Essays in Monetary Policy With Informational Frictions

Essays on Monetary Policy with Informational Frictions

A dissertation presented

by

David Choi

to

The Committee for the PhD in Business Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

April 2018

Dissertation Advisors:
Professor N. Gregory Mankiw
Professor Benjamin Friedman

Author:
David Choi

Essays on Monetary Policy with Informational Frictions

Abstract

This dissertation presents three essays addressing the role of informational frictions in monetary policy. In particular, motivated by the observation that central banks around the world are often concerned about their reputation for knowledgeability, I study why this might be the case and how this affects how central banks set monetary policy. In the first essay, "*Perceptions of Competence: Monetary Policy and the Reputational Accelerator,*" I characterize the role that perceptions of a central bank's knowledgeability play in affecting monetary policy's ability to stabilize aggregate outcomes. In the second essay, "*Monetary Policy Reversal Aversion,*" I analyze the monetary policy distortions that might arise when central banks care about their reputation for knowledgeability. The third and last essay, "*Perceptions of Central Bank Knowledgeability and the Signaling Channel: An Empirical Analysis*," empirically tests for the novel mechanisms outlined in the first essay, providing evidence that perceptions of central bank knowledgeability can affect the transmission of monetary policy. Together, these essays shed light on the unanswered questions of how the public's uncertainty of the central bank's ability to read and understand the economy affects aggregate economic outcomes and the way central banks set monetary policy, and why central banks might care about their reputation for knowledgeability in the first place.

Contents

List of Tables

List of Figures

Acknowledgments

I am grateful to my advisers Emmanuel Farhi, Benjamin Friedman, Xavier Gabaix, and N. Gregory Mankiw for their invaluable guidance and support. Greg and Ben welcomed me into their offices despite my rather late discovery of the fascinating world of macroeconomics, and my meandering around without a dissertation topic. Emmanuel and Xavier joined later still, but they too made me feel welcome, and together the suggestions and comments from my advisers improved the dissertation greatly. I would also like to thank David Scharfstein, who checked up on me periodically throughout graduate school, and Robin Greenwood, who went out of his way to help me and provide encouragement, and who gave me the motivation and prodding I needed to keep going forward.

The most meaningful part of graduate school has been the wonderful friendships I've made. Beyond the stimulating intellectual discussions with classmates, my fondest memories of graduate school will always be spending time with incredible friends who made the Ph.D. experience so worthwhile. I give special thanks to Ben and Vivek for all the memorable times spent together as roommates, and for their continued friendship after life made us move out. I give most special thanks to Stephanie for her love, support, and patience, and for never failing to cheer me up when graduate school took all of its twists and turns.

Finally, I am extremely thankful for the support of my family. My loving grandparents raised me as a young infant and cheered for me halfway across the world. My late grandfather always wanted me to pursue a Ph.D., and I thought about him constantly as I wrote this dissertation. I hope I have made him proud. Jane, my amazing big sister who I will always look up to, has taken care of me from the day I was born (the pictures prove it!). Our bond is inseparable. My mother sacrificed her dreams to allow me to pursue mine, and no matter how difficult the times, did everything she could possibly do to give me the best opportunities in life. My father passed on his intellectual curiosity and love of economics, and has been my mentor and role model all my life. My parents have put up with everything and given me everything, and I dedicate my dissertation to them.

To my parents

Introduction

> "There is an absolute yawning gap between the general perception of non-economist outsiders that reversals of policy, changes of mind, are to be deplored and castigated, as evidence of error, irresolution and general incompetence, and the apparent findings from our economic models"
>
> **- Charles Goodhart, former member of the Bank of England Monetary Policy Committee**

Around the world, central banks appear to care greatly about their reputation for competence and knowledgeability. One can find many statements from policymakers, past and present, emphasizing the importance that central banks place in guarding their reputation for near omniscience, and central banks are often reluctant to admit having made any type of error. Despite this, existing academic work has very little to say about why this might be the case. Our models and intuition might suggest that if a central bank has imperfect knowledge, it would be good if the public knew the extent to which the central bank lacks information, yet monetary authorities do not appear to readily admit this. Why might this be the case? Why do central banks care about their reputation for knowledgeability, and how is this reputation determined? How might this affect both the way monetary policy is set as well as equilibrium outcomes?

In this dissertation I make progress on these questions in three essays on the role of informational frictions in monetary policy. In the first essay, "*Perceptions of Competence: Monetary Policy and the Reputational Accelerator,*" I develop a general equilibrium macroeconomic model that characterizes the role that perceptions of a central bank's knowledgeability play in affecting monetary policy's ability to stabilize aggregate outcomes, providing a reason for why a purely benevolent central bank might care about its reputation for knowledgeability. In the second essay, "*Monetary Policy Reversal Aversion,*" I show that if a central bank does indeed care about its reputation for knowledgeability, this can lead to distortions in how monetary policy is set, such as excess reversal aversion and conservatism. The third and last essay, "*Perceptions of Central Bank Knowledgeability and the Signaling Channel: An Empirical Analysis*," empirically tests for the novel mechanisms outlined in the first essay, relating publicly announced forecasts previously made by a central bank with heterogeneity in the strength of the signaling channel in monetary policy, providing evidence that perceptions of central bank knowledgeability can affect the transmission of monetary policy. Together, these essays shed light on the unanswered questions of how the public's uncertainty of the central bank's ability to read and understand the economy affects aggregate economic outcomes and the way central banks set monetary policy, and why central banks might care about their reputation for knowledgeability in the first place.

In the first essay, I develop a new macroeconomic model in which perceptions of a central bank's competence affect monetary policy's ability to stabilize aggregate outcomes. The central bank receives imperfect signals on the state of the economy, with the precision of its signals private information to the central bank. A reputation for competence is formed rationally by agents in equilibrium through the observation of aggregate outcomes. I show that regardless of the true underlying knowledgeability of the central bank, the mere *perception* that the central bank is knowledgeable helps improve equilibrium outcomes, through a coordination channel and confidence channel that alleviates informational externalities generated by the dispersion of information in the economy. The endogenous formation of perceptions amplifies the effects of monetary errors; favorable monetary outcomes endogenously lead to more favorable outcomes in the future, and vice-versa, an effect I call the *reputational accelerator*. The model also exhibits excess output volatility, and endogenously generates large downside tail risk, particularly after long periods of stability. Such central bank 'Minsky moments' arise from the decoupling of equilibrium responses to signals from their true informative content. Breaking this sharp link also generates a time-varying inefficiency wedge between the equilibrium and efficient responses to public signals, adding to the theory of dispersed information.

In the second essay, I build a game-theoretic reputation model in which a central bank cares about its reputation for knowledgeability. I show that because more knowledgeable central banks are less likely to reverse course in equilibrium, reversal aversion can arise because frequent reversals might show that the central bank has likely erred, damaging its reputation for knowledgeability.

Two distortions arise, relative to a frictionless world in which the knowledgeability of the central bank is perfectly known by the public. First, the central bank may not reverse course even when doing so is appropriate given the fundamentals of the economy. Second, foreseeing the possibility of wanting to reverse in the future, and thus having to choose between taking a reputational hit or setting an inappropriate interest rate, a central bank may choose to wait until more uncertainty is resolved before changing the interest rate. Taken together, these two distortions, which I call *reversal aversion* and *conservatism*, lead to central banks not reversing their rates frequently, and also reacting slowly ("falling behind the curve") to shocks. These are both observed looking at central banks around the world, but are hard to explain jointly in existing theoretical frameworks. A key novel result is that conservatism arises from the desire to protect the central bank's reputation for knowledgeability, which has not yet been explored in the context of an economic model.

In the third essay, I use data on individual level private forecasts to find empirical support for the model laid out in the first essay. Focusing on the U.S. and the Federal Reserve, I show that information effects — the provision of private information about the state of the economy through monetary actions — appear substantially stronger when publicly announced forecasts made by the Federal Reserve were more accurate in the recent past, a key prediction made by the model. I find not only larger changes in mean forecasts (following a monetary shock) after a period of more accurate Federal Reserve public forecasts, but a larger reduction in the variance of forecasts as well. In addition, by tracking the same forecasters before and after a monetary shock, I show heterogeneity in information effects across forecasters, depending on the relative position of a private forecast to that of the Federal Reserve's internal forecast prior to the monetary shock. The nature of this heterogeneity is shown to be predicted by dispersed information models. Thus, there appears to be heterogeneity in information effects cross-sectionally, consistent with dispersed information models, and heterogeneity across time depending on the public's perception of the central bank's knowledgeability, a unique prediction of the model presented in the first essay.

Together, these essays fill the gap between existing academic work and the observed behavior of central banks, by first providing a reason for *why* central banks are concerned about their reputation for knowledgeability, and also showing that a few key consistently observed patterns in how monetary policy is conducted around the world, such as reversal aversion and conservatism, can be explained by central banks that are concerned about their reputation for knowledgeability.

Chapter 1

Perceptions of Competence: Monetary Policy and the Reputational Accelerator

1.1 Introduction

Central banks are often concerned about their reputation for competence. For example, monetary authorities appear to go to great lengths to avoid admitting error, in an attempt to cultivate an aura of near omniscience.[1] In speeches, op-eds, and papers, numerous current and former policymakers point to the reason why: a common theme is that the mere *perception* of having erred leads to a reputational cost that affects the central bank's ability to effectively stabilize the economy going forward.[2] Under this view, a central bank's failure to achieve satisfactory aggregate outcomes consequently leads not only to short-term welfare loss, but long-term welfare costs coming from reputational damage as well.

Despite this apparent concern of policymakers, the reputation for competence — defined here as the ability to correctly infer the fundamentals of the economy — has remained mostly unexplored in

[1] Alan Greenspan in particular created what has been described as the "cult of omniscience" of central banking. The recent financial crisis and subsequent recession has done much damage to the reputation of central bankers and to the notion of central banks' superiority of knowledge.

[2] This reputational effect is often described in the context of policy reversal aversion. The following quotes illustrate this fear: "A series of poor published forecasts might undermine the bank's credibility, and thereby hamper its ability to conduct policy" " —Alan Blinder, former vice chairman of the Fed.

"Policymakers might be reluctant to reverse course in that doing so would damage their reputation, perhaps because the public would lose confidence in the central bank's ability to understand and stabilize the economy" —Charles Evans, Federal Reserve Bank of Chicago President, with coauthors.

existing academic work. Models involving reputational concerns of the central bank instead focus on cases where the monetary authority may have different preferences than that of the public,[3] or cases where central banks develop a reputation for following a good policy rule, e.g., resisting the temptation to generate surprise inflation[4] or following the Taylor Principle.[5] However, even with benevolent intentions, the history of central banking includes many mistakes and blunders, not due to any deliberate intent, but rather from the general inability to correctly read the state of the economy. A central bank may *want* to prescribe the socially optimal policy action, but not know what exactly this right action is. In this paper, I explore this new approach to reputation: the central bank's ability to correctly read the economy's state is unknown to private agents, and a reputation for competence or incompetence is built from observed aggregate outcomes.

I develop a model in which a central bank's competence (ability) is defined as the precision of the imperfect signals it receives on the state of the underlying economy. This precision is time-varying, and private information to the central bank. Naturally, a central bank receiving more precise signals is more likely to achieve desirable aggregate outcomes, given its better information about the state of the economy. Perceptions about the central bank are thus endogenously determined: when aggregate outcomes are relatively favorable, agents believe the central bank to be relatively competent, and vice-versa when outcomes are poor. More surprisingly, I show in the model that regardless of the *true* competence of the central bank, social welfare increases from the *perception* that the central bank receives precise signals of the economy's shocks. Thus, good aggregate outcomes, leading to an upward revision of beliefs about the central bank's ability, result in persistently higher expected welfare; similarly, bad aggregate outcomes lead to persistently worse expected outcomes, through this perceptions channel. These endogenously formed beliefs, which amplify monetary policy's effects, are the key driving force in the model, which I call the *reputational accelerator*.

Taking a standard workhorse model of monopolistic competition with informational frictions, I show how the dynamics of monetary shocks are altered when the central bank's latent ability is imperfectly known by firms, providing a reason for central banks to care deeply about their

[3] See Backus and Driffill (1985) and Cukierman and Meltzer (1986)

[4] See Barro and Gordon (1983a; 1983b) and Barro (1986)

[5] See Hansen and McMahon (2016), Eusepi and Preston (2010), Bianchi and Melosi (2016)

reputation for superior knowledge. In a model with dispersed information throughout the economy — the setting I study in this paper — monetary policy provides a public signal about the economy's fundamentals, in additional to its traditional effects. In its attempt to offset shocks in order to stabilize output, the central bank's policy provides valuable information used to determine a firm's optimal price. In the dispersed information macroeconomic model I build on, the relative price distortions that arise from dispersed information lead to an inefficient allocation of resources and insufficient price coordination in equilibrium, stemming from an informational externality that creates a wedge between the socially optimal level of coordination and the private incentive to coordinate. When the central bank is simply *perceived* as more competent, the central bank's actions serve as a stronger coordination device, reducing disagreement and thus reducing inefficient price dispersion. This *coordination channel*, which brings the equilibrium closer to an appropriately defined constrained efficient benchmark, is the first channel by which a reputation for competence leads to better monetary relevant outcomes. The second channel, which I call the *confidence channel*, comes from the self-fulfilling nature of inflation when information is dispersed. If the central bank has a price level or inflation target (e.g., due to Calvo frictions), to achieve the desired target, it must set policy in such a way that firms are incentivized to set prices in line with that target. When firms perceive the central bank as competent, they believe the central bank knows exactly what this optimal policy is; thus, they trust that the policy set by the central bank is indeed such that their ex-post privately optimal price is consistent with the central bank's implicit target, leading to more price level stability and thus increased welfare.

In this framework, the ability to coordinate and dictate prices through policy actions are valuable tools in the central bank's arsenal. Yet the strength of these tools depend crucially on how informative firms believe the central bank's policy is in providing credible information on the underlying fundamentals of the economy. An error by the central bank — inferred from a subpar monetary welfare outcome — and the subsequent fallout in reputation leads to a weakening of these tools, and provides additional channels through which monetary shocks (errors), with their concomitant alteration of beliefs about the central bank's competence, lead to persistence and amplification of business cycles and a permanently altered path of expected welfare towards the stochastic steady state. Aggregate outcomes generated by monetary policy continually feed into the beliefs channel, sharpening or blunting the coordination and confidence tools at the central bank's disposal.

The core of this paper lies in the decoupling of equilibrium responses to signals from their true informative content. This model exhibits excess output volatility, relative to a benchmark model where the central bank's competence is fully observed, and endogenously generates large downside tail events, particularly after long periods of stability. Such central bank 'Minsky' moments where stability leads to instability arise when a central bank starts getting imprecise signals after having achieved a string of good monetary outcomes; because of the unobservability of the central bank's signal precision, firms continue to perceive the central bank as competent. While such beliefs increase expected welfare — due to the confidence and coordination channel — this leads to the risk of potentially very high levels of output volatility, and states of the world where welfare is very low, generating a fat downside tail of welfare outcomes and elevated risk. The breaking of the sharp link between agents' responses to public signals with the signals' true informative content also adds to the vast existing literature on dispersed information. While one key comparative static the literature has focused on involves the precision of the public signal, two separate effects are conflated: when the precision increases, one effect comes from the actual increase in informativeness of the public signal (facilitating information provision), whereas another effect comes from the increased equilibrium response to a more precise signal (vice-versa when the precision decreases). When the precision of the public signal is common knowledge, these two channels must move in perfect lockstep. By allowing the underlying precision to change, and making the change imperfectly observed, this paper distinguishes the two effects and quantifies their respective roles in affecting each macroeconomic aggregate outcome. The informational wedge — the difference between the socially optimal and equilibrium level of coordination — becomes time-varying, with the socially optimal level dependent on the true exogenously-changing underlying precision, and the equilibrium level dependent on the endogenously-changing beliefs. In contrast, this wedge is constant in the existing literature, with the wedge solely a function of time-invariant parameters.

The paper is organized as follows. Section 2 lays out the model and the informational structure of the economy. Section 3 characterizes the equilibrium conditions, showing the key optimality conditions and the resulting equilibrium properties. Section 4 illustrates the reputational accelerator and breaks down the dynamics of the model. Section 5 extends the model to allow for Calvo frictions, providing illustration of the mechanics of the confidence channel. Section 6 characterizes the equilibrium and efficient use of information, quantifying the time-varying inefficiency wedge. Section 7 concludes.

1.1.1 Related Literature

The role of informational frictions in generating real effects from monetary shocks was first explored in Phelps (1970) and Lucas (1972), in their since widely used model of island economies. Other early influential papers in this literature include King (1982), Townsend (1983), Phelps (1983), and Woodford (2003a), and recently, the literature on monetary shocks with imperfect information has taken off in many new directions: most similar to this paper, Hellwig (2005) analyzes a model of monopolistic competition with imperfect information about monetary shocks, Angeletos and La'O (2009) analyzes the the response of prices to nominal shocks in a variant of the Calvo model, and Baeriswyl and Cornand (2010) endogenizes the monetary shock by introducing a discretionally optimizing central bank, who also possesses imperfect information about the economy's fundamentals. For recent surveys on informational and coordination frictions in macroeconomics, see Mankiw and Reis (2010) and Angeletos and Lian (2016).

In practice, central banks make policy decisions with noisy information, and the role that imperfect knowledge plays in monetary policy has been explored in a wide variety of models.[6] The approach I take in modeling imperfect knowledge most closely follows Adam (2007) and Baeriswyl and Cornand (2010), who model a central bank receiving imperfect signals directly on variables that capture all the composite shocks of the economy. Orphanides (2003) and Hubert (2014, 2015) provide compelling empirical evidence that imperfect knowledge plays a large role in monetary policy; they show that central banks appear to make decisions with considerable uncertainty, and looking at both internal and externally published now-casts and forecasts, they show periods of both accurate and inaccurate information. This paper defines the *competence* of the central bank as the precision of the signals it receives on the fundamentals of the economy, a definition previously used by Moscarini (2007) and Frankel and Kartik (2016). Both papers analyze in a stylized game-theoretic setting a model where the competence of the central bank is known,

[6]This includes the analysis of the performance of simple rules (e.g., Orphanides and Williams (2002; 2007) and the discussion in Taylor and Williams (2010), linear-quadratic frameworks with forward looking variables (e.g., Aoki (2003; 2006) and Svensson and Woodford (2003; 2004)), and recently, the role incomplete information plays at the zero lower bound (e.g., Gust et al. (2015) and Evans et al. (2016)).

and show how differences in competence alter the credibility of monetary policy when the central bank faces time-inconsistency with an inflation bias. In contrast, I embed this notion of competence in a fully dynamic micro-founded model, where the competence is private information of the central bank, and where the central bank faces neither time inconsistency nor an inherent desire to push output above its natural level.

This paper also complements the theory literature on the role of public signals in games of dispersed information, starting with Morris and Shin (2002) and extended substantially by Angeletos and Pavan (2007). Many applications of this setting have subsequently been studied: investment complementarities, asset markets, fiscal policy, and monetary shocks are a few prominent applications.[7] While these papers all focus on the effects of imperfect knowledge, the precisions of the signals themselves are common knowledge; this paper adds to the literature by adding uncertainty to the precision of the public signal, endogenizing the evolution of beliefs about the precision, and showing how these beliefs affect equilibrium dynamics. The public signal in this model is the central bank's policy action; this ties into another literature which looks at the role of the central bank's policy instrument as a signal of the fundamentals of the economy, which has received much recent interest. Adam (2007), Baeriswyl and Cornand (2010), Roca (2010), Berkelmans (2011), Tang (2015), and Melosi (2017) all focus on how information on the shocks of the economy are conveyed to price-setting firms via the central bank's policy. Empirical support of the signaling channel is found in Romer and Romer (2000), Campbell et al. (2012), El-Shagi et al. (2014), and Nakamura and Steinsson (2017), and empirical support of central banks as a coordination device is found in Hubert (2014; 2015).

1.2 Model

I begin with the baseline dynamic general equilibrium model of Woodford (2003b), adding modifications to the information structure. I first consider a model where informational heterogeneity leads to incomplete nominal adjustment, and later add sticky-price frictions as in Calvo (1983) in section 5. I begin by outlining the basic setup, and then explain in detail the informational structure of the economy.

[7] See, e.g., Allen et al. (2006), Angeletos and Pavan (2004), Bacchetta and van Wincoop (2006)

1.2.1 Starting Elements

A representative household maximizes the expected discounted sum of utility given by

$$\mathbb{E}_0 \left\{ \sum_{t=0}^{\infty} \beta^t U_t \right\} = \mathbb{E}_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left[u\left(C_t; \xi_t\right) - \int_0^1 v\left(H_t(i); \xi_t\right) di \right] \right\} \tag{1.1}$$

subject to the budget constraint

$$P_t C_t + B_t \leq \int_0^1 \left[W_t(i) H_t(i) + \Pi_t(i) \right] di + (1 + R_t) B_{t-1} - T_t \tag{1.2}$$

where $0 < \beta < 1$ denotes the discount factor, $u(\cdot)$ and $v(\cdot)$ represent the utility from consumption and disutility from labor, respectively, C_t is a consumption index, $H_t(i)$ is the amount of labor supplied to firm i, and ξ_t is a vector of stochastic shocks. I assume that for each possible value of ξ, $u(\cdot; \xi)$ is concave and strictly increasing, and $v(\cdot; \xi)$ is convex and increasing. Aggregate consumption C_t is a Dixit-Stiglitz aggregator of the variety of goods indexed by i,

$$C_t \equiv \left[\int_0^1 C_t(i)^{\frac{\eta-1}{\eta}} d_i \right]^{\frac{\eta}{\eta-1}} \tag{1.3}$$

where $\eta > 1$ governs the elasticity of substitution across varieties. From the budget constraint, B_t denotes the nominal value of end of period financial assets, $W_t(i)$ is the wage given for labor in production of variety i, R_t is the return on the one-period bond, T_t is a lump-sum transfer to the government, $\Pi_t(i)$ denotes nominal profits of firm i, and P_t is a price index given by

$$P_t \equiv \left[\int_0^1 P_t(i)^{1-\eta} \right]^{\frac{1}{1-\eta}} \tag{1.4}$$

with the property that $\int P_t(i) C_t(i) di = P_t C_t$.

There is a continuum of monopolistically competitive firms indexed by $i \in [0, 1]$. Each firm hires firm-specific labor $H_t(i)$ and produces output according to the general production function

$$Y_t(i) = A_t F\left(H_t(i)\right)$$

where A_t denotes the stochastic economy-wide technological coefficient, and $F(\cdot)$ is an increasing concave function that satisfies the usual Inada conditions. Each firm i hires labor variety i in a competitive labor market,[8] and thus takes $W_t(i)$ as given. As the monopolistic supplier of good variety i, it chooses the price of good i, and given (1.3) and (1.4), in equilibrium the firm faces relative demand

$$Y_t(i) = Y_t \left(\frac{P_t(i)}{P_t} \right)^{-\eta} \tag{1.5}$$

The firm sets prices each period to maximize that period's expected profits, given by

$$\mathbb{E}_t \left(m(Y_t) \left[(1+\tau) P_t(i) Y_t(i) - W_t(i) Y_t(i) \right] \mid \mathcal{I}_t^i \right) \tag{1.6}$$

where $m(Y_t)$ denotes the stochastic discount factor used to weight profits in each state, τ denotes the time-invariant output subsidy, and $\mathcal{I}_t^i$ denotes the information set of firm i when choosing the price at time period t. I assume that the output subsidy τ fully offsets the distortion that arises from monopoly power, that is $\tau = 1/(\eta - 1)$, and that fiscal policy is fully Ricardian, simply taxing the consumer on a lump-sump basis to pay for the sales subsidy, so that $T_t = \tau \int P_t(i) Y_t(i) di$.

There is a monetary authority that controls nominal aggregate demand $Q_t = P_t C_t$ directly. This assumption can be micro-founded with a cash-in-advance constraint and a monetary authority that pays interest on money holdings.[9] Letting the central bank control nominal demand directly, rather than manipulating the interest rate, is done to sidestep issues related to price-level determinacy.[10] The central bank's objective is to maximize the utility of the representative household, and so society's preferences and that of the monetary authority are perfectly aligned. Finally, I assume that the central bank cannot credibly commit to future policy actions; thus, it sets nominal demand Q_t each period on a discretionary basis.

---

[8]Note that I assume that each producer is a wage-taker, despite the differentiation of various types of labor. One way to justify this is to the approach taken by Woodford (2003b) and assume a double continuum of differentiated goods indexed by (I, j), where the elasticity of substitution between any two goods remain η. One can assume that all goods with the same index I(say, an "industry") receive the same signals and change their prices at the same time (in the case of Calvo frictions). Assuming all firms in I are produced using the same type of labor (type I labor), it follows that each firm has the same degree of market power assumed in this model, but without any market power in the labor market. I take this approach for the remainder of the paper.

[9]This is the assumption made in Hellwig (2005), Angeletos and La'O(2009; 2012), among others. See Appendix A for a micro-foundation of this assumption.

[10]This is long literature starting from Sargent and Wallace (1975); see also Atkeson et al. (2010) and Cochrane (2011)

1.2.2 Optimization

Firms set prices at the beginning of each period, before aggregate uncertainty about the underlying shocks is realized. Appendix A shows that linearizing the first order condition for firm i yields an expression for the profit-maximizing price

$$p_t^i = \mathbb{E}_t \left[p_t + \zeta(y_t - y_t^N) \mid \mathcal{I}_t^i \right] \tag{1.7}$$

where lower case letters denote logs of upper case variables, e.g., $p_t^i = \log P_t(i)$. y_t^N denotes the log of the *natural rate of output*, which is defined in terms of model primitives in Appendix A; importantly, y_t^N is a function of purely real disturbances in $\overline{\xi} = (\xi, A_t)$, and is independent of monetary policy. An individual firm i's optimal price is thus a function of its expectation of the overall price level p_t as well as the *output gap* $\widetilde{y}_t \equiv y_t - y_t^N$. Movements in the output gap capture the movements in demand and cost conditions relevant for a given firm's optimal price, and each firm uses its information set to infer both the price level and the output gap, which together determine marginal costs. The sensitivity of the optimal price to the output gap is given by ζ, which is purely a function of model primitives, depending on the preferences of the representative household and the production function of the firms. Weak assumptions are made such that $\zeta < 1$, so that prices are strategic complements, although this model does permit cases in which $\zeta > 1$, in which cases prices are strategic substitutes.[11]

Plugging in $y_t = q_t - p_t$, one can rewrite (1.7) as

$$p_t^i = \mathbb{E}_t \left[(1-\zeta) p_t + \zeta(q_t - y_t^N) \mid \mathcal{I}_t^i \right] \tag{1.8}$$

Thus $(1-\zeta)$ is a measure of strategic complementarities present in the model. Using $p_t = \int p_t^j dj$ and iterating (1.8) forward yields

$$p_t^i = \sum_{k=0}^{\infty} (1-\zeta)^k \mathbb{E}_t \left[\overline{\mathbb{E}}^k \left[\zeta \left(q_t - y_t^N \right) \right] \middle| \mathcal{I}^i \right] \tag{1.9}$$

where average expectations and higher order expectations are defined as

$$\overline{\mathbb{E}}(\cdot) = \int_0^1 \mathbb{E}(\cdot \mid \mathcal{I}^i) di$$

[11] $\zeta > 0$ holds for all admissible values of the underlying parameters; for details, see Appendix A.

$$\overline{\mathbb{E}}^k(\cdot) = \overline{\mathbb{E}}\left[\overline{\mathbb{E}}^{k-1}(\cdot)\right]$$

The specific role higher order expectations play is discussed Appendix A; what matters crucially here is that each higher-order belief of the exogenous disturbances is used to determine a firm's optimal price, and each higher-order belief in turn depends on perceptions about the central bank's degree of knowledge of these disturbances, as will be made clear later.

The central bank is benevolent and maximizes the utility of the representative household. Appendix A derives the period welfare function, which can be written as a quadratic approximation to the representative household's utility

$$U_t = -\left[\widetilde{y}_t^2 + \kappa \mathcal{V}(p_t^i) + t.i.p. + \mathcal{O}\left(\| \widehat{Y}, \widetilde{\xi} \|^3\right)\right] \tag{1.10}$$

where $\kappa = \frac{\eta}{\zeta}$, $\mathcal{V}(p_t^i)$ denotes the cross-sectional variance of prices across all firms, and the last two terms denote terms independent of policy and higher-order terms. While various exogenous shocks affect welfare, the focus of this paper is on *monetary* welfare, that is, welfare affected by monetary policy. As standard, deviations of output from its natural level lead to gaps between the marginal rate of substitution and the marginal product of labor, leading to aggregate inefficiency. The adverse consequences of dispersed prices come from the misalignment of relative prices with relative productivities, leading to misallocation of labor across firms and thus aggregate TFP loss. The central bank's objective function to be maximized can be written as

$$W_t = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_t \left[U_t \mid \mathcal{I}_t^{CB}\right] \tag{1.11}$$

which the central bank maximizes with its information set in the current period.

A few things are worth pointing out at this stage. Monetary policy analysis often uses a quadratic loss function involving the squared output gap and *inflation*. This is often done in an ad-hoc manner, but also comes as the relevant welfare criterion in the standard New Keynesian model with Calvo frictions. However, welfare losses from inflation or deflation do not come inherently from movements in the price level, but rather from the price dispersion that arises from inflation if some firms are unable or unwilling to change their prices.[12] Here, price dispersion arises

[12] When Calvo frictions are the only nominal friction, all firms who change their price choose the same price, and thus the squared inflation term captures movements in the cross-sectional variance of prices.

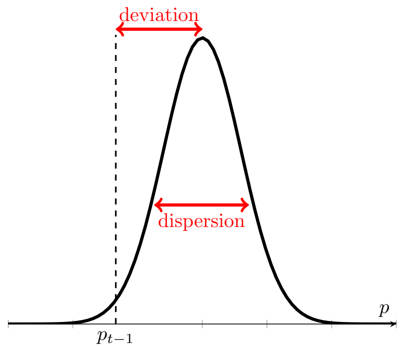


Figure 1.1: Components of cross-sectional price variability: (1) the price dispersion among firms changing their prices (due to dispersed information), and (2) movements in the price level when a fraction of firms do not change their prices (standard New Keynesian framework).

due to informational heterogeneity: firms all condition their prices on different information sets, naturally leading to price dispersion. I later combine this with Calvo frictions, in which case two sources of cross-sectional price variance arise each period: the dispersion among firms changing their prices in the current period, as well changes in the overall price level. Figure 1.1 provides an illustration.

1.2.3 Information Structure

Both the firms and the central bank get imperfect signals about the state of the economy. Because the demand and supply shocks in $\overline{\xi} = (\xi, A_t)$ are captured by the natural rate of output y_t^N, I model agents as receiving signals directly on y_t^N.[13] I depart from the existing literature by assuming the central bank can be of multiple types, representing the central bank's *competence.* In particular, a central bank of type θ receives a signal

$$S_t^\theta = y_t^N + \varepsilon_t^\theta$$

where $\theta \in \Theta = \{H, L\}$, and

$$\varepsilon_t^\theta \sim \mathcal{N}\left(0, \left(\tau^\theta\right)^{-1}\right)$$

[13]This can be micro-founded in a model where agents can choose which shocks to use their limited informational capacity on. While different explanations have been used to motivate informational frictions, this paper's main focus is not on these micro-foundations but rather its welfare effects when the degree of informational frictions is itself imperfect information.

where $\tau^H > \tau^L$. Thus, τ^θ denotes the precision of signals that a central bank of type θ receives, following the definition of central bank competence used by Walsh (1995), Moscarini (2007), and Frankel and Kartik (2016), among others.

The central bank's competence is time-varying, with the type θ of the central bank evolving as as an exogenous Markov chain with transition matrix given by

$$\Lambda = \begin{bmatrix} 1-\lambda & \lambda \\ \lambda & 1-\lambda \end{bmatrix}$$

where $\lambda < 1/2$ denotes the probability of a regime shift. Because $\lambda < 1/2$, the state variable θ is persistent, with this persistence decreasing in λ. Importantly, I also assume that the central bank's precision is *private information* of the central bank. The central bank observes past aggregate outcomes and its own private signals, and thus its information set is

$$\mathcal{I}_t^{CB} = \left\{\theta_{t-k}, S_{t-k}^{CB}, q_{t-1-k}, y_{t-1-k}, p_{t-1-k}\right\}_{k=0}^{\infty}$$

Firms also each receive their own idiosyncratic signal; firm i receives a signal

$$S_t^i = y_t^N + \varepsilon_t^i$$

where

$$\varepsilon_t^i \sim \mathcal{N}\left(0, \left(\tau^F\right)^{-1}\right)$$

which is assumed to be i.i.d. across firms and across time. Firms directly observe only their own private signals as well as past policy actions and aggregate outcomes, and also observe the current period's policy action. Their information set is thus represented by

$$\mathcal{I}_t^i = \left\{S_{t-k}^i, q_{t-k}, y_{t-1-k}, p_{t-1-k}\right\}_{k=0}^{\infty}$$

The key state variable that drives the dynamics of the model is a variable denoted by $\mathcal{B}_t$, denoting the probability that firms put on the central bank being of type $\theta_t = H$ heading into period t. Each firm i uses its information set $\mathcal{I}_t^i$ to form its own belief $\mathcal{B}_t^i$; I show later that because idiosyncratic signals provide no additional information after aggregate outcomes y_t and p_t are realized, all firms share the same beliefs heading into a given period, and thus $\mathcal{B}_t^i = \mathcal{B}_t$. Furthermore, I show that $\mathcal{B}$ satisfies the Markov property, and thus the beliefs heading into period $t+1$ can be written as

$$\mathcal{B}_{t+1} = \mathcal{G}(p_t, y_t, q_t \mid \mathcal{B}_t)$$

where $\mathcal{G}(\cdot)$ is the law of motion for beliefs.

1.2.4 Timing

The timing within each period t is as follows:

1. Central bank type θ determined (hit with information shock with prob λ), shocks ξ_t drawn, determining y_t^N
2. Central bank receive private signal S_t^θ and firms receive idiosyncratic signals S_t^i
3. Central bank sets nominal demand q_t (on a discretionary basis), observed by firms
4. Firms set prices with information set $\mathcal{I}_t^i$
5. Consumers demand products, production takes place at given prices
6. Firms observe aggregate variables, update beliefs $\mathcal{B}_{t+1}$ about the central bank

1.2.5 Equilibrium Definition

I focus on symmetric equilibria that satisfy optimality of households, firms, and the central bank, as well as market clearing and consistency of beliefs. I restrict myself to Markov Perfect Equilibria (henceforth MPE) where the central bank acts on a discretionary basis, and optimizes over the expected discounted sum of welfare that takes into account both that its type may change as well as how its policy actions affect the backward-looking variable of beliefs. Lastly, both firms and the central bank must best respond to each other, taking each other's equilibrium actions into account. The description of equilibrium is as follows:

Definition 1. *A symmetric rational expectations Markov Perfect Equilibrium is defined as a set of functions* $C(\cdot)$, $q^H(\cdot)$, $q^L(\cdot)$, $\mathcal{G}(\cdot)$, $p(\cdot)$ *such that:*

1. *Households, firms, and each central bank type all maximize, respectively, (1.1), (1.6), and (1.11), with symmetry implying all firms use the same pricing rule $p(\cdot)$. Firms and each central bank type optimizes conditional on their respective information sets.*

2. *Beliefs of firms are consistent: $\mathcal{G}(\cdot)$ is such that $\mathcal{B}(\cdot)$ best incorporates firms' information set $\mathcal{I}(\cdot)$.*

3. *All markets clear.*

1.3 Equilibrium Characterization

I now analyze one period in isolation, looking at equilibrium actions taken each period and the resulting equilibrium outcomes. Within each period itself is embedded a sequential game; the central bank of type θ_t sets policy first, and the firms set prices next, each doing so with incomplete information. Both best respond to each other's actions, with optimization over the continuation payoffs within the dynamic structure of the game. For tractability, I begin by making a few key assumptions:

Assumption 1. *Beliefs $\mathcal{B}_t$ are updated once each period after the realization of aggregate variables in time t.*

Assumption 2. *The central bank and firms enter each period with a diffuse prior about the natural level of output.*

Assumption 1 greatly simplifies the analysis, preserving the Gaussian information structure of the game. If firms instead continuously update beliefs throughout the interim period, idiosyncratic private signals would lead to a distribution of individual beliefs, leading to much more complicated best response functions that would cease to be linear.[14] This assumption allows beliefs to be captured by the single-valued random variable $\mathcal{B}$ capturing aggregate beliefs; as I show later, beliefs are then always identical for each firm, as idiosyncratic signals provide no additional information about θ_t after aggregate outcomes are realized.

Assumption 2 is a common assumption in the literature on public signals,[15] allowing to clearly

[14] In the Appendix, I show that this assumption can additionally be thought of as a rough approximation, as given the parameter restrictions of the model, interim beliefs are unlikely to change much in the interim period.

[15] see Morris and Shin (2002) as well Amato et al. (2002), and James and Lawler (2012; 2016), and Roca (2010) in the context of monetary policy.

isolate the the optimal use of private information relative to public information, a main focus of this paper. Importantly in this model, the diffuse prior assumption also allows for a particularly easy solution to the central bank's optimization problem, shutting off strategic behavior for a central bank wishing to manipulate its reputation for competence. Because the focus of this paper is not on the strategic considerations of a central bank developing its reputation, and how this affects optimal policy, but rather on the effects of the reputation for competence channel on equilibrium outcomes and welfare, this assumption allows me to make direct comparisons with existing models while still ensuring the model is consistent with central bank optimization in equilibrium.[16]

The next two assumptions are not crucial for the analysis, but are mild assumptions used to make sharper predictions of the model:

Assumption 3. *The signal precisions of central bank and the elasticity of substitution across goods satisfy* $\frac{\tau^H}{\tau^L} < \eta$.

Assumption 4. *A fraction* $\alpha \to 0$ *of firms set prices at the steady state price level* $p = 0$.

Assumption 3 is a mild restriction on the parameter space. Typical estimates of η are around 5-10, which generates steady-state markups (without subsidies) of about 10% - 25%. While τ^G and τ^B are not directly estimable, empirical work looking at variation in central bank forecasting errors show substantial, but much less variation in forecasting errors, than that which would generate a violation of this assumption. Assumption 4 is made for the central bank to prefer an equilibrium outcome with less price level variability, which leads to a unique equilibrium; because $\alpha \to 0$, there are no major implications of this assumption. This assumption is dropped when adding Calvo frictions, as then the central bank has an inherent reason to target a particular price level.

Given these assumptions, I now characterize the unique linear Markov Perfect Equilibrium of the game. Before doing so, it is useful to define the following variable $\mathcal{H}_t$, denoting the weighted harmonic mean of $\tau^\theta + \tau_F$, with weights given by firm beliefs:

$$\mathcal{H}_t = \left[\frac{\mathcal{B}_t}{\tau^H + \tau_F} + \frac{1 - \mathcal{B}_t}{\tau^L + \tau_F} \right]^{-1}$$

[16]This assumption can additionally be thought of as an approximation when the variance of underlying shocks are substantially larger than the variance of the signal errors, which seems to be a reasonable assumption.

Heuristically, this variable captures the perceived total amount of combined precision of the central bank and firm's signal, and is higher if the possible central bank precisions τ^H and τ^L are larger, and if the firm's precision τ_F is larger. More importantly, for given precisions, $\mathcal{H}_t$ will be larger if $\mathcal{B}_t$ is larger: if the central bank is perceived as being type H, the amount of total precision is perceived to be higher as well. Given this variable, the equilibrium is characterized by the following proposition:

Proposition 1. *Under Assumptions 1, 2, 3 in equilibrium each central bank type $\theta \in \Theta$ plays $q_t^\theta = S_t^\theta$, thus setting nominal demand equal to its best estimate of natural output. Furthermore, firms set their prices as $p_t^i = \gamma_{1t} q_t + \gamma_{2t} S_t^i$ where the equilibrium pricing coefficients are given by*

$$\gamma_{1t} = 1 - \underbrace{\frac{\mathcal{H}_t - \tau_F}{\mathcal{H}_t - (1-\zeta)\tau_F}}_{\text{signaling channel}}$$

$$\gamma_{2t} = -\frac{\zeta \tau_F}{\mathcal{H}_t - (1-\zeta)\tau_F}$$

which implies $\gamma_{1t} + \gamma_{2t} = 0$, and thus $\gamma_{1t} = -\gamma_{2t} \equiv \gamma_t$. This is the unique linear Markov Perfect Equilibrium.

Proof. See Appendix. □

Given the assumptions of the model, the optimal action for each central bank type θ is to set nominal demand in order to attempt to track the natural level of output; this is the action it would take if its type was perfectly observed.[17] Intuitively, because the complete information equilibrium is efficient,[18] in a static sense the central bank has no incentive to deviate from truth-telling by altering its action in a way that manipulates the information of firms. In addition, if firms believe that $q_t^\theta = S_t^\theta$, it is shown in the proof of the proposition that any alternative action by the central bank leads to a more unfavorable distribution of beliefs in the future, a point made clearer later. Lastly, given the diffuse prior assumption, attempting to 'mimic' the competent type becomes a

[17] More generally, without a diffuse prior, the optimal action when the type is known is for the central bank of type θ to set q_t^θ equal to the posterior mean of the central bank's estimate of y_t^N given it's signal S_t^θ.

