share a common prior over λ but draw an individual signal when forming their forecasts, reverting to the common prior once all information is publicly revealed. Under this interpretation, cross-sectional disagreement about the central bank's preferences maps to the precision of beliefs over λ. As we show in Appendix A, higher precision corresponds to better reputation. Therefore, time-varying disagreement about λ can itself be interpreted as variation in reputation, rather than noise.

Model Heterogeneity When forecasters use heterogeneous models to forecast, derived beliefs over $\psi^{\pi} = \frac{1}{1+\kappa^2\lambda}$ will be heterogeneous even if they shared a common prior. In particular, each forecaster's estimate of reputation depends on their own value of the NKPC slope, $\mathbb{E}_t^i[\psi_i^{\pi}] = \mathbb{E}_t^i\left[\frac{1}{1+\kappa_i^2\lambda}\right]$. Model heterogeneity is therefore a special case of belief heterogeneity, and our estimate $\gamma_{2,t}$ will reflect both average reputation and the degree of model disagreement across forecasters. Time-varying changes in the models used by forecasters can generate spurious variation in $\gamma_{2,t}$. However, the key comparative static is preserved: a shift in beliefs towards a more hawkish central bank will unambiguously decrease the estimate of $\gamma_{2,t}$.

Optimal Policy A crucial assumption for interpretability is that the private sector believes the central bank is myopic. In Appendix D, we relax this assumption and show that when they believe the bank follows the optimal policy, we can no longer summarize the entire distribution of expectations with a single statistic. Instead, our measure $\gamma_{2,t}$ becomes a function of the full belief distribution, and a shift toward more hawkish expectations generally leads to a more negative $\gamma_{2,t}$, though the magnitude depends on the specific shape of the belief distribution.

5.4 Extensions

We now consider a number of extensions. We examine the robustness of our results to alternative identification assumptions, and relate our measure to existing estimates of private sector perceptions about monetary policy.

COMMITMENT In the baseline model we assumed the private sector believes the central bank optimizes period-by-period under discretion. Suppose instead that they believe the central bank operates under commitment, designing optimal policy from the timeless perspective. The first-order condition is, then

$$\tilde{y}_t - y_{t-1} + \kappa\lambda\tilde{\pi}_t = 0 \tag{22}$$

Lemma 3. *Under the timeless perspective, the k-periods ahead forecast is*

$$\begin{aligned}
\mathbb{E}_t^i\left[y_{t+k}\right] =& \mathbb{E}_t^i\left[\varphi^y y_{t+k-1}\right] - \alpha\mathbb{E}_t^P\left[\psi^y\right]\sum_{s=0}^{\infty}\left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right] \\
\mathbb{E}_t^i\left[\pi_{t+k}\right] =& \mathbb{E}_t^i\left[\varphi^\pi y_{t+k-1}\right] + \alpha\mathbb{E}_t^P\left[\psi^\pi\right]\sum_{s=0}^{\infty}\left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]
\end{aligned}$$

where $\alpha < 1$, and φ^y and φ^π are positive constants.

Proof. See Appendix D ■

Lemma 3 highlights two differences relative to the myopic benchmark. First, shocks have a smaller direct impact ($\alpha < 1$): under commitment, the central bank smooths policy to reduce volatility. Second, the lagged output gap becomes a new state variable, reflecting the central bank honoring its past promises.

Empirically, this implies we can recover reputation under commitment by estimating

$$\mathbb{E}_t^i\left[y_{t+k}\right] = \gamma_{1,t} + \gamma_{2,t}\mathbb{E}_t^i\left[\pi_{t+k}\right] + \gamma_{3,t}\mathbb{E}_t^i\left[y_{t+k-1}\right] + u_{i,t,k}$$

By the Frisch-Waugh-Lovell Theorem, $\gamma_{2,t}$ coincides with the estimand in (19), so Proposition 3 continues to hold. In practice, the correlation coefficient between the commitment-based and baseline estimates is 0.84, confirming the robustness of our measure to the assumed policy behavior.

INERTIA Suppose that past inflation feeds into current inflation, so that the NKPC exhibits inertia. In this case, the New Keynesian Phillips Curve (2) becomes

$$\tilde{\pi}_t = \kappa \tilde{y}_t + \delta \pi_{t-1} + \beta \mathbb{E}_t^P \left[\tilde{\pi}_{t+1}\right] + \varepsilon_t \quad \delta + \beta < 1.$$[22]

Lemma 4. *Under inertia, the k-periods ahead forecast is*

$$\mathbb{E}_t^i \left[y_{t+k}\right] = - \mathbb{E}_t^i \left[\varphi^y \pi_{t+k-1}\right] - \alpha \mathbb{E}_t^P \left[\psi^y\right] \sum_{s=0}^{\infty} \left(\alpha \beta \mathbb{E}_t^P \left[\psi^\pi\right]\right)^s \mathbb{E}_t^i \left[\varepsilon_{t+k+s}\right]$$
$$\mathbb{E}_t^i \left[\pi_{t+k}\right] = \mathbb{E}_t^i \left[\varphi^\pi \pi_{t+k-1}\right] + \alpha \mathbb{E}_t^P \left[\psi^\pi\right] \sum_{s=0}^{\infty} \left(\alpha \beta \mathbb{E}_t^P \left[\psi^\pi\right]\right)^s \mathbb{E}_t^i \left[\varepsilon_{t+k+s}\right]$$

where $\alpha > 1$, and φ^y and φ^π are positive constants.

Proof. See Appendix D ■

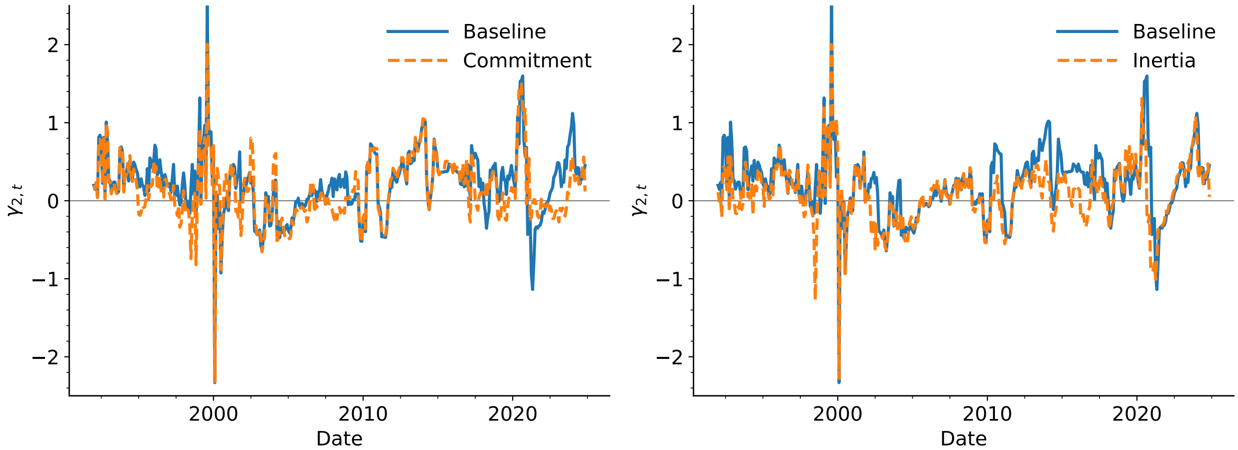
Lemma 4 highlights two differences relative to the myopic benchmark. First, shocks have a larger direct impact ($\alpha > 1$): inertia amplifies the impact of shocks on inflation. Second, the lagged inflation becomes a new state variable, reflecting that past inflation acts as a cost-push shock in the presence of inertia.

Empirically, this implies we can recover reputation in the presence of inertia by estimating

$$\mathbb{E}_t^i \left[y_{t+k}\right] = \gamma_{1,t} + \gamma_{2,t} \mathbb{E}_t^i \left[\pi_{t+k}\right] + \gamma_{3,t} \mathbb{E}_t^i \left[\pi_{t+k-1}\right] + u_{i,t,k}$$

By the Frisch-Waugh-Lovell theorem, the coefficient γ_{2t} coincides with the estimand in (19), so Proposition 3 continues to hold. In practice, the correlation coefficient between the inertia and baseline estimates is 0.82, confirming the robustness of our measure to the presence of inertia.

[22] In conventional macro models with inertia, the sum of the coefficients of the NKPC is weakly smaller than one. See Werning (2022).



(a) Commitment

(b) Inertia

Figure 11: Monthly estimates for $\gamma_{2,t}$ under Commitment and Inertia

Zero Lower Bound A further concern is that the private sector's model abstracts from the Zero Lower Bound (ZLB). When the ZLB binds, Lemma 1 no longer applies, and forecasts are not given by (8) and (9). To address this issue, we re-estimate our baseline specification (19) after excluding forecasts that assume the policy rate will remain at the ZLB—that is, those for which $\mathbb{E}_t^i[i_{t+k}] = 0$— during the 2009-2015 period, when the constraint was binding.[23] Figure 12 compares the two series, which display a correlation coefficient of 0.78, confirming robustness with respect to the ZLB. What matters for our measure is whether forecasters *expect* to be at the ZLB, not whether the economy *currently* is. For this reason, both estimates coincide.

Perceived Taylor Rule Coefficients A growing body of literature estimates the private sector's perceived Taylor rule coefficients using survey data (Bauer et al. 2024) or bond market data (Bocola et al. 2024). How do these estimates relate to perceived central bank preferences? In Appendix D, we establish two results.

First, when demand shocks have real effects, there is a one-to-one mapping between the perceived Taylor rule coefficient on the output gap and reputation for gradualism: a central bank perceived as less gradualist will lead to a higher perceived coefficient on output. Second, no such mapping exists for the perceived coefficient on inflation. A central bank with a strong reputation and one with a weak reputation can share the same perceived

[23]In practice, we exclude observations with$\mathbb{E}_t^i[i_{t+k}] < 0.20$. Since policy rates typically move in increments of 0.25, values below 0.20 are best interpreted as ZLB.

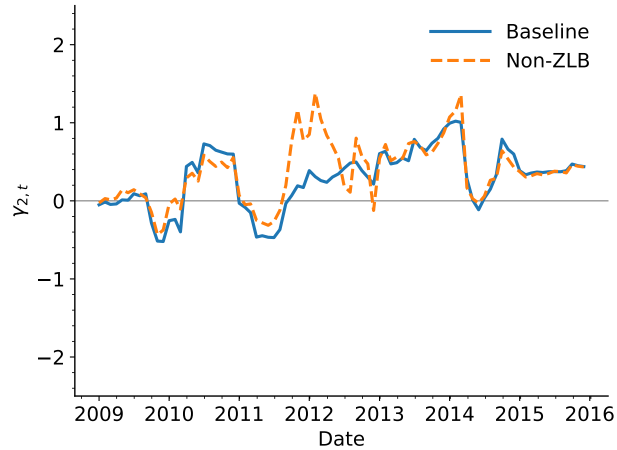


Figure 12: Monthly estimates for $\gamma_{2,t}$ at the ZLB

inflation coefficient. The reason is that a hawkish central bank is perceived as willing to move rates aggressively, but may not *need* to do so as much in response to the same shock. In practice, there is an inverse-U relationship between reputation and the perceived Taylor rule coefficient on inflation

6 Quantitative Analysis

Our theoretical analysis shows that reputation shapes the central bank's stabilization trade-off. We now bring the model to the data. We estimate the structural parameters by matching the myopic model to U.S. macroeconomic moments from two distinct macroeconomic episodes, the Great Moderation (1992-2019) and the Great Inflation (1973-1982), and then solve for the optimal policy under each estimated parameterization. Comparing myopic and optimal outcomes across both episodes provides a disciplined assessment of how the value of reputation management varies with the macroeconomic environment.

6.1 Data and Targeted Moments

We use four quarterly time series for the United States obtained from FRED: the Consumer Price Index (CPI), the effective federal funds rate (FFR), real GDP, and the Congressional Budget Office estimate of real potential GDP. The CPI is reported at monthly frequency and the FFR at daily frequency; we aggregate both to quarterly averages.[24] From these series we construct three observable variables. The output gap is

$$y_t^{\text{data}} = 100 \times \log\left(\frac{Y_t}{Y_t^n}\right),$$

where Y_t is real GDP and Y_t^n is real potential GDP. Annualized CPI inflation is

$$\pi_t^{\text{data}} = 400 \times \left[\log(\text{CPI}_t) - \log(\text{CPI}_{t-1})\right],$$

and the annualized federal funds rate is taken directly from the data.

Two estimation samples. We estimate the model on two subsamples chosen to bracket the range of environments in which reputation may matter.

The first sample covers 1992:Q1–2019:Q4, the Great Moderation. Inflation expectations were well anchored, output volatility declined markedly, and the federal funds rate operated away from the zero lower bound for most of the period. The low and stable inflation is consistent with a central bank that has accumulated substantial reputation for price stability, while the residual fluctuations discipline the structural shock processes. Estimating on this sample tests whether the model can rationalize the joint dynamics of a "quiet" economy where credibility is high and the stabilization trade-off is comparatively mild.

The second sample covers 1973:Q1–1982:Q4, encompassing the OPEC oil embargo, the Iranian revolution, and the onset of the Volcker disinflation. This episode is dominated by large, persistent cost-push shocks that generated severe stagflationary pressure, precisely the type of disturbance at the center of our model. The period also witnessed a dramatic shift in the Federal Reserve's perceived commitment to price stability, from the accommodative stance of the Burns era to the aggressive tightening under Volcker. Estimating on this sample tests whether the reputation channel can quantitatively account for the joint behavior of output, inflation, and interest rates when supply shocks and evolving credibility were first-order macroeconomic forces.

To ensure comparability across samples, all second moments (standard deviations, autocorrelations, and cross-correlations) are computed by demeaning around fixed long-run constants from the full 1960-2024 sample, rather than sample-specific means.[25]

[24]Results are robust to using end-of-period CPI or end-of-period FFR instead.

[25]This fixed-point demeaning prevents shifts in sample means from contaminating the comparison of second

6.2 Model Specification

Under the myopic assumption, the central bank takes the private sector's expectations as given and optimizes period by period. A myopic central bank with preference parameter λ implements the output gap and inflation

$$\tilde{y}_t = -\psi^y X_t, \qquad \tilde{\pi}_t = \psi^\pi X_t, \tag{23}$$

where $\psi^y = \kappa\lambda/(1+\kappa^2\lambda)$, $\psi^\pi = 1/(1+\kappa^2\lambda)$, and $X_t = \varepsilon_t/\big(1 - \beta\rho_\varepsilon \mathbb{E}_t^P[\psi^\pi]\big)$ captures the present value of cost-push shocks scaled by the private sector's expected pass-through. The realized (observed) output gap and inflation incorporate the central bank's forecast error η_t: $y_t = \tilde{y}_t + \eta_t$ and $\pi_t = \tilde{\pi}_t + \kappa\,\eta_t$. The nominal interest rate follows from the dynamic IS equation (**??**):

$$i_t = r_t^n + \mathbb{E}_t^P[\pi_{t+1}] - \frac{1}{\sigma}\big(y_t - \mathbb{E}_t^P[y_{t+1}]\big), \tag{24}$$

where the natural rate $r_t^n \equiv \rho + \upsilon\,\mathbb{E}_t^P[\Delta a_{t+1}]$ depends on productivity and preference shocks. Beliefs evolve according to the Bayesian learning rule in Section 3. The prior location $\overline{\psi}_t$ is updated via equation (13), with a Kalman gain ω_t that depends on the signal-to-noise ratio.

The model features two exogenous forcing processes, cost-push shocks ε_t and productivity shocks a_t, each following an AR(1):

$$\varepsilon_t = \rho_\varepsilon\,\varepsilon_{t-1} + \nu_t^\varepsilon, \qquad a_t = \rho_a\,a_{t-1} + \nu_t^a, \tag{25}$$

with $\nu_t^\varepsilon \sim \mathcal{N}(0,\sigma^2_{\nu^\varepsilon})$, $\nu_t^a \sim \mathcal{N}(0,\sigma^2_{\nu^a})$, and the forecast error $\eta_t \sim \mathcal{N}(0,\sigma^2_\eta)$ drawn independently. To match data units, we scale the model's output gap by 100 and annualize inflation and the interest rate by a factor of 400.

Calibrated parameters. Table 6.1 reports the parameters fixed prior to estimation. The discount factor $\beta = 0.99$ and the NKPC slope $\kappa = 0.17$ are standard in the quarterly New Keynesian literature. The elasticity of intertemporal substitution is $\sigma = 1$ (log utility). Productivity shocks follow a standard RBC calibration ($\rho_a = 0.95$, $\sigma_{\nu^a} = 0.0065$).

We fix $\overline{\psi}_0 = 0.6$ because it is only weakly identified by the moments targeted in the baseline estimation exercise. Since these moments are primarily long-run objects, they con-

moments across episodes. Details on the bootstrap procedure used to estimate the sampling variability of these moments are provided in Section **??**.

tain limited information about initial conditions. The initial prior location is instead more relevant for transition dynamics, short historical episodes, and counterfactual exercises conditional on specific starting states.

Table 6.1: Calibrated Parameters

Parameter	Description	Value	Source
β	Discount factor	0.99	Standard
κ	NKPC slope	0.17	Galí (2003)
σ	Inverse EIS	1.00	Log utility
υ	Natural rate loading	1.00	Woodford (2003a)
ρ_a	Persistence, productivity	0.95	Standard RBC
σ_{ν^a}	Std. dev., productivity innovation	0.0065	Standard RBC
$\overline{\psi}_0$	Initial belief's location parameter	0.60	—

Estimated parameters and parsimony. For each sample period, we estimate the parameter vector $\theta = (\lambda, \rho_\varepsilon, \sigma_{\nu^\varepsilon}, \sigma_\eta, \tau)$ by Simulated Method of Moments (SMM), taking the parameters in Table 6.1 as fixed. For a candidate value of θ, we simulate the myopic model, compute the model-implied counterparts of the targeted moments, and choose the parameter vector that minimizes the weighted distance between simulated and empirical moments.

Identification relies on the distinct roles each parameter plays. The preference weight λ governs the inflation–output trade-off and primarily shapes the relative volatility $\sigma(\pi_t)/\sigma(y_t)$. The cost-push parameters ρ_ε and σ_{ν^ε} control the persistence and scale of all endogenous fluctuations. The forecast error σ_η introduces a wedge between intended and realized allocations since η_t enters the output gap additively but inflation only through $\kappa\eta_t$, it disproportionately raises output volatility. The belief parameters τ and $\overline{\psi}_0$ shape the learning dynamics: a lower τ implies a higher Kalman gain and faster updating; $\overline{\psi}_0$ pins down the initial state of reputation.

Our model is deliberately parsimonious. Three observable time series (output gap, inflation, and the interest rate) are driven by only two structural shocks (cost-push and productivity) plus a forecast error affecting the learning process. This parsimony is a feature, not a limitation: it forces the reputation mechanism to do the heavy lifting in rationalizing the joint dynamics. However, it also means we should not expect the model to match every moment of the data equally well.[26] The model's comparative advantage lies in matching

[26]Our model has no independent demand shock (i.e., the output gap and inflation are driven by the same

the *relative* magnitudes of volatilities and the comovement between inflation and interest rates—the moments most directly shaped by the reputation channel.

6.3 Estimation Results

Table 6.2 reports the estimated parameters for both samples side by side. Across both episodes, the estimated λ implies a central bank that places substantial weight on inflation stabilization. The two episodes yield economically distinct parameterizations, as expected. For the Great Moderation, cost-push persistence is higher but innovation volatility is lower, and the belief precision τ implies a tighter prior. This is consistent with well-anchored expectations. For the Great Inflation, the cost-push process is moderately persistent and volatile, the forecast error is non-negligible, and the initial prior location $\overline{\psi}_0$ places the private sector's beliefs away from the true pass-through. This is consistent with an era of uncertain and evolving credibility.

Table 6.2: Estimation Results Across Episodes: Parameters

Parameter	Description	Great Moderation	Great Inflation
Estimated parameters			
λ	CB weight on inflation	22.4735	12.8073
ρ_ε	Persistence, cost-push	0.9636	0.9616
σ_{ν^ε}	Std. dev., cost-push innov.	0.0009	0.0023
σ_η	Std. dev., forecast error	0.0001	0.0003
τ	Prior precision	9.7114	7.9320

Note: Great Moderation: 1992-Q1 to 2019-Q4 (112 obs.). Great Inflation: 1973-Q1 to 1982-Q4 (40 obs.). All second moments are computed by demeaning around fixed long-run means from the 1960–2024 sample.

The moment fit reflects the parsimony of the model (See Table 6.3). Volatilities are generally well matched in both episodes. Autocorrelations pose a greater challenge. Cross-correlations are the most demanding targets. The model generically produces a strong negative correlation between output and inflation which matches the Great Inflation data ($\text{Corr}(y_t, \pi_t) < 0$) but conflicts with the positive correlation observed during the Great

cost-push process) so it generically produces $\text{Corr}(y_t, \pi_t) < 0$. This is correct for supply-shock-dominated episodes (the Great Inflation) but not for demand-dominated ones (the Great Moderation, where demand-side forces dominated). Rather than introduce additional shocks that would dilute the reputation mechanism, we accept the model's inability to match certain cross-correlations and focus the estimation on the moments most informative about the parameters of interest.

Moderation, when demand-side forces dominated. This is a well-understood limitation of supply-shock-driven New Keynesian models and underscores why we focus on the $\text{Corr}(\pi_t, i_t)$ moment, which the model matches well in both episodes.

Table 6.3: Estimation Results Across Episodes: Model Fit

	Great Moderation		Great Inflation	
Moment	Data	Model	Data	Model
Volatilities				
$\sigma(y_t)$	1.76	1.68	2.99	2.94
$\sigma(\pi_t)$	2.26	1.76	5.72	5.40
$\sigma(i_t)$	2.95	4.41	6.33	6.73
Autocorrelations				
$\text{Corr}(y_t, y_{t-1})$	0.96	0.95	0.85	0.92
$\text{Corr}(\pi_t, \pi_{t-1})$	0.48	0.95	0.93	0.92
$\text{Corr}(i_t, i_{t-1})$	1.00	0.99	0.98	0.93
Cross-correlations				
$\text{Corr}(\pi_t, i_t)$	0.56	0.36	0.83	0.77
Objective $\frac{1}{2}\mathbf{g}'\mathbf{W}\mathbf{g}$	1.33		0.14	

Note: Great Moderation: 1992-Q1 to 2019-Q4 (112 obs.). Great Inflation: 1973-Q1 to 1982-Q4 (40 obs.). All second moments are computed by demeaning around fixed long-run means from the 1960–2024 sample.

6.4 Counterfactual Exercise: Optimal vs. Myopic Policy

Having estimated the structural parameters under the myopic assumption, we solve for the optimal policy (where the central bank internalizes the effect of its actions on private sector beliefs) using the *same* parameter values, and simulate both models under identical shock realizations. This comparison isolates the pure effect of reputation management: any difference in outcomes is attributable solely to the central bank's forward-looking behavior.

The optimal policy is computed via Howard's policy iteration algorithm on a discretized state space $(\overline{\psi}, \varepsilon)$, with belief precision τ held fixed at its estimated value.[27] The solution yields an optimal output gap policy $y^*(\overline{\psi}, \varepsilon)$; inflation then follows from the NKPC. For

[27] The state space uses 400 grid points for $\overline{\psi} \in [-1, 1.5\overline{\psi}_0]$ and 101 grid points for ε, with 31 quadrature nodes for the forecast error η. Convergence is achieved when the sup-norm of successive output gap iterates falls below 10^{-6}.

simulation, we interpolate the policy function and feed it into the same data-generating process used for the myopic model, ensuring that beliefs update endogenously along the equilibrium path.

Table 6.4 compares the simulated moments under myopic and optimal policies with the data, side by side for both episodes. Table 6.5 reports the welfare comparison.

