Mis-specified Forecasts and Myopia in an Estimated New Keynesian Model*

Ina Hajdini†

April 3, 2025

A Partial Equilibrium New Keynesian Pricing Problem

A.1 Model

Firms face nominal rigidities a la Calvo: they cannot reset the price with probability $\alpha \in (0,1)$ each period. Every firm seeks to maximize the present discounted value of real profits, i.e.,

$$\max_{P^*_{jt}} \widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{t,t+h} \left(\frac{P^*_{jt}}{P_{t+h}} y_{j,t+h} - mc_{t+h} y_{j,t+h} \right) \tag{A.1}$$

where Q_t is a generic stochastic discount factor; P^*_{jt} is the optimal price set by the j^{th} firm; P_t is the aggregate price level; y_{jt} is the demand for the j^{th} firm's good; mc_t is the marginal cost; $\beta \in (0,1)$ is a deterministic discount factor. The demand each firm faces and the aggregate price level are given by

$$y_{jt} = \left(\frac{P^*_{jt}}{P_t} \right)^{-\zeta} y_t \qquad\qquad P_t = \left[\int_{j=0}^{1} P_{jt}^{1-\zeta} \right]^{\frac{1}{\zeta-1}} \tag{A.2}$$

where $\zeta > 1$ is the elasticity of substitution among the differentiated goods. Substituting for $y_{j,t+h}$ into (A.1), we have

$$\max_{P^*_{jt}} \widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{t,t+h} \left(\left(\frac{P^*_{jt}}{P_{t+h}} \right)^{1-\zeta} - \left(\frac{P^*_{jt}}{P_{t+h}} \right)^{-\zeta} mc_{t+h} \right) y_{t+h} \tag{A.3}$$

The first-order condition with respect to P^*_{jt} is

$$\frac{P^*_{jt}}{P_t} = \frac{\zeta}{\zeta-1} \frac{\widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{t,t+h} y_{t+h} mc_{t+h} \pi^{\zeta}_{t,t+h}}{\widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{t,t+h} y_{t+h} \pi^{\zeta-1}_{t,t+h}} \tag{A.4}$$

*The views expressed herein are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

†Federal Reserve Bank of Cleveland. Email: Ina.Hajdini@clev.frb.org.

where $\pi_{t,t+h} = \frac{P_{t+h}}{P_t} = \prod_{l=0}^{h} \pi_{t+l}$. Due to Calvo pricing, the aggregate price level in (A.2) can be rewritten as

$$P_t = \left[\alpha P_{t-1}^{1-\zeta} + (1-\alpha)(P_t^*)^{1-\zeta}\right]^{\frac{1}{\zeta-1}} \tag{A.5}$$

Assume that the steady state for inflation is $\bar{\pi} = 1$. From (A.5), we have that in the steady state, $P^*/P = 1$. Then, from the optimality condition in (A.4) it follows that in the steady-state equilibrium $\bar{mc} = \frac{\zeta-1}{\zeta}$. Log-linearizing the first-order condition around steady-state values and dropping the subscript j, since every firm has the same optimality condition, we have

$$\hat{p}_t^* = \widetilde{\mathbb{E}}_t \sum_{h=0}^{\infty} (\alpha\beta)^h \left((1-\alpha\beta)\hat{mc}_{t+h} + \alpha\beta\hat{\pi}_{t+h+1}\right) \tag{A.6}$$

where $\hat{p}_t^* = log(P_t^*/P_t)$ and $\hat{\pi}_{t+1} = log(P_{t+1}/P_t)$ is inflation in period $(t+1)$.

A.2 Convergence

Following Hommes and Zhu (2014), stability under SAC learning is determined by the associated Ordinary Differential Equations:

$$\begin{aligned} \frac{d\delta}{d\tau} &= \delta\left(\frac{\beta n(1-\alpha)/(1-\alpha\beta n) - b}{1-b} - 1\right) \\ \frac{d\gamma}{d\tau} &= F(\gamma) - \gamma = \frac{b+\rho}{1+b\rho} - \gamma \end{aligned} \tag{A.7}$$

For convergence, we need $\frac{\partial \frac{d\delta}{d\tau}}{\partial\delta}|_{\gamma^*} < 0$ and $\frac{\partial \frac{d\gamma}{d\tau}}{\partial\gamma}|_{\gamma^*} < 0$. The first inequality holds true if and only if $\beta n(1-\alpha)/(1-\alpha\beta n) < 1$, which is true for any parameterization of the model. The second inequality is true if and only if $F'(\gamma)|_{\gamma^*} < 1$. We know that F is increasing in γ, that it intersects the 45° line once for $\gamma \in (0,1)$, and that $F(0) = \rho \geq 0$ and $F(1) < 1$. From the proof of Proposition 1 in Section D.1, we know that F is convex when it crosses 45° line, implying that $F'(\gamma)|_{\gamma^*} < 1$. As a result, the second inequality is also true for any parameterization.