[18] When firms have complete information about the underlying shocks, all firms set the same price $p_t^i = p_t$ and it follows that $y_t = y_t^N$. Thus, $p_t = q_t - y_t^N$, and the optimal monetary policy is indeterminate: aggregate demand q_t can be arbitrarily chosen by the central bank, leading the price level to adjust automatically one for one. It follows that $U_t = 0$, and the equilibrium is perfectly efficient.

fruitless endeavor; Assumption 2 thus shuts off strategic considerations of a central bank attempting to build its reputation, which is outside the scope of this paper.

The firm pricing coefficients have a fairly simple interpretation. γ_{1t}, representing the response of individual prices to nominal demand set by the central bank q_t, can be decomposed into two components. The first is the *direct effect*, arising from the optimal response to increase prices one-for-one with any exogenous shock to nominal income. The second effect comes from a *signaling channel*: because the central bank sets nominal demand in an attempt to track the natural output level, its action directly reveals its own signal. Because of the publicly observed nature of the central bank's action, it serves as a public signal about the natural level of output. If nominal demand is set high, the natural level of output is also likely to be high, muting the optimal price increase effect of an increase in nominal demand.

$\gamma_{2t} < 0$ is the counterpart to the public signal effect, capturing the response firms have to their own private signal about natural output. As expected, the weights on the public and private signal, given by the signaling component of γ_{1t} and γ_{2t}, respectively, add up to 1.[19] Also note that in the case that θ is common knowledge to all agents in the model, $H_t = \tau^\theta + \tau_F$, and γ_{1t} and γ_{2t} reduce to familiar formulas as in the existing literature of public and private information with strategic complementarities.

Crucially, a distinguishing feature of the model are the time-varying price coefficients, whose movements depend solely on the movement in the beliefs $\mathcal{B}_t$ about the central bank's competence, which enter through the $\mathcal{H}_t$ term defined earlier. The following property is used to derive many implications of the model:

Lemma 1. *γ_t is decreasing in $\mathcal{B}_t$. Thus, firms put less weight on their own private signal, when they perceive the central bank to be more competent.*

Proof. Because $\tau^H > \tau^L$, it follows that $\mathcal{H}_t$ is increasing $\mathcal{B}_t$. and it immediately follows that $-\gamma_{2t}$, which is decreasing in $\mathcal{H}_t$, is decreasing in $\mathcal{B}_t$. □

[19] The responses to the signals add up to -1; this is negative because ceteris paribus, a higher level of natural output leads to a lower output gap and thus a lower optimal price

1.3.1 Equilibrium Properties

As standard in models with dispersed information, changes in the equilibrium response to public and private signals lead to a tradeoff between aggregate volatility and dispersion in actions (here price dispersion). However, in the model presented here, equilibrium responses are *not* captured solely by the precisions of signals, but rather by *beliefs* about the precisions, which move separately from the true underlying precisions. I now look at the aggregate outcomes relevant for monetary relevant welfare, starting with the output gap:

Proposition 2. *The equilibrium output gap is* $\widetilde{y}_t = (1-\gamma_t)\,\varepsilon_t^{\theta}$ *, and thus* $\mathbb{E}_t\left(\tilde{y}_t^2 \mid \theta_t, \mathcal{B}_t\right)$ *is*

(i) decreasing in τ_t^{θ} *(lower when* $\theta_t = H$*)*

(ii) increasing in $\mathcal{B}_t$*.*

Proof. See Appendix. □

In terms of the output gap, perceptions of high competence are *unfavorable* for social welfare. When firms perceive the central bank as more competent, they increase weight put on the public signal, magnifying any monetary error's impact on the output gap. When the true underlying competence is higher, these errors are likely to be smaller, lowering output gap volatility. This stands in contrast to existing models with public information, where an increase in the precision of a public signal may in fact *increase* aggregate volatility. There, the effect comes from the increase in the equilibrium response to the public signal, which moves according to the true underlying precision; here, the response is dictated solely by the perceptions $\mathcal{B}_t$ which moves imperfectly with θ_t. Thus, in this model an increase in the precision of the public signal unambiguously *decreases* volatility.

The expression derived for the output gap also helps illustrate central bank 'Minsky moments' where periods of stability lead to instability. As later shown, long periods of stable monetary outcomes will lead to a higher perception of the central bank's competence, leading to a decrease in γ_t. If the true underlying type reverts to $\theta = L$, γ_t remains low, due to the imperfect observability of the type θ. In such a scenario, $(1-\gamma_t)$ is high, and the variance of the monetary errors ε_t^{θ} is also high, potentially leading to very large movements in the output gap, and possible states of the world where welfare is very low. Such an effect can only come due to the decoupling of the true

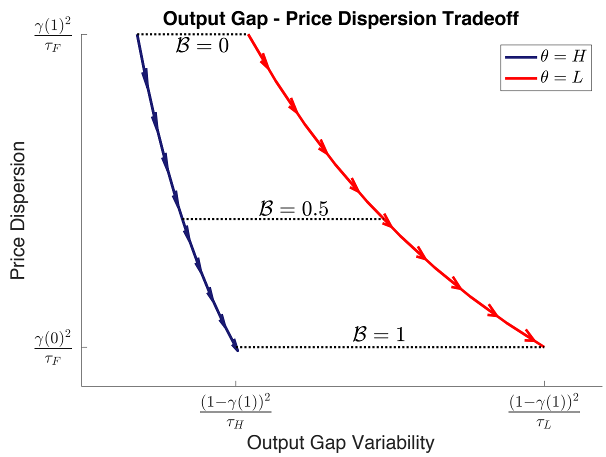


Figure 1.2: Price dispersion and output gap volatility for different perceptions of competence $\mathcal{B}$. As $\mathcal{B}$ increases (direction of arrows), price dispersion decreases as output gap variability increases.

competence with beliefs, which leads to sluggish adjustment of γ. As later shown, expected welfare remains higher with higher $\mathcal{B}_t$, when the underlying type is $\theta = L$; such a scenario, however, leads to large downside tail risk.

I now analyze price dispersion, the second component of monetary welfare:

Proposition 3. *$\mathcal{V}(p_t^i)$ is decreasing in $\mathcal{B}_t$ and is independent of the true type θ.*

Proof. Without Calvo frictions,

$$\mathcal{V}(p_t^i) = \int_0^1 \left(p_t^i - p_t\right)^2 di = \gamma_t^2 \int \left(S_t^i - y_t^N\right)^2 di = \gamma_t^2 \left(\tau_F\right)^{-1}$$

Because γ_t is decreasing in $\mathcal{B}_t$ by Lemma 1, it immediately follows that $\frac{\partial \mathcal{V}(p_t^i)}{\partial \mathcal{B}_t} < 0$. □

In terms of price dispersion, perceptions of high competence are favorable; when firms perceive the central bank as more competent, they decrease their weight on their own idiosyncratic private signals, the source of disagreement that leads to price dispersion in the first place. Unlike the output gap, the degree of price dispersion is independent of the monetary error, as prices are set at the beginning of the period and are thus pre-determined, before the other aggregate outcomes (e.g., output) adjust. What emerges is thus a tradeoff to perceptions of competence: a better reputation for competence leads to lower price dispersion, at the expense of larger output gap volatility. This is illustrated in Figure 1.2, which plots price dispersion and output gap volatility as a function of beliefs $\mathcal{B}$. In terms of welfare, the positive effects of a better reputation for competence always outweigh the costs, *in every state of the world,* as shown in the following proposition:

Proposition 4. *Equilibrium expected welfare is*

$$\mathcal{U}\left(\gamma_t, \theta_t\right) = -\left[\left(1-\gamma_t\right)^2 \frac{1}{\tau^{\theta_t}} + \frac{\eta}{\zeta \tau_F} \gamma_t^2\right]$$

This is increasing in $\mathcal{B}_t$, regardless of the true type θ_t and beliefs $\mathcal{B}_t$, given the parameter restriction in Assumption 3.

Proof. See Appendix. □

One key implication from this is that a central bank's attempts to announce its type is merely cheap talk and not credible, as the central bank would always claim to be the $\theta = H$ type. The reasoning behind this result comes from the inefficient use of information by firms. This inefficiency leads to *insufficient* coordination in equilibrium, and an increase in $\mathcal{B}_t$ induces a smaller response to idiosyncratic private signals, leading to more coordination in equilibrium and an allocation closer to the constrained efficient benchmark (discussed in Section 6). This *coordination channel* is one of two channels in this model that provide a reason for central banks to want to be perceived as competent. Another important observation is that the $\theta = H$ type faces a more favorable trade-off, in that while both types benefit equally from a marginal increase in $\mathcal{B}_t$ (because price dispersion is predetermined ex-ante from beliefs), the $\theta = H$ type faces a smaller cost, as in expectation it makes smaller errors; thus the amplification of errors from a higher $\mathcal{B}_t$ is less costly.

The final property provides a key result for *how* these beliefs are determined:

Proposition 5. *In equilibrium, the central bank's error ε^θ is correctly inferred after the observation of aggregate outcomes (q_t, y_t, p_t). The likelihood functions are*

$$\mathcal{L}(\theta \mid q_t, y_t, p_t, \mathcal{B}_t) = f\left(\frac{q_t - y_t}{\gamma_t(\mathcal{B}_t)} \,\middle|\, \theta\right)$$

where $f(\cdot \mid \theta)$ is the pdf of $\mathcal{N} \sim \left(0, \left(\tau^\theta\right)^{-1}\right)$. In terms of the output gap:

$$\mathcal{L}(\theta \mid q_t, y_t, p_t, \mathcal{B}_t) = f\left(\frac{\widetilde{y}_t}{1-\gamma_t(\mathcal{B}_t)} \,\middle|\, \theta\right)$$

Proof. See Appendix. □

Thus, from the observation of aggregate outcomes, firms can form posteriors about the type of the central bank through standard Bayesian updating. Note that firms do not directly observe the output gap $\widetilde{y}_t$ (or the natural level of output y_t^N); they instead infer these through the observable aggregate outcomes $(q_t, y_t, p_t, \mathcal{B}_t)$. In addition, if the output gap is large in magnitude, the central bank is relatively more likely to be of a poorer type, as intuitively a central bank with worse signals will have a harder time stabilizing output.

1.4 Equilibrium Dynamics

The previous section illustrated how perceptions of competence affect equilibrium outcomes in any given period, by taking as fixed beliefs $\mathcal{B}_t$ and analyzing one period in isolation. I now move onto the dynamics of the model, first showing how these perceptions are formed endogenously from previous monetary outcomes. Posterior beliefs are given by Bayes' rule; the updated beliefs at the end of period t are

$$\widehat{\mathcal{B}}_t = \frac{\mathcal{L}(\theta = H \mid q_t, y_t, p_t, \mathcal{B}_t)}{\mathcal{B}_t \mathcal{L}(\theta = H \mid q_t, y_t, p_t, \mathcal{B}_t) + (1 - \mathcal{B}_t)\mathcal{L}(\theta = L \mid q_t, y_t, p_t, \mathcal{B}_t)} \mathcal{B}_t$$

Furthermore, because the type of the central bank changes with an exogenous probability λ, it follows that heading into period $t + 1$ the updated prior is

$$\mathcal{B}_{t+1} = \lambda + \widehat{\mathcal{B}}_t(1 - 2\lambda)$$

Given this, I show the following key proposition:

Proposition 6. *Given prior beliefs $\mathcal{B}_t$, the beliefs entering the next period $\mathcal{B}_{t+1}$ are increasing in monetary welfare:*

$$\left.\frac{\partial \mathcal{B}_{t+1}}{\partial U_t}\right|_{\mathcal{B}_t} \leq 0$$

Proof. See Appendix. □

Proposition 6 states an intuitive result: when the central bank achieves a good monetary outcome (measured by monetary welfare U_t), the perceived competence of the central bank becomes higher than if the central bank's actions led to poor welfare outcomes. Heuristically, because price dispersion is fixed in terms of beliefs, the unpredictable variation in monetary welfare comes

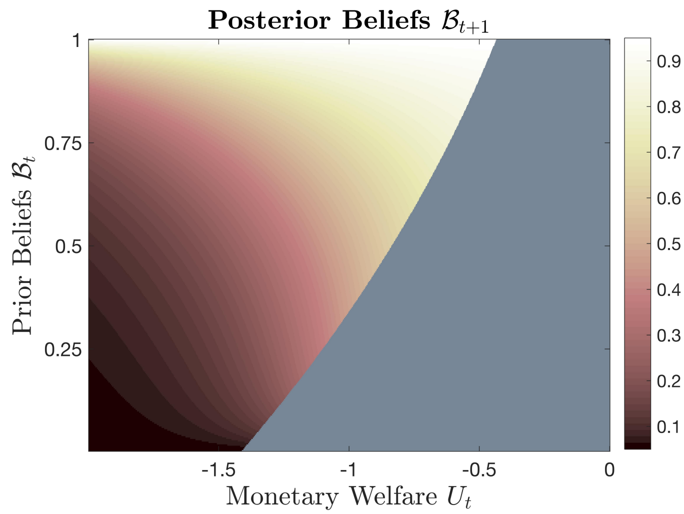


Figure 1.3: Updated beliefs as a function of monetary welfare outcomes. Lighter shades represent higher posteriors, with gray denoting unachievable outcomes.

from the output gap; large magnitudes of the output gap stem from large monetary errors, and because the monetary error is revealed through the observation of aggregate outcomes (q_t, y_t, p_t) (Proposition 5), Proposition 6 holds. Because a higher perception of competence leads to better expected welfare outcomes in all possible states, this beliefs channel leads to an *accelerator* effect, whereby good monetary outcomes (stable output) lead to further good monetary outcomes (reduced price dispersion) in future periods. This *reputational accelerator* serves to amplify the total effects of any given monetary error.

Figure 1.3 provides an illustration for how posterior beliefs about the central bank's type change as a result of monetary outcomes. As Proposition 6 shows, for any given prior $\mathcal{B}_t$, better monetary outcomes lead to higher beliefs $\mathcal{B}_{t+1}$ in the next period. As expected, updated beliefs are less sensitive to monetary outcomes when priors are near the extremes. In addition, during the updating of beliefs, the central bank's performance is evaluated *relative* to the expected outcome given beliefs; when $\mathcal{B}_t$ is high, expected monetary welfare is also higher, and the central bank is thus judged accordingly.

1.4.1 Illustration: the Reputational Accelerator

In order to analyze the dynamics of the model, I first begin with a few illustrative examples that provide intuition, before providing a full mathematical analysis. The following example illustrates

	Case 1 (accurate)	Case 2 (inaccurate)
Signal S_t^θ	0.5	0.5
Policy q_t	0.5	0.5
Natural Output r_t^N	0.7	1.5
Output gap $\widetilde{y}_t$	-0.118	-0.591
Period Welfare U_t	-0.572	-0.907
New belief $\mathcal{B}_{t+1}$	0.766	0.588

Table 1.1: Numerical example. In both cases, the true underlying type, the reputation for competence, and the signal are the exact same; in Case 1, the signal is accurate (leading to near optimal policy), and in Case 2, the signal is inaccurate (leading to suboptimal policy)

the mechanics of the reputational accelerator, by considering two situations where the true central bank type θ_t and beliefs $\mathcal{B}_t$ are the same: for concreteness, suppose $\theta_t = H$, $\mathcal{B}_t = 0.7$, and the central bank receives a signal of natural output $S_t^H = 0.5$, leading to the same equilibrium action $q_t = 0.5$. In one case, this signal is relatively accurate, and the central bank's level of nominal demand is thus appropriate given underlying fundamentals. In the other case, however, the signal is relatively inaccurate, leading to a suboptimal action. The latter case will exhibit a large welfare loss at time t compared to the former case, which also leads to damage to its reputation for competence. Table 1.1 provides numerical details given the baseline parameterization of the model.

Figure 1.4 shows the path of expected welfare after the shock at $t = 0$, also illustrating the case where the beliefs channel is not present (i.e., the type is common knowledge to the firms). The dotted lines represent an alternative model in which the type is perfectly observed by firms, with the blue line representing the accurate signal case and the red line representing the inaccurate case. When the type is fully known, the monetary error at $t = 0$ leads to a large effect on welfare in period 0, but the path of expected welfare remains unaltered: expected welfare continues on its steadily declining original path, with the downward slope coming from the central bank potentially changing to the unfavorable case $\theta = L$ with an exogenous probability λ each period.[20] Thus, when the type is fully observed, the effect of a shock in period t is contained within that period.

When the type is unknown, however, the effects of a shock at $t = 0$ on expected welfare are very different: now, the shock's effects persist into future periods, due to the accelerator effect. As shown, expected welfare in all future periods at time $t + 1, t + 2, ...$ is higher under the case

[20]as $t \to \infty$, expected welfare is such that the probability of each type approaches 0.5 (stationary distribution of Λ), that is it converges to the arithmetic mean of $\mathbb{E}_t(\mathcal{U}_t \mid \theta_t = H)$ and $\mathbb{E}_t(\mathcal{U}_t \mid \theta_t = L)$.

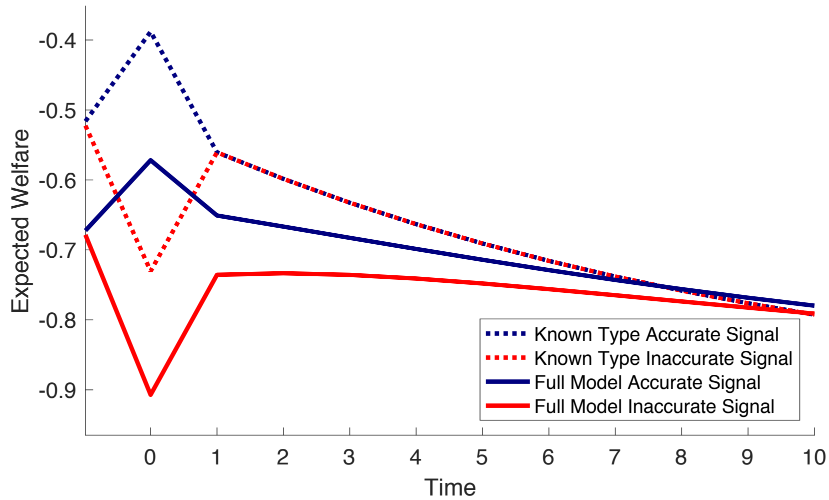


Figure 1.4: Expected welfare dynamics after a shock. Dotted lines indicate the model in which the type is fully observed, and solid lines denote the model in which the type is imperfectly observed. Blue denotes the case where the central bank's signal was accurate, and red denotes the case where the signal was inaccurate.

when the signal at $t = 0$ was accurate than under the case when it was inaccurate, with the path of expected welfare before the shock hits in between. That is,

$$\mathbb{E}_t^{inacc}[\mathcal{U}_{t+j}] < \mathbb{E}_t^{original}[\mathcal{U}_{t+j}] < \mathbb{E}_t^{acc}[\mathcal{U}_{t+j}]$$

for all j. Thus, monetary errors have a *permanent* effect on expected welfare, with this effect declining over longer horizons. Note that again, the paths of expected welfare are declining in the long run, because of the potential for the true underlying θ to change to L. However, in the inaccurate signal case, the path is increasing for a few periods: because the true underlying type is indeed H, in the short to medium run the central bank is expected to somewhat undo the large damage done to its reputation. Despite this, expected welfare always remains lower than if the central bank had not created as subpar a monetary outcome at the onset; thus, monetary errors have spillover effects that persistently last into the future under the presence of type uncertainty.

Figure 1.4 also illustrates other key implications of the model. At time $t = -1$, before the shock hits, expected welfare is higher when the type is known; this is because the *true* type is $\theta = H$, and thus the corresponding 'true belief' is $\mathcal{B}_{t-1} = 1$; when the true type is high competence, the central bank would prefer its type to be perfectly observed. In addition, the decline in the path

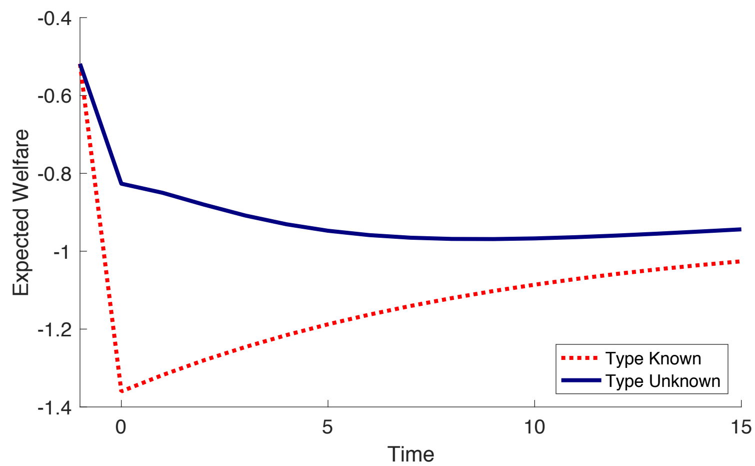


Figure 1.5: Expected welfare dynamics under a type change from $\theta = H$ to $\theta = L$, for both the model in which the type is fully observed (dotted red line) and the model in which the type is imperfectly observed (solid blue line).

of expected welfare is more gradual when the type is imperfect information, due to two separate effects. First, because the true type is $\theta = H$ at $t = -1$, the expected path of *beliefs* trends upwards, as the passage of time slowly reveals the central bank's true type. Because a higher $\mathcal{B}_t$ is beneficial for expected welfare at time t, this serves to shift the path of expected welfare upwards. Second, when the type is unknown, an actual change to type $\theta = L$ is less detrimental to welfare, due to the slow learning of firms about the central bank's type; this is shown in more detail in the example below. Both these effects serve to mute the declining path of expected welfare that stems from a potential change to type $\theta = L$. Interestingly, note that while expected welfare is initially lower when the type is unknown, it eventually becomes *higher* than the case when the type is known. The long run properties of the model, and the convergence to the stochastic steady state are discussed in detail in section 5.

1.4.2 Illustration: Type Shocks

I now consider a case where the true type θ changes from type $\theta = H$ to $\theta = L$, a shock that occurs with probability λ. In the model where the type fully observed, this has a large effect on expected welfare at the time of the shock: the public realizes the central bank's inability to stabilize the economy, leading to a large loss in price coordination. However, in the case where the public must infer the central bank's type, the effect on welfare is muted: expected welfare immediately drops as the central bank has a harder time choosing the right level of nominal demand to stabilize output, but the coordination channel is not affected immediately. Thus, imperfect information about the central bank's type acts as a temporary buffer to stem the loss in expected welfare, which is illustrated in Figure 1.5. Note that expected welfare from $t+1$ continues to decrease, as in expectation, the type will slowly be revealed to the public, with the central bank struggling to minimize output gap volatility due to its relatively imprecise signal.

To summarize, each shock in the model — both monetary errors (ε^{θ}) and shocks which change the type θ — have direct effects that take place in the current period, as well as indirect effects on the path of expected welfare going forward. Large monetary shocks (or the lack thereof) cause large welfare loss (higher than expected welfare) in the current period, with reputation effects leading to welfare loss (welfare gain) in future periods, through a reputation for competence channel. A type shock also causes movements in expected welfare, coming from both a direct effect from the change in precision of signals affecting the central bank's ability to stabilize the economy, as well as indirect effects through the expected path of beliefs.

1.4.3 Value Function

I now define the value function $V(\theta, \mathcal{B})$, corresponding to (1.11), the discounted sum of monetary welfare when the type is θ and $\mathcal{B}$ in period t. I show two main results: for any given belief, total future welfare is higher when the central bank's current true type is the competent type, and for any true type, total future welfare is increasing in current beliefs about the central bank's type. This is captured in the following two propositions:

Proposition 7. *(Welfare Increasing in True Precision) For any belief $\mathcal{B}$, expected discounted welfare is higher when the central bank is the high-competence type:*

$$V(\theta = H, \mathcal{B}) > V(\theta = L, \mathcal{B})$$

Proof. See Appendix. □

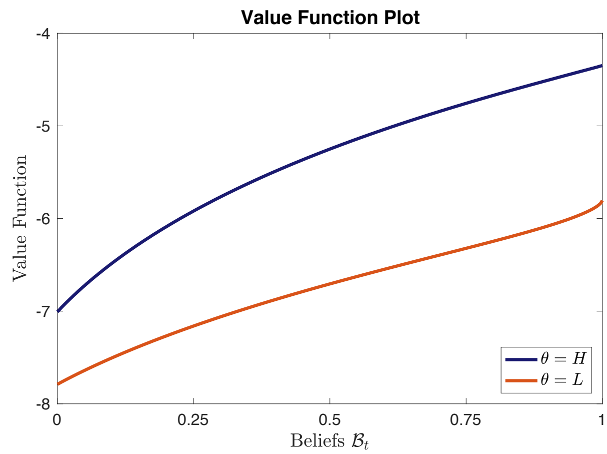


Figure 1.6: Value function illustration

Proposition 8. *(Monotonicity in Beliefs): For any true type θ, expected discounted welfare is increasing in beliefs $\mathcal{B}$:*

$$\frac{\partial V(\theta, \mathcal{B})}{\partial \mathcal{B}} \geq 0$$

Proof. See Appendix. □

These propositions are illustrated in Figure 1.6. Proposition 7 and the monotonicity result from Proposition 8 are apparent: the value function curve representing the high-competence type is always above the curve representing the low-competence type, and each curve is strictly increasing. While the formal proofs of these propositions require an analysis of the joint distribution of the evolution of θ and $\mathcal{B}$, the results are intuitive based on the preceding discussion focusing on the static case, combined with the persistence of types. While these two properties also hold for the period welfare function $\mathcal{U}(\gamma(\theta), \mathcal{B})$ as well, the *shape* of the value function looks markedly different from that of the period welfare function. For instance, in the baseline parameterization used throughout the paper, the curve corresponding to $\theta = L$ goes from a concave function at low values of $\mathcal{B}$ to a convex function at high values of $\mathcal{B}$, and the difference between the $\theta = H$ and $\theta = L$ curves initially increases before decreasing again. Explaining these peculiarities, and formally analyzing the mechanics of the reputational accelerator, are the focus of the remainder of this section.

1.4.4 Inspecting the Mechanism: Decomposing Dynamic Effects

In order to fully illustrate the key dynamics of the model, I now decompose the main mechanisms of the model, quantifying them in terms of model parameters. I begin by analyzing the effects of a change in prior beliefs on expected welfare in the immediate following period, before looking at the entire path of dynamics of expected welfare. First begin by noting that expected welfare in period $t+1$ at time t is

$$\mathbb{E}_t\left[\mathcal{U}_{t+1} \mid \theta_t, \mathcal{B}_t\right] = (1-\lambda)\int \mathcal{U}\Big(\gamma\left(\mathcal{B}'\left(\mathcal{B}_t, \varepsilon_t\right)\right), \theta_t\Big) dF^{\theta_t}(\varepsilon_t) + \lambda \int \mathcal{U}\Big(\gamma\left(\mathcal{B}'\left(\mathcal{B}_t, \varepsilon_t\right)\right), \widetilde{\theta_t}\Big) dF^{\theta_t}(\varepsilon_t) \tag{1.12}$$

where

$$\mathcal{B}'(\mathcal{B}, \varepsilon) = \lambda + \frac{f(\varepsilon \mid \theta = H)}{\mathcal{B} f(\varepsilon \mid \theta = H) + (1-\mathcal{B}) f(\varepsilon \mid \theta = L)} \mathcal{B}(1-2\lambda)$$

where F^{θ} denotes the normal cumulative density function with $(\mu, \sigma^2) = \left(0, \left(\tau^{\theta}\right)^{-1}\right)$, $f(\cdot \mid \theta)$ the corresponding likelihood function, and

$$\gamma(\mathcal{B}) = \frac{\tau_F \zeta}{\left[\frac{\mathcal{B}}{\tau^H + \tau_F} + \frac{1-\mathcal{B}}{\tau^L + \tau_F}\right]^{-1} - (1-\zeta)\tau_F}$$

$$\mathcal{U}(\gamma, \theta) = -\left[(1-\gamma)^2 \frac{1}{\tau^{\theta}} + \frac{\eta}{\zeta \tau_F}\gamma^2\right]$$

Here $\mathcal{B}'(\mathcal{B}, \varepsilon)$ shows how beliefs change according to the prior belief $\mathcal{B}$ and the realized shock ε. $\gamma(\mathcal{B})$ is the expression found earlier determining the weight agents place on their private signal as a function of beliefs about the central bank, and $\mathcal{U}(\gamma, \theta)$ is expected welfare in the period as a function of the true type θ and the private response γ. To see how a marginal change in $\mathcal{B}_t$ affects expected welfare at period $t+1$, I first look at how beliefs in the next period are affected by a marginal change in the prior in the previous period.

It is easily shown that

$$\frac{\partial \mathcal{B}'(\mathcal{B}, \varepsilon)}{\partial \mathcal{B}} > 0, \; \frac{\partial \mathcal{B}'(\mathcal{B}, \varepsilon)}{\partial \mid \varepsilon \mid} < 0 \tag{1.13}$$

and so unsurprisingly, for any realization of the shock ε, the posterior belief is higher when the prior is higher; conversely, for any given prior belief, a higher magnitude of the realized shock induces a reduction in the posterior beliefs. The following key properties also hold of the period expected welfare function:[21]

Expected welfare is increasing, and concave in beliefs:

$$\frac{\partial \mathcal{U}(\mathcal{B},\theta)}{\partial \mathcal{B}} > 0,\ \frac{\partial^2 \mathcal{U}(\mathcal{B},\theta)}{\partial \mathcal{B}^2} < 0$$

For any given belief, a marginal increase in beliefs increases expected welfare more for the $\theta = H$ type:

$$\frac{\partial \mathcal{U}(\mathcal{B},\theta = H)}{\partial \mathcal{B}} > \frac{\partial \mathcal{U}(\mathcal{B},\theta = L)}{\partial \mathcal{B}} \tag{1.14}$$

These properties are unsurprising, given the discussion in previous sections: a higher perception of competence increases period welfare, and this has a larger effect when the true type is $\theta = H$. The key new property here is the diminishing marginal benefit of beliefs that comes from the concavity of the expected welfare function.

While an increase in beliefs in the current period benefits the $\theta = H$ type more in the *current* period, one interesting question is which type benefits more from this (current) marginal increase in beliefs in *future* periods, which helps understand the shape of the value function. Using (1.12), and taking partial derivatives with respect to the prior beliefs, the marginal effect on expected welfare in the next period is given by

$$\int \left[(1-\lambda) \frac{\partial \mathcal{U}(\mathcal{B}',\theta = H)}{\partial \mathcal{B}} + \lambda \frac{\partial \mathcal{U}(\mathcal{B}',\theta = L)}{\partial \mathcal{B}} \right] dF^{\theta=H}(\varepsilon) \tag{1.15}$$

if the type is $\theta = H$ in the current period and

$$\int \left[\lambda \frac{\partial \mathcal{U}(\mathcal{B}',\theta = H)}{\partial \mathcal{B}} + (1-\lambda) \frac{\partial \mathcal{U}(\mathcal{B}',\theta = L)}{\partial \mathcal{B}} \right] dF^{\theta=L}(\varepsilon) \tag{1.16}$$

if the type is $\theta = L$; note that $\mathcal{B}'$ depends on the realization of ε, where I suppress notation for simplicity. Decomposing the terms in the integrand further, note that

$$\frac{\partial \mathcal{U}(\mathcal{B}',\theta')}{\partial \mathcal{B}} = \underbrace{\frac{\partial \mathcal{U}(\mathcal{B}',\theta')}{\partial \mathcal{B}'}}_{\text{WB channel}} \underbrace{\frac{\partial \mathcal{B}'(\mathcal{B},\varepsilon)}{\partial \mathcal{B}}}_{\text{PP channel}}$$

where the first term denotes how welfare is affected by beliefs — which I call the welfare-belief (WB) channel — and the second term denotes how posterior beliefs are affected by prior beliefs,

[21] see Appendix for proofs

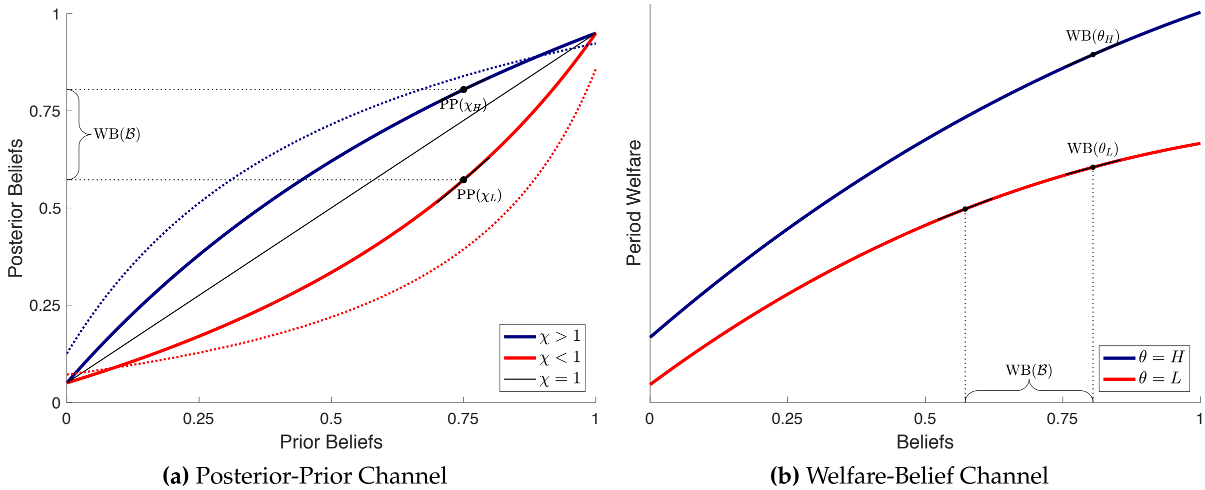


Figure 1.7: Posterior-prior channel and welfare-beliefs channel. The figure on the left illustrates posterior beliefs as a function of priors, plotting $\mathcal{B}'(\cdot, \varepsilon)$ for different values of ε. The case where identical shocks of ε are repeated is also shown (dotted lines), showing the increased curvature at the tail end of beliefs. The slope of this function determines the effect of the PP channel. The figure on the right plots the period welfare function $\mathcal{U}(\cdot, \theta)$ for both possible θ's. The slope of this function determines the effect of the WB channel; at a given $\mathcal{B}'$, the slope is always higher for the $\theta = H$ type (first effect), but ε draws lead $\mathcal{B}$ to generally be lower for the $\theta = L$ type (second effect), as shown in figure (a).

which I call the posterior-prior (PP) channel.

The WB channel itself has two competing effects: first note that conditional on a given $\mathcal{B}'$, the first effect is higher when $\theta = H$, as given by (1.14). Because $\lambda < 1/2$, this effect thus serves to increase the integral (1.15) associated with the high type relatively more than the integral associated with the low type (1.16). Note that regardless of the period's draw of ε (which fixes $\mathcal{B}'$), the integrand will be higher with more weight on the $\theta = H$ term. The second competing effect comes from concavity: when $\mid \varepsilon \mid$ is high, $\mathcal{B}'$ is low (see (1.13)), increasing both terms in the integrand. Because the distribution of $\mid \varepsilon \mid$ when ε is drawn from $\mathcal{N}\left(0, \left(\tau^L\right)^{-1}\right)$ first-order stochastically dominates the distribution when ε is drawn from $\mathcal{N}\left(0, \left(\tau^H\right)^{-1}\right)$, this effect increases (1.16) relative to (1.15). These effects are shown in Figure 1.7b.

To analyze the PP channel, first note that the magnitude of the effect depends on both the realization of the shock ε and the prior belief $\mathcal{B}$, as shown by Figure 1.7a. For low values of $\mid \varepsilon \mid$, the derivative is large when $\mathcal{B}$ is low, and small when $\mathcal{B}$ is high. The opposite is true when $\mid \varepsilon \mid$ is high; in this case, the derivative is large when $\mathcal{B}$ is high, and small when $\mathcal{B}$ is low. The intuition is that priors have particularly strong effects on posteriors when realized outcomes are at odds with prior beliefs: the priors serve as anchors preventing movements in posterior beliefs.[22]Because $\mid \varepsilon \mid$ is more likely to be high when the type is $\theta = L$ (in a FOSD sense), at low values of $\mathcal{B}$ the PP channel is relatively larger when $\theta = H$, and at high values of $\mathcal{B}$ the PP channel is relatively larger when $\theta = L$. Figure 1.7a shows how posterior beliefs depend on prior beliefs (PP channel), for different realizations of ε, with Figure 1.7b showing how these effects translate into effects on the period welfare function (through the WB channel). Note that the convexity in the posteriors for high realizations of $\mid \varepsilon \mid$ explains why the value function turns convex at high values of $\mathcal{B}$ for type $\theta = L$, the type more likely to receive such shocks.

What about the effects of an increase in $\mathcal{B}_t$ on expected welfare starting from period $t + 2$? The calculations become substantially more complicated, arising from the fact that the true type may change in the interim period $t + 1$, changing the distribution that the shocks are drawn from. In addition, the effect of a shock at period t (ε_t) affects how the shock in the next period (ε_{t+1}) affects the welfare-relevant variables. It is important to note that due to the fact that the true type might change ($\lambda > 0$), and the fact that firms rationally know this and update beliefs accordingly, that the order of shocks is *not* commutative, that is,

$$\mathcal{B}'(\mathcal{B}'(\mathcal{B}, \varepsilon_1), \varepsilon_2) \neq \mathcal{B}'(\mathcal{B}'(\mathcal{B}, \varepsilon_2), \varepsilon_1)$$

if $\varepsilon_1 \neq \varepsilon_2$, and so the order of shocks matters in determining beliefs. Intuitively, when the distribution the shocks are drawn from may change over time, more recent shocks are given more weight in the determination of beliefs.

To analyze the effects of an increase in $\mathcal{B}_t$ on expected welfare for any $t + j$, consider any arbitrary path of true underlying types from periods t to $t + j$ with the restriction that θ_t is given. I define the j-tuple $\mathbf{\Theta}^j = (\theta_{t+1}, ..., \theta_{t+j})$ as a sequence of underlying true types and let $\mathbf{\Theta}^j_k$ denote the k-th element of $\mathbf{\Theta}^j$, and further define $\Delta^j(\mathcal{B}, \theta; \mathbf{\Theta}^j)$ as the change in ex-ante expected welfare in period $t + j$ given a marginal change in beliefs $\mathcal{B}$ in period t, *if the sequence of true underlying types follows* $\mathbf{\Theta}^j$.

[22]One can show formally that when ε is such that the likelihood ratio $\chi(\varepsilon) = f(\varepsilon \mid \theta^H)/f(\varepsilon \mid \theta^L)$ is greater than one, $\mathcal{B}'$ is concave in $\mathcal{B}$, and convex when $\chi(\varepsilon)$ is less than one. When $\chi(\varepsilon) = 1$, the relationship is linear: the posterior moves one for one with the prior, with an adjustment factor depending on λ.

Thus for any arbitrary $j = 1, 2, ...$ and sequence $\mathbf{\Theta}^j$,

$$\Delta^j(\mathcal{B}, \theta; \mathbf{\Theta}^j) \equiv \underbrace{\int \cdots \iint}_{j \text{ times}} \frac{\partial \mathcal{U}\Big(\gamma\left(\mathcal{B}'\left(\cdots \mathcal{B}'\left(\mathcal{B}'(\mathcal{B}, \varepsilon^0), \varepsilon^1\right) \cdots \varepsilon^{j-1}\right)\right), \mathbf{\Theta}^j_j\Big)}{\partial \mathcal{B}} dF^{\theta_t} \varepsilon^0 dF^{\mathbf{\Theta}^j_1} \varepsilon^1 \cdots dF^{\mathbf{\Theta}^j_{j-1}} \varepsilon^{j-1}$$

and the *total* marginal effect on expected welfare at $t + j$ is

$$\sum_{\mathbf{\Theta}^j} P(\mathbf{\Theta}^j) \Delta^j(\mathcal{B}, \theta; \mathbf{\Theta}^j)$$

where $P(\mathbf{\Theta}^j)$ denotes the probability of the sequence $\mathbf{\Theta}^j$, given θ_t.[23]

The WB and PP channels discussed in the case for $j = 1$ have direct analogs for general j.[24] Again the WB channel has two competing effects: conditional on any given realization of $\mathcal{B}'\left(\cdots \mathcal{B}'\left(\mathcal{B}'(\mathcal{B}, \varepsilon^0), \varepsilon^1\right) \cdots \varepsilon^{j-1}\right)$, the marginal benefit of increased beliefs is larger when $\mathbf{\Theta}^j_j = H$. When the type in period t is θ, the type in any period $t + j$ for j is more likely to be θ, with this probability converging to $1/2$ as $j \to \infty$.[25] Looking at the equations above, this increased effect on $\Delta(\mathcal{B}, \theta; \mathbf{\Theta}^j)$ works through the $P(\mathbf{\Theta}^j)$ term: sequences of $\mathbf{\Theta}^j$ where $\mathbf{\Theta}^j_j = \mathbf{\Theta}^j_0$ are more likely. Note that this increased effect via the WB channel for the $\theta_t = H$ type decreases as j increases.