Table 6.4: Simulated Moments: Myopic vs. Optimal Policy Across Episodes

	Great Moderation			Great Inflation		
Moment	Data	Myopic	Optimal	Data	Myopic	Optimal
Volatilities						
$\sigma(y_t)$	1.76	1.68	2.14	2.99	2.93	4.31
$\sigma(\pi_t)$	2.26	1.76	0.76	5.72	5.39	2.90
$\sigma(i_t)$	2.95	4.39	6.20	6.33	6.71	12.95
Autocorrelations						
$\text{Corr}(y_t, y_{t-1})$	0.96	0.95	0.95	0.85	0.92	0.92
$\text{Corr}(\pi_t, \pi_{t-1})$	0.48	0.95	0.95	0.93	0.92	0.92
$\text{Corr}(i_t, i_{t-1})$	1.00	0.99	0.96	0.98	0.93	0.92
Cross-correlations						
$\text{Corr}(\pi_t, i_t)$	0.56	0.35	0.74	0.83	0.77	0.92
$\text{Corr}(y_t, \pi_t)$	0.36	−1.00	−0.99	−0.09	−1.00	−0.98
$\text{Corr}(y_t, i_t)$	0.70	−0.35	−0.75	−0.32	−0.77	−0.95
Means						
$E(y_t)$	−0.62	−0.02	−0.02	−1.26	−0.06	−0.08
$E(\pi_t)$	2.24	0.02	0.01	8.38	0.11	0.06
$E(i_t)$	2.62	4.04	4.07	9.68	4.13	4.26
$E(\hat{\psi}_t^\pi)$	—	0.56	0.37	—	0.61	0.45
Welfare						
$\bar{u}/(1-\beta)$	—	−0.04	−0.03	—	−0.19	−0.15
$\bar{u}_{\text{vol}}/(1-\beta)$	—	−0.04	−0.03	—	−0.19	−0.15

Note: Great Moderation: 1992-Q1 to 2019-Q4 (112 obs.). Model averages over 10,000 simulations of $T = 112$ periods. Great Inflation: 1973-Q1 to 1982-Q4 (40 obs.). Model averages over 10,000 simulations of $T = 40$ periods.

We evaluate welfare using the central bank's per-period loss $u(y_t, \pi_t) = -(y_t^2 + \lambda\pi_t^2)/2$. The tables report three complementary measures: the discounted-sum welfare $W^s = \sum_{t=0}^{T-1} \beta^t u_t$; the perpetuity-equivalent per-period welfare $\bar{u}/(1-\beta)$, which converts the average flow loss into the present value of a permanent stream; and a volatility-based measure $\bar{u}_{\text{vol}} =$

$-\frac{1}{2}\left(\sigma_y^2 + \lambda\,\sigma_\pi^2\right)$, which decomposes welfare into output and inflation components. Reputation management improves welfare if and only if the reduction in $\lambda\,\sigma_\pi^2$ exceeds the increase in σ_y^2.

Table 6.5: Welfare Comparison: Myopic vs. Optimal Policy Across Episodes

	Great Moderation			Great Inflation		
	Myopic	Optimal	Gain (%)	Myopic	Optimal	Gain (%)
Panel A: Discounted-sum welfare $W = \sum_t \beta^t u_t$						
Mean	−0.03	−0.02	25.09	−0.06	−0.05	20.22
Panel B: Per-period welfare $\bar{u}/(1-\beta)$						
$\bar{u}/(1-\beta)$	−0.04	−0.03	25.09	−0.19	−0.15	20.22
Panel C: Volatility-based welfare $\bar{u}_{\text{vol}} = -(\sigma_y^2 + \lambda\,\sigma_\pi^2)/2$						
$\bar{u}_{\text{vol}}/(1-\beta)$	−0.04	−0.03	25.09	−0.19	−0.15	20.22

Note: Per-period utility is $u(y_t, \pi_t) = -(y_t^2 + \lambda\pi_t^2)/2$. Great Moderation: $\beta = 0.99$, $\lambda = 22.5$. Great Inflation: $\beta = 0.99$, $\lambda = 12.8$. Panel A reports $W = \sum_{t=0}^{T-1} \beta^t u_t$. Panel B reports $\bar{u}/(1-\beta)$. Panel C computes $\bar{u}_{\text{vol}} = -(\sigma_y^2 + \lambda\,\sigma_\pi^2)/2$, rescaled by $1/(1-\beta)$. Gain is $100 \times (W_{\text{opt}} - W_{\text{myo}})/|W_{\text{myo}}|$.

In both episodes, the optimal central bank reduces inflation volatility relative to the myopic benchmark at the cost of higher output gap volatility—the hallmark of active reputation management. By leaning more aggressively against cost-push shocks, the optimal policy steers beliefs toward the true pass-through, compressing the present-value multiplier X_t and easing the stabilization trade-off in future periods.

Comparing the two episodes. The welfare gains from reputation management are present in both episodes but differ in magnitude, reflecting the distinct economic environments. During the Great Moderation, the longer sample (112 quarters) and highly persistent cost-push process allow the reputation channel to compound over time, yielding larger percentage gains. During the Great Inflation, the gains are modest in percentage terms: the signal extraction problem is severe and cost-push shocks are large, but the short sample (40 quarters) limits the cumulative benefit of improved beliefs.

The mechanism is the same in both cases. When the central bank responds more aggressively to a positive cost-push shock, tolerating a larger output loss, observed inflation falls below the private sector's expectation. This "inflation surprise" pulls the posterior location $\overline{\psi}_{t+1}$ toward a lower value, improving reputation. The improved reputation reduces expected pass-through, which compresses the present-value multiplier and flattens future trade-offs.

The optimal policy trades a *static* loss (larger output gap today) for a *dynamic* gain (better reputation and smaller future trade-offs).

The magnitude of the welfare gain depends on the estimated parameters in economically interpretable ways. A larger σ_η makes the signal extraction problem more severe, amplifying the scope for reputation management. A lower τ implies that beliefs are more responsive, increasing the return to forward-looking behavior. Higher cost-push persistence ρ_ε raises the present-value multiplier, making the dynamic gains from improved reputation more valuable.

7 New Problems, Same Old Solutions?

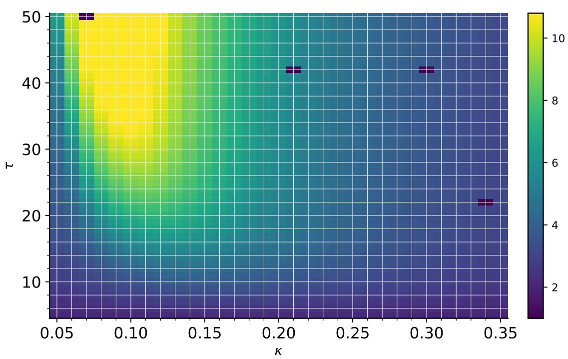
Implementing the optimal policy is not straightforward. It requires the central bank to track reputation in real time and correctly anticipate how its actions affect economic outcomes and influence the private sector's beliefs. This raises a natural question: can a simple rule approximate the optimal policy, and can it do so with greater robustness under misspecification? In his seminal contribution, Rogoff (1985) proposed appointing a more hawkish central banker to offset inflation bias. In this section, we ask whether a similar delegation can work in our setting. We show that delegating policy to a hawkish myopic central banker closely approximates the optimal policy.

We start with a simple observation: if the central bank had to choose among myopic central bankers, it would always prefer one with a higher preference for inflation stabilization. For any myopic central banker of type $\tilde{\lambda}$, the steady state welfare is

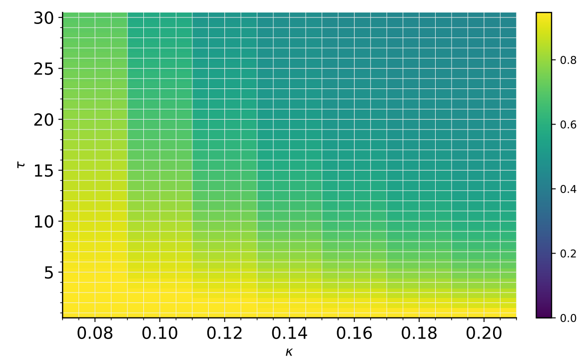
$$\begin{aligned}\mathcal{W}\left(\tilde{\lambda},\lambda\right) &= -\frac{1}{2}\mathbb{E}^{CB}\left[\tilde{y}\left(\tilde{\lambda}\right)^2 + \lambda\tilde{\pi}\left(\tilde{\lambda}\right)^2\right] \\ &= -\frac{1}{2}\overbrace{\left(\psi^y\left(\tilde{\lambda}\right)^2 + \lambda\psi^\pi\left(\tilde{\lambda}\right)^2\right)}^{\mathcal{W}_1\left(\tilde{\lambda},\lambda\right)} \times \overbrace{\frac{1}{\left(1-\beta\mathbb{E}^P\left[\psi^\pi \middle| \tilde{\lambda}\right]\rho\right)^2} \times \mathbb{E}^{CB}\left[\varepsilon^2\right]}^{\mathcal{W}_2\left(\tilde{\lambda}\right)} \qquad (26)\end{aligned}$$

Welfare combines two terms: a static trade-off between inflation and output gap stabilization, $\mathcal{W}_1\left(\tilde{\lambda},\lambda\right)$, and a dynamic term capturing how policy shapes reputation in the long run, $\mathcal{W}_2\left(\tilde{\lambda}\right)$. A myopic central bank optimizes only the first, whereas the optimal policy internalizes both. Since reputation improves with hawkishness we have $\left.\frac{d\mathcal{W}\left(\tilde{\lambda},\lambda\right)}{d\tilde{\lambda}}\right|_{\tilde{\lambda}=\lambda} > 0$. This result is robust and holds regardless of the specific learning model.

With this insight, we numerically evaluate the policy of delegating to a myopic central banker with relative preference $\tilde{\lambda}$ that generates the same long-run reputation as the optimal policy. We calculate this value across a wide grid of parameters of the slope of the NKPC, κ, and the precision of prior beliefs, τ.[28] To this end, we simulate the economy for 5000 periods, and burn the first 1000 to eliminate any dependence on initial conditions.



(a) $\frac{\tilde{\lambda}}{\lambda}$



(b) $\frac{\tilde{\psi}^{\pi}}{\psi^{\pi}}$

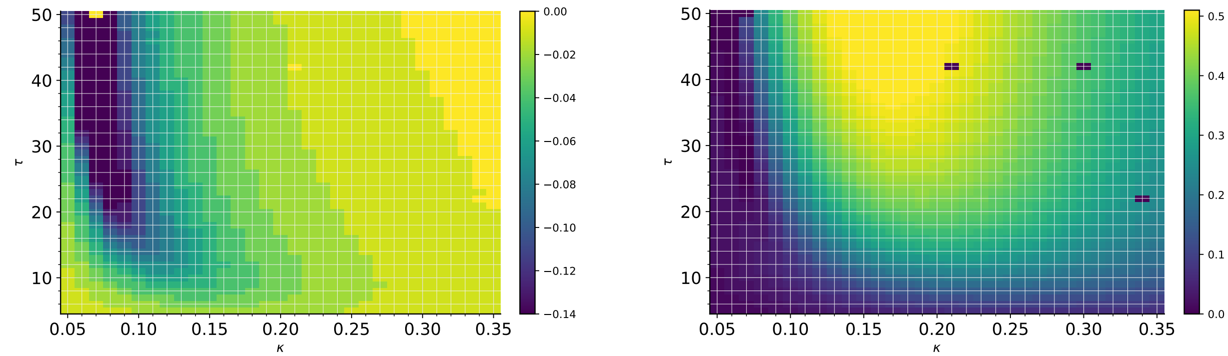
Figure 13: Heatmap across different values of τ and κ

Figure 13a displays the ratio of $\tilde{\lambda}$ to the true preference of the central bank, λ, which is fixed at $\lambda = 10$ in our calibration. The appointed central banker must always be more hawkish. It increases when κ decreases and τ increases, that is, along the negative-slope diagonal. Even though the value of $\tilde{\lambda}$ varies substantially with κ and τ, the its effect on $\psi^{\pi} = \frac{1}{1+\kappa^2\lambda}$: the movements in $\tilde{\lambda}$ are partially offset by movements in κ in the opposite direction. Everything else equal, an economy with a flatter NKPC will require a much more hawkish central bank to implement the same allocation for inflation. Figure 13b plots the ratio of $\tilde{\psi}^{\pi} = \frac{1}{1+\kappa^2\tilde{\lambda}}$ to the true preference of the central bank $\psi^{\pi} = \frac{1}{1+\kappa^2\lambda}$.

How well does this hawkish myopic policy perform? To answer this question, we simulate the optimal, myopic, and hawkish myopic policies and calculate welfare under the dual mandate. Figure 14 plots the relative welfare of the hawkish delegate and the optimal policy (Figure 14a) and the myopic policy (Figure 14b).

Consistent with our insight, the hawkish myopic delegate always outperforms the myopic policy for any pair of κ and τ. Moreover, it closely tracks welfare of the optimal policy. To understand why, Figure 15a illustrates the policy responses when reputation is at its stochas-

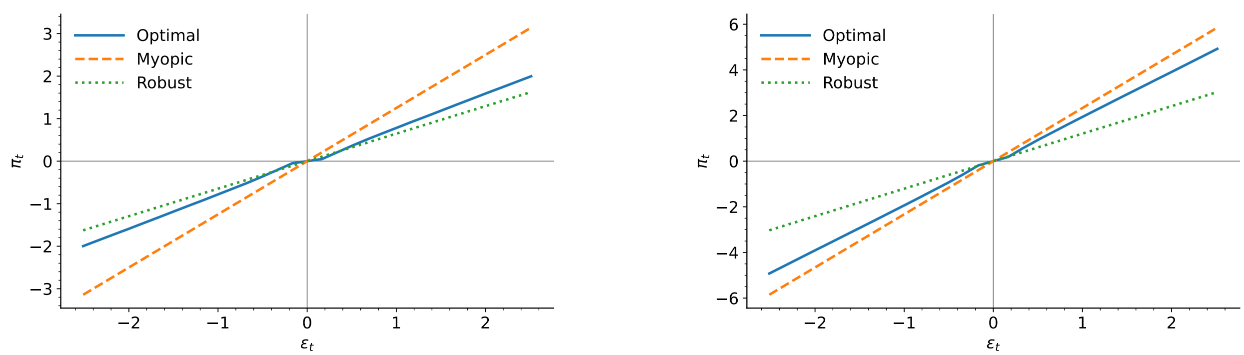
[28] To avoid beliefs collapsing to a single value, we consider a modified economy in which precision remains constant. We provide details on the calibration in Appendix F.



(a) $\frac{W^{Robust} - W^{Opt}}{|W^{Opt}|}$ **(b)** $\frac{W^{Robust} - W^{Myopic}}{|W^{Myopic}|}$

Figure 14: Welfare comparison of hawkish myopic policy

tic steady state: both the hawkish-myopic and optimal policies coincide at the deterministic steady state (where $X_t = 0$), but their slopes differ—they diverge in their responses to large shocks. Figure 15b shows that, when reputation deviates from its stochastic steady state, the two policies diverge further. Since the economy spends most of its time near the stochastic steady state of reputation and large shocks are infrequent, both policies yield similar welfare despite having different marginal responses.



(a) Steady-state Reputation **(b)** Reputation away from Steady State

Figure 15: Policy Functions: Optimal Policy vs Hawkish Myopic

To provide more insight, Table 7.1 summarizes the outcomes for $\kappa = 0.10$ and $\tau = 15$.[29] Consistent with the theory and quantitative exercise, and relative to the myopic policy, the optimal policy trades off larger volatility of the output gap for lower volatility of

[29] We show in Appendix F that these conclusions do not depend qualitatively on the particular choice of κ and τ.

inflation. Consequently, the optimal policy yields a more hawkish reputation and reduces the negative relationship between inflation and the output gap. This overreaction to shocks under the optimal policy pushes the covariance between the output gap and inflation to be more negative than under the myopic policy. The hawkish myopic policy delivers similar welfare, reflected by a relatively small welfare loss. However, the policy mix appears different: it prioritizes inflation stability relatively more, at the cost of larger volatility of the output gap. The reason is that, while both policies coincide at the deterministic steady state ($X_t = 0$), their derivatives differ—the hawkish-myopic rule responds more aggressively to shocks. This more aggressive response is why the covariance between the output gap and inflation is more negative under the hawkish myopic policy than under the optimal policy. However, since large shocks are infrequent, these differences in marginal responses do not translate into meaningful welfare losses.

Table 7.1: Performance of alternative policies

Policy	Welfare Loss	$\mathrm{Cov}(y_t, \pi_t)$	$\sigma_y/\sigma_y^{\mathrm{opt}}$	$\sigma_\pi/\sigma_\pi^{\mathrm{opt}}$	Average $\mathbb{E}_t^P[\psi^\pi]$
Myopic	−0.203	−1.459	0.473	1.473	0.757
Optimal	0.000	−1.955	1.000	1.000	0.589
Robust	−0.128	−2.050	1.348	0.689	0.589

Notes: Welfare Loss is defined as $\frac{W^{\mathrm{opt}} - W}{|W^{\mathrm{opt}}|}$ with W^{opt} the baseline Optimal welfare.

Our analysis demonstrates that delegating monetary policy to a hawkish myopic central banker provides a suitable approximation to the optimal policy. In addition, we show in Appendix F that it is also robust to situations where the central bank is not reacting enough to shocks. These findings suggest that Rogoff's (1985) insight extends beyond offsetting inflation bias. Even without time inconsistency, delegating policy to a conservative central banker helps manage reputation and provides insurance against model uncertainty. Simple institutional solutions may outperform sophisticated optimal policies.

8 Concluding Remarks

This paper studies the design of monetary policy when the private sector is uncertain about the central bank's relative preference for inflation stability. We develop a tractable framework in which this uncertainty—summarized by reputation—shapes both the propagation of shocks and the optimal policy response.

Our main theoretical results follow from two properties of reputation. First, a hawkish reputation reduces stabilization costs by dampening the pass-through of cost-push shocks to short-run inflation expectations. Second, reputation evolves endogenously: the private sector learns from the bank's actions, so policy choices shape future beliefs. Because of this, the optimal central bank overreacts to cost-push shocks relative to a myopic benchmark—it internalizes the value of its reputation. Furthermore, the bank manages reputation as an asset: it invests in it when perceived as dovish, and draws it down when perceived as hawkish.

Empirically, we construct a measure of reputation from cross-sectional variation in U.S. professional forecasts and find broad support for the theoretical mechanisms: a hawkish reputation dampens the pass-through of cost-push shocks to inflation expectations, and unexpected monetary tightenings improve it. Our measure captures a dimension of credibility distinct from long-run inflation expectations, the standard empirical proxy.

Quantitatively, the welfare gains from optimal policy are larger during the Great Moderation than the Great Inflation. Importantly, these gains can be closely approximated by delegating policy to a conservative but myopic central banker—a simple and robust institutional solution in the spirit of Rogoff (1985).

References

Aikman, D., F. Monti, and S. Zhang (2024): "In the Fed we Trust? Measuring Trust in Central Banking and its Effects on the Macroeconomy," Discussion Paper 19811, CEPR, Paris & London, cEPR Press. 1

Backus, D. and J. Driffill (1985): "Inflation and Reputation," *The American Economic Review*, 75, 530–538. 4, 13

Barnichon, R. and G. Mesters (2023): "A Sufficient Statistics Approach for Macro Policy," *American Economic Review*, 113, 2809–2845. 8

Barro, R. J. (1986): "Reputation in a Model of Monetary Policy with Incomplete Information," *Journal of Monetary Economics*, 17, 3–20. 4

Barro, R. J. and D. B. Gordon (1983): "Rules, discretion and reputation in a model of monetary policy," *Journal of Monetary Economics*, 12, 101–121. 1, 4

Bauer, M. D., C. E. Pflueger, and A. Sunderam (2024): "Perceptions About Monetary Policy," *The Quarterly Journal of Economics*, 139, 2227–2278. 1, 4, 25, 34, 80, 95

Bauer, M. D. and E. T. Swanson (2022): "A Reassessment of Monetary Policy Surprises and High-Frequency Identification," NBER Working Paper 29939, National Bureau of Economic Research. 29, 88

——— (2023): "An Alternative Explanation for the "Fed Information Effect"," *American Economic Review*, 113, 664–700. 29, 88

Bernanke, B. S. (2007): "Inflation Expectations and Inflation Forecasting," in *Speech at the Monetary Economics Workshop of the NBER Summer Institute*, "Anchored" means relatively insensitive to incoming data. 2

Blinder, A. S., M. Ehrmann, J. De Haan, and D.-J. Jansen (2024): "Central bank communication with the general public: Promise or false hope?" *Journal of Economic Literature*, 62, 425–457. 22

Blinder, A. S., M. Ehrmann, M. Fratzscher, J. De Haan, and D.-J. Jansen (2008): "Central bank communication and monetary policy: A survey of theory and evidence," *Journal of Economic Literature*, 46, 910–945. 22

Bocola, L., A. Dovis, K. Jørgensen, and R. Kirpalani (2025): "Monetary Policy without an Anchor," Working Paper 34436, National Bureau of Economic Research. 5, 13, 27, 76

Bocola, L., A. Dovis, K. Jørgensen, and R. Kirpalani (2024): "Bond Market Views of the Fed," Working Paper 32620, National Bureau of Economic Research. 1, 4, 34

Caballero, R. J. and A. Simsek (2022): "Monetary Policy with Opinionated Markets," *American Economic Review*, 112, 2353–92. 58

Canzoneri, M. B. (1985): "Monetary Policy Games and the Role of Private Information," *The American Economic Review*, 75, 1056–1070, publisher: American Economic Association. 4

Cukierman, A. and A. H. Meltzer (1986): "A Theory of Ambiguity, Credibility, and Inflation under Discretion and Asymmetric Information," *Econometrica*, 54, 1099–1128. 4

Eggertsson, G. B. and D. P. Kohn (2023): "The Inflation Surge of the 2020s: The Role of Monetary Policy," *Brookings Papers on Economic Activity*, Summer, 45–131. 23

Esponda, I. and D. Pouzo (2016): "Berk-Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models," *Econometrica*, 84, 1093–1130. 1

Eusepi, S. and B. Preston (2018): "The Science of Monetary Policy: An Imperfect Knowledge Perspective," *Journal of Economic Literature*, 56, 3–59. 9

Evans, G. W. and S. Honkapohja (2001): *Learning and Expectations in Macroeconomics*, Princeton University Press. 9, 56

Federal Reserve Board (2016): "Tealbook A: Economic and Financial Conditions: Current Situation and Outlook," Tech. rep., Board of Governors of the Federal Reserve System, report to the FOMC on Economic Conditions and Monetary Policy. 8

Galí, J. (2003): "New Perspectives on Monetary Policy, Inflation, and the Business Cycle," in *Advances in Economic Theory and Econometrics, Eighth World Congress*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, Cambridge, UK: Cambridge University Press, vol. III, 151–197. 97

García-Schmidt, M. and M. Woodford (2019): "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis," *American Economic Review*, 109, 86–120. 76

Gertler, M. and P. Karadi (2015): "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 7, 44–76. 29

Hamilton, J. D., S. Pruitt, and S. Borger (2011): "Estimating the Market-Perceived Monetary Policy Rule," *American Economic Journal: Macroeconomics*, 3, 1–28. 1, 4

Hicks, J. R. (1946): *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory*, Oxford: Clarendon Press, 2nd ed., originally published in 1939. 76

Känzig, D. R. (2021): "The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements," *American Economic Review*, 111, 1092–1125. 27, 89

King, R. G., Y. K. Lu, and E. S. Pastén (2008): "Managing Expectations," *Journal of Money, Credit and Banking*, 40, 1625–1666. 4

Kostadinov, R. and F. Roldán (2025): "Reputation and the credibility of inflation plans," *Journal of Economic Theory*, 229, 106062, elsevier. 4

Kreps, D. M. (1998): "Anticipated Utility and Dynamic Choice," in *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983–1997*, ed. by D. P. Jacobs, E. Kalai, M. I. Kamien, and N. L. Schwartz, Cambridge University Press, 242–274. 9

Krugman, P. R. (1998): "It's Baaack: Japan's Slump and the Return of the Liquidity Trap," *Brookings Papers on Economic Activity*, 29, 137–206, economic Studies Program. 79

Kydland, F. E. and E. C. Prescott (1977): "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85, 473–491. 1, 4

Lu, Y. K., R. G. King, and E. Pasten (2016): "Optimal reputation building in the New Keynesian model," *Journal of Monetary Economics*, 84, 233–249. 4

Marcet, A. and T. J. Sargent (1989): "Convergence of least squares learning mechanisms in self-referential linear stochastic models," *Journal of Economic Theory*, 48, 337–368. 9

Nakamura, E., V. Riblier, and J. Steinsson (2025): "Beyond the Taylor Rule," NBER Working Paper 34200, National Bureau of Economic Research. 4

Nakamura, E. and J. Steinsson (2018): "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect," *The Quarterly Journal of Economics*, 133, 1283–1330. 29

Powell, J. H. (2020): "New Economic Challenges and the Fed's Monetary Policy Review," in *Jackson Hole Economic Policy Symposium: Navigating the Decade Ahead: Implications for Monetary Policy*, Jackson Hole, Wyoming: Federal Reserve Bank of Kansas City. 23

Ramey, V. A. and S. Zubairy (2018): "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy*, 126, 850–901. 28

Rogoff, K. (1985): "The Optimal Degree of Commitment to an Intermediate Monetary Target," *The Quarterly Journal of Economics*, 100, 1169–1189. 1, 4, 5, 43, 47

Sargent, T. J. (1999): *The Conquest of American Inflation*, Princeton, NJ: Princeton University Press. 1

Sastry, K. (2022): “Disagreement About Monetary Policy,” *SSRN Electronic Journal*, available at SSRN: `https://ssrn.com/abstract=3421723` or `http://dx.doi.org/10.2139/ssrn.3421723`. 58

Svensson, L. E. O. (2003): “What Is Wrong with Taylor Rules? Using Judgment in Monetary Policy through Targeting Rules,” *Journal of Economic Literature*, 41, 426–477. 4

Vickers, J. (1986): “Signalling in a Model of Monetary Policy with Incomplete Information,” *Oxford Economic Papers*, 38, 443–455, publisher: Oxford University Press. 4

Werning, I. (2022): “Expectations and the Rate of Inflation,” Working Paper 30260, National Bureau of Economic Research. 33

Woodford, M. (2003a): *Interest and prices: foundations of a theory of monetary policy*, Princeton University Press. 7

——— (2003b): *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press. 56

——— (2003c): “Optimal Interest-Rate Smoothing,” *The Review of Economic Studies*, 70, 861–886. 22

Reassesing Central Bank Reputation: Long-Run Expectations

Online Appendix

Contents

A The Economy

A.1 Private Sector and Government

Households There is a continuum of identical households. In each period, the representative household derives utility from consuming a continuum of differentiated final goods, $c_t(j)$ for $j \in [0,1]$, and working N_t units. Assuming separability between consumption and labor and isoelastic functions, the lifetime utility is given by

$$\mathcal{U}_0 \equiv \mathbb{E}_0^P \left[\sum_{t=0}^{\infty} \beta^t Z_t \left(\frac{C_t^{1-\sigma} - 1}{1-\sigma} - \frac{N_t^{1+\varphi}}{1+\varphi} \right) \right]$$

where $C_t = \left[\int_0^1 c_t(j)^{1-\frac{1}{\epsilon_t}} dj \right]^{\frac{1}{1-\frac{1}{\epsilon_t}}}$ is the Dixit-Sitglitz consumption aggregator, ϵ_t is the elasticity of substitution across final goods varieties, and $\mathbb{E}_t^P[\cdot]$ is the private sector's expectation operator, which we assume satisfies the Law of Iterated Expectations.[30] In our analysis below, we allow the elasticity of substitution between varieties to be time-varying to allow for markup shocks that will simulate an inflationary episode.

