Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

The Natural Rate of Interest Through a Hall of Mirrors

Phurichai Rungcharoenkitkul and Fabian Winkler

2022-010

Please cite this paper as:
Rungcharoenkitkul, Phurichai, and Fabian Winkler (2022). "The Natural Rate of Interest Through a Hall of Mirrors," Finance and Economics Discussion Series 2022-010. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2022.010.

The Natural Rate of Interest Through a Hall of Mirrors*

Phurichai Rungcharoenkitkul† Fabian Winkler‡

March 7, 2022

Abstract

Prevailing explanations of persistently low interest rates appeal to a secular decline in the natural interest rate, or r-star, due to factors outside monetary policy's control. We propose informational feedback via learning as an alternative explanation for persistently low rates, where monetary policy plays a crucial role. We extend the canonical New Keynesian model to an incomplete information setting where the central bank and the private sector learn about r-star and infer each other's information from observed macroeconomic outcomes. An informational feedback loop emerges when each side underestimates the effect of its own action on the other's inference, possibly leading to large and persistent changes in perceived r-star disconnected from fundamentals. Monetary policy, through its influence on the private sector's beliefs, endogenously determines r-star as a result. We simulate a calibrated model and show that this 'hall-of-mirrors' effect can explain much of the decline in real interest rates since 2008.

JEL Classification: E43, E52, E58, D82, D83

Keywords: Natural rate of interest, learning, misperception, overreaction, dispersed information, long-term rates, demand shocks, monetary policy shocks.

*We thank Claudio Borio, Jenny Chan, Giovanni Favara, Etienne Gagnon, Ben Johannsen, Esa Jokivuolle, Nigel McClung, Afrasiab Mirza, Hyun Song Shin, Min Wei, and participants at the 2022 AEA annual meeting, Bank of Finland, the Conference Expectations in Dynamic Macroeconomic Models at the Czech National Bank, CEF 2021 Conference, the EEA Annual Congress 2021, the 52nd MMF Annual Conference, and the European Winter Meeting of the Econometric Society 2021 for helpful comments. All remaining errors are ours. The views herein are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System, the Bank for International Settlements, or of any other person associated with the Federal Reserve System or the Bank for International Settlements.

†Bank for International Settlements (Basel, Switzerland): phurichai.rungcharoenkitkul@bis.org

‡Board of Governors of the Federal Reserve System (Washington, DC, United States.): fabian.winkler@frb.gov

1 Introduction

Few concepts have had greater influence on monetary policy in the past decade than the natural rate of interest, or r-star—the real interest rate consistent with output equaling potential and stable inflation. Interest rates fell sharply in the wake of the Great Financial Crisis (GFC) in 2008, as major central banks cut their policy interest rates to record lows in a bid to support economic recovery. Nominal interest rates then stayed persistently low in the subsequent decade, but global inflation remained subdued even after demand recovered. Standard macroeconomic theory can rationalize the co-existence of a persistent lack of price pressure and very low interest rates by a decline in r-star. A fall in r-star to very low levels is a challenge to central banks because it limits the scope of monetary policy accommodation that policymakers can provide. Such concerns have led some central banks to introduce unconventional policy measures, and more recently to review their monetary policy frameworks with a view to regaining policy space.

Several factors driving a persistent fall in r-star have been proposed, including a fall in trend productivity growth, population aging and higher demand for safe assets, among others. These explanations invoke different changes in economic *fundamentals* that raise real desired savings or lower desired investment, putting downward pressure on the equilibrium real interest rate. Empirically, there is little consensus, however, on the relative importance of these factors. The literature that evaluates these competing explanations without imposing a priori theoretical restrictions is relatively scant, and tends to find only limited explanatory power of various saving-investment factors consistently over long samples (see Borio et al. (2017) and Lunsford and West (2019)). The lack of solid empirical evidence is perhaps not surprising given the inherent identification challenge: Not only is r-star unobservable, but it is also a theoretical construct—it can only be estimated by taking a stand on the correct model of the economy. This leaves open the possibility that other factors may well be relevant secular drivers of real interest rates.

This paper proposes an alternative explanation of persistent movements in r-star that is based on endogenous *beliefs* and informational feedback. The central idea is that r-star depends on beliefs which can evolve in a persistent way when the central bank and the private sector learn from each other. When the central bank adjusts the policy rate, it sends a signal about r-star that the private sector incorporates into consumption-saving decisions. This in turn affects macroeconomic dynamics, which feed back into the central bank's inference about r-star. When agents underestimate the importance of this informational feedback, r-star can become endogenous to cyclical perturbations including those of monetary policy.

We formalize the idea by adding stochastic trends to the canonical New Keynesian model.

The exogenous, fundamental determinants of r-star are difficult to observe and may change over time. Both the private sector and the central bank learn about these fundamentals from their own information, as well as from observing output, inflation and interest rates, which partially reveal the information of the other side. This setup, which is new to the literature, is arguably a good description of a world in which central banks infer r-star at least partly from macroeconomic outcomes, while markets, firms and households also form their expectations of future interest rates at least partly from current policy rates and central bank communications.

While our extension to incomplete information is simple, it has strong macroeconomic implications, in particular when the central bank and the private sector overestimate the information quality of the other.[1] We show that with this misperception, r-star beliefs overreact to cyclical macroeconomic shocks. This overreaction can be very persistent and quantitatively substantial. In particular, r-star becomes highly endogenous to monetary policy because the private sector systematically mistakes policy actions as revelations of useful information about long-run fundamentals.

The overreaction effect can be likened to a *hall of mirrors* in the spirit of Bernanke (2004): The central bank's expectations excessively reflect the private sector's expectations and vice-versa. To illustrate, suppose that the central bank cuts interest rates sharply in response to a recession. In our model, private agents mistakenly attribute a part of this policy adjustment to the central bank reevaluating its views about the long-run real interest rate in the economy. In response, the private sector lowers their own estimate of r-star, prompting output and inflation to fall. The central bank in turn mistakenly interprets this demand shortfall as an indication that r-star has fallen, thus further cutting interest rates. The private sector then lowers its own r-star beliefs further and so on. Both sides end up misperceiving the macroeconomic effects of their own actions as genuine information: They are staring into a hall of mirrors.

Despite its simplicity, our model is capable of explaining a range of salient empirical facts in the post-GFC period, including some that are otherwise difficult to rationalize. In particular, the excess sensitivity of long-term forward real rates to short-term interest rate movements runs counter to the natural rate hypothesis, which postulates a convergence of real interest rates to r-star in the long run. Such sensitivity is a general property of our model, as the private sector (i.e. financial market participants) learns about the long-run real rate from the central bank's actions. Moreover, the model can quantitatively explain

[1]The psychology literature refers to this type of misperception as *informational influence* or *social proof* (Cialdini et al., 1999), a cognitive bias stemming from the belief that others have more accurate information than oneself, thought to be more likely in situations of high uncertainty. Another related phenomenon is *herding behavior*, where agents discard own information in favor of others.

the evolution of several macroeconomic variables after the GFC, including the entire decline of US long-term real interest rates between 2008 and 2019.

The hall-of-mirrors effect has important implications for current monetary policy debates. The extraordinary monetary policy measures over the past decade were guided in no small part by policymakers' beliefs that r-star had substantially fallen, for reasons outside their control. But with the hall-of-mirrors effect, an aggressive policy strategy is less effective in reviving spending, and, worse, exacerbates the very problem policymakers are trying to solve. When the central bank cuts interest rate because it believes r-star has fallen, it not only lowers the short-term interest rate, but also endogenously prompts a decline in the natural rate that weakens the degree of policy accommodation. Our model thus calls for greater recognition of the unintended consequences of policy communications. These consequences are more severe the more the private sector and the central bank overestimate each other's knowledge of the economy.

To our knowledge, our paper is the first in which both the central bank and the private sector are learning about uncertain long-run economic fundamentals from each other. The macroeconomic literature has extensively studied cases in which only the central bank learns about economic fundamentals from the private sector. Prominent contributions in this area are Orphanides (2003), Cukierman and Lippi (2005) and Primiceri (2006) and Nimark (2008). Orphanides and Williams (2007, 2008) additionally allow for imperfect information on behalf of the private sector, though only about the short-run dynamics of the economy. On the empirical side, the well-known r-star estimation procedures of Laubach and Williams (2003) and others (e.g. Holston et al., 2017; Johannsen and Mertens, 2021) also belong in this category, since they estimate r-star from macroeconomic and financial outcomes, which depend on the private sector's information and expectations. Crucially, these empirical studies implicitly assume that r-star is exogenous to monetary policy.

On the flip side, a more recent strand of the literature has examined the case in which only the private sector learns about economic fundamentals from the central bank. This direction of learning is the signaling channel of monetary policy, which has been prominently documented empirically by Nakamura and Steinsson (2018). On the theory side, there have been several structural models of the information channel, for example Tang (2015), Melosi (2016), Angeletos et al. (2020) and Hillenbrand (2022). Our paper forms a bridge between the two strands of the literature by considering the case in which the information sets of the central bank and the private sector are not nested within each other, thus giving rise to learning by both sides.

An older literature in monetary economics has discussed different versions of a hall-of-mirrors effect, in which the central bank relies too much on private sector expectations for its actions. Bernanke and Woodford (1997) argue that if the central bank targets private sector inflation forecasts to steer actual inflation, indeterminacy can obtain from a positive feedback loop between expectations. In our model, the equilibrium is always determinate because observed r-star fundamentals anchor expectations, but amplification of noise can still be unbounded. Morris and Shin (2002) argue that the information provided by monetary policy communications can crowd out dispersed information in the private sector, preventing an efficient aggregation of information. In our model, the main source of inefficient information aggregation comes from misperception about the quality of information, with potentially much more powerful consequences than the crowding-out effect in Morris and Shin (2002). More generally, our paper is the first to argue that a hall-of-mirrors effect may apply to r-star.

Our model also relates to an emerging literature on the possibility that r-star could be endogenous to monetary policy. In Rungcharoenkitkul et al. (2019), a monetary policy regime that focuses unduly on short-term output can exacerbate the financial boom-bust cycle, resulting in lower equilibrium output and interest rates in the long run. In Mian et al. (2020), the natural rate of interest is lower when demand is constrained by over-indebtedness, which can result from monetary policy accommodation. Similarly in Beaudry and Meh (2021), low interest rates can push the economy into an ELB trap in which r-star is endogenously low. In our model, r-star is endogenous not because of fundamental economic mechanisms, but because of mutual learning and endogenous information acquisition. The notorious practical difficulties in assessing r-star speak to the importance of having a model where learning is a central feature.

The remainder of this paper is organized as follows. The next section discusses empirical evidence that motivates our analysis. Section 3 sets up the basic macroeconomic framework modified to accommodate incomplete information, and establishes the modified r-star concept. Section 4 builds intuition by analyzing a tractable static version of the model and deriving key qualitative results. The full dynamic version of our model is laid out in Section 5, and Section 6 discusses our quantitative simulation results. Section 7 concludes.

2 Motivating evidence

Is our proposed hall-of-mirrors effect on r-star simply a theoretical curiosity? The answer would be yes if the natural rate hypothesis held true and if private sector agents did not need to learn about the natural rate from the central bank. The natural rate hypothesis states that the short-term real interest rate will converge in the long run to a natural rate that is independent of monetary policy, which we call r-star. This hypothesis is essentially a restatement of long-run monetary neutrality. Now, if private sector agents believed that central banks do not have any information about r-star that is valuable to them, then monetary policy actions and communications should have no bearing on their r-star expectations. This prediction, if validated empirically, would indeed rule out the hall-of-mirrors effect.

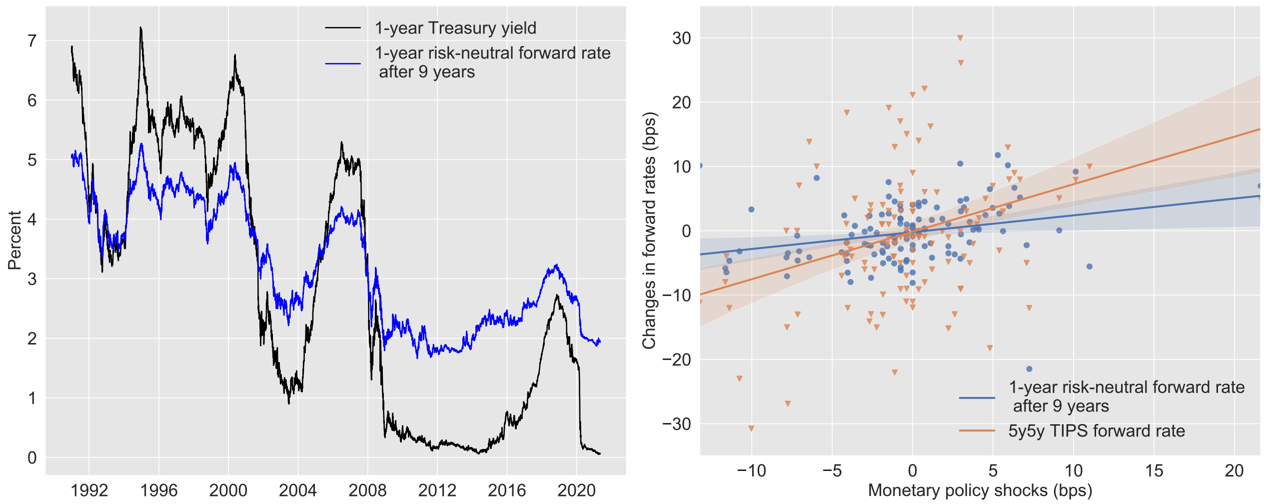
There is, however, strong evidence that monetary policy affects market expectations of interest rates over very long horizons. Hanson and Stein (2015) and Hanson et al. (2018) document that monetary policy news have a surprisingly strong effects on forward real interest rates in the distant future.[2] Hanson and Stein (2015) explain this by movements in term premia. But there is also evidence of expected short-term rate in the long future being sensitive to monetary policy. The left panel in Figure 1 plots the one-year nominal bond yield (black line), a proxy for the policy interest rate and its near-term outlook, against the forward real rate of the same maturity nine years ahead (blue line). This forward rate is constructed from a risk-neutral yield curve, where risk premia have been removed as in Adrian et al. (2013), and thus contains only the expected short-term interest rate component. Nine years are arguably long enough for cyclical shocks to dissipate and inflation to return to target in expectations. This long forward rate thus serves as a reasonable proxy for r-star. There is a high correlation between the two series, which is evidently driven by variations of interest rates over the monetary policy cycles. In each episode of persistent tightening or loosening of the policy interest rate, the forward rate follows suit. If r-star shocks had been responsible for moving the short rate and the long forward rate in tandem, one would expect a much weaker correlation of the two at the policy cycle frequency.

A better way of gauging the causal impact of monetary policy is to use high frequency-identified monetary policy shocks to examine how the long forward rate responds to monetary policy surprises immediately after FOMC meetings. The right panel in Figure 1 shows significant positive responses of long forward rates to monetary policy surprises.[3] In a recent paper, Hillenbrand (2022) documents that the change in 10-year nominal yields around FOMC meetings explains the *entire* decline in 10-year yields over the last thirty years. Monetary policy thus seems to impart a significant effect on the market expectations of steady-state interest rate.

There is also evidence that expectations about long-term rates do not conform to the

[2] Hanson and Stein (2015) estimate a regression of changes in forward interest rates on changes in 2-year nominal yields, on FOMC announcement dates, and find that a 100 bps change in 2-year nominal yields translates to a 40 bps change in real forward rate at the 10-year horizon.

[3] The positive response is stronger if one uses the 5-year 5-year real forward rates from the TIPS market, though part of this responsiveness may owe to the risk premium component as noted in Hanson and Stein (2015). At the same time, the result rules out excess sensitivity of long-run inflation expectations to monetary policy as an explanation.