The other two effects, which depend on the sequence of draws $(\varepsilon^0, \varepsilon^1, ..., \varepsilon^{j-1})$, tend to initially *increase* as j increases before eventually decreasing. Consider first the countervailing effect on the WB channel. Note that not only is $\mid \varepsilon^0 \mid$ likely to be higher when $\theta_t = L$, but the same is true for $\mid \varepsilon^1 \mid, ..., \mid \varepsilon^{j-1} \mid$, due to the persistence of the underlying type. However, unlike the first effect on the WB channel, the repeated draws of ε compound, combining to affect the ultimate prevailing

[23]Defining $\mathbf{\Theta}^j_0 = \theta_t$, and denoting by ι_k an indicator variable that equals 1 if $\mathbf{\Theta}^j_k = \mathbf{\Theta}^j_{k-1}$, the probability of observing any sequence $\mathbf{\Theta}^j$ is given by

$$P(\mathbf{\Theta}^j) = (1 - \lambda)^{\sum \iota_k} \lambda^{j - \sum \iota_k}$$

[24]for the case $j = 1$ analyzed above,

$$\Delta^1(\mathcal{B}, \theta; \mathbf{\Theta}^1) = \int \frac{\partial \mathcal{U}\Big(\mathcal{B}'(\mathcal{B}, \varepsilon), \mathbf{\Theta}^1_1\Big)}{\partial \mathcal{B}} dF^\theta(\varepsilon)$$

and the total marginal effect on expected welfare $(1 - \lambda)\Delta^1(\mathcal{B}, \theta; \mathbf{\Theta}^1 = \theta_t) + \lambda \Delta^1(\mathcal{B}, \theta; \mathbf{\Theta}^1 = \widetilde{\theta}_t)$, which corresponds exactly to (1.15) and (1.16).

[25]It can be shown that

$$P(\theta_{t+j} = \theta \mid \theta_t = \theta) = \frac{1}{2}\left[1 + (1 - 2\lambda)^j\right] > \frac{1}{2}$$

belief $\mathcal{B}'\left(\cdots\mathcal{B}'\left(\mathcal{B}'(\mathcal{B},\varepsilon^0),\varepsilon^1\right)\cdots\varepsilon^{j-1}\right)$. Thus, for low values of j, the gap between the distributions of beliefs for the two different types for a given starting prior belief $\mathcal{B}$ increasingly widens,[26] and therefore this effect on the WB channel — which tends to be stronger for the $\theta = L$ type — may in fact increase for increasingly future periods.

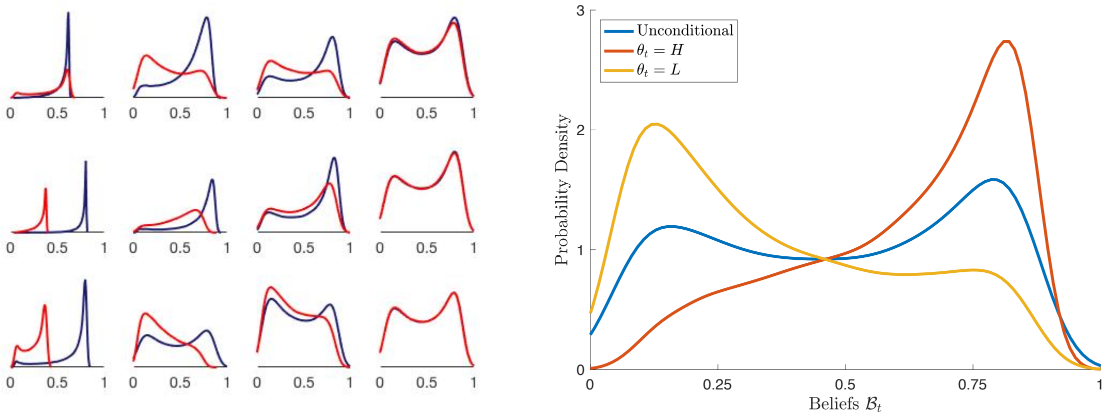
Likewise, the repeated draws of ε compound to affect the PP channel. For instance, if draws continue where $\chi(\varepsilon) < 1$, the beliefs in terms of the prior further "convexifies" as shown in Figure 1.7b. Note that the slope of the PP curve in this case increases for large $\mathcal{B}$ and lowers for small $\mathcal{B}$; informally, the steep part of the curve shifts to the right for repeated draws where $\chi(\varepsilon) < 1$, and vice versa for repeated shocks where $\chi(\varepsilon) > 1$. Thus, as j increases, the effect of the PP channel becomes increasingly large for increasingly low (for $\theta = H$) and high (for $\theta = L$) priors, and decreases for intermediate values of the prior, explaining why in Figure 1.6 the value function turns convex only at the right-most tail for the $\theta = L$ type.

To analyze the overall effect of beliefs, note that the effect of beliefs on the value function is the sum of these effects for each period $t + j$, discounted by a factor β^j. The dynamics in the short-to-medium run are what drives the shape of the value function. This is due not only to the discount factor β, but also due to the eventual convergence to the stochastic steady state. This discussion also provides further intuition for the dynamics of the model after a shock in period t, and visually represented earlier in Figure 1.4. Again, recall that the monetary error shock in period t affects expected welfare in periods $t + 1, t + 2, ...$ through its effect on $\mathcal{B}_{t+1}, \mathcal{B}_{t+2}, ...$ only, and so the framework laid out can be applied directly: a monetary error shock affects the beliefs, which in turn affect the expected welfare dynamics of the model through the WB and PP channels explained above.

1.4.5 Long Run Properties and Fat Tails

It is also of interest to understand the long-run properties of the model, in order to make welfare comparisons between the various benchmark models and the full model when the underlying type is imperfectly observed. The Appendix shows that given any starting state vector $\mathbf{S} = (\theta, \mathcal{B})$, the

[26] One can show that for $j < K$ for some $K > 1$, the distribution of $\mathcal{B}'\left(\cdots\mathcal{B}'\left(\mathcal{B}'(\mathcal{B},\varepsilon^0),\varepsilon^1\right)\cdots\varepsilon^{j-1}\right)$ FOSD $\mathcal{B}'\left(\cdots\mathcal{B}'\left(\mathcal{B}'(\mathcal{B},\varepsilon^0),\varepsilon^1\right)\cdots\varepsilon^{j-2}\right)$ for $\theta = L$ and vice-versa for $\theta = H$. See Appendix for details.



(a) Belief Dynamics

(b) Stationary Distribution of Beliefs

Figure 1.8: Belief dynamics and long-run stationary distribution of beliefs. Figure (a) illustrates the evolution of the distribution of beliefs over time, with each column representing 1, 5, 10, and 25 periods ahead. The first row illustrates the case where the initial true type varies, starting from the same prior belief. The second row illustrates two different priors, for a starting type $\theta = H$, and the last row the same conditions but with a starting type $\theta = L$. Figure (b) illustrates the marginal stationary distribution of beliefs, as well as the stationary conditional distributions for each type θ.

distribution of the state eventually converges to a unique stationary distribution over the state space $\mathbb{S} = \Theta \times [0, 1]$.[27] This result can be seen in Figures 1.8a and 1.8b, where Figure 1.8a shows how the distribution of beliefs evolves given different initial conditions. Note that the distribution of beliefs converges to the distribution shown in the right-most column of the figure, regardless of the initial state $\mathbf{S}_t = (\theta_t, \mathcal{B}_t)$. Also note that this distribution is double-peaked: intuitively, due to the persistence of the true type θ, and eventual learning by firms, beliefs tend to get pulled towards the tails of the distribution, depending on the true underlying type. While the peaks of these conditional distributions are near 0 and 1 — the beliefs that would prevail in the absence of imperfect information about the central bank's type — there is substantial mass throughout the belief space, showing that this informational friction can have a potentially large bite. Figure 1.8b also shows the conditional distribution of beliefs given a true type θ; note that the conditional distributions are themselves double-peaked, and so large misalignment from beliefs and the true type are not only possible, but occur with substantial frequency. Such cases are likely to arise after a type change unobserved by firms.

[27]Note that due to the continuous beliefs space $[0, 1]$, the usual ergodic theorems that apply to discrete state Markov chains or density Markov chains will not apply.

The conditional distributions of beliefs provide the key in understanding the excess volatility generated in this model, which in turn generate fat downside tails of welfare outcomes. Because price dispersion is fixed ex-ante in terms of beliefs, the distribution of price dispersion in any given period is bounded from both below and above. The worst possible welfare outcomes, therefore, are generated by large magnitudes of output gaps, which in turn are generated by monetary errors; because monetary errors are unbounded, these errors are the driving force behind downside tail risk. This risk is greatest when the central bank has done a relatively good job in stabilizing monetary outcomes for long periods of time: this leads to a high $\mathcal{B}$, which leads to large amplification of any monetary error on the output gap. If the central bank type is $\theta = L$, this leads to potentially very large errors that in turn get magnified by the high perception of competence (most likely after a recent type change from $\theta = H$). This is the essence of central bank ''Minsky moments', where periods of long stability can endogenously lead to instability. If the type is known, however, once the type changes to $\theta = L$, there is an immediate concomitant adjustment of beliefs that in turn minimize the effects of monetary errors on output gap volatility. Note, however, that the central bank still prefers to be known as the competent type when its true type is $\theta = L$ as expected welfare is then higher; however, downside tail risks become larger, leading to potential states of the world where welfare is extremely low.

Figure 1.9 illustrates this phenomenon by plotting the cumulative distribution function of monetary welfare outcomes under the model where the type (competence) is fully observed versus the case where the type is imperfectly observed, focusing on the case when the true type is $\theta = L$. As the preceding discussion showed, the model with imperfect observability of type generates a fatter downside tail. However, under the baseline parameterization of this model, expected welfare is in fact *higher* when the type is imperfectly observed. Denoting by $\mathbf{U}_A$ and $\mathbf{U}_B$ the long run average welfare when the type is imperfectly known and perfectly known, respectively, we have

$$\mathbf{U}_A = \int_S \psi^*(dx)\mathcal{U}\Big(\gamma(\mathcal{B}), \theta\Big), \; \mathbf{U}_B = \frac{1}{2}\Big[\mathcal{U}\Big(\gamma(1), \theta = H\Big) + \Big(\gamma(0), \theta = L\Big)\Big]$$

where ψ^* denotes the stationary distribution of $\mathcal{B}$ and θ, and noting that in the stochastic steady state, the probability of $\theta = H$ and $\theta = L$ is $1/2$ each. The precise conditions under which $\mathbf{U}_A$ is

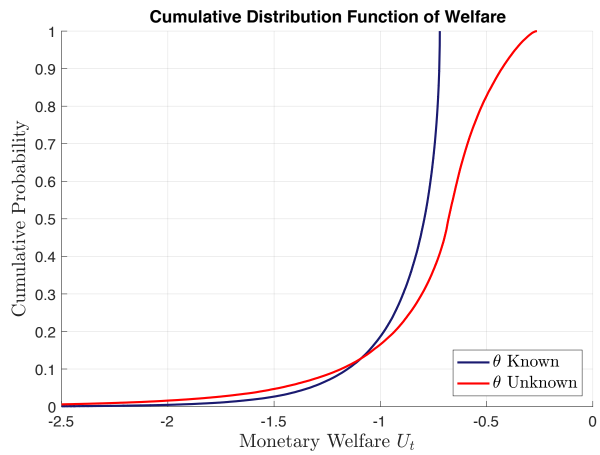


Figure 1.9: Cumulative distribution function of monetary welfare outcomes when $\theta = L$, comparing the models in which the type is fully observed (blue) and imperfectly unobserved (red). Note that here long run expected welfare is higher when the type is unknown, but there is substantially more downside tail risk.

bigger or smaller than $\mathbf{U}_B$ are beyond the scope of this paper, but both $\mathbf{U}_A > \mathbf{U}_B$ and $\mathbf{U}_B > \mathbf{U}_A$ are possible under different parameterizations. This brings up a key point: while the central bank has no ability to credibly reveal its type, this shows that the central bank may in fact not want to *commit* to truthfully revealing its type, *even if it had access to such a technology*. In essence, the gains from type unobservability when the type is incompetent may outweigh the benefits of its type being known when its type is competent.

To summarize:

1. For a given $\mathcal{B}$, the $\theta = H$ type benefits more from an additional unit of $\mathcal{B}$ in a static setting (the WB(θ) channel). However, because the $\theta = L$ type is more likely to face low values of $\mathcal{B}$ in future periods, it potentially benefits more from an additional unit of $\mathcal{B}$ due to the concavity of $\mathcal{U}(\cdot, \theta)$ (the WB($\mathcal{B}$) channel). At high values of $\mathcal{B}$, due to the concavity and convexity of $\mathcal{B}'(\cdot, \varepsilon)$ when $\mid \chi \mid > 1$ and $\mid \chi \mid < 1$, respectively, the benefit of higher initial $\mathcal{B}$ due to its effect on higher $\mathcal{B}$ in the future is larger for the $\theta = H$ type when the initial $\mathcal{B}$ is low, and larger for the $\theta = L$ type when the initial $\mathcal{B}$ is high (the PP channel).

2. The WB(θ) channel's effect diminishes as j increases, while the WB($\mathcal{B}$) channel and PP channel may increase initially before decreasing, due to the compounding effect of repeated draws of ε^{θ}.

3. This model generates larger output volatility and fatter downside tails of monetary outcomes,

through the decoupling of the true type θ with the beliefs about the type $\mathcal{B}$.

4. The joint distribution of $\mathbf{S} = (\theta, \mathcal{B})$ converges to a unique invariant distribution ψ^*, which determines the long-run properties of the model and the stochastic steady state. Long-run welfare if the central bank's type is imperfect information may be higher or lower than if the central bank's type is fully known, depending on various parameter configurations.

1.5 Sticky Prices à la Calvo: the Confidence Channel

I now augment the model by adding nominal stickiness in prices, following Calvo (1983). The dynamics of the model are substantially affected by the interaction of the reputation for competence channel with this additional source of nominal frictions. I follow a similar approach to Angeletos and La'O (2009), one of the few papers to simultaneously combine Calvo frictions with incomplete information in a hybrid model. Two main differences arise, relative to their model: the public signal's precision remains unknown to firms, adding an additional degree of uncertainty, and monetary shocks follow not an exogenous stochastic process but come from errors made by an optimizing central bank.[28]

As standard, suppose now that a fraction ϕ of firms cannot change their firms in any given period. The firms that can change their prices must then take two things into account: they may not be able to change prices in future periods, and because of strategic complementarities, their optimal price will tilt towards the existing price level p_{t-1}. As shown in previous sections, in the flexible price equilibrium, firms would optimally choose

$$p_t^{i*} = p_t + \zeta \widetilde{y}_t$$

As in standard Calvo-pricing models, the optimal price in period t can be written as

$$p_t^i = \mathbb{E}_t \left[(1 - \beta\phi) \sum_{j=0}^{\infty} (\beta\phi)^j p_{t+j}^{i*} \middle| \mathcal{I}_t^i \right]$$

[28]One potential difficulty arises from the changing linear pricing coefficients each period; this complicates the analysis, as when prices are inter-temporally sticky, the optimal price depends on price levels in *future* periods, which are determined by pricing coefficients in that period. However, as shown in the appendix, these effects wash away in equilibrium when a central bank fully optimizes, conditional on its information set.

With sticky prices, the central bank's period welfare function can be written as[29]

$$U_t = -\left[\widehat{y}_t^2 + \kappa\left(\frac{\phi}{1-\phi}\pi_t^2 + (1-\phi)\triangle_t + \phi\mathcal{V}(p_{t-1}^i)\right)\right] \tag{1.17}$$

where $\triangle_t$ denotes the dispersion of prices among firms changing their prices in period t, $\pi_t = p_t - p_{t-1}$ denotes inflation, and as before $\kappa = \frac{\eta}{\zeta}$. Again, the determinants of welfare come from output gap variability, as well as cross-sectional price dispersion, but now with Calvo frictions there are three different sources of price dispersion at period t: the residual price dispersion carrying over from the previous period ($\mathcal{V}(p_{t-1}^i)$), price dispersion coming from movements in the price level (π_t, generating differences between firms setting their prices today and those that are not), and the additional price dispersion coming from dispersed information, which leads to dispersion among firms setting their prices in the current period ($\triangle_t$). Note that with Calvo frictions, price dispersion today affects price dispersion in future periods, which must be taken into account when analyzing welfare implications. In particular, a marginal increase in dispersion today of $d\mathcal{V}(p_t^i)$ leads to a $\phi^j d\mathcal{V}(p_t^i)$ increase of dispersion in period j, for a total long-run effect on welfare of $\frac{d\mathcal{V}(p_t^i)}{1-\beta\phi}$.

I now show that the following analog to Proposition 1 holds:

Proposition 9. *There exists a linear equilibrium in which the central bank of type θ sets $q_t^\theta = S_t^\theta + p_{t-1}$ in each period, and where the pricing strategy of a firm is given by*

$$p_{i,t} = (1-\widehat{\gamma}_{1t})\,p_{t-1} + \widehat{\gamma}_{1t}q_t + \widehat{\gamma}_{2t}S_t^i$$

where

$$\widehat{\gamma}_t \equiv \widehat{\gamma}_{1t} = -\widehat{\gamma}_{2t} = \frac{\zeta(1-\beta\phi)\tau_F}{\mathcal{H}_t - \big(1-\zeta(1-\beta\phi)\big)(1-\phi)\tau_F}$$

and the aggregate price level is given by

$$p_t = p_{t-1} + \Psi_t\varepsilon_t^\theta$$

where $\Psi_t = (1-\phi)\widehat{\gamma}_t$. The coefficients satisfy the following properties:

(i) $\widehat{\gamma}_t$ are Ψ_t are decreasing in $\mathcal{B}_t$

(ii) $\widehat{\gamma}_t$ are Ψ_t are increasing in ζ and decreasing in ϕ

Proof. See Appendix. □

[29] Appendix A provides a derivation.

The policy $q_t^\theta = S_t^\theta + p_{t-1}$ has a similar interpretation to the action $q_t^\theta = S_t^\theta$ policy without Calvo frictions: again, the central bank would like to track the natural level of output, and thus nominal demand is set while taking into account the prevailing price level. Here, the divine coincidence holds, and so minimizing output gap variability also minimizes movements in the price level that generate inefficient price dispersion. The linear pricing coefficients are similar to those in the case without Calvo frictions, with a slight adjustment taking into account the additional nominal friction; as $\phi \to 0$, the coefficients converge to the ones found earlier, that is $\lim_{\phi\to 0} \widehat{\gamma}_{1t} = \gamma_{1t}$ and $\lim_{\phi\to 0} \widehat{\gamma}_{2t} = \gamma_{2t}$. Property (i) simply states that when firms put more weight on the central bank being of high competence, they put less weight on their own idiosyncratic signals to set prices. Property (ii) states the relative weight on private signals also decreases with the degree of strategic complementarities (captured by $1-\zeta$) and that Calvo frictions add to the degree of 'effective' strategic complementarity, as the mass of firms stuck at the previous period's price level acts as a gravitational force for more price sluggishness.[30] Plugging in the findings from Proposition 9 yields the following proposition regarding the expected welfare function:

Proposition 10. *In equilibrium, expected welfare is*

$$\mathcal{U}_t\left(\theta_t, \gamma(\mathcal{B}_t)\right) = -\left[(1-(1-\phi)\widehat{\gamma}_t)^2 \frac{1}{\tau^{\theta_t}} + \kappa\left(\frac{\phi}{1-\phi}\left((1-\phi)\widehat{\gamma}_t \varepsilon_t^\theta\right)^2 + (1-\phi)\frac{\widehat{\gamma}_t^2}{\tau_F} + \phi\mathcal{V}(p_{t-1}^i)\right)\right]$$

As $\mathcal{B}_t$ increases, output gap variability increases, while inflation variability and price dispersion among firms changing prices decreases, that is

$$\frac{\partial \mathbb{E}_t\left(\widetilde{y}_t^2 \mid \theta, \mathcal{B}_t\right)}{\partial \mathcal{B}_t} > 0, \ \frac{\partial \mathbb{E}_t\left(\pi_t^2 \mid \theta, \mathcal{B}_t\right)}{\partial \mathcal{B}_t} < 0, \ \frac{\partial(\triangle_t \mid \theta, \mathcal{B}_t)}{\partial \mathcal{B}_t} < 0$$

Taking into account the latter two's effect on future cross-sectional price dispersion, the latter two's effects on welfare outweigh the first, that is

$$-\left\{\frac{\partial \mathbb{E}_t\left(\widetilde{y}_t^2 \mid \theta, \mathcal{B}_t\right)}{\partial \mathcal{B}_t} + \frac{\kappa}{1-\beta\phi}\left(\frac{\phi}{1-\phi}\frac{\partial \mathbb{E}_t\left(\pi_t^2 \mid \theta, \mathcal{B}_t\right)}{\partial \mathcal{B}_t} + (1-\phi)\frac{\partial(\triangle_t \mid \theta, \mathcal{B}_t)}{\partial \mathcal{B}_t}\right)\right\} > 0$$

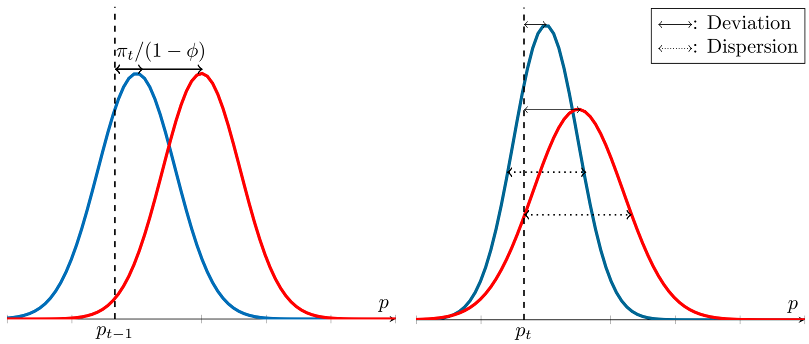
Proof. See Appendix. □

[30] Comparing the coefficients to the case without Calvo frictions shows two main effects that arise from sticky prices: noting from before that $(1-\zeta)$ was a measure of the degree of strategic complementarity, the presence of sticky prices acts as a multiplier on ζ by scaling it down by a multiplicative factor of $(1-\beta\phi) < 1$. This effect is compounded by the possibility that a firm's price will be stuck: because in expectation the price level remains fixed, there is additional impetus for a firm to set prices near the price level that obtains at the end of the period.

As in the model without Calvo frictions, output gap variability increases as firms put more weight on the central bank having high competence, while the price dispersion among firms changing their prices in the current period decreases, and is predetermined in terms of beliefs; the rationale remains exactly the same. The key change relative to before is the squared inflation term, arising from the beneficial welfare effects that come from a stable price level when Calvo frictions are present. Movements in the price level are a function of two factors: monetary errors, where the central bank misjudges how its nominal demand policy incentivizes firms to set prices in conjunction with their own idiosyncratic signals, and crucially, the time-varying coefficient Ψ_t, which is decreasing in $\mathcal{B}_t$. When $\mathcal{B}_t$ is higher, the effects of any monetary error on inflation is reduced. The reasoning comes from a 'confidence' effect: with imperfect information, if the central bank attempts to target a price level (here the no inflation price level), it must incentivize firms through its nominal demand policy to set prices in line with that target.[31] When the central bank has a high reputation for competence, firms are more inclined to believe that whatever action taken by the central bank is indeed such that their own ex-post optimal price is near the implicit zero-inflation target of the central bank, and they thus set prices accordingly. In essence, because *optimal* policy leads each firm's own optimal price to be equal to the previous period's price level, faith in the knowledgeability of the central bank to enact this optimal policy causes firms to ignore their private signals and simply set their prices close to p_{t-1}, generating minimal movements in the price level. This *confidence channel* is the second channel through which a reputation for competence is beneficial for welfare.

Figure 1.10 illustrates the essence of the reputational accelerator in terms of both the coordination and confidence channel, by again considering an example where the state variables remain fixed heading into a given period, but in one case the central bank makes a relatively small error and in the second case it makes a relatively large error. While the dispersion of prices among firms changing their prices is the same (because $\mathcal{B}_t$ is the same), the latter case leads to a larger change in the price level. Heading into next period, $\mathcal{B}_{t+1}$ increases under the former case, and decreases in the second case. In the next period, price dispersion among firms changing their prices is smaller under the former case (coordination channel), and *given the same monetary error* in the next period

[31]Note here that the divine coincidence holds, so the central bank indeed attempts to set policy that generates zero inflation.



(a) Possible outcomes **(b)** Future inflation and dispersion

Figure 1.10: Illustration of the coordination and confidence channels. In figure (a), the central bank enters period t with given beliefs $\mathcal{B}_t$. In case 1 (blue), the central bank achieves a good monetary outcome and in case 2 (red) the central bank makes a large error, leading to a bad monetary outcome. Under case 1, in period $t+1$, the dispersion of prices among firms changing their prices is lower (coordination channel), and the change in the price level is also smaller, given the same monetary error (confidence channel).

the magnitude of inflation is also smaller (confidence channel).

Given Proposition 10, the main conclusions from the previous section remain; in particular, a variant of Proposition 7 and Proposition 8 can be shown to hold, and the dynamics of the reputational accelerator are also similar. In summary, Calvo frictions provide an additional channel through which the reputation for competence can help achieve better welfare outcomes, via a confidence effect.

1.6 Equilibrium and Efficient Coordination

I now tie the model to the existing literature on public and private information, by analyzing the inefficient use of information that gives rise to the existence of the coordination channel, where a higher reputation for competence helps achieve better welfare outcomes by inducing more price coordination. In order to illustrate the mechanism more clearly, and to make comparisons with existing work as simple as possible, I again assume the case with no Calvo frictions ($\phi = 0$).

I consider a constrained efficient benchmark widely used in dispersed information models, where a social planner can dictate to agents how to act conditional on their information set, but cannot move information around across agents. As shown in Angeletos and Pavan (2007), with a quadratic welfare function and a Gaussian signal structure, the optimal allocation under this benchmark can be achieved via linear decision rules of the agents in response to the public and private signals.[32] Following Hellwig (2005), this linear decision rule, which makes socially optimal use of locally available information, is called the *Decentralized Information Optimum*. The DIO is solved by using a benevolent social planner's welfare function (which is simply the central bank's welfare function, and the welfare of the representative household), and taking first order conditions with respect to the pricing coefficients γ_{1t} and γ_{2t}, which I then compare with the FOC that arise from the firm's optimal pricing equation. The Appendix shows that the coefficients solving the social planner's problem satisfy

$$\gamma_{1t}^{*} + \gamma_{2t}^{*} = 0$$

$$(1 - \gamma_{1t}^{*})\frac{1}{\tau^{\theta}} + \gamma_{2t}^{*}\frac{\eta}{\zeta\tau_F} = 0$$

where the first and second equations are the FOC corresponding to γ_{1t} and γ_{2t}, respectively, whereas the equilibrium coefficients satisfy

$$\gamma_{1t}^{EQ} + \gamma_{2t}^{EQ} = 0$$

$$(1 - \gamma_{1t}^{EQ})\frac{1}{\mathcal{H}_t - \tau_F} + \gamma_{2t}^{EQ}\frac{1}{\zeta\tau_F} = 0$$

Thus, the FOCs γ_{1t} hold in each period t. Firms thus make optimal use of public information in order to minimize aggregate output gap volatility (note that γ_{1t} is the response to a common shared variable q_t, and thus only affects output gap volatility and not price dispersion). In contrast, rearranging the FOC for γ_{2t} yields

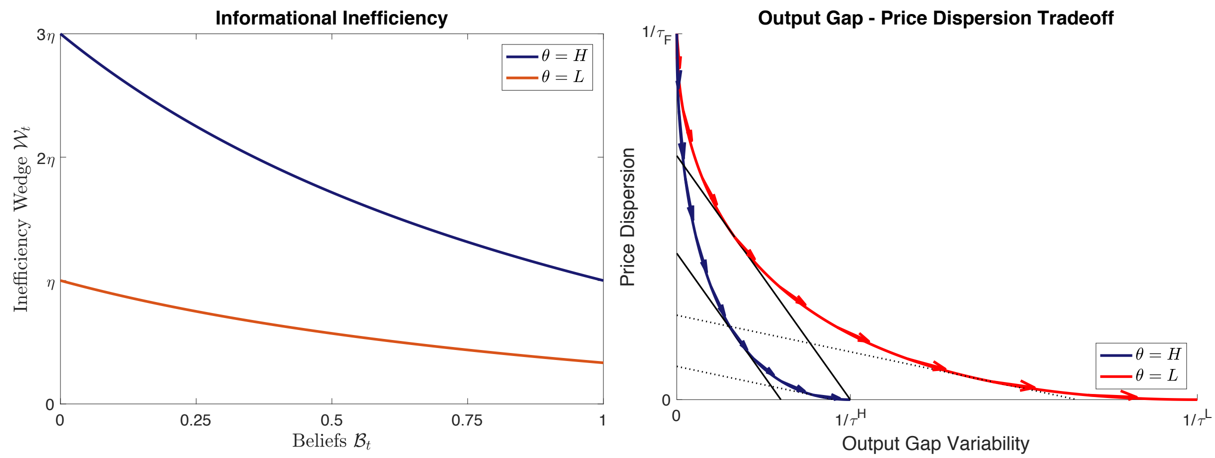
$$\gamma_{2t}^{*} = -\left(1 - \gamma_{1t}^{*}\right)\frac{\zeta\tau_F}{\eta\tau^{\theta}}$$

$$\gamma_{2t}^{EQ} = -\left(1 - \gamma_{1t}^{EQ}\right)\underbrace{\frac{\zeta\tau_F}{\mathcal{H}_t - \tau_F}}_{\omega^{EQ}(\mathcal{B}_t)} < -\left(1 - \gamma_{1t}^{EQ}\right)\underbrace{\frac{\zeta\tau_F}{\eta\tau^{\theta}}}_{\omega^{*}(\theta_t)}$$

where the inequality holds for *all* $\mathcal{B}_t$ and θ_t, provided condition $\frac{\tau^G}{\tau^B} < \eta$ holds, which is guaranteed by Assumption 3.[33] Thus while firms make optimal use of public information, they make inefficient

[32] By Assumption 1, the linear-Gaussian structure is preserved even when the underlying precision of the public signal is unknown.

[33] See Appendix for a formal proof.



(a) Inefficiency Wedge **(b)** Output gap and price dispersion as a function of beliefs

Figure 1.11: Inefficiency wedge and the output gap - price dispersion tradeoff. The figure on the left illustrates the inefficiency wedge $\mathcal{W}$ as a function of the true type θ and beliefs $\mathcal{B}$. The figure on the right shows how changes in γ affect output gap variability $\mathbb{E}\widetilde{y}^2$ and price variability $\mathcal{V}(p_t^i)$; as γ decreases (less relative weight on the private signal), the outcome pair moves in the direction of the arrows. Parallel lines indicate indifference curves for the central bank, with different slopes reflecting different possibilities of $\kappa = \eta/\zeta$.

use of *private* information: recalling that $1 - \gamma_{1t}$ represents the response to the public signal component of q_t, and that this component and γ_{2t} and both negative, it follows that ω represents the weight that firms put on their own private signal, relative to the public signal. Because $\omega^{EQ}(\mathcal{B}) > \omega^*(\theta)$ for all possible $\mathcal{B}$ and θ, it follows that in each state of the world, firms put *too much* relative weight on their own private signals. The intuition remains similar to that of Hellwig (2005) and Angeletos and Pavan (2007): changes in γ_{2t} affect both output gap variability as well as price dispersion; firms do not fully internalize the costs arising from socially inefficient price dispersion (captured by η) that arises from conditioning prices on private information, creating a wedge between the equilibrium and socially optimal amount of coordination. This can be characterized by what I call the *inefficiency wedge*, defined by

$$\mathcal{W}_t\left(\theta_t, \mathcal{B}_t\right) \equiv \underbrace{\frac{\zeta\tau_F}{\mathcal{H}_t - \tau_F}}_{\omega^{EQ}(\mathcal{B}_t)} \Big/ \underbrace{\frac{\zeta\tau_F}{\eta\tau^{\theta_t}}}_{\omega^*(\theta_t)} = \frac{\eta\tau_t^{\theta}}{\mathcal{H}_t - \tau_F} > 1$$

Figure 1.11a illustrates this inefficiency wedge as a function of the true type and of the beliefs of firms. This wedge illustrates a few novel properties of this model. First, both the equilibrium and socially optimal degree of coordination are time-varying, whereas in the existing literature, both are fixed. Here, time-variation in the true type θ leads to changes in the socially optimal degree of coordination $\omega^*(\theta)$ through τ^{θ_t}, whereas time-variation in the beliefs $\mathcal{B}_t$ affect the equilibrium degree of coordination $\omega^{EQ}(\mathcal{B})$ through $\mathcal{H}_t$. Furthermore, because the true underlying type and beliefs about the type do not move together one for one (due to the imperfect observability of the type), it follows that the wedge itself must be time-varying. If instead the true type is perfectly observed, then $\mathcal{H}_t = \tau^{\theta_t} + \tau_F$ and it follows that $\mathcal{W}_t = \eta$, and the inefficiency wedge is perfectly captured by elasticity of substitution across goods, a time-invariant parameter.[34] Thus, even if the true type moves around, what generates time-variation in the inefficiency wedge is the *decoupling* of beliefs with the true type.

Because $\mathcal{W}_t(\theta_t, \mathcal{B}_t) > 1$ (obtaining a minimum at $\mathcal{W}^{min} = \eta \frac{\tau^B}{\tau^G}$ and a maximum at $\mathcal{W}^{max} = \eta \frac{\tau^G}{\tau^B}$), it follows that the inefficiency wedge is always present, and always in the same direction: in each possible state the social planner would like to alter the equilibrium response in such a way that firms respond less to their own private information, conditional on their response to the public signal. This is how a reputation for competence is beneficial to the central bank: by increasing $\mathcal{H}_t$, it decreases the firms' conditioning on private information, inducing more coordination and alleviating the inefficiency in the use of private information, which brings the equilibrium closer to the constrained efficient benchmark.

For a given belief, the inefficiency wedge is also greater for the high-competence central bank, as shown Figure 1.11a. While both types benefit from an increase in their reputation for competence, the inefficiency wedge is more sensitive to a marginal change in beliefs for the $\theta = H$ type, that is

$$\left.\frac{\partial \mathcal{W}_t}{\partial \mathcal{B}_t}\right|_{\theta=H,\mathcal{B}_t} < \left.\frac{\partial \mathcal{W}_t}{\partial \mathcal{B}_t}\right|_{\theta=L,\mathcal{B}_t}$$

noting this derivative is negative. This result is the exact counterpart to a result derived earlier, where the central bank of type $\theta = H$ benefits more from a marginal increase in $\mathcal{B}$ then the $\theta = L$ type, given the same starting beliefs: while both types benefit equally from increased equilibrium coordination from an increase in $\mathcal{B}_t$, the cost in terms of potential output gap variability is higher for the $\theta = L$ term, due to output gap variability coming from monetary errors. It additionally follows that the socially optimal degree of coordination $\omega^*(\theta)$ is higher when τ^θ is higher, as again

[34]Recall that the welfare costs of price dispersion are increasing in η; this social cost is not taken into account by firms, which leads to inefficient price dispersion. It follows that this inefficiency is then increasing with η.

the potential harm of coordination is less present when the central bank receives more precise signals.

Figure 1.11b illustrates the effects that changes in γ_t have on the equilibrium output gap variability $\mathbb{E}_t\tilde{y}_t^2$ and price dispersion $\mathcal{V}(p_t^i)$. As γ_t decreases, there is less price dispersion at the expense of increased variability in the output gap, shown by the direction of the arrows in the figure; the tradeoff is more favorable for the $\theta = H$ type, as described earlier. The optimal γ^*, that is, the one that arises as the solution to the social planner's decentralized information optimum, is given by the point where this curve is tangent to the central bank's indifference curves, which are linear in price dispersion - output gap variability space. The higher $\kappa = \frac{\eta}{\zeta}$ is, the flatter these indifference curves are, and the more the social planner emphasizes price dispersion, leading to a social optimum further along the γ curve.

1.7 Conclusion

While central banks appear to care greatly about their reputation for competence, the reasons as to why have remained elusive. This paper built a model in which the central bank's reputation for competence is endogenously determined by past outcomes, and whereby this reputation has real effects on the transmission of monetary policy through its effects on the pricing behavior of firms. Taking a standard macroeconomic model with dispersed information, along with a simple, often-used characterization of central bank competence, I embedded the imperfect observability of this competence and derived several new implications. When the competence of the central bank, defined as the precision of the signals it receives about the state of the economy, is imperfectly known by agents, the beliefs that firms have about the precision of the central bank's signals play a large role in how firms respond to any action taken by the central bank. Two channels through which a reputation for competence improves welfare outcomes emerge: the coordination channel, whereby firms coordinate their prices on what they believe to be a more precise signal, and the confidence channel, whereby firms trust that the central bank's implemented policy is such that the price consistent with the central bank's implicit inflation target is indeed privately ex-post optimal. The theory also reconciled the observation that the reputation for competence appears heavily influenced by macroeconomic aggregate outcomes, which reveal information about the central bank's proficiency in reading the state of the economy. This paper illustrated what I call the *reputational accelerator*, whereby undesirable macroeconomic outcomes endogenously lead to further undesirable outcomes, and vice-versa, through a reputational effect. Lastly, I provided empirical evidence in support of the model, showing that the strength of information effects in the U.S. appears stronger when forecasts by the Fed were more accurate in the recent past.[35]

This paper thus provides a new theory, separate from private incentives,[36] for why central banks might care about their reputation for competence.[37] Anecdotally, this may provide an additional explanation for why many central banks place strong emphasis on unanimous support or a strong consensus of decisions, as well as provide an additional reason for why central banks appear to be so averse to reversals, which may come with the implicit conclusion that the central bank has erred.[38]

This paper also tied into the general theory of public information with dispersed information, providing a framework to analyze the effects of public information when the informativeness of the public signal is imperfectly known. The decoupling of beliefs about the precision with the true precision of the public signal leads to a time-varying wedge between the equilibrium and efficient use of information, with this wedge depending crucially on past outcomes. In addition, this decoupling generated excess output volatility and large downside tail risk of welfare outcomes, and endogenously generated central bank 'Minsky moments' whereby long periods of stability

[35]This is also consistent with the observation that private sector inflation expectations continue to be substantially misaligned from the FOMC's inflation expectations announced via the Summary of Economic Projections; the Fed's inflation forecasts have now been inaccurate for a considerable period of time.

[36]As central bankers are human, the innate human desire to appear competent and receive praise is one simple reason central bankers may care about their reputation. A natural alternative story could be career-concerns as in Holmstrom (1999). Another possibility is an environment where contracts implicitly or explicitly link pay for performance. For a theoretical model studying optimal contracts for central bankers where competency is unknown, see Walsh (1995).

[37]A very different explanation is offered by Moscarini (2007), another paper that looks at the beneficial effects of a central bank's reputation for competence, in a game-theoretic model with a central bank wishing to inflate. There, competence helps credibility, as there the central bank faces a tradeoff between taking the appropriate action and communicating this to align expectations accordingly, while also wanting to induce low expectations and surprise the economy with high inflation. With higher competence, the central bank puts more weight on the former, and thus its credibility in not engaging in surprise inflation is strengthened. While one could also interpret the gains from competence in the model here as a credibility channel, the notion of credibility is very different. In that model, the central bank's competence acts as a counterweight to the desire to inflate; here, the competence affects the central bank's credibility as an informative public signal. In the former case, the importance is the central bank's perception of its *own* competence; thus, for instance, central bank overconfidence in its own ability is beneficial. In my model, what matters is the *private sector*'s perception of the central bank's competence, with the true competence also playing a large allocative role.

[38]See the second essay. For motivation, see the discussion in Evans et al. (2016)

could potentially lead to large instability.

While this paper focused on a particular macroeconomic setting, the developed framework can be applied in the myriad situations where public information plays a central role. While a few parametric assumptions were made in order to apply the theory to this macroeconomic setting, focusing on the general case yields multiple possible extensions. For instance, one can extend this model to cases with a continuum of types, with general welfare functions, and a relaxation of the diffuse prior assumption.[39] In addition, while the model in the paper was purposely stripped down in order to focus on the reputation for competence channel, this framework can also be embedded into a larger-scale quantitative model, or a more elaborate model in which the unobservability of the central bank's type affects the strategic actions taken by the central bank. These are the subjects of ongoing and future research.

[39] In another paper focusing on a more general setting, I analyze these cases. One key assumption that can be relaxed is the rigidity of beliefs within a period, which in this model was imposed to provide clean analytic solutions for this particular case.