Labor markets are competitive, and the representative household takes nominal wages W_t as given. Moreover, households can trade one-period nominal risk-free bonds, B_t, which the government issues. The representative household's budget constraint is given by

$$P_t C_t + B_t = W_t N_t + \Pi_t + (1 + i_{t-1}) B_{t-1} + T_t$$

where $P_t = \left[\int_0^1 P_t(j)^{1-\epsilon_t} dj \right]^{\frac{1}{1-\epsilon_t}}$ is the aggregate price index, i_{t-1} is the short-term nominal interest rate from period $t-1$ to t, Π_t denotes the nominal firms' profits, and T_t are lump sum transfers. Then, given B_{-1}, aggregate prices, government policy, and firms' profits, households choose a path for consumption, labor, and asset portfolio that maximizes their utility $\mathcal{U}_0$ subject to the budget constraint at every period. As a result of this optimization process, households' optimal behavior is captured by an aggregate Euler equation, consumption-labor optimal allocation and the budget constraint.

[30] We fully describe the household's information set and belief updating process later in this appendix.

Firms Firms and households share the same information and understanding of the functioning of the economy, so their expectation operator is also $\mathbb{E}_t^P[\cdot]$. There is monopolistic competition in the final goods market. Each producer of variety $j \in [0,1]$ has access to the same production technology

$$Y_t(j) = A_t N_t(j)^{1-\alpha}$$

where A_t is stochastic aggregate productivity and $\alpha \in (0,1)$ controls the degree of decreasing returns to scale in this economy. Firms set prices à la Calvo, with a probability $1-\theta \in [0,1]$ of changing prices every period. There is a production subsidy to offset the firms' market power so that the non-stochastic steady state is efficient.

Government Each period, the government issues short-term nominal bonds to finance lump-sum transfers and past government debt. We abstract from government expenditure. Then, the government's budget constraint is $(1+i_{t-1})B_{t-1} + T_t = B_t$. In addition, there is a monetary authority (Central Bank) that sets the path of nominal interest, $\{i_t\}_{t\geqslant 0}$.

Log-linear model Throughout the paper, we study the optimal monetary policy using a linear-quadratic approach. Thus, we log-linearize the model and obtain a version around its deterministic, efficient steady state.[31] Hence, up to first-order approximation, the aggregate Euler Equation and goods market clearing condition (i.e., $Y_t = C_t$) leads to the dynamic IS equation:

$$y_t = \mathbb{E}_t^P\left[y_{t+1}\right] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t^P\left[\pi_{t+1}\right] - r_t^n\right) \tag{A.1}$$

where y_t is the log deviation of the output concerning the efficient level of output (output gap), $\pi_t = P_t/P_{t-1}$ is the price inflation rate between $t-1$ and t, and

$$r_t^n \equiv \rho + \vartheta \mathbb{E}_t^P\left[\Delta a_{t+1}\right] - \mathbb{E}_t^P\left[\Delta z_{t+1}\right]$$

with $z_t \equiv \log Z_t$, $a_t \equiv \log A_t$, and ϑ being a function of the model's parameters. Similarly, up to first order, the solution to the firms' problem, together with households' labor supply,

[31] When prices are flexible, reputation has no bite in the economy, so the log-linear approximation does not rely on assuming a particular value for reputation in the long run.

implies a New-Keynesian Phillips Curve[32]:

$$\pi_t = \kappa y_t + \beta \mathbb{E}_t^P \left[\pi_{t+1}\right] + \varepsilon_t \tag{A.2}$$

where ε_t denotes the log deviation of the efficient output level (*frictionless* concept) with respect to the natural output level (*flexible-prices* concept), and κ is a function of the model's parameters. In our model, these disturbances arise from markup shocks, $\frac{\epsilon_t}{\epsilon_t - 1} \geqslant 1$, but, in general, our analysis does not change if we consider different sources of these cost-push shocks.

Second-order Approximation of Welfare We approximate welfare to second order following Woodford (2003b). A key observation is that incomplete information affects only the transmission of monetary policy — it has no bearing on the flexible-price equilibrium, and therefore does not affect the non-stochastic steady state. This allows us to proceed as in the textbook case. Since there is a production subsidy that offsets firms' market power, the steady state is efficient, and only second-order terms survive the approximation.

The relevant object is the evolution of price dispersion, $\Delta_t \equiv \int_0^1 \left(\frac{P_t(j)}{P_t}\right)^{-\epsilon} dj$, which follows

$$\Delta_t = \theta \Delta_{t-1} + (1-\theta) \left(\frac{P_t^*}{P_t}\right)^{-\epsilon}$$

Crucially, this law of motion is purely backward-looking and depends only on realized prices — it does not involve households' expectations or beliefs. As a result, the dynamics of price dispersion are independent of the private sector's information set and learning parameters. Following Woodford (2003b), the second-order approximation of welfare is then

$$\mathcal{W} \approx -\frac{1}{2} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[y_t^2 + \lambda^{Struct} \pi_t^2\right]$$

where λ^{Struct} is a function of the preference parameters of the household, and the average frequency of price adjustment. This expression is identical to the full-information case — learning does not alter the welfare criterion.

[32] This is a direct consequence of the expectation process satisfying the Law of Iterated Expectations. See Evans and Honkapohja (2001).

A.2 Central Bank and Monetary Policy Regime

The monetary authority has a dual mandate over output gaps and inflation, given by the following welfare loss function:

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t^{CB}\left[\sum_{k=0}^{\infty}\beta^k\left(y_{t+k}^2 + \lambda\pi_{t+k}^2\right)\right] \tag{A.3}$$

where $\mathbb{E}_t^{CB}[\cdot]$ is the Central Bank's rational expectations operator. This loss function captures the idea that the Central Bank wishes to stabilize inflation around zero and output around its efficient level. The importance of inflation stabilization, relative to output stabilization, is controlled by the fixed parameter $\lambda \geqslant 0$. Notice that λ is arbitrary and idiosyncratic to the central banker, and may or may not be equal to λ^{Struct}.

Central Bank's Information Set and Policy Regime Each period t, the Central Bank announces the nominal interest rate i_t before observing the realization of the demand shocks hitting the economy that period. Thus, the information set at t of the Central Bank is given by the history of output gaps and inflation, the cost-push shocks, and the private sector's behavioral equations: $\mathcal{I}_t^{CB} \equiv \{\mathcal{M}^{CB}, h_{t-1}, \mu_t, \boldsymbol{\varepsilon}_t\}$ where $\mathcal{M}^{CB}$ contains the structure of the economy, i.e., equations (A.1) and (2) plus the data generator process of the exogenous variables, $h_{t-1} \equiv \{y_s, \pi_s, i_s, \varepsilon_s, z_s, a_s\}_{s\leqslant t-1}$, μ_t encodes the private sector's beliefs at t, and $\boldsymbol{\varepsilon}_t = \{\mathbb{E}_t[\varepsilon_{t+s}]\}_{s\geqslant 0}$ the cost-push shocks.

Even though the Central Bank's model for the economic structure is correctly specified, it sets monetary policy at t with imperfect information on the realization of z_t, and a_t. Formally, given $\mathcal{I}_t^{CB}$, the Central Bank privately forecasts exogenous variable realizations, $\tilde{r}_t^n \equiv \mathbb{E}_t^{CB}[r_t^n]$, and chooses i_t to maximize (A.3) subject to the private sector's ex-ante equilibrium conditions, (A.1) and (A.2):

$$\tilde{y}_t = \mathbb{E}_t^{CB}\left[\mathbb{E}_t^P[y_{t+1}]\right] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t^{CB}\left[\mathbb{E}_t^P[\pi_{t+1}]\right] - \tilde{r}_t^n\right)$$

$$\tilde{\pi}_t = \kappa\tilde{y}_t + \beta\mathbb{E}_t^{CB}\left[\mathbb{E}_t^P[\pi_{t+1}]\right] + \varepsilon_t$$

where $\tilde{y}_t \equiv \mathbb{E}_t^{CB}[y_t]$ and $\tilde{\pi}_t \equiv \mathbb{E}_t^{CB}[\pi_t]$ are the Central Bank's ex-ante expected values of output and inflation, respectively.[33] Under the learning structure of our model (described

[33]The Central Bank's ex-ante expectations of the private sector's forward looking terms is present in the

below), it turns out that the terms with double expectations are simply the private sector's expectation of the Central Bank's choices:

$$\mathbb{E}_t^{CB}\left[\mathbb{E}_t^{P}\left[y_{t+1}\right]\right]=\mathbb{E}_t^{P}\left[\tilde{y}_{t+1}\right] \qquad \text{and} \qquad \mathbb{E}_t^{CB}\left[\mathbb{E}_t^{P}\left[\pi_{t+1}\right]\right]=\mathbb{E}_t^{P}\left[\tilde{\pi}_{t+1}\right].$$

Under this monetary policy regime, differences between the output gap and its Central Bank's ex-ante expected value, $\eta_t \equiv y_t - \tilde{y}_t$, may arise from forecasting errors about the realization of aggregate demand, productivity, or markup shocks. Note that after the realization of error η_t, the resulting inflation rate is $\pi_t = \tilde{\pi}_t + \kappa\eta_t$.

Re-writing the Central Bank's problem Given the quadratic objective function and the i.i.d. property of η_t with respect to the Central Bank's expectation operator, we can rewrite the Central Bank's objective function in terms of the ex-ante output and inflation (i.e., $\tilde{y}_{t+k}$ and $\tilde{\pi}_{t+k}$),

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t^{CB}\left[\sum_{k=0}^{\infty}\beta^k\left(\tilde{y}_{t+k}^2+\lambda\tilde{\pi}_{t+k}^2\right)\right]+t.i.p. \tag{A.4}$$

where $t.i.p.$ are terms independentt of policy. Finally, under the assumption that disturbances are properly bounded, zero lower bound is never binding so that we can re-express the Central Bank's problem as directly choosing the (ex-ante) allocation path $\{\tilde{y}_{t+k}, \tilde{\pi}_{t+k}\}_{k=0}^{\infty}$ to maximize (3) subject to

$$\tilde{\pi}_t = \kappa\tilde{y}_t + \beta\mathbb{E}_t^{P}\left[\tilde{\pi}_{t+1}\right]+\varepsilon_t \tag{A.5}$$

and the below-described public's learning process. It is worth emphasizing that if the Central Bank chooses to implement output $\tilde{y}_t$ and inflation $\tilde{\pi}_t$, the final equilibrium outcome at t are given by $\tilde{y}_t + \eta_t$ and $\tilde{\pi}_t + \kappa\eta_t$.

Although this Central Bank's problem, given by (3) and (A.5), seems to be a linear-quadratic one, the below-described expectation formation process introduces a nonlinearity into the Phillips Curve. In particular, $\mathbb{E}_t^P[\tilde{\pi}_{t+1}]$ depends on $\{\tilde{\pi}_s\}_{s\leqslant t}$ in a nonlinear way. The main objective of this paper is to characterize how this nonlinear dependence shapes the optimal monetary policy.

literature of monetary policy with disagreements. See, e.g. Caballero and Simsek (2022) and Sastry (2022).

A.3 Proof of Lemma 1

Proof. The first-order condition of a Central Bank with parameter $\lambda = \tilde{\lambda}$ is

$$\tilde{y}_t + \kappa\tilde{\lambda}\tilde{\pi}_t = 0$$

Taking expectations as given, this yields

$$\begin{aligned}\tilde{y}_t\left(\tilde{\lambda}\right) =& -\psi^y\left(\varepsilon_t + \beta\mathbb{E}_t^{CB}\left[\mathbb{E}_t^P\left[\pi_t\right]\right]\right)\\ \tilde{\pi}_t\left(\tilde{\lambda}\right) =& \psi^\pi\left(\varepsilon_t + \beta\mathbb{E}_t^{CB}\left[\mathbb{E}_t^P\left[\pi_t\right]\right]\right)\end{aligned}$$

We conjecture that the Central Bank follows a linear policy rule, and they are not aware of the private sector's perceived bias:

$$\begin{aligned}\tilde{y}_t\left(\tilde{\lambda}\right) =& -\psi^y\left(\tilde{\lambda}\right)\sum_{s=0}^{\infty}\Theta_s\mathbb{E}_t^P\left[\varepsilon_{t+s}\right]\\ \tilde{\pi}_t\left(\tilde{\lambda}\right) =& \psi^\pi\left(\tilde{\lambda}\right)\sum_{s=0}^{\infty}\Theta_s\mathbb{E}_t\left[\varepsilon_{t+s}\right]\end{aligned}$$

Matching coefficients for $s = 0$

$$\Theta_0 = 1$$

For $s = 1$, using that there is no anticipated learning

$$\Theta_1 = \beta\mathbb{E}_t^P\left[\psi^\pi\right]$$

For any arbitrary s, we have

$$\Theta_s = \beta\mathbb{E}_t^P\left[\psi^\pi\right]\Theta_{s-1} = \left(\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s$$

Putting everything together,

$$\begin{aligned}\tilde{y}_t\left(\tilde{\lambda}\right) =& -\psi^y\sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)\mathbb{E}_t\left[\varepsilon_{t+s}\right]\\ \tilde{\pi}_t\left(\tilde{\lambda}\right) =& \psi^\pi\sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)\mathbb{E}_t\left[\varepsilon_{t+s}\right]\end{aligned}$$

■

A.4 Comparative Statics

To formalize how beliefs about λ translate into reputation, we establish two results:

Lemma 5. *Suppose the private sector has beliefs μ_t over λ. Consider a different set of beliefs μ_t' that first-order stochastically dominates μ_t. Then, $\mathbb{E}_t^P[\psi^\pi]$ is lower under μ_t' than under μ_t.*

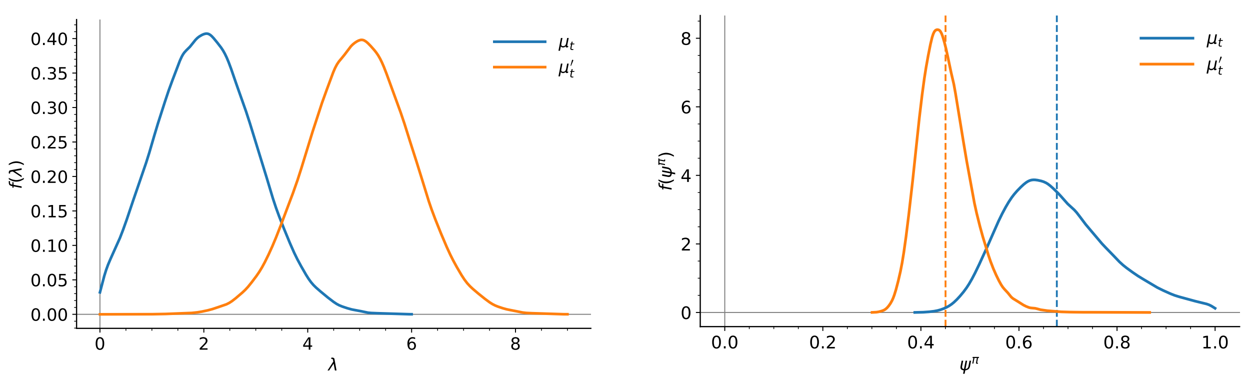
Proof. Rewrite the expected value as

$$\mathbb{E}_t^P[\psi^\pi] = \int_0^1 \psi^\pi f_\mu^{\psi^\pi}(\psi^\pi)\, d\psi^\pi = \int_0^1 \left(\int_0^{\psi^\pi} ds\right) f_\mu^{\psi^\pi}(\psi^\pi)\, d\psi^\pi$$

$$= \int_0^1 \left(\int_s^{\psi} f_\mu^{\psi^\pi}(\psi^\pi)\, d\psi^\pi\right) ds = \int_0^1 \left(1 - F_\mu^{\psi^\pi}(\psi^\pi)\right) d\psi^\pi$$

If μ_t' first-order stochastically dominates μ_t then $F_{\mu_t}^\lambda(\lambda) \geqslant F_{\mu_t'}^\lambda(\lambda)$. Since $\psi^\pi(\lambda)$ is a decreasing function of λ then $F_{\mu_t'}^{\psi^\pi}(\psi^\pi) \geqslant F_{\mu_t}^{\psi^\pi}(\psi^\pi)$ and therefore $\mathbb{E}_t^P[\psi^\pi]$ is lower under μ_t' than under μ_t ■

Lemma 5 establishes that if beliefs place more weight on larger values of λ, the central bank is perceived as more hawkish. Therefore, expected inflation is lower for any given sequence of shocks. Figure 16 plots the case where first-order stochastic dominance leads to an improvement in reputation.



(a) Beliefs over λ **(b)** Beliefs over ψ^π

Figure 16: First-order stochastic dominance: $\mu_t' \overset{FOSD}{\gtrsim} \mu_t$

Lemma 6. *Suppose the private sector has beliefs μ_t over λ. Consider a different set of beliefs μ_t' that is a mean-preserving spread of μ_t. Then, $\mathbb{E}_t^P[\psi^\pi]$ is lower under μ_t that under μ_t'*

Proof. This is a direct consequence of Jensen's inequality and the convexity of $\psi^\pi(\lambda)$. ■

Lemma 6 says that greater uncertainty about the central bank's preferences raises expected inflation. Because ψ^π is convex in λ, the private sector prices in the possibility of a dovish central bank. When in doubt, they expect softer responses to shocks.

B Solution to the Optimal Policy Problem

B.1 Learning Structure

In this subsection, we analytically characterize the learning structure of our model. Using our functional form assumptions in (10) yields

$$\begin{aligned}\mu_{t+1}\left(\tilde{\psi}^\pi\right) &\propto \sqrt{\tau_\eta} exp\left(-\frac{\tau_\eta}{2}\left(\kappa^{-1}\left(\tilde{\pi}_t - \tilde{\psi}^\pi X_t + \kappa\eta_t\right)\right)^2\right) \times \sqrt{\tau_t} exp\left(-\frac{\tau_t}{2}\left(\tilde{\psi}^\pi - \overline{\psi}_t\right)^2\right) \\ &= exp\left(-\frac{\tau_\eta\left(\kappa^{-1}X_t\right)^2 + \tau_t}{2}\left(\tilde{\psi}^\pi - \overline{\psi}_{t+1}\right)^2\right)\end{aligned}$$

where

$$\overline{\psi}_{t+1} = \omega_t X_t^{-1}\left(\tilde{\pi}_t + \kappa\eta_t\right) + \left(1-\omega_t\right)\overline{\psi}_t \quad \text{where} \quad \omega_t = \frac{\left(\kappa^{-1}X_t\right)^2\tau_\eta}{\tau_t + \left(\kappa^{-1}X_t\right)^2\tau_\eta}$$

therefore the model exhibits a conjugate prior and $\mu_{t+1}\left(\tilde{\psi}^\pi\right) \sim \Psi_{t+1} | \Psi_{t+1} \in [0,1]$ where $\Psi_{t+1} \sim \mathcal{N}\left(\overline{\psi}_{t+1}, \tau_{t+1}^{-1}\right)$, with $\tau_{t+1} = \tau_t + \left(\kappa^{-1}X_t\right)^2\tau_\eta$.

The properties of this learning structure are illustrated in Figure 17.

The first panel depicts that the evolution of beliefs has a mean-reversion component: $\mathbb{E}_t^P[\psi^\pi] > \overline{\psi}_t$ only when $\overline{\psi}_t < \frac{1}{2}$. To see its effect on the learning structure, rewrite (13) as

$$\overline{\psi}_{t+1} = \overline{\psi}_t + \omega_t\left[\left(\mathbb{E}_t^P[\psi^\pi] - \overline{\psi}_t\right) + \left(\frac{\pi_t}{X_t} - \mathbb{E}_t^P[\psi^\pi]\right)\right] \tag{B.1}$$

The first term in this equation captures the effect of mean reversion on the dynamics of reputation. When reputation is poor ($\overline{\psi}_t > \frac{1}{2}$), this force pushes toward an improvement, and the opposite occurs when reputation is good ($\overline{\psi}_t < \frac{1}{2}$).