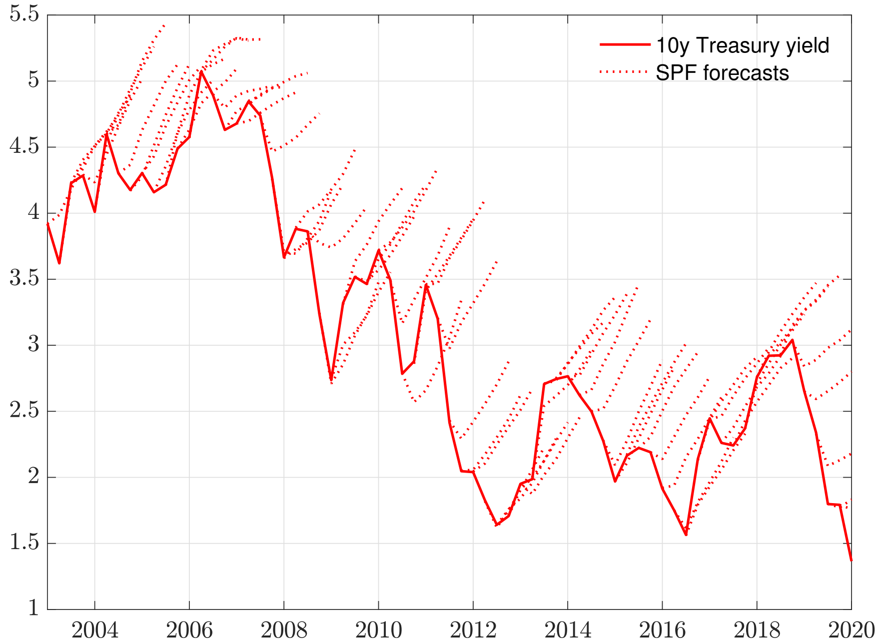
(a) Correlation between long and short rates. (b) Sensitivity of long rates to policy surprises.

Note: Left panel: The black line shows the nominal one-year US Treasury bond yield. The blue line shows the 9y1y forward interest rate. Right panel: Blue dots and fitted line trace the relationship between changes in the 9y1y-neutral forward interest rate around policy surprise events (2-day window), and the size of monetary policy shocks during those events. Orange dots and line use the 5y5y real forward interest rate from the TIPS market as the forward rate. Shaded areas are 90 percent confidence bands. The period covered in both cases is December 2003 to June 2019. Policy shock series are from Kearns et al. (2018). The forward rate in both panels is from the risk-neutral yield curve as in Adrian et al. (2013).

Figure 1: Relation of long-term rates and monetary policy.

rational expectations hypothesis, as has been documented previously in the literature (e.g. Coibion and Gorodnichenko, 2015). Figure 2 plots the time-series of 10-year US Treasury yield alongside its projection from the Survey of Professional Forecasters. The long-term interest rate declined continuously throughout the sample, by over 3 percentage points from its peak. Yet, analysts consistently expected the decline in long-term yields to reverse each time they were surveyed, leading to systematic forecast errors. These persistent errors appear all the more puzzling considering that some proposed r-star drivers, such as life expectancy or dependency ratios, follow slow-moving and predictable trends.

The model we propose provides a parsimonious joint explanation of these stylized facts, as well as the secular decline in real interest rates and the slow economic recovery after the GFC. To be clear, our hypothesis is not the only possible explanation of these facts. The excess sensitivity of long-term yields to monetary policy could stem from financial market participants being unduly attentive to short-term factors. Meanwhile, slow output recovery and falling real interest rates may have owed to other unrelated factors. What our model offers is a complementary and unified perspective tying each described phenomenon to a common cause. It can moreover be formalized within a standard workhorse macroeconomic



Note: The solid line is the 10-year Treasury yield, while dotted lines represent the projected paths of 10-year Treasury yield according to the Survey of Professional Forecasters. The start of each line marks the current yield as of the survey date, and is hence on the solid line by definition.

Figure 2: Trend decline in long-term yield was largely unforeseen

model and the usual natural interest rate concept, as the next section illustrates.

3 Macroeconomic environment

We now introduce the model and show how expectations are the determinant of the de facto natural interest rate. By influencing how agents make consumption and saving decisions, these expectations also dictate how inflation and the output gap respond to monetary policy.

3.1 The New Keynesian model with unobserved r-star

Our model is the standard New Keynesian model, but with incomplete information about stochastic trends in the rate of growth and the rate of time preference. A representative household solves the utility maximization problem

$$\max_{\{C_t,N_t,B_t\}_{t=0}^{\infty}} E_0^h \sum_{t=0}^{\infty} \beta^t \Xi_t \left(\frac{C_t^{1-\sigma}}{1-\sigma} - A_t^{1-\sigma} \frac{N_t^{1+\varphi}}{1+\varphi} \right)$$

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

by choosing consumption C_t, labor supply N_t and nominal bond holdings B_t which are in zero net supply and yield a nominal return of i_t. Consumption C_t is an aggregate of differentiated goods:

$$C_t \equiv \left(\int_0^1 C_{it}^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}}$$

which gives rise to a standard CES demand function. The household takes as given the price level P_t, the real wage W_t, dividends from firms Π_t and any lump-sum transfers from the government T_t. The utility function is affected by shocks to the rate of time preference Ξ_t. To ensure a balanced growth path with trend productivity shocks, the disutility of labor also depends on $A_t^{1-\sigma}$, where A_t is the aggregate productivity level.

Importantly, the expectation operator E^h is conditional on the information set of private agents, namely the household and firms. This information set is potentially incomplete and different from that of the central bank. The expectations also need not coincide with rational expectations once we introduce misperception later on.

Differentiated goods are produced by a continuum of firms $i \in [0, 1]$ with the technology

$$Y_{it} = A_t N_{it}^{1-\alpha}$$

and sold at the price P_{it} (whose CES sum over i equals P_t). Firms are subject to Calvo pricing frictions and can only re-optimize their prices with probability $1 - \theta$. Firms' revenues are subsidized at a rate τ_t. In steady state, this subsidy is set to the value $\bar{\tau}$ that ensures efficiency, while random fluctuations around this value act as cost-push shocks. Firms distribute their profits to the household.

The government consists of a central bank that sets the nominal interest rate i_t, and a fiscal authority that collects taxes on firms and distributes the proceeds lump-sum to the household.

Productivity growth is made up of permanent and temporary components (respectively a random walk and an iid process). Defining $a_t \equiv \log(A_t)$, we posit:

$$\begin{aligned} \Delta a_{t+1} &= g_t + \epsilon_{at+1}, & \epsilon_{at+1} &\sim \mathcal{N}\left(0, \sigma_a^2\right) \\ \Delta g_{t+1} &= \epsilon_{gt+1}, & \epsilon_{gt+1} &\sim \mathcal{N}\left(0, \sigma_g^2\right). \end{aligned}$$

The shock to the discount factor $\xi_t \equiv \log(\Xi_t)$ similarly consists of permanent and temporary components, but also a persistent part:[4]

$$\begin{aligned} \Delta\xi_{t+1} &= -z_t - u_{ht} - \epsilon_{\xi,t+1}, & \epsilon_{\xi,t+1} &\sim \mathcal{N}\left(0,\sigma_\xi^2\right) \\ \Delta z_{t+1} &= \epsilon_{zt+1}, & \epsilon_{z,t+1} &\sim \mathcal{N}\left(0,\sigma_z^2\right) \\ u_{ht+1} &= \rho_h u_{ht} + \epsilon_{ht+1}, & \epsilon_{h,t+1} &\sim \mathcal{N}\left(0,\sigma_h^2\right). \end{aligned}$$

A key departure from the standard setup arises from agents' incomplete information about the a_t and ξ_t processes. In particular, the household and firms can observe the current productivity level a_t and the current preference shifter ξ_t as well as u_{ht}. But they cannot separately observe the subcomponents $g_t, \epsilon_{at}, z_t, \epsilon_{\xi t}$. As a result, agents cannot disentangle movements in a_t and ξ_t that are attributable to the permanent components g_t and z_t from those that are due to the temporary shocks ϵ_{at} and $\epsilon_{\xi t}$.

3.2 The belief-driven natural interest rate

We now derive the natural interest rate with incomplete information. Log-linearizing the first-order conditions and solving the model leads to the familiar Euler equation:

$$E_t^h\left[\Delta y_{t+1}\right] = \frac{1}{\sigma}\left(i_t - E_t^h\left[\pi_{t+1}\right] + E_t^h\left[\Delta\xi_{t+1}\right]\right) \tag{3.1}$$

where $i_t - E_t^h\left[\pi_{t+1}\right]$ is the ex-ante real interest rate from the perspective of private agents. Evaluating this equation under flexible prices, where log output is at its natural level $y_t^* = a_t$ and $E_t^h\left[\Delta y_{t+1}^*\right] = E_t^h\left[g_t\right]$, one can back out out the corresponding level of the real interest rate under flexible prices as

$$E_t^h\left[\sigma g_t + z_t\right] + u_{ht}.$$

We define r-star as the private sector's expectation of this real interest rate in the long run. If trend growth g_t and the trend discount rate z_t were fully observed, then r-star would be:

$$r_t^{**} = \sigma g_t + z_t. \tag{3.2}$$

This notation follows Laubach and Williams (2003), but we denote this object "r-double-star" because in our setup it is not observable. One can think of r-double-star as the *fundamentals* driving long run real interest rates. These fundamentals are truly exogenous in our model. However, they are distinct from the *de facto* natural rate in the economy under incomplete

[4]Technically, ξ_t has to be bounded in order to guarantee that expected discounted utility remains finite. Imposing such bounds would introduce a non-linearity that would render the filtering problems in the model computationally prohibitive. We abstract from this constraint in what follows.

information, which equals:

$$r_t^* = E_t^h [\sigma g_t + z_t] = E_t^h [r_t^{**}] . \tag{3.3}$$

The *de facto* r-star is an expectation, and as such it is endogenous to changes in the private sector's information. It is this endogenous expectation that actually determines aggregate demand. Denoting the output gap by $\tilde{y}_t \equiv y_t - y_t^*$, one obtains the familiar IS curve:

$$E_t^h [\Delta \tilde{y}_{t+1}] = \frac{1}{\sigma} \left(i_t - E_t^h [\pi_{t+1}] - r_t^* - u_{ht} \right) . \tag{3.4}$$

The second equation of the linearized model is the Phillips curve, which also takes the standard form up to the expectation operator:

$$\pi_t = \beta E_t^h [\pi_{t+1}] + \kappa \tilde{y}_t + u_{pt} \tag{3.5}$$

where $\kappa > 0$ is a function of other primitive parameters (see Appendix A and Galí (2015) for detailed derivation). The cost-push shock $u_{pt} \sim \log \tau_t$, which is observed by private sector agents, is assumed to follow a normal AR(1) process with autocorrelation ρ_p and innovation variance σ_p^2.

We close the model by assuming that the central bank sets the nominal interest rate according to a standard Taylor-type rule with inertia:

$$i_t = \rho_i i_{t-1} + (1 - \rho_i) (\hat{r}_t^* + \phi_\pi \pi_t + \phi_y \tilde{y}_t + u_{ct}) . \tag{3.6}$$

The monetary policy shock u_{ct} is observed by the central bank but not by the private sector. It is assumed to follow a normal AR(1) process with autocorrelation ρ_c and innovation variance σ_c^2.

Like private agents, the central bank cannot directly observe r_t^{**} and must form an estimate to set policy:

$$\hat{r}_t^* \equiv E_t^c [r_t^{**}] \tag{3.7}$$

where E_t^c denotes the expectation with respect to the central bank's information set and inference. In general, these do not necessarily coincide with those of private agents, and hence E_t^c and E_t^h need not be identical.

This recast of the New Keynesian model to incomplete information yields two key insights:

1. The *de facto* natural rate of interest relevant to the economy, r_t^*, is belief-dependent. It is whatever private agents expect the long-run level of interest rates to be. It is only in the special case where private agents perfectly observe the fundamentals r_t^{**} (and understand the model correctly) that r_t^* is exogenous.

2. The *de facto* natural interest rate r^* is not necessarily the same as the estimate $\hat{r}^*$ used by the central bank to guide monetary policy. The two coincide only when the central bank and the private sector share the same beliefs, so that $E_t^c = E_t^h$.

In the next section, we let both the central bank and the private sector learn from each other through observing the macroeconomic outcomes, such that their beliefs evolve in an interdependent way.

4 The hall-of-mirror effect: Building intuition

In this section, we illustrate the hall-of-mirror effect under the simplest possible macroeconomic setting: a 2-period version of the New Keynesian model discussed above. This allows us to focus on the mutual learning problem and develop intuition for how the mechanism operates. The central insights carry over to the dynamic setting, which we deal with in the next section.

4.1 The 2-period model

Assume that the economy returns to full employment from period 1 onward, so that $E_0^h[\tilde{y}_t] = E_0^c[\tilde{y}_t] = 0$ for all $t \geq 1$, and that the Taylor rule has no inertia, $\rho_i = 0$. By the Phillips curve equation 5.8, both the central bank and households expect inflation to return to zero in period 1. Also assume that the cost-push shock u_{pt} is absent. The model then becomes effectively static, and can be summarized in terms of period-0 variables, omitting time subscripts:

$$\tilde{y} = -\frac{1}{\sigma}(i - r^* - u_h) \tag{4.1}$$

$$\pi = \kappa\tilde{y} \tag{4.2}$$

$$i = \hat{r}^* + \phi_\pi\pi + \phi_y\tilde{y} + u_c \tag{4.3}$$

where $r^* \equiv E^h[r^{**}]$ and $\hat{r}^* \equiv E^c[r^{**}]$. We assume that the fundamentals r^{**} and macroeconomic shocks are white noise:

$$r^{**} \sim \mathcal{N}(0, 1) \tag{4.4}$$

$$u_i \sim \mathcal{N}\left(0, \sigma_{ui}^2\right), \quad i = c, h. \tag{4.5}$$

The stochastic terms (r^{**}, u_c, u_h) are mutually independent. The prior on r^{**} has zero mean and unit variance without loss of generality.

Solving (4.1)–(4.3) gives

$$\tilde{y} = \frac{1}{\lambda}\left(r^* - \hat{r}^* + u_h - u_c\right) \tag{4.6}$$

$$i = \frac{\sigma}{\lambda}\left(\hat{r}^* + u_c\right) + \left(1 - \frac{\sigma}{\lambda}\right)\left(r^* + u_h\right) \tag{4.7}$$

where $\lambda = \sigma + \phi_\pi \kappa + \phi_y$. The output gap (and hence inflation) increases with the difference $r^* - \hat{r}^*$, because a higher r-star belief by the central bank implies a tighter monetary policy, all else equal. As a result, disagreement about r-star can cause the output gap to deviate from zero, even in the absence of demand and monetary policy shocks. Furthermore, the interest rate that prevails in equilibrium becomes a weighted average between the beliefs of the private sector r^* and those of the central bank $\hat{r}^*$.

To form beliefs about the natural interest rate, the private sector and the central bank rely on different sources of information. First, the private sector "h", and the central bank "c" each receives a signal about r^{**}:

$$s_i = r^{**} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}\left(0, \sigma^2_{\epsilon i}\right),\ i = c, h \tag{4.8}$$

where ϵ_c and ϵ_h are mutually independent.[5] The variance $\sigma^2_{\epsilon i}$ can be zero, which corresponds to i having full information about r^{**}; it can also be infinity, which corresponds to i having no private information about r^{**}. Each side can only observe their own signal.