Chapter 2

Monetary Policy Reversal Aversion

2.1 Introduction

Throughout the world, reversals of monetary policy decisions by central banks are rare. After a movement in the policy interest rate in one direction, the subsequent change in the policy interest rate is in the opposite direction at a much smaller fraction than changes in the policy interest rate in the same direction. Figure 2.1 displays the target interest rate (Federal Funds Rate) of the U.S. Federal Reserve since 1982, with reversals marked in red. As seen in the figure, continuations also appear to occur in rapid succession, whereas reversals rarely take place within a short time-frame. In the time frame shown in Figure 2.1, an average of 49 days separate continuations in the Federal Funds target rate, compared to 239 days separating reversals.[1] In addition, the prevalence of reversals has decreased over time, and the days separating interest rate movements during reversals have increased.

To explain these empirical facts, academics have mostly focused on *gradualism*, or why central banks might want a gradual path of interest rates on the way to hitting an end target rate. Explanations for gradualism include uncertainty about the economy, from uncertainty about underlying model parameters[2] to uncertainty due to potential mismeasurement in relevant macroeconomic

[1] The long period near the zero lower bound before the recent hike in interest rates is an outlier. Excluding this period, the average time separating reversals of decisions is 150 days.

[2] See Brainard (1967) and Sack (1998)

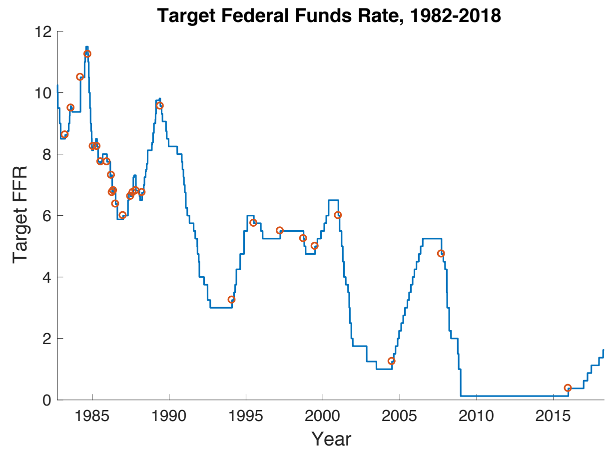


Figure 2.1: Federal Funds Target Rate, 1982-2018. Reversals marked in red circles. Since December 2016, the target range had an upper and lower limit, in which case the figure shows the average of the two.

variables,[3] the ability for more control over long-term rates with partial adjustment of short-term rates,[4] and aversion to bond market volatility.[5] Due to gradualism, interest rate movements are likely to be followed by interest rate movements in the same direction, thus making reversals appear relatively less likely. This paper, however, focuses on a separate, distinct issue: central banks may be averse to reversals themselves, rather than the lack of reversals coming as a byproduct of gradualism.

Why might this be the case? In this paper, I show that if a central bank cares about its reputation for being knowledgable, reversal aversion can arise because frequent reversals might show that the central bank has likely erred. As the quote in the introduction of the dissertation shows, central bank policymakers are often loath to reverse rates due to outsiders perceiving such reversals as evidence that the policymaker is incompetent and prone to mistakes.[6] Even as central banks have

---

[3]See Orphanides (2003)

[4]See Goodfriend (1991) and Woodford (2003c)

[5]See Stein and Sunderam (2017)

[6]A few other quotes nicely illustrate this concern:

> "If we are perceived to have tightened and then have been compelled by market forces to quickly reverse, our reputation for professionalism will suffer a severe blow"
>
> - Alan Greenspan (former Chairman of the Fed), FOMC speech July 1996

> "Citizens, and perhaps even markets...might misinterpret a sequence in which interest rates first rise and then fall as prima facie evidence that the bank had erred. That belief, in return, might undermine the

improved their credibility over the years, reputation-based reversal aversion appears to continue to play a large role in the monetary policy decision making process, with exceedingly rare interest rate reversals in conjunction with harsh criticism for the reversals enacted by the ECB, Swedish Riksbank, and Bank of New Zealand in recent years, and talks of painful blows to credibility if the Fed had to reverse course after its first rate increase in over a decade. Surprisingly, this issue, while frequently discussed by policymakers and media analysts of monetary policy, has received scant attention in the academic literature. No formal model exists which captures the intricate dynamics between reversal aversion, conservatism, and central bank reputation formation in the context of imperfect information about the central bank's knowledgeability or competence. This paper represents a first attempt at building a theoretical framework to understand these issues.

I develop a model in which central banks differ in their ability to correctly gauge the state of the economy. Central banks receive imperfect signals about the underlying state, following a burgeoning literature analyzing monetary policy in the context of imperfect information.[7] Crucially, the precision of the signals that the central bank receives is private information to the central bank. A more knowledgable central bank receives a more precise signal that guides where interest rates should be set, and the central bank would like to develop a reputation for being the more knowledgable type. While this desire is an ad-hoc assumption in this paper, a central bank may wish to signal its knowledgeability for myriad reasons; for instance, a reputation for making interest rate mistakes can send mixed messages, not least due to financial markets which may cause excess asset price volatility. As discussed briefly (but not modeled) in an influential paper by Evans et al. (2016), policymakers may also be reluctant to reverse course because doing so may cause the public to lose confidence in the central bank's ability to understand and stabilize the economy. In Choi (2018) I show that social welfare may be improved when the central bank is seen as knowledgable, through coordination and confidence effects that reduce relative price distortions. There, the mere *perception* that the central bank is knowledgable can help improve real outcomes, regardless of

bank's credibility"

- Alan Blinder (former vice Chairman of the Fed), 2006

[7]Influential papers include Phelps (1970), Lucas (1972), King (1982), Townsend (1983), Phelps (1983), and Woodford (2003a), and Svensson and Woodford (2003, 2004)

the true underlying knowledgeability of the central bank. Yet another, and potentially very likely, possibility is that policymakers are susceptible to the natural human desire to be seen as competent, and praised rather than criticized. While these are all possibilities, it remains unclear why central banks appear to care deeply about their reputation for knowledgeability. This paper takes no stand on whether or not the central bank's desire to be seen as knowledgable stems from personal benefits or from trying to improve social welfare; it simply takes as given that the central bank cares, and analyzes how this affects equilibrium actions. While the model's results are not dependent on the source of the central bank's welfare function, it does change how the model's results might be interpreted from a normative standpoint. If the perception that the central bank is knowledgable improves equilibrium outcomes, reversal aversion can arise from a purely benevolent central bank with preferences perfectly in line with that of society; on the other hand, if it is purely for private reasons that the central bank wants to be perceived as knowledgable, reversal aversion can lead to distortions that ultimately have detrimental effects on society.

In the model, because more knowledgable central banks (i.e., those with more precise signals about the appropriate nominal interest rate to set) are less likely to reverse course in equilibrium, central banks may avoid reversals in order to mimic a more knowledgable monetary authority. Two distortions arise, relative to a frictionless world in which the knowledgeability of the central bank is perfectly known by the public. First, the central bank may not reverse course even when doing so is appropriate when reputational considerations are not present. Second, foreseeing the possibility of wanting to reverse in the future, and thus having to choose between taking a reputational hit or setting an inappropriate interest rate, a central bank may choose to wait until more uncertainty is resolved before changing the interest rate.[8] The intuition has close parallels to the real options literature, in which there is option value to waiting until uncertainty is resolved before taking an action that is irreversible or costly to reverse (e.g., investment). A key difference here is that the

[8]The following quote by policymakers illustrates that the desire to avoid interest rate reversals may indeed be a reason for reacting slowly to shocks:

> "Reversals could be misunderstood, creating uncertainty both in financial markets and in the wider economy and damaging the credibility of the MPC process...the desire to minimize the risk of policy reversals was likely to mean that interest rate changes would, on average, be made too late"
>
> - Bank of England Monetary Policy Committee, March 1998 Minutes

costs to reversing are endogenous to the model, and arise from the equilibrium actions taken by agents; the costs differ depending on which actions occur in equilibrium. In contrast, in the real options literature the costs are exogenous and unaffected by the actions taken by the agent. Taken together, these two distortions, which I call *reversal aversion* and *conservatism*, lead to central banks not reversing their rates frequently, and also reacting slowly ("falling behind the curve") to shocks. These are both observed looking at central banks around the world, but are hard to explain jointly in existing theoretical frameworks. A key novel feature of the model is that conservatism comes from the desire to protect the central bank's reputation for knowledgeability, which has not yet been explored in the context of an economic model.

More generally, this paper aims to address how monetary policy is affected when a central bank is concerned with the public's perception of its knowledgeability. Outside of reversal aversion and conservatism, a central bank may dynamically shape its reputation for knowledgeability in other dimensions; for instance, reacting aggressively to a shock may show confidence that bolsters a central bank's reputation for knowledgeability going forward, and such considerations may dampen the incentives for conservatism. I aim to provide a simple model that takes these myriad forces into account, illustrating a few key mechanisms.

The remainder of the paper is organized as follows. Section 2 outlines the model and discusses the key assumptions. Section 3 analyzes the model's equilibria and the distortions that arise from reversal aversion. Section 4 extends the model to allow for different assumptions and informational structures. Section 5 concludes and discusses the key takeaways to be drawn from this paper.

2.1.1 Related Literature

Other papers that model reversal aversion include Kobayashi (2010), who simply takes monetary policy aversion as given by imposing an irreversibility constraint on the control space, and then analyzes the resulting effects this constraint on policy has on the economy, and Ellison (2006), who analyzes a model with data and parameter uncertainty in a learning model and shows that monetary policy with frequent reversals makes learning about key parameters more difficult, thus showing that optimal policy has a lower number of reversals. In Woodford (2003b), a central bank may choose not to reverse in order to build a reputation for gradualism, but this is in order to maintain leverage over long-term rates with minimal movements in short-term rates, rather than from any desire to come across as knowledgable. The model in this paper differs from the previous models in that the desire to form a reputation for knowledgeability and inerrancy—a desire mentioned frequently by policymakers past and present—takes front and center stage.

In a concurrent paper, Choi (2018) analyzes the role that perceptions of central bank knowledgeability have on the effects of monetary policy and aggregate outcomes. While that paper provides a model in which central banks care about their reputation for knowledgeability in the context of a dynamically optimizing central bank, the optimal monetary policy remains undistorted relative to a benchmark model in which the knowledgeability of the central bank is common knowledge, and the focus of the paper is to explain *why* central banks may want to come across as being knowledgable, and potentially much more knowledgable than they really are. This paper instead takes as *given* this desire, and in a simple game theoretic setting, shows how this desire can help explain a few salient empirical patterns in how monetary policy is set throughout the world, such as the lack of reversals, conservatism, and reacting slowly to shocks. Another paper where the perceptions of the central bank's knowledgeability affects aggregate outcomes is Moscarini (2007), who shows that a central bank perceived as "competent" (knowledgable) is expected to use its discretion relatively more to pursue its stated targets, rather than to engineer surprise expectations and stimulate output, thus improving the output-inflation tradeoff in a setting in which a central bank faces time-inconsistency with an inflation bias. However, in his model the competence of the central bank is common knowledge, and thus the model has no role for a central bank to manipulate the perception of its competence in any way, which is the focus of this paper.

Lastly, this paper is related to the large literature of developing reputation in game theoretic settings, most often couched in terms of a game in which some players are uncertain about the payoff structures of their opponents. Leading examples in the context of monetary policy include Barro and Gordon (1983a) and Backus and Driffill (1985), in which a central bank must attempt to improve its reputation for aggressively fighting inflation in order to alleviate an inflation bias that worsens outcomes with rational agents. The focus on reputation in this paper instead concerns a central bank attempting to improve its reputation for *knowledgeability*, with its payoff structure completely known by the public. A closely related paper is Prendergast and Stole (1996), who also analyze a model in which competence (knowledgeability) is unobserved, with the decision maker taking into account how its actions affect its reputation for knowledgeability, which enter into the payoff function for the decision maker. Like this paper, reversing a prior decision provides evidence to others that it may have made a mistake in the past, damaging its reputation. However, the authors assume that the decision maker only cares about its payoff in a single contemporaneous period, and is thus fully myopic. A crucial element of this paper is how actions in one period may affect payoffs in the *future*, and in particular how the possibility of reversing in the future may lead a central bank to be more conservative than otherwise. This dynamic shaping of reputation for knowledgeability is a key new feature of the model.[9]

2.2 The Model

I begin by laying out the structure of the model, and later discuss the key assumptions.

2.2.1 Structure of Model

In each period $t = 1, 2$, the economy can be in one of two states, $\theta_t = L$ or H, representing "low" and "high" states, respectively. For simplicity, I assume that there are only two possible interest rate policies that the central bank (CB) can set, i^L or i^H. Without reputational considerations, the appropriate interest rate i_t in each period t is $i_t = i^{\theta_t}$ in state θ_t. For concreteness, $i^L < i^H$ and so by the "high" state ($\theta = H$) one can think of an economy in which output and inflation are above target, and a socially-efficient response requires a higher nominal interest rate $(i = i^H)$ set by the central bank.

The central bank's can be of one of two types, denoted by G ('good' or knowledgable) and B ('bad' or unknowledgeable). The G type observes the state perfectly each period, whereas the B type observes the state imperfectly. In particular, in period $t = 1$ the central bank gets an imperfect signal of the state of the economy s_t, where $Prob(s_t = \theta_t) = \tau$ and thus τ represents the precision of the signal that type B receives. I make varying assumptions about the central bank's signal

[9]As Prendergast and Stole (1996) show, with a fully myopic agent, agents may be tempted to be more aggressive instead of more conservative. Aggression shows confidence in a decision maker's information set, and because the agent is fully myopic, the potential of making a mistake having to reverse in the future has no bearing on the actions in the current period. While the confidence affect is present in the model of this paper as well, the possibility of this reversal in the future puts a counterweight on this incentive, muting this effect.

precision in period $t = 2$; in the simplest version of the model, the CB observes the state perfectly in period 2, which is meant to capture the resolution of uncertainty over time in the simplest manner possible. Crucially, the type is *private information* to the central bank, and is not directly observed by the public. In addition, I make the assumption that the public does not observe the true state of the world θ at any point, including past states of the world, although it knows the underlying stochastic process that governs the true state of the world; it thus must make inferences about the central bank's type through observing its actions only. On the other hand, both central bank types observe *past* states of the world perfectly.

The central bank attempts to maximize a welfare function consisting of two separate components: gains from setting the appropriate interest rate given no reputational considerations, as well as gains from being seen as a knowledgable central bank. The central bank's objective function is

$$W^{CB} = \sum_{t=1}^{2} \left\{ I\left(i_t = i^{\theta_t}\right) + \Gamma \widehat{\mu}_t \right\}$$

where $I(\cdot)$ is an indicator function, $\widehat{\mu}_t$ denotes the probability that the public puts on the central bank being of type G, and $\Gamma \geq 0$ denotes the relative weight that the central bank puts on being seen as the knowledgable type. Henceforth, I denote the first component of the welfare function the *fundamental component* and the second component the *reputation component.*

I make varying assumptions on how the state evolves. I first analyze a very simple structure, and in a later section analyze other possible setups. For now, I consider a case in which in period zero ($t = 0$), the economy is in state L, and the action by the central bank is $i_0 = i^L$; for simplicity, I assume that this initial state is perfectly known by each central bank type and the public. In future periods $t = 1, 2$, if the state has not changed from L to H in a previous period, the state changes to $\theta = H$ with probability λ. On the other hand, once the state changes to $\theta = H$, it remains there until the end of the game. Thus, there are three possible combinations of states of the world in periods 1 and 2; these are $\Theta : \{(H, H), (L, H), (L, L)\}$. Heuristically, one can think of this setup representing a case in which a hiking cycle is imminent, but the precise timing of when to optimally begin increasing interest rates is uncertain; once the economy reaches a critical juncture, a sequence of hikes is appropriate which can be broadly captured by playing i^H (that is, instead of representing a single interest rate, i^H may represent a regime in which the central bank begins to gradually hike rates). Such scenarios seem to accurately capture many interest rate policy decisions that monetary authorities face, such as the recent and current decisions by central banks on when to begin liftoff after a period of exceedingly low interest rates.

Lastly, I make the assumption that the bad central bank still receives a signal precise enough that it has an incentive to follow its signal, that is $\tau > 1 - \lambda$. This ensures that without reputational considerations, it sets policy according to its signal, as will be shown shortly.

2.2.2 Discussion of Assumptions

The model's setup is deliberately simple, with just two true states, two periods, and two types of central banks, in order to deliver the intuition of reversal aversion and conservatism in as straightforward a manner as possible. In addition, the true sequence of states does not allow for a change to state $\theta = H$ and then back to $\theta = L$; doing so simplifies the expressions in the remainder of the paper considerably, as it immediately gets rid of one possible sequence of states, thus making the calculation of posterior probabilities—both of the B type central bank's assessment of the possible sequence of states, and of the public's assessment of the central bank's type—much easier. Note that when $\lambda < 1/2, \theta_1 = H, \theta_2 = L$ is the least likely sequence of states of the world, with this probability going to zero quickly as λ decreases. Making this assumption does little to change the analysis at the benefit of a great deal of simplification. Again, one may think of this model as representing a situation in which a central bank must decide when to begin a hiking cycle; once the economy reaches a critical threshold for when a gradual sequence of hikes is appropriate, it is unlikely that the optimal interest policy requires instead a sequence of interest rate decreases. In the main extension of the model, I relax this assumption by allowing the true state to change back from $\theta = H$ to $\theta = L$ and show that the main results hold.

As discussed in the introduction, why the central bank wants to be seen as the G type is an ad-hoc assumption; it is not critical for this model, which attempts to explain what distortions might arise in monetary policy *if* a central bank cares about being perceived as being more knowledgable. The easiest interpretation is that reputation is a purely private gain or loss to the central bank(er), and is separate from the possible reputational effects on the real economy. As such, one may view the central bank's welfare function as having two distinct components: one representing the benevolent side of the central bank, that coincides with society's welfare function, and another component that represents a central banker's own selfish desire to appear knowledgable. This convenient separation allows one to define the 'optimal' interest rate as the one that a perfectly benevolent central banker would try to pick, and thus any deliberate attempt to deviate from this action can be interpreted as a pure welfare cost to society due to personal reputational motives. In a model in which the real economy itself is affected by the public's perception of the central bank's reputation, the interpretation of the 'optimal rate' becomes substantially more nuanced. Note also that the central bank's actions do not affect the state in the next period; the true state moves exogenously and independently of the central bank's decisions, and the job of the central bank is to simply respond to these exogenous movements in the economy. In reality, monetary policy likely has persistent effects on the economy in the future, and so the optimal rate in the future may be a function of the rate that is chosen today. I abstract away from this to keep the model as simple as possible, but later introduce an alternative setup that can accommodate this possibility.

Note that I assume that *both* types want to be known as the G type, and so both types are strategic players; nevertheless, I later restrict myself to parameter values such that the G type never distorts its actions in equilibrium. A very key assumption is that the public never knows the true state of the world, even states in the past (with the exception of θ_0 as given). This assumption broadly captures the idea that central banks tend to have informational advantages[10] as well as the idea that the appropriate interest rate a central bank should have set is often ambiguous and subject to much debate, and this debate often carries on long after an interest rate policy has been set.[11]

The stylized nature of the game aims to provide illustration of the key effects that may arise when central banks care about their reputation for knowledgeability. In practice, the degree to which this concern affects monetary policy is surely changing over time, depending on the environment that the central bank faces. At times, central banks may be more or less concerned

[10]See for example Romer and Romer (2000).

[11]Note that if the central bank cares about its reputation for knowledgeability because it improves aggregate outcomes, this assumption makes sense as long as whether or not the central bank made the right decision is not known until a sufficient lag. Because the benefits to reputation then accrue immediately, the central bank will have an incentive to distort actions to appear knowledgeable today, even if this is revealed to be the incorrect action later. If the central bank cares mostly its reputation for knowledgeability mostly for private benefits, it is likely that these benefits accrue long after the central banker makes its decision, in which case it is less likely that actions will be distorted today if this will be found out to be the wrong action in a later period.

about their reputation, have varying degrees of true underlying knowledgeability, face a changing amount of uncertainty, and may have already built a solid reputation (or vice-versa) through its actions in the past. These can be represented by changes in the underlying parameters Γ, τ, λ, and $\widehat{\mu}_0$, and the interplay of these parameter values together determine the extent to which distortions may arise from reputational considerations. One can think of these parameter values as time-varying in practice, with the short two period version of the model capturing a snapshot in time.

2.3 Equilibrium Analysis

2.3.1 Benchmark Case: No Reputational Concerns

To highlight the impact of reputational concerns on the central bank's equilibrium actions, I first consider a benchmark case in which there are no reputational concerns ($\Gamma = 0$, or the central bank type is fully observed and so $\widehat{\mu}$ is fixed in both periods at either 0 or 1) and the central bank's welfare function is thus simply given by

$$\widetilde{W}^{CB} = \sum_{t=1}^{2} \{[I(i_t = \theta_t)]\}$$

Without reputational concerns, each CB type will simply play the action corresponding to the state which has a higher posterior probability. In the baseline model, both types fully observe the true state in period 2, and play the corresponding action; in the first period, the good CB plays the action corresponding to the true state, which it fully observes. After receiving signal in period 1, the bad type puts probability

$$\widetilde{P}\,(\theta_1 = H \mid s_1 = H) = \frac{\tau\lambda}{\tau\lambda + (1-\tau)(1-\lambda)} > \frac{1}{2} \tag{2.1}$$

$$\widetilde{P}\,(\theta_1 = L \mid s_1 = L) = \frac{\tau(1-\lambda)}{\tau(1-\lambda) + (1-\tau)\lambda} > \frac{1}{2} \tag{2.2}$$

given that $\tau > 1 - \lambda$ and $\lambda \geq 1/2$. Thus, in period 1, the B type central bank plays i^H when its signal is H, and i^L when its signal is L. While reputation is not a concern for the central bank in this model, it is useful to calculate the posterior probabilities the public puts on the central bank being of each type given observed actions, in order to set the stage for analysis of the full model with reputational concerns. Given the equilibrium actions of the benchmark case, it is easily shown that

$$\widehat{\mu}_1 \left(i_1 = H\right) = \frac{\widetilde{\mu}_0 \lambda}{\widetilde{\mu}_0 \lambda + (1 - \widetilde{\mu}_0) \left[\lambda \tau + (1 - \lambda)(1 - \tau)\right]} \tag{2.3}$$

$$\widehat{\mu}_1(i_1 = L) = \frac{\widetilde{\mu}_0(1 - \lambda)}{\widetilde{\mu}_0(1 - \lambda) + (1 - \widetilde{\mu}_0) \left[\lambda(1 - \tau) + (1 - \lambda)\tau\right]} \tag{2.4}$$

Thus, with reputational considerations, the central bank's optimal action may change in the first period even in a one period model. Because the focus of the paper in period 1 is on conservatism generated by reversal aversion, I show later that the main results still hold in an alternative version of the model where the central bank only cares about the reputation it obtains at the end of the game, after period $t = 2$.[12] Nevertheless, in a general setting one must also take into account reputational effects unrelated to reversal aversion when solving for the equilibrium, and I analyze these forces together later in the paper.

What about the posterior probabilities after observing actions i_1 and i_2 by the central bank? Because the true state never reverts from H then back down to L, the G type will never play $i_1 = H$ and $i_2 = L$ and it immediately follows that

$$\widehat{\mu}_2 \left(i_1 = H,\, i_2 = L\right) = 0$$

This provides some intuition behind the driving forces of the model: playing $\left(i^H, i^L\right)$ is associated with being the less knowledgable type, and thus when there are reputational considerations, there are potential distortions that arise from incentives to avoid playing $\left(i^H, i^L\right)$. This can manifest itself in two ways: the central bank can avoid playing H in the first period (*conservatism*), and if H is played in the first period, avoid playing L in the following period (*reversal aversion*). Note again that the desire to play L in period 1 can come from both reputational considerations in period 1 and in period 2; it is the desire to protect reputation in period 2—that is the ability to affect $\widetilde{\mu}_2$, through conservatism and reversal aversion—that is the focus of this paper.

[12]Note that this assumption is more difficult to reconcile if the reputation for knowledgeability affects real outcomes in the contemporaneous period, as in Choi (2018).

2.3.2 Reputation Concerns

I now consider the case in which the central bank not only cares about aligning interest rates appropriately, but also about its reputation as being considered as knowledgable, that is $\Gamma > 0$ and the central bank's type is unobservable to the public. From earlier, we know that without reputational concerns, there is a unique pure strategy equilibrium in which each central bank type sets its interest in accordance of its best guess of the true state of the world each period. As will be shown, this pure strategy equilibrium generally fails to exist when reputational concerns are sufficiently high (Γ sufficiently high). The need to search for alternative equilibria demonstrate the effect that reputational considerations have on the actions of central banks.

2.3.2.1 Parameter Restrictions

Because of the large number of free parameters (λ, τ, $\widehat{\mu}_0$, Γ), it is difficult to analyze the model without more structure on the model's parameters. In conjunction with the Bayesian incomplete information nature of the game, a wide range of equilibria are possible; throughout the analysis, I often restrict the parameter space in order to isolate the main effects as clearly as possible. I also focus the search to equilibria in which there are minimal distortions (relative to the benchmark case with no reputational concerns) outside of reversal aversion and conservatism. In particular, I focus on equilibria in which the G type central bank (which has perfect information) always sets an interest rate in line with the true state of the world; this will always be the case if Γ is sufficiently low.[13] I thus analyze equilibrium actions by the B type central bank taken as given this assumption of the G type's behavior, and later check to make sure that this assumption holds in equilibrium (i.e., that the G type has no incentive to deviate from this truth-telling behavior). I also restrict myself to equilibria with minimal distortions in the B type's behavior. The possible subgames at $t = 2$ can be enumerated as follows:

Case 1: $(\theta_1, \theta_2) = (H, H)$, action in period 1 is i^H

Case 2: $(\theta_1, \theta_2) = (L, H)$, action in period 1 is i^H

[13] In particular, this is guaranteed with the assumption that $\Gamma < 1$, but this condition can be relaxed substantially for many parameter values.

Case 3: $(\theta_1, \theta_2) = (L, L)$, action in period 1 is i^H

Case 4: $(\theta_1, \theta_2) = (H, H)$, action in period 1 is i^L

Case 5: $(\theta_1, \theta_2) = (L, H)$, action in period 1 is i^L

Case 6: $(\theta_1, \theta_2) = (L, L)$, action in period 1 is i^L

Note that in period 2, the B type also has perfect information about the state. I look at equilibria in which the B type always sets an interest rate in line with the state unless it has set i^H in period 1, and the true state in period 2 is $\theta_2 = L$; this captures the idea of reversal aversion, and arises in Case 3 as enumerated above. In addition, in period $t = 1$, I assume that the B type central bank only considers not playing according to its signal if $s_1 = H$, playing $i_1 = i^H$ with probability σ_1 which may be less than one (playing $i_1 = i^L$ with probability 1 if it gets signal $s_1 = L$). As will be seen, searching for equilibria with no distortions in the other cases is not very restrictive, as it is difficult to generate equilibria in which these conditions do not hold. Nevertheless, for all numerical examples I make sure these conditions in periods hold in equilibrium. Given these restrictions, I now analyze the model, using backward induction and starting from period $t = 2$.

2.3.2.2 Period 2 Analysis: Reversal Aversion

In period 2, and taking the equilibrium actions in other cases into account (verifying later that they hold in equilibrium), I now analyze Case 3. Here, the B type central bank has played H in the first period, and is faced with a dilemma: it knows the true state of the world is $\theta_2 = L$, and would like to move its interest rate back to i^L, but doing so would reveal itself to be the B type with probability 1. In order to see if this reputation cost is enough to distort the central bank's actions, I first calculate $\widehat{\mu}_2 \left(i_1 = i^H, i_2 = i^H\right)$, the posterior belief that the public will have of the central bank after the observation of the sequence of actions i^H followed by i^H again; this is the reputation gain the central bank has by playing i^H instead of i^L in period 2 in Case 3, noting that $\widehat{\mu}_2 \left(i_1 = i^H, i_2 = i^H\right) = 0$. In order to calculate the posterior beliefs, I first assume that in Case 3 the B type central bank plays L with probability σ_2; given this equilibrium action, I enumerate all the possible state and signal combinations, and calculate the probability that this state occurs and that the B type indeed plays $i_1 = i^H, i_2 = i^H$. Letting $\Theta = (\theta_1, \theta_2)$ denote the sequence of true states, we have

1. $\Theta = (H, H)$: probability is $\underbrace{\lambda}_{\text{state prob.}} \underbrace{\tau\sigma_1}_{\text{period 1}}$
2. $\Theta = (L, H)$: probability is $\underbrace{(1-\lambda)\lambda}_{\text{state prob.}} \underbrace{(1-\tau)\sigma_1}_{\text{period 1}}$
3. $\Theta = (L, L)$: probability $\underbrace{(1-\lambda)^2}_{\text{state prob.}} \underbrace{(1-\tau)\sigma_1}_{\text{period 1}} \left[\underbrace{(1-\sigma_2)}_{\text{Period 2 Mix}} \right]$

Call the sum of these three terms $A(\sigma_1, \sigma_2)$. We then have

$$\widehat{\mu}_2 \left(i_1 = i^H, i_2 = i^H; \sigma_1, \sigma_2 \right) = \frac{\widehat{\mu}_0 \lambda}{\widehat{\mu}_0 \lambda + (1 - \widehat{\mu}_0) A(\sigma_1, \sigma_2)}$$

Now consider the pure strategy in which the B type plays its best assessment of the state, that is it plays L with probability 1 in Case 3, so that $\sigma_2 = 1$ (no distortion strategy). This pure strategy equilibrium exists if

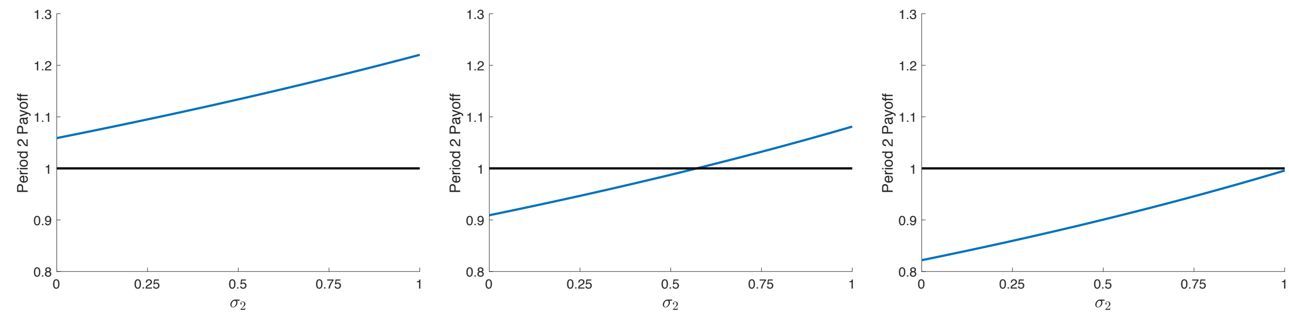
$$\underbrace{1}_{\text{Payoff to } i_2 = i^L} > \underbrace{\Gamma \left[\frac{\widehat{\mu}_0 \lambda}{\widehat{\mu}_0 \lambda + (1 - \widehat{\mu}_0) A(\sigma_1, \sigma_2 = 1)} \right]}_{\text{Reputation Gain from } i_2 = i^H} \tag{2.5}$$

Note that the term in brackets is always less than one; thus, if we make the assumption that $\Gamma < 1$, this condition will always hold. However, if Γ is sufficiently high, this equilibrium in general *fails* to exist. Thus, it must be the case that Γ is sufficiently higher than 1.[14] It may even be the case that a pure strategy equilibrium exists in which there is complete reversal aversion, that is the B type never moves back down to i^L after moving to i^H at period 1 ($\sigma_2 = 0$). This will be the case if

$$1 < \Gamma \left[\frac{\widetilde{\mu}_0 \lambda}{\widetilde{\mu}_0 \lambda + (1 - \widetilde{\mu}_0) A(\sigma_1, \sigma_2 = 0)} \right]$$

where I maintain the assumption that $\widehat{\mu}_2 \left(i_1 = i^H,\ i_2 = i^L; \sigma_1, \sigma_2 = 0 \right) = 0$, that is the public always puts zero weight on the central bank being of type G after the observation of $i_1 = i^H, i_2 = i^L$; while neither the G nor the B type ever play these pairs of actions when $\sigma_2 = 0$, one must still specify

[14]As discussed later in more detail, while Γ must be sufficiently higher than 1 to generate interesting reputational equilibria, if Γ is too high there may be strange behavior arises due to the severe distortions from reputational considerations
.



(a) Pure strategy $\sigma_2 = 0$ equilibrium **(b)** Mixed strategy equilibrium **(c)** Pure strategy $\sigma_2 = 1$ equilibrium

Figure 2.2: Graphs of payoffs from playing i^H and i^L in Case 3 when the probability of playing i^L is σ_1. Figure (a) illustrates the case where the pure strategy $\sigma_2 = 0$ is the unique equilibrium strategy; here, even though σ_2 is low (making i^H relatively less attractive), playing i^H strictly dominates playing i^L, which yields a sure payoff of 1. Figure (b) illustrates the case where a mixed strategy $0 < \sigma_2 < 1$ is the unique equilibrium, given by the intersection of the two lines. Given σ_2, the central bank is indifferent between playing i^H and i^L. Figure (c) illustrates the case where $\sigma_1 = 1$ is the unique equilibrium strategy; even though σ_2 is high (making i^H relatively more attractive), playing i^L strictly dominates playing i^H. Shown using the baseline parameterization of the model, with $\lambda = 0.25$, $\tau = 0.8$, $\Gamma = 1.75$, $\widehat{\mu}_0 = 0.5$. Figures (a), (b), and (c) show the cases of $\sigma_1 = 0.5$, $\sigma_1 = 0.7$, and $\sigma_1 = 0.85$, respectively.

beliefs on off-the-equilibrium paths, and so I maintain this assumption. I discuss off-equilibrium beliefs in more detail later. Note that a mixed strategy in Case 3 is also possible, in which case the payoff from playing H and L must be the same, given mixing probability σ_2. Note crucially that the actions taken in period 1 affect the optimal action in Case 3, through the effect σ_1 has on $A(\sigma_1, \sigma_2)$. This interplay between how actions in period 1 affect period 2 is discussed in more detail later.

One issue that arises is the potential for multiple equilibria in Case 3. Given σ_1, however, $A(\sigma_1, \sigma_2)$ is a decreasing function of σ_2 and thus $\widehat{\mu}_2\left(i_1 = i^H, i_2 = i^H; \sigma_1, \sigma_2\right)$ is an increasing function of σ_2. Because the equilibrium is determined where $\Gamma\widehat{\mu}_2\left(i_1 = i^H, i_2 = i^H; \sigma_1, \sigma_2\right)$ is relative to 1, the equilibrium given σ_1 is unique. Using the baseline parameterization of the model, Figures 2.2a and 2.2c show cases in which the pure strategies of $\sigma_2 = 1$ and $\sigma_2 = 0$ are the unique equilibrium given σ_1, with Figure 2.2b showing the case in which a unique mixed strategy equilibrium obtains. Each figure shows cases of increasingly larger σ_1, showing that as σ_1 increases, the incentive to play i^H in Case 3 increases. While it must be the case that given σ_1, the equilibrium action in period 2 is unique, as we shall see later the complementarity between σ_1 and σ_2 does potentially lead to multiple equilibria of the broader game, with either low amounts of conservatism and reversal aversion, or high amounts of each.

While it is assumed that in the other possible cases outside of Case 3, the central banks simply play the interest rate that matches up with the state in period $t = 2$, in order to check whether this holds we must also derive the expressions for $\widehat{\mu}_2\left(i_1 = i^L,\ i_2 = i^H\right)$ and $\widehat{\mu}_2\left(i_1 = i^L,\ i_2 = i^L\right)$. I again enumerate all the possible state and signal combinations, and calculate the probability that this state occurs and that the B type indeed plays $i_1 = i^L,\ i_2 = i^H$ or .$i_1 = i^L,\ i_2 = i^L$ We have

1. $\Theta = (H, H)$: probability of $i_1 = i^L,\ i_2 = i^H$ is $\underbrace{\lambda}_{\text{state prob.}} \left[\underbrace{(1-\tau)}_{\text{L signal period 1}} + \underbrace{\tau(1-\sigma_1)}_{\text{H signal period 1}} \right]$,
   of $i_1 = i^L,\ i_2 = i^L$ is 0
2. $\Theta = (L, H)$: probability $i_1 = i^L,\ i_2 = i^H$ is $\underbrace{(1-\lambda)\lambda}_{\text{state prob.}} \left[\underbrace{\tau}_{\text{L signal period 1}} + \underbrace{(1-\tau)(1-\sigma_1)}_{\text{H signal period 1}} \right]$,
   of $i_1 = i^L,\ i_2 = i^L$ is 0
3. $\Theta = (L, L)$: probability of $i_1 = i^L,\ i_2 = i^H$ is 0,
   of $i_1 = i^L,\ i_2 = i^L$ is $\underbrace{(1-\lambda)^2}_{\text{state prob.}} \left[\underbrace{\tau}_{\text{L signal period 1}} + \underbrace{(1-\tau)(1-\sigma_1)}_{\text{H signal period 1}} \right]$

Calling the sum of these terms C for the case of $i_1 = i^L,\ i_2 = i^H$ and the sum of these terms D for the case of $i_1 = i^L,\ i_2 = i^L$, we have

$$\widehat{\mu}_2\left(i_1 = i^L, i_2 = i^H; \sigma_1\right) = \frac{\widehat{\mu}_0(1-\lambda)\lambda}{\widehat{\mu}_0(1-\lambda)\lambda + (1-\widehat{\mu}_0)\,C}$$

$$\widehat{\mu}_2\left(i_1 = i^L, i_2 = i^L; \sigma_1\right) = \frac{\widehat{\mu}_0(1-\lambda)^2}{\widehat{\mu}_0(1-\lambda)^2 + (1-\widehat{\mu}_0)\,D}$$

noting that these terms are independent of σ_2 .

2.3.2.3 Period 1 Analysis: Conservatism

To analyze period 1, I consider the case when the B type gets a signal of $s_1 = i^H$, and considers whether or not to follow its signal. The posterior probabilities that the B type gives to each state are given by (2.1) and (2.2). In order to complete the expected period 1 payoffs of playing H and L, I now derive $\widehat{\mu}_1\left(i_1 = i^H\right)$ and $\widehat{\mu}_1\left(i_1 = i^L\right)$, given that the G type always plays in accordance to the true state, the B type plays L with probability 1 when it receives a signal of $s_1 = L$, and plays i^H

with probability σ_1 when it receives a signal of $s_1 = H$. We then have

$$\widehat{\mu}_1\left(i_1 = i^H; \sigma_1\right) = \frac{\widehat{\mu}_0 \lambda}{\widehat{\mu}_0 \lambda + (1-\widehat{\mu}_0)\,\sigma_1\left[\lambda\tau + (1-\lambda)(1-\tau)\right]} \tag{2.6}$$

$$\widehat{\mu}_1\left(i_1 = i^L; \sigma_1\right) = \frac{\widehat{\mu}_0(1-\lambda)}{\widehat{\mu}_0(1-\lambda) + (1-\widehat{\mu}_0)\left[(1-\lambda)(\tau + (1-\tau)(1-\sigma_1)) + \lambda\left((1-\tau) + \tau(1-\sigma_1)\right)\right]} \tag{2.7}$$

Note that $\widehat{\mu}_1\left(i_1 = i^H; \sigma_1\right)$ and $\widehat{\mu}_1\left(i_1 = i^L; \sigma_1\right)$ are, respectively, increasing and decreasing functions of σ_1. The intuition is clear: as σ_1 increases, the B type is more likely to play i^H, and so the posterior probability that the central bank is of type G when i^H is observed decreases relative to if σ_1 is smaller (and vice-versa for when i^L is observed). Note that when $\sigma_1 = 1$ the expressions above collapse down to (2.3) and (2.4). As seen from these expressions, one complication that arises in the analysis of conservatism is that even when there are no distortions in the actions of the B type in period $t = 1$, the posterior probabilities that the public puts on the type of the central bank are affected by the action of the central bank. Thus, it becomes difficult to disentangle whether distortions in actions in period 1 arise from true conservatism (foreseeing the possibility of reversal aversion) or the mechanical effect of avoiding an action due to its partial revelation of the central bank's type. This effect goes away with the parameter restriction $\lambda = 1/2$, in which case

$$\widehat{\mu}_1\left(i_1 = i^H; \sigma_1 = 1\right) = \widehat{\mu}_0$$

$$\widehat{\mu}_1\left(i_1 = i^L; \sigma_1 = 1\right) = \widehat{\mu}_0$$

and so given a candidate equilibrium with no distortions, there would be no revelation of the central bank's type from the action in the period 1. Despite this, because the actions in period $t = 1$ also affect the reputation posteriors $\widehat{\mu}_2$ of the next period, even the parameter restriction of $\lambda = 1/2$ does not completely shut off the reputational incentives that would still arise assuming a candidate equilibrium in which there are no distortions.