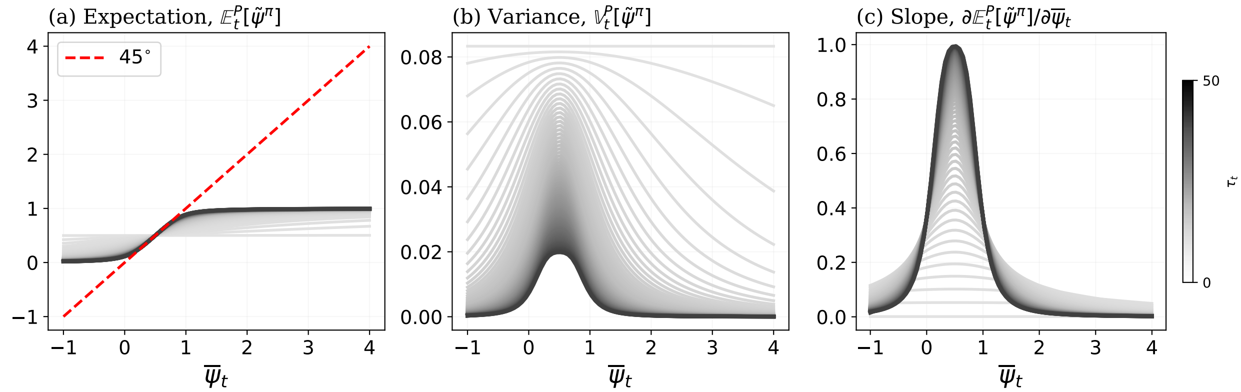


Figure 17: Prior beliefs as a function of $\overline{\psi}_t$

The second term reflects the effect of policy surprises on reputation. In the presence of inflationary shocks, $X_t > 0$, inflation below what the private sector expected, $\pi_t < \mathbb{E}_t^P[\psi^\pi] X_t$, counterfactually shifts beliefs toward more hawkish types. However, this by itself is not sufficient to improve reputation. If reputation is already high, mean reversion implies that only a sufficiently large hawkish surprise can shift beliefs further. By contrast, when reputation is low, mean reversion works in the opposite direction: reputation may improve even if inflation turns out to be higher than expected.

The second and third panels of Figure 17 show that uncertainty and sensitivity in the private sector's beliefs go hand in hand: states in which beliefs have higher variance are also those in which they are more sensitive to the central bank's actions.

B.2 The Central Bank's Problem

After observing the cost push shocks $\{\mathbb{E}_t[\varepsilon_{t+s}]\}_{s\geqslant 0}$ the central bank maximizes

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t\left[\sum_{s=0}^{\infty}\big(\tilde{y}_{t+s}^2 + \lambda\tilde{\pi}_t^2\big]\right]$$

subject to

(i) The New-Keynesiean Phillips Curve,

$$\tilde{\pi}_t = \kappa\tilde{y}_t + \beta\mathbb{E}_t^P[\tilde{\pi}_{t+1}] + \varepsilon_t$$

(ii) The private sector's expectation formation process and learning structure,

$$\begin{aligned}
\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right] =& \mathbb{E}_t^P\left[\psi^\pi\right]\mathbb{E}_t^P\left[X_{t+1}\right]\\
\mathbb{E}_t^P\left[\psi^\pi\right] =& \overline{\psi}_t - \frac{1}{\sqrt{\tau}}\frac{\phi\left(\tau\left(1-\overline{\psi}_t\right)\right)-\phi\left(-\tau\overline{\psi}_t\right)}{\Phi\left(\tau\left(1-\overline{\psi}_t\right)\right)-\Phi\left(-\tau\overline{\psi}_t\right)}\\
\overline{\psi}_{t+1} =& \omega_t X_t^{-1}\left(\tilde{\pi}_t+\kappa\eta_t\right)+\left(1-\omega_t\right)\overline{\psi}_t
\end{aligned}$$

where $\omega_t = \frac{\left(\kappa^{-1}X_t\right)^2\tau_\eta}{\tau+\left(\kappa^{-1}X_t\right)^2\tau_\eta}$ and $\tau_{t+1} = \tau_t + \left(\kappa^{-1}X_t\right)^2\tau_\eta$.

Let μ_t denote the Lagrange multipliers. The first-order conditions are

$$\begin{aligned}
\tilde{y}_t =& \kappa\mu_t\\
\lambda\tilde{\pi}_t =& -\mu_t + \sum_{s=1}^{\infty}\beta^s\mathbb{E}_t^{CB}\left[\beta\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial\tilde{\pi}_t}\mu_{t+s}\right]\\
\tilde{\pi}_t =& \kappa\tilde{y}_t + \beta\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right]+\varepsilon_t
\end{aligned}$$

Combining the three equations, we obtain

$$\tilde{y}_t = \tilde{y}_t^M + \psi^\pi\sum_{s=1}^{\infty}\beta^s\mathbb{E}_t^{CB}\left[\beta\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial\tilde{\pi}_t}\tilde{y}_{t+s}\right] \tag{B.2}$$

$$\tilde{\pi}_t = \tilde{\pi}_t^M + \kappa\psi^\pi\sum_{s=1}^{\infty}\beta^s\mathbb{E}_t^{CB}\left[\beta\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial\tilde{\pi}_t}\tilde{y}_{t+s}\right] \tag{B.3}$$

B.3 Solution Method

We assume the cost-push shock follows an AR(1) process, with persistence ρ. Given $\left(\overline{\psi}_t, \varepsilon_t\right)$, together with $\eta_t \sim F$ being white noise, the central bank solves

$$V\left(\overline{\psi}_t,\varepsilon_t\right) = \min_{\{y_{t+s},\pi_{t+s}\}_{s\geqslant 0}}\frac{1}{2}\mathbb{E}_t^{CB}\left[\sum_{s=0}^{\infty}\beta^s\left(y_{t+s}^2+\lambda\pi_{t+s}^2\right)\right]$$

subject to

$$\begin{aligned}
\varepsilon_{t+1} =& \rho\varepsilon_t + \nu_{t+1}\\
\pi_t =& \kappa y_t + \beta\mathbb{E}_t^P\left[\pi_{t+1}\right]+\varepsilon_t\\
\mathbb{E}_t^P\left[\pi_{t+s}\right] =& \mathbb{E}_t^P\left[\psi^\pi\right]\mathbb{E}_t^P\left[X_{t+s}\right]
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}^P\left[\psi^\pi\right] =& \overline{\psi}_t - \frac{1}{\sqrt{\tau}} \frac{\phi\left(\sqrt{\tau_t\left(1-\overline{\psi}_t\right)}\right) - \phi\left(-\sqrt{\tau_t \overline{\psi}_t}\right)}{\Phi\left(\sqrt{\tau_t\left(1-\overline{\psi}_t\right)}\right) - \Phi\left(-\sqrt{\tau_t \overline{\psi}_t}\right)} \\
X_t =& \frac{1}{1-\beta \mathbb{E}_t^P\left[\psi^\pi\right]\rho} \varepsilon_t \\
\overline{\psi}_{t+1} =& \overline{\psi}_t + \omega_t X_t^{-1}\left(\pi_t + \kappa \eta_t - \overline{\psi}_t X_t\right) \\
\omega_t =& \frac{\left(\kappa^{-1} X_t\right)^2 \tau_\eta}{\left(\kappa^{-1} X_t\right)^2 \tau_\eta + \tau}
\end{aligned}
$$

We can rewrite the problem recursively as

$$
V\left(\overline{\psi}, \varepsilon\right) = \min_y \quad \frac{1}{2}\left(y^2 + \lambda\left\{\kappa y + G\left(\overline{\psi}, \epsilon\right)\right\}^2\right) + \beta \mathbb{E}\left[V\left(\overline{\psi}', \epsilon'\right)\middle|\varepsilon\right]
$$

subject to

$$
\begin{aligned}
\overline{\psi}' =& \frac{\omega\left(\overline{\psi}, \varepsilon\right)}{G\left(\overline{\psi}, \varepsilon\right)}\left(\kappa\left(y+\eta\right) + G\left(\overline{\psi}, \varepsilon\right)\right) + \left(1-\omega\left(\overline{\psi}, \varepsilon\right)\right)\overline{\psi} \\
\varepsilon' =& \rho\varepsilon + \nu
\end{aligned}
$$

where

$$
\begin{aligned}
g\left(\overline{\psi}\right) =& \overline{\psi} - \frac{1}{\sqrt{\tau}} \frac{\phi\left(\sqrt{\tau}\left(1-\overline{\psi}\right)\right) - \phi\left(-\sqrt{\tau}\overline{\psi}\right)}{\Phi\left(\sqrt{\tau}\left(1-\overline{\psi}\right)\right) - \Phi\left(-\sqrt{\tau}\overline{\psi}\right)} \\
\omega\left(\overline{\psi}, \varepsilon\right) =& \frac{\left(\kappa^{-1} G\left(\overline{\psi}, \varepsilon\right)\right)^2 \tau_\eta}{\left(\kappa^{-1} G\left(\overline{\psi}, \varepsilon\right)\right)^2 \tau_\eta + \tau} \\
G\left(\overline{\psi}, \varepsilon\right) =& \left(1 - \beta\rho g\left(\overline{\psi}, \varepsilon\right)\right)^{-1} \varepsilon
\end{aligned}
$$

We discretize ε using the Rowenhorst method: $\boldsymbol{\varepsilon}$ (S-states vector) and $\boldsymbol{P}$ (transition matrix). We define a grid for the forecast error $\boldsymbol{\eta}$. We build a grid for the parameters of the initial prior $\overline{\boldsymbol{\psi}}$ (I elements). We propose an $I \times S$ matrix $\boldsymbol{V}^{old}$ where their elements (i, s) specify the value for $V\left(\overline{\psi}_i, \varepsilon_s\right)$. Given $\boldsymbol{V}^{old}$:

1. Compute
$$
\tilde{\boldsymbol{V}} = \boldsymbol{V}^{old} \times \boldsymbol{P}^T
$$

2. For each s:

i. Take $\tilde{\boldsymbol{V}}_{\bullet,s}$

ii. Compute $G\left(\overline{\psi}_i, \varepsilon_s\right)$, $\omega\left(\overline{\psi}_i, \varepsilon_s\right)$ and $g\left(\overline{\psi}_i\right)$.

iii. Compute $\overline{\psi}'(y,\eta)$.

iv. Solve

$$V\left(\overline{\psi}_i, \varepsilon_i\right) = \max_y \left\{ \frac{1}{2}\left(y^2 + \lambda\left(\kappa y + G\right)^2\right) + \beta \sum_\eta interp\left(\tilde{\boldsymbol{V}}, \overline{\psi}'(y,\eta)\right) \boldsymbol{F}_\eta \right\}$$

v. Store the solution in $\boldsymbol{V}^{new}$.

3. Check $\left\|\boldsymbol{V}^{new} - \boldsymbol{V}^{old}\right\|_\infty$ and update/stop.

B.4 Proof of Proposition 1

Proof. From (B.2) and (B.3), it suffices to show that

$$\mathcal{I}_t = \sum_{s=0}^{\infty} \beta^s \mathbb{E}_t^{CB}\left[\beta \frac{\partial \mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial \tilde{\pi}_t} \tilde{y}_{t+s}\right] \tag{B.4}$$

is negative when $X_t > 0$, and positive when $X_t < 0$. Recall that

$$\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right] = \mathbb{E}_{t+s}^P\left[\psi^\pi\right]\mathbb{E}_{t+s}^P\left[X_{t+s+1}\right]$$

Then, we have

$$\begin{aligned}\frac{\partial \mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial \tilde{\pi}_t} &= \mathbb{E}_{t+s}^P\left[\psi^\pi\right]\mathbb{E}_{t+s}^P\left[X_{t+s+1}\right]\left(\frac{\frac{\partial \mathbb{E}_{t+s}^P[\psi^\pi]}{\partial \tilde{\pi}_t}}{\mathbb{E}_{t+s}^P\left[\psi^\pi\right]} + \frac{\frac{\partial \mathbb{E}_{t+s}^P[X_{t+s+1}]}{\partial \tilde{\pi}_t}}{\partial \tilde{\pi}_t}\right) \\ &= \mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]\left(\frac{\frac{\partial \mathbb{E}_{t+s}^P[\psi^\pi]}{\partial \tilde{\pi}_t}}{\mathbb{E}_{t+s}^P\left[\psi^\pi\right]} + \frac{\frac{\partial \mathbb{E}_{t+s}^P[X_{t+s+1}]}{\partial \tilde{\pi}_t}}{\partial \tilde{\pi}_t}\right)\end{aligned}$$

Since $\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]\tilde{y}_{t+s} < 0$, we need to show that the term in parenthesis is positive when $X_t > 0$ and negative when $X_t < 0$. From the learning structure (13) and the definition of X_t: with a positive cost-push shock, higher inflation signals a lower commitment with inflation stability, whereas the opposite occurs with a negative shocks.

■

B.5 State-Dependence

Proposition 4. *The output gap overreaction* $\left|\tilde{y}_t - \tilde{y}_t^M\right|$ *is increasing in* $\mathbb{E}_t^P\left[\psi^\pi\right]$ *for small values of* $\mathbb{E}_t^P\left[\psi^\pi\right]$ *and decreases for large values of* $\mathbb{E}_t^P\left[\psi^\pi\right]$*.*

Proof. Under our functional form assumptions, (15) becomes

$$\frac{\partial \mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial \tilde{\pi}_t} = \frac{\partial \mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial \mathbb{E}_{t+s}^P\left[\psi^\pi\right]} \tau_{t+s} \mathbb{V}_{t+s}^P\left[\psi^\pi\right] \frac{\partial \overline{\psi}_{t+s}}{\partial \tilde{\pi}_t}$$

where $\mathbb{V}_{t+s}^P\left[\psi^\pi\right]$ is the dispersion of beliefs about ψ^π. Taking current uncertainty as a proxy for future uncertainty, this expression shows that expectations are most sensitive when belief dispersion is high. Since $\frac{\partial \mathbb{E}_{t+s}^P[\tilde{\pi}_{t+s+1}]}{\partial \mathbb{E}_{t+s}^P[\psi^\pi]}$ varies smoothly with reputation, and $\frac{\overline{\psi}_{t+s}}{\partial \tilde{\pi}_t}$ is also inverse U-shaped, then $\frac{\partial \mathbb{E}_{t+s}^P[\tilde{\pi}_{t+s+1}]}{\partial \tilde{\pi}_t}$ is inverse U-shaped. Then, $\frac{\partial \mathbb{E}_{t+s}^P[\tilde{\pi}_{t+s+1}]}{\partial \tilde{\pi}_t}$ increases with $\overline{\psi}_t$ is small, and decreases when $\overline{\psi}_t$ is large. Everything else equal, this implies that $\mathcal{I}_t$ increases with $\overline{\psi}_t$ at first, and then decreases. ■

B.6 Proof of Proposition 2

Proof. Suppose for simplicity that $X_t > 0$. From the private sector learning, we have

$$\overline{\psi}_{t+1} = \overline{\psi}_t + \omega_t\left[\left(\mathbb{E}_t^P\left[\psi^\pi\right] - \overline{\psi}_t\right) - X_t^{-1}\left(\pi_t - \tilde{\pi}_t^M\right)\right]$$

The first term is negative when $\mathbb{E}_t^P\left[\psi^\pi\right] > \frac{1}{2}$ and positive when $\mathbb{E}_t^P\left[\psi^\pi\right] < \frac{1}{2}$. From Proposition 4, there exist a value $\widehat{\psi} < \psi^\pi$ such that $\varepsilon_t^m > 0$ when $\mathbb{E}_t^P\left[\psi^\pi\right] > \widehat{\psi}$. Then, there exist a threshold such that reputation improves, from the ex-ante perspective of the central bank.

■

B.7 Sensitivity of expectations to τ

Recall the sensitivity of beliefs is given by (15)

$$\frac{\partial \mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial \overline{\psi}_t} = \frac{\partial \mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial \mathbb{E}_{t+s}^P\left[\psi^\pi\right]} \frac{\partial \mathbb{E}_{t+s}^P\left[\psi^\pi\right]}{\partial \overline{\psi}_{t+s}} \frac{\partial \overline{\psi}_{t+s}}{\partial \tilde{\pi}_t}$$

Conditional on a certain reputation, $\mathbb{E}_t^P\left[\psi^\pi\right]$, the sensitivity of expectations depends on two elements:

1. The sensitivity of reputation to the prior's location. Under our functional form assumptions,
$$\frac{\partial \mathbb{E}^P_{t+s}\left[\psi^\pi\right]}{\partial \overline{\psi}_{t+s}} = \tau_{t+s}\mathbb{V}^P_{t+s}\left[\psi^\pi\right]$$
which is increasing in τ_{t+s}. The tighter the beliefs are around the prior's location, the more sensitive they become to changes in it.

2. The sensitivity of the central bank's choices on the prior's location
$$\frac{\partial \overline{\psi}_{t+s}}{\partial \tilde{\pi}_t} = \left(\prod_{j=1}^{s-1}(1-\omega_{t+s-j})\right)\omega_t X_t^{-1} \qquad s \geqslant 1$$
which, for $s \geqslant 2$ increases in τ when τ is small, and decreases when τ is large.[34] If $\overline{\psi}_{t+1}$ depends almost entirely on actions at period t, then those actions will have little impact on $\overline{\psi}_{t+s}$ for $s \geqslant 2$.

Taken together, these facts imply that when τ is relatively small, an increase in τ increases the sensitivity of expectations to the central bank's actions. And when τ is large, an increase in τ may decrease the sensitivity of expectations to the central bank's actions.

C Extensions of the Baseline Model

C.1 Uncertainty about Interest Rate Smoothing

Suppose for simplicity that prices are fixed. We slightly modify the information structure. The central bank and the private sector have the same information about the natural rate, but there are random monetary policy shocks η_t. Let $\tilde{i}_t$ denote the ex-ante interest rate. We have $i_t = \tilde{i}_t - \frac{1}{\sigma}\eta_t$. Then, the (ex-ante) economy private sector is only characterized by the Euler Equation
$$\tilde{y}_t = \mathbb{E}^P_t\left[\tilde{y}_{t+1}\right] - \frac{1}{\sigma}\left(\tilde{i}_t - \mathbb{E}_t\left[r^n_t\right]\right) \tag{C.1}$$

[34] For $s = 1$ is it always decreasing in τ.

Under the baseline model, the Divine Coincidence holds, and the central bank can always implement the first best. Instead, suppose the central bank has the triple mandate (16):

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t^{CB}\left[\sum_{s=0}^{\infty}\left(\tilde{y}_{t+s}^2 + \lambda\tilde{\pi}_{t+s}^2 + \varphi\tilde{i}_{t+s}^2\right)\right]$$

If $\varphi > 0$, then the divine coincidence does not hold. In particular, the larger φ the less the central bank is willing to move interest rates to stabilize the economy. Since prices are fixed, then the private sector's perception of λ does not matter. Instead, suppose the private sector does not know how much is the central bank is willing to move rates to stabilize the economy. That is, they have beliefs μ_t over φ. To form their forecasts, they also assume the central bank optimizes under discretion.

A myopic central bank of type $\varphi = \tilde{\varphi}$ has the following first-order condition

$$\tilde{y}_t = \sigma\tilde{\varphi}\tilde{i}_t$$

Define $\psi^y(\tilde{\varphi}) := \frac{\sigma\tilde{\varphi}}{1+\sigma^2\tilde{\varphi}}$, taking expectations as exogenous, then the central bank implements

$$\tilde{y}_t(\tilde{\varphi}) = \psi^y(\tilde{\varphi})\left(\mathbb{E}_t\left[r_t^n\right] + \sigma\mathbb{E}_t^P\left[\tilde{y}_{t+1}\right]\right)$$

We conjecture that the central bank follows linear policy rule, and they are not aware of the private sector's perceived bias:

$$\tilde{y}_t(\tilde{\varphi}) = \psi(\tilde{\varphi})\sum_{s=0}^{\infty}\Theta_s\mathbb{E}_t\left[r_{t+s}^n\right]$$

Matching coefficients for $s = 0$

$$\Theta_0 = 1$$

For $s = 1$, using that there is no anticipated learning

$$\Theta_1 = \sigma\mathbb{E}_t^P\left[\psi^y\right]$$

For any arbitrary s, we have

$$\Theta_s = \sigma\mathbb{E}_t^P\left[\psi\right]\Theta_{s-1} = \left(\sigma\mathbb{E}_t^P\left[\psi^y\right]\right)^s$$

Putting everything together,

$$\tilde{y}_t(\tilde{\varphi}) = \psi^y(\tilde{\varphi}) \sum_{s=0}^{\infty} \left(\sigma \mathbb{E}_t^P[\psi^y]\right)^s \mathbb{E}_t\left[r_{t+s}^n\right] = \psi^y(\tilde{\varphi}) X_t \tag{C.2}$$

$$\tilde{i}_t(\tilde{\varphi}) = (1 - \sigma\psi^y(\tilde{\varphi})) \sum_{s=0}^{\infty} \left(\sigma \mathbb{E}_t^P[\psi]\right)^s \mathbb{E}_t\left[r_{t+s}^n\right] = \psi^i(\tilde{\varphi}) X_t \tag{C.3}$$

In this model, we define $\mathbb{E}_t^P[\psi^y]$ as the central bank's reputation. A smaller value of $\mathbb{E}_t^P[\psi^y]$ signals a more hawkish central bank. When the private sector's perception about φ switches to a smaller value, $\mathbb{E}_t^P[\psi^y]$ decreases and the solution moves closer to the one from the Divine Coincidence. When the perception of φ goes to zero, then we recover $\mathbb{E}_t^P[\psi^y] = 0$ and we recover the Divine coincidence. When $\varphi \to \infty$, then $\mathbb{E}_t^P[\psi^y] \to \frac{1}{\sigma}$

Optimal Policy We keep the private sector's learning structure. The central bank observes a forecast of the natural rate and optimizes under discretion. The final allocation will be $y_t = \tilde{y}_t + \eta_t$ and i_t from (C.3). The private sector is unsure whether the final allocation was due to the central bank's preference φ, the monetary policy shock, η_t.

Each period, given μ_t, the private sector updates their prior distribution to their posterior as follows:

1. Given μ_t, the private sector believes that a central bank of type $\varphi = \tilde{\varphi}$ chooses the allocations $\tilde{y}^M(\tilde{\varphi})$ and $i_t(\tilde{\varphi})$ given by (C.2) and (C.3), leading to the ex-post realizations

$$y_t = \tilde{y}^M(\tilde{\varphi}) + \eta_t(\tilde{\varphi})$$

2. The true realization of output gap is given by the allocation chosen by the central bank plus the forecast error

$$y_t = \tilde{y}_t + \eta_t$$

3. Upon observing y_t and i_t, the private sector does not know whether the current realizations are due to the monetary surprise η_t or a (myopic) central bank's preferences $\varphi = \tilde{\varphi}$. Then, from the private sector's point of view, the monetary policy shock of a central bank of type $\varphi = \tilde{\varphi}$ must be

$$\eta_t(\tilde{\varphi}) = \tilde{y}_t - \psi^y(\tilde{\varphi}) X_t + \eta_t$$

Under the same functional assumptions for η_t and beliefs over φ^y, beliefs over φ^y also have a conjugate prior, and the updating of the location parameter is

$$\overline{\psi}_{t+1} = \omega_t X_t^{-1} \left(\tilde{y}_t + \eta_t\right) + \left(1 - \omega_t\right) \overline{\psi}_t \quad \text{where} \quad \omega_t = \frac{X_t^2 \tau_\eta}{\tau_t + X_t^2 \tau_\eta}$$

Then, the same properties will hold. After observing the demand shocks $\left\{\mathbb{E}_t\left[r_{t+s}^n\right]\right\}_{s\geqslant 0}$ the central bank maximizes

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t\left[\sum_{s=0}^{\infty} \beta^s \left(\tilde{y}_{t+s} + \varphi \tilde{i}_{t+s}^2\right)\right]$$

subject to

(i) The Euler Equation

$$\tilde{y}_t = \mathbb{E}_t^P\left[\tilde{y}_{t+1}\right] - \frac{1}{\sigma}\left(\tilde{i}_t - \mathbb{E}_t\left[r_t^n\right]\right)$$

(ii) The private sector's expectation formation process and learning structure,

$$\begin{aligned}
\mathbb{E}_t^P\left[\tilde{y}_{t+1}\right] =& \mathbb{E}_t^P\left[\psi^y\right]\mathbb{E}_t^P\left[X_{t+1}\right] \\
\mathbb{E}_t^P\left[\psi^y\right] =& \overline{\psi}_t - \frac{1}{\sqrt{\tau}}\frac{\phi\left(\tau\left(1-\overline{\psi}_t\right)\right) - \phi\left(-\tau\overline{\psi}_t\right)}{\Phi\left(\tau\left(1-\overline{\psi}_t\right)\right) - \Phi\left(-\tau\overline{\psi}_t\right)} \\
\overline{\psi}_{t+1} =& \omega_t X_t^{-1}\left(\tilde{y}_t + \eta_t\right) + \left(1-\omega_t\right)\overline{\psi}_t
\end{aligned}$$

where $\omega_t = \frac{X_t^2 \tau_\eta}{\tau + X_t^2 \tau_\eta}$.