The second information source comes from macroeconomic outcomes $\tilde{y}$, π and i. Observing these outcomes allows each side to extract information about the private signal of the other. The information content can be summarized easily by rearranging the equilibrium conditions (4.6) and (4.7) in terms of two sufficient statistics:

$$a_h \equiv E^h\left[r^{**}\right] + u_h = i - (\phi_y - \sigma)\tilde{y}. \tag{4.9}$$

$$a_c \equiv E^c\left[r^{**}\right] + u_c = i - (\phi_y + \phi_\pi \kappa)\tilde{y} \tag{4.10}$$

Here, a_c and a_h are noisy signals of $E^c[r^{**}]$ and $E^h[r^{**}]$ respectively. These endogenous signals are publicly observable through observations of the nominal interest rate and the output gap.

4.2 Inference problem

We can cast the mutual learning problem in terms of the two "agents" in our model—the private sector and the central bank—forming expectations about the random variable r^{**}

[5] In the fully dynamic model, we allow for correlated private signals as well as public signals about r-star.

conditional on (s_i, u_i, a_j) where $j \neq i$. The inference problem of agent i is non-trivial because a_j is endogenous to i's expectations.

Due to the Gaussian structure of the fundamentals and signals, beliefs in equilibrium will depend linearly on the signals and noises. We therefore conjecture, and subsequently verify, that agent i's belief of agent j's expectation takes the following linear form:

$$E^j\left[r^{**}\right] = \alpha_j s_j + \beta_j s_i + \gamma_j u_j + \delta_j u_i. \tag{4.11}$$

We solve agent i's signal extraction problem given this belief. Agent i's expectation is

$$E^i\left[r^{**}\right] = E\left[r^{**} \mid s_i, u_i, a_j\right]$$

where $a_j = E^j\left[r^{**}\right] + u_j$ with $E^j\left[r^{**}\right]$ given by (4.11). We can simplify this problem by transforming a_j into $\tilde{a}_j$ using a linear combination of agent i's observables:

$$\tilde{a}_j \equiv a_j - \beta_j s_i - \delta_j u_i. \tag{4.12}$$

Under agent i's beliefs, the vector $(r^{**}, s_i, \tilde{a}_j)'$ is normally distributed. The optimal filtering solution then obtains as

$$E^i\left[r^{**}\right] = g_{si} s_i + g_{ai} \tilde{a}_j \tag{4.13}$$

with the following gain parameters:

$$\begin{pmatrix} g_{si} \\ g_{ai} \end{pmatrix} = \frac{1}{\alpha_j^2\left(\sigma_{\epsilon i}^2 + \sigma_{\epsilon j}^2 + \sigma_{\epsilon i}^2 \sigma_{\epsilon j}^2\right) + \left(1+\gamma_j\right)^2 \sigma_{uj}^2\left(\sigma_{\epsilon i}^2 + 1\right)} \begin{pmatrix} \alpha_j^2 \sigma_{\epsilon j}^2 + \left(1+\gamma_j\right)^2 \sigma_{uj}^2 \\ \alpha_j \sigma_{\epsilon i}^2 \end{pmatrix}. \tag{4.14}$$

4.3 Common knowledge equilibrium

In an equilibrium with common knowledge, agent i's conjecture (4.11) coincides with the actual expectation formation of agent j. Substituting (4.11) and (4.12) into (4.13) yields:

$$E^i\left[r^{**}\right] = g_{si} s_i + g_{ai}\left(\alpha_j s_j + \left(1+\gamma_j\right) u_j\right). \tag{4.15}$$

Indeed, this has the functional form of the conjecture in (4.11). Comparing coefficients for $i = c, h$ yields the following equilibrium conditions:

$$\begin{pmatrix} \alpha_i \\ \beta_i \\ \gamma_i \\ \delta_i \end{pmatrix} = \begin{pmatrix} g_{si} \\ g_{ai}\alpha_j \\ 0 \\ g_{ai}\left(1+\gamma_j\right) \end{pmatrix}. \tag{4.16}$$

The equilibrium expectation under common knowledge is thus given by:

$$E^i\left[r^{**}\right] = g_{si}s_i + g_{ai}\left(g_{sj}s_j + u_j\right). \tag{4.17}$$

The equilibrium conditions (4.16) are a non-linear system of equations because g_{si} and g_{ai} depend on α_j through (4.14). The following proposition shows that an equilibrium always exists[6] and that the parameters of the reaction functions are bounded.

Proposition 1 (Common knowledge equilibrium)**.** *The equilibrium defined by* (4.14) *and* (4.16) *exists and satisfies* $0 \leq g_{si} \leq 1$ *and* $0 \leq g_{ai} < 1$*. Furthermore,* $g_{si} = 1$ *if and only if* $\sigma_{\epsilon i} = 0$*.*

Proof. See Appendix B.1 □

It is instructive to consider some special cases. The first arises when the private sector has perfect information, while the central bank has no direct source of information and must only rely on macroeconomic outcomes to infer r-star. This case underlies the empirical approach of filtering r-star with a macroeconomic model to gauge the neutral stance of monetary policy (e.g. Laubach and Williams, 2003, Holston et al., 2017 and others). In our setting, this situation is captured by $\sigma^2_{\epsilon h} = 0$ and $\sigma^2_{\epsilon c} = \infty$, yielding:

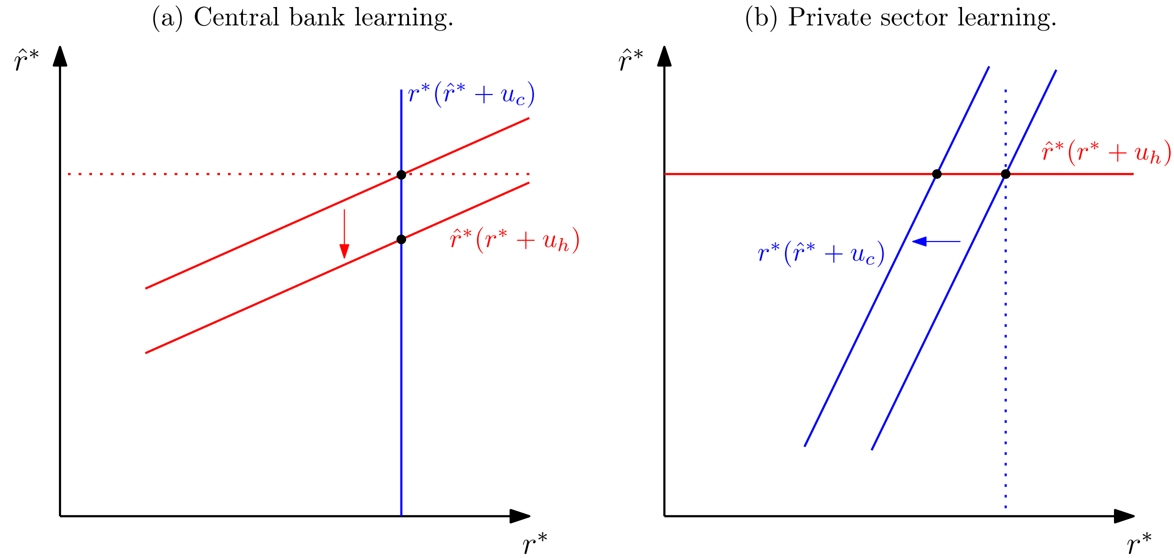
$$\text{Central bank learning:} \quad r^* = r^{**} \tag{4.18}$$

$$\hat{r}^* = g_{ac}\left(r^* + u_h\right). \tag{4.19}$$

The left panel of Figure 3 depicts the two reaction functions above. The red line traces the central bank's estimate $\hat{r}^*$ as a function of $= r^*$, and has a positive slope g_{ac}. Intuitively, when a_h increases, the central bank observes higher output and inflation for a given level of interest rate and revises its own estimate $\hat{r}^*$ upwards. Meanwhile, the blue line plots the *de facto* r-star r^* as a function of $\hat{r}^*$, a vertical line because the private sector is already perfectly

[6]We can also rule out "nonfundamental" equilibria in which expectations would not conform to the form conjectured in (4.11) and instead coordinate on a sunspot variable (Benhabib et al., 2015; Chan, 2020).

Figure 3: Determination of expectations with one-sided learning.



Note: Each panel shows the central bank's estimate of r-star $\hat{r}^* \equiv E^c[r^{**}]$ as a function of the noisy observation of the private sector expectation $a_h \equiv r^* + u_h$ (red lines), and the private sector's *de facto* r-star $r^* \equiv E^c[r^{**}]$ as a function of the noisy observation of the central bank expectation $a_c \equiv \hat{r}^* + u_c$ (blue lines).

informed about r^{**}. Note how the central bank's estimate $\hat{r}^*$ fluctuates with cyclical shocks. In the figure, a negative demand shock u_h shifts the red line down, resulting in lower $\hat{r}^*$. As in Laubach and Williams (2003), the central bank cannot readily distinguish between cyclical and permanent economic forces, and as a result assigns some weight to the possibility of a reduction in the natural interest rate.

The second special case is the reverse of the second: The central banks has perfect information, while the private sector has no direct information and has to rely on the central bank's policy actions to infer r-star. This situation gives rise to the signaling channel of monetary policy (Nakamura and Steinsson, 2018).[7] In this instance, we have $\sigma^2_{\epsilon h} = \infty$ and $\sigma^2_{\epsilon c} = 0$, yielding:

$$\text{Private sector learning:} \qquad r^* = g_{ah}\left(\hat{r}^* + u_c\right) \tag{4.20}$$

$$\hat{r}^* = r^{**} \tag{4.21}$$

The right panel of Figure 3 plots these reaction functions. In this case, the reaction function $\hat{r}^*$ is flat, as the central bank forms expectation independently. Meanwhile, the schedule r^* has a positive slope $1/g_{ah}$. Intuitively, when a_h increases, the private sector observes higher interest rates for given levels of output and inflation and revises up its own beliefs of long-

[7]This special case also applies if the central bank has imperfect information, but can perfectly observe the private sector's expectation. This situation arises in Hillenbrand (2022).

term interest rates. This *de facto* r-star now becomes endogenous to monetary policy shocks. In the figure, a negative interest rate shock u_c shifts the blue line to the left, leading to a fall in r^*. The private sector cannot readily differentiate between monetary policy shocks and changes to the central bank's information about r-star, and thus assigns some weight to the possibility that the natural real interest rate has declined.

The general case, where both agents learn about the determinants of r-star from each other, is unexplored in the literature as of yet. Under common knowledge, we can rearrange the equilibrium expression (4.15) to write agent i's expectation as a function of her observables:

$$E^i\left[r^{**}\right] = \left(1 - g_{ai}g_{aj}\right)g_{si}s_i - g_{ai}g_{aj}u_i + g_{ai}\left(E^j\left[r^{**}\right] + u_j\right). \tag{4.22}$$

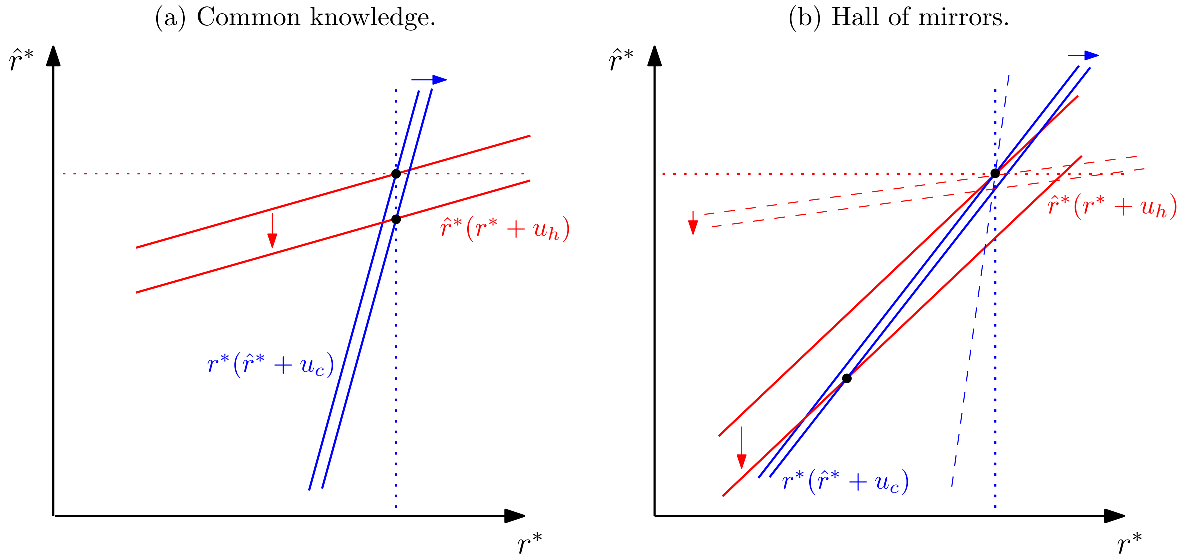
This general case is depicted in the left panel of Figure 4, where the *de facto* r^* of the private sector and the central bank's estimate $\hat{r}^*$ are both increasing functions of each other. In this case, both sides have useful private information about r-star and try to learn from each other. The r-star beliefs of both sides now depend on cyclical shocks. In the figure, a negative demand shock u_h shifts the red line down, as the imperfectly informed central bank assigns some weight to the possibility that r-star has fallen. At the same time, the private sector observes the demand shock and knows that it has no bearing on r-star. It rationally corrects its reaction function in anticipation of the decline in $\hat{r}^*$: The blue line shifts slightly to the right. In equilibrium, r^* is unchanged, as shown in equation (4.15). Thus, only the central bank's expectations are affected by cyclical demand shocks.

The equilibrium under common knowledge illustrates how an agent can misconstrue an unobserved cyclical perturbation as a r-star shock. But the informational feedback loop is limited because each side correctly understands the other's reaction function. Common knowledge of the informational environment thus limits the informational feedback loop.

4.4 Hall-of-mirrors equilibrium

We now consider a case where agents misperceive the quality of information about the determinants of r-star. Specifically, they are overly optimistic about the information that they can only observe indirectly: The central bank overestimates the private sector's knowledge of economic fundamentals, and the private sector likewise overestimates the central bank's knowledge. As we show, this kind of misperception has two effects: First, agents *overestimate* the amount of information contained in observable macroeconomic outcomes, and thus pay too much attention to them when they form their beliefs. Second, agents *underestimate* how much attention the other agents pay to macroeconomic outcomes,

Figure 4: Determination of expectations with two-sided learning.



Note: Each panel shows the central bank's estimate of r-star $\hat{r}^* \equiv E^c[r^{**}]$ as a function of the noisy observation of the private sector expectation $a_h \equiv r^* + u_h$ (red lines), and the private sector's *de facto* r-star $r^* \equiv E^c[r^{**}]$ as a function of the noisy observation of the central bank expectation $a_c \equiv \hat{r}^* + u_c$ (blue lines).

and thus fail to internalize how much their own actions influence the actions of others. As a result, everyone's actions end up reinforcing their incorrect subjective beliefs, generating a positive feedback loop that distorts the exchange of information and amplifies noise.