A.3 Relaxing Consistent Expectations Equilibrium

In this section, I relax the assumption of CE equilibrium and re-evaluate the parametric space for which the three facts of forecast errors are matched. The structure of the ALM remains as it is in the main text, with the difference that now $0 \leq \gamma < 1$ is a free parameter. The delayed over-shooting condition of Proposition 2 does not depend on whether the CE equilibrium concept is imposed, therefore, for late over-shooting to occur $n^h\gamma^{h+1} > \max(b^{h+1}, \rho^{h+1})$ has to hold. The CE equilibrium constrains γ^*, therefore, assuming instead that γ is a free parameter would widen the range of myopia for which delayed over-shooting occurs. Clearly, in the case of no myopia, imposing the CE equilibrium implies that one should have $\gamma > \max(b, \rho)$. Along the CE equilibrium, $\gamma^* > \max(\rho, b)$ always holds. Thus, in the case of no myopia, whether one imposes or not a CE equilibrium is not relevant for the occurrence of delayed over-shooting.

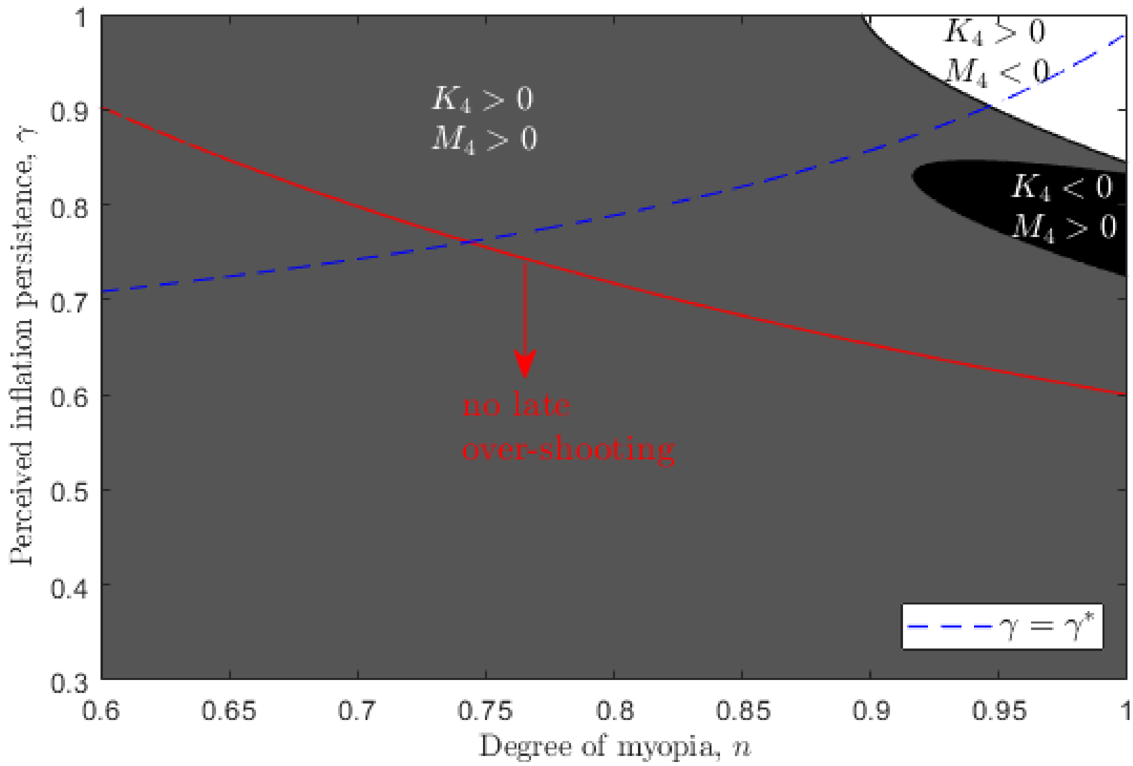


Figure 1: Regions of delayed over-shooting, under-reaction to ex-ante forecast revisions, and over-reaction to current realizations in the (n, γ) parametric space. The dashed blue curve plots $\gamma = \gamma^*$, whereas the red dashed curve plots $\gamma = n^{-h/h+1}\max(b, \rho)$. Forecasting horizon: $h = 4$. The area for which $n < 0.6$ is truncated for better visibility, but it is a region of no late over-shooting and characterized by $K_4 > 0$, $M_4 > 0$. Parameterization: $\alpha = 0.5$, $\beta = 0.99$, $\sigma_\varepsilon = 1$.