When deciding on its optimal action in period 1 after receiving signal $s_1 = i^H$, the B type central bank takes into account how its actions affect its current payoffs today—both from playing the 'optimal' action as well as from reputational considerations in the current period given by (2.6) and (2.7)—as well as the payoffs from the next period. As such, the central bank must take into account the probabilities of being in each possible sequence of states of the world. Denoting by $\widetilde{P}\left(\cdot \mid s_1\right)$ the subjective probability that the B type puts on the true state of the world given a signal s_1 in period 1, we have

$$\widetilde{P}(\theta_1 = H,\ \theta_2 = H \mid s_1) = \widetilde{P}(\theta_1 = H \mid s_1) \tag{2.8}$$

$$\widetilde{P}(\theta_1 = L,\ \theta_2 = H \mid s_1) = \lambda \widetilde{P}(\theta_1 = L \mid s_1) \tag{2.9}$$

$$\widetilde{P}(\theta_1 = L,\ \theta_2 = L \mid s_1) = (1-\lambda)\widetilde{P}(\theta_1 = L \mid s_1) \tag{2.10}$$

where the probabilities on the right hand side are given by (2.1) and (2.2) and so the probabilities have an easily defined recursive structure.

The first natural question to ask is whether or not a pure strategy of no distortion in period 1 ($\sigma_1 = 1$) exists. In order for this to be the case, it must be that given the public assumes $\sigma_1 = 1$, and given the optimal reaction function of the B type in period 2 (in particular, in case 3), there is no incentive for the B type to deviate from the pure strategy of playing i^H when given signal s_1^H. To see whether or not this holds, the expected payoff from playing i^H is

$$\begin{aligned}
&\widetilde{P}(\theta_1 = H,\ \theta_2 = H \mid s_1 = H)\left[2 + \Gamma\Big(\widehat{\mu}_1\left(i^H;\sigma_1 = 1\right) + \widehat{\mu}_2\left(i^H, i^H;\sigma_1 = 1, \sigma_2^*\right)\Big)\right] \\
&+\widetilde{P}(\theta_1 = L,\ \theta_2 = H \mid s_1 = H)\left[1 + \Gamma\Big(\widehat{\mu}_1\left(i^H;\sigma_1 = 1\right) + \widehat{\mu}_2\left(i^H, i^H;\sigma_1 = 1, \sigma_2^*\right)\Big)\right] \\
&+\widetilde{P}(\theta_1 = L,\ \theta_2 = L \mid s_1 = H)\left[\Gamma\widehat{\mu}_1\left(i^H;\sigma_1 = 1\right) + \max\Big\{1,\ \Gamma\widehat{\mu}_2\left(i^H, i^H;\sigma_1 = 1, \sigma_2^*\right)\Big\}\right]
\end{aligned} \tag{2.11}$$

where $\sigma_2^* \equiv \sigma_2(\sigma_1 = 1)$ denotes the unique period $t = 2$ action (in case 3) that obtains when the B type central bank plays the pure strategy $\sigma_2 = 1$ in period $t = 1$. The payoff from deviating and playing i^L after receiving signal s_1^H is

$$\begin{aligned}
&\widetilde{P}(\theta_1 = H,\ \theta_2 = H \mid s_1 = H)\left[1 + \Gamma\Big(\widehat{\mu}_1\left(i^L;\sigma_1 = 1\right) + \widehat{\mu}_2\left(i^L, i^H;\sigma_1 = 1, \sigma_2^*\right)\Big)\right] \\
&+\widetilde{P}(\theta_1 = L,\ \theta_2 = H \mid s_1 = H)\left[2 + \Gamma\Big(\widehat{\mu}_1\left(i^L;\sigma_1 = 1\right) + \widehat{\mu}_2\left(i^L, i^H;\sigma_1 = 1, \sigma_2^*\right)\Big)\right] \\
&+\widetilde{P}(\theta_1 = L,\ \theta_2 = L \mid s_1 = H)\left[2 + \Gamma\Big(\widehat{\mu}_1\left(i^L;\sigma_1 = 1\right) + \widehat{\mu}_2\left(i^L, i^L;\sigma_1 = 1, \sigma_2^*\right)\Big)\right]
\end{aligned} \tag{2.12}$$

Given these expressions, the following Lemma is immediate:

Lemma 2. *If Γ is sufficiently small, there is a unique pure strategy equilibrium in which the B type central bank plays the no distortion actions $\sigma_1 = 1$ and $\sigma_2 = 1$.*

Proof. For any given σ_1, for sufficiently small Γ we have that $1 > \Gamma \widehat{\mu}_2 \left(i^H, i^H; \sigma_1 = 1, \sigma_2 = 1\right)$ and so the pure strategy $\sigma_2 = 1$ is indeed the optimal action when the sequence of states is $\theta_1 = L$, $\theta_2 = L$ and the B type central bank plays i^H in period 1. To compare the payoffs from playing i^H and i^L in period $t = 1$ after receiving the signal $s_1 = H$, we can take the total expected payoff (2.11) from playing i^H and subtract the total expected payoff (2.12) from playing i^L. We have that the difference in expected payoff excluding reputation-related payoff as

$$\widetilde{P}\left(\theta_1 = H,\ \theta_2 = H \mid s_1 = H\right) - \widetilde{P}\left(\theta_1 = L,\ \theta_2 = H \mid s_1 = H\right) - \widetilde{P}\left(\theta_1 = L,\ \theta_2 = L \mid s_1 = H\right)$$

$$= \widetilde{P}\left(\theta_1 = H \mid s_1 = H\right) - \widetilde{P}\left(\theta_1 = L \mid s_1\right) > 0$$

using (2.1) and (2.2) as well as (2.8)-(2.10). We can group reputation-related payoff terms, which enter linearly in terms of Γ. For *any* possible set of posterior beliefs held by the public $\widehat{\mu}_1(\cdot)$ and $\widehat{\mu}_2(\cdot)$, as Γ decreases, the reputational terms eventually become less than $\widetilde{P}\left(\theta_1 = H \mid s_1 = H\right) - \widetilde{P}\left(\theta_1 = L \mid s_1\right)$, in which case the optimal action in period 1 after receiving H is to play $i_1 = i^H$. □

This is a very unsurprising result, which simply states that for small reputational effects, there is a unique equilibrium that converges to the unique equilibrium in the benchmark case when reputational effects are not present. Generally, however, there may be distortions, as we have already seen in the case of period $t = 2$. The optimal strategy in period $t = 1$ depends on all the terms given by (2.11) and (2.12). A brief discussion of each of these terms will provide intuition for when a pure strategy of $\sigma_1 = 1$ is *not* an equilibrium.

To begin with, it is instructive to analyze the expected payoffs (2.11) and (2.12) assuming that all of the non-zero $\widehat{\mu}$ terms are similar, later adding in variations to the $\widehat{\mu}$ terms to analyze how incentives would change. This assumption holds when actions reveal little information about the type, unless i^H is followed by i^L in which case the public puts zero weight on the central bank being of type G. For simplicity, suppose for now that all of the $\widehat{\mu}$ terms are equal to μ^*; we then have that the expected gain from playing i^H after receiving signal $s_1 = H$ is

$$\widetilde{P}\left(H,\ H \mid s_1 = H\right) - \widetilde{P}\left(L, H \mid s_1 = H\right) - \widetilde{P}\left(L, L \mid s_1 = H\right)\left[2 + \Gamma\mu^* - \max\left\{1, \Gamma\mu^*\right\}\right]$$

$$= \underbrace{\widetilde{P}\left(H,\ H \mid s_1 = H\right) - \widetilde{P}\left(L, H \mid s_1 = H\right) - \widetilde{P}\left(L, L \mid s_1 = H\right)}_{\text{Expected P1 payoff gain}} \tag{2.13}$$

$$- \underbrace{\widetilde{P}\left(L, L \mid s_1 = H\right)\left[1 + \Gamma\mu^* - \max\left\{1, \Gamma\mu^*\right\}\right]}_{\text{Expected P2 payoff loss}}$$

where $\widetilde{P}\left(H,\ H \mid s_1 = H\right) = \widetilde{P}\left(\theta_1 = H,\ \theta_2 = H \mid s_1 = H\right)$ and similarly for the other probabilities. As shown, the difference from playing i^H over i^L can be decomposed into the expected period $t = 1$ gain and the period $t = 2$ loss. Intuitively, because the actions in all cases other than Case 3 are the same—that is, distortions in actions only arise in Case 3—and we assume for now that the public's posterior of the central bank's type remains the same unless the fully-revealing actions of $i_1 = i^H$ and $i_2 = i^L$ are played, there are only relative gains or losses at $t = 2$ when Case 3 occurs. As shown by Lemma 2, the expected P1 payoff gain is positive, as given the central bank's signal the H state is more likely than the L state at $t = 1$. However, note that under these assumptions the difference in expected payoffs in period $t = 2$ from playing $i_1 = i^H$ is *negative*, as the only differences in actions come when the true sequence of states is revealed to be $\theta_1 = L$, $\theta_2 = L$. In this situation, a central bank that has played $i_1 = i^H$ faces a decision: it can either go back to i^L, and reveal itself to be the B type with probability 1, or it can stick with i^H and thus set an interest rate suboptimal for the state. Meanwhile, a central bank having played $i_1 = i^L$ can simply set $i_2 = i^L$ and set an appropriate interest while maintaining its reputation. Note that the expected loss at $t = 2$ can be substantial, and is increasing in $\widetilde{P}\left(L, L \mid s_1 = H\right)$, the probability that the central bank finds itself in Case 3, and in Γ, the reputational cost $\Gamma\mu^*$; this reputational cost reaches a maximum of 1, after which the central bank protects its reputation and takes a loss of 1 (the mismatch of interest rate) instead.

Of course, in any actual equilibrium it will not be the case that all positive $\widehat{\mu}$ terms are exactly the same. Under the pure strategy $\sigma_1 = 1$, we have that the period 1 posteriors about the central bank's type are given by (2.3) and (2.4), and it straightforward to show that $\widehat{\mu}_1\left(i_1 = H\right) \leq \widehat{\mu}_0$ and $\widehat{\mu}_1\left(i_1 = L\right) \geq \widehat{\mu}_0$. Intuitively, when a change of state is less likely than not—given by assumption $\lambda \leq 1/2$—the B type is relatively more likely to move to i^H due to the possibility of getting a misinformed signal of $s_1 = H$ when the true state is $\theta_1 = L$.[15] This effect provides additional incentive for the B type to be conservative in the first period, in addition to the conservatism coming reversal aversion. These reinforcing effects in conjunction can lead to substantial conservatism

[15]Note that an error in the other direction is possible, that is the B type gets a signal of $s_1 = L$ when the true state is $\theta_1 = H$. However, because the conditional probability of getting an error is symmetric, and the state $\theta_1 = L$ is (weakly) more likely than $\theta_1 = H$, it it follows that an error of $s_1 = H$ when $\theta_1 = L$ is more likely.

(inertia) in actions.

In addition to these effects, we must take into account how equilibrium actions in period $t = 1$ affect the equilibrium actions in $t = 2$. Recall from earlier that $\widehat{\mu}_1\left(i_1 = i^H; \sigma_1\right)$ and $\widehat{\mu}_1\left(i_1 = i^L; \sigma_1\right)$ are, respectively, increasing and decreasing functions of σ_1; as σ_1 decreases, playing i^H becomes more attractive, muting the desire to play i^L in the first place. This effect is similar to that seen in Prendergast and Stole (1996): playing an 'aggressive' action shows confidence in taking an action that a conservative central bank with imperfect knowledge may not take, due to the potential of large costs in future periods if the assessment of the state proves to be incorrect. Note, however, that in contrast to that model, these effects are second order and arise only because of conservatism on the part of the B type in the first place; there, conservatism from dynamic reputational considerations do not arise, as the agent is completely myopic when making decisions.

We thus have a broad outline of the myriad new forces that arise when central banks care about their reputation for knowledgeability, relative to the benchmark model with no reputational concerns. Without reputational concerns, there is a tendency of uninformed central banks to start hiking (cutting) cycles prematurely,[16]and so with reputational effects conservatism can arise from both the fear of being seen as unknowledgeable *prima facie* (by starting a new cycle) or by revelation if a need to reverse a prior decision arises later. In this model, either of these forces can lead to conservatism, with stronger conservatism when both effects are present. Such conservatism leads to potential distortions in period $t = 1$ in which σ_1 is less than one; as the forces of conservatism get stronger (σ_1 decreases), the association between playing i^H and being the of type B weakens, which then leads to i^H becoming relatively more attractive, in addition to having the highest expected payoff when reputation for knowledgeability is not a concern. A mixed strategy equilibrium arises when the multiple forces on each side balance out.

One potentially surprising result concerns whether there can be a pure strategy equilibrium

[16]Note that this is counterfactual to what is generally observed in monetary policy, in which central banks tend to react slowly and fall behind the curve. Thus, this model provides a potential reputation-based explanation for why this might be the case. A particularly salient example of this may be the recent deliberations by central banks of when to begin liftoff from a long period of low interest rates near the effective lower bound. This debate remains contentious with a lot of uncertainty on the optimal timing. The European Central Bank began raising interest rates fairly early, and potentially suffered a hit to its reputation for competence almost immediately after making this decision. It was forced to decrease interest rates shortly thereafter, and the whole ordeal dented much of the credibility of the central bank. The Federal Reserve Bank of the United States also had many meetings where it considered hiking rates. As the quote by the president of the Federal Reserve Bank of Chicago in footnote 2 of the first essay shows, the desire to not reverse rates, due to reputational considerations, may have played a considerable role in the decision-making process.

in which $\sigma_1 = 0$, that is there is complete conservatism in that the B type central bank never sets $i_1 = i^H$. At first glance, the answer may appear to be no: the central bank, after receiving signal $i_1 = i^H$, can get a higher expected payoff in terms of setting the appropriate interest rate at $t = 1$, and also be revealed to be type i^H with probability 1. However, assuming that $i_1 = i^H$ and $i_2 = i^L$ still reveals the central bank to be of type B with probability 1 (or a sufficiently high probability),[17] $\sigma_1 = 0$ may in fact be a pure strategy equilibrium action. If it is the case that the prior $\widehat{\mu}_0$ is already very high, then the reputational gains from playing $i_1 = i^H$ will be minimal, with the same tradeoff in terms of expected gains and losses illustrated in (2.13) still present; note that if the prior $\widehat{\mu}_0$ is indeed very high, actions other than the pair $i_1 = i^H$, $i_2 = i^L$ reveal very little about the type (due to the anchoring effect of strong priors), and so the implicit assumptions underlying (2.13) will approximately hold. In addition, in this situation the pair of actions $i_1 = i^H$, $i_2 = i^L$ may be potentially very costly, due to the a large drop in the public's posterior probabilities in the central bank being of type G. Thus, the loss if the central bank finding itself in states $\theta_1 = L$, $\theta_2 = L$ can be very large, up to a maximum of 1.

We have thus seen the main forces underlying the central bank's equilibrium actions in period 1. While a more formal analysis outlining all possibilities (including the possibility of multiple equilibria) and specific conditions under which each type of equilibrium arises is outside the scope of this paper, I provide a more heuristic discussion in the following section, which provides in more detail when certain types of equilibria might arise.

2.3.2.4 Checking Imposed Equilibrium Conditions

The previous section restricted the equilibrium search to 'reasonable' equilibria in which there are minimal distortions in actions, in order to focus on reversal aversion and conservatism. This first

[17]Note that if $\sigma_1 = 0$, $i_1 = i^H$, $i_2 = i^L$ never arises in equilibrium. Nevertheless, one must specify the posterior probabilities on off-the-equilibrium paths in order to check for an equilibrium. The assumption of $i_1 = i^H$ and $i_2 = i^L$ leading to $\widehat{\mu}_2 = 0$ is a reasonable assumption if we assume that the B type may instead play $i_1 = i^H$ with very small probability $\varepsilon > 0$, as in the the trembling hand perfect equilibrium refinement concept of Selten (1975). We could also assume that the G type also trembles with symmetric probability ε, in which case the posterior probabilities could be such that $\widehat{\mu}_2 \left(i_1 = i^H, i_2 = i^L\right) > 0$. However, for the rest of the paper I maintain the assumption that the G type never trembles; in essence, this paper takes as given the G type's behavior as fixed (noting that it never deviates from the benchmark case actions in equilibrium), and analyzes actions from the perspective of the B type, who may make a mistake with infinitesimal probability ε. A natural interpretation of this model is one of an imperfect central bank attempting to mimic a perfect central bank, that knows the state perfectly and takes actions accordingly without error. Making this assumption also leads to continuity in posterior beliefs as a function of σ_2.

requires that in the period 2, in all cases other than Case 3, the B type central bank sets an interest in accordance to the true state of the world, which is by then revealed perfectly. In order for this to hold, it must be the case that the differences in reputation from taking an alternative action is not sufficient to overturn this desire, which will hold as long Γ is not too high. In addition, when receiving a signal $s_1 = i^L$, it must be the case that the B type only plays $i_1 = i^L$; this is an extremely weak assumption, as playing i^H in such a situation yields a lower expected (non-reputational) payoff in terms period $t = 1$, as well as generally lower expected payoff in period $t = 2$, as seen in (2.13). In addition, in most cases playing $i_1 = i^H$ is associated with being the *bad* type, and so this further disincentives playing $i_1 = i^H$ when receiving signal $s_1 = L$. [18] As I show, in this model reversal aversion and conservatism are first order distortions, in that as Γ increases, distortions tend to first arise in Case 3 and when receiving signal $s_1 = H$.

For completeness, we must have that

$$1 + \Gamma \widehat{\mu}_2 \left(i^H, i^H; \sigma_1^*, \sigma_2^* \right) \geq 0 \tag{2.14}$$

$$1 + \Gamma \widehat{\mu}_2 \left(i^L, i^H; \sigma_1^*, \sigma_2^* \right) \geq \Gamma \widehat{\mu}_2 \left(i^L, i^L; \sigma_1^*, \sigma_2^* \right) \tag{2.15}$$

$$1 + \Gamma \widehat{\mu}_2 \left(i^L, i^L; \sigma_1^*, \sigma_2^* \right) \geq \Gamma \widehat{\mu}_2 \left(i^L, i^H; \sigma_1^*, \sigma_2^* \right) \tag{2.16}$$

On the LHS are the payoffs from playing an interest rate in accordance with the true state, with the RHS representing the payoffs from deviating to the other interest rate. Using the enumeration outlined before, (2.14) corresponds to Case 1 and Case 2, (2.15) corresponds to Case 4 and 5, and (2.16) corresponds to Case 6. Note again that these conditions are likely to hold as long as Γ is not very large or the difference in nonzero $\widehat{\mu}_2$ terms are not very large, and with (2.14) holding for sure.

The period $t = 1$ condition that is needed is that the expected payoff from playing $i_1 = i^L$ is greater than the expected payoff from playing $i_1 = i^H$ when receiving signal $s_1 = i^L$. The expected payoff to playing i^H and i^L can be seen by (2.11) and (2.12), respectively, replacing $s_1 = H$ with

[18] Intuitively, playing $i_1 = i^H$ in this scenario can only hold if σ_1 is extremely close to 0, that is the B type rarely plays i^H when receiving a signal of $s_1 = H$. In such a case, playing $i_1 = i^H$ may be attractive as it is no longer associated with being the B type. Of course, if this is the case, then there is an equal reputational incentive for the central bank to play $i_1 = i^H$ in the case of receiving a signal $s_1 = H$ (in additional to the gain in terms of expected non-reputational payoffs, which is higher than when receiving a signal of $s_1 = L$), in which case σ_1 can no longer be small in equilibrium. Given this, it is unlikely that σ_1 will be close to zero but the central bank does not play i^L with probability 1 when receiving a signal of $i_1 = i^L$. While I do not provide a formal analysis, It is extremely unlikely that this assumption is violated for any reasonable parameter values.

$s_1 = L$. Heuristically, to show this is likely to hold, consider the equivalent of (2.13) which assumes that all non-zero $\widehat{\mu}$ terms are equal to μ^*. The needed condition is that

$$\widetilde{P}(L, H \mid s_1 = L) + \widetilde{P}(L, L \mid s_1 = L) - \widetilde{P}(H,\ H \mid s_1 = L) \tag{2.17}$$
$$+\widetilde{P}(L, L \mid s_1 = L)\left[1 + \Gamma\mu^* - \max\{1, \Gamma\mu^*\}\right] \geq 0$$

which holds as $\widetilde{P}(L, H \mid s_1 = L) + \widetilde{P}(L, L \mid s_1 = L) > \widetilde{P}(H,\ H \mid s_1 = L)$ by (2.2) and (2.8) - (2.10).

The second set of equilibrium conditions that are needed is that the G type always plays an action according to the true state of the world (which it knows perfectly in each time period). Note that in period $t = 2$, the G and B type are identical, as they then have perfect knowledge about the state; thus, the conditions for the G type are identical to that of the B type given in (2.14) - (2.16), noting that in equilibrium Case 3 and Case 6 never arise for the G type if it sets the interest rate in line with the true state at $t = 1$. Furthermore, if the condition in period $t = 1$ for the B type when $s_1 = i^L$ holds, it holds for the G type as well, given that it knows the state perfectly in time period $t = 1$ and thus the assumption is strictly weaker. To see this heuristically, the condition (2.17) would be

$$1 + (1 - \lambda)\left[1 + \Gamma\mu^* - \max\{1, \Gamma\mu^*\}\right] \geq 0 \tag{2.18}$$

where the LHS of (2.18) is bigger than the LHS of (2.17). One last condition to check is that the G type plays $i_1 = i^H$ when the state is revealed to be H at $t = 1$; because the state is guaranteed to be H at $t = 2$, this condition is

$$2 + \Gamma\left[\widehat{\mu}_1\left(i^H; \sigma_1^*, \sigma_2^*\right) + \Gamma\widehat{\mu}_2\left(i^H, i^H; \sigma_1^*, \sigma_2^*\right)\right]$$
$$\geq \Gamma\widehat{\mu}_1\left(i^L; \sigma_1^*, \sigma_2^*\right) + \max\left\{1 + \Gamma\widehat{\mu}_2\left(i^L, i^H; \sigma_1^*, \sigma_2^*\right), \widehat{\mu}_2\left(i^L, i^L; \sigma_1^*, \sigma_2^*\right)\right\}$$

which again holds assuming that the difference in nonzero $\widehat{\mu}_2$ terms are not very large.[19]

Naturally, all of these conditions above hold for small Γ, but Γ has to be sufficiently large for any deviations in actions from the benchmark case to arise. Note, however, that Γ must be quite large (or the parameters of the model such that the structure of the model is a very particular way)

[19]In particular, playing i^H in period $t = 1$ cannot be too heavily associated with being the B type if reputational costs are large. Again, however, if this is the case there will be less incentive for the B type itself to play i^H in period 1, mitigating the association of $i_1 = i^H$ with the B type in the first place. Thus this assumption is quite weak.

for the relatively weak conditions above to be violated; on the other hand, the forces of reversal aversion and conservatism are likely to arise even when Γ is not particularly large. Thus, reversal aversion and conservatism are likely to be some of the first distortions that arise when a central bank cares about its reputation for knowledgeability within this model.

2.3.3 Discussion of Results

The above analysis provides conditions in which reversal aversion and conservatism are likely to arise. I now analyze these conditions in more detail, showing under which parameter configurations these conditions will hold, and later provide some numerical examples for illustration. It was already shown that Γ has to be sufficiently high in order for distortions to arise from reversal aversion and conservatism, as otherwise reputational considerations have no bite. The interplay of parameters λ, τ, and $\widehat{\mu}_0$ also affect the likelihood that reversal and conservatism will arise, and I now provide an explanation for each.

I first begin by looking at $t = 1$ when the B type central bank receives a signal $s_1 = H$. The key equations are once again (2.11) and (2.12). Playing i^H in period 1 leads to the potential of cases 1, 2, and 3, whereas playing i^L leads to the potential of cases 4, 5, and 6, where the probabilities of each of these cases are given by $\widetilde{P}\left(\theta_1 = H,\ \theta_2 = H \mid s_1 = H\right)$, $\widetilde{P}\left(\theta_1 = L,\ \theta_2 = H \mid s_1 = H\right)$, and $\widetilde{P}\left(\theta_1 = L,\ \theta_2 = L \mid s_1 = H\right)$, respectively. There are two components in each period to the payoffs of the central bank: the payoff from getting the 'correct' interest, and the reputation that obtains in each period, for a total of four components with two periods. The imposed equilibrium conditions ensure that because both central bank types play according to the true state of the world in all cases other than Case 3, the direct payoff from setting the correct interest rate is the same in Cases 1 and 4, as well as in Cases 2 and 5. In addition, if the difference in payoffs from reputation is similar between all non-zero $\widehat{\mu}_2$, the expected payoffs in these pairs of cases will be very similar. This is heuristically shown by (2.13), where the expected difference in payoffs are then simply given by the probabilities of each of these possible cases. In the cases where the B type central bank 'guesses' correctly and plays the action corresponding to the true state of the world—Cases 1, 5, and 6—it receives near the maximum amount of total payoff possible, by maxing out the fundamental component and not revealing itself to be the B type. In the cases where the central bank takes the wrong action in period 1—Cases 2, 3, and 4—damage is limited in Cases 2 and 4 because the central bank can credibly move to the correct action in the next period, receiving a payoff of 1, while still maintaining its reputation. However, Case 3, where i^H is played but the true sequence of states is revealed to have stayed at L in both periods, is a particularly bad situation for the central bank to find itself in; it will have to decide whether to forgo the payoff of 1 and maintain its reputation, or move back to i^L, earning a payoff of 1 but revealing itself as the B type with probability 1. It cannot have both, as in Cases 2 and 4, in which the damage is contained within the fundamental mismatch cost at $t = 1$. Hence, playing $i_1 = i^L$ can be an optimal strategy even if the probability that $\theta_1 = L$ is less than $1/2$, because of the ability to hedge against the possibility of Case 3 arising, which has the *lowest* possible payoff. Heuristically, looking at (2.13), if the non-zero $\widehat{\mu}$ terms are the same, there is a difference of 1 in the corresponding pairs of Case 1 and Case 4, and of Case 2 and 5, but a difference of 2 between Case 3 and Case 6. It is this desire to avoid Case 3 that drives the conservatism from reversal aversion in this paper.

As derived earlier, this conservatism is more likely when $\widetilde{P}\left(\theta_1 = L, \theta_2 = L \mid s_1 = H\right)$ is relatively high; in such a case Case 3 is relatively more likely when i^H is played at $t = 1$. When is this likely to be the case? Looking at (2.1) and (2.10), note that (fixing other parameters), when τ gets closer to the imposed lower bound of $1 - \lambda$, the probability of $\widetilde{P}\left(\theta_1 = L \mid s_1 = H\right) = 1 - \widetilde{P}\left(\theta_1 = H \mid s_1 = H\right)$ goes down. Intuitively, as the precision of the signal decreases, the central bank becomes less sure that the true state is indeed $\theta_1 = H$, reducing the expected gain of playing $i_1 = i^H$ and increasing the chances of finding itself in Case 3 with the lowest possible payoff. As τ gets closer to 1, the probability of an error goes to 0, and the B type's actions converge towards that of the G type, which always plays the true state in equilibrium. Similarly, a decrease in λ has a large effect on conservatism, for multiple reasons. Again, a lower λ today increases $\widetilde{P}\left(\theta_1 = L \mid s_1 = H\right)$; intuitively, $(1 - \lambda)$ is the prior probability the central bank puts on the state staying at $\theta_1 = L$ (before its signal is received), and as λ decreases, the more likely the B type has received an incorrect signal which it thus down-weights. Secondly, if the true state at $t = 1$ is $\theta_1 = L$, when λ is low it becomes relatively more likely that $\theta_2 = L$ is also low, as given by (2.10). Thus, λ increases the probability that Case 3 arises when $i_1 = i^H$ in reinforcing ways. Lastly, when λ is relatively low, moving to i^H is tends to be associated with the B type,[20] the type more likely receive a false signal and thus react incorrectly. This direct reputational effect is yet another reason why low values of λ are likely to lead to more conservatism.

Lastly, the prior probability that the public attaches to the central bank's type, given by $\widehat{\mu}_0$, can have a large effect on the equilibrium actions of the model. Again, when $\widehat{\mu}_0$ is very high, actions other than the pair $i_1 = i^H$ and $i_2 = i^L$ reveal little about the type, and the big potential drop in reputation that leads to distortions in Case 3 can lead to substantial reversal aversion and conservatism. If $\widehat{\mu}_0$ is instead very low, the central bank finding itself in Case 3 rather than Case 6 is not that costly; it can simply move its action back to i^L, and because its reputation was low to begin with, the loss relative to having played $i_1 = i^L$ in the first is minimized as both situations have minimal levels of reputational capital. As such, there is little need for conservatism at $t = 1$, which is further reinforced by the fact that actions in period 1 reveal little about the central bank's type when $\widehat{\mu}_0$ is close to 0 or 1 due to the anchoring effect of priors.

Analyzing $t = 2$ (in particular, Case 3) is relatively straightforward, as the state of the world is perfectly revealed and thus the payoffs are thus simple to calculate. However, the desire to reverse or not depends crucially on the equilibrium strategy in period 1, which are determined my the model's parameters, as shown above. A pure strategy, no-distortion equilibrium in Case 3 arises if the inequality given by (2.5) holds. I first note that $A(\sigma_1, \sigma_2 = 1)$ is increasing in σ_1, and thus $\widehat{\mu}_2 \left(i_1 = i^H, i_2 = i^H; \sigma_1, \sigma_2 = 1\right)$ is decreasing in σ_1. Thus, when there is more conservatism at $t = 1$ (smaller σ_1), we have that this inequality is less likely to hold in equilibrium, and thus some form of reversal aversion is more likely; this was illustrated earlier by Figures 2.2a, 2.2b, and 2.2c. Intuitively, as the B type central bank is more conservative in period 1, the action of $i_1 = i^H$ becomes more indicative of the G type, and thus *if* the central bank finds itself in Case 3 (noting that more conservatism makes this less likely), it becomes more tempting to play $i_2 = i^H$ even when the true state is $\theta_2 = L$ in order to protect its (relatively high) reputation. Thus, conservatism and reversal aversion become *complements*, in that more conservatism leads to less reversals, conditional on the central bank finding itself with a need to reverse (of course, with conservatism, there are less

[20] See, e.g., the public's posteriors in Period 1 in the benchmark case given by (2.3) and (2.4); in the reputational equilibrium, the posteriors are more complicated due to the presence of strategic actions on the part of the B type, as shown in (2.6) and (2.7), but the tendency for i^H to be associated with the B type remains.

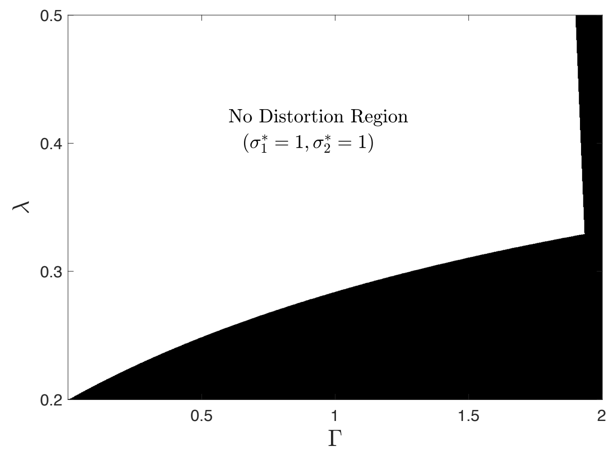


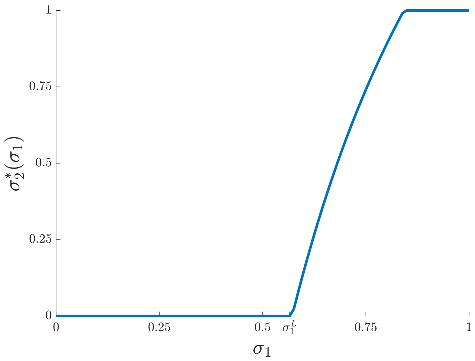
Figure 2.3: Illustration of (Γ, λ) space, showing when a no distortion equilibrium can exist. The white region shows λ and Γ configurations in which an equilibrium with no distortions exists, and the black region shows configurations in which no such equilibrium exists. $\tau = 0.8$ and $\widehat{\mu}_0 = 0.5$ is fixed throughout.

situations where the central bank will need to reverse—this is the point of conservatism in the first place).

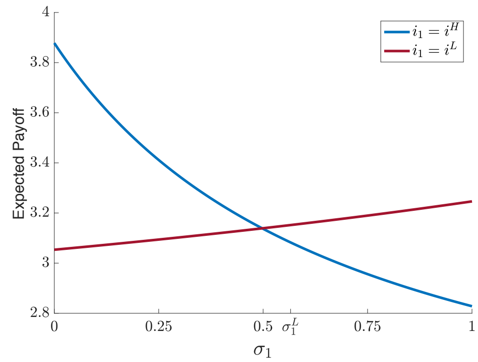
The main goal of this model is to show conditions under which reputational considerations can lead to distortions in optimal monetary policy, rather than provide an exhaustive set of conditions showing when each type of equilibrium arises. As such, I focus on identifying conditions where the pure strategy equilibrium actions of no distortions—the equilibrium under the benchmark case—fails to exist. As the preceding discussion shows, all else equal low values of λ and high values of Γ are likely to lead to conservatism and reversal aversion in actions. Figure 2.3 provides an illustration, dividing up the (λ, Γ) parameter space into regions where the pure strategy equilibrium exists and fails to exist. As shown, in the upper west region, with relatively high λ and low Γ are parameter combinations such that the no distortion equilibrium exists. Generally, as λ increases, a higher Γ is needed for distortions to arise.[21]

I now turn to a numerical example using the baseline parameterization used throughout the paper. The parameter values used are $\lambda = 0.25$, $\tau = 0.8$, $\Gamma = 1.8$, $\widehat{\mu}_0 = 0.5$. With this

[21] As shown, at higher levels of λ, (where high Γ is needed for distortions to arise), an increase in λ in fact *decreases* the threshold Γ. With higher λ, distortions are more likely to arise from reversal aversion in Case 3, rather than from conservatism at $t = 1$; recall that lower λ increases the *likelihood* of Case 3, as well as makes it less likely that the state has indeed changed to $\theta_1 = H$. Thus, lower λ is likely to affect actions at $t = 1$, whereas it has little effect on $t = 2$ actions other than through its influence on the public's posterior beliefs via the $\widehat{\mu}_2$ terms. At higher levels λ, distortions instead come from reversal aversion in Case 3. Here, higher levels of λ are associated with higher levels of $\widehat{\mu}_2\left(i_1 = i_2, i_2 = i^H; \sigma_1 = 1, \sigma_2 = 1\right)$, and thus the reputation cost for moving back down to $i_2 = i^L$ having played $i_1 = i^H$ is larger, requiring a lower level of Γ for reversal aversion to arise.



(a) Equilibrium strategy σ_2 given σ_1



(b) Expected total payoff at $t = 1$

Figure 2.4: Illustration of equilibrium with both reversal aversion and conservatism. Figure (a) shows the unique equilibrium strategy σ_2 for given $t = 1$ equilibrium action σ_1. σ_1^L denotes the threshold value of σ_1 such that when $\sigma_1 < \sigma_1^L$, the pure strategy equilibrium in Case 3 is $\sigma_2 = 0$. Figure (b) shows expected total payoffs from playing i^H and i^L after receiving signal $s_1 = H$, for given equilibrium strategy σ_1 and assuming $\sigma_2 = 0$. Note that there is no pure strategy equilibrium; if $\sigma_1 = 0$ (pure strategy of always playing $i_1 = i^L$), the B type central bank has an incentive to deviate $i_1 = i^H$, and when $\sigma_1 = 1$ (pure strategy of always playing $i_1 = i^H$) it has an incentive to deviate to $i_1 = i^L$. The intersection of the two lines indicate the mixed strategy equilibrium, noting crucially that the intersection comes at a point $\sigma_1 < \sigma_1^L$, so is consistent with the equilibrium action $\sigma_2 = 0$ at $t = 2$. Graphs shown using the baseline parameterization of the model, with $\lambda = 0.25$, $\tau = 0.8$, $\Gamma = 1.8$, $\widehat{\mu}_0 = 0.5$.

parameterization, λ is fairly low; as the previous analysis showed, this is a situation in which some degree of both reversal aversion and conservatism are likely. After receiving a signal of $s_1 = H$, the B type central bank puts 0.57 probability that the true state has moved to $\theta_1 = H$. However, despite the fact that this change in state is more likely than not, there is still considerable probability that the state has not changed. Furthermore, given that the state has not changed, it is very likely that the state remains at L at $t = 2$, leading to a state of the world that is costly to the central bank. The high degree of conservatism, in order to avoid the worst possible state of the world, itself leads to more reversal aversion *should* the central bank find itself in that state.

As seen by Figure 2.3, this is a parameter configuration such that the no-distortion equilibrium fails to exist. Figures 2.4a and 2.4b show how this baseline parameterization leads to distortions both at $t = 2$ (reversal aversion) and at $t = 1$ (conservatism). Figure 2.4a shows the (unique) equilibrium action at $t = 2$ (Case 3) given the $t = 1$ strategy σ_1 (after receiving signal $s_1 = H$). Recall from Figure 2.2, that higher values of σ_1 lead to higher values of σ_2, and thus σ_1 and σ_2 are strategic complements; when $\sigma_1 > \sigma_1^H$ for some σ_1^H the pure strategy $\sigma_2 = 1$ is optimal, and when $\sigma_1 < \sigma_1^L$ for some σ_1^L the pure strategy $\sigma_2 = 0$ is optimal. Conjecturing that in equilibrium this parameterization leads to pure reversal aversion at $t = 2$, Figure 2.4b shows the total expected payoff from playing $i_1 = i^H$ and $i_1 = i^L$ given σ_1, assuming that $\sigma_2 = 0$. As shown, in equilibrium the central bank mixes with probability σ_1^*, noting that $\sigma_1^* < \sigma_1^L$ and so the initial conjecture holds true. In this equilibrium, we have $\sigma_1 < 1$ and $\sigma_2 = 0$, and so both conservatism and reversal aversion arise. A quick inspection looking at strategies where $\sigma_2 = 1$ or where $0 < \sigma_2 < 1$ shows that this is the only equilibrium; while theoretically there is potential for multiple equilibria due to the strategic complementarity between σ_1 and σ_2, with this baseline parameterization, the equilibrium—featuring both reversal aversion and conservatism—is unique.

2.4 Extensions

I now briefly turn to an extension of the model with alternative assumptions. One assumption earlier, made for illustrative purposes and simplification, was that the state never reverts back immediately if it moves to $\theta = H$, in which case playing $i_1 = i^H$ and $i_2 = i^L$ reveals the type to be B with probability 1 immediately given that the G type central bank plays according to the state in each period. I now relax this assumption, by assuming that the state changes in each period with probability λ, independent of the previous state; thus, if the state moves to $\theta_1 = H$, we have that $\theta_2 = L$ with probability λ rather than zero. Note that when λ is fairly low, the relaxation of this assumption makes little difference. In addition, I assume that the B type central bank does *not* know the state perfectly at $t = 2$, and again receives a signal with identical precision τ as in the first period. I maintain the assumption, however, that the B type central bank still learns of the previous period's true state (with the G type also knowing the current period's true state) to maintain tractability. Note also that making this assumption obviates the need to specify off-the-equilibrium path beliefs, as the G type may also reverse with some positive probability.

One additional possibility to consider are potential nonlinearities in the central bank's objective function. So far, the central bank's objective function has been linear in terms of the fundamental component as well as the posterior beliefs. I make that assumption again for simplicity, and to illustrate the main forces in an analytically tractable form. Another key assumption is that the the central bank's assumption does not affect the state of the economy, and thus does not affect the optimal interest rate in the next period in any way through fundamentals. In reality, it may be the case that setting too low (high an interest) rate exacerbates the need to set a higher (lower) interest rate in the next period. Alternatively, it may simply be the case that the losses to inappropriate interest rates are convex; one mistake may not cause a lot of damage to the economy, but multiple mistakes may lead to substantial welfare loss. To address both of these issues, while maintaining the basic structure of the model intact, I assume that there are potentially convex losses to setting the interest rate misaligned with the true fundamentals in consecutive periods. To do this, I assume that the central bank's welfare function is now

$$W^{CB} = \sum_{t=1}^{2} \left\{ I\left(i_t = i^{\theta_t}\right) + \Gamma \widehat{\mu}_t \right\} \; - k \; I\left(i_1 \neq i^{\theta_1}, i_2 \neq i^{\theta_2}\right)$$

where $k > 0$ represents the additional loss from setting the interest rate incorrectly in terms of fundamentals in consecutive periods.