As an example, suppose that the central bank revises its r-star estimate downward and reduces interest rates. The private sector reacts by strongly lowering its own estimate of r-star because it overestimates the precision of the central bank's information about the economy. Output and inflation fall. Because the central bank is ignorant to the private sector's misperception and itself overestimates the precision of the private sector's information, it interprets this demand shortfall as a further indication that r-star has fallen and further lowers its own estimate. But in reality, the central bank is merely reacting to a reflection of its own initial revision. Worse, the ensuing further reduction in interest rates prompts the private sector to lower its own r-star beliefs a second time, even though it was entirely prompted by the initial reduction in aggregate demand. Both sides end up misperceiving reactions to their own actions as genuine information: They are staring into a hall of mirrors.[8]

We postulate that each agent $i = c, h$ has a subjective beliefs $\sigma_{\epsilon j}^{|i} < \sigma_{\epsilon j}$ about the noise in the private signals of the other side. As before, we conjecture that agent i's belief of agent j's expectation takes the form in (4.11), so that the solution to her signal extraction problem

[8] Thinking of the equilibrium as a sequence of revisions and reactions is a useful but only narrative device. In the model, the equilibrium is the fixed point of agents' reaction functions and the revisions are all realized within one period.

is still given by Equations (4.13) and (4.14). In addition, agent i mistakenly believes that her beliefs are shared by agent j, and that the economy will be in a common knowledge equilibrium. Thus, she also believes that agent j's expectation in (4.11) follow (4.16), but where the coefficients are given by the values that would obtain if agent j had the same beliefs about the private signal precision as agent i herself.[9]

In sum, the solution of agent i's filtering problem in this misperception equilibrium is represented by the following modification of (4.22):

$$\hat{E}^i\left[r^{**}\right]=\left(1-g_{ai}^{|i}g_{aj}^{|i}\right)g_{si}^{|i}s_i-g_{ai}^{|i}g_{aj}^{|i}u_i+g_{ai}^{|i}\left(\hat{E}^j\left[r^{**}\right]+u_j\right) \tag{4.23}$$

Here, $\hat{E}^i\left[r^{**}\right]$ denotes the expectation in the misperception equilibrium, and superscripts i on the gain parameters denote the subjective beliefs of agent i about the gain parameters of either agent. The subjective beliefs of these parameters under misperception are given by a modification[10] of (4.14):

$$\begin{pmatrix} g_{si}^{|i} \\ g_{ai}^{|i} \end{pmatrix}=\frac{1}{\left(g_{sj}^{|i}\right)^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^{2|i}+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^{2|i}\right)+\sigma_{uj}^2\left(\sigma_{\epsilon i}^2+1\right)}\begin{pmatrix}\left(g_{sj}^{|i}\right)^2\sigma_{\epsilon j}^{2|i}+\sigma_{uj}^2 \\ g_{sj}^{|i}\sigma_{\epsilon i}^2\end{pmatrix} \tag{4.24}$$

$$\begin{pmatrix} g_{sj}^{|i} \\ g_{aj}^{|i} \end{pmatrix}=\frac{1}{\left(g_{si}^{|i}\right)^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^{2|i}+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^{2|i}\right)+\sigma_{ui}^2\left(\sigma_{\epsilon j}^{2|i}+1\right)}\begin{pmatrix}\left(g_{si}^{|i}\right)^2\sigma_{\epsilon i}^{2}+\sigma_{ui}^2 \\ g_{si}^{|i}\sigma_{\epsilon j}^{2|i}\end{pmatrix} \tag{4.25}$$

The first equation describes the solution of the filtering problem of agent i under misperception, while the second line describes agent i's perceived solution of the filtering problem of agent j. Agent j herself also solves her own problem and guesses agent i's solution according to the same formula.

In the equilibrium with misperception, neither agent has the correct beliefs about how fundamentals and beliefs are related. Equation (4.23) relates the equilibrium beliefs of agent $i=c$ to those of agent $j=h$, and those of agent $i=h$ to those of agent $j=c$. This system of two linear equations allows to solve for the equilibrium expectations of both agents, leading to the following expression:

[9] Unlike in Caballero and Simsek, 2021, agents in our model do not "agree to disagree": They are unaware that they disagree about the distribution of information in the economy, and mistakenly attribute all their observable disagreements to differences in private information.

[10] Relative to (4.14), we have substituted α_j with $g_{sj}^{|i}$, using the equilibrium relation (4.16) as perceived by agent i; and, switching around the roles of i and j in (4.14), we have substituted α_i with $g_{si}^{|i}$.The variances of ϵ_i and $\epsilon_j$as well as the gain parameters carry superscripts to denote the subjective beliefs of agent i.

$$\hat{E}^i[r^{**}] = g_{si}^{|i} s_i + g_{ai}^{|i}\left(g_{sj}^{|j} s_j + u_j\right)$$
$$+ \frac{g_{ai}^{|i}}{1 - g_{ai}^{|i} g_{aj}^{|j}}\left[\left(g_{aj}^{|j} - g_{aj}^{|i}\right)\left(g_{si}^{|i} s_i + u_i\right) + \left(g_{ai}^{|i} - g_{ai}^{|j}\right) g_{aj}^{|j}\left(g_{sj}^{|j} s_j + u_j\right)\right] \tag{4.26}$$

The equilibrium beliefs differ from beliefs under common knowledge (4.17) in two ways. First, the gain parameters in the first line of (4.26) differ because agents misjudge the informativeness of signals.

Second, each side also misperceives the reaction function of the other, which gives rise to the terms on the second line of (4.26). Consider a temporary positive demand shock $u_h > 0$. The private sector ($i = h$) does not react to this shock, knowing that it is unrelated to r-star. The expectation of the central bank ($j = c$), however, increases by $g_{aj}^{|j} u_i$ and the policy rate rises by the same amount. Because the private sector misperceives the central bank's reaction function, it only expects the policy rate to rise by $g_{aj}^{|i} u_i$, resulting in a surprise of $\Delta = \left(g_{aj}^{|j} - g_{aj}^{|i}\right) u_i$. This surprise is positive when agents over-estimate the central bank's signal quality. The initial misperception is then subject to a multiplier effect: The private sector adjusts its expectation by $g_{ai}^{|i}\Delta$. The central bank, seeing this change, adjusts its expectation further by an amount $g_{aj}^{|j} g_{ai}^{|i}\Delta$, resulting in a further adjustment $g_{ai}^{|i} g_{aj}^{|j} g_{ai}^{|i}\Delta$ of the private sector and so on. In equilibrium, the initial error Δ is multiplied by the term $g_{ai}^{|i}/\left(1 - g_{ai}^{|i} g_{aj}^{|j}\right)$ in (4.26).

This multiplier embodies what we call the *hall-of-mirrors effect.* Its strength depends on how much attention agents are paying to each other's expectations in forming their own beliefs and grows unbounded with the degree of misperception.

Proposition 2 ([Hall-of-mirrors effect)**.** *If $\sigma_{\epsilon j}^{|i}$ is sufficiently small, then:*

1. *Beliefs overreact to demand and monetary policy shocks u_h and u_c relative to the common knowledge equilibrium: $g_{ai}^{|i} > g_{ai}$ and $g_{aj}^{|j} > g_{aj}^{|i}$.*
2. *This overreaction can be arbitrarily large: $g_{ac}^{|c}$ and $g_{ah}^{|h}$ can be arbitrarily close to one.*

Proof. See Appendix B.2. □

The hall-of-mirrors effect is graphically illustrated in the right panel of Figure 4. The slope of the central bank's reaction function (red line) is $g_{ac}^{|c}$ and the slope of the private sector's reaction function (blue line) is $1/g_{ah}^{|h}$. In the hall-of-mirrors equilibrium, both slopes are close to one. After a negative demand shock u_h, the central bank's reaction function shifts down. The private sector's reaction function shifts to the right, but unlike in the common knowledge case the shift is insufficient to offset the decrease in r-star expectations. The reason is that the private sector mistakenly thinks that the central bank has very good private information and will not react much to private sector expectations. This misperceived reaction function is represented by the red dashed line in Figure 4, which has slope $g_{ac}^{|h} < g_{ac}^{|c}$. The private sector adjusts its own expectation to this perceived reaction function. In equilibrium, however, the central bank pays a lot of attention to private sector expectations. The result is a decrease in r-star expectations by both sides that can in principle become arbitrarily large.

4.5 Macroeconomic implications

We now turn to the macroeconomic implications of imperfect knowledge and misperception. Recall that (4.6) relates the output gap $\tilde{y}$ (and inflation $\pi = \kappa\tilde{y}$) to the macroeconomic shocks as well as the difference of r-star beliefs between the household and the central bank:

$$\tilde{y} = \frac{1}{\lambda}\left(r^* - \hat{r}^* + u_h - u_c\right).$$

This difference in beliefs depends in turn on the macroeconomic shocks as well as the signals:

$$r^* - \hat{r}^* = b_h g_{sh}^{|h} s_h - b_c g_{sc}^{|c} s_c - (1 - b_h)\, u_h + (1 - b_c)\, u_c \tag{4.27}$$

$$\text{with } b_i = \left(1 - g_{aj}^{|j}\right) \frac{1 - g_{ai}^{|i} g_{aj}^{|i}}{1 - g_{ai}^{|i} g_{aj}^{|j}}.$$

The following proposition shows that, while expectations can overreact to shocks due to the hall-of-mirrors effect, the output gap and inflation underreact to shocks.

Proposition 3. *If $\sigma_{\epsilon h}^{|c}$ and $\sigma_{\epsilon h}^{|i}$ are sufficiently small, then the difference of r-star beliefs $r^* - \hat{r}^*$ comoves negatively with demand shocks u_h and positively with policy shocks u_c: $0 < b_c, b_h < 1$. Moreover, the effects of these shocks on the output gap and inflation are dampened relative to the full information case where $b_h = b_c = 1$.*

Proof. See Appendix B.3. □

As an illustration, consider again a negative demand shock $u_h < 0$. The direct effect of this shock is to lower the output gap. The indirect informational effect is that the central bank revises down its r-star estimate $\hat{r}^*$ as it sees demand falling. The private sector will also revise down its estimate r^*, but by less than the central bank. Even though the economy has weakened due to lower r^*, the more accommodative policy due to lower $\hat{r}^*$ more than offsets this weakness, so that output and inflation fall less than under full information.[11]

[11] This output-dampening effect would reverse once policy is constrained by the ELB, as the interest rate

By the same token, misperception weakens the transmission of monetary policy accommodation to output. As usual, a negative interest rate shock $u_c < 0$ has a direct expansionary effect on the output gap. But the indirect learning effect induces a fall in the term $r^* - \hat{r}^*$, as the private sector revises down their r-star belief by more than the central bank. This has a negative impact on output as the central bank lowers the interest rate by less than needed to offset the shock. As a result, the shock stimulates aggregate demand less than under full information.

While incomplete information and the hall-of-mirrors effect tend to dampen movements in output and inflation, they also amplify movements in expectations of r-star and interest rates. An outside observer may conclude that the structural relation between interest rates and real activity is weak. However, this pattern in our model is entirely consistent with a standard Euler equation and particular correlations between r-star beliefs and cyclical shocks.

5 Dynamic model

We now turn to the full, dynamic version of our model. The information structure and the associated qualitative equilibrium results all carry over from the static model presented above. To make the model more general, we add a public source of information observable by all agents, and also allow for a general autocorrelation structure of economic fundamentals and signal noise. We will show that the hall-of-mirrors effect not only amplifies noise to r-star beliefs, but also generates misperception that can be very persistent under plausible parameters.

5.1 Fundamentals and exogenous signals

In the dynamic model, the central bank and the private sector need to form expectations of the fundamental determinants of real interest rates, $r_t^{**} = \sigma g_t + z_t$, which forms a random walk process

$$r_t^{**} = r_{t-1}^{**} + v_t, \; v_t \sim \mathcal{N}\left(0, \sigma_r^2\right) \tag{5.1}$$

with $\sigma_r^2 = \sigma^2 \sigma_g^2 + \sigma_z^2$.

At the start of each period t, the central bank and the private sector $(i = c, h)$ receive

could no longer fall to compensate for a lower r-star. The interaction of the hall-of-mirrors effect and the ELB is an interesting question which we leave to future research.

privately observed signals s_{it} about the fundamentals:

$$s_{it} = r_t^{**} + e_{it} \tag{5.2}$$

$$e_{it} = \rho_{ei} e_{it-1} + \epsilon_{it},\ \epsilon_{it} \sim \mathcal{N}\left(0, \sigma_{\epsilon i}^2\right). \tag{5.3}$$

In addition, both observe a public signal x_t of the same form:

$$x_t = r_t^{**} + f_t \tag{5.4}$$

$$f_t = \rho_f f_{t-1} + \eta_t,\ \eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right). \tag{5.5}$$

Apart from these signals, there are three transient macroeconomic shocks in the model: The demand shock u_{ht}, the cost-push shock u_{pt} and the monetary policy shock u_{ct}. Each follows an AR(1) process:

$$u_{kt} = \rho_k u_{kt-1} + \nu_{kt},\ \nu_{kt} \sim \mathcal{N}\left(0, \sigma_{uk}^2\right).$$

The private sector is assumed to observe the demand and cost-push shocks u_{ht} and u_{pt} but not the policy shock u_{ct}. Meanwhile, the central bank observes the policy shock u_{ct}, but not the demand and cost-push shocks u_{ht} and u_{pt}.

We collect the vector of exogenous states in $Z_t = (r_t^{**}, e_{ht}, e_{ct}, f_t, u_{ht}, u_{pt}, u_{ct})'$ and the vector of exogenous shocks in $q_t = (v_t, \epsilon_{sht}, \epsilon_{ct}, \eta_t, \nu_{ht}, \nu_{pt}, \nu_{ct})'$. Then we can write

$$Z_t = A_z Z_{t-1} + q_t,\ q_t \sim \mathcal{N}\left(0, \Sigma_q\right) \tag{5.6}$$

with $A_z = \text{diag}\left(1, \rho_{e1}, \rho_{e2}, \rho_f, \rho_{uh}, \rho_{up}, \rho_{uc}\right)$ and $\Sigma_q = \text{diag}\left(\sigma_r^2, \sigma_{\epsilon 1}^2, \sigma_{\epsilon 2}^2, \sigma_\eta^2, \sigma_{uh}^2, \sigma_{up}^2, \sigma_{uc}^2\right)$.

5.2 Macroeconomic outcomes and endogenous signals

The private sector determines inflation and output according to the system of equations consisting of (3.4)–(3.6). We write this system of equations entirely from the perspective of the private sector information set:

$$\tilde{y}_t = E_t^h\left[\tilde{y}_{t+1}\right] + \frac{1}{\sigma} E_t^h\left[\pi_{t+1} + r_t^{**}\right] + \frac{1}{\sigma}\left(u_{ht} - i_t\right) \tag{5.7}$$

$$\pi_t = \beta E_t^h\left[\pi_{t+1}\right] + \kappa \tilde{y}_t + u_{pt} \tag{5.8}$$

$$i_t = \rho_i i_{t-1} + (1 - \rho_i)\left(E_t^h E_t^c r_t^{**} + \phi_\pi \pi_t + \phi_y \tilde{y}_t + E_t^h\left[u_{ct}\right]\right) \tag{5.9}$$

The private sector observes the nominal interest rate i_t in addition to its private signal s_{ht} and the public signal x_t, and it also observes the demand and cost-push shocks u_{ht} and u_{pt}. But it does not separately observe $E_t^c r_t^{**}$ and u_{ct} in the policy rule (5.9) and has to estimate these objects.

The central bank observes current inflation π_t and the output gap $\tilde{y}_t$ in addition to its private signal s_{ct} and the public signal x_t. It determines the nominal interest rate i_t according the Taylor rule as a function of inflation, the output gap, its current-period estimate of r-star, as well as the monetary policy shock u_{ct}. We assume that the coefficients ϕ_π and ϕ_y in the monetary policy rule yield a unique solution to (5.7)–(5.9), given expectations of the exogenous fundamentals.

5.3 Solution with common knowledge

Solving the dynamic model requires keeping track of higher-order beliefs explicitly. We define the "zero-th order beliefs" as the true fundamentals: $E_{it}^{(0)}[Z_t] = Z_t$. The first-order beliefs of agent $i = c, h$ are her expectations about the fundamentals: $E_{it}^{(1)}[Z_t] = E_{it}[Z_t]$. For $n \geq 1$, her $n+1$-th order belief is defined as the belief about the n-th order belief of agent j: $E_{it}^{(n+1)}[Z_t] = E_{it}\left[E_{jt}^{(n)}[Z_t]\right]$, $j \neq i$.