Let μ_t denote the Lagrange multipliers, the first-order conditions are

$$\begin{aligned}
\varphi \tilde{i}_t =& \frac{1}{\sigma}\mu_t \\
\tilde{y}_t =& \mu_t - \sum_{s=1}^{\infty} \mathbb{E}_t^{CB}\left[\frac{\partial \mathbb{E}_{t+s}^P\left[\tilde{y}_{t+s+1}\right]}{\partial \tilde{y}_t}\mu_{t+s}\right] \\
\tilde{y}_t =& \mathbb{E}_t^P\left[\tilde{y}_{t+1}\right] - \frac{1}{\sigma}\left(\tilde{i}_t - \mathbb{E}_t\left[r_t^n\right]\right)
\end{aligned}$$

Combining the three equations we obtain

$$\tilde{y}_t = \tilde{y}_t^M - \psi^y \sum_{s=1}^{\infty} \mathbb{E}_t^{CB}\left[\frac{\partial \mathbb{E}_{t+s}^P\left[\tilde{y}_{t+s+1}\right]}{\partial \tilde{y}_t}\tilde{i}_{t+s}\right] \tag{C.4}$$

$$\tilde{i}_t = \tilde{i}_t^M + \sigma\psi^y \sum_{s=1}^{\infty} \mathbb{E}_t^{CB}\left[\frac{\partial \mathbb{E}_{t+s}^P\left[\tilde{y}_{t+s+1}\right]}{\partial \tilde{y}_t}\tilde{i}_{t+s}\right] \tag{C.5}$$

The insurance term in this case is given by

$$\mathcal{I}_t = \sum_{s=1}^{\infty} \mathbb{E}_t^{CB}\left[\frac{\partial \mathbb{E}_{t+s}^P\left[\tilde{y}_{t+s+1}\right]}{\partial \tilde{y}_t}\tilde{i}_{t+s}\right]$$

Notice that this model is isomorphic to a model where the central bank tries to learn about the relative weight given to inflation. Thus, the same principles will hold. The main difference is that the central bank will now prioritize more output gap stability relative to the myopic benchmark. However, the policy prescriptions for the interest rate remains unchanged.

C.2 Uncertainty about Inflation Target

Suppose that the central bank's objective function is given by

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t^{CB}\left[\sum_{s=0}^{\infty}\beta^s\left(\tilde{y}_{t+s}^2 + \lambda\left(\pi_{t+s} - \varphi\right)^2\right)\right] \tag{C.6}$$

In contrast with the main model, the private sector is uncertain about the central bank's inflation target φ. We also assume the private sector is aware of the value of λ. Like our main model, the price sector believes the central bank is myopic, and we maintain the information structure. A myopic central bank with inflation target $\varphi = \tilde{\varphi}$ has the following first-order condition

$$\tilde{y}_t + \lambda\kappa\left(\tilde{\pi}_t - \tilde{\varphi}\right) = 0$$

Define $\psi^\pi = \frac{1}{1+\lambda\kappa^2}$ and $\psi^y = \frac{\lambda\kappa}{1+\lambda\kappa^2}$. Taking expectations as exogenous, the central bank implements

$$\begin{aligned}\tilde{y}_t\left(\tilde{\varphi}\right) =& -\psi^y\left(\varepsilon_t + \beta\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right] - \tilde{\varphi}\right)\\ \tilde{\pi}_t =& \psi^\pi\left(\varepsilon_t + \beta\mathbb{E}_t^P\left[\tilde{\pi}_{t+s}\right]\right) + \kappa\psi^y\tilde{\varphi}\end{aligned}$$

We conjecture that the central bank follows a linear policy rule, and they are not aware of the private sector's perceived bias:

$$\tilde{y}_t\left(\tilde{\varphi}\right) = \Psi_1^y\tilde{\varphi} - \Psi_2^y\mathbb{E}_t^P\left[\varphi\right] - \Psi_3^y\sum_{s=0}^{\infty}\Theta_s^y\mathbb{E}_t\left[\varepsilon_{t+s}\right]$$

$$\tilde{\pi}_t\left(\tilde{\varphi}\right) = \Psi_1^\pi \tilde{\varphi} + \Psi_2^\pi \mathbb{E}_t^P\left[\varphi\right] + \Psi_3^\pi \sum_{s=0}^{\infty} \Theta_s^\pi \mathbb{E}_t\left[\varepsilon_{t+s}\right]$$

Then

$$\Psi_1^y = \psi^y \qquad \Psi_2^y = \frac{\kappa\beta\psi^y}{1-\beta\psi^\pi}\psi^y \qquad \Psi_3^y = \psi^y$$

$$\Psi_1^\pi = \kappa\psi^y \qquad \Psi_2^\pi = \frac{\kappa\beta\psi^y}{1-\beta\psi^\pi}\psi^\pi \qquad \Psi_3^\pi = \psi^\pi$$

Finally, we match the coefficients for $\{\Theta_s^y, \Theta_s^\pi\}_{s\geqslant 0}$. Start with $s = 0$

$$\Theta_0^y = \Theta_0^\pi = 1$$

For $s = 1$

$$\Theta_1^y = \beta\psi^\pi \qquad \Theta_1^\pi = \beta\psi^\pi$$

For an arbitrary s, we have

$$\Theta_s = \beta\psi^\pi\Theta_{s-1} = \left(\beta\psi^\pi\right)^s$$

Putting everything together,

$$\tilde{y}_t\left(\tilde{\varphi}\right) = \Psi_1^y\tilde{\varphi} - \Psi_2^y\mathbb{E}_t^P\left[\varphi\right] - \Psi_3^y\sum_{s=0}^{\infty}\left(\beta\psi^\pi\right)^s\mathbb{E}_t\left[\varepsilon_{t+s}\right] = \Psi_1^y\tilde{\varphi} - \Psi_2^y\mathbb{E}_t^P\left[\varphi\right] - \Psi_3^y X_t \tag{C.7}$$

$$\tilde{\pi}_t\left(\tilde{\varphi}\right) = \Psi_1^\pi\tilde{\varphi} + \Psi_2^\pi\mathbb{E}_t^P\left[\varphi\right] - \Psi_3^\pi\sum_{s=0}^{\infty}\left(\beta\psi^\pi\right)^2\mathbb{E}_t\left[\varepsilon_{t+s}\right] = \Psi_1^\pi\tilde{\varphi} + \Psi_2^\pi\mathbb{E}_t^P\left[\varphi\right] + \Psi_3^\pi X_t \tag{C.8}$$

The private sector's perception about the inflation target acts as a cost-push shock: it raises inflation expectations and decreases output. In contrast to the case where λ is uncertain, it does not depend on the shocks: according to the private sector, the central bank will try to boost the output gap to inflate the economy. Therefore, they will raise their inflation expectations. In this model we can interpret $\mathbb{E}_t^P\left[\varphi\right]$ as the reputation of the central bank: a good reputation is tied to a low inflation target.

Upon observing the final allocations π_t and y_t, the private sector does not know whether the current realizations are due to the forecast error η_t or the myopic central bank's inflation target $\tilde{\varphi}$. From the private sector's point of view, the first-order condition of a myopic central bank with inflation target $\varphi = \tilde{\varphi}$ is

$$\tilde{y}_t(\tilde{\varphi}) + \lambda\kappa(\tilde{\pi}_t(\tilde{\varphi}) - \tilde{\varphi}) = 0$$

Then, the forecast error of a central bank with inflation bias $\varphi = \tilde{\varphi}$ must be

$$\eta_t(\tilde{\varphi}) = \psi^{\pi}\left(\tilde{y}_t + \lambda\kappa(\tilde{\pi}_t - \tilde{\varphi}) + \left(1 + \lambda\kappa^2\right)\eta_t\right)$$

We assume that the forecast error is normally distributed, i.e., $\eta_t \sim \mathcal{N}\left(0, \tau_\eta^{-1}\right)$ for all t; and, the prior belief about φ is also normally distributed, i.e., $\tilde{\varphi} \sim \mathcal{N}\left(\mathbb{E}_t^P[\varphi], \tau_t^{-1}\right)$. Under these assumptions, beliefs have a conjugate prior, and

$$\mathbb{E}_{t+1}^P[\varphi] = \omega_t(\lambda\kappa)^{-1}\left(\tilde{y}_t + \lambda\kappa\tilde{\pi}_t + \left(1 + \lambda\kappa^2\right)\eta_t\right) + (1 - \omega_t)\,\mathbb{E}_t^P[\varphi]$$

where

$$\omega_t = \frac{(\psi^y)^2\tau_\eta}{(\psi^y)^2\tau_\eta + \tau_t}$$

In contrast to our main model, the attention weight ω_t does not depend on the size of the shocks. To understand the implications of the learning process, rewrite this expression as

$$\mathbb{E}_{t+1}^P[\varphi] = \mathbb{E}_t^P[\varphi] + \omega_t(\lambda\kappa)^{-1}\left(\tilde{y}_t + \lambda\kappa\left(\tilde{\pi}_t - \mathbb{E}_t^P[\varphi]\right) + \left(1 + \lambda\kappa^2\right)\eta_t\right)$$

Then, the only way to improve the central bank can improve its reputation is by acting more hawkish than expected. This way, the central bank signals a lower inflation target and reduces inflation expectations in the following period. As in our main model, we assume the Kalman gain is constant, so $\omega_t = \omega$ for all t. After observing the cost push shocks $\{\mathbb{E}_t[\varepsilon_{t+s}]\}_{s\geqslant 0}$ the central bank maximizes

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t\left[\sum_{s=0}^{\infty}\beta^s\left(\tilde{y}_{t+s}^2 + \lambda(\tilde{\pi}_{t+s} - \varphi)^2\right)\right]$$

subject to

(i) The New Keynesian Phillips Curve

$$\tilde{\pi}_t = \kappa\tilde{y}_t + \beta\mathbb{E}_t^P[\tilde{\pi}_{t+1}] + \varepsilon_t$$

(ii) The private sector's learning structure,

$$\begin{aligned}\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right] &= \left(\Psi_1^\pi + \Psi_2^\pi\right)\mathbb{E}_t^P\left[\varphi\right] + \Psi_3^\pi \mathbb{E}_t\left[X_{t+1}\right] \\ \mathbb{E}_{t+1}^P\left[\varphi\right] &= \omega_t\left(\lambda\kappa\right)^{-1}\left(\tilde{y}_t + \lambda\kappa\tilde{\pi}_t + \left(1+\lambda\kappa^2\right)\eta_t\right) + \left(1-\omega_t\right)\mathbb{E}_t^P\left[\varphi\right]\end{aligned}$$

where $\quad \omega_t = \frac{(\psi^\pi)^2\tau_\eta}{(\psi^\pi)^2\tau_\eta + \tau_t}$.

Let μ_t denote the Lagrange multipliers, the first-order conditions are

$$\begin{aligned}\tilde{y}_t &= \kappa\mu_t + \sum_{s=1}^{\infty}\beta^s\mathbb{E}_t^{CB}\left[\beta\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial\tilde{y}_t}\mu_{t+s}\right] \\ \lambda\left(\tilde{\pi}_t - \varphi\right) &= -\mu_t + \sum_{s=1}^{\infty}\beta^s\mathbb{E}_t^{CB}\left[\beta\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial\tilde{\pi}_t}\mu_{t+s}\right] \\ \tilde{\pi}_t &= \kappa\tilde{y}_t + \beta\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right] + \varepsilon_t\end{aligned}$$

Combining the first and second equation

$$\tilde{y}_t + \lambda\kappa\left(\tilde{\pi}_t - \varphi\right) = \sum_{s=1}^{\infty}\beta^s\mathbb{E}_t^{CB}\left[\beta\left(\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial\tilde{y}_t} + \kappa\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial\tilde{\pi}_t}\right)\mu_{t+s}\right]$$

Using the private sector's learning structure

$$\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial\tilde{y}_t} + \kappa\frac{\partial\mathbb{E}_{t+s}^P\left[\tilde{\pi}_{t+s+1}\right]}{\partial\tilde{\pi}_t} = \frac{\omega}{\psi^y}\left(1-\omega\right)^{s-1}$$

Combining with the New Keynesian Phillips Curve we have

$$\begin{aligned}\tilde{y}_t &= \tilde{y}_t^M + \frac{\omega}{1-\omega}\sum_{s=1}^{\infty}\left(\beta\left(1-\omega\right)\right)^s\mathbb{E}_t^{CB}\left[\mu_{t+s}\right] \\ \tilde{\pi}_t &= \tilde{\pi}_t^M + \kappa\frac{\omega}{1-\omega}\sum_{s=1}^{\infty}\left(\beta\left(1-\omega\right)\right)^s\mathbb{E}_t^{CB}\left[\mu_{t+s}\right]\end{aligned}$$

The unique equilibrium is linear, and given by

$$\tilde{y}_t = \Theta_1^y\varphi - \Theta_2^y\mathbb{E}_t^P\left[\varphi\right] - \sum_{s=0}^{\infty}\Theta_{3,s}^y\mathbb{E}_t\left[\varepsilon_{t+s}\right] \tag{C.9}$$

$$\tilde{\pi}_t = \Theta_1^\pi \varphi + \Theta_2^\pi \mathbb{E}_t^P[\varphi] + \sum_{s=0}^{\infty} \Theta_{3,s}^\pi \mathbb{E}_t[\varepsilon_{t+s}] \tag{C.10}$$

How does the optimal policy compare to the myopic central bank? First, the inflation target has a smaller influence on the optimal policy, that is, $\Theta_1^y < \Psi_1^y$ and $\Theta_1^\pi < \Psi_1^\pi$. The central bank internalizes that raising output will raise inflation expectations.

Second, the optimal policy for output gap (inflation) is to overreact (underreact) to the central bank's reputation $\mathbb{E}_t^P[\varphi]$, that is, $\Theta_2^y > \Psi_2^y$ and $\Theta_2^\pi < \Psi_2^\pi$. When the private sector learns about the inflation target, inflation expectations are a cost-push shock. The optimal response to a cost-push shock is to induce a recession to mitigate its impact on inflation. For the case of the perceived inflation target, it is a shock that is endogenous to policy. By overreacting, the central bank reduces the future recession's size.

Third, the optimal policy for current shocks is to react exactly like the myopic central bank, that is, $\Theta_{3,0}^y = \Psi_3^y$ and $\Theta_{3,0}^\pi = \Psi_3^\pi$. This is no longer true when considering the reaction to persistent shocks. When shocks are persistent, a current shock is also informative about a future recession. To smooth the size of the future recession, the central bank overreacts to improve its reputation. Therefore, we have $\Theta_{3,s}^y > \Psi_{3,s}^y$ and $\Theta_{3,1}^\pi < \Psi_{3,s}^\pi$ for $s \geqslant 1$.

As in the main model, the central bank overreacts to persistent shocks. However, there is no insurance. There is no incentive to improve reputation in response to an *iid* shock. In response to a persistent shock, the central bank smooths the size of the recession. In response, the central bank trades-off some of the future recession by inducing a current recession. In the main model, this mechanism holds regardless of the nature of the shock. For that reason, there is no overreaction in response to contemporary cost-push shocks. In this sense, when the central bank is concerned about the private sector's perception of the inflation target, reputation and stabilization are independent from each other. This is no longer true in the main model: there, the role of reputation is precisely to reduce the cost of stabilization.

C.3 Correctly Specified Beliefs

So far, we assumed for tractability that the private sector held misspecified beliefs. They believe the central bank does not internalize their belief updating process. In reality, financial institutions are aware that the central bank is trying to manipulate its beliefs. To add this dimension, we assume the private sector is fully aware that the central bank internalizes their learning process. For simplicity, we assume the central bank has no commitment. The learning process is still the same as (11), but we do not place any functional form assumptions . The first-order condition of a central bank of type $\tilde{\psi}^{\pi}$ is:

$$\tilde{y}_t\left(\tilde{\psi}^{\pi}\right) = \tilde{y}_t^M\left(\tilde{\psi}^{\pi}\right) + \tilde{\psi}^{\pi}\sum_{s=1}^{\infty}\beta^s\mathbb{E}_t^{CB}\left[\beta\frac{\partial\mathbb{E}_{t+s}^P\left[\pi_{t+1+s}\left(\psi^{\pi}\right)\right]}{\partial\tilde{\pi}_t\left(\tilde{\psi}^{\pi}\right)}\tilde{y}_{t+s}\left(\tilde{\psi}^{\pi}\right)\right]$$

where

$$\tilde{y}_t^M\left(\tilde{\psi}^{\pi}\right) = -\,\psi^y\left(\tilde{\psi}^{\pi}\right)\left(\varepsilon_t + \beta\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\left(\tilde{\psi}^{\pi}\right)\right]\right)$$

$$\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\left(\tilde{\psi}^{\pi}\right)\right] = \int_0^1\mathbb{E}_t\left[\tilde{\pi}_{t+1}\left(\tilde{\psi}^{\pi}\right)\right]\mu_t\left(\tilde{\psi}^{\pi}\right)d\tilde{\psi}^{\pi}$$

Current inflation expectations, $\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\left(\psi^{\pi}\right)\right]$, are endogenous, which means that we cannot interpret $\tilde{y}_t^M$ as the equilibrium allocation implemented by a myopic central bank. However, $\tilde{y}_t^M$ is the *temporary* equilibrium allocation (Hicks 1946, García-Schmidt and Woodford 2019). That is, the equilibrium allocation taking expectations of the next period $\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right]$ as given. We can define overreaction and underreaction with respect to this new benchmark, which is analog of a myopic central bank when beliefs are correctly specified.

Since there is a continuum of states $\psi^{\pi} \in (0,1)$, the whole distribution of prior beliefs is a state variable, making the problem intractable. One possibility to make the problem tractable is to assume types are finite, as in Bocola et al. (2025). However, this would eliminate the extensive margin of reputation we consider in our analysis. We take a different route, and consider the three-period version of our model from Section 2. Given any initial distribution of beliefs at $t = 0$, $\mathbb{E}_1^P\left[\tilde{\pi}_2\right] = \mathbb{E}_1^P\left[\psi^{\pi}\right]\mathbb{E}_1\left[\varepsilon_2\right]$. Therefore, all that matters for optimal policy is the central bank's reputation in $t = 1$, $\mathbb{E}_1^P\left[\psi^{\pi}\right]$. Using backward induction, we have

$$\tilde{y}_0\left(\tilde{\psi}^{\pi}\right) = \tilde{y}_0^M\left(\tilde{\psi}^{\pi}\right) + \tilde{\psi}^{\pi}\beta\mathbb{E}_0^{CB}\left[\beta\frac{\partial\mathbb{E}_1^P\left[\tilde{\pi}_2\right]}{\partial\tilde{\pi}_0\left(\tilde{\psi}^{\pi}\right)}\tilde{y}_1\left(\tilde{\psi}^{\pi}\right)\right]$$

$$\tilde{\pi}_0\left(\tilde{\psi}^{\pi}\right) = \tilde{\pi}_0^M\left(\tilde{\psi}^{\pi}\right) + \kappa\tilde{\psi}^{\pi}\beta\mathbb{E}_0^{CB}\left[\beta\frac{\partial\mathbb{E}_1^P\left[\tilde{\pi}_2\right]}{\partial\tilde{\pi}_0\left(\tilde{\psi}^{\pi}\right)}\tilde{y}_1\left(\tilde{\psi}^{\pi}\right)\right]$$

Let $\mu_0\left(\psi^{\pi};\tilde{\psi}^{\pi}\right)$ denote the prior density of the central bank of type ψ^{π}, from the point of view of a central bank of type $\tilde{\psi}^{\pi}$. The posterior density is characterized by

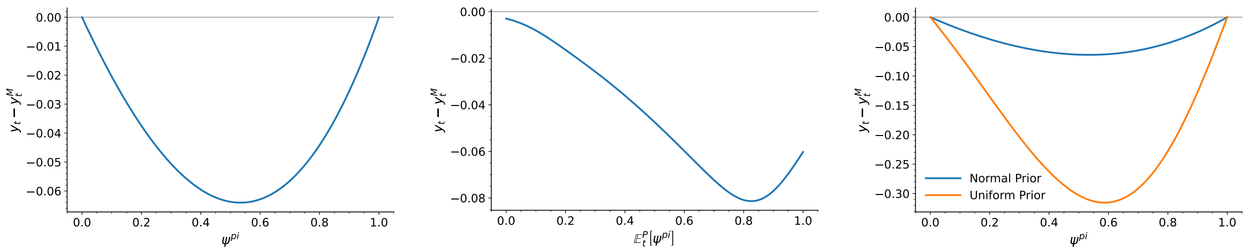
$$\mu_1\left(\psi^{\pi};\tilde{\psi}^{\pi}\right)=\frac{f_\eta\left(\kappa^{-1}\left(\tilde{\pi}_0\left(\psi^{\pi}\right)-\tilde{\pi}_0\left(\tilde{\psi}^{\pi}\right)\right)+\eta_0\right)\mu_0\left(\psi^{\pi};\tilde{\psi}^{\pi}\right)}{\int_0^1 f_\eta\left(\kappa^{-1}\left(\tilde{\pi}_0\left(\psi^{\pi}\right)-\tilde{\pi}_0\left(\tilde{\psi}^{\pi}\right)\right)+\eta_0\right)\mu_0\left(\psi^{\pi};\tilde{\psi}^{\pi}\right)d\psi^{\pi}}$$
$$=\frac{f_\eta\left(\psi^{\pi};\tilde{\psi}^{\pi},\eta_0\right)\mu_0\left(\psi^{\pi};\tilde{\psi}^{\pi}\right)}{\int_0^1 f_\eta\left(\psi^{\pi};\tilde{\psi}^{\pi},\eta_0\right)\mu_0\left(\psi^{\pi};\tilde{\psi}^{\pi}\right)d\psi^{\pi}}$$

Then, we have
$$\frac{\partial\mathbb{E}_1^P\left[\psi^{\pi}\right]}{\partial\tilde{\pi}_0\left(\tilde{\psi}^{\pi}\right)}=\kappa^{-1}\mathbb{COV}_1^P\left[\psi^{\pi},s_\eta\left(\psi^{\pi};\tilde{\psi}^{\pi},\eta_0\right)\right]$$

where $s_\eta\left(\psi^{\pi};\tilde{\psi}^{\pi},\eta_0\right)$ is the score of the likelihood of a central bank of type ψ^{π} from the point of view of central bank of type $\tilde{\psi}^{\pi}$. Since because central banks with a larger value of ψ^{π}, will choose higher (lower) inflation in response to a positive (negative) cost-push shock, $\tilde{\pi}_0\left(\psi^{\pi}\right)$ increases (decreases) with ψ^{π}. This expression formalizes that the central bank's reputation is more sensitive when its actions are more informative about its type. The score is the sensitivity of the likelihood of a type ψ^{π} to the central bank's actions. The sensitivity of expectations averages over all the hypothetical types ψ^{π}. This expression holds regardless of the functional form assumptions.

For any value of η_0, the score is nondecreasing in ψ^{π} when the cost-push shock is positive, and negative when it is negative. As a result, the covariance will be positive when the cost-push shock is positive, and negative otherwise. Since $\mathbb{E}_1\left[\varepsilon_2\right]\tilde{y}_1<0$, then Proposition 1 still holds, and there will be overreaction in the output gap and underreaction in inflation. Overreaction will be smaller compared to the case where the private sector believes the central bank is myopic: since the private sector internalizes that the central bank would like to increase its reputation, overreaction becomes less informative about the central bank's type. As a result, overreaction is smaller.

Figure 18a plots the overreaction as a function of the central bank's type, ψ^{π}. The magnitude of overreaction is u-shaped: reputation allows the central bank to improve its inflation–output trade-off, whose importance increases when the central bank's preferences are more balanced. Figure 18b displays overreaction of a central bank with $\lambda=10$ as a function of reputation. Consistent with our main model, overreaction is also u-shaped as a function of reputation. Finally, Figure 18c compares the extent of overreaction across priors. When the private sector's prior is uniform, uncertainty is maximal—every type is equally likely—and the central bank overreacts more than under the truncated normal prior with the same mean.