First note that the addition of the k term makes reversal aversion *less* likely, conditional on reaching a state of the world where $i_1 = i^H$ was played but the previous state was revealed to be $\theta_1 = L$, and $\theta_2 = L$ is more likely than $\theta_2 = H$, because not reversing leads to potentially higher costs due to the added loss of consecutive mismatches between the interest rate and true underlying fundamentals. To see this, first for notational simplicity, let $\tilde{p}_1(H, L) \equiv \widetilde{P}(\theta_2 = H \mid \theta_1 = L, s_2 = L) < 1/2$ be the probability that the B type puts on the state being H at $t = 2$, given that $\theta_1 = L$ and it receives a signal of L. Note that this probability is exactly equal to $1 - \widetilde{P}\left(\theta_1 = L \mid s_1 = L\right)$, where $\widetilde{P}\left(\theta_1 = L \mid s_1 = L\right)$ is given by (2.2). Again, consider a *pure strategy* where the CB plays its best assessment of the signal, that is the CB plays L with probability 1, so that $\sigma_2 = 1$. For this pure strategy to exist, it must be the case that

$$\underbrace{1 - \tilde{p}_1(H, L) + \Gamma \widehat{\mu}_2\left(i_1 = H, i_2 = L; \sigma_1, \sigma_2 = 1\right) - k \tilde{p}_1(H, L)}_{\text{Expected payoff playing } i_1 = i^L} \tag{2.19}$$

$$\geq \underbrace{\tilde{p}_1(H, L) + \Gamma \widehat{\mu}_2\left(i_1 = H, i_2 = H; \sigma_1, \sigma_2 = 1\right) - k\left[1 - \tilde{p}_1(H, L)\right]}_{\text{Expected payoff playing } i_1 = i^H}$$

Noting that because $\tilde{p}_1(H, L) < 1/2$, the presence of the k term makes this inequality more likely to hold, and thus reversals more likely. Looking at (2.19), the new setup of the model requires recalculation of $\widehat{\mu}_2\left(i_1 = H, i_2 = H; \sigma_1, \sigma_2\right)$, and crucially calculation of the term $\widehat{\mu}_2\left(i_1 = H, i_2 = L; \sigma_1, \sigma_2\right)$ which was 0 for all values of σ_1 and σ_2 in the previous setup. In order to calculate $\widehat{\mu}_2\left(i_1 = H, i_2 = H; \sigma_1, \sigma_2\right)$, I again enumerate all possible states, and calculate the probability that a B type plays $i_1 = i^H$, and now noting that $(\theta_1, \theta_2) = (H, L)$ occurs with positive probability. We have

1. $\Theta = (H, H)$: probability is $\underbrace{\lambda(1-\lambda)}_{\text{state prob.}} \underbrace{\tau\sigma_1}_{\text{period 1}} \left[\underbrace{\tau}_{\text{H signal period 2}} + \underbrace{(1-\tau)(1-\sigma_2)}_{\text{L signal period 2}} \right]$

2. $\Theta = (H, L)$: probability is $\underbrace{\lambda^2}_{\text{state prob.}} \underbrace{\tau\sigma_1}_{\text{period 1}} \left[\underbrace{(1-\tau)}_{\text{H signal period 2}} + \underbrace{\tau(1-\sigma_2)}_{\text{L signal period 2}} \right]$

3. $\Theta = (L, H)$: probability is $\underbrace{(1-\lambda)\lambda}_{\text{state prob.}} \underbrace{(1-\tau)\sigma_1}_{\text{period 1}} \left[\underbrace{\tau}_{\text{H signal period 2}} + \underbrace{(1-\tau)(1-\sigma_2)}_{\text{L signal period 2}} \right]$

4. $\Theta = (L, L)$: probability $\underbrace{(1-\lambda)^2}_{\text{state prob.}} \underbrace{(1-\tau)\sigma_1}_{\text{period 1}} \left[\underbrace{(1-\tau)}_{\text{H signal period 2}} + \underbrace{\tau(1-\sigma_2)}_{\text{L signal period 2}} \right]$

Letting the sum of these terms equal $A(\sigma_1, \sigma_2)$, we have

$$\widehat{\mu}_2 (i_1 = H, i_2 = H; \sigma_1, \sigma_2) = \frac{\widehat{\mu}_0 \lambda(1-\lambda)}{\widehat{\mu}_0 \lambda(1-\lambda) + (1-\widehat{\mu}_0) A(\sigma_1, \sigma_2)}$$

Similarly to calculate $\widehat{\mu}_2 (i_1 = H, i_2 = L; \sigma_1, \sigma_2)$, the probability the B type plays $i_1 = i^H$ and $i_2 = i^L$ is given by

1. $\Theta = (H, H)$: probability is $\underbrace{\lambda(1-\lambda)}_{\text{state prob.}} \underbrace{\tau\sigma_1}_{\text{period 1}} \left[\underbrace{(1-\tau)\sigma_2}_{\text{L signal period 2}} \right]$

2. $\Theta = (H, L)$: probability is $\underbrace{\lambda^2}_{\text{state prob.}} \underbrace{\tau\sigma_1}_{\text{period 1}} \left[\underbrace{\tau\sigma_2}_{\text{L signal period 2}} \right]$

3. $\Theta = (L, H)$: probability is $\underbrace{(1-\lambda)\lambda}_{\text{state prob.}} \underbrace{(1-\tau)\sigma_1}_{\text{period 1}} \left[\underbrace{(1-\tau)\sigma_2}_{\text{L signal period 2}} \right]$

4. $\Theta = (L, L)$: probability $\underbrace{(1-\lambda)^2}_{\text{state prob.}} \underbrace{(1-\tau)\sigma_1}_{\text{period 1}} \left[\underbrace{\tau\sigma_2}_{\text{L signal period 2}} \right]$

Letting the sum of these terms equal $B(\sigma_1, \sigma_2)$, we have

$$\widehat{\mu}_2 (i_1 = H, i_2 = L; \sigma_1, \sigma_2) = \frac{\widehat{\mu}_0 \lambda^2}{\widehat{\mu}_0 \lambda^2 + (1-\widehat{\mu}_0) B(\sigma_1, \sigma_2)}$$

Rearranging terms and plugging into (2.19), the needed condition for $\sigma_2 = 1$ to be a pure strategy equilibrium is

$$\underbrace{1 - \frac{2(1-\tau)\lambda}{(1-\tau)\lambda + \tau(1-\lambda)}}_{i_2 = i^L \text{expected fundamental gain}} > \underbrace{\Gamma\left[\frac{\widehat{\mu}_0\lambda(1-\lambda)}{\widehat{\mu}_0\lambda(1-\lambda) + (1-\widehat{\mu}_0)\,A\left(\sigma_1,\sigma_2\right)} - \frac{\widehat{\mu}_0\lambda^2}{\widehat{\mu}_0\lambda^2 + (1-\widehat{\mu}_0)\,B\left(\sigma_1,\sigma_2\right)}\right]}_{i_2 = i^H \text{ reputation gain}}$$
$$\underbrace{-k\left[1 + \frac{2(1-\tau)\lambda}{(1-\tau)\lambda + \tau(1-\lambda)}\right]}_{i_2 = i^H \text{expected additional loss}} \tag{2.20}$$

As shown by (2.20)When deciding between playing $i_2 = i^H$ and $i_2 = i^L$, the B type must take into account three different components: the expected relative gain in fundamentals by playing $i_2 = i^L$, the relative gain in reputation by not reversing and playing $i_2 = i^H$, and the expected loss from consecutive mismatches, which is higher when $i_2 = i^H$.

Despite the fact that there may be less reversals in equilibrium, this does not mean that conservatism from reversal aversion is less likely to arise as a distortion. Recall from before in the original setup that the maximum possible loss when arriving at a state in which the central bank must decide between setting an interest rate in line with fundamentals or saving its reputation was 1; if the reputation cost of moving back to i^L was too costly, the central bank would simply forgo the fundamental payoff and lose 1. Here in this alternative setup, because of the presence of k, the potential loss can be much greater than 1; the central bank may find itself having to reverse (because of large costs of consecutive mismatches with fundamentals), and lose a large amount of reputation, or it may find itself not reversing (because of large reputation costs) and a potentially large amount of loss from from consecutive mismatches, in addition to the original fundamental loss. Thus, although reversal aversion may be *less* likely in period $t = 2$ due to the presence of consecutive mismatch costs, the potential costs of arriving at such a state in which a reversal is needed can be very high (potentially much larger than the previous cap of 1), making it more likely that the central bank at $t = 1$ chooses to be *more* conservative in order to avoid any chance of landing at this state of the world. Note that this conservatism still arises from the costs of reversing (loss of reputation), and thus conservatism from the forces of reversal aversion is still present. This conservatism also may make reversals as a whole more unlikely; even though the central bank reverses when it needs to, it avoids this situation in the first place through conservatism.

Lastly, while in the analysis of the previous model I did not formally show comparative statics in terms of the prior $\widehat{\mu}_0$, we see that changes in $\widehat{\mu}_0$ can have large effects on the equilibrium actions through its effects on the $\widehat{\mu}_1$ and $\widehat{\mu}_2$ terms. In particular, at the extreme values of $\widehat{\mu}_0$, the equilibrium actions provide little changes to the beliefs about the central bank's type, unless $i_1 = i^H$ is followed directly after by $i_1 = i^L$. In the context of central banking in practice, when central banks have either developed a reputation (or lack thereof) of knowledgeability, it may be able take actions with little effect on the public's perception of its knowledgeability, and building reputation for knowledgeability may lead to less need for reversal aversion and conservatism. In the original model, however, reversing is extremely costly if $\widehat{\mu}_0$ is high, as then the public's posterior that the central bank is the G type goes from very high to very low. In the extension of the model presented here, note that when $\widehat{\mu}_0$ is high, then $\widehat{\mu}_2\left(i_1 = H, i_2 = H; \sigma_1, \sigma_2\right)$ and $\widehat{\mu}_2\left(i_1 = H, i_2 = L; \sigma_1, \sigma_2\right)$ will both tend to be high as well, meaning there is not much reputation loss from reversing. This is because here the G type may still reverse with some (small) probability, so if the public believes the central bank is very likely to be the G type, reversing changes that perception only slightly. It is in this sense that a central bank may be able to 'build up reputational capital,' in which case it can conduct monetary policy with minimal distortions, and simultaneously enjoy a high perception of its knowledgeability.

2.5 Conclusion

While central banks appear to care considerably about their reputation for knowledgeability, how this affects affects monetary policy has remained unexplored in the academic literature. In this paper, I develop a model in which central banks attempt to shape the public's perception of their knowledgeability in a favorable direction, which endogenously leads to reversal aversion. This leads to distortions in monetary policy and potential welfare costs, as central banks may not reverse their previous decisions ex-post, even when it is appropriate to do so, and this possibility of future reversal aversion may in turn also cause central banks to react slowly in response to shocks in the economy. As shown in the stylized model presented in the paper, distortions from reversal aversion can arise in situations in which central banks are uncertain about the state of the economy; central banks may then become conservative, in order to hedge against the potential state of the world in which they must decide between either setting an inappropriate interest rate or losing their reputation for knowledgeability.

This paper aims to demonstrate how reputation for knowledgeability considerations can affect the behavior of monetary policy, doing so in a very stylized and simple model. A natural way to extend the model is to consider a setting with a longer time horizon, as in the reputational models of Kreps et al. (1982) and Backus and Driffill (1985). Like those papers, the model in this paper features an agent attempting to mimic a 'behavioral' type, with higher payoffs accruing to the agent when outside players falsely believe that the agent is the behavioral type. However, in those models, the posterior probabilities put on the agent being the behavioral type are straightforward to calculate, given the mechanical actions taken by the behavioral type and the recursive structure of the models. In this paper, the state itself is constantly changing, and so in order to calculate these posterior probabilities, one must take into account the entire sequence of possible true states, as well as all the actions taken by the central bank in the past. Doing so makes these probabilities quickly intractable. Modifying the model and simplifying in other respects in order to make the analysis amenable to a finitely-repeated games framework is likely to yield interesting results.

Another possible direction for research is to merge the findings presented in this paper in a larger scale macroeconomic model. Choi (2018) shows in a dynamic general equilibrium model that in a setting in which the central bank's knowledgeability is private information to the central bank, welfare outcomes are improved when the central bank is perceived as more knowledgeable. However, in that setting the optimal monetary policy, which takes into account how the actions of the central bank might affect the dynamic evolution of beliefs about the central bank's knowledgeability, remains unchanged relative to a benchmark model in which the central bank's knowledgeability is common knowledge. In the model presented here, the central bank's concern about its reputation for knowledgeability is taken as given, and it is shown that optimal monetary policy from the perspective of the central bank may feature considerable distortions relative to a benchmark case in which its knowledgeability is common knowledge. Merging the ideas of these two papers—showing how and why optimal monetary policy changes in a micro-founded general equilibrium macroeconomic model in which the central bank's knowledgeability is not fully observed—remains work for future research.

Chapter 3

Perceptions of Central Bank Knowledgeability and the Signaling Channel: An Empirical Analysis

3.1 Introduction

A burgeoning literature has identified the presence of a signaling channel of monetary policy, in which a central bank's actions reveal information about its own assessment of the economy, with private sector expectations following suit. Often called *information effects*, a consistently found empirical result is a general increase (decrease) in private expectations of output and inflation after a contractionary (expansionary) monetary shock, in contrast with standard macroeconomic models which would predict an opposite response. The general reasoning provided is that the central bank would only increase (decrease) rates if confident about the state of the economy, and private sector expectations, internalizing this, follow suit. Despite this compelling story and strong empirical findings, many questions remain. What is the nature of these information effects? What determines the strength of information effects, and which private sector expectations are most reactive to these effects? How does the distribution of forecasts change after a monetary shock?

This paper aims to make progress on these questions by identifying heterogeneity in information effects both cross-sectionally and over time, providing evidence in support of dispersed information models as well as the model presented in Choi (2018), a dispersed information model in which perceptions of a central bank's knowledgeability affects the strength of information effects. As that paper claims, because the essence of information effects is the provision of the central bank's private information about the state of the economy through monetary actions, the strength of these information effects should depend crucially on whether or not the private sector believes the central bank has valuable information to start with. Looking at the U.S. and the Federal Reserve, and the response of professional forecasts to monetary shocks, I show that information effects appear substantially stronger when publicly announced forecasts made by the Federal Reserve were more accurate in the recent past, a key prediction of the model. The accuracy of recently announced public forecasts is used as a proxy for the public's perception of the Federal Reserve's knowledgeability, which has a direct analog in Choi (2018).

This paper is also able to answer broader questions on the nature of information effects by using data at the individual forecast level, as opposed to just a simple time series of mean forecasts that have so far been used in the literature. Looking at individual level forecast data allows one to look at how the entire distribution of forecasts changes after a monetary shock, and also allows one to track individual forecasts before and after a shock. Together, this can help distinguish between different theories of information effects, and also to provide evidence in favor of dispersed information models more generally. In particular, I find not only larger changes in mean private forecasts following a monetary shock after a period of more accurate public forecasts by the Federal Reserve, but a larger reduction in the variance of forecasts as well. In addition, by tracking the same forecasters before and after a monetary shock, I show heterogeneity of information effects across forecasters, depending on the relative position of a forecast to that of the Federal Reserve's internal forecast prior to the monetary shock. In particular, I find that after a contractionary (expansionary) monetary shock, only forecasters below the Federal Reserve's Greenbook forecast upwardly (downwardly) revise their forecast of output; such behavior is a prediction generated by dispersed information models, and contrary to many existing information effects models that predict a uniform shift in expectations across forecasters. Thus, there appears to be heterogeneity in information effects cross-sectionally, consistent with dispersed information models, and heterogeneity across time depending on the public's perception of the central bank's knowledgeability, a unique prediction of the dispersed information model of Choi (2018).

3.1.1 Related Literature

Empirical evidence of the signaling channel of monetary policy and information effects is found in Romer and Romer (2000), Campbell et al. (2012), El-Shagi et al. (2014), Hubert (2014; 2015), and Nakamura and Steinsson (2017). These papers look at the responses of professional forecasts to Federal Open Market Committee (FOMC) announcements, focusing on output, inflation, and unemployment forecasts, concluding that professional forecasters infer that unexpected FOMC policy adjustments are responses to nonpublic information that the FOMC has regarding the future economy. While this paper follows a similar approach, it differs by first using individual level forecast data, providing a more nuanced description of how information effects operate, and also by looking at the heterogeneity of information effects, both cross-sectionally and over time.

Theoretical models of the signaling channel of monetary policy include Adam (2007), Baeriswyl and Cornand (2010), Roca (2010), Berkelmans (2011), Tang (2015), and Melosi (2017). This paper most closely provides evidence of the model presented in Choi (2018), in which information effects depend on whether or not a central bank's forecasts were accurate or not in the recent past.

3.2 Empirical Evidence

I now turn to empirical evidence in support of the theoretical model presented in Choi (2018). The core idea of the model is that central bank "competence" — defined here as the accuracy of a central bank's information set — is publicly unobservable, and that the perception of this competence comes from the accuracy (or lack thereof) of the central bank's information in the past. The resulting implication, and the focus of the model, is that when the central bank's information is perceived as having been inaccurate in the recent past, this will lead to a weakening signaling channel and less coordination of expectations in the future.

While the reputation for central bank competence is perhaps more relevant in emerging market economies, where the ability of the central bank is potentially considerably more uncertain, in this section I provide empirical support for this model by looking at the U.S. Federal Reserve (Fed). I use the accuracy of past published forecasts as a proxy for the perceived past precision of the Fed's information, and use this as an empirical counterpart for the key variable $\mathcal{B}$ in the model, the beliefs about the central bank's competence. Focusing on the response of professional forecasters to monetary announcements, I show that information effects — both a signaling effect as well as a coordination/focal point effect — are stronger when the Fed's published forecasts were more accurate in the recent past, a key prediction in line with the theoretical model. In addition, by looking at *individual* level forecasts, I show that information effects appear to work within the context of a dispersed information model, as opposed to more homogenous information effects that have been proposed in this literature.

The baseline specifications follow the existing literature on Fed information effects; in particular, I most closely follow Nakamura and Steinsson (2017), who also look at the response of private sector forecasts to monetary shocks. They show that after a *contractionary* monetary shock, private sector expectations of output *increase*, and vice-versa for an expansionary shock. This is counter to the standard prediction that expectations of output would decrease, due to the contractionary nature of the shock. The explanation given is an information story; the Fed would only increase rates if it was confident about the state of the economy, and this information is implicitly revealed to the public, with their expectations following suit. In my empirical analysis, I document not only changes in mean forecasts, but a convergence in expectations *relative* to the Fed's internal forecasts, with large shocks providing large amounts of information and thus heavily reducing the dispersion of forecasts around the Fed's forecasts; I then show that these myriad information effects appear substantially stronger when the Fed's public forecasts were more accurate in the recent past.

3.2.1 Data

For private sector forecasts, I use the *Blue Chip Economic Indicators* (BCEI) forecasts, a monthly report conducted by Blue Chip Publications. In the first few days of every month, professional forecasters are surveyed on their projections of various economic indicators; following the literature on information effects, I focus on forecasts made on real GDP growth. This data set begins in May 1985, with approximately 50 individual forecasts for real GDP growth per month, with the identities of forecasters tracked over time.

In order to construct a proxy for the perceived precision of the Fed's information, I use recent forecasts made public by the Federal Open Market Committee (FOMC). Since 1979, the Fed has been mandated to make public twice a year (February and July) a report on the state of the economy through the *Monetary Policy Reports* (*MPR*) delivered in the Humphrey-Hawkins testimony.[1] Included in the *Monetary Policy Reports* are heavily followed forecasts for real and nominal GDP growth, inflation, as well as unemployment. While each member of the FOMC provides their own forecasts, only the median and "central tendency" (a measure of the range excluding outliers) are reported; I use the median forecast as my measure of the FOMC's forecast. The forecasts for GDP growth are forecasts for Q4/Q4 growth, so are not entirely comparable to the GDP forecasts from the BCEI; nevertheless, they are a measure of the Fed's ability to forecast GDP growth, which is all that is needed to construct the proxy. In order to construct the real-time accuracy measure, I supplement the *MPR* with vintage data from the Bureau of Economic Analysis (BEA), which include the BEA's best estimates of past GDP at a given date. Vintage data is used so that the perceived accuracy measure uses only information available to private forecasters at the time of their forecasts.

In order to analyze information effects, I use the Greenbook (Fed staff) forecasts for real GDP growth. Each forecast of RGDP in the BCEI data has a direct counterpart in the Greenbooks. The Greenbooks are known to represent the Fed's best forecasts,[2] and are thus representative of the Fed's information set at the time they are constructed. The Greenbooks are only available with a 5-year lag, and are thus not observable to the public at the time they are constructed. However, the essence of information effects are about the private sector's information moving towards the central bank's information; thus, as an econometrician, one can use the Greenbooks to analyze if information effects are indeed present.

Lastly, in order to analyze the effects of monetary news shocks, I follow Nakamura and Steinsson (2017), and use their measure of 'Policy News Shocks,' which looks at unexpected changes in interest rates in a 30-minute window surrounding FOMC announcements as a measure of news about monetary policy. In particular, they look at five different interest rates at different maturities (spanning the first year of the term structure), and extract the first principle component as a composite measure of the monetary shock, which they then scale so that the effect of an increase of one unit of the policy shock on the one-year Treasury yield is 100 basis points.

[1] Starting from 2010, these have since been replaced by the Summary of Economic Projections.

[2] see Romer and Romer (2000)

	Greenbook Inaccuracy
Past Inaccuracy	0.372*
	(0.190)
Observations	103

Table 3.1: Results from simple regression of Greenbook accuracy on the proxy for *perceived* accuracy (past accuracy of publicly announced FOMC forecasts).

For the analysis, I use the time period from 1995-2011, where data is available from all sources. Following the existing literature, I exclude the peak of the financial crisis (second half of 2008) for the baseline analysis, although the results are robust to including this period.

3.2.2 Empirical Specifications

There are many possible ways to construct the perceived accuracy measure, which translates to the key variable $\mathcal{B}$ in the model; one could use just the most recent public forecasts, or also use forecasts made from the more distant past, weighted appropriately. Another difficulty arises from the *MPR* forecasts being fixed event forecasts, and thus forecasts made in February and July for current year GDP growth are not comparable; naturally, the forecasts made in July are substantially more accurate. As such, I use as my perceived accuracy measure one of the simplest possible options, looking at the most recent February RGDP growth forecast made by the Fed, such that the forecasted variable has been realized with corresponding estimates by the BEA;[3] I then simply take the magnitude of the error as my measure of past inaccuracy in period t: $pastinacc_t = \mid HH_{recent} - BEA_{recent} \mid$. This measure is used for simplicity, and when constructing other alternative accuracy measures the empirical results are broadly consistent.

3.2.2.1 Preliminary Check

In the model, the type of the central bank θ is persistent, meaning that the precision of the central bank's information is persistent. Empirically, this should mean that accuracy in the *past* should correlate with accuracy *today*. In addition, if the proxy constructed for the perceived precision is valid, it should be the case that inaccuracy of public forecasts made in the past should be correlated with inaccuracy of Greenbook forecasts (the Fed's current information) in the current period. To test

[3] I use February due to the longer horizon of the forecasts.

this, I do a simple regression of $pastinacc_t$ on the Greenbook accuracy (measured ex-post, after the forecasted variable is realized), defined as $gbaccuracy_t =\mid GB_t - RGDPG_t \mid$, and cluster standard errors by the observations that share the same past inaccuracy measure. The results are shown in (Table 3.1); past accuracy does indeed appear to have substantial predictive power of current Fed information precision, with strong statistical significance.

3.2.2.2 Main Specifications

Given the construction of this measure, I now test the main empirical prediction of the model: that information effects are stronger (weaker) when the perceived precision of the central bank's information — proxied by past accuracy of recent public forecasts — is higher. I estimate the following regression:

$$\triangle y_t = \beta_0 + \beta_1 \triangle m_t + \beta_2 \triangle m_t \times pastinacc_t + \lambda' X_t + \epsilon_t \tag{3.1}$$

where the dependent variable $\triangle y_t$ denotes the monthly change in the mean Blue Chip survey expectations about output growth, and the independent variable $\triangle m_t$ denotes the policy news shock variable, which is interacted with the past inaccuracy measure to test for dependence of the information effect on the past public forecasts made by the Fed. The key finding in Nakamura and Steinsson (who use a simple regression framework) is a positive coefficient β_1: a contractionary monetary shock is followed by an upward revision in output growth forecasts. The new prediction of this model is a *negative* coefficient β_2, indicating that information effects are muted when the perceived accuracy (proxied by past accuracy) of the Fed is lower.

Another possible way to test for information effects is to instead look at the dispersion of forecasts, relative to the Greenbook forecasts. One simple approach is to look at changes in the mean squared error $\triangle MSE_t = \frac{1}{N}\sum_i \left(\widehat{y}_t^i - \widehat{y}_t^{GB}\right)^2 - \frac{1}{N}\sum_i \left(\widehat{y}_{t-1}^i - \widehat{y}_t^{GB}\right)^2$, which captures both the dispersion of private forecasts, as well as the mean difference between the private forecasts and the Greenbook forecasts. I use the *magnitude* of policy shocks as the independent variable; if information effects are first-order, then a large magnitude of the policy shock is likely to be an event where a lot of new information is conveyed by the Fed. Due to the large informational surprise, one would then expect a large decrease in the MSE. To test for this, I run the following specification:

	(1) Change RGDP	(2) Change Sq. Error	(3) Change Abs. Error
Policy News Shock	2.366**		
	(1.098)		
Policy News Shock × Past Inaccuracy	−1.341**		
	(0.596)		
Abs(Policy News Shock)		−2.083***	−1.805***
		(0.597)	(0.549)
Abs(Policy News Shock) × Past Inaccuracy		0.787*	0.713*
		(0.453)	(0.415)
Cal. Month FE	*Y*	*Y*	*Y*
Controls	*Y*	*Y*	*Y*
N	92	92	92

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.2: Estimates from main specifications. First column contains estimates from (3.1), second column contains estimates from (3.2), and third column contains estimates from (3.2) but using the change in absolute deviation as the dependent variable.

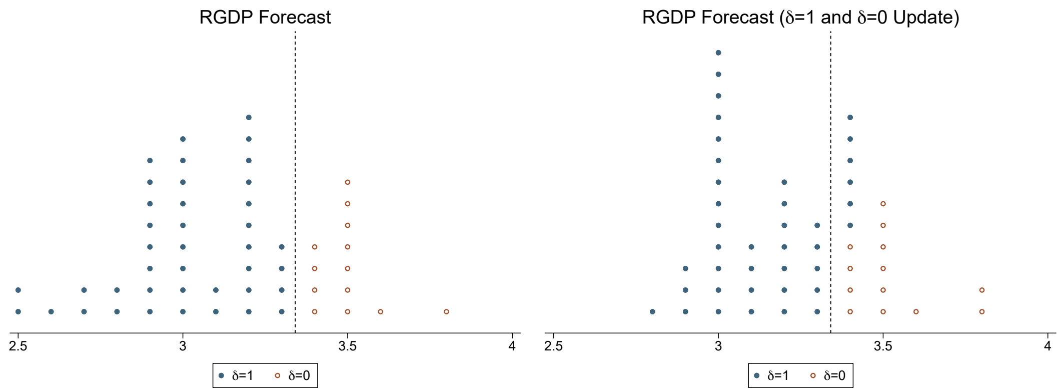
$$\triangle MSE_t = \beta_0 + \beta_1 \left|\triangle m_t\right| + \beta_2 \left|\triangle m_t\right| \times pastinacc_t + \lambda' X_t + FE_{m(t)} + \epsilon_t \quad (3.2)$$

this time interacting the past inaccuracy measure with the magnitude of the policy shock. Here, information effects would lead to a *negative* coefficient β_1, and a *positive* coefficient β_2 if this effect is muted when the recent forecast of the Fed was inaccurate. I include calendar month fixed effects in this specification due to the potential that more information about output may be revealed in some months (e.g., a BEA report that comes out every April).

The last specification I run replaces the dependent variable $\triangle MSE_t$ in the above specification with the change in the absolute deviation, or $\frac{1}{N}\sum_i \left|\widehat{y}_t^i - \widehat{y}_t^{GB}\right| - \frac{1}{N}\sum_i \left|\widehat{y}_{t-1}^i - \widehat{y}_t^{GB}\right|$, an alternative measure for if the private forecasts move toward the Greenbook forecasts.

3.2.3 Results

Table 3.2 presents the regression results from the main specifications. The first column shows the estimates from the specification given by (3.1). As shown, the significant positive coefficient on the policy news shock variable indicates the presence of information effects; when the Fed's recent forecast has been accurate, an increase in the monetary shock tends to increase forecasts of output growth, whereas the negative coefficient on the interaction term indicates that this effect tends to be smaller when the Fed's forecasts have been inaccurate in the recent past. This coefficient is



(a) Pre-Shock Forecasts **(b)** Post-Shock Forecasts

Figure 3.1: Distribution of RGDP growth forecasts in April 2010 and May 2010, where in between these forecasts was a contractionary policy news announcement. Vertical line denotes Greenbook RGDP growth forecast. $\delta = 1$ group denotes forecasts below the Greenbook forecast at the time of shock, and $\delta = 0$ denotes those above.

significant both statistically and economically: the standard deviation of the inaccuracy measure is 1.22, and a one standard deviation increase in inaccuracy from the mean is enough to wipe out the information effect. The second and third columns, representing the estimates from specification (3.2), tell a similar story. The negative coefficient on the magnitude of the policy shock variable indicates the presence of information effects, with the statistically significant positive coefficient on the interaction term consistent with a muting of these effects after a recent inaccurate forecast. All of these results are consistent with the predictions of the theoretical model: after large forecasting errors by the central bank, private sector forecasts are subsequently less influenced by the signaling component of monetary policy going forward.

3.2.4 Other Supplementary Evidence

As mentioned, one new contribution of this empirical section is to use *individual-level* private forecasts, as opposed to simply mean forecasts that have been analyzed in the literature.[4] Looking at mean forecasts alone cannot capture the potentially rich distributional structure of information effects: while a contractionary shock may lead to an overall upward revision of forecasts of output growth, how does the *distribution* of forecasts change? In particular, in a dispersed information

[4] Note that the main specifications looking at the mean-squared error is only possible using individual level data.

model, public signals have very heterogenous effects on private expectations, depending on where a private expectation is relative to the public signal. The empirical counterpart here is that information effects should differ depending on where a private forecast is relative to the Greenbook forecast. A contractionary shock is more likely when the Fed's forecast is higher than the mean private sector forecast (this is heavily supported in the data); after the shock, however, those *below* the Greenbook forecasts should revise their estimates upwards, with those above the Fed's forecasts much less likely to do so. In essence, in a dispersed information model, information provided by a public signal—here the monetary shock from Fed policy—provides a larger revision of beliefs for agents with information furthest away from the given public signal. This *focal point* effect is distinct from an alternative simple story whereby an information shock induces a uniform increase across the distribution of forecasts.

Figure 3.1 illustrates these heterogenous effects after a moderate contractionary policy news shock in April 2010. After the policy news announcement, there is large upward revision in forecasts for those below the Greenbook forecast, with almost no movement at all for those above the forecast. To test for heterogenous effects , I split the sample in a very simple manner: for contractionary shocks, I let $\delta_{it} = 1$ if the private forecast before the shock is below the Greenbook forecast, and 0 otherwise, and for expansionary shocks, I let $\delta_{it} = 1$ if the forecast is above the Greenbook forecast, and 0 otherwise. I then test if information effects are heterogenous between the two groups: after splitting the sample, I run

$$\triangle y_{it} = \beta_0 + \beta_1 \triangle m_t + \lambda' X_t + FE_{m(t)} + \epsilon_t$$

$$\triangle MSE_t = \beta_0 + \beta_1 \left| \triangle m_t \right| + \beta_2 \left| \triangle m_t \right| + \lambda' X_t + FE_{m(t)} + \epsilon_t$$

separately for the two groups. Table 3.3 presents the results. Information effects appear very strong for the $\delta = 1$ group, and non-existent for the $\delta = 0$ group, consistent with the dispersed information model and a focal point effect story. For the $\delta = 0$ group, revisions in expectations in fact go in the *opposite* direction of that predicted in a standard information effects story, consistent with all private expectations converging towards that of the public signal. Broadly, the empirical evidence points in favor of dispersed information models, and within the class of dispersed information models, is also consistent with the unique predictions borne out in the model developed in Choi (2018).

	Change RGDP		Change Abs. Error		Change Sq. Error	
	δ=1	δ=0	δ=1	δ=0	δ=1	δ=0
Policy News Shock	2.393***	−2.481**				
	(0.380)	(1.015)				
Abs(Policy News Shock)			−1.516***	−1.662	−2.066***	−0.906
			(0.397)	(1.628)	(0.518)	(1.341)
RGDP Revision	0.213***	0.197				
	(0.054)	(0.119)				
Month FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
N	103	103	103	103	103	103

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: Regression results testing for information effects after splitting sample. $\delta = 1$ corresponds to forecasts below (above) the Greenbook forecast when a contractionary (expansionary) policy news shock hits, and $\delta = 0$ corresponds to the reverse.

3.3 Conclusion

This paper provided evidence of heterogeneity in information effects both cross-sectionally, in support of dispersed information models, as well as over time, depending on the public's perception of the central bank's knowledgeability, in support of the model presented by Choi (2018). Focusing on the U.S. and the Federal Reserve, and constructing a proxy for the public's perception of the Federal Reserve's knowledgeability by using previously announced forecasts made by the FOMC, this paper showed that information effects — the provision of private information about the state of the economy through monetary actions — appear substantially stronger when these public FOMC forecasts were more accurate in the recent past. When recent forecasts were more accurate, private expectations of output increased (decreased) after a contractionary (expansionary) monetary shock substantially more relative to when recent forecasts were inaccurate. In addition, the distribution of forecasts around the Federal Reserve's internal Greenbook forecasts shrunk relatively more after any given monetary shock.

Cross-sectionally, information effects were much stronger for individual forecasts that dispersed information models would predict to react most strongly to monetary shocks. Individual forecasts of output below (above) the FOMC's internal Greenbook forecast after a contractionary (expansionary) monetary shock increased (decreased) considerably, with revision of the forecasts towards that of the Greenbook forecast, with no such effect for those above (below) the FOMC's internal Greenbook forecast. Such heterogeneity is consistent with the signaling channel of monetary policy in a dispersed information framework.

While this paper provided a preliminary analysis of the drivers of information effects, more work should be done. Looking at private sector expectations of other macroeconomic variables, as well as constructing other possible proxies for the public's perception of the central bank's knowledgeability are different avenues to explore. In addition, this analysis can be studied in other countries, which may have stronger heterogenous effects due to the potentially greater amount of uncertainty about the central bank's knowledgeability. This is left for future research.

References

ADAM, K. (2007): "Optimal Monetary Policy with Imperfect Common Knowledge," *Journal of Monetary Economics*, 54, 267–301.

ALLEN, F., S. MORRIS, AND H. S. SHIN (2006): "Beauty Contests and Iterated Expectations in Asset Markets," *Review of Financial Studies*, 19, 719–752.

AMATO, J. D., S. MORRIS, AND H. S. SHIN (2002): "Communication and Monetary Policy," *Oxford Review of Economic Policy*, 18, 495–503.

ANGELETOS, G.-M. AND J. LA'O (2009): "Incomplete Information, Higher-Order Beliefs and Price Inertia," *Journal of Monetary Economics*, 56, S19–S37.

——— (2012): "Optimal Monetary Policy with Informational Frictions," *Working Paper*.

ANGELETOS, G.-M. AND C. LIAN (2016): "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination," *Handbook of Macroeconomics*, 2, 1065–1240.

ANGELETOS, G.-M. AND A. PAVAN (2004): "Transparency of Information and Coordination in Economies with Investment Complementarities," *American Economic Review*, 94, 91–98.

——— (2007): "Efficient Use of Information and Social Value of Information," *Econometrica*, 75, 1103–1142.

AOKI, K. (2003): "On the Optimal Monetary Policy Response to Noisy Indicators," *Journal of Monetary Economics*, 50, 501–523.

——— (2006): "Optimal Commitment Policy Under Noisy Information," *Journal of Economic Dynamics and Control*, 30, 81–109.

ATKESON, A., V. V. CHARI, AND P. J. KEHOE (2010): "Sophisticated Monetary Policies," *The Quarterly Journal of Economics*, 125, 47–89.

BACCHETTA, P. AND E. VAN WINCOOP (2006): "Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?" *The American Economic Review*, 552–576.

BACKUS, D. AND J. DRIFFILL (1985): "Inflation and Reputation," *The American Economic Review*, 75, 530–538.

BAERISWYL, R. AND C. CORNAND (2010): "The Signaling Role of Policy Actions," *Journal of Monetary Economics*, 57, 682–695.

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

BARRO, R. J. AND D. B. GORDON (1983a): “A Positive Theory of Monetary Policy in a Natural Rate Model,” *Journal of Political Economy*, 91, 589–610.

——— (1983b): “Rules, Discretion and Reputation in a Model of Monetary Policy,” *Journal of Monetary Economics*, 12, 101–121.

BERKELMANS, L. (2011): “Imperfect Information, Multiple Shocks, and Policy’s Signaling Role,” *Journal of Monetary Economics*, 58, 373–386.

BIANCHI, F. AND L. MELOSI (2016): “Constrained Discretion and Central Bank Transparency,” *Working Paper*.

BRAINARD, W. C. (1967): “Uncertainty and the Effectiveness of Policy,” *The American Economic Review*, 57, 411–425.

CALVO, G. A. (1983): “Staggered Prices in a Utility-Maximizing Framework,” *Journal of Monetary Economics*, 12, 383–398.

CAMPBELL, J. R., C. L. EVANS, J. D. FISHER, AND A. JUSTINIANO (2012): “Macroeconomic Effects of Federal Reserve Forward Guidance,” *Brookings Papers on Economic Activity*, 2012, 1–80.

CHOI, D. (2018): “Perceptions of Competence: Monetary Policy and the Reputational Accelerator,” *Working Paper*.

COCHRANE, J. H. (2011): “Determinacy and Identification with Taylor Rules,” *Journal of Political economy*, 119, 565–615.

CUKIERMAN, A. AND A. H. MELTZER (1986): “A Theory of Ambiguity, Credibility, and Inflation under Discretion and Asymmetric Information,” *Econometrica*, 54, 1099–1128.

EL-SHAGI, M., S. GIESEN, AND A. JUNG (2014): “Does the Federal Reserve Staff Still Beat Private Forecasters?” *Working Paper*.

ELLISON, M. (2006): “The Learning Cost of Interest Rate Reversals,” *Journal of Monetary Economics*, 53, 1895–1907.

EUSEPI, S. AND B. PRESTON (2010): “Central Bank Communication and Expectations Stabilization,” *American Economic Journal: Macroeconomics*, 2, 235.

EVANS, C., J. FISHER, F. GOURIO, AND S. KRANE (2016): “Risk management for monetary policy near the zero lower bound,” *Brookings Papers on Economic Activity*, 2015, 141–219.

FRANKEL, A. AND N. KARTIK (2016): “What Kind of Central Bank Competence?” *Working Paper*.

GOODFRIEND, M. (1991): “Interest Rates and the Conduct of Monetary Policy,” *Carnegie-Rochester Conference Series on Public Policy*, 34, 7–30.

GUST, C. J., B. K. JOHANNSEN, AND D. LOPEZ-SALIDO (2015): “Monetary Policy, Incomplete Information, and the Zero Lower Bound,” *Working Paper*.

HANSEN, S. AND M. MCMAHON (2016): “First impressions matter: Signalling as a source of policy dynamics,” *The Review of Economic Studies*, 83, 1645–1672.

HELLWIG, C. (2005): “Heterogeneous Information and the Welfare Effects of Public Information Disclosures,” *Working Paper*.

HOLMSTROM, B. (1999): "Managerial Incentive problems: A Dynamic Perspective," *The Review of Economic studies*, 66, 169–182.

HUBERT, P. (2014): "FOMC Forecasts as a Focal Point for Private Expectations," *Journal of Money, Credit and Banking*, 46, 1381–1420.

——— (2015): "Do Central Bank Forecasts Influence Private Agents? Forecasting Performance versus Signals," *Journal of Money, Credit and Banking*, 47, 771–789.

JAMES, J. G. AND P. LAWLER (2012): "Strategic Complementarity, Stabilization policy, and the Optimal Degree of Publicity," *Journal of Money, Credit and Banking*, 44, 551–572.

——— (2016): "Optimal Transparency and Policy Intervention with Heterogeneous Signals and Information Stickiness," *The Manchester School*, n/a–n/a.

KING, R. G. (1982): "Monetary Policy and the Information Content of Prices," *Journal of Political Economy*, 90, 247–279.

KOBAYASHI, T. (2010): "Policy Irreversibility and Interest Rate Smoothing," *The BE Journal of Macroeconomics*, 10.

KREPS, D. M., P. MILGROM, J. ROBERTS, AND R. WILSON (1982): "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic theory*, 27, 245–252.

LUCAS, R. E. (1972): "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 4, 103–124.

MANKIW, N. G. AND R. REIS (2010): "Imperfect Information and Aggregate Supply," *Handbook of Monetary Economics*, 3, 183 – 229.