For each agent $i = c, h$, we denote with X_{it} the states that agent i does not observe, which are the beliefs of all orders $n = 0, 1, 2, \ldots$ of Z_t of the other agent j:

$$X_{it} = \left(E_{jt}^{(n)} Z_t\right)_{n=0}^{\infty}. \tag{5.10}$$

Agent i has to form beliefs about X_{it}. She enters period t with a prior belief about X_{it-1} which is distributed as $\mathcal{N}(m_{it-1}, P_i)$.[12] She then observes her own private signal s_{it}, the public signal x_t, as well as the macroeconomic outcomes $(\tilde{y}_t, \pi_t, i_t)$. Additionally, the central bank observes u_{ct}, while the private sector observes u_{ht} and u_{pt}.

Agent i's posterior belief takes the form

$$X_{it} \mid i, t-1 \sim \mathcal{N}(m_{it}, P_i). \tag{5.11}$$

To characterize this belief, we solve the signal extraction problem using a conjecture on the other agent's belief m_{jt}. We guess, and later verify, that agent j's belief evolves according to:

$$m_{jt} = \Phi_j m_{jt-1} + \Psi_j m_{it-1} + \Omega_j Z_t. \tag{5.12}$$

[12] We assume that enough time has passed for the prior variance to reach its time-invariant level, in keeping with much of the literature.

We can rewrite the state X_{it} that player i has to learn about in a recursive form:

$$X_{it} = A_i X_{it-1} + B_i q_t + C_i m_{it-1}. \tag{5.13}$$

where the matrices A_i, B_i and C_i depend on the guess in (5.12) as well as A_z in (5.6). The exact expressions are provided in Appendix C.

We also guess, and later verify, that the signals agent i receives in period t are a linear combination of the current state:

$$Y_{it} = H_i X_{it}. \tag{5.14}$$

Equations (5.13) and (5.14) form a standard linear filtering problem, the solution of which is given by the Kalman filter. The optimal filtering equation describing the evolution of beliefs is:

$$m_{it} = (I - G_i H_i)(A_i + C_m \Psi_j) m_{it-1} + G_i Y_{it} \tag{5.15}$$

The Kalman gain G_i, as well as the time-invariant posterior covariance matrix P_i, can be computed using standard formulas, also detailed in Appendix C.

We can now find the equilibrium with common knowledge and verify our conjectures (5.12) and (5.14). In a common knowledge equilibrium, agent i's beliefs of (5.12) and of the signal matrix H_i are correct. We can thus express the vector of signals Y_{it} in terms of past beliefs and the current state:

$$\begin{aligned} Y_{it} &= H_i (C_z Z_t + C_m m_{jt}) \\ &= H_i ((C_z + C_m \Omega_j) Z_t + C_m \Phi_j m_{jt-1} + C_m \Psi_j m_{it-1}). \end{aligned} \tag{5.16}$$

Substituting this expression into (5.15) gives an expression for m_{it} that verifies our guess (5.12). The equilibrium coefficients can be found using the following system of equations:

$$\Phi_i = (I - G_i H_i) A_i + C_m \Psi_j \tag{5.17}$$

$$\Psi_i = G_i H_i C_m \Phi_j \tag{5.18}$$

$$\Omega_i = G_i H_i (C_z + C_m \Omega_j). \tag{5.19}$$

Finally, we need to compute the observation matrices H_i. This step is more involved than in the static model because the endogenous signals provided by observations of macroeconomic outcomes now depend on the fundamentals as well as first- and higher-order beliefs. For example, the output gap, which the central bank uses as an endogenous signal about r-star, depends on the private sector's expectation of the future interest rate path, which depend on its expectations about the central bank's expectation of r-star. Appendix C solves for the macroeconomic outcomes of the model as a function of private sector beliefs, and shows that the signal matrix for the central bank H_c is a function of the macroeconomic model parameters, as well as the matrices $A_z, \Phi_h, \Psi_h, \Omega_h$. The signal matrix for the private sector H_h does not depend on other parameters in the model.

To compute the equilibrium numerically, we use the following iterative algorithm:

1. Start with an initial guess $(\Phi_i, \Psi_i, \Omega_i, H_i)_{i=c,h}$.
2. For $i = c, h$:
   (a) compute the law of motion for X_{it} from (C.4);
   (b) compute the Kalman matrices P_i^-, S_i, G_i, P_i from (C.6)–(C.9);
   (c) compute Φ_i, Ψ_i, Ω_i from (5.17)–(5.19).
3. Compute the signal matrix H_c according to (C.11). The matrix H_h stays the same across iterations.
4. Iterate on steps 2. and 3. until convergence.

For our computations, we have to truncate the infinite sequence of higher-order beliefs contained in m_{it} to some finite level N. Our numerical results show that when N is sufficiently large, the choice of N does not affect the equilibrium dynamics.

5.4 Misperception equilibrium

As in the static version of our model, we can compute the corresponding equilibrium with misperception. In this case, each agent $i = c, h$ has own beliefs about the properties A and Σ_q of the fundamentals and/or signals Z_t in (5.6). We denote these beliefs with $A^{|i}$ and $\Sigma_q^{|i}$. For the purpose our simulations, we only consider the "hall-of-mirrors" case in which each agent believes that $\sigma_{sj}^{|i} < \sigma_{sj}$. Furthermore, agent i believes that the other agent $j = h, c$ shares her own beliefs about the fundamentals. That is, both agents mistakenly assume common knowledge of their own beliefs about the fundamentals, when in fact they disagree about them.

To solve agent i's perceived law of motion of beliefs, we first solve a common knowledge equilibrium where we substitute subjective beliefs $A^{|i}$ and $\Sigma_q^{|i}$ for the true values A and Σ_q, respectively. This solution yields a perceived law of motion (4.11) for m_{kt} with coefficients $\Phi_k^{|i}$, $\Psi_k^{|i}$ and $\Omega_k^{|i}$, $k = c, h$, as well as perceived gains $G_k^{|i}$ and signal matrices $H_k^{|i}$.

To proceed from the perceived law of motions to the equilibrium, we then write the filtering equation (5.15) as:

$$\begin{aligned} m_{it} &= \left(\Phi_i^{|i} - G_i^{|i} H_i^{|i} C_m \Psi_j^{|i}\right) m_{it-1} + G_i^{|i} Y_{it} \\ &= \left(\Phi_i^{|i} - G_i^{|i} H_i^{|i} C_m \Psi_j^{|i}\right) m_{it-1} + G_i^{|i} H_i^{|j} \left(C_z Z_t + C_m m_{jt}\right). \end{aligned} \tag{5.20}$$

To obtain the second line, we substitute out the actual signals Y_{it} that agent i receives, which depend on the expectations of agent j and are hence given by $Y_{it} = H_i^{|j} X_{it}$. The above expression also holds when the roles of i and j are reversed, and we can use this fact to substitute out m_{jt}. We then obtain the actual law of motion of beliefs describing the equilibrium under misperception:

$$m_{it} = \hat{\Phi}_i m_{it-1} + \hat{\Psi}_i m_{jt-1} + \hat{\Omega}_i Z_t \tag{5.21}$$

where the actual transition coefficients are given by:

$$\hat{\Phi}_i = \left(I - G_i^{|i} H_i^{|j} C_m G_j^{|j} H_j^{|i} C_m\right)^{-1} \left(\Phi_i^{|i} - G_i^{|i} H_i^{|i} C_m \Psi_j^{|i}\right) \tag{5.22}$$

$$\hat{\Psi}_i = \left(I - G_i^{|i} H_i^{|j} C_m G_j^{|j} H_j^{|i} C_m\right)^{-1} G_i^{|i} H_i^{|j} C_m \left(\Phi_j^{|j} - G_j^{|j} H_j^{|j} C_m \Psi_i^{|j}\right) \tag{5.23}$$

$$\hat{\Omega}_i = \left(I - G_i^{|i} H_i^{|j} C_m G_j^{|j} H_j^{|i} C_m\right)^{-1} G_i^{|i} H_i^{|j} \left(I + C_m G_j^{|j} H_j^{|i}\right) C_z. \tag{5.24}$$

The macroeconomic outcomes in the misperception equilibrium are again determined by the private sector's beliefs, and we relegate this formula to Appendix C.

6 Simulation results

We will now use a calibrated version of our dynamic model to explore the potential quantitative implications of the hall-of-mirrors effect. How much can cyclical shocks to aggregate demand and monetary policy affect the *de facto* r-star? As the preceding analysis suggests, the answer will depend in part on the degree of misperception. Our benchmark calibration will imply a large degree of misperception, but we also document the sensitivity of our results to the information parameters including misperception.

To conduct our simulations, we calibrate the model parameters as in Table 1. Macroeconomic parameters are standard in the literature (e.g. Billi, 2011). The standard deviation of changes to r-star fundamentals is set to 0.05 percent quarterly, in line with the estimates by Holston et al. (2017). The public signal about these fundamentals is

Table 1: Calibrated parameters.

Parameter	Symbol	Value	Parameter	Symbol	Value
Inverse EIS	σ	6	Initial value of r_t^{**}	r_0^{**}	2.4 %
Phillips curve slope	κ	0.015	S.d. of r_t^{**} shock	σ_r	0.05
Discount factor	β	0.9941	Steady-state inflation	π^*	2 %
Rule coefficient on inflation	ϕ_π	1.5	Autocorr. of cost-push shock	$\rho_{u\pi}$	0.8
Rule coefficient on output gap	ϕ_y	0.125	S.d. of cost-push shock	$\sigma_{u\pi}$	0.1
Rule coefficient on lagged rate	ρ_i	0.7	Autocorr. of public signal noise	ρ_f	0
Autocorr. of policy shock	ρ_{uc}	0.7	Autocorr. of private signal noise	ρ_{ei}	0
S.d. of policy shock	σ_{uc}	0.1	S.d. of public signal noise	σ_η	3
Autocorr. of demand shock	ρ_{uh}	0.8	S.d. of private signal noise	$\sigma_{\epsilon i}$	∞
S.d. of demand shock	σ_{uh}	0.2	Perceived —	$\sigma_{\epsilon i}^{\vert j}$	0.2

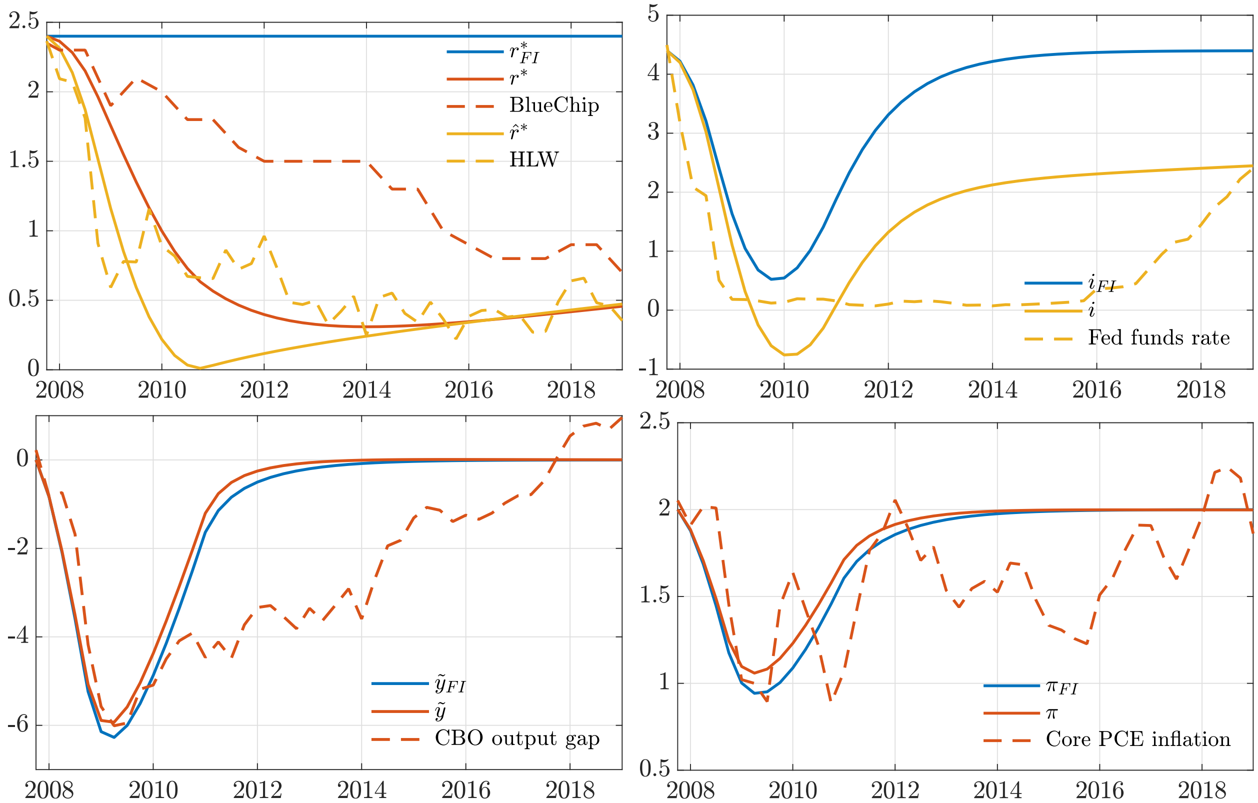
assumed to be quite noisy, with a standard deviation of 3, and true private information is assumed to be absent ($\sigma_{\epsilon i} = \infty$). The resulting uncertainty about r-star in the absence of misperception corresponds to 90% confidence intervals of subjective r-star beliefs of ±2.5%, which is large but within the range of empirical studies.[13] In addition, we assume a large degree of misperception: Despite there being no useful private information at all, each side believes that the other side has valuable private information ($\sigma_{\epsilon i}^{|j} = 0.2$). The resulting uncertainty about r-star corresponds to a 90% confidence interval of ±0.9% for the private sector, and ±1.4% for the central bank, which is again close to the empirical estimates by Holston et al..

6.1 A demand-driven recession

In the first simulation exercise, we focus on the decade following the GFC, a period commonly associated with a notable decline in the natural interest rate, large adverse demand shocks, and extraordinary monetary policy accommodation. We simulate a sequence of adverse transient demand shocks lasting 12 quarters. The size of the shocks is chosen to reduce inflation by about 1% and output gap by 5% at their peaks, yielding a reasonable approximation of the strong demand headwinds in the immediate aftermath of the GFC.

Figure 5 depicts the simulation outcomes of key variables in our benchmark calibration (solid red and yellow lines), and also in a counterfactual simulation where the central bank and the private sector both have full information about the determinants of r-star (solid blue lines). In this full information counterfactual, r-star is unchanged throughout the simulation

[13]The equilibrium subjective variance of the states is P_i and the first element of that matrix is the variance of r_t^{**}, which corresponds to the variance of r-star at quarterly rate. The size of a symmetric confidence interval of size α at annual rate is then given by $4z_\alpha\sqrt{P_{i,11}}$. For $\alpha = 0.9$, $z_\alpha \approx 1.68$.



Note: Simulation of macroeconomic variables based on a sequence of negative demand shocks over 8 quarters. Parameters used for the calibration are shown in Table 1.

Figure 5: Responses of macroeconomic variables to an adverse demand shock.

as everyone understands that the adverse demand shocks are not permanent. Inflation and the output gap decline and the central bank responds by lowering the nominal interest rate.

With incomplete information and misperception, the perceived r-star from the the central bank's perspective (solid yellow line) declines steeply, by almost 2.5 percentage points within the first few years. The reason is that the central bank misinterprets the transient shock to output as being partly driven by a decline in r-star, prompting the central bank to lower its r-star estimate as well as to cut its policy rate by more than the policy rule reaction to inflation and the output gap would imply (top right panel). This policy action is observed and interpreted by the private sector as signaling a fall in r-star, prompting the private sector to revise its r-star estimate as well. As a result, the *de facto* r-star (red line) declines steadily, though not as sharply as the one perceived by the central bank. This sets in motion a positive learning feedback that keeps both agents' estimates of r-star low throughout the following decade. This result shows that the hall-of-mirrors effect does not only affect r-star perceptions temporarily, but in fact very persistently.