The results of Proposition 3 remain similar, with the difference that γ^* would be substituted with γ in the expressions for K_h^+, K_h^-, M_h^+, and M_h^-. The condition for $K_h > 0$ and $M_h < 0$ absent of myopia in Corollary 2 becomes

$$\underline{\gamma} < \gamma^{h+1} < \bar{\gamma} \tag{A.8}$$

with $\underline{\gamma}$ is as defined in Corollary 2 and $\bar{\gamma}$ changes to $\bar{\gamma} = b^{h+1} + \rho(1-b^2)(1-\rho\gamma)\frac{\rho^{h+1}-b^{h+1}}{(\rho-b)(1+\rho b-\gamma(b+\rho))}$. Since an equilibrium condition is not imposed on γ, the parametric space that would be consistent with $K_h > 0$ and $M_h < 0$ would generally widen, all else equal.

Figure 1 here plots the parametric space of the free behavioral parameters (n, γ) for which different facts are matched, for $\rho = 0.6$. The blue dashed curve plots $\gamma = \gamma^*$ whereas the red dashed curve plots a threshold below which there is no delayed over-shooting. Now, imagine slicing Figure 1 in the main text at $\rho = 0.6$: one would get $K_4, M_4 > 0$ for some degree of myopia

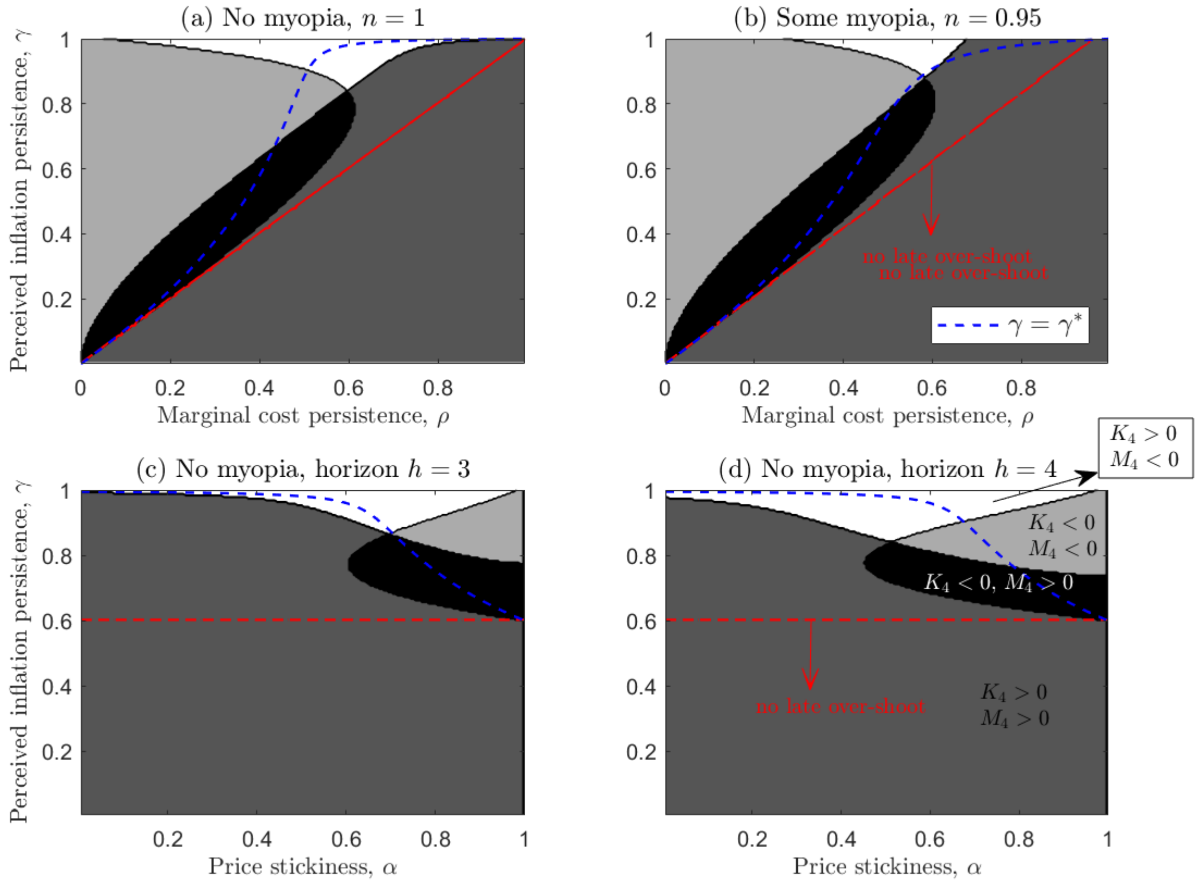


Figure 2: Panels (a) and (b): regions of delayed over-shooting, under-reaction to ex-ante forecast revisions, and over-reaction to current realizations in the (ρ, γ) parametric space with $\alpha = 0.5$ and $h = 4$. Panels (c) and (d): regions of delayed over-shooting, under-reaction to ex-ante forecast revisions, and over-reaction to current realizations in the (α, γ) parametric space with $n = 1$ and $\rho = 0.6$. The dashed blue curve plots $\