MELOSI, L. (2017): "Signalling Effects of Monetary Policy," *The Review of Economic Studies*, 84, 853–884.

MEYN, S. P. AND R. L. TWEEDIE (2012): *Markov Chains and Stochastic Stability*, Springer Science & Business Media.

MORRIS, S. AND H. S. SHIN (2002): "Social Value of Public Information," *The American Economic Review*, 92, 1521–1534.

MOSCARINI, G. (2007): "Competence Implies Credibility," *American Economic Review*, 97, 37–63.

NAKAMURA, E. AND J. STEINSSON (2017): "High Frequency Identification of Monetary Non-Neutrality: the Information Effect," *Working Paper*.

ORPHANIDES, A. (2003): "Monetary Policy Evaluation with Noisy Information," *Journal of Monetary Economics*, 50, 605–631.

ORPHANIDES, A. AND J. C. WILLIAMS (2002): "Robust Monetary Policy Rules with Unknown Natural Rates," *Brookings Papers on Economic Activity*, 63–146.

——— (2007): "Robust Monetary Policy with Imperfect Knowledge," *Journal of Monetary Economics*, 54, 1406–1435.

PHELPS, E. S. (1970): *Microeconomic Foundations of Employment and Inflation Theory*, Norton.

——— (1983): “The Trouble with Rational Expectations and the Problem of Inflation Stabilization,” *Individual Forecasting and Aggregate Outcomes*.

PRENDERGAST, C. AND L. STOLE (1996): “Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning,” *Journal of Political Economy*, 104, 1105–1134.

ROCA, M. (2010): *Transparency and Monetary Policy with Imperfect Common Knowledge*, 10-91, International Monetary Fund.

ROMER, C. D. AND D. H. ROMER (2000): “Federal Reserve Information and the Behavior of Interest Rates,” *American Economic Review*, 90, 429–457.

SACK, B. (1998): “Uncertainty, Learning, and Gradual Monetary Policy,” *Working Paper*.

SARGENT, T. J. AND N. WALLACE (1975): “"Rational" Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule,” *Journal of Political Economy*, 83, 241–254.

SELTEN, R. (1975): “Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games,” *International Journal of Game Theory*, 4, 25–55.

STEIN, J. C. AND A. SUNDERAM (2017): “The Fed, the Bond Market, and Gradualism in Monetary Policy,” *Working Paper*.

SVENSSON, L. E. AND M. WOODFORD (2003): “Indicator Variables for Optimal Policy,” *Journal of Monetary Economics*, 50, 691–720.

——— (2004): “Indicator Variables for Optimal Policy Under Asymmetric Information,” *Journal of Economic Dynamics and Control*, 28, 661–690.

TANG, J. (2015): “Uncertainty and the Signaling Channel of Monetary Policy,” *Working Paper*.

TAYLOR, J. B. AND J. C. WILLIAMS (2010): “Simple and Robust Rules for Monetary Policy,” *Handbook of Monetary Economics*, 3, 829–859.

TOWNSEND, R. M. (1983): “Forecasting the Forecasts of Others,” *Journal of Political Economy*, 91, 546–588.

WALSH, C. E. (1995): “Optimal Contracts for Central Bankers,” *The American Economic Review*, 150–167.

WOODFORD, M. (2003a): “Imperfect Common Knowledge and the Effects of Monetary Policy,” *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, 25.

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

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

Appendix A

Appendix to Chapter 1

A.1 Firm Price Setting Equation

The profit maximization problem of firm i is given by

$$\max_{P_t^i} \mathbb{E}_t \left[m(Y_t)\Pi(P_t^i, P_t; Y_t, \bar{\xi}_t) \mid \mathcal{I}_t^i \right]$$

where $m(Y_t)$ denotes the stochastic discount factor and $\Pi = (1+\tau)P_t(i)Y_t(i) - W_t(i)H_t(i)$. Clearing of the labor market implies that

$$\frac{v_h(H_t(i);\xi_t)}{u_c(C_t;\xi_t)} = \frac{W_t(i)}{P_t}$$

and plugging in the relative demand function from (1.5), we can rewrite the profit function as

$$\Pi(P_t^i, P_t; Y_t, \bar{\xi}_t) = (1+\tau)\left(\frac{P_t(i)}{P_t}\right)^{-\eta} P_t(i) - \frac{v_h(H_t;\xi_t)}{u_c(C_t;\xi_t)} P_t H_t(i)$$

The FOC of firm i is then[1]

$$\mathbb{E}_t\left[m(Y_t)\eta\left[\frac{1}{f'(f^{-1}(\frac{Y_t(i)}{A_t}))}\frac{1}{A_t}\frac{v_h(H_t(i);\xi_t)}{u_c(C_t;\xi_t)}\left(\frac{P_t(i)}{P_t}\right)^{-\eta-1}-\left(\frac{P_t(i)}{P_t}\right)^{-\eta}\right]\,\middle|\,\mathcal{I}_t^i\right]=0$$

which can be rewritten as

$$\mathbb{E}_t\left[m(Y_t)\eta\left(\frac{P_t(i)}{P_t}\right)^{-(1+\eta)}\left[\left(\frac{P_t(i)}{P_t}\right)-s\left(Y_t(i),Y_t;\tilde{\xi}_t\right)\right]\,\middle|\,\mathcal{I}_t^i\right]=0 \tag{A.1}$$

where

$$s\left(y,Y;\tilde{\xi}\right)\equiv\frac{v_h(f^{-1}(y/A);\xi)}{u_c(Y;\xi)A}\Psi(y/A) \tag{A.2}$$

represents the real marginal cost function and

$$\Psi(y)\equiv\frac{1}{f'(f^{-1}(y))}$$

As shown in Woodford (2003b), the (log) of the natural rate of output can be rewritten as a linear combination of the real disturbances,

$$\hat{Y}_t^N=\frac{\sigma^{-1}g_t+\omega q_t}{\sigma^{-1}+\omega}+\mathcal{O}\left(\|\,\tilde{\xi}\,\|^2\right) \tag{A.3}$$

where ω and σ^{-1} denote the elasticity of real marginal cost with respect to own output and aggregate output, respectively, and $g_t\equiv-\frac{u_{c\xi}\tilde{\xi}_t}{\overline{Y}u_{cc}}$ and $q_t=-\frac{v_{y\xi}\tilde{\xi}_t}{\overline{Y}v_{yy}}$ denote the percentage variation in output required to keep the marginal utility of expenditure u_c and the marginal disutility of supply v_y at its steady-state level, given the exogenous disturbances in the economy.

Log-linearizing (A.1) around the steady state $\xi_t=0$, $A_t=1$, $P_t(i)=P_t=\overline{P}$, $Y_t=\overline{Y}$ yields the pricing rule

$$p_t^i=\mathbb{E}_t\left[p_t+\hat{s}_t(i)\mid\mathcal{I}_t^i\right]$$

where $\hat{s}_t(i)$ denotes the log-deviation of real marginal cost from its steady state. As shown in

[1] I make use of the fact that $(1+\tau)*(1-\eta)=-\eta$ and that

$$H_t(i)=f^{-1}\left(\frac{Y_t(i)}{A_t}\right)$$

And so

$$\frac{\partial H_t(i)}{\partial P_t(i)}=-\eta\frac{1}{f'\left(f^{-1}\left(\frac{Y_t(i)}{A_t}\right)\right)}*\frac{1}{A_t}*\frac{1}{P_t}\left(\frac{P_t(i)}{P_t}\right)^{-\eta-1}$$

Woodford (2003b), a log-linear approximation of (A.2) yields

$$\hat{s}_t(i) = \omega \widehat{Y}_t(i) + \sigma^{-1}\widehat{Y}_t - (\omega + \sigma^{-1})\widehat{Y}_t^N$$

where $\widehat{Y}_t(i) = \log\left(\frac{Y_t(i)}{\overline{Y}}\right)$, $\widehat{Y}_t = \log\left(\frac{Y_t}{\overline{Y}}\right)$, $\widehat{Y}_t^N = \log\left(\frac{Y_t^N}{\overline{Y}}\right)$. Noting that $\widehat{Y}_t(i) = \widehat{Y}_t - \eta\left[p_t^i - p_t\right]$

$$p_t^i = \mathbb{E}_t\left[p_t + \zeta\left(\widehat{Y}_t - \widehat{Y}_t^N\right) \mid \mathcal{I}_t^i\right]$$

where $\zeta = \frac{\omega+\sigma^{-1}}{1+\omega\eta}$ captures the degree of strategic complementarities between price-setters in the model (I assume that $\zeta \in (0,1)$ so that prices are strategic complements).

A.2 Welfare Approximation

The first part of the welfare analysis follows Woodford (2003b). Defining the indirect (dis)utility of labor as

$$\widetilde{v}(y;\widetilde{\xi}) \equiv v\left(f^{-1}(y/A);\xi\right)$$

the period utility of the representative household can be written as

$$U_t = \widetilde{u}(Y_t;\widetilde{\xi}_t) - \int_0^1 \widetilde{v}(Y_t(i);\widetilde{\xi}_t)di \tag{A.4}$$

we can rewrite the real marginal cost given by (A.2) as

$$s\left(y, Y;\widetilde{\xi}\right) = \frac{v_y(y;\widetilde{\xi})}{u_c(Y;\widetilde{\xi})}$$

The denominator can be approximated around the deterministic steady state $\xi = 0$,$A_t = 1$, $Y_t = \overline{Y}$ as

$$u_c(Y;\widetilde{\xi}) = \overline{u} + u_c\widetilde{Y}_t + u_\xi\widetilde{\xi}_t + \frac{1}{2}u_{cc}\widetilde{Y}_t^2 + U_{c\xi}\widetilde{\xi}_t\widetilde{Y}_t + \frac{1}{2}\widetilde{\xi}_t' u_{\xi\xi}\widetilde{\xi}_t + \mathcal{O}\left(\| \widehat{Y}, \tilde{\xi} \|^3\right)$$

where $\overline{u} \equiv u(\overline{Y};0)$ and $\widetilde{Y}_t \equiv Y_t - \overline{Y}$. Substituting $\widehat{Y}_t$ in for $\widetilde{Y}_t$ using the Taylor series expansion

$$Y_t/\overline{Y} = 1 + \widehat{Y}_t + \frac{1}{2}\widehat{Y}_t^2 + \mathcal{O}\left(\| \widehat{Y} \|^3\right)$$

and collecting all terms that are independent of policy, leads to

$$u_c(Y;\widetilde{\xi}) = \overline{Y}u_c\widehat{Y}_t + \frac{1}{2}\left[\overline{Y}u_c + \overline{Y}^2 u_{cc}\right]\widehat{Y}_t^2 - \overline{Y}^2 u_{cc} g_t \widehat{Y}_t + t.i.p + \mathcal{O}\left(\| \widehat{Y}, \tilde{\xi} \|^3\right)$$

$$= \overline{Y} u_c \left[\widehat{Y}_t + \frac{1}{2}(1-\sigma^{-1})\widehat{Y}_t^2 + \sigma^{-1} g_t \widehat{Y}_t \right] + t.i.p + \mathcal{O}\left(\| \widehat{Y}, \tilde{\xi} \|^3 \right)$$

by noting that $\sigma^{-1} = -\frac{\overline{Y} u_{cc}}{u_c}$. We can similarly approximate the indirect disutility of labor function as[2]

$$\widetilde{v}(Y_t(i); \widetilde{\xi}) = \overline{Y} u_c \left[\widehat{Y}_t(i) + \frac{1}{2}(1+\omega)\widehat{Y}_t(i)^2 - \omega q_t \widehat{Y}_t(i) \right]$$

Integrating over the differentiated goods i yields

$$\int_0^1 \widetilde{v}(Y_t(i); \widetilde{\xi}) di = \overline{Y} u_c \left[\mathbb{E}_i \widehat{Y}_t(i) + \frac{1}{2}(1+\omega) \left[\left(\mathbb{E}_i \widehat{Y}_t(i) \right)^2 + \mathbb{V}\mathrm{ar}_i \widehat{Y}_t(i) \right] - \omega q_t \mathbb{E}_i \widehat{Y}_t(i) \right] + t.i.p + \mathcal{O}\left(\| \widehat{Y}, \tilde{\xi} \|^3 \right)$$

where $\mathbb{E}_i$ and $\mathbb{V}ar_i$ denote the cross-sectional mean and variance across firms at a given time. Noting that $Y_t \equiv \left[\int_0^1 Y_t(i)^{(\eta-1)/\eta} di \right]^{\eta/(\eta-1)}$ and taking a Taylor approximation leading to

$$\widehat{Y}_t = \mathbb{E}_i \widehat{Y}_t(i) + \frac{1}{2}(1-\eta^{-1}) \mathbb{V}\mathrm{ar}_i \widehat{Y}_t(i) + \mathcal{O}\left(\| \widehat{Y}, \tilde{\xi} \|^3 \right)$$

and plugging in for $\mathbb{E}_i \widehat{Y}_t(i)$ yields

$$\int_0^1 \widetilde{v}(Y_t(i); \widetilde{\xi}) di = \overline{Y} u_c \left[(1-\omega q_t)\widehat{Y}_t + \frac{1}{2}(1+\omega)\widehat{Y}_t^2 + \frac{1}{2}\left[\eta^{-1} + \omega \right] \mathbb{V}\mathrm{ar}_i \widehat{Y}_t(i) \right] + t.i.p + \mathcal{O}\left(\| \widehat{Y}, \tilde{\xi} \|^3 \right)$$

Putting everything together, it follows that a second-order approximation of (A.4) yields

$$U_t = \overline{Y} u_c \left[(\sigma^{-1} g_t + \omega q_t)\widehat{Y}_t - \frac{1}{2}(\sigma^{-1} + \omega)\widehat{Y}_t^2 - \frac{1}{2}(\epsilon^{-1} + \omega) \mathbb{V}ar_i \widehat{Y}_t(i) \right] + t.i.p + \mathcal{O}\left(\| \widehat{Y}, \tilde{\xi} \|^3 \right)$$

Using the expression for the natural rate given by (A.3) leads to

$$U_t = -\frac{\overline{Y} u_c}{2} \left[(\sigma^{-1} + \omega)\widetilde{y}_t^2 + (\eta^{-1} + \omega) \mathbb{V}ar_i \widehat{Y}_t(i) \right] + t.i.p + \mathcal{O}\left(\| \widehat{Y}, \tilde{\xi} \|^3 \right)$$

where $\widetilde{y}_t^2 \equiv \widehat{Y}_t - \widehat{Y}_t^N$ denotes the output gap. Noting from (1.5) that

$$\widehat{Y}_t(i) = \log Y_t - \eta \log P_t(i) + \eta \log P_t - \log \overline{Y}_t$$

and thus the cross sectional-variance is simply

$$\mathbb{V}ar_i \widehat{Y}_t(i) = \eta^2 \mathbb{V}ar_i p_t(i)$$

[2] In this model note that τ is such that the overall distortion in the steady-state output level is 0; in the notation of Woodford (2003b), $\Phi_y = 0$ and it follows that $\overline{Y}/Y^* = 0$ and $u_c = v_y$ in the deterministic steady state.

and so

$$U_t = -\frac{\overline{Y}u_c}{2}\left[(\sigma^{-1}+\omega)\widetilde{y}_t^2 + (\eta+\omega\eta^2)\mathbb{V}ar_i p_t(i)\right] + t.i.p + \mathcal{O}\left(\parallel \widehat{Y}, \tilde{\xi} \parallel^3\right)$$

$$= -K\left[\zeta\widetilde{y}_t^2 + \eta\mathbb{V}ar_i p_t(i)\right] + t.i.p + \mathcal{O}\left(\parallel \widehat{Y}, \tilde{\xi} \parallel^3\right)$$

where $K = \frac{\overline{Y}u_c(1+\omega\eta)}{2}$. It follows that (after scaling and discarding higher-order terms) the welfare component affected by monetary policy is quadratic in the output gap and the variance of prices:

$$\mathcal{U}_t = \widetilde{y}_t^2 + \kappa\mathcal{V}(p_t^i)$$

where $\kappa = \eta/\zeta$.

Without Calvo Frictions

Without Calvo Frictions, we have that

$$\mathcal{V}(p_t^i) = \int_0^1 \left(p_t^i - p_t\right)^2 di = \gamma_{2t}^2 \int (S_t^i - S_t)^2 di = \gamma_{2t}^2 \left(\tau_F\right)^{-1}$$

And thus

$$\mathcal{U}_t = -\left[\widetilde{y}_t^2 + \frac{\eta}{\zeta\tau_F}\gamma_{2t}^2\right]$$

With Calvo Frictions

With Calvo frictions, a fraction ϕ of firms cannot change their price in a given period. We have

$$p_t - p_{t-1} = \mathbb{E}_i p_t^i - p_{t-1}$$

$$= \phi p_{t-1} + (1-\phi)p_t^* - p_{t-1}$$

$$= (1-\phi)(p_t^* - p_{t-1})$$

where p_t^* denotes the mean of the log prices of firms that change their prices in period t. The total dispersion of prices can be written as

$$\mathcal{V}(p_t^i) = \mathbb{V}ar_i\left[p_t^i - p_{t-1}\right]$$

$$= \mathbb{E}_i\left[\left(p_t^i - p_{t-1}\right)^2\right] - \left[\mathbb{E}_i\left(p_t^i - p_{t-1}\right)\right]^2$$

$$= \phi \mathbb{E}_i \left[\left(p^i_{t-1} - p_{t-1} \right)^2 \right] + (1-\phi) \mathbb{E}_i \left[\left(p^{i*}_t - p_{t-1} \right)^2 \right] - \left[\mathbb{E}_i \left(p^i_t - p_{t-1} \right) \right]^2$$

$$= \phi \mathbb{E}_i \left[\left(p^i_{t-1} - p_{t-1} \right)^2 \right] + (1-\phi) \left[\triangle_t + (p^*_t - p_{t-1})^2 \right] - \left[\mathbb{E}_i \left(p^i_t - p_{t-1} \right) \right]^2$$

$$= \phi \mathcal{V}(p^i_{t-1}) + (1-\phi) \triangle_t + \frac{\phi}{1-\phi} \left[\mathbb{E}_i \left(p^i_t - p_{t-1} \right) \right]^2$$

$$= \phi \mathcal{V}(p^i_{t-1}) + \frac{\phi}{1-\phi} \pi^2_t + (1-\phi) \widehat{\gamma}^2_{2t} \left(\tau_F \right)^{-1}$$

where $\triangle_t = \widehat{\gamma}^2_{2t} \left(\tau_F \right)^{-1}$ denotes the dispersion (variance) of prices among firms change their prices, and using the linearity of means in line 3 and the properties of mean squared errors in line 4. Putting everything together, it follows that

$$\mathcal{U}_t = - \left[\widetilde{y}^2_t + \frac{\eta}{\zeta} \left[\phi \mathcal{V}(p^i_{t-1}) + \frac{\phi}{1-\phi} \pi^2_t + (1-\phi) \widehat{\gamma}^2_{2t} \left(\tau_F \right)^{-1} \right] \right]$$

$$= - \left[\widetilde{y}^2_t + \kappa \frac{\phi}{1-\phi} \pi^2_t + \frac{(1-\phi)\kappa}{\tau_F} \widehat{\gamma}^2_{2t} + \kappa \phi \mathcal{V}(p^i_{t-1}) \right]$$

where $\kappa = \eta / \zeta$.

A.3 Higher-Order Beliefs

The firm's optimal price depends on higher-order beliefs, as shown in (1.9). Supposing that each central bank type θ optimally plays a linear strategy $q^\theta_t = a^\theta_t \left(\mathcal{B}_t \right) S^\theta_t$ (which is does in equilibrium, as shown in (Proposition 1)), it is useful to define

$$\Omega_t = \left[\mathcal{B}_t \frac{\left(\tau^H / a^H_t \right)}{\tau^H + \tau_F} + (1-\mathcal{B}_t) \frac{\left(\tau^L / a^L_t \right)}{\tau^L + \tau_F} \right]$$

Given this definition, the following Lemma holds:

Lemma 3. *Higher order beliefs are such that*

$$\overline{\mathbb{E}}^k_t \left(y^N_t \right) = q_t \left[\sum_{j=0}^{k-1} \Omega_t \left(\mathcal{H}^{-1}_t \tau_F \right)^j \right] + y^N_t \left(\mathcal{H}^{-1}_t \tau_F \right)^k$$

$$\mathbb{E}_t \left(\overline{\mathbb{E}}^k_t \left(y^N_t \right) \mid \mathcal{I}^i_t \right) = q_t \left[\sum_{j=0}^{k} \Omega_t \left(\mathcal{H}^{-1}_t \tau_F \right)^j \right] + S^i_t \left(\mathcal{H}^{-1}_t \tau_F \right)^{k+1}$$

Proof. Proof is by induction on k. First consider $k = 1$. In this case, by the Law of Iterated

Expectations,

$$\mathbb{E}\left(y_t^N \mid \mathcal{I}_t^i\right) = \mathcal{B}_t \frac{\tau^H\left(\frac{q_t}{a^H}\right) + \tau_F S_t^i}{\tau^H + \tau_F} + (1 - \mathcal{B}_t^I)\frac{\tau^H\left(\frac{q_t}{a^B}\right) + \tau_F S_t^i}{\tau^H + \tau_F} = \Omega_t q_t + \mathcal{H}_t^{-1}\tau_F S_t^i$$

and so

$$\overline{\mathbb{E}}\left(y_t^N\right) = \Omega_t q_t + \mathcal{H}_t^{-1}\tau_F y_t^N$$

$$\mathbb{E}_t\left(\overline{\mathbb{E}}\left(y_t^N\right) \mid \mathcal{I}_t^i\right) = \Omega_t q_t + \mathcal{H}_t^{-1}\tau_F\left[\Omega_t q_t + \mathcal{H}_t^{-1}\tau_F S_t^i\right]$$

$$= q_t \sum_{j=0}^{1} \Omega_t\left(\mathcal{H}_t^{-1}\tau_F\right)^j + S_t^i\left(\mathcal{H}_t^{-1}\tau_F\right)^2$$

So this holds for $k = 1$. Now suppose this holds for k. We have

$$\mathbb{E}_t\left(\overline{\mathbb{E}}_t^k\left(y_t^N\right) \mid \mathcal{I}_t^i\right) = q_t\left[\sum_{j=0}^{k} \Omega_t\left(\mathcal{H}_t^{-1}\tau_F\right)^j\right] + S_t^i\left(\mathcal{H}_t^{-1}\tau_F\right)^{k+1}$$

by assumption, and so

$$\overline{\mathbb{E}}_t^{k+1}\left(y_t^N\right) = q_t\left[\sum_{j=0}^{k} \Omega_t\left(\mathcal{H}_t^{-1}\tau_F\right)^j\right] + y_t^N\left(\mathcal{H}_t^{-1}\tau_F\right)^{k+1}$$

by the Law of Large Numbers, and

$$\mathbb{E}_t\left(\overline{\mathbb{E}}_t^{k+1}\left(y_t^N\right) \mid \mathcal{I}_t^i\right) = q_t\left[\sum_{j=0}^{k} \Omega_t\left(\mathcal{H}_t^{-1}\tau_F\right)^j\right] + \left[\Omega_t q_t + \mathcal{H}_t^{-1}\tau_F S_t^i\right]\left(\mathcal{H}_t^{-1}\tau_F\right)^{k+1}$$

$$= q_t\left[\sum_{j=0}^{k+1} \Omega_t\left(\mathcal{H}_t^{-1}\tau_F\right)^j\right] + S_t^i\left(\mathcal{H}_t^{-1}\tau_F\right)^{k+2}$$

and so this is true for $k + 1$. Thus by induction, this Lemma holds. □

Figure A.1 illustrates the role that beliefs about the central bank play in determining higher-order beliefs about the underlying shocks. The figure plots the relative weight that firms put on the public signal,[3] for varying levels of higher-order beliefs and different beliefs about the central bank's competence. Note that each hierarchy of belief depends solely on the reputation (belief) about the central bank's type, and not the actual precision of the public signal itself; this is

[3]This is defined by

$$\frac{\sum_{j=0}^{k} \Omega_t\left(\mathcal{H}_t^{-1}\tau_F\right)^j}{\sum_{j=0}^{k} \Omega_t\left(\mathcal{H}_t^{-1}\tau_F\right)^j + \left(\mathcal{H}_t^{-1}\tau_F\right)^{k+1}}$$

for the simplest possible cast when $a_t^H = a_t^L = 1$.

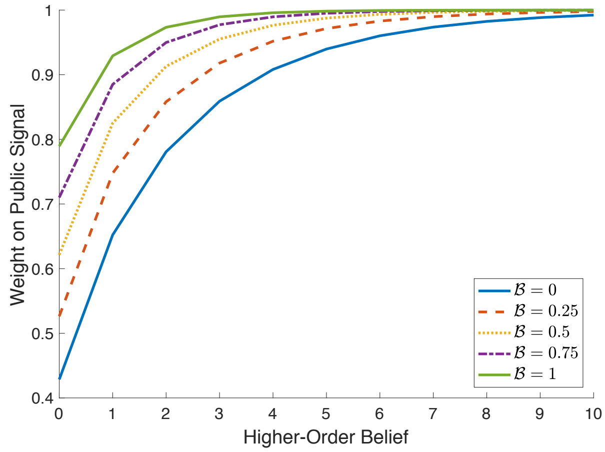


Figure A.1: Higher order beliefs given different beliefs about central bank competence

a fundamental departure from the existing literature, where the precision of signals is perfectly known. In this limit case, each higher-order belief would be represented by the two solid lines in Figure A.1, representing $\mathcal{B} = 0$ and $\mathcal{B} = 1$ (corresponding to τ^L and τ^H, respectively). Note that the weight placed on the public signal is always bounded below and above by these cases; when the true type is τ^H, the presence of type uncertainty reduces the weight placed on the public signal, and vice versa for when the true type is τ^L, for each higher-order belief. In all cases, there is convergence in the higher-order beliefs, as each successive higher-order puts increasingly more weight on the public signal, to the common limit of all weight on the public signal; unsurprisingly, convergence happens more quickly for higher levels of $\mathcal{B}$. These higher-order beliefs, successively discounted by a factor of $(1 - \zeta)$, lead to the public signal q_t receiving higher than Bayesian weight in the firm's optimal pricing condition, as seen in (1.9).

A.4 Markov Dynamics and Convergence to Stochastic Steady State

Letting $\mathbf{S} = (\theta, \mathcal{B})$ denote the random state vector of the true type and beliefs and denoting $\mathbb{S} = \Theta \times [0, 1]$ as the state space, first note that

$$\mathbf{S}_{t+1} = H(\mathbf{S}_t, W_t \mid \mathbf{S}_t^1, V_t)$$

for some function H, $\mathbf{S}^1$ denotes the first element of $\mathbf{S}$, and where $W_t \sim N\left(0, \left(\tau^{\theta_t}\right)^{-1}\right)$ and $V_t \sim^{IID} \text{Bern}(\lambda)$. It is clear that the system is Markov process on $\mathbb{S}$, with stochastic kernel $P(x, y)$

for $x, y \in \mathbb{S}$ defined as

$$P(x,y) = \Big[(1-\lambda)\mathbb{1}\{y^1 = x^1\} + \lambda \mathbb{1}\{y^1 \neq x^1\}\Big] h_{x^1}(y^2 \mid x^2)$$

where $h_{x^1}(y^2 \mid x^2)$ is a conditional density function that depends on the true state x^1 according to which the Gaussian shock is drawn.[4] Denoting by ψ_t as any initial distribution on $\mathbb{S}$, the sequence$(\psi_{t+j})_{j\geq 0}$ satisfies

$$\psi_{t+1}(y) = \int_{\mathbb{S}} \psi_t(dx) P(x,y)$$

where we integrate with respect to the measure ψ_t, and defining ψ_{t+j} for any j analogously in terms of ψ_{t+j-1}.

Convergence

Defining the measure space associated with the true type θ as $(\mathbb{S}^1, \mathscr{B}^1)$ where $\mathbb{S}^1 = \{H, L\}$ and $\mathscr{B}^1$ denotes all subsets of $\mathbb{S}^1$, and likewise defining the measure space associated with beliefs as $(\mathbb{S}^2, \mathscr{B}^2)$ where $\mathbb{S}^2 = [0,1]$ and $\mathscr{B}^2$ is the Borel σ-algebra on $[0,1]$. we can then define the measure space $(\mathbb{S}, \mathscr{B})$ for the entire state space, where $\mathbb{S} = \mathbb{S}^1 \times \mathbb{S}^2$, $\mathscr{B} = \{A \times B;\ A \in \mathscr{B}^1, B \in \mathscr{B}^2\}$. Note that for any probability measures $\mathscr{P}^1$ and $\mathscr{P}^2$ on $(\mathbb{S}^1, \mathscr{B}^1)$ and $(\mathbb{S}^2, \mathscr{B}^2)$ respectively, we can define the probability space $(\mathbb{S}, \mathscr{B}, \mathscr{P})$, where $\mathscr{P}$ is the product measure $\mathscr{P} = \mathscr{P}^1 \times \mathscr{P}^2$.

The relevant stochastic kernel $P(x, \cdot)$ for any $x \in \mathbb{S}$ is constructed as follows. First defining $P^1(x^1, \cdot)$ as 1 for $\mathbb{S}^1$, $(1-\lambda)$ for x^1, λ for $\mathbb{S}^1 \setminus x^1$ and 0 for $\emptyset$ and defining $P^2(x^2, \cdot)$ as the Borel measure associated with the density of $\mathcal{B}'(x^2, \varepsilon)$ where ε is drawn from $N\left(0, \left(\tau^{x^1}\right)^{-1}\right)$, we can then define $P(x, \cdot)$ as the product measure of P^1 and P^2 . It easily verified that $P(x, \cdot)$ is a probability measure on $(\mathbb{S}, \mathscr{B})$ for all $x \in \mathbb{S}$, and $P(\cdot, \mathcal{A})$ is a measurable function on $\mathbb{S}$ for any $\mathcal{A} \in \mathscr{B}$, and so this construction yields a valid stochastic kernel. Note that this is equivalent to the the stochastic

[4]Note that because $\mathcal{B}'(\mathcal{B}, \varepsilon) = \mathcal{B}'(\mathcal{B}, \mid \varepsilon \mid)$ is differentiable and monotonically decreasing in $\mid \varepsilon \mid$ for $\mid \varepsilon \mid \in (0, \infty)$, then this density is given by

$$h_{x^1}(y^2 \mid x^2) = f_{x^1}\left[\mathcal{B}'^{-1}(\mathcal{B}, y^2)\right] \cdot \left| \frac{\partial \mathcal{B}'^{-1}(\mathcal{B}, y^2)}{\partial y^2} \right|$$

defining $\mathcal{B}'^{-1}(\mathcal{B}, \cdot)$ as the inverse of $\mathcal{B}'$ and where f_{x^1} is the density of the folded normal distribution with parameters $(\mu, \sigma^2) = \left(0, \left(\tau^{x^1}\right)^{-1}\right)$.

kernel laid out directly above, and captures the dynamics of the model.

Now let φ_1 be any probability measure on $\mathbb{S}^1$, and let φ_2 be the Lebesgue measure restricted to the open set (a, b) such that $a > \lambda$ and $b < K$ for some $\lambda < K < 1 - \lambda$. Defining φ as the product measure $\varphi_1 \times \varphi_2$ it is readily shown that for every $x \in \mathbb{S}$, we have that $\varphi(\mathcal{A}) > 0 \Rightarrow P_x(\tau_{\mathcal{A}} < \infty) > 0$ where $\tau_{\mathcal{A}} = \inf\{n \geq 0; \mathbf{S}_n \in \mathcal{A})$ is the first hitting time of the subset $\mathcal{A}$ when starting from x. This implies that the stochastic process $\{\mathbf{S}_{t+j}\}_{j=0}^{\infty}$ is φ-irreducible, as defined in Meyn and Tweedie (2012), and thus it follows that there exists a maximal irreducibility measure ϕ such that $\{\mathbf{S}_{t+j}\}_{j=0}^{\infty}$ is ϕ-irreducible. Letting $\mathscr{B}^+ \equiv \{\mathcal{A} \in \mathscr{B} : \phi(\mathcal{A}) > 0\}$, and noting that if $\varphi(\mathcal{A}) > 0$ then $\phi(\mathcal{A}) > 0$ for any φ such that $\{\mathbf{S}_{t+j}\}_{j=0}^{\infty}$ is φ-irreducible, it is straightforward to show that for every $x \in \mathbb{S}$ and $\mathcal{A} \in \mathscr{B}^+$, we have $\sum_{n=1}^{\infty} P^n(x, A) = \infty$.[5] It thus follows that $\{\mathbf{S}_{t+j}\}_{j=0}^{\infty}$ is recurrent, which by Theorem 10.0.1 in Meyn and Tweedie (2012) implies that there exists a unique invariant measure ψ, and furthermore for every $x \in \mathbb{S}$ we have

$$\lim_{n\to\infty} \sup_{\mathcal{A}\in\mathscr{B}} \Big| \, P^n(x, \mathcal{A}) - \psi^*(\mathcal{A}) \, \Big| \to 0$$

and

$$\frac{1}{n}\sum_{t=1}^{n} \mathcal{U}\Big(\gamma(\mathcal{B}_t), \theta_t\Big) \to \int_{\mathbb{S}} \psi^*(dx)\mathcal{U}\Big(\gamma(x^2), x^1\Big) \text{ as } n \to \infty$$

with probability one.

A.5 Decentralized Information Optimum

First note that for any linear pricing coefficients γ_{1t} and γ_{2t}, the output gap can be written as

$$\widetilde{y}_t = (1 - \gamma_{1t})q_t - (1 + \gamma_{2t})y_t^N$$

Because $q_t = S_t^\theta = y_t^N + \varepsilon_t^\theta$ it follows that

$$\mathbb{E}_t \left[\widetilde{y}_t^2 \mid \mathcal{I}_t^{CB}\right] = (1 - \gamma_{1t})^2 \left(S_t^\theta\right)^2 + (1 + \gamma_{2t})\mathbb{E}_t^2 \left[\left(y_t^N\right)^2 \mid S_t^\theta\right] - 2(1 - \gamma_{1t})(1 + \gamma_{2t})S_t^\theta \mathbb{E}_t \left[y_t^N \mid S_t^\theta\right]$$

Because $\mathbb{E}_t \left[y_t^N \mid S_t^\theta\right] = S_t^\theta$ and $\mathbb{E}_t \left[\left(y_t^N\right)^2 \mid S_t^\theta\right] = \left(S_t^\theta\right)^2 + \left(\tau^\theta\right)^{-1}$ by the diffuse prior assumption,

[5] Heuristically, $\mathcal{B}$ does not drift unboundedly; thus, if the probability of reaching $\mathcal{A}$ in finite time is greater than zero, this will happen infinitely often as $n \to \infty$

$$\mathbb{E}_t\left[\widetilde{y}_t^2 \mid \mathcal{I}_t^{CB}\right] = (1-\gamma_{1t})^2\left(S_t^\theta\right)^2 + (1+\gamma_{2t})^2\left[\left(S_t^\theta\right)^2 + \left(\tau^\theta\right)^{-1}\right] - 2(1-\gamma_{1t})(1+\gamma_{2t})\left(S_t^\theta\right)^2$$

Adding the price dispersion term, we thus have

$$\mathbb{E}_t\left[\mathcal{U}_t \mid S_t^\theta\right] = -\left\{(1-\gamma_{1t})^2\left(S_t^\theta\right)^2 + (1+\gamma_{2t})^2\left[\left(S_t^\theta\right)^2 + \left(\tau^\theta\right)^{-1}\right] - 2\left(1-\gamma_{1t}\right)\left(1+\gamma_{2t}\right)\left(S_t^\theta\right)^2 + \frac{\epsilon}{\zeta\tau_F}\gamma_{2t}^2\right\}$$

Taking FOC with respect to γ_{1t} yields

$$-\left[-2(1-\gamma_{1t})\left(S_t^\theta\right)^2 + 2(1+\gamma_{2t})\left(S_t^\theta\right)^2\right] = 0$$

$$\Rightarrow \gamma_{1t}^* + \gamma_{2t}^* = 0$$

Taking FOC with respect to γ_{2t} yields

$$-\left[2(1+\gamma_{2t})\left[\left(S_t^\theta\right)^2 + \left(\tau^\theta\right)^{-1}\right] - 2(1-\gamma_{1t})\left(S_t^\theta\right)^2 + 2\frac{\eta}{\zeta\tau_F}\gamma_{2t}\right] = 0$$

$$\Rightarrow (1+\gamma_{2t})\left[\left(S_t^\theta\right)^2 + \left(\tau^\theta\right)^{-1}\right] - (1-\gamma_{1t})\left(S_t^\theta\right)^2 + \frac{\eta}{\zeta\tau_F}\gamma_{2t} = 0$$

Plugging in $\gamma_{1t}^* + \gamma_{2t}^* = 0$ yields

$$(1-\gamma_{1t}^*)\frac{1}{\tau^\theta} + \gamma_{2t}^*\frac{\eta}{\zeta\tau_F} = 0$$

It was shown earlier that the equilibrium coefficients satisfy

$$\gamma_{1t}^{EQ} + \gamma_{2t}^{EQ} = 0$$

In addition, one can show that

$$\frac{(1-\gamma_{1t})}{\gamma_{2t}} = -\Omega_t\frac{\mathcal{H}_t}{\zeta\tau_F}$$

and since

$$\Omega_t\mathcal{H}_t = \left[\frac{\mathcal{B}_t\tau_H\left(\tau^L+\tau_F\right) + \tau_L(1-\mathcal{B}_t)(\tau^H+\tau_F)}{\left(\tau^H+\tau_F\right)\left(\tau^L+\tau_F\right)}\right]\left[\frac{\mathcal{B}_t(\tau^L+\tau_F) + (1-\mathcal{B}_t)(\tau^H+\tau_F)}{(\tau^H+\tau_F)(\tau^L+\tau_F)}\right]^{-1}$$

$$= \frac{\mathcal{B}_t\tau_H\left(\tau^L+\tau_F\right) + \tau_L(1-\mathcal{B}_t)(\tau^H+\tau_F)}{\mathcal{B}_t(\tau^L+\tau_F) + (1-\mathcal{B}_t)(\tau^H+\tau_F)}$$

$$= \frac{(\tau^H+\tau_F)(\tau^L+\tau_F) - \tau_F\left[\mathcal{B}_t(\tau^L+\tau_F) + (1-\mathcal{B}_t)(\tau^H+\tau_F)\right]}{\mathcal{B}_t(\tau^L+\tau_F) + (1-\mathcal{B}_t)(\tau^H+\tau_F)}$$

$$= \mathcal{H}_t - \tau_F$$

It follows that

$$\frac{(1-\gamma_{1t}^{EQ})}{\gamma_{2t}^{EQ}} = -\frac{\mathcal{H}_t - \tau_F}{\zeta\tau_F}$$

$$\Rightarrow (1-\gamma_{1t}^{EQ})\frac{1}{\mathcal{H}_t - \tau_F} + \gamma_{2t}^{EQ}\frac{1}{\zeta\tau_F} = 0$$

Finally, to show that

$$\mid \gamma_{2t}^{EQ} \mid = \left(1-\gamma_{1t}^{EQ}\right)\frac{\zeta\tau_F}{\mathcal{H}_t - \tau_F} > \left(1-\gamma_{1t}^{EQ}\right)\frac{\zeta\tau_F}{\eta\tau^\theta}$$

Note that since $(1-\gamma_{1t}^{EQ}) > 0$, it suffices to show that $\eta\tau^\theta > \mathcal{H}_t - \tau_F$. The LHS attains a minimum at $\eta\tau^L$ whereas the RHS attains a maximum when $\mathcal{B}_t = 1$, as $\mathcal{H}_t$ is increasing in $\mathcal{B}_t$. $\mathcal{B}_t = 1$ implies $\mathcal{H}_t = \tau_G + \tau_F$, and so the RHS attains a maximum at τ_G. It immediately follows that if $\eta > \tau^G/\tau^B$, the inequality is always satisfied.

A.6 Proofs of Propositions

Proposition 1

Proof. First suppose that each central bank type plays $q_t^\theta = S_t^\theta$ in equilibrium; I then show that firms set prices according to the equilibrium coefficients in Proposition 1 and that neither central bank type θ has an incentive to deviate from the proposed strategy.