The paths of the simulated variables under misperception exhibit the same patterns as the data in the aftermath of the GFC quite well. By construction of the shocks in the simulation, the output gap and inflation fall by similar amounts as in the data. They subsequently recover more quickly, however, partly due to the simplicity of the standard New-Keynesian model that does not feature intrinsic inertia, and partly due to the fact that our model is linear and cannot capture the contractionary effects of the binding zero lower bound in the aftermath of the GFC. Indeed, our simulated path of the nominal interest rate becomes negative for four quarters. At the end of the period shown, however, the levels of the simulated nominal rate and the federal funds rate align well.

Most strikingly, the r-star estimate based on Holston et al. (2017), a popular benchmark measure of r-star estimated from inflation and output, lines up well with the model's prediction of the central bank's estimate of r-star. At the same time, the private sector's expectation of long-term real rates, e.g. based on the Blue Chip survey, declines in tandem but at a slower pace than the Holston et al. (2017) measure, consistent with the model's prediction. This result confirms that our relatively simple model can quantitatively explain the entire fall in perceived r-star from persistent misperception of temporary shocks due to the hall-of-mirrors effect.

It is also noteworthy that the reductions in inflation and the output gap are slightly less pronounced under misperception than under full information. As in the static model, these differences reflect two countervailing forces. First, misperception generates a deflationary force by lowering de facto r-star. Conditional on the same interest rate path, the hall-of-mirrors case is therefore associated with lower output and inflation. Second, the central bank's perceived r-star declines by more than that of the private sector, which makes the monetary policy stance more accommodative than what the inertial Taylor rule would prescribe. This second force dominates in the simulation. If our model incorporated an effective lower bound on the nominal interest rate, only the first force would operate and the hall-of-mirrors equilibrium would likely result in an unambiguously inferior macroeconomic outcome.

We can also use the model to simulate longer-term interest rates and their subjective expectations, providing another avenue for empirical evaluation. We use the simulated interest rate paths to generate a series of yield curves.[14] We compare these simulated yield curves with the data. The model is able to replicate the stylized facts discussed in section 2. The top-left panel of Figure 6 shows that, just as in the data, yield forecasts systematically fail to predict the persistent decline in the actual yield. Note that expectations of interest rates *underreact*, even though expectations of r-star *overreact* in our model. The reason is that private sector agents in the model are ignorant of the predictable overreaction of their

[14]The model implicitly embeds the assumption that term premia are zero.

own r-star expectations, and therefore expect interest rates to recover for a long time. It is only when r-star expectations stop falling that the bias in interest rate forecasts all but disappears in our model.

The top right panel of Figure 6 shows that long-term forward real rates reacts to monetary policy surprises, as in the data.[15] In the model, this correlation arises from the signaling channel: The private sector interprets downward surprises in nominal interest rates as signaling information on behalf of the central bank that r-star has declined.

Finally, the bottom two panels of the figure show that the broad movements of the yield curve match up with the data as well. As discussed in section 2, these stylized facts collectively are difficult to explain if one maintains that r-star must be independent from cyclical phenomena.

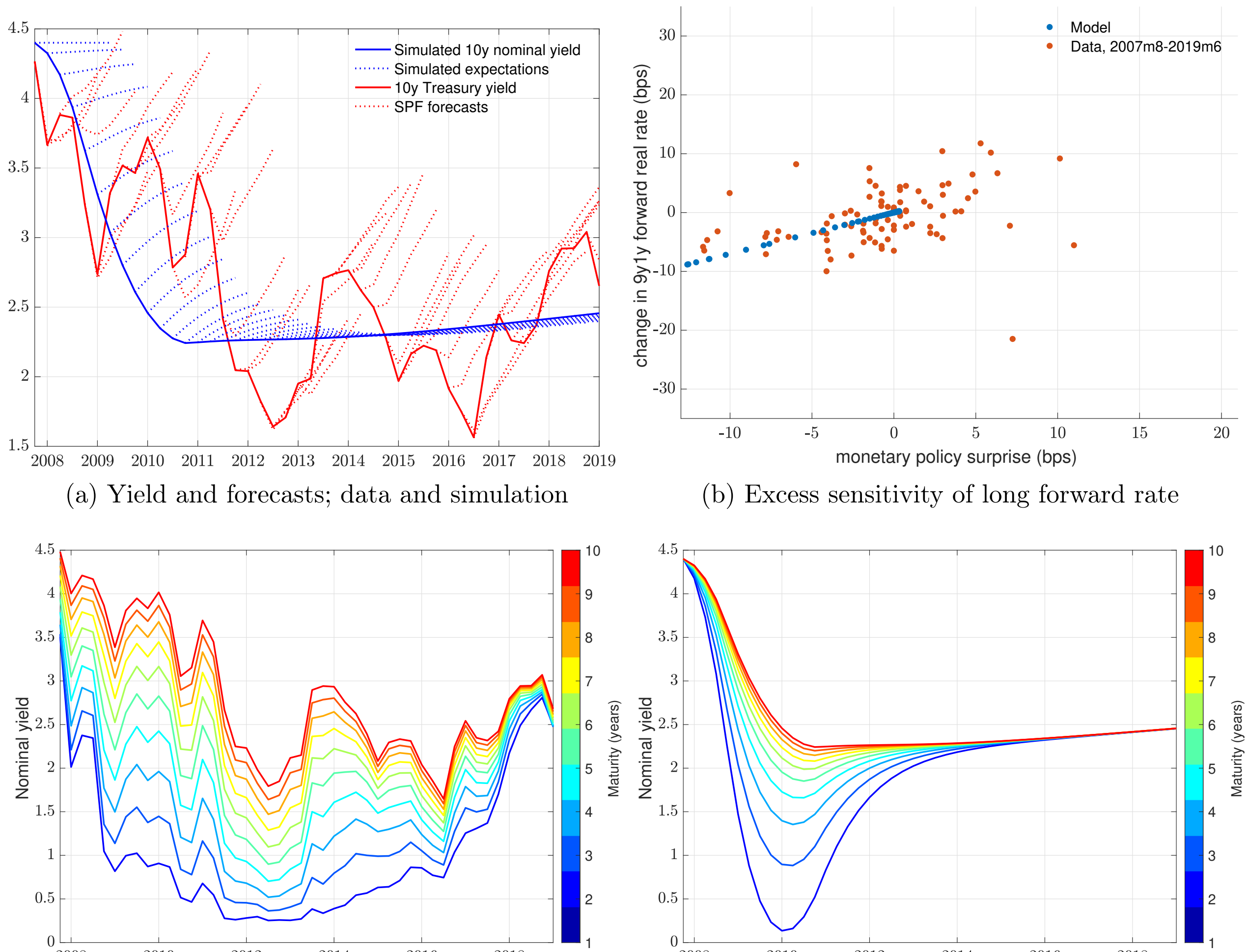
6.2 Monetary policy shocks

Temporary monetary policy shocks can set off a similar chain reaction that prompts a persistent fall in r-star. We show the effects of an expansionary shock to the Taylor rule in Figure 7.

The central bank initially believes correctly that r-star remains constant. But as it eases policy, the private sector reacts by revising down its perception of r-star, pushing the relevant natural rate for the economy lower. This creates demand headwinds, which the central bank then attributes partly to a decline in r-star, not realizing the endogenous impact of its own policy action. Again, the information feedback is set in motion, leading to persistently low r-star as perceived by both parties.

The figure also shows that the hall-of-mirrors effect in this case is unambiguously contractionary relative to the full information case, limiting the effectiveness of the shock in stimulating aggregate demand. The supposed tell-tale signs of overly accommodative policy such as inflation pressure become unreliable, as the hall-of-mirrors effect distorts the private sector's beliefs. What's more, the misperception set in motion by expansionary policy takes very long to dissipate. A dilemma consequently emerges: A more aggressive monetary policy accommodation may boost output in the short run, but may also worsen demand headwinds over time.

[15]In the data, monetary policy surprises are defined as changes in short-term interest rates around announcement dates. In the model, macroeconomic shocks and monetary policy surprises are realized simultaneously. Therefore, we use a decomposition of the overall surprise in each period, and of the corresponding movements in yields, that best isolates the policy surprise component. The construction of this decomposition is documented in Appendix C.1.



(a) Yield and forecasts; data and simulation

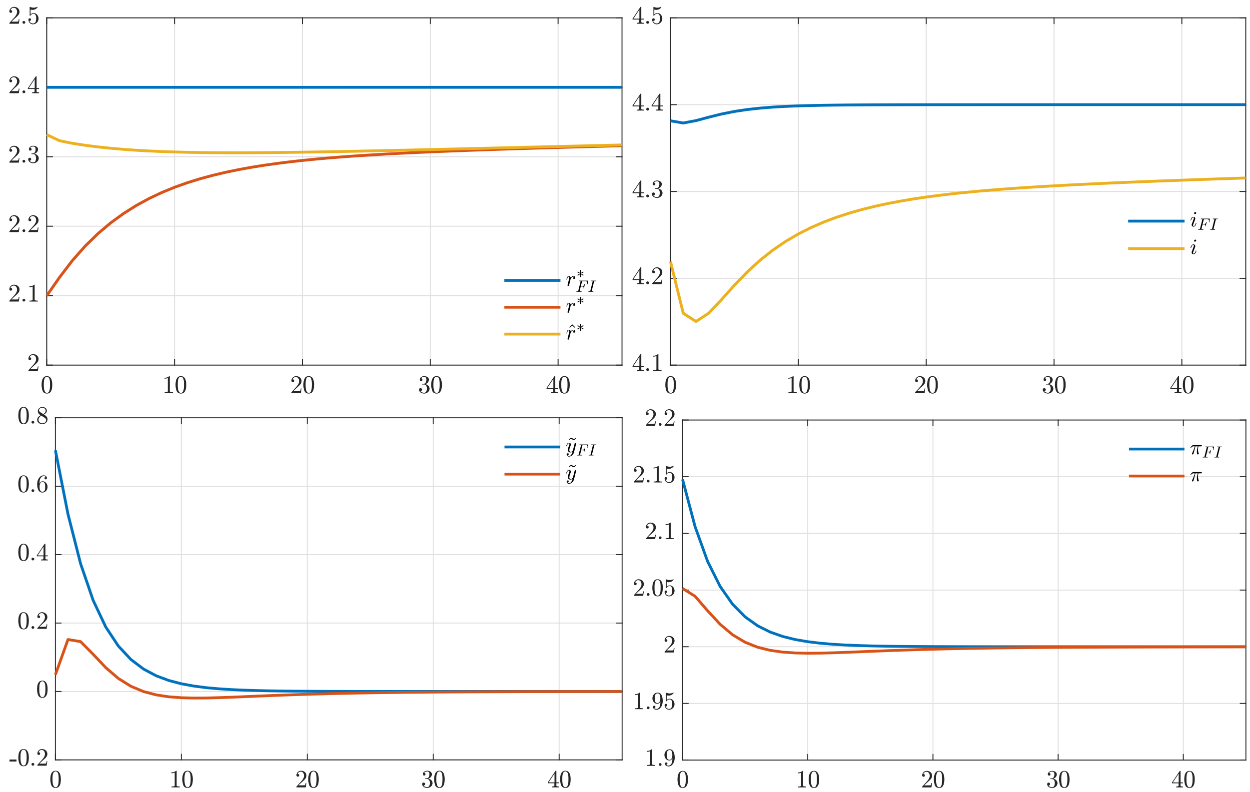
(b) Excess sensitivity of long forward rate

(c) Yield curve data

(d) Yield curve simulation

Note: Simulation of yield curves and forward interest rates, constructed using expected path of interest rate and assuming that the expectation hypothesis holds. For the construction of monetary policy surprises and associated yield movements see Appendix C.1. Results are based on a sequence of negative demand shocks over 8 quarters. Parameters used for the calibration are shown in Table 1.

Figure 6: Responses of yield curve to an adverse demand shock.



Note: Simulation results based on an accommodative monetary policy shock. Parameters used for the calibration are shown in Table 1.

Figure 7: Responses to an expansionary monetary policy shock.

6.3 Sensitivity to alternative information structures

How much misperception is needed to generate a notable fall in r-star? The answer is, not much, as a sensitivity analysis to key information parameters reveals. We repeat the simulation of Figure 5, keeping the shocks constants, but changing the information parameters. The corresponding simulated outcomes on the private sector's de factor r-star and the output gap are shown in Figure 8.

In the first simulation labeled "less misperception", we reduce the *perceived* quality of private information by increasing $\sigma_j^{|i}$ tenfold, thus reducing the amount of misperception considerably. In this simulation, r-star still falls by over a percentage point within five years, about half the decline in the baseline simulation.

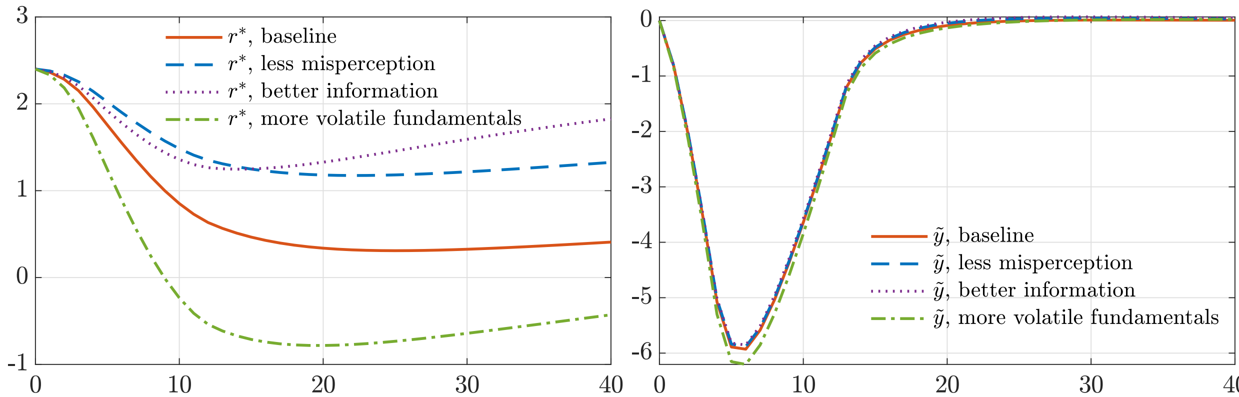
In the second simulation labeled "better information", we improve the *actual* quality of private information, by lowering σ_i drastically from infinity to one. This change, too, reduces the amount of misperception, but also leads both the private sector and the central bank to pay more attention to their own information. The result is a decline in r-star of a similar magnitude as in the first simulation, but which dissipates more quickly.

The third simulation labeled "more volatile fundamentals" shows a case in which the underlying fundamentals determining r-star (that is, the trends of productivity and discount factor changes) become more volatile. Our baseline calibration has $\sigma_r = 0.05$, corresponding to quarterly changes in annualized r-star of 0.2 percent, which is at the lower end of the estimates in Holston et al. (2017). If we increase this value to $\sigma_r = 0.07$, towards the upper bound of their estimates, this would push r-star down by a further full percentage point. Intuitively, when it is harder to pin down the true drivers of r-star, agents rely more on learning from each other, strengthening the hall-of-mirrors effect.

Finally, we note that the impact on output and inflation (not shown) of these changes to the information structure is small. These outcomes are mainly determined by the difference between the central bank's and the private sector's estimates of r-star, which stay relatively similar across the simulations.

7 Conclusion

We have extended the canonical New-Keynesian model to an incomplete information setup where agents have to learn about the determinants of r-star. Our analysis highlights the potentially important role of beliefs and two-way learning feedback as an *independent* driver of persistent changes in real interest rates. Crucially, it is difficult to determine if r-star shifts come from structural saving and investment factors, or instead from endogenous



Note: The simulation labeled "baseline" refers to the simulation shown in Figure 5. In each of the other simulations, one parameter is changed relative to this baseline. For "less misperception", the perceived noise in private signals is set to $\sigma_i^{|j} = 2$ instead of $\sigma_i^{|j} = 0.2$. For "better information", the actual noise in private signals is set to $\sigma_i = 1$ instead of $\sigma_i = \infty$. For "more volatile fundamentals", the volatility of r-star fundamentals is set to $\sigma_r = 0.07$ instead of $\sigma_r = 0.05$.