(a) Overreaction and ψ^{π} **(b)** Overreaction and $\mathbb{E}_t^P [\psi^{\pi}]$ **(c)** Normal vs Uniform priors

Figure 18: Overreaction in the Markov equilibrium

Figure 19 then plots the dynamics of reputation at $t = 1$. The pattern mirrors the main model: reputation improves when the central bank is perceived as dovish, and deteriorates when it is perceived as hawkish.

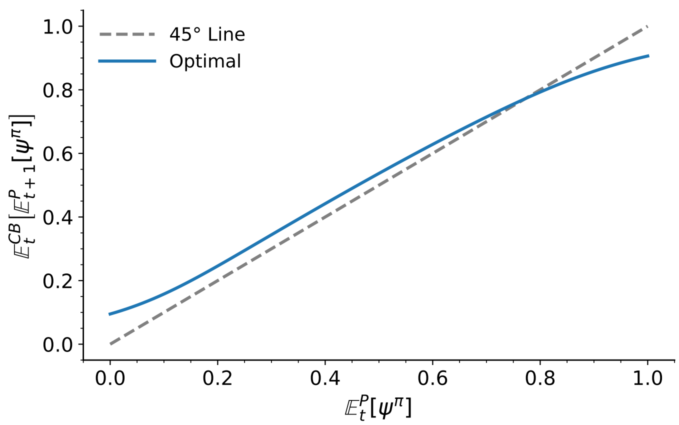


Figure 19: Reputation dynamics in the Markov equilibrium

Taken together, these figures show that the qualitative results of our economy do not change if we consider a Markov perfect equilibrium. Overreaction remains a robust feature, its magnitude varies with type and reputation in a u-shaped fashion, and the dynamics of credibility follow the same logic as in the baseline environment. Even though the equilibrium in this finite-horizon economy is unique, we cannot rule out the possibility of multiple equilibria in the infinite-horizon case. If the central bank's type λ changes over time, then the posterior distribution does not converge to a singleton around its true value, but rather to a stationary distribution. In that setting, if private sector beliefs are sufficiently rigid, there may exist a steady state with hawkish reputation and another with dovish reputation.

When reputation is dovish, the central bank lacks strong incentives to improve it; when it is hawkish, the cost of sustaining credibility is not high enough to induce deteriorating it. Multiple equilibria can therefore arise. In this sense, our assumption that the private sector believes the central bank is myopic can be viewed as an equilibrium refinement that restores uniqueness.

C.4 Zero Lower Bound

If the ZLB binds then

$$\tilde{y}_t + \lambda\kappa\tilde{\pi}_t = -\mu_t$$

Where μ_t is the shadow value of decreasing interest rates below 0. Not being able to decrease rates below zero leads to recession and, through the NKPC, deflation. In this environment, a dovish reputation can allow the central bank to get out of the ZLB. Recall the IS curve is given by

$$\tilde{y}_t = \mathbb{E}_t^P\left[\tilde{y}_{t+1}\right] - \frac{1}{\sigma}\left(\mathbb{E}_t^{CB}\left[\tilde{r}_t^n\right] - \mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right]\right)$$

Using (8) and (9)

$$\tilde{y}_t = \left(\frac{1}{\sigma}\mathbb{E}_t^P\left[\psi^\pi\right] - \mathbb{E}_t^P\left[\psi^y\right]\right)\mathbb{E}_t^P\left[X_{t+1}\right] - \frac{1}{\sigma}\mathbb{E}_t^{CB}\left[\tilde{r}_t\right]$$

Define $\Psi := \mathbb{E}_t^P\left[\psi^\pi\right] - \sigma\mathbb{E}_t^P\left[\psi^y\right])$, it is increasing in $\mathbb{E}_t^P\left[\psi^\pi\right]$. If the central bank has sufficiently dovish reputation then $\Psi > 0$, which can help the central bank get out of the ZLB in response to expansionary shocks. This echoes Krugman (1998), which proposes central banks to "credibly promisse to be irresponsible".

D Proofs and Robustness Checks for Empirical Section

D.1 Proof of Proposition 3

Proof. The estimand of $\gamma_{2,t}$ is

$$\gamma_{2,t} = \frac{Cov\left(\mathbb{E}_t^i\left[y_{t+k}\right], \mathbb{E}_t^i\left[\pi_{t+k}\right]\right)}{Var\left(\mathbb{E}_t^i\left[\pi_{t+k}\right]\right)} = -\frac{\mathbb{E}_t^P\left[\psi^y\right]\mathbb{E}_t^P\left[\psi^\pi\right]Var\left(\mathbb{E}_t^i\left[X_{t+k}\right]\right)}{\mathbb{E}_t^P\left[\psi^\pi\right]^2 Var\left(\mathbb{E}_t^i\left[X_{t+k}\right]\right)} = -\frac{\mathbb{E}_t^P\left[\psi^y\right]}{\mathbb{E}_t^P\left[\psi^\pi\right]} \quad \text{(D.1)}$$

Finally, from the definition of ψ^π and ψ^y we know that $\psi^y = \kappa^{-1}(1-\psi^\pi)$. Replacing into (D.1) completes the proof. ■

D.2 Demand Shocks

To allow for situations where the Divine Coincidence does not hold, suppose the central bank follows the triple mandate in (16)

$$\mathcal{W}_t = -\frac{1}{2}\mathbb{E}_t^{CB}\left[\sum_{s=0}^{\infty}\beta^s\left(\tilde{y}_{t+s}^2 + \lambda\tilde{\pi}_{t+s}^2 + \varphi i_{t+s}^2\right)\right]$$

Assume the private sector does not know λ or φ, but continues to assume the central bank is myopic. For simplicity, assume there are only demand shocks.[35] In this environment, the correlation between inflation and the output gap can be either positive or negative.

Lemma 7. *Suppose the central bank's preferences are the dual mandate from* (16) *and there are only demand shocks. Define* $\widehat{\psi} := \frac{\varphi\sigma}{1+\lambda\kappa^2+\sigma^2\varphi}$. *Then, the estimand of* $\gamma_{2,t}$ *in equation* (19) *can be positive or negative, and increasing in* $\mathbb{E}_t^P\left[\widehat{\psi}\right]$

Proof. The first-order condition with respect to output gap, $\tilde{y}_t$, of a central bank with parameters λ and φ is

$$\tilde{y}_t + \lambda\kappa\tilde{\pi}_t = \varphi\sigma i_t$$

Define $\widehat{\psi} = \frac{\varphi\sigma}{1+\lambda\kappa^2+\varphi\sigma^2}$, $\widehat{\psi}^y := \frac{\lambda\kappa}{1+\lambda\kappa^2+\varphi\sigma^2}$ and $\widehat{\psi}^\pi := \frac{1}{1+\lambda\kappa^2+\varphi\sigma^2}$. Our new measure of reputation will be $\mathbb{E}_t^P\left[\widehat{\psi}\right]$. As elsewhere, a lower $\mathbb{E}_t^P\left[\widehat{\psi}\right]$ signals a more hawkish central bank. Taking expectations as given, the central bank implements

$$\begin{aligned}
\tilde{y}_t\left(\tilde{\lambda},\tilde{\varphi}\right) =& \widehat{\psi}\left(\tilde{\lambda},\tilde{\varphi}\right)\left(r_t^n + \sigma\mathbb{E}_t^P\left[\tilde{y}_{t+1}\right] + \mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right]\right) - \widehat{\psi}^y\left(\tilde{\lambda},\tilde{\varphi}\right)\beta\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right] \\
\tilde{\pi}_t\left(\tilde{\lambda},\tilde{\varphi}\right) =& \kappa\widehat{\psi}\left(\tilde{\lambda},\tilde{\varphi}\right)\left(r_t^n + \sigma\mathbb{E}_t^P\left[\tilde{y}_{t+1}\right] + \mathbb{E}_t\left[\tilde{\pi}_{t+1}\right]\right) + \widehat{\psi}^\pi\left(\tilde{\lambda},\tilde{\varphi}\right)\beta\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right]
\end{aligned}$$

We conjecture that the Central Bank follows a linear policy rule

$$\tilde{y}_t\left(\tilde{\lambda},\tilde{\varphi}\right) = \sum_{s=0}^{\infty}\Theta_s^y\left(\tilde{\lambda},\tilde{\varphi}\right)\mathbb{E}_t^i\left[\tilde{r}_{t+k+s}^n\right]$$

[35] In this setting, the assumption $\mathbb{E}_t^i[\eta_{t+k}] = \mathbb{E}_t^P[\eta_{t+k}]$ can also be read as no disagreement about future monetary policy shocks, as in Bauer et al. (2024).

$$\tilde{\pi}_t\left(\tilde{\lambda},\tilde{\varphi}\right)=\sum_{s=0}^{\infty}\Theta_s^{\pi}\left(\tilde{\lambda},\tilde{\varphi}\right)\mathbb{E}_t^i\left[\tilde{r}_{t+k+s}^n\right]$$

Matching coefficients for $k=0$

$$\Theta_0^y\left(\tilde{\lambda},\tilde{\varphi}\right)=\widehat{\psi}\left(\tilde{\lambda}\right)\qquad\Theta_0^{\pi}\left(\tilde{\lambda},\tilde{\varphi}\right)=\kappa\Theta_0^y\left(\tilde{\lambda},\tilde{\varphi}\right)$$

For $k=1$

$$\begin{aligned}\Theta_1^y\left(\tilde{\lambda},\tilde{\varphi}\right)=&\widehat{\psi}\left(\tilde{\lambda},\tilde{\varphi}\right)(\sigma+\kappa)\,\mathbb{E}_t^P\left[\widehat{\psi}\right]-\kappa\beta\widehat{\psi}^y\left(\tilde{\lambda},\tilde{\varphi}\right)\mathbb{E}_t^P\left[\widehat{\psi}\right]\\ \Theta_1^{\pi}\left(\tilde{\lambda},\tilde{\varphi}\right)=&\kappa\widehat{\psi}\left(\tilde{\lambda},\tilde{\varphi}\right)(\sigma+\kappa)\,\mathbb{E}_t^P\left[\widehat{\psi}\right]+\kappa\beta\widehat{\psi}^{\pi}\left(\tilde{\lambda},\tilde{\varphi}\right)\mathbb{E}_t^P\left[\widehat{\psi}\right]\end{aligned}$$

For an arbitrary s, we have

$$\Theta_s^y\left(\tilde{\lambda},\tilde{\varphi}\right)=\widehat{\psi}\left(\tilde{\lambda},\tilde{\varphi}\right)\left(\sigma\mathbb{E}_t^P\left[\Theta_{s-1}^y\right]+\mathbb{E}_t^P\left[\Theta_{s-1}^{\pi}\right]\right)-\widehat{\psi}^y\left(\tilde{\lambda},\tilde{\varphi}\right)\beta\mathbb{E}_t^P\left[\Theta_{s-1}^{\pi}\right]\qquad\text{(D.2)}$$

$$\Theta_s^{\pi}\left(\tilde{\lambda},\tilde{\varphi}\right)=\kappa\widehat{\psi}\left(\tilde{\lambda},\tilde{\varphi}\right)\left(\sigma\mathbb{E}_t^P\left[\Theta_{s-1}^y\right]+\mathbb{E}_t^P\left[\Theta_{s-1}^{\pi}\right]\right)+\widehat{\psi}^{\pi}\left(\tilde{\lambda},\tilde{\varphi}\right)\beta\mathbb{E}_t^P\left[\Theta_{s-1}^{\pi}\right]\qquad\text{(D.3)}$$

Then, k-period ahead forecast is given by

$$\begin{aligned}\mathbb{E}_t^i\left[y_{t+k}\right]&=\sum_{s=0}^{\infty}\mathbb{E}_t^P\left[\Theta_s^y\right]\mathbb{E}_t^i\left[\tilde{r}_{t+k+s}^n\right]+\mathbb{E}_t^P\left[\eta_{t+k}\right]\\ \mathbb{E}_t^i\left[\pi_{t+k}\right]&=\sum_{s=0}^{\infty}\mathbb{E}_t^P\left[\Theta_s^{\pi}\right]\mathbb{E}_t^i\left[\tilde{r}_{t+k+s}^n\right]+\kappa\mathbb{E}_t^P\left[\eta_{t+k}\right]\end{aligned}$$

The estimand of $\gamma_{2,t}$ from equation (19) is

$$\gamma_{2,t}=\frac{\sum_{s=0}^{\infty}\mathbb{E}_t^P\left[\Theta_s^{\pi}\right]\mathbb{E}_t^P\left[\Theta_s^y\right]Var\left(\mathbb{E}_t^i\left[\tilde{r}_{t+k+s}^n\right]\right)}{\sum_{s=0}^{\infty}\mathbb{E}_t^P\left[\Theta_s^{\pi}\right]\mathbb{E}_t^P\left[\Theta_s^{\pi}\right]Var\left(\mathbb{E}_t^i\left[\tilde{r}_{t+k+s}^n\right]\right)}$$

Assume that $Var\left(\mathbb{E}_t^i\left[\tilde{r}_{t+k+s}\right]\right)=\rho^s Var\left(\mathbb{E}_t^i\left[\tilde{r}_{t+k}\right]\right)$ with $\rho<1$. Implicitly, we assume that the private sector reaches consensus about the long run. Then, we can rewrite the estimand as

$$\gamma_{2,t}=\sum_{s=0}^{\infty}\omega_s\frac{\mathbb{E}_t^P\left[\Theta_s^y\right]}{\mathbb{E}_t^P\left[\Theta_s^{\pi}\right]}\qquad\omega_s=\frac{\mathbb{E}_t^P\left[\Theta_s^{\pi}\right]^2}{\sum_{s=0}^{\infty}\mathbb{E}_t^P\left[\Theta_s^{\pi}\right]^2}$$

Now we study the sign of $\gamma_{2,t}$ by studying each one of the terms in the sum. First,

$$\frac{\mathbb{E}_t^P\left[\Theta_0^y\right]}{\mathbb{E}_t^P\left[\Theta_0^\pi\right]} = \frac{1}{\kappa} > 0$$

which does not depend on the beliefs. This is the result of the contemporaneous effect of a demand shock that is not fully stabilized: an unit increase of output gap directly implies a contemporaneous increase in inflation of κ. For $k = 1$,

$$\frac{\mathbb{E}_t^P\left[\Theta_1^y\right]}{\mathbb{E}_t^P\left[\Theta_1^\pi\right]} = \frac{\left(1+\sigma\frac{\mathbb{E}_t^P\left[\Theta_{s-1}^y\right]}{\mathbb{E}_t^P\left[\Theta_{s-1}^\pi\right]}\right) - \beta\frac{\mathbb{E}_t^P\left[\widehat{\psi}^y\right]}{\mathbb{E}_t^P\left[\widehat{\psi}\right]}}{\kappa\left(1+\sigma\frac{\mathbb{E}_t^P\left[\Theta_{s-1}^y\right]}{\mathbb{E}_t^P\left[\Theta_{s-1}^\pi\right]}\right) + \beta\frac{\mathbb{E}_t^P\left[\widehat{\psi}^\pi\right]}{\mathbb{E}_t^P\left[\widehat{\psi}\right]}}$$

Observe that

$$\lim_{\lambda\to 0}\frac{\mathbb{E}_t^P\left[\widehat{\psi}^y\right]}{\mathbb{E}_t^P\left[\widehat{\psi}\right]} = 0 \quad \lim_{\lambda\to \infty}\frac{\mathbb{E}_t^P\left[\widehat{\psi}^y\right]}{\mathbb{E}_t^P\left[\widehat{\psi}\right]} \to \infty \quad \lim_{\lambda\to 0}\frac{\mathbb{E}_t^P\left[\widehat{\psi}^\pi\right]}{\mathbb{E}_t^P\left[\widehat{\psi}\right]} > 0 \quad \lim_{\lambda\to \infty}\frac{\mathbb{E}_t^P\left[\widehat{\psi}^\pi\right]}{\mathbb{E}_t^P\left[\widehat{\psi}\right]} > 0$$

Then, we have

$$\lim_{\lambda\to 0}\frac{\mathbb{E}_t^P\left[\Theta_1^y\right]}{\mathbb{E}_t^P\left[\Theta_1^\pi\right]} > \frac{\mathbb{E}_t^P\left[\Theta_0^y\right]}{\mathbb{E}_t^P\left[\Theta_0^\pi\right]} \qquad \lim_{\lambda\to \infty}\frac{\mathbb{E}_t^P\left[\Theta_1^y\right]}{\mathbb{E}_t^P\left[\Theta_1^\pi\right]} \to -\infty$$

With anticipated demand shocks there are two opposite effects. First, the expansionary effect of a demand shock that is not fully stabilized. Second, the increase in inflation expectations acts as a cost-push shock for inflation. When the Central Bank is dovish, the first effect dominates. When the Central Bank is hawkish, the second effect dominates.

This logic also holds for an arbitrary horizon s. To see this, use (D.2) and (D.3)

$$\frac{\mathbb{E}_t^P\left[\Theta_s^y\right]}{\mathbb{E}_t^P\left[\Theta_s^\pi\right]} = \frac{\left(1+\sigma\frac{\mathbb{E}_t^P\left[\Theta_{s-1}^y\right]}{\mathbb{E}_t^P\left[\Theta_{s-1}^\pi\right]}\right) - \beta\frac{\mathbb{E}_t^P\left[\widehat{\psi}^y\right]}{\mathbb{E}_t^P\left[\widehat{\psi}\right]}}{\kappa\left(1+\sigma\frac{\mathbb{E}_t^P\left[\Theta_{s-1}^y\right]}{\mathbb{E}_t^P\left[\Theta_{s-1}^\pi\right]}\right) + \beta\frac{\mathbb{E}_t^P\left[\widehat{\psi}^\pi\right]}{\mathbb{E}_t^P\left[\widehat{\psi}\right]}}$$

When $\lambda \to 0$ we have

$$\lim_{\lambda\to 0}\frac{\mathbb{E}_t^P\left[\Theta_s^y\right]}{\mathbb{E}_t^P\left[\Theta_s^\pi\right]} > \lim_{\lambda\to 0}\frac{\mathbb{E}_t^P\left[\Theta_{s-1}^y\right]}{\mathbb{E}_t^P\left[\Theta_{s-1}^\pi\right]}$$

and for $\lambda \to \infty$ we have

$$\lim_{\lambda \to \infty} \frac{\mathbb{E}_t^P \left[\Theta_s^y\right]}{\mathbb{E}_t^P \left[\Theta_s^\pi\right]} = \begin{cases} \frac{1}{\kappa}\left(1 + \frac{\beta}{\sigma}\right) & s \in \{2, 4, \ldots, 2n\} \\ -\infty & s \in \{1, 3, \ldots, 2n+1\} \end{cases}$$

Then, we conclude that $\gamma_{2t} > 0$ as $\lambda \to 0$, and $\gamma_{2t} \to -\infty$ as $\lambda \to \infty$. This completes the proof. ■

Lemma 7 extends the logic of our results to demand shocks. As in the main model, improvements in reputation are also captured by a smaller value of $\gamma_{2,t}$. Unlike the cost-push shock case, the sign of the correlation can flip. Contemporaneous demand shocks push inflation and the output gap in the same direction, yielding a positive correlation. In addition to this effect, anticipated demand shocks also act like cost-push shocks, raising inflation expectations and generating a negative comovement. When the central bank is perceived as dovish, the first effect dominates; when it is perceived as hawkish, the second one dominates. As in the cost-push shock case, a stronger reputation dampens the impact of news shocks.

This extension also modifies the reputation measure by incorporating φ, which captures gradualism. When the private sector believes the central bank is less willing to adjust rates (higher perceived φ), $\mathbb{E}_t^P\left[\widehat{\psi}\right]$ increases. For the perceived inflation-output trade-off, the logic remains the same: a higher perceived weight on inflation stabilization implies a more negative slope. If reputation improves, whether through a higher perceived weight on inflation or lower perceived gradualism, the estimand of $\gamma_{2,t}$ decreases.

D.3 Belief Heterogeneity

Suppose that beliefs over λ, μ_t^i, differ across agents, while still assuming that forecasts of shocks are independent of beliefs about the central bank's preferences.

Lemma 8. *When beliefs are heterogeneous and $Var\left(\mathbb{E}_t^i\left[\varepsilon_{t+k}\right]\right) = \phi^k Var\left(\mathbb{E}_t^i\left[\varepsilon_t\right]\right)$ with $\phi \leqslant 1$, then*

$$\gamma_{2,t} = -\overline{\mathbb{E}}\left[\omega^i \frac{\mathbb{E}_t^i\left[\psi^y\right]}{\mathbb{E}_t^i\left[\psi^\pi\right]}\right] \qquad \omega^i = \frac{\sum_{s=0}^{\infty}\left(\beta^s \mathbb{E}_t^i\left[\psi^\pi\right]^{s+1}\right)^2 \phi^s}{\overline{\mathbb{E}}\left[\sum_{s=0}^{\infty}\left(\beta^s \mathbb{E}_t^i\left[\psi^\pi\right]^{s+1}\right)^2 \phi^s\right]}$$

where $\overline{\mathbb{E}}\left[\cdot\right]$ denotes the cross-sectional mean. Then, $\gamma_{2,t}$ is increasing in $Var\left(-\frac{\mathbb{E}_t^i[\psi^y]}{\mathbb{E}_t^i[\psi^\pi]}\right)$.

Proof. Since the cross-sectional variation in beliefs about the Central Bank is orthogonal to the cross-sectional variation in forecasts of the shocks we have

$$\begin{aligned} Cov\left(\mathbb{E}_t^i\left[y_{t+k}\right], \mathbb{E}_t^i\left[\pi_{t+k}\right]\right) &= -Cov\left(\mathbb{E}_t^i\left[\psi^y\right]\sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^i\left[\psi^\pi\right]\right)^s\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right], \mathbb{E}_t^i\left[\psi^\pi\right]\sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^i\left[\psi^\pi\right]\right)^s\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]\right) \\ &= -\sum_{s=0}^{\infty}\beta^{2s}\overline{\mathbb{E}}\left[\mathbb{E}_t^i\left[\psi^\pi\right]^{2s+1}\mathbb{E}_t^i\left[\psi^y\right]\right]Var\left(\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]\right) \\ &= -\overline{\mathbb{E}}\left[\left(\sum_{s=0}^{\infty}\left(\beta^s\mathbb{E}_t^i\left[\psi^\pi\right]^{s+1}\right)^2 Var\left(\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]\right)\right)\frac{\mathbb{E}_t^i\left[\psi^y\right]}{\mathbb{E}_t^i\left[\psi^\pi\right]}\right] \end{aligned}$$

$$\begin{aligned} Var\left(\mathbb{E}_t^i\left[\pi_{t+k}\right]\right) =& Cov\left(\mathbb{E}_t^i\left[\psi^\pi\right]\sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^i\left[\psi^\pi\right]\right)^s\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right], \mathbb{E}_t^i\left[\psi^\pi\right]\sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^i\left[\psi^\pi\right]\right)^s\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]\right) \\ =& \sum_{s=0}^{\infty}\beta^{2s}\overline{\mathbb{E}}\left[\mathbb{E}_t^i\left[\psi^\pi\right]^{2s+2}Var\left(\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]\right)\right] \\ =& \overline{\mathbb{E}}\left[\left(\sum_{s=0}^{\infty}\left(\beta^s\mathbb{E}_t^i\left[\psi^\pi\right]^{s+1}\right)^2 Var\left(\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]\right)\right)\right] \end{aligned}$$

Then we have

$$\frac{Cov\left(\mathbb{E}_t^i\left[y_{t+k}\right], \mathbb{E}_t^i\left[\pi_{t+k}\right]\right)}{Var\left(\mathbb{E}_t^i\left[\pi_{t+k}\right]\right)} = -\overline{\mathbb{E}}\left[\omega^i\frac{\mathbb{E}_t^i\left[\psi^y\right]}{\mathbb{E}_t^i\left[\psi^\pi\right]}\right]$$

which depends on the horizon k. Under the assumption $Var\left(\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]\right) = \phi^s Var\left(\mathbb{E}_t^i\left[\varepsilon_{t+k}\right]\right)$ that expression does not depend on the horizon k and therefore $\gamma_{2,t} = -\overline{\mathbb{E}}\left[\omega^i\frac{\mathbb{E}_t^i[\psi^y]}{\mathbb{E}_t^i[\psi^\pi]}\right]$. Since $\overline{\mathbb{E}}\left[\omega^i\right] = 1$ we have

$$\gamma_{2,t} = -\overline{\mathbb{E}}\left[\frac{\mathbb{E}_t^i\left[\psi^y\right]}{\mathbb{E}_t^i\left[\psi^\pi\right]}\right] + Cov\left(\omega^i, -\frac{\mathbb{E}_t^i\left[\psi^y\right]}{\mathbb{E}_t^i\left[\psi^\pi\right]}\right) \tag{D.4}$$

Notice that ω^i is increasing in $\mathbb{E}_t^P\left[\psi^\pi\right]$, which implies $\gamma_{2,t}$ puts more weight on individuals for which the Central Bank has worse reputation. Then, from (D.4), $\gamma_{2,t}$ is biased towards worse reputation compared to the cross-sectional average reputation.