The optimal price can be written as

$$p_t^i = \sum_{k=0}^{\infty}(1-\zeta)^k \mathbb{E}_t\left[\overline{\mathbb{E}}^k\left[\zeta\left(q_t - y_t^N\right)\right]\middle|\mathcal{I}^i\right]$$

Plugging in from Lemma 3, we have

$$p_t^i = \zeta\sum_{k=0}^{\infty}(1-\zeta)^k\left[q_t\left[1-\sum_{j=0}^{k}\Omega_t\left(\mathcal{H}_t^{-1}\tau_F\right)^j\right] - S_t^i\left(\mathcal{H}_t^{-1}\tau_F\right)^{k+1}\right]$$

$$= \zeta\frac{1}{1-(1-\zeta)}q_t - \zeta\sum_{k=0}^{\infty}(1-\zeta)^k\Omega_t q_t\sum_{j=0}^{k}\left(\mathcal{H}_t^{-1}\tau_F\right)^j - \zeta\left[\frac{\mathcal{H}_t^{-1}\tau_F}{1-(1-\zeta)(\mathcal{H}_t^{-1}\tau_F)}\right]S_t^i$$

Noting that

$$\sum_{k=0}^{\infty}(1-\zeta)^k\sum_{j=0}^{k}\left(\mathcal{H}_t^{-1}\tau_F\right)^j = \sum_{k=0}^{\infty}(1-\zeta)^k + \left(\mathcal{H}_t^{-1}\tau_F\right)\sum_{k=1}^{\infty}(1-\zeta)^k + \left(\mathcal{H}_t^{-1}\tau_F\right)^2\sum_{k=2}^{\infty}(1-\zeta)^k + \cdots$$

$$= \frac{1}{\zeta}\sum_{k=0}^{\infty}\left(\mathcal{H}_t^{-1}\tau_F\right)^k(1-\zeta)^k = \frac{1}{\zeta}\frac{1}{1-(1-\zeta)\mathcal{H}_t^{-1}\tau_F}$$

And so plugging this in, we have

$$p_t^i = q_t - q_t\Omega_t\left[\frac{1}{1-(1-\zeta)\mathcal{H}_t^{-1}\tau_F}\right] - \left[\frac{\zeta\tau_F}{\mathcal{H}_t - \mathcal{H}_t(1-\zeta)(\mathcal{H}_t^{-1}\tau_F)}\right]S_t^i$$

$$= \left[1-\Omega_t\frac{\mathcal{H}_t}{\mathcal{H}_t-(1-\zeta)\tau_F}\right]q_t - \left[\frac{\zeta\tau_F}{\mathcal{H}_t-(1-\zeta)\tau_F}\right]S_t^i$$

Under the equilibrium strategies $q_t^\theta = S_t^\theta$, $\Omega_t = \left[\mathcal{B}_t\frac{\tau^H}{\tau^H+\tau_F} + (1-\mathcal{B}_t)\frac{\tau^L}{\tau^L+\tau_F}\right]$, and so

$$p_t^i = \left[1-\frac{\mathcal{H}_t-\tau_F}{\mathcal{H}_t-(1-\zeta)\tau_F}\right]q_t - \left[\frac{\zeta\tau_F}{\mathcal{H}_t-(1-\zeta)\tau_F}\right]S_t^i$$

$$= \gamma_{1t}q_t + \gamma_{2t}S_t^i$$

as desired. I now analyze the actions of a central bank of type θ_t, given beliefs $\mathcal{B}_t$ and having received signal S_t^θ. I show that given this proposed equilibrium, the optimal action that maximizes (1.11) is indeed $q_t^\theta = S_t^\theta$. To do so, I begin by analyzing the period welfare function. Given the proposed equilibrium above, we can write the period (monetary) welfare in all future possible states

$$\mathcal{U}(\gamma(\mathcal{B}),\theta) = -\left[(1-\gamma(\mathcal{B}))^2\frac{1}{\tau^\theta} + \frac{\epsilon}{\zeta\tau_F}\gamma(\mathcal{B})^2\right]$$

where

$$\gamma(\mathcal{B}) = \frac{\tau_F\zeta}{\left[\frac{\mathcal{B}}{\tau^H+\tau_F}+\frac{1-\mathcal{B}}{\tau^L+\tau_F}\right]^{-1} - (1-\zeta)\tau_F}$$

It is readily shown that $\mathcal{U}_t$ is decreasing in $\gamma(\mathcal{B})$ which is itself decreasing in $\mathcal{B}$. Thus, it follows that per-period (monetary) welfare is increasing in the probability that private agents assign to the central bank being type H. Now note that for any possible sequence of true types $\left\{\theta_{t+j}\right\}_{j=1}^\infty$ that at any given period $t+j$ the distribution of $\mathcal{B}_{t+j}$ when period $t+1$'s belief is $\mathcal{B}_{t+1}$ FOSD the distribution of $\mathcal{B}_{t+j}$ when period t's belief is $\widetilde{\mathcal{B}}_{t+1}$, when $\mathcal{B}_{t+1} > \widetilde{\mathcal{B}}_{t+1}$. It thus follows that under the proposed equilibrium (1.11) is increasing in $\mathcal{B}$, and that the (expected) continuation payoff in the infinite horizon welfare function is increasing in $\mathcal{B}_{t+1}$.

I now analyze the central bank's actions at period t. The perceived error by the central bank is

$$\varepsilon_t^P = q_t + \frac{p_t - \gamma_t q_t}{\gamma_t} = q_t - \left(S_t^\theta - \varepsilon_t^\theta\right)$$

Given S_t^θ and action q_t, the distribution of ε_t^P is thus $\mathcal{N}\left(q_t - S_t^\theta, \left(\tau^\theta\right)^{-1}\right)$. It thus follows immediately that for any $q_t \neq S_t^\theta$ the distribution of $\mid \varepsilon_t^P \mid$ first order stochastically dominates the distribution when $q_t = S_t^\theta$. Because $\mathcal{B}_{t+1}$ is decreasing in $\mid \varepsilon_t^P \mid$, putting everything together, this implies that under this proposed equilibrium the action $q_t = S_t^\theta$ maximizes the continuation payoff in the infinite horizon welfare function (1.11).

I now consider the current period welfare function. The price dispersion term is pre-determined, so it suffices to look at the output gap

$$\widetilde{y}_t = (1 - \gamma_{1t})q_t - (1 + \gamma_{2t})y_t^N$$

and so

$$\begin{aligned}\mathbb{E}_t\left[\widetilde{y}_t^2 \mid S_t^\theta\right] &= (1-\gamma_t)^2 q_t^2 + (1-\gamma_t)^2 \mathbb{E}_t\left[\left(y_t^N\right)^2 \mid S_t^\theta\right] - 2(1-\gamma_t)^2 q_t \mathbb{E}_t\left[y_t^N \mid S_t^\theta\right] \\ &= (1-\gamma_t) q_t^2 + (1-\gamma_t)^2\left[\left(S_t^\theta\right)^2 + \left(\tau^\theta\right)^{-1}\right] - 2(1-\gamma_t)^2 q_t S_t^\theta\end{aligned}$$

using the fact that $\mathbb{E}_t\left[y_t^N \mid S_t^\theta\right] = S_t^\theta$ and $\mathbb{E}_t\left[\left(y_t^N\right)^2 \mid S_t^\theta\right] = \left(S_t^\theta\right)^2 + \left(\tau^\theta\right)^{-1}$ by the diffuse prior assumption. Taking FOC w.r.t. to q_t yields

$$2(1-\gamma_t)^2 q_t - 2(1-\gamma_t)^2 S_t^\theta = 0$$

Implying that $q_t = S_t^\theta$. Thus $q_t = S_t^\theta$ maximizes the period welfare function. Thus given proposed equilibrium actions $q_t = S_t^\theta$ in all future periods, the action $q_t = S_t^\theta$ maximizes the value function, as it maximizes the period welfare as well as the continuation welfare. Putting everything together, this implies that the equilibrium outlined in Proposition 1 is indeed an equilibrium. □

Proposition 2

Proof. Begin by noting that$p_t = \gamma_{1t} q_t + \gamma_{2t} y_t^N$ by the Law of Large Numbers, and thus

$$\widetilde{y}_t = y_t - y_t^N = q_t - p_t - y_t^N$$

$$= (1-\gamma_{1t})q_t - (\gamma_{2t}+1)y_t^N$$

By (Proposition 1), $q_t = S_t^\theta = y_t^N + \varepsilon_t^\theta$ and $\gamma_{1t} = -\gamma_{2t} \equiv \gamma_t$. Thus,

$$\widetilde{y}_t = (1-\gamma_t)\,\varepsilon_t^\theta$$

as desired. In addition,

$$\mathbb{E}_t\left(\tilde{y}_t^2 \mid \theta_t, \mathcal{B}_t\right) = (1-\gamma_t)^2 \frac{1}{\tau^{\theta_t}}$$

and (i) and (ii) immediately follow. □

Proposition 4

Proof. Combining Proposition 2 and Proposition 3, and plugging this into the social welfare function (1.10) yields

$$\mathcal{U}(\gamma_t, \theta_t) = -\left[(1-\gamma_t)^2 \frac{1}{\tau^{\theta_t}} + \frac{\eta}{\zeta\tau_F}\gamma_t^2\right]$$

Taking the derivative with respect to $\mathcal{B}_t$ yields

$$\frac{\partial \mathcal{U}(\gamma_t,\theta_t)}{\partial \mathcal{B}_t} = -\left[-2\left(1-\gamma_t\right)\frac{1}{\tau^{\theta_t}}\frac{\partial \gamma_t}{\partial \mathcal{B}_t} + 2\frac{\eta}{\zeta\tau_F}\gamma_t\frac{\partial \gamma_t}{\partial \mathcal{B}_t}\right]$$
$$= -2\frac{\partial \gamma_t}{\partial \mathcal{B}_t}\left[\frac{\eta}{\zeta\tau_F}\gamma_t - \left(1-\gamma_t\right)\frac{1}{\tau^{\theta_t}}\right]$$

By (Lemma 1), $\frac{\partial \gamma_t}{\partial \mathcal{B}_t} < 0$, and thus the sign of $\frac{\partial \mathcal{U}(\gamma_t,\theta_t)}{\partial \mathcal{B}_t}$ depends on the sign of the bracketed terms. This term takes a minimum when γ_t is at its smallest possible value, and when τ^{θ_t} is at its smallest possible value (noting that $0 < \gamma_t < 1$ and $\tau^{\theta_t} > 0$). The minimum of γ_t attains when $\mathcal{B}_t = 1$ and thus

$$\gamma_t^{min} = \frac{\zeta\tau_F}{\tau^H + \zeta\tau_F}$$

while the minimum possible value of τ^{θ_t} is τ^L. Thus the minimum of the bracketed terms is

$$\frac{\eta}{\zeta\tau_F}\frac{\zeta\tau_F}{\tau^H+\zeta\tau_F} - \left(1 - \frac{\zeta\tau_F}{\tau^H+\zeta\tau_F}\right)\frac{1}{\tau^L}$$
$$= \frac{\eta}{\tau^H+\zeta\tau_F} - \left(\frac{\tau^H}{\tau^H+\zeta\tau_F}\right)\frac{1}{\tau^L} > 0$$

if $\eta > \frac{\tau^H}{\tau^L}$ as guaranteed by Assumption 3. It immediately follows that

$$\frac{\partial \mathcal{U}\left(\gamma_t, \theta_t\right)}{\partial \mathcal{B}_t} > 0$$

as desired. □

Proposition 5

Proof. From the proof of Proposition 2, $\widetilde{y}_t = (1-\gamma_t)\,\varepsilon_t^\theta$, where $\widetilde{y}_t = (1-\gamma_t)q_t - (1-\gamma_t)y_t^N$. In addition, from the LLN, $p_t = \gamma_t q_t - \gamma_t y_t^N$ and thus $y_t^N = \frac{\gamma_t q_t - p_t}{\gamma_t}$. Putting these two equations yield

$$\varepsilon_t^\theta = q_t + \frac{p_t - \gamma_t q_t}{\gamma_t} = \frac{q_t - y_t}{\gamma_t}$$

where each term is perfectly observable. The error in terms of the output gap is immediate from $\widetilde{y}_t = (1-\gamma_t)\,\varepsilon_t^\theta$. □

Proposition 6

Proof. First note that we have

$$\mathcal{B}_{t+1} = \lambda + \widehat{\mathcal{B}}_t(1-2\lambda)$$

which is increasing in $\widehat{\mathcal{B}}_t$.

Given beliefs, by Proposition 3, price dispersion is fixed; thus, the variability of monetary welfare comes solely from the output gap. From before, $\widetilde{y}_t = (1-\gamma_t)\,\varepsilon_t^\theta$ where ε_t^θ is perfectly inferred by firms after aggregate outcomes are realized, by Proposition 5. Thus,

$$\begin{aligned}\widehat{\mathcal{B}}_t &= \frac{\mathcal{L}(\theta = H \mid q_t, y_t, p_t, \mathcal{B}_t)}{\mathcal{B}_t\mathcal{L}(\theta = H \mid q_t, y_t, p_t, \mathcal{B}_t) + (1-\mathcal{B}_t)\mathcal{L}(\theta = L \mid q_t, y_t, p_t, \mathcal{B}_t)}\mathcal{B}_t \\ &= \frac{f\left(\varepsilon_t^\theta \mid \theta = H\right)}{\mathcal{B}_t f\left(\varepsilon_t^\theta \mid \theta = H\right) + (1-\mathcal{B}_t) f\left(\varepsilon_t^\theta \mid \theta = L\right)}\mathcal{B}_t\end{aligned}$$

where

$$f\left(\varepsilon^\theta \middle| \theta\right) = \frac{1}{\sqrt{2\left(\tau^\theta\right)^{-1}\pi}} \exp\left(-\frac{(\varepsilon^\theta)^2}{2\left(\tau^\theta\right)^{-1}}\right)$$

As well known, $\widehat{\mathcal{B}}_t$ is in turn increasing in the likelihood ratio

$$\chi\left(\varepsilon_t^\theta\right) = \frac{f\left(\varepsilon_t^\theta \mid \theta = H\right)}{f\left(\varepsilon_t^\theta \mid \theta = L\right)}$$

$$= \frac{\frac{1}{\sqrt{2(\tau^H)^{-1}\pi}} \exp\left(-\frac{\left(\varepsilon_t^\theta\right)^2}{2(\tau^H)^{-1}}\right)}{\frac{1}{\sqrt{2(\tau^L)^{-1}\pi}} \exp\left(-\frac{\left(\varepsilon_t^\theta\right)^2}{2(\tau^L)^{-1}}\right)}$$

$$= \sqrt{\frac{\tau^H}{\tau^L}} \exp\left(\frac{1}{2}\left(\varepsilon_t^\theta\right)^2\left(\frac{1}{(\tau^L)^{-1}} - \frac{1}{(\tau^H)^{-1}}\right)\right)$$

Note that

$$\frac{1}{(\tau^L)^{-1}} - \frac{1}{(\tau^H)^{-1}} = \tau^L - \tau^H < 0$$

and thus the likelihood ratio is decreasing in $\left(\varepsilon_t^\theta\right)^2$, and is thus decreasing in $\widetilde{y}_t^2 = (1-\gamma_t)^2\left(\varepsilon_t^\theta\right)^2$. Since $U_t = -\left[\widetilde{y}_t^2 + \kappa\mathcal{V}(p_t^i)\right]$, and $\mathcal{V}(p_t^i)$ is fixed, it follows that the likelihood ratio is increasing in U_t. Putting everything together, it follows that $\widehat{\mathcal{B}}_t$ and thus $\mathcal{B}_{t+1}$ is increasing in U_t. □

Proposition 7

Proof. In order to analyze expected period welfare in future periods, I need to consider the joint distribution of $\mathbf{S}_{t+j} = (\theta_{t+j}, \mathcal{B}_{t+j})$. As already shown in the proof of Proposition 1, in equilibrium expected current period welfare is higher when $\theta = H$. Furthermore, it can be easily shown that

$$P(\theta_{t+j} = H \mid \theta_t = H) = \frac{1}{2}\left[1 + (1-2\lambda)^j\right] > 1 - \frac{1}{2}\left[1 + (1-2\lambda)^j\right] = P(\theta_{t+j} = H \mid \theta_t = L)$$

and so $\theta_t = H$ implies θ_{t+j} for all $j = 1, 2, 3, ...$ is more likely to be H than if $\theta_t = L$.

Given the above expressions, we can analyze the joint distribution of the state vector by consider the conditional densities $f_{\mathcal{B}}(\mathcal{B}_{t+j} \mid \theta_{t+j} = H)$ and $f_{\mathcal{B}}(\mathcal{B}_{t+j} \mid \theta_{t+j} = L)$. I first show that the conditional distribution $f_{\mathcal{B}}(\mathcal{B}_{t+j} \mid \theta_t = H, \theta_{t+j}, \mathcal{B}_t)$ FOSD $f_{\mathcal{B}}(\mathcal{B}_{t+j} \mid \theta_t = L, \theta_{t+j}, \mathcal{B}_t)$ for both $\theta_{t+j} = H, L$. To show this is the case, consider any arbitrary sequence starting from $\theta_t = L$ and ending in θ_{t+j}. There exists the exact same sequence corresponding to $\theta_t = H$, with all states in each period but the initial state the same; that is, the sequence starting from θ_{t+1} can be denoted by $(\theta_{t+1}, \theta_{t+2}, .., \theta_{t-j}, \theta_t)$. It immediately follows that because each of these cases draws from the same distribution starting from $t+1$ (period-by-period), and because the distribution of $\mid \varepsilon^L \mid$ FOSD $\mid \varepsilon^H \mid$, it follows that the distribution $f_{\mathcal{B}}(\mathcal{B}_{t+j} \mid \theta_t = H, \theta_{t+1}, \theta_{t+2}, .., \theta_{t-j}, \theta_{t+j}, \mathcal{B}_t)$ FOSD $f_{\mathcal{B}}(\mathcal{B}_{t+j} \mid \theta_t = L, \theta_{t+1}, \theta_{t+2}, .., \theta_{t+j}, \mathcal{B}_t)$. Because $f_{\mathcal{B}}(\mathcal{B}_{t+j} \mid \theta_t, \theta_{t+j}, \mathcal{B}_t)$ is a weighted sum of each possible sequence such that the true type in period $t+j$ is θ_{t+j}, it follows that $f_{\mathcal{B}}(\mathcal{B}_{t+j} \mid \theta_t = H, \theta_{t+j}, \mathcal{B}_t)$

FOSD $f_{\mathcal{B}}(\mathcal{B}_{t+j} \mid \theta_t = L, \theta_{t+j}, \mathcal{B}_t)$.

Summarizing, we have that in all future periods $t+j$, $P(\theta_{t+j} = H \mid \theta_t = H) > P(\theta_{t+j} = H \mid \theta_t = L)$ and conditional on each true type θ_{t+j}, the distribution of $\mathcal{B}_{t+j}$ when the initial true type is $\theta_t = H$ FOSD the distribution of $\mathcal{B}_{t+j}$ when the initial true type is $\theta_t = L$. Because expected current period welfare is higher when $\theta = H$ and when $\mathcal{B}$ is higher, it follows that for a given $\mathcal{B}_t$, $V(\theta = H, \mathcal{B}_t) > V(\theta = L, \mathcal{B}_t)$. □

Proposition 8

Proof. This follows from the argument made in the proof of Proposition 1. First note that sequence of true types $\{\theta_{t+j}\}_{j=1}^{\infty}$ is exogenously determined. This also implies that in equilibrium, at any given period $t+j$ the distribution of $\mathcal{B}_{t+j}$ when period t's belief is $\mathcal{B}$ FOSD the distribution of $\mathcal{B}_{t+j}$ when period t's belief is $\widetilde{\mathcal{B}}_t$, when $\mathcal{B}_t > \widetilde{\mathcal{B}}_t$. Because $\mathcal{U}(\gamma(\mathcal{B}), \theta)$ is increasing in $\mathcal{B}$, taking these facts together imply that $\mathbb{E}_t\mathcal{U}_{t+j}$ is higher for all $j = 0, 1, 2, ...$ when $\mathcal{B}_t > \widetilde{\mathcal{B}}_t$,which implies that for any θ_t, $\frac{\partial V(\theta, \mathcal{B}_t)}{\partial \mathcal{B}_t} \geq 0$. □

Proposition 9

Proof. To show this, I conjecture this to be the equilibrium and show that each agent has no incentive to deviate from this equilibrium. I first look at the price set by a firm changing its price in period t, having received signal S_t^i. Conjecturing further that $\mathbb{E}_t\left[p_{t+j}^* \mid \mathcal{I}_t^i\right] = \mathbb{E}_t\left[p_t \mid \mathcal{I}_t^i\right]$, and later showing this to be true, we have

$$
\begin{aligned}
p_t^i &= \mathbb{E}_t\left[(1-\beta\phi)p_t^{i*} \mid \mathcal{I}_t^i\right] + \mathbb{E}_t\left[(1-\beta\phi)\sum_{j=1}^{\infty}(\beta\phi)^j p_t \middle| \mathcal{I}_t^i\right] \\
&= (1-\beta\phi)\mathbb{E}_t\left[p_t + \zeta\widetilde{y}_t \mid \mathcal{I}_t^i\right] + \beta\phi\mathbb{E}_t\left[p_t \mid \mathcal{I}_t^i\right] \\
&= \mathbb{E}_t\left[p_t \mid \mathcal{I}_t^i\right] + \zeta(1-\beta\phi)\mathbb{E}_t\left[\widetilde{y}_t \mid \mathcal{I}_t^i\right]
\end{aligned}
$$

Since $p_t = \varphi_{0t} p_{t-1} + \varphi_{1t} y_t^N + \varphi_{2t}\varepsilon_t^\theta$, for some coefficients $\varphi_0, \varphi_{1t}, \varphi_{2t}$[6] we have

$$\mathbb{E}_t\left[p_t \mid \mathcal{I}_t^i\right] = \varphi_{0t} p_{t-1} + \varphi_{1t}\mathbb{E}_t\left[y_t^N \mid \mathcal{I}_t^i\right] + \varphi_{2t}\mathbb{E}_t\left[\varepsilon_t^\theta \mid \mathcal{I}_t^i\right]$$

Because

$$\mathbb{E}_t\left[y_t^N \mid \mathcal{I}_t^i\right] = \mathcal{B}_t \frac{\tau^H (q_t - p_{t-1}) + \tau_F S_t^i}{\tau^H + \tau_F} + (1-\mathcal{B}_t)\frac{\tau^L (q_t - p_{t-1}) + \tau_F S_t^i}{\tau^B + \tau_F} = \Omega_t (q_t - p_{t-1}) + \mathcal{H}_t^{-1}\tau_F S_t^i$$

and

$$\mathbb{E}_t\left[\varepsilon_t^\theta \mid \mathcal{I}_t^i\right] = q_t - p_{t-1} - \left(\Omega_t (q_t - p_{t-1}) + \mathcal{H}_t^{-1}\tau_F S_t^i\right) = (q_t - p_{t-1})(1-\Omega_t) - \mathcal{H}_t^{-1}\tau_F S_t^i$$

Thus

$$\mathbb{E}_t\left[p_t \mid \mathcal{I}_t^i\right] = \varphi_{0t} p_{t-1} + \varphi_{1t}\left\{\Omega_t (q_t - p_{t-1}) + \mathcal{H}_t^{-1}\tau_F S_t^i\right\} + \varphi_{2t}\left\{(q_t - p_{t-1})(1-\Omega_t) - \mathcal{H}_t^{-1}\tau_F S_t^i\right\}$$

Similarly, we have

$$\begin{aligned}\widetilde{y}_t &= \varepsilon_t^\theta + p_{t-1} - p_t \\ &= \varepsilon_t^\theta + p_{t-1} - \left(\varphi_{0t} p_{t-1} + \varphi_{1t} y_t^N + \varphi_{2t}\varepsilon_t^\theta\right) \\ &= (1-\varphi_{0t}) p_{t-1} - \varphi_{1t} y_t^N + (1-\varphi_{2t})\varepsilon_t^\theta\end{aligned}$$

and so

$$\mathbb{E}_t\left[\widetilde{y}_t \mid \mathcal{I}_t^i\right] = (1-\varphi_{0t}) p_{t-1} - \varphi_{1t}\left\{\Omega_t (q_t - p_{t-1}) + \mathcal{H}_t^{-1}\tau_F S_t^i\right\} + (1-\varphi_{2t})\left\{(q_t - p_{t-1})(1-\Omega_t) - \mathcal{H}_t^{-1}\tau_F S_t^i\right\}$$

Putting everything together,

$$p_t^i = \widehat{\gamma}_{0t} p_{t-1} + \widehat{\gamma}_{1t}(q_t - p_{t-1}) + \widehat{\gamma}_{2t} S_t^i$$

---

[6]This can be justified as follows. Because a fraction of firms cannot change their prices, and due to strategic complementarities, p_{t-1} is an important factor driving a firm's optimal price. Each firm must also attempt to infer the natural rate of output, as well as other firms' inferences of the natural rate of output, and this information is provided by the idiosyncratic signal S_t^i as well as $q_t - p_{t-1}$, which is the central bank's signal when it plays $q_t = S_t + p_{t-1}$. Given this guess, and the fact that only a randomly selected fraction $1-\phi$ of firms can adjust their prices, the aggregate price level must then satisfy

$$p_t = \phi p_{t-1} + (1-\phi)\int P(p_{t-1}, q_t, S_t^j; y_t^N)\, dF_t(S_t^j)$$

Because of the linearity of $P(\cdot)$ and by the Law of Large Numbers implying that $\int S_t^j dF_t(S_t^j) = y_t^N$, it can be shown that

$$p_t = \varphi_{0t} p_{t-1} + \varphi_{1t} y_t^N + \varphi_{2t}\varepsilon_t^\theta$$

where

$$\widehat{\gamma}_{0t} = \varphi_{0t} + \zeta(1-\beta\phi)(1-\varphi_{0t})$$

$$\widehat{\gamma}_{1t} = \varphi_{1t}\Omega_t + \varphi_{2t}(1-\Omega_t) + \zeta(1-\beta\phi)\left\{-\varphi_{1t}\Omega_t + (1-\varphi_{2t})(1-\Omega_t)\right\}$$

$$\widehat{\gamma}_{2t} = \varphi_{1t}\mathcal{H}_t^{-1}\tau_F - \varphi_{2t}\mathcal{H}_t^{-1}\tau_F + \zeta(1-\beta\phi)\left\{-\varphi_{1t}\mathcal{H}_t^{-1}\tau_F - (1-\varphi_{2t})\mathcal{H}_t^{-1}\tau_F\right\}$$

$$\varphi_{0t} = \phi + (1-\phi)\widehat{\gamma}_{0t}$$

$$\varphi_{1t} = (1-\phi)\left(\widehat{\gamma}_{1t} + \widehat{\gamma}_{2t}\right)$$

$$\varphi_{2t} = (1-\phi)\widehat{\gamma}_{1t}$$

where the last three equations come from $p_t = \phi p_{t-1} + (1-\phi)\left[\widehat{\gamma}_{0t}p_{t-1} + \widehat{\gamma}_{1t}\left(y_t^N + \varepsilon_t^\theta\right) + \widehat{\gamma}_{2t}y_t^N\right]$. Solving this system of equations yields

$$\widehat{\gamma}_{0t} = 1,\ \widehat{\gamma}_{1t} = \frac{\zeta(1-\beta\phi)\tau_F}{\mathcal{H}_t - \big(1-\zeta(1-\beta\phi)\big)(1-\phi)\tau_F},\ \widehat{\gamma}_{2t} = -\frac{\zeta(1-\beta\phi)\tau_F}{\mathcal{H}_t - \big(1-\zeta(1-\beta\phi)\big)(1-\phi)\tau_F}$$

$$\varphi_{0t} = 1,\ \varphi_{1t} = 0,\ \varphi_{2t}{=}(1-\phi)\frac{\zeta(1-\beta\phi)\tau_F}{\mathcal{H}_t - \big(1-\zeta(1-\beta\phi)\big)(1-\phi)\tau_F}$$

which correspond to the coefficients in the main text. Because $\widehat{\gamma}_t$ and φ_{2t} are decreasing in $\mathcal{H}_t$, and $\mathcal{H}_t$ is increasing in $\mathcal{B}_t$, property (i) follows. Noting that $\mathcal{H}_t > \tau_F > (1-\phi)\tau_F$, we also have

$$\frac{\partial\widehat{\gamma}_t}{\partial\zeta} = \frac{(1-\beta\phi)\tau_F\left[\mathcal{H}_t - \big(1-\zeta(1-\beta\phi)\big)(1-\phi)\tau_F\right] - \zeta(1-\beta\phi)\tau_F(1-\beta\phi)(1-\phi)\tau_F}{\left[\mathcal{H}_t - \big(1-\zeta(1-\beta\phi)\big)(1-\phi)\tau_F\right]^2} > 0$$

$$\frac{\partial\varphi_{2t}}{\partial\zeta} = (1-\phi)\frac{\partial\widehat{\gamma}_t}{\partial\zeta} > 0$$

and

$$\frac{\partial\widehat{\gamma}_t}{\partial\phi} = \frac{\Lambda'(\phi)\tau_F\left[\mathcal{H}_t - (1-\Lambda(\phi))(1-\phi)\tau_F\right] - \left[\Lambda'(\phi)(1-\phi) + (1-\Lambda(\phi))\phi\right]\tau_F\Lambda(\phi)\tau_F}{\left[\mathcal{H}_t - \big(1-\zeta(1-\beta\phi)\big)(1-\phi)\tau_F\right]^2}$$

where $\Lambda(\phi) = \zeta(1-\beta\phi)$ with $\frac{d\Lambda}{d\phi} = -\zeta\beta < 0$. Because the numerator is equal to

$$\tau_F\left\{\Lambda'(\phi)\left[\mathcal{H}_t - (1-\Lambda(\phi))(1-\phi)\tau_F\right] - \left[\Lambda'(\phi)(1-\phi) + (1-\Lambda(\phi))\phi\right]\Lambda(\phi)\tau_F\right\}$$

$$\tau_F\left\{\Lambda'(\phi)\mathcal{H}_t - \Lambda'(\phi)(1-\phi)\tau_F - (1-\Lambda(\phi))\phi\tau_F\Lambda(\phi)\right\}$$

$$= \tau_F\left\{\underbrace{\Lambda'(\phi)}_{-}\left[\underbrace{\mathcal{H}_t - (1-\phi)\tau_F}_{+}\right] - (1-\Lambda(\phi))\phi\tau_F\Lambda(\phi)\right\} < 0$$

It follows that

$$\frac{\partial \widehat{\gamma}_t}{\partial \phi} < 0$$

$$\frac{\partial \varphi_{2t}}{\partial \phi} = -\widehat{\gamma}_t + (1-\phi)\frac{\partial \widehat{\gamma}_t}{\partial \phi} < 0$$

and thus property (ii) follows.

Lastly, I now verify that initial conjecture $\mathbb{E}_t\left[p^*_{t+j} \mid \mathcal{I}^i_t\right] = \mathbb{E}_t\left[p_t \mid \mathcal{I}^i_t\right]$ holds. Since in the conjectured equilibrium the central bank plays $q_t = S^\theta_t + p_{t-1}$ we have

$$\mathbb{E}_t\left[p^*_{t+j} \mid \mathcal{I}^i_t\right] = \mathbb{E}_t\left[p_{t+j} + (1-\alpha)\widetilde{y}_{t+j} \mid \mathcal{I}^i_t\right]$$

$$= \mathbb{E}_t\left[\varphi_{0t+j}p_{t+j-1} + \varphi_{1t+j}y^N_{t+j} + \varphi_{2t+j}\varepsilon^\theta_{t+j} + (1-\alpha)\left[(1-\varphi_{0t+j})p_{t+j-1} - \varphi_{1t+j}y^N_{t+j} + (1-\varphi_{2t+j})\varepsilon^\theta_{t+j}\right] \mid \mathcal{I}^i_t\right]$$

$$= \mathbb{E}_t\left[p_{t+j-1} \mid \mathcal{I}^i_t\right]$$

And thus it follows that $\mathbb{E}_t\left[p^*_{t+j} \mid \mathcal{I}^i_t\right] = \mathbb{E}_t\left[p_t \mid \mathcal{I}^i_t\right]$ for all $j = 1, 2, \ldots$.

I now look at the central bank's equilibrium actions. First consider the period welfare function

$$\mathcal{U}_t = -\left[\widetilde{y}^2_t + \kappa\frac{\phi}{1-\phi}\pi^2_t + \frac{(1-\phi)\kappa}{\tau_F}\widehat{\gamma}^2_{2t} + \kappa\phi\mathcal{V}(p^i_{t-1})\right]$$

The central bank has no control over both the third and fourth terms, which are predetermined as of time t. The 'controllable' part of period welfare is a weighted function of both the output gap and inflation. Assuming firm equilibrium actions derived above, the output gap can be written as

$$\widetilde{y}_t = q_t - \left(\phi p_{t-1} + (1-\phi)\left[\widehat{\gamma}_{0t}p_{t-1} + \widehat{\gamma}_{1t}\left(q_t - p_{t-1}\right) + \widehat{\gamma}_{2t}y^N_t\right]\right) - y^N_t$$
$$= \left(1-(1-\phi)\widehat{\gamma}_t\right)q_t - \left(1-(1-\phi)\widehat{\gamma}_t\right)p_{t-1} - \left(1-(1-\phi)\widehat{\gamma}_t\right)y^N_t$$

And so

$$\mathbb{E}_t\left[\widetilde{y}^2_t \mid S^\theta_t\right] = \left(1-(1-\phi)\widehat{\gamma}_t\right)^2\left(q_t - p_{t-1}\right)^2$$
$$+\left(1-(1-\phi)\widehat{\gamma}_t\right)^2\mathbb{E}_t\left[\left(y^N_t\right)^2 \mid S^\theta_t\right] - 2\left(1-(1-\phi)\widehat{\gamma}_t\right)^2\left(q_t - p_{t-1}\right)\mathbb{E}_t\left[y^N_t \mid S^\theta_t\right]$$
$$= \left(1-(1-\phi)\widehat{\gamma}_t\right)^2\left(q_t - p_{t-1}\right)^2 + \left(1-(1-\phi)\widehat{\gamma}_t\right)^2\left[\left(S^\theta_t\right)^2 + \left(\tau^\theta\right)^{-1}\right] - 2\left(1-(1-\phi)\widehat{\gamma}_t\right)^2\left(q_t - p_{t-1}\right)S^\theta_t$$

Taking FOC w.r.t. q_t gives

$$2\left(1-(1-\phi)\widehat{\gamma}_t\right)^2\left(q_t - p_{t-1}\right) - 2\left(1-(1-\phi)\widehat{\gamma}_t\right)^2 S^\theta_t = 0$$

Leading to $q_t = S_t^\theta + p_{t-1}$. The inflation term is

$$\pi_t = p_t - p_{t-1} = \phi p_{t-1} + (1-\phi)\left[\widehat{\gamma}_{0t} p_{t-1} + \widehat{\gamma}_{1t}\left(q_t - p_{t-1}\right) + \widehat{\gamma}_{2t} y_t^N\right] - p_{t-1}$$
$$= \widehat{\gamma}_{1t}\left(q_t - p_{t-1}\right) + \widehat{\gamma}_{2t} y_t^N$$

and so

$$\mathbb{E}_t\left[\pi_t^2 \mid S_t^\theta\right] = \widehat{\gamma}_t^2\left(q_t - p_{t-1}\right)^2 + \widehat{\gamma}_t^2 \mathbb{E}_t\left[\left(y_t^N\right)^2 \mid S_t^\theta\right] + 2\widehat{\gamma}_t^2\left(q_t - p_{t-1}\right)\mathbb{E}_t\left[y_t^N \mid S_t^\theta\right]$$
$$= \widehat{\gamma}_t^2\left(q_t - p_{t-1}\right)^2 + \widehat{\gamma}_t^2\left[\left(S_t^\theta\right)^2 + \left(\tau^\theta\right)^{-1}\right] + 2\widehat{\gamma}_t^2\left(q_t - p_{t-1}\right) S_t^\theta$$

Taking FOC w.r.t. q_t gives

$$2\widehat{\gamma}_t^2\left(q_t - p_{t-1}\right) + 2\widehat{\gamma}_t^2 S_t^\theta = 0$$

leading to $q_t = S_t^\theta + p_{t-1}$. Thus $q_t = S_t^\theta + p_{t-1}$ minimizes both the output gap and inflation term, and the divine coincidence holds. This is thus the optimal action for maximizing current period welfare (noting that minimizing inflation also minimizes future residual price dispersion). In order to analyze the dynamics, I first consider the price dispersion that obtains in period t, and carries into period $t+1$'s welfare function. This can be expressed as

$$\mathcal{V}(p_t^i) = \phi\mathcal{V}(p_{t-1}^i) + \frac{\phi}{1-\phi}\pi_t^2 + (1-\phi)\widehat{\gamma}_{2t}^2\left(\tau_F\right)^{-1}$$

Because the central bank can only affect the squared inflation term, and because this term is minimized in expectation at $q_t = S_t^\theta + p_{t-1}$, the central bank minimizes expected price dispersion by setting the proposed equilibrium action.[7]In addition, if the firms believe the central bank takes equilibrium actions $q_t = S_t^\theta + p_{t-1}$, then using the same argument as shown in Proposition 1, given S_t^θ for any action $q_t \neq S_t^\theta + p_{t-1}$ the distribution of $\mid \varepsilon_t^P \mid$ first order stochastically dominates the distribution when $q_t = S_t^\theta + p_{t-1}$. Because $\mathcal{B}_{t+1}$ is decreasing in $\mid \varepsilon_t^P \mid$, it follows that $q_t = S_t^\theta + p_{t-1}$ maximizes beliefs heading into the future period. Because $\mathcal{U}_t$ is decreasing in $\gamma(\mathcal{B})$ which is itself decreasing in $\mathcal{B}$, and because for any possible sequence of true types $\left\{\theta_{t+j}\right\}_{j=1}^\infty$, for given period $t+j$ the distribution of $\mathcal{B}_{t+j}$ when period $t+1$'s belief is $\mathcal{B}_{t+1}$ FOSD the distribution of $\mathcal{B}_{t+j}$ when

[7]To be more precise, one needs to show that the distribution of π_t^2 when $q_t = S_t^\theta + p_{t-1}$ is first order stochastically dominated by the distribution for any other q_t, a stronger condition than the expectation is minimized. I do not show this here.

period t's belief is $\widetilde{\mathcal{B}}_{t+1}$, when $\mathcal{B}_{t+1} > \widetilde{\mathcal{B}}_{t+1}$, it follows that $q_t = S_t^\theta + p_{t-1}$ is indeed the dynamically optimal action for each central bank type θ. □

Proposition 10

Proof. The expected welfare term comes from plugging in the equilibrium actions found in Proposition 9 and plugging into (1.17). The second part of the proposition follows the proofs of Proposition 2 and Proposition 3. To show the last part of the proposition, note that

$$-\left\{\frac{\partial \mathbb{E}_t\left(\tilde{y}_t^2 \mid \theta, \mathcal{B}_t\right)}{\partial \mathcal{B}_t} + \frac{\kappa}{1-\beta\phi}\left(\frac{\phi}{1-\phi}\frac{\partial \mathbb{E}_t\left(\pi_t^2 \mid \theta, \mathcal{B}_t\right)}{\partial \mathcal{B}_t} + (1-\phi)\frac{\partial(\triangle_t \mid \theta, \mathcal{B}_t)}{\partial \mathcal{B}_t}\right)\right\}$$

$$= -\left\{-2(1-\Psi_t)\frac{1}{\tau^\theta} + \kappa\frac{\phi}{1-\phi}2\Psi_t\frac{1}{\tau^\theta} + \frac{(1-\phi)\kappa}{\tau_F}2\frac{1}{1-\phi}\left(\frac{\Psi_t}{1-\phi}\right)\right\}$$

which is less than zero if

$$-(1-\Psi_t)(1-\phi)\frac{1}{\tau^\theta} + \frac{1}{1-\beta\phi}\frac{\eta}{\zeta}\phi\Psi_t\frac{1}{\tau^\theta} + \left(\frac{1}{1-\beta\phi}\right)\frac{(1-\phi)\eta}{\zeta\tau_F}\left(\frac{\Psi_t}{1-\phi}\right) > 0$$

Rearranging terms, this is true if

$$\Psi_t > \frac{(1-\phi)\zeta\tau_F}{\frac{1}{1-\beta\phi}\tau^\theta\eta + \frac{\phi}{1-\beta\phi}\tau_F\eta + (1-\phi)\zeta\tau_F}$$

Since $\mathcal{H}_t$ attains a maximum at $\tau_G + \tau_F$, we know that

$$\Psi_t = (1-\phi)\frac{\zeta(1-\beta\phi)\tau_F}{\mathcal{H}_t - \big(1-\zeta(1-\beta\phi)\big)(1-\phi)\tau_F} \geq (1-\phi)\frac{\zeta(1-\beta\phi)\tau_F}{\tau_G + \tau_F - \big(1-\zeta(1-\beta\phi)\big)(1-\phi)\tau_F}$$

Thus what we need to show is

$$(1-\phi)\frac{\zeta(1-\beta\phi)\tau_F}{\tau^H + \tau_F - \big(1-\zeta(1-\beta\phi)\big)(1-\phi)\tau_F} > \frac{(1-\phi)\zeta\tau_F}{\frac{1}{1-\beta\phi}\tau^L\eta + \frac{\phi}{1-\beta\phi}\tau_F\eta + (1-\phi)\zeta\tau_F}$$

Rearranging and simplifying, this condition is simply

$$\tau_F\left[\phi(\eta-1)\right] > \tau^H - \tau^L\eta$$

Since by Assumption 3 the RHS is negative, and since $\tau_F, \phi > 0$ and $\eta > 1$, this condition always holds. □