Figure 8: Alternative simulations with an adverse demand shock.

misperception by both the private sector and the central bank. Both cases produce similar macroeconomic outcomes, but the policy implications could not be more different. If misperception is indeed responsible for a trend decline in real interest rates, then reacting to negative demand shocks with an overly aggressive policy accommodation could be counterproductive. Such a strategy would prompt the private sector to revise down its perception of r-star, and further raise the bar for what it takes to keep the economy on a sustainable path. The problem likely becomes more severe the closer the economy is to the policy lower bound, as the very act of aggressive policy easing could lower the natural interest rate and make that constraint more likely to bind. Our finding sounds caution on the conventional policy recommendation that central banks should ease policy aggressively in the vicinity of the effective lower bound to avoid a liquidity trap.

Our results point to several promising avenues of further research. First, exploring further ways of taking the model to the data and quantifying the hall-of-mirrors effect in practice would enhance our understanding of its importance. Second, it could be fruitful to explore normative implications of the hall-of-mirrors effect for the design of monetary policy. Third, our model abstracts from ways in which r-star misperception may affect decisions such as investment or R&D, decisions that have a long-term impact. Adding these channels would likely imply that even temporary misperception of r-star could have longer-lasting economic consequences that are hard to reverse.

References

ADRIAN, T., R. CRUMP, AND E. MOENCH (2013): "Pricing the Term Structure with Linear Regressions," *Journal of Financial Economics*, 110, 110–38.

ANGELETOS, G.-M., L. IOVINO, AND J. LA'O (2020): "Learning over the business cycle: Policy implications," *Journal of Economic Theory*, 190, 105115.

BEAUDRY, P. AND C. MEH (2021): "Monetary policy, trends in real interest rates and depressed demand," Bank of Canada Staff Working Papers 2021-27.

BENHABIB, J., P. WANG, AND Y. WEN (2015): "Sentiments and Aggregate Demand Fluctuations," *Econometrica*, 83, 549–585.

BERNANKE, B. (2004): "What Policymakers Can Learn from Asset Prices," Speech before the Investment Analysts Society of Chicago, Chicago, Illinois.

BERNANKE, B. AND M. WOODFORD (1997): "Inflation Forecasts and Monetary Policy," *Journal of Money, Credit and Banking*, 29, 653–84.

BILLI, R. M. (2011): "Optimal Inflation for the US Economy," *American Economic Journal: Macroeconomics*, 3, 29–52.

BORIO, C., P. DISYATAT, J. JUSELIUS, AND P. RUNGCHAROENKITKUL (2017): "Why so Low for so Long? A Long-Term View of Real Interest Rates," BIS Working Paper 685.

CABALLERO, R. J. AND A. SIMSEK (2021): "Monetary Policy with Opinionated Markets," NBER Working Papers 27313.

CHAN, J. (2020): "Monetary Policy and Sentiment-Driven Fluctuations," working paper.

CIALDINI, R. B., W. WOSINSKA, D. W. BARRETT, J. BUTNER, AND M. GORNIK-DUROSE (1999): "Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists," *Personality and Social Psychology Bulletin*, 25, 1242–1253.

COIBION, O. AND Y. GORODNICHENKO (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105, 2644–78.

CUKIERMAN, A. AND F. LIPPI (2005): "Endogenous monetary policy with unobserved potential output," *Journal of Economic Dynamics and Control*, 29, 1951–1983.

GALÍ, J. (2015): *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*, Princeton University Press.

HANSON, S. G., D. O. LUCCA, AND J. H. WRIGHT (2018): "The Excess Sensitivity of Long-term Rates: A Tale of Two Frequencies," Staff Report 810, Federal Reserve Bank of New York.

Hanson, S. G. and J. C. Stein (2015): "Monetary Policy and Long-term Real Rates," *Journal of Financial Economics*, 115, 429–48.

Hillenbrand, S. (2022): "The Fed and the Secular Decline in Interest Rates," Working paper.

Holston, K., T. Laubach, and J. C. Williams (2017): "Measuring the Natural Rate of Interest: International Trends and Determinants," *Journal of International Economics*, 108, S59–S75.

Johannsen, B. K. and E. Mertens (2021): "A Time-Series Model of Interest Rates with the Effective Lower Bound," *Journal of Money, Credit and Banking*, 53, 1005–1046.

Kearns, J., A. Schrimpf, and D. Xia (2018): "Explaining Monetary Spillovers: The Matrix Reloaded," BIS Working Papers 757.

Laubach, T. and J. C. Williams (2003): "Measuring the Natural Rate of Interest," *Review of Economics and Statistics*, 85, 1063–1070.

Lunsford, K. G. and K. D. West (2019): "Some Evidence on Secular Drivers of US Safe Real Rates," *American Economic Journal: Macroeconomics*, 11, 113–39.

Melosi, L. (2016): "Signalling Effects of Monetary Policy," *Review of Economic Studies*, 84, 853–884.

Mian, A., L. Straub, and A. Sufi (2020): "Indebted demand," NBER Working Papers 26940.

Morris, S. and H. S. Shin (2002): "Social Value of Public Information," *American Economic Review*, 92, 1521–1534.

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

Nimark, K. (2008): "Monetary policy with signal extraction from the bond market," *Journal of Monetary Economics*, 55, 1389–1400.

Orphanides, A. (2003): "Monetary policy evaluation with noisy information," *Journal of Monetary Economics*, 50, 605–631.

Orphanides, A. and J. C. Williams (2007): "Robust monetary policy with imperfect knowledge," *Journal of Monetary Economics*, 54, 1406–1435, carnegie-Rochester Conference Series on Public Policy: Issues in Current Monetary Policy Analysis November 10-11, 2006.

——— (2008): "Learning, expectations formation, and the pitfalls of optimal control monetary policy," *Journal of Monetary Economics*, 55, S80–S96.

Primiceri, G. E. (2006): "Why Inflation Rose and Fell: Policy-Makers' Beliefs and U. S. Postwar Stabilization Policy*," *Quarterly Journal of Economics*, 121, 867–901.

Rungcharoenkitkul, P., C. Borio, and P. Disyatat (2019): "Monetary policy hysteresis and the financial cycle," BIS Working Papers 817.

Tang, J. (2015): "Uncertainty and the Signaling Channel of Monetary Policy," Working Paper 15-8, Federal Reserve Bank of Boston.

Appendix A New Keynesian Model with Incomplete Information

A.1 Model building blocks

Households

The representative household solves

$$\max_{C_t,N_t} E_0^h \sum_{t=0}^{\infty} \beta^t e^{\xi_t} \left(\frac{C_t^{1-\sigma}}{1-\sigma} - A_t^{1-\sigma} \frac{N_t^{1+\varphi}}{1+\varphi} \right) \tag{A.1}$$

$$\text{s.t.} \quad P_t C_t + Q_t B_t \leq B_{t-1} + W_t N_t + D_t \tag{A.2}$$

where C_t is aggregate consumption. Importantly, E^h is the mathematical expectations taken with respect to households' information set, economic model and beliefs about the relative accuracy of their information. We elaborate on this in the next section.

Solving this problem leads to the first-order conditions

$$\frac{W_t}{P_t} = C_t^{\sigma} A_t^{1-\sigma} N_t^{\varphi} \tag{A.3}$$

$$Q_t = \beta E_t^h \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\sigma} \frac{P_t}{P_{t+1}} e^{\Delta \xi_{t+1}} \right] \tag{A.4}$$

plus a transversality condition. These conditions can be written in log-linear form as

$$w_t - p_t = \sigma c_t + (1-\sigma) a_t + \varphi n_t \tag{A.5}$$

$$c_t = E_t^h(c_{t+1}) - \frac{1}{\sigma} \left[i_t - E_t^h(\pi_{t+1}) - \rho + \Delta \xi_{t+1}) \right] \tag{A.6}$$

where $i_t \equiv -\log Q_t$, $\rho \equiv -\log(\beta)$, $\pi_{t+1} \equiv p_{t+1} - p_t$, and small letters denote logs of relevant variables. Households also solve a sub-problem, which arises from their preference for variety. The aggregate consumption is posited to be a CES sum of differentiated goods

$$C_t \equiv \left(\int_0^1 C_t(i)^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} \tag{A.7}$$

Households maximize this for a given level of expenditure

$$\int_0^1 P_t(i) C_t(i) di \tag{A.8}$$

which gives rise to

$$C_t(i) = \left(\frac{P_t(i)}{P_t} \right)^{-\epsilon} C_t \tag{A.9}$$

and can be used to verify that equation A.8 is equal to P_tC_t, consistent with the budget constraint A.2.

Production

Firms $i \in [0, 1]$ produce differentiated goods, with an identical technology

$$Y_t(i) = A_t N_t(i)^{1-\alpha} \tag{A.10}$$

where we leave the process for $a_t \equiv \log(A_t)$ unspecified for now (this will be one source of potential misperception, along with z_t).

Price setting

Assume Calvo pricing where firms can re-optimize and adjust prices with probability $1 - \theta$. This gives rise to the sticky price formulation of aggregate price in a log-linearized term (around steady-state inflation of zero):

$$\pi_t = (1 - \theta)(p_t^* - p_{t-1}) \tag{A.11}$$

where p_t^* is the log price set by re-optimizing firms, which must take into account how long they will remain with the price once it has been reset. This can be shown to be given by

$$p_t^* = \mu + (1 - \beta\theta) \sum_{k=0}^{\infty} (\beta\theta)^k E_t^h(\psi_{t+k|t}) \tag{A.12}$$

where $\psi_{t+k|t}$ is the log marginal cost in period $t + k$ for a firm that last reset in period t, and $\mu \equiv \log(\frac{\epsilon}{\epsilon-1})$ is desired gross markup. Note that again we use the expectations operator E_t^h, with the assumption that firms and households share the same information set, economic model, and perception about the accuracy of their signals relative to the central bank's.

Equilibrium

Consider first the derivation of the Phillips curve. The individual firm's marginal cost $\psi_{t+k|t}$ is wage at $t + k$ minus marginal product of labor in $t + k$ for a firm resetting in t. These need to be solved in the general equilibrium, and will depend on future employment, hence output and price. It can be shown that

$$\psi_{t+k|t} = \psi_{t+k} - \frac{\alpha\epsilon}{1-\alpha}(p_t^* - p_{t+k}) \tag{A.13}$$

where $\psi_{t+k} \equiv \int_0^1 \psi_t(i)di$ is the cross-sectional average marginal cost.

Combining equations A.12 and A.13, one can write p_t^* in a recursive form as

$$p_t^* = \beta\theta E_t^h(p_{t+1}^*) + (1 - \beta\theta)\left(p_t - \frac{1-\alpha}{1-\alpha+\alpha\epsilon}\hat{\mu}\right) \tag{A.14}$$

where $\hat{\mu} \equiv \mu_t - \mu$ is the deviation between the average markup $\mu_t \equiv p_t - \psi_t$ and the desired markup. Plugging this into equation A.11, we get

$$\pi_t = \beta E_t^h(\pi_{t+1}) - \left(\frac{(1-\theta)(1-\beta\theta)(1-\alpha)}{\theta(1-\alpha+\alpha\epsilon)} \right) \hat{\mu_t} \tag{A.15}$$

To derive $\hat{\mu}$, note that the average markup μ_t depends on output and productivity:

$$\mu_t = \left(\sigma + \frac{\varphi + \alpha}{1-\alpha} \right) (a_t - y_t) \, a_t + \log(1-\alpha) \tag{A.16}$$

Under flexible prices, $\mu_t = \mu$ obtains as firms can set markup frictionlessly. Inverting the equation leads to the natural output y_t^n definition:

$$y_t^n = a_t + \psi_y \tag{A.17}$$

where $\varphi_y \equiv -(1-\alpha)(\mu - \log(1-\alpha))/(\sigma(1-\alpha) + \varphi + \alpha)$. Thus:

$$\hat{\mu_t} = -\left(\sigma + \frac{\varphi + \alpha}{1-\alpha} \right) (y_t - y_t^n) \tag{A.18}$$

Substituting this into A.15 results in the New Keynesian Phillips curve

$$\pi_t = \beta E_t^h(\pi_{t+1}) + \kappa \tilde{y}_t \tag{A.19}$$

where $\tilde{y}_t \equiv y_t - y_t^n$ is the output gap and $\kappa \equiv \left(\frac{(1-\theta)(1-\beta\theta)(1-\alpha)}{\theta(1-\alpha+\alpha\epsilon)} \right) \left(\sigma + \frac{\varphi+\alpha}{1-\alpha} \right)$.

To derive the IS curve, note that the goods market equilibrium condition

$$Y_t = C_t \tag{A.20}$$

implies that the Euler equation A.6 is given by

$$y_t = E_t^h(y_{t+1}) - \frac{1}{\sigma} \left[i_t - E_t^h(\pi_{t+1}) - \rho + \Delta\xi_{t+1} \right] \tag{A.21}$$

We use the following specification for productivity process a_t and that of the stochastic discount factor ξ_t:

$$\begin{aligned}
a_t &= a_{t-1} + g_{t-1} + \epsilon_{y^*t} \\
g_t &= g_{t-1} + \epsilon_{gt} \\
\xi_t &= \xi_{t-1} - z_{t-1} - \sigma u_{ht-1} - \epsilon_{\xi t} \\
u_{ht} &= \rho_h u_{ht-1} + \epsilon_{ht} \\
z_t &= z_{t-1} + \epsilon_{zt}
\end{aligned}$$

We can now rewrite the Euler equation in terms of output gaps to make r-star appear:

$$\tilde{y}_t = E_t^h(\tilde{y}_{t+1}) + E_t^h(y_{t+1}^n) - y_t^n - \frac{1}{\sigma}\left[i_t - E_t^h \pi_{t+1} + E_t^h \Delta\xi_{t+1} - \rho\right] \tag{A.22}$$

$$= E_t^h(\tilde{y}_{t+1}) + \psi_{ya} E_t^h \Delta a_{t+1} - \frac{1}{\sigma}\left[i_t - E_t^h \pi_{t+1} + E_t^h \Delta\xi_{t+1} - \rho\right] \tag{A.23}$$

$$= E_t^h(\tilde{y}_{t+1}) - \frac{1}{\sigma}\left[i_t - E_t^h \pi_{t+1} - E_t^h(r_t^*)\right] + \varepsilon_t^y \tag{A.24}$$

where the natural interest rate is defined by

$$r_t^* = \rho + \sigma g_{t+1} + z_{t+1} \tag{A.25}$$

Remark 1. *The natural rate r_t^* depends on how agents perceive the z_t and a_t processes. There is thus a distinction between the de facto r_t^* as defined here, and the perfect-information counterpart, r_t^{**}.*

The model is closed by the Taylor rule

$$i_t = E_t^c(r_t^*) + \phi_\pi \pi_t + \phi_y \tilde{y}_t + \varepsilon_t^m \tag{A.26}$$

where E^c is the mathematical expectations taken with respect to the central bank's information set, economic model and beliefs about the accuracy of its information relative to that of households.