Focusing on the second term, $Cov\left(\omega^i, -\frac{\mathbb{E}_t^i[\psi^y]}{\mathbb{E}_t^i[\psi^\pi]}\right)$. It is positive, and increasing in the cross-sectional dispersion of the Cenral Bank's reputation. Then, when the disagreement between forecasters increases so does $\gamma_{2,t}$. ■

Lemma 8 shows that with heterogeneous beliefs, the estimand γ_{2t} is a weighted average of the individual estimands $\gamma_{2t}^i = -\frac{\mathbb{E}_t^i[\psi^y]}{\mathbb{E}_t^i[\psi^\pi]}$. Formally,

$$\gamma_{2,t} = -\overline{\mathbb{E}}\left[\frac{\mathbb{E}_t^i\left[\psi^y\right]}{\mathbb{E}_t^i\left[\psi^\pi\right]}\right] + Cov\left(\omega^i, -\frac{\mathbb{E}_t^i\left[\psi^y\right]}{\mathbb{E}_t^i\left[\psi^\pi\right]}\right)$$

The second term is positive, so belief heterogeneity biases γ_{2t} upward relative to the cross-sectional mean. Disagreement creates an additional channel through which γ_{2t} can change: a reduction in the disagreement about the central bank's preferences lowers γ_{2t}. If we interpret cross-sectional disagreement as different draws from the same prior beliefs, μ_t, then this result is consistent with Lemma 6.

D.4 Model Heterogeneity

Suppose instead that each forecaster i relies on a different model. In particular, each forecaster uses a different value of the slope of the NKPC, κ_i. While beliefs about the central bank's preference λ remain homogeneous, the derived beliefs over $\psi^\pi = \frac{1}{1+\kappa^2\lambda}$ are heterogeneous.

Under this assumption, Lemma 8 holds: the estimand of $\gamma_{2,t}$ is a weighted average of the individual estimands, $\gamma_{2,t}^i = -\frac{\mathbb{E}_t^i[\psi^y]}{\mathbb{E}_t^i[\psi^\pi]}$, placing more weight on forecaster with a flatter NKPC. Since beliefs over λ are still homogeneous, the comparative statics from Lemma 5 and Lemma 6 continue to hold. A shift in beliefs toward a more hawkish central bank makes each $\gamma_{2,t}^i$ more negative, and therefore their weighted average, $\gamma_{2,t}$, shifts in the same direction.

D.5 Optimal Policy

A central bank of type λ has policies

$$y_t\left(\lambda, \left\{\mathbb{E}_t^{CB}\left[\varepsilon_{t+s}\right]\right\}_{t+s}\right) \quad \pi_t\left(\lambda, \left\{\mathbb{E}_t^{CB}\left[\varepsilon_{t+s}\right]\right\}_{t+s}\right)$$

Forecasters differ in their forecasts of shocks, $\{\mathbb{E}_t^i\left[\varepsilon_{t+s}\right]\}$ and have the same beliefs over the central bank's type, $\mu_t\left(\lambda\right)$. Assume further that, conditional on the type, forecasts are 'certainty equivalent', namely

$$\begin{aligned}
\mathbb{E}_t^i\left[y_{t+k}\left(\lambda\right)\right] =& \mathbb{E}_t^i\left[y_{t+k}\left(\lambda, \left\{\mathbb{E}_{t+s}^{CB}\left[\varepsilon_{t+k+s}\right]\right\}_{s\geqslant 0}\right)\right] = y_{t+k}\left(\lambda, \left\{\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]\right\}_{s\geqslant 0}\right) \\
\mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\right] =& \mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda, \left\{\mathbb{E}_{t+s}^{CB}\left[\varepsilon_{t+k+s}\right]\right\}_{s\geqslant 0}\right)\right] = \pi_{t+k}\left(\lambda, \left\{\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]\right\}_{s\geqslant 0}\right)
\end{aligned}$$

Then

$$\mathbb{E}_t^i\left[y_{t+k}\right] = \int_0^\infty \mathbb{E}_t^i\left[y_{t+k}\left(\lambda\right)\right]\mu_t\left(\lambda\right)d\lambda$$

$$\mathbb{E}_t^i\left[\pi_{t+k}\right]=\int_0^\infty \mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\right]\mu_t\left(\lambda\right)d\lambda$$

We estimate the following regression in the cross-section

$$\mathbb{E}_t^i\left[y_{t+k}\right]=\gamma_{1t}+\gamma_{2t}\mathbb{E}_t^i\left[\pi_{t+k}\right]+u_{t,k}^i$$

Then

$$\gamma_{2,t}=\frac{Cov\left(\mathbb{E}_t^i\left[y_{t+k}\right],\mathbb{E}_t^i\left[\pi_{t+k}^i\right]\right)}{Var\left(\mathbb{E}_t^i\left[\pi_{t+k}\right]\right)}$$

We proceed with each term separately

$$\begin{aligned}
Cov\left(\mathbb{E}_t^i\left[y_{t+k}\right],\mathbb{E}_t^i\left[\pi_{t+k}\right]\right) &= \int_i \mathbb{E}_t^i\left[y_{t+k}\right]\mathbb{E}_t^i\left[\pi_{t+k}\right]dF_i-\int_i \mathbb{E}_t^i\left[y_{t+k}\right]dF_i\int_i \mathbb{E}_t^i\left[\pi_{t+k}\right]dF_i \\
&= \int_i\left(\int_0^\infty \mathbb{E}_t^i\left[y_{t+k}\left(\lambda\right)\right]\mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\right]\mu_t\left(\lambda\right)d\lambda\right)dF_i<0 \\
Var\left(\mathbb{E}_t^i\left[\pi_{t+k}\right]\right) &= \int_i \mathbb{E}_t^i\left[\pi_{t+k}\right]^2 dF_i \\
&= \int_i\left(\int_0^\infty \mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\right]^2\mu_t\left(\lambda\right)d\lambda\right)dF_i>0
\end{aligned}$$

Then

$$\gamma_{2,t}=\int_i \alpha_t^i\left(\int_0^\infty \omega_t^i\left(\lambda\right)\frac{\mathbb{E}_t^i\left[y_{t+k}\left(\lambda\right)\right]}{\mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\right]}d\lambda\right)dF_i$$

where

$$\omega_t^i\left(\lambda\right)=\frac{\mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\right]^2\mu_t}{\int_0^\infty \mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\right]^2\mu_t d\lambda}\qquad \alpha_t^i=\frac{\int_0^\infty \mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\right]^2\mu_t d\lambda}{\int_i\left(\int_0^\infty \mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\right]^2\mu_t d\lambda\right)dF_i}$$

Fix i, $\frac{\mathbb{E}_t^i[y_{t+k}(\lambda)]}{\mathbb{E}_t^i[\pi_{t+k}(\lambda)]}$ is negative, and the magnitude decreases with λ: a central bank with higher priority for inflation stability. Now let's focus on

$$\omega_t^i\left(\lambda\right)=\frac{\mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\right]^2\mu_t}{\int_0^\infty \mathbb{E}_t^i\left[\pi_{t+k}\left(\lambda\right)\mu_t\right]^2\mu_t d\lambda}$$

Suppose $\mu_t' \gtrsim_{\text{FOSD}} \mu_t$, then

$$\int_0^x \mu_t'(\lambda)\,d\lambda \leqslant \int_0^x \mu_t(\lambda)\,d\lambda.$$

Since $\mathbb{E}_t^i[\pi_{t+k}(\lambda)]^2$ is decreasing in λ, we can integrate by parts to obtain

$$\int_0^\infty \mathbb{E}_t^i[\pi_{t+k}(\lambda)]^2\,\mu_t'(\lambda)\,d\lambda = \int_0^\infty \left(-\frac{d\,\mathbb{E}_t^i[\pi_{t+k}(\lambda)]^2}{d\lambda}\right)\left(\int_0^\lambda \mu_t'(s)\,ds\right)d\lambda$$

$$\leqslant \int_0^\infty \left(-\frac{d\,\mathbb{E}_t^i[\pi_{t+k}(\lambda)]^2}{d\lambda}\right)\left(\int_0^\lambda \mu_t(s)\,ds\right)d\lambda = \int_0^\infty \mathbb{E}_t^i[\pi_{t+k}(\lambda)]^2\,\mu_t(\lambda)\,d\lambda.$$

Define

$$\omega_t^i(\lambda) := \frac{h_t^i(\lambda)\,\mu_t(\lambda)}{\int_0^\infty h_t^i(s)\,\mu_t(s)\,ds}, \qquad \omega_t^{i'}(\lambda) := \frac{h_t^i(\lambda)\,\mu_t'(\lambda)}{\int_0^\infty h_t^i(s)\,\mu_t'(s)\,ds},$$

and the ratio

$$R^i(\lambda) := \frac{\omega_t^{i'}(\lambda)}{\omega_t^i(\lambda)} = c\,\frac{\mu_t'(\lambda)}{\mu_t(\lambda)}, \qquad c := \frac{\int_0^\infty h_t^i(s)\,\mu_t(s)\,ds}{\int_0^\infty h_t^i(s)\,\mu_t'(s)\,ds} < 1.$$

Since

$$\int_0^\infty \omega_t^{i'}(\lambda)\,d\lambda = \int_0^\infty \omega_t^i(\lambda)\,d\lambda = 1,$$

it follows that $R^i(\lambda)$ must be greater than one for some values of λ and smaller than one for others.

Now, assume $\mu_t' \gtrsim_{\text{MLR}} \mu_t$ (which is stronger than FOSD). Then $R^i(\lambda)$ is increasing in λ. Because both weights integrate to one, they must cross. Taken together, $R^i(\lambda)$ crosses exactly once (from below to above).

Hence, if $\mu_t' \gtrsim_{\text{MLR}} \mu_t$, then

$$\omega_t^{i'} \gtrsim_{\text{FOSD}} \omega_t^i, \qquad \text{i.e.} \qquad \int_0^x \omega_t^{i'}(\lambda)\,d\lambda \leqslant \int_0^x \omega_t^i(\lambda)\,d\lambda.$$

Now let's focus on

$$\gamma_{2t}^i = \int_0^\infty \omega_t^i(\lambda)\,\frac{\mathbb{E}_t^i[y_{t+k}(\lambda)]}{\mathbb{E}_t^i[\pi_{t+k}(\lambda)]}d\lambda = \int_0^\infty \left(1-\int_0^\lambda \omega_t^i(x)\,d\lambda\right)\left(-\frac{d}{d\lambda}\frac{\mathbb{E}_t^i[y_{t+k}(\lambda)]}{\mathbb{E}_t^i[\pi_{t+k}(\lambda)]}\right)d\lambda$$

Then, since $\frac{\mathbb{E}_t^i[y_{t+k}(\lambda)]}{\mathbb{E}_t^i[\pi_{t+k}(\lambda)]}$ is negative and its magnitude increasing in λ, we have that if $\mu_t' \gtrsim_{MLR} \mu_t$ then $\gamma_{2t}^{i'} < \gamma_{2t}^i$. Finally, we have

$$\gamma_{2t} = \int_i \alpha_t^i\gamma_{2t}^i dF_i = \mathbb{E}_t\left[\gamma_{2,t}^i\right] + Cov_t\left(\alpha_t^i, \gamma_{2t}^i\right)$$

where the second term is positive: forecasters who forecasts larger shocks are given more weight (larger α_t^i), and also they have a worse estimate of reputation, as the allocation becomes less dependent on the central bank's preferences and more on the economic conditions (smaller magnitude of γ^i_{2t}). However, with better reputation, shocks have a smaller pass-through on forecasts. Therefore, when $\mu'_t \gtrsim_{MLR} \mu_t$, under some general conditions, γ_{2t} becomes more negative.

D.6 Fed Information Effect

For robustness, we follow Bauer and Swanson (2022) and Bauer and Swanson (2023), and orthogonalize the shocks with respect to the public information that became available between FOMC meetings. In particular, we orthogonalize the shocks using the same six variables in Bauer and Swanson (2022): Nonfarm payrolls surprise, employment growth, change in the S&P 500, change in the slope of the Yield curve, change in Commodity prices, and implied skewness of the ten-year Treasury yield.

Figure 20 plots the impulse response to an orthogonalized monetary tightening. The effect prevails.

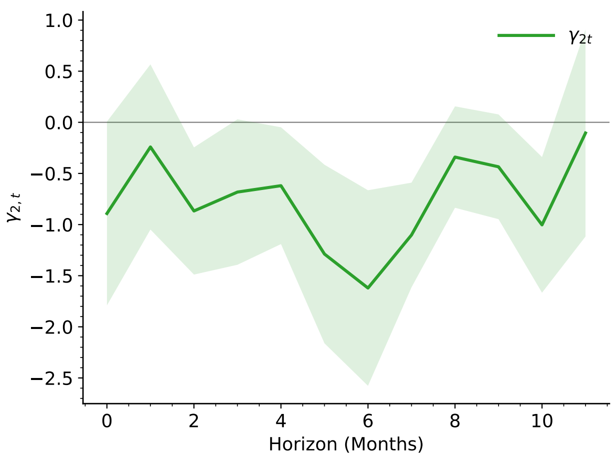


Figure 20: Impulse response to an orthogonalized monetary tightening.
Shaded areas denote 68% confidence intervals.

D.7 Survey of Professional Forecasters

We repeat the empirical exercise using the Survey of Professional Forecasters (SPF), which collects individual forecasts at the quarterly frequency. In particular, we collect the quarterly estimates of $\gamma_{2,t}$ from (19). The correlation coefficient between both series is 0.60, showing that both series capture similar time-series variation in the private sector's beliefs.[36] Figure 21 plots the estimates from both datasets. Qualitatively, both series look similar after the dot-com bubble.

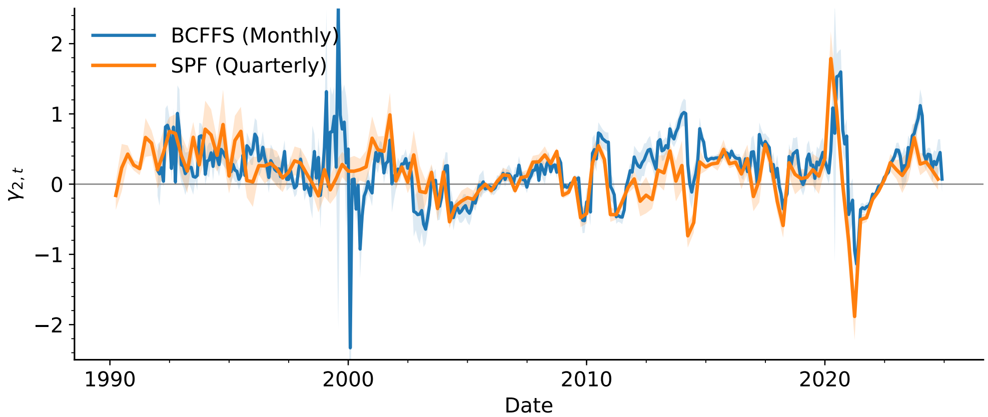


Figure 21: Estimates from SFP and BCFF.

?? plots the impulse response to an oil price news cost-push shock from Känzig (2021). Consistent with the estimates from BCFF, a more hawkish reputation dampens the pass-through of cost-push shocks on inflation expectations. However, the point estimates suggest that when reputation is high, a positive cost-push shock may even reduce inflation expectations, which is implausible. We interpret this as evidence that while the qualitative pattern—stronger reputation leads to lower pass-through—is robust across datasets, the quantitative magnitudes in SPF should be interpreted with caution.

Figure 23 plots the impulse response to an unexpected monetary tightening. We obtain a quarterly time series for monetary surprises by aggregating the monthly series. We find no effect of a monetary policy shock on reputation. This contrasts with the significant effect found in the estimates using BCFF data. One potential explanation is that aggregating monthly monetary surprises to the quarterly frequency may obscure the learning dynamics documented in the main text, particularly if belief updating occurs primarily in response to individual FOMC announcements rather than cumulative quarterly policy changes.

Taken together, these findings support that the facts we document are consistent across datasets, and time-varying perceptions about the central bank's hawkishness are present in the data.

[36] We take the quarterly mean of the monthly estimates from BCFF.

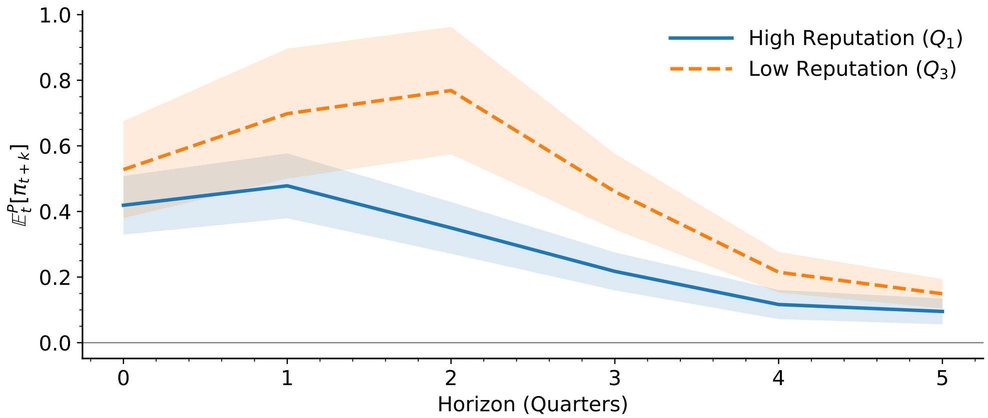


Figure 22: Impulse response of inflation expectations to an oil price news shock for high and low reputation.
Shaded areas denote 95% confidence intervals.

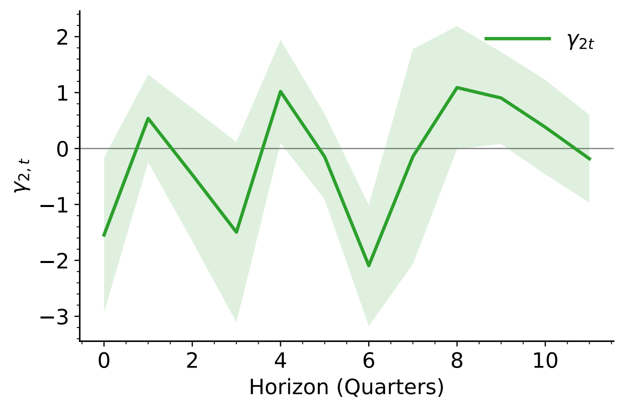


Figure 23: Impulse response to a monetary tightening.
Shaded areas denote 68% confidence intervals.

D.8 Proof of Lemma 3

Proof. For simplicity, suppose there is no disagreement about the forecast of demand. Taking expectations as exogenous, the Central Bank implements

$$\tilde{y}_t = \psi^\pi y_{t-1} - \psi^y \left(\varepsilon_t + \beta \mathbb{E}_t^P \left[\tilde{\pi}_{t+1}\right]\right)$$
$$\tilde{\pi}_t = \kappa \psi^\pi y_{t-1} + \psi^\pi \left(\varepsilon_t + \beta \mathbb{E}_t^P \left[\tilde{\pi}_{t+1}\right]\right)$$

Conjecture the Central Bank follows a linear rule

$$\begin{aligned}\tilde{y}_t =& \varphi^y y_{t-1} - \alpha\psi^y \sum_{s=0}^{\infty} \left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s \mathbb{E}_t\left[\varepsilon_{t+s}\right] \\ \tilde{\pi}_t =& \varphi^\pi y_{t-1} + \alpha\psi^\pi \sum_{s=0}^{\infty} \left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s \mathbb{E}_t\left[\varepsilon_{t+s}\right]\end{aligned}$$

Replacing into both equations

$$\begin{aligned}&\varphi^\pi y_{t-1} + \alpha\psi^\pi \sum_{s=0}^{\infty} \left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s \mathbb{E}_t\left[\varepsilon_{t+s}\right] \\ &\quad = \kappa\psi^\pi y_{t-1} + \psi^\pi \sum_{s=0}^{\infty} \left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s \mathbb{E}_t\left[\varepsilon_{t+s}\right] + \psi^\pi\beta\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right] y_{t-1} - \alpha\psi^\pi\beta\mathbb{E}_t^P\left[\varphi^\pi\psi^y\right] \sum_{s=0}^{\infty} \left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right) \mathbb{E}_t\left[\varepsilon_{t+s}\right]\end{aligned}$$

$$\begin{aligned}&\varphi^y y_{t-1} - \alpha\psi^y \sum_{s=0}^{\infty} \left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s \mathbb{E}_t\left[\varepsilon_{t+s}\right] \\ &\quad = \psi^\pi y_{t-1} - \psi^y \sum_{s=0}^{\infty} \left(\alpha\beta\mathbb{E}\left[\psi^\pi\right]\right)^s \mathbb{E}_t\left[\varepsilon_{t+s}\right] - \psi^y\beta\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right] y_{t-1} + \alpha\psi^y\beta\mathbb{E}_t^P\left[\varphi^\pi\psi^y\right] \sum_{s=0}^{\infty} \left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s \mathbb{E}_t\left[\varepsilon_{t+s}\right]\end{aligned}$$

Matching coefficients yields

$$\begin{aligned}\alpha =& \frac{1}{1+\beta\mathbb{E}^P\left[\varphi^\pi\psi^y\right]} \\ \varphi^\pi =& \left(\kappa + \beta\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right]\right)\psi^\pi \\ \varphi^y =& \left(1 - \beta\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right]\right)\psi^y\end{aligned}$$

Notice that $\alpha < 1$. From the equations for φ^π and φ^y, $\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right]$ is pinned down by

$$\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right] = \left(\kappa + \beta\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right]\right)\left(1 - \beta\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right]\right)\mathbb{E}_t^P\left[\psi^\pi\psi^y\right]$$

We conjecture φ^π and φ^y are always positive. This pins down a unique solution for $\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right]$. Defining the right hand side as a polynomial in $\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right]$, it has one positive and one negative root. In addition, it is positive at zero. Since the left hand side is a linear function, it follows there is a unique value $\mathbb{E}_t^P\left[\varphi^\pi\varphi^y\right]$ such that the equality holds. Plugging into our expression for φ^π and φ^y completes the proof. ■

D.9 Proof of Lemma 4

Proof. For simplicity, suppose there is no disagreement about the forecast of demand. Taking expectations as exogenous, the Central Bank implements

$$\begin{aligned}\tilde{y}_t =& -\psi^y\left(\varepsilon_t + \beta\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right] + \delta\pi_{t-1}\right)\\ \tilde{\pi}_t =& \psi^\pi\left(\varepsilon_t + \beta\mathbb{E}_t^P\left[\tilde{\pi}_{t+1}\right] + \delta\pi_{t-1}\right)\end{aligned}$$

Conjecture the Central Bank follows a linear rule

$$\begin{aligned}\tilde{y}_t =& -\varphi^y\pi_{t-1} - \alpha\psi^y\sum_{s=0}^{\infty}\left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s\mathbb{E}_t\left[\varepsilon_{t+s}\right]\\ \tilde{\pi}_t =& \varphi^\pi\pi_{t-1} + \alpha\psi^\pi\sum_{s=0}^{\infty}\left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s\mathbb{E}_t\left[\varepsilon_{t+s}\right]\end{aligned}$$

Replacing into both equations

$$\begin{aligned}&\varphi^\pi\pi_{t-1} + \alpha\psi^\pi\sum_{s=0}^{\infty}\left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s\mathbb{E}_t\left[\varepsilon_{t+s}\right]\\ &\quad = \psi^\pi\delta\pi_{t-1} + \psi^\pi\sum_{s=0}^{\infty}\left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s\mathbb{E}_t\left[\varepsilon_{t+s}\right] + \psi^\pi\beta\mathbb{E}_t^P\left[\varphi^\pi\varphi^\pi\right]\pi_{t-1} + \alpha\psi^\pi\beta\mathbb{E}_t^P\left[\varphi^\pi\psi^\pi\right]\sum_{s=0}^{\infty}\left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)\mathbb{E}_t\left[\varepsilon_{t+s}\right]\end{aligned}$$

$$\begin{aligned}&-\varphi^y\pi_{t-1} - \alpha\psi^y\sum_{s=0}^{\infty}\left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s\mathbb{E}_t\left[\varepsilon_{t+s}\right]\\ &\quad = -\psi^y\delta\pi_{t-1} - \psi^y\sum_{s=0}^{\infty}\left(\alpha\beta\mathbb{E}\left[\psi^\pi\right]\right)^s\mathbb{E}_t\left[\varepsilon_{t+s}\right] - \psi^y\beta\mathbb{E}_t^P\left[\varphi^\pi\varphi^\pi\right]\pi_{t-1} - \alpha\psi^y\beta\mathbb{E}_t^P\left[\varphi^\pi\psi^\pi\right]\sum_{s=0}^{\infty}\left(\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s\mathbb{E}_t\left[\varepsilon_{t+s}\right]\end{aligned}$$

Matching coefficients yields

$$\begin{aligned}\alpha =& \frac{1}{1-\beta\mathbb{E}_t^P\left[\varphi^\pi\psi^\pi\right]}\\ \varphi^\pi =& \left(\delta + \beta\mathbb{E}_t^P\left[\varphi^\pi\varphi^\pi\right]\right)\psi^\pi\\ \varphi^\pi =& \left(\delta + \beta\mathbb{E}_t^P\left[\varphi^\pi\varphi^\pi\right]\right)\psi^y\end{aligned}$$

Therefore we have $\alpha > 1$. From the second and third equation, notice that $\varphi^\pi = \phi\psi^\pi$ and $\varphi^y = \phi\psi^y$. Solving for ϕ yields

$$\phi = \delta + \beta\mathbb{E}_t^P\left[\psi^\pi\psi^\pi\right]\phi^2$$

We can rewrite this expression to obtain

$$\phi = \frac{\delta}{1 - \beta \mathbb{E}_t^P \left[\varphi^\pi \psi^\pi\right]} = \alpha\delta > \delta$$

Then, we can find α as follows

$$\alpha = 1 + \beta \mathbb{E}_t^P \left[\psi^\pi \psi^\pi\right] \delta \alpha^2$$

There are two solutions, but only one of them ensures $\alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right] < 1$, which is needed for a well-defined policy function. To see this, define $\tilde{\alpha} := \alpha\beta\mathbb{E}_t^P\left[\psi^\pi\right]$. It solves

$$\tilde{\alpha} = \beta \mathbb{E}_t^P \left[\psi^\pi\right] + \delta \frac{\mathbb{E}_t^P \left[\psi^\pi \psi^\pi\right]}{\mathbb{E}_t^P \left[\psi^\pi\right]} \tilde{\alpha}^2$$

This equation has two solutions $0 < \tilde{\alpha}_1 < 1 < \tilde{\alpha}_2$. We pick the first one to ensure the policy functions are well-defined. This completes the proof. ■

D.10 Perceived Taylor Rule Coefficients

We consider two perceived Taylor rules. The first regresses forecast interest rates on forecast inflation:

$$\mathbb{E}_t^i \left[i_{t+k}\right] = \alpha_{1t}^\pi + \alpha_{2t}^\pi \mathbb{E}_t^i \left[\pi_{t+k}\right] + u_{i,t+k} \tag{D.5}$$

The second regresses forecast interest rates on forecast output:

$$\mathbb{E}_t^i \left[i_{t+k}\right] = \alpha_{1t}^y + \alpha_{2t}^y \mathbb{E}_t^i \left[y_{t+k}\right] + u_{i,t+k} \tag{D.6}$$

We establish two results relating these coefficients to reputation.

Lemma 9. *The perceived Taylor rule coefficient on inflation $\alpha_{2,t}^\pi$ is U-shaped in $\mathbb{E}_t^P\left[\psi^\pi\right]$: it is decreasing when reputation is strong (low $\mathbb{E}_t^P\left[\psi^\pi\right]$) and increasing when reputation is weak (high $\mathbb{E}_t^P\left[\psi^\pi\right]$).*

Proof. The estimate of the slope of the Taylor Rule is given by

$$\alpha_{2,t}^\pi = \frac{Cov\left(\mathbb{E}_t^i\left[i_{t+k}\right], \mathbb{E}_t^i\left[\pi_{t+k}\right]\right)}{Var\left(\mathbb{E}_t^i\left[\pi_{t+k}\right]\right)}$$

The IS curve implies the following equilibrium behavior of the policy rate

$$\begin{aligned}\mathbb{E}_t^i\left[i_{t+k}\right] =&\mathbb{E}_t^i\left[r_{t+k}^n\right]+\mathbb{E}_i^P\left[\pi_{t+k+1}\right]+\sigma\left(\mathbb{E}_t^i\left[y_{t+k+1}\right]-\mathbb{E}_t^i\left[y_{t+k}\right]\right)\\ =&\mathbb{E}_t^P\left[r_{t+k}^n\right]+\mathbb{E}_t^P\left[\psi^\pi\right]\mathbb{E}_t^P\left[X_{t+k+1}\right]-\sigma\mathbb{E}_t^P\left[\psi^y\right]\left(\mathbb{E}_t^P\left[X_{t+k+1}\right]-\mathbb{E}_t^P\left[X_{t+k}\right]\right)\end{aligned}$$

For simplicity assume that inflation forecasts are not correlated with forecasts of the natural rate, we have

$$\begin{aligned}Cov\left(\mathbb{E}_t^i\left[i_{t+k}\right],\mathbb{E}_t^i\left[\pi_{t+k}\right]\right) =& \left(\mathbb{E}_t^P\left[\psi^\pi\right]\right)^2 Cov\left(\mathbb{E}_t^i\left[X_{t+k+1}\right],\mathbb{E}_t^i\left[X_{t+k}\right]\right)\\ &-\sigma\mathbb{E}_t^P\left[\psi^y\right]\mathbb{E}_t^P\left[\psi^\pi\right]\left(Cov\left(\mathbb{E}_t^i\left[X_{t+k+1}\right],\mathbb{E}_t^i\left[X_{t+k}\right]\right)-Var\left(\mathbb{E}_t^i\left[X_{t+k}\right]\right)\right)\end{aligned}$$

Since $\mathbb{E}_t^i\left[X_{t+k}\right]=\sum_{s=0}^{\infty}\left(\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^s\mathbb{E}_t^i\left[\varepsilon_{t+k+s}\right]$ then we have

$$\begin{aligned}Var\left(\mathbb{E}_t^i\left[X_{t+k}\right]\right) =&Var\left(\mathbb{E}_t^i\left[\varepsilon_{t+k}\right]\right)+\left(\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)^2 Var\left(\mathbb{E}_t^i\left[X_{t+k+i}\right]\right)\\ Cov\left(\mathbb{E}_t^i\left[X_{t+k+1}\right],\mathbb{E}_t^i\left[X_{t+k}\right]\right) =&\beta\mathbb{E}_t^P\left[\psi^\pi\right]Var\left(\mathbb{E}_t^i\left[X_{t+k+1}\right]\right)\end{aligned}$$

because we assumed reputation was constant between forecasters, so $\mathbb{E}_t^P\left[\psi^\pi\right]$ is a constant. Assume there is no heteroskedasticity in the dispersion of forecasts, so that $Var\left(\mathbb{E}_t^i\left[\varepsilon_{t+k}\right]\right)$ does not depend on the horizon k. Therefore, we have

$$Cov\left(\mathbb{E}_t^i\left[X_{t+k+1}\right],\mathbb{E}_t^i\left[X_{t+k}\right]\right)=\beta\mathbb{E}_t^P\left[\psi^\pi\right]Var\left(\mathbb{E}_t^i\left[X_{t+k}\right]\right)$$

Then, we have

$$\begin{aligned}Cov\left(\mathbb{E}_t^i\left[i_{t+k}\right],\mathbb{E}_t^i\left[\pi_{t+k}\right]\right) =& \left(\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)\left(\mathbb{E}_t^P\left[\psi^\pi\right]\right)^2 Var\left(\mathbb{E}_t^i\left[X_{t+k}\right]\right)\\ &+\sigma\mathbb{E}_t^P\left[\psi^y\right]\mathbb{E}_t^P\left[\psi^\pi\right]\left(1-\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)Var\left(\mathbb{E}_t^i\left[X_{t+k}\right]\right)\end{aligned}$$

And the estimand of the taylor rule coefficient is

$$\alpha_{2,t}^\pi=\beta\mathbb{E}_t^P\left[\psi^\pi\right]+\sigma\frac{\mathbb{E}_t^P\left[\psi^y\right]}{\mathbb{E}_t^P\left[\psi^\pi\right]}\left(1-\beta\mathbb{E}_t^P\left[\psi^\pi\right]\right)=\beta\left(1+\frac{\sigma}{\kappa}\right)\mathbb{E}_t^P\left[\psi^\pi\right]+\frac{\sigma}{\kappa}\frac{1}{\mathbb{E}_t^P\left[\psi^\pi\right]}-\frac{\sigma}{\kappa}\left(1+\beta\right)$$

The first term is increasing in $\mathbb{E}_t^P\left[\psi^\pi\right]$, whereas the second one is decreasing. Then, $\alpha_{2,t}$ is strictly convex and U-shaped.

Under heteroskedascitiy of forecasts errors, the dispersion of the forecast changes with the horizon. This adds an extra term to the estimand of $\alpha_{2,t}^\pi$ that varies both because of the change in reputation and because of changes in the dispersion of short-term forecasts; further complicating the structural interpretation of the Taylor rule coefficient. ■

Lemma 10. *In the economy of Appendix C.1, where the central bank has a triple mandate, prices are fixed, and only demand shocks are present, the perceived Taylor rule coefficient on output is*

$$\alpha^{y}_{2,t} = \frac{1}{\mathbb{E}^P_t\left[\psi^y\right]} - \sigma \geqslant 0$$

and is increasing in reputation.

Proof. Under these assumptions, the $k-$periods ahead forecast of output gap and interest rates is

$$\mathbb{E}^i_t\left[y_{t+k}\right] = \mathbb{E}^P_t\left[\psi^y\right]\mathbb{E}^i_t\left[X_{t+k}\right] \tag{D.7}$$

$$\mathbb{E}^i_t\left[i_{t+k}\right] = \mathbb{E}^P_t\left[\psi^i\right]\mathbb{E}^i_t\left[X_{t+k}\right] \tag{D.8}$$

In this model, the assumption is that there is no disagreement about the future monetary policy surprises. Suppose we follow Bauer et al. (2024) and estimate the perceived Taylor-rule coefficient

$$\mathbb{E}^i_t\left[i_{t+k}\right] = \alpha^{\pi}_{1,t} + \alpha^{\pi}_{2,t}\mathbb{E}^i_t\left[y_{t+k}\right]$$

Then the coefficient becomes

$$\alpha^{\pi}_{2,t} = \frac{1}{\mathbb{E}^P_t\left[\psi^y\right]} - \sigma \geqslant 0$$

which is increasing in the central bank's reputation. ■

E Details for the Quantitative Exercise

This section collects the technical details of the estimation procedure.

E.1 Simulated Method of Moments (SMM)

We estimate the model by SMM. For a candidate parameter vector θ, we simulate $S = 10{,}000$ paths of the myopic model, each of length matching the data sample, and compute the cross-simulation average of each targeted moment. The estimator solves

$$\hat{\theta} = \arg\min_{\theta}\ \frac{1}{2}\,\mathbf{g}(\theta)'\,W\,\mathbf{g}(\theta), \tag{E.1}$$

where $\mathbf{g}(\theta)$ is the vector of moment deviations $g_j(\theta) = \bar{m}_j(\theta) - m_j^{\text{data}}$, $\bar{m}_j(\theta)$ is the cross-simulation mean of moment j, and m_j^{data} is the corresponding data moment. The weight matrix W is computed once from the data (see below) and held fixed throughout the optimization.

The minimization is performed using a constrained interior-point algorithm (`fmincon` in MATLAB), initialized at multiple random starting points to guard against local minima. Parameter bounds are enforced throughout. All stochastic processes are discretized via the Rouwenhorst method on finite grids, ensuring that the simulated Markov chains preserve the unconditional variance and persistence of the underlying AR(1) processes.

E.2 Bootstrap weight matrix

The weight matrix W is constructed from a moving-block bootstrap that preserves the serial dependence in the quarterly data. We draw $B = 10{,}000$ overlapping blocks of length $\ell = \lfloor T^{1/3} \rfloor$ quarters from the joint (y_t, π_t, i_t) vector and recompute all targeted moments on each bootstrap sample. The resulting bootstrap variance–covariance matrix $\hat{\Sigma}$ yields two natural choices:

$$W_{\text{diag}} = \text{diag}\big(1/\hat{\sigma}_1^2, \ldots, 1/\hat{\sigma}_J^2\big), \qquad W_{\text{opt}} = \hat{\Sigma}^{-1}, \tag{E.2}$$

where $\hat{\sigma}_j$ is the bootstrap standard error of moment j. The diagonal matrix W_{diag} gives each moment influence proportional to its precision in the data, providing transparent, scale-free weighting. The optimal matrix W_{opt} achieves asymptotic efficiency but can be sensitive to estimation of the off-diagonal terms, particularly with a small number of moments. We use W_{diag} as our baseline.

Importantly, the bootstrap respects the fixed-point demeaning convention described in Section 6.1: each resampled block is demeaned using the same long-run constants from the 1960–2024 sample, not the bootstrap-sample mean. This ensures that the estimated sampling variability of the moments corresponds to the moments actually targeted in the SMM objective.

E.3 Simulation design and burn-in

Each evaluation of the SMM objective simulates S independent paths of the discretized model. Each path is initialized at the candidate parameter vector's initial conditions (or at the ergodic distribution for the Great Moderation specification) and run for $T_{\text{burn}} + T$

periods, where $T_{\text{burn}} = 200$ quarters and T matches the data sample length. The first T_{burn} periods are discarded.

The burn-in serves to eliminate the dependence of simulated moments on initial conditions. Without a burn-in, the simulated paths would reflect transient dynamics as the Markov chains converge to their ergodic distributions, biasing the cross-simulation averages away from the model's stationary moments. This is particularly important for the belief state $\overline{\psi}_t$: since the Kalman gain ω_t depends on the history of signals, starting all simulations at the same initial prior would generate artificial cross-simulation correlation in the early periods. The 200-quarter burn-in ensures that, by the time the sample window begins, each simulation has had roughly 50 years to "forget" its starting point. A fixed random seed (`rng(1234)`) guarantees that the myopic and optimal models face identical shock realizations, so that any difference in outcomes is attributable solely to the policy rule.

F Details and Extensions for the Delegation Exercise

F.1 Calibration

We benchmark against the optimal policy of a utilitarian planner with $\lambda = 10$, a standard value in the literature (see Galí 2003). The model is calibrated at a quarterly frequency. The discount factor is $\beta = 0.99$; and the intertemporal elasticity of substitution is $\sigma = 1$. The cost-push shock follows an AR(1) process with persistence $\rho = 0.90$ and standard deviation $\sigma_u = 0.2$. Finally, we arbitrarily set $\sigma_\eta = 1$. Table F.1 summarizes the full set of parameter values.

Table F.1: Calibrated Parameter Values

Category	**Parameter**	**Value**	**Description / Target**
Preferences	β	0.99	Subjective discount factor
Policy Preferences	λ	10	Weight on inflation in loss function
Shocks	ρ	0.9	Persistence of cost-push shock
	σ_u	0.2	Std. dev. of cost-push shock
	σ_η	1	Std. dev. of forecast error

To avoid beliefs collapsing into a mass point, we assume the precision τ is constant.

Then, the evolution of the private sector's beliefs is dictated by

$$\overline{\psi}_{t+1} = \omega_t X_t^{-1}\left(\tilde{\pi} + \kappa\eta_t\right) + \left(1-\omega_t\right)\overline{\psi}_t \quad \text{where} \quad \omega_t = \frac{\left(\kappa^{-1}X_t\right)^2 \tau_\eta}{\tau + \left(\kappa^{-1}X_t\right)^2 \tau_\eta}$$

F.2 Robustness

The optimal policy requires precise knowledge of the model and how the private sector learns. In practice, central banks face substantial uncertainty along both dimensions. We now examine whether hawkish delegation remains effective under various forms of misspecification.

Consider first a central bank that computes $\tilde{\lambda}$ assuming parameters (κ_0, τ_0). Keeping this $\tilde{\lambda}$ fixed, we simulate economies whose true parameters (κ, τ) deviate from the baseline. We measure welfare losses relative to the optimal policy under the true parameters. This exercise reveals how the performance of the hawkish myopic delegate deteriorates when the model used by the bank differs from the economy's actual structure.

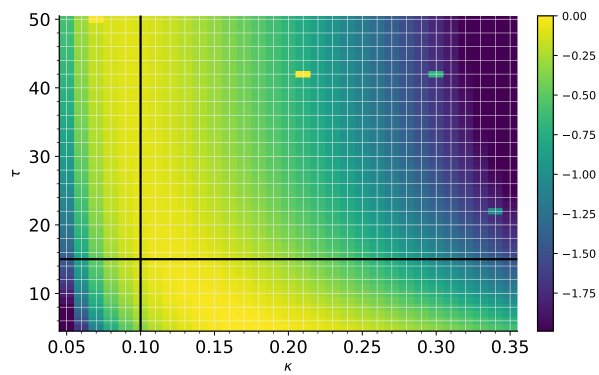


Figure 24: Welfare losses $\frac{W - W_{true}^{opt}}{\left|W_{true}^{opt}\right|}$ when $\tilde{\lambda}$ is fixed but true parameters vary

Figure 24 plots the welfare loss of the hawkish myopic delegate, relative to the optimal policy. Welfare losses can be high if the central bank both underestimates the precision of the private sector's beliefs, τ, and the slope of the NKPC, κ.

Next, consider a central bank that has the incorrect model of how the private sector learns. In particular, the central bank thinks τ has a value, but the true value is different.

Figure 25 plots the relative welfare of the hawkish myopic policy and the optimal policy when the central bank has an incorrect value of τ for different values of κ. The black horizontal line displays the value of τ used by the central bank. For relatively small values κ, the

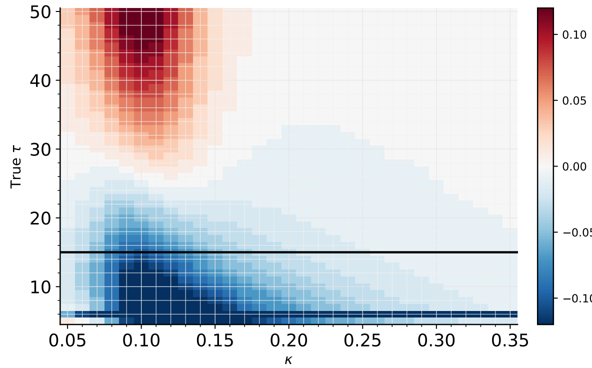
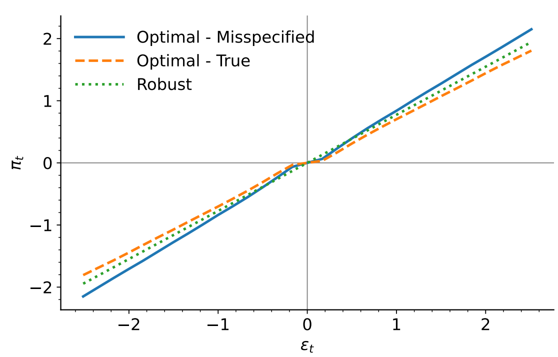
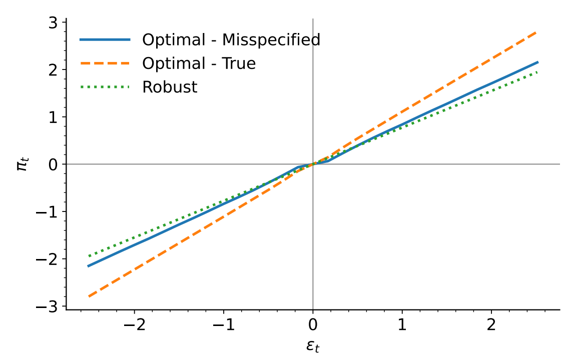


Figure 25: Welfare losses $\frac{W - W^{opt}_{misspecified}}{|W^{opt}_{misspecified}|}$ when τ is misspecified

hawkish myopic policy can reduce the welfare losses when the central bank underestimates the true value of τ, but amplifies them when it overestimates it. Figure 26 shows that when the central bank underestimates the actual value of τ it is not reacting enough to shocks. Then, appointing a hawkish delegate who reacts more can reduce welfare losses. However, for the same reason, the delegate amplifies losses when the central bank overestimates τ.[37]



(a) $\tau < \tau^{True}$



(b) $\tau > \tau^{True}$

Figure 26: Policy Functions under Misspecification of τ ($\kappa = 0.13$)

Similarly, suppose the central bank has the incorrect model of the economy. In particular, assume that the value of κ, the slope of the NKPC, is different from its true value.

Figure 27 plots the relative welfare of the hawkish myopic policy and the optimal policy when the central bank has an incorrect value of κ for different values of τ. The black vertical line displays the value of κ used by the central bank. The hedging benefit of the hawkish

[37]Explore the mechanisms behind this result in Appendix E.

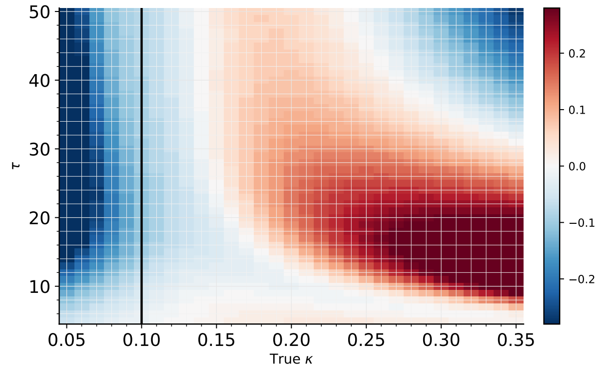


Figure 27: Welfare losses $\frac{W - W^{opt}_{misspecified}}{|W^{opt}_{misspecified}|}$ when κ is misspecified

myopic policy appears quite knife-edge: the policy only dominates when the true value of κ is substantially larger than the central bank's belief.[38]

Finally, we consider simultaneous misspecification of both κ and τ. Figure 28 reveals that both effects are present. Importantly, while the hedging benefit against underestimating κ alone appeared knife-edge, the interaction between the two sources of misspecification amplifies this effect. Now, the hawkish myopic policy now dominates across a broader region.

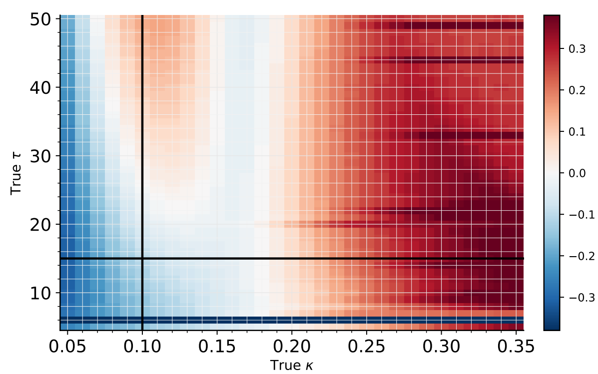


Figure 28: Welfare losses $\frac{W - W^{opt}_{misspecified}}{|W^{opt}_{misspecified}|}$ when κ and τ are misspecified

[38]We also consider the case where the central bank has the incorrect value of the IES, σ. Here, the optimal policy always outperforms the hawkish delegate.