Appendix B Proofs

B.1 Proof of Proposition 1

Proof. The coefficients g_{si} and g_{ai} are given by (4.14); moreover, (4.16) implies $\gamma_j = 0$ and $\alpha_j = g_{sj}$. Therefore:

$$\begin{pmatrix} g_{si} \\ g_{ai} \end{pmatrix} = \frac{1}{g_{sj}^2\left(\sigma_{\epsilon i}^2 + \sigma_{\epsilon j}^2 + \sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right) + \sigma_{uj}^2\left(\sigma_{\epsilon i}^2 + 1\right)} \begin{pmatrix} g_{sj}^2\sigma_{\epsilon j}^2 + \sigma_{uj}^2 \\ g_{sj}\sigma_{\epsilon i}^2 \end{pmatrix}.$$

It is immediate that $g_{si} \geq 0$. By symmetry then, $g_{sj} \geq 0$ and this in turn implies $g_{ai} \geq 0$ as well. It is also clear that

$$g_{sj}^2\sigma_{\epsilon j}^2 + \sigma_{uj}^2 \leq g_{sj}^2\left(\sigma_{\epsilon i}^2 + \sigma_{\epsilon j}^2 + \sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right) + \sigma_{uj}^2\left(\sigma_{\epsilon i}^2 + 1\right)$$

and therefore $g_{si} \leq 1$. The above inequality is binding if and only if

$$0 = \sigma_{\epsilon i}^2\left(g_{sj}^2\left(1 + \sigma_{\epsilon j}^2\right) + \sigma_{uj}^2\right).$$

Because we assume $\sigma_{uj}^2 > 0$, this can be the case if and only if $\sigma_{\epsilon i}^2 = 0$. This establishes that $g_{si} = 1$ if and only if $\sigma_{\epsilon i}^2 = 0$.

We now show that $g_{ai} < 1$. Because $\sigma_{uj}^2 > 0$, we have that

$$
\begin{aligned}
g_{ai} &< \frac{g_{sj}\sigma_{\epsilon i}^2}{g_{sj}^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)} \\
&= \frac{1}{g_{sj}}\frac{\sigma_{\epsilon i}^2}{\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)} \\
&= \frac{g_{si}^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)+\sigma_{ui}^2\left(\sigma_{\epsilon j}^2+1\right)}{g_{si}^2\sigma_{\epsilon i}^2+\sigma_{ui}^2}\frac{\sigma_{\epsilon i}^2}{\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)} \\
&\leq \frac{\sigma_{\epsilon i}^2 g_{si}^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)+\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)\sigma_{ui}^2}{\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)\left(g_{si}^2\sigma_{\epsilon i}^2+\sigma_{ui}^2\right)} \\
&= 1.
\end{aligned}
$$

To show existence, note that $g_{si} \in [0,1]$ and that it is a non-increasing function of g_{sj}, in fact:

$$\frac{\partial g_{si}}{\partial g_{sj}} = -\frac{2g_{sj}\sigma_{uj}^2\sigma_{\epsilon i}^2}{\left(g_{sj}^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)+\sigma_{uj}^2\left(\sigma_{\epsilon i}^2+1\right)\right)^2}.$$

Because this holds for $i = c, h$, there exists at least one pair (g_{sh}, g_{sc}) that satisfies the equilibrium conditions. The second derivative of g_{si} with respect to g_{sj} is:

$$\frac{\partial^2 g_{si}}{\partial g_{sj}^2} = -2\sigma_{uj}^2\sigma_{\epsilon i}^2\frac{\sigma_{uj}^2\left(\sigma_{\epsilon i}^2+1\right)-3g_{sj}^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)}{\left(g_{sj}^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)+\sigma_{uj}^2\left(\sigma_{\epsilon i}^2+1\right)\right)^3}$$

Therefore, g_{si} has at most one inflection point. It follows that there are at most three pairs (g_{sh}, g_{sc}) that satisfy the equilibrium conditions. □

B.2 Proof of Proposition 2

Proof. Suppose that $\sigma_{\epsilon j}^{2|i} = 0 < \sigma_{\epsilon j}^2$ for $i = c, h$ and $j \neq i$. Because $\sigma_{\epsilon j}^2 > 0$, $g_{sj} < 1$ in the case without misperception by Proposition 1. The same proposition also implies $g_{sj}^{|i} = 1$. Therefore:

$$g_{ai}^{|i} = \frac{\sigma_{\epsilon i}^2}{\sigma_{\epsilon i}^2+\sigma_{uj}^2\left(\sigma_{\epsilon i}^2+1\right)}.$$

As $\sigma_{uj}^2 \to 0$, $g_{ai}^{|i} \to 1$ and this shows that both $g_{ac}^{|c}$ and $g_{ah}^{|h}$ can be arbitrarily close to one.

Next, we compare the corresponding parameter without misperception and obtain:

$$
\begin{aligned}
g_{ai} &= \frac{g_{sj}\sigma_{\epsilon i}^2}{g_{sj}^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)+\sigma_{uj}^2\left(\sigma_{\epsilon i}^2+1\right)} \\
&= \frac{1}{g_{sj}}\frac{\sigma_{\epsilon i}^2}{\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2+\frac{\sigma_{uj}^2}{g_{sj}^2}\left(\sigma_{\epsilon i}^2+1\right)} \\
&= \frac{g_{si}^2\sigma_{\epsilon i}^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2\right)+\sigma_{ui}^2\left(\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2\sigma_{\epsilon i}^2\right)}{g_{si}^2\sigma_{\epsilon i}^2+\sigma_{ui}^2}\frac{1}{\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2+\frac{\sigma_{uj}^2}{g_{sj}^2}\left(\sigma_{\epsilon i}^2+1\right)} \\
&< \frac{\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2}{\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2+\frac{\sigma_{uj}^2}{g_{sj}^2}\left(\sigma_{\epsilon i}^2+1\right)} \\
&< \frac{\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2}{\sigma_{\epsilon i}^2+\sigma_{\epsilon j}^2+\sigma_{\epsilon i}^2\sigma_{\epsilon j}^2+\sigma_{uj}^2\left(\sigma_{\epsilon i}^2+1\right)} \\
&< \frac{\sigma_{\epsilon i}^2}{\sigma_{\epsilon i}^2+\sigma_{uj}^2\left(\sigma_{\epsilon i}^2+1\right)} \\
&= g_{ai}^{|i}.
\end{aligned}
$$

Finally, $\sigma_{\epsilon i}^{2|j}=0$ also implies that $g_{ai}^{|j}=0$, so $g_{ai}^{|i}>g_{ai}^{|j}$. By a continuity argument, these inequalities also hold for $\sigma_{\epsilon i}^{2|j}$ close enough to zero. □

B.3 Proof of Proposition 3

Proof. Recall the expression for b_i:

$$
b_i=\left(1-g_{aj}^{|j}\right)\frac{1-g_{ai}^{|i}g_{aj}^{|i}}{1-g_{ai}^{|i}g_{aj}^{|j}}
$$

Because $g_{ai}^{|i},g_{aj}^{|i}<1$ for $i=c,h$ and $j\neq i$, it is immediate that $b_i>0$. For sufficiently strong misperception ($\sigma_{\epsilon j}^{2|i}$ close enough to zero), $g_{aj}^{|i}<g_{aj}^{|j}$ by Proposition 2, and therefore

$$
b_i<\left(1-g_{aj}^{|j}\right)<1.
$$

In the case of full information of both agents, $g_{ai}^{|i}=g_{aj}^{|i}=1$ for $i=c,h$ and $j\neq i$. Therefore $b_i=1$. □

Appendix C Details on the solution of the dynamic model

The solution to the macro side of the model (5.7)–(5.9) can be written in the following form:

$$\begin{pmatrix} \tilde{y}_t \\ \pi_t \\ i_t \end{pmatrix} = \Gamma i_{t-1} + \Theta E_t^h Z_t + \sum_{s=0}^{\infty} \theta_s E_t^h \left[E_{t+s}^c \left[Z_{t+s} \right] \right]. \tag{C.1}$$

The coefficient matrices Γ, Θ and θ_s depend on the parameters as well as on the matrix A_z. Their values can be computed using standard solution methods. We use Chris Sims's Gensys procedure, which has the advantage that it directly yields the coefficients θ_s without the need to specify a process for $E_t^h \left[E_{t+s}^c \left[Z_{t+s} \right] \right]$. Future values $E_t^h \left[E_{t+s}^c \left[Z_{t+s} \right] \right]$ for $s \geq 0$ need to be carried over because the evolution of these expectations will be endogenous to the belief formation process.

The coefficient matrices in 5.13 can be found as follows. Start by observing that $X_{it} = \left(Z_t', m_{jt}' \right)'$, so that we can define linear maps C_z,C_m, D_z and D_m for which

$$X_{it} = C_z Z_t + C_m m_{jt} \tag{C.2}$$

$$\begin{pmatrix} Z_t \\ m_{jt} \end{pmatrix} = \begin{pmatrix} D_z \\ D_m \end{pmatrix} X_{it}. \tag{C.3}$$

With this notation and the guess (5.12), we can write

$$\begin{aligned} X_{it} &= C_z Z_t + C_m \left(\Phi_j m_{jt-1} + \Psi_j m_{jt-1} + \Omega_j Z_t \right) \\ &= \left(\left(C_z + C_m \Omega_j \right) A_z D_z + C_m \Phi_j D_m \right) X_{it-1} + \left(C_z + C_m \Omega_j \right) q_t + C_m \Psi_j m_{it-1} \\ &= A_i X_{it-1} + B_i q_t + C_i m_{it-1}. \end{aligned} \tag{C.4}$$

Equations (5.13) and (5.14) form a standard linear filtering problem, the solution of which is given by the Kalman filter. The optimal filtering equation describing the evolution of beliefs is:

$$m_{it} = \left(I - G_i H_i \right) \left(A_i + C_m \Psi_j \right) m_{it-1} + G_i Y_{it} \tag{C.5}$$

The Kalman gain G_i, as well as the time-invariant posterior covariance matrix P_i, can be computed using the following formula:

$$P_i^- = A_i P_i A_i' + B_i \Sigma_q B_i' \tag{C.6}$$

$$S_i = H_i P_i^- H_i' \tag{C.7}$$

$$G_i = P_i^- H_i' S_i^{-1} \tag{C.8}$$

$$P_i = P_i^- - G_i S_i G_{it}' \tag{C.9}$$

In practice, one iterates on these four equations to find the fixed point of this system of equations.

Finally, we will describe how to find the signal matrices H_i. The household's observation problem is straightforward. The household observes $s_{ht}, x_t, u_{ht}, u_{pt}$, as well as the nominal interest rate i_t. From the monetary policy rule (5.9), one can see that observing i_t, as well as $\tilde{y}_t$ and π_t (which are the household's own choice variables) is equivalent to observing $E_t^c z_t + u_{ct}$ each period. We can therefore write the household observation as $Y_{ht} = H_h X_{ht}$ with a matrix H_h that is independent of equilibrium beliefs:

$$H_h = \begin{pmatrix} 1 & 1 & 0 & & & & \\ 1 & 0 & 0 & 1 & 0 & & \\ & & & 0 & 1 & 0 & \\ & & & & 0 & 1 & 0 \\ & & & & & 0 & 1 \end{pmatrix} D_z + \begin{pmatrix} 0 & 0 & \cdots \\ 0 & 0 & \cdots \\ 0 & 0 & \cdots \\ 0 & 0 & \cdots \\ 1 & 0 & \cdots \end{pmatrix} D_m. \tag{C.10}$$

The central bank's signaling problem is more complicated. Its information about the household's expectation comes from observing $\tilde{y}_t$ and π_t, which are themselves equilibrium outcomes that depend on the household's beliefs in a non-trivial way. We first note that i_{t-1} is also in the central bank's information set. Using (C.1), we can express the macroeconomic outcomes of the model as:

$$\begin{aligned} \begin{pmatrix} \tilde{y}_t \\ \pi_t \\ i_t \end{pmatrix} - \Gamma i_{t-1} &= \Theta E_t^h [Z_t] + \sum_{s=0}^{\infty} \theta_s E_t^h \left[E_{t+s}^c [Z_{t+s}] \right] \\ &= \Theta D_z m_{ht} + \sum_{s=0}^{\infty} \theta_s D_z E_t^h [m_{ct+s}] \\ &= \Theta D_z m_{ht} + \sum_{s=0}^{\infty} \theta_s D_z (A_h + C_m \Psi_c)^s m_{ht} \\ &= \underbrace{\left(\Theta D_z + \sum_{s=0}^{\infty} \theta_s D_z (A_h + C_m \Psi_c)^s \right)}_{=M_h} D_m X_{ct}. \end{aligned}$$

Therefore $Y_{ct} = H_c X_{ct}$, where the matrix H_c is endogenous to the belief formation process:

$$H_c = \begin{pmatrix} 1 & 0 & 1 & & & & \\ 1 & 0 & 0 & 1 & 0 & & \\ & & & & 0 & 0 & 1 \\ & & & & & & 0 \\ & & & & & & 0 \end{pmatrix} D_z + \begin{pmatrix} 0 & & \\ 0 & & \\ 0 & & \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \left(\Theta D_z + \sum_{s=0}^{\infty} \theta_s D_z M_s \right) D_m. \tag{C.11}$$

The household's expectations of future macroeconomic outcomes, which we use to compute nominal and real yields in the simulations, can be found through the recursion:

$$E_t^h \begin{pmatrix} \tilde{y}_{t+s} \\ \pi_{t+s} \\ i_{t+s} \end{pmatrix} = \Gamma E_t^h i_{t+s-1} + M_h (A_h + C_m \Psi_c)^s m_{ht}, \; s \geq 0. \tag{C.12}$$

Finally, macroeconomic outcomes in a misperception equilibrium are again given by (C.12), but where A_h and Ψ_c are replaced by $A_h^{|h}$ and $\Psi_c^{|h}$, respectively, in (C.12) and in the formula for M_h above.

C.1 Model counterpart to high-frequency policy surprises

In the empirical literature, monetary policy surprises and their effects on asset prices are constructed by measuring changes in short- term interest rates and other asset prices in narrow intra-day windows around announcement dates. Identification obtains from the making sure that during the window, no macroeconomic news other than the policy announcement are realized. In our model, there is no direct counterpart to an announcement window, as all macroeconomic shocks are realized simultaneously, at the same time as expectations are updated.

To construct a counterpart to policy surprises in the model, we proceed as follows. In each period, we construct the household's belief about the state X_{ht} after observing the public signal x_t, the private signal s_{ht}, as well as the demand and cost-push shocks u_{ht}, u_{pt}, but before observing the current interest rate i_t. Denote this belief by $\bar{m}_{ht}$. When the prior is distributed as $\mathcal{N}(m_{ht-1}, P_h)$, then the posterior $\bar{m}_{ht}$ can be found using the analogous equations to (C.5)–(C.8) of the Kalman filter:

$$\bar{m}_{ht} = \left(I - \bar{G}_h \bar{H}_h\right)(A_h + C_m \Psi_c)\, m_{ht-1} + \bar{G}_h \bar{Y}_{ht} \tag{C.13}$$

$$\bar{G}_h = P_h^- \bar{H}_h' \left(\bar{H}_h P_h^- \bar{H}_h'\right)^{-1}. \tag{C.14}$$

Here, the signals before the observation of the interest rate are:

$$\bar{Y}_{ht} = \begin{pmatrix} s_{ht} \\ x_t \\ u_{ht} \\ u_{pt} \end{pmatrix} = \bar{H}_h X_{ht} = \begin{pmatrix} 1 & 1 & 0 & & & & \\ 1 & 0 & 0 & 1 & 0 & & \\ & & & 0 & 1 & 0 & \\ & & & & 0 & 1 & 0 \\ & & & & & 0 & 1 \end{pmatrix} D_z X_{ht}. \tag{C.15}$$

We now take the information that is revealed through observing the current interest rate as the policy surprise, which comprises both information about the policy shock u_{ct} as well as about the central bank's beliefs $\hat{r}_t^*$. The effect of the surprise on private sector expectations is simply $m_{ht} - \bar{m}_{ht}$. The surprise in $E_t^h(\tilde{y}_{t+s}, \pi_{t+s}, i_{t+s})$, which can be used to compute the short-term policy surprise in i_t as well as the announcement effects on the nominal and real yield curve, then takes the form:

$$M_h (A_h + C_m \Psi_c)^s (m_{ht} - \bar{m}_{ht}),\; s \geq 0. \tag{C.16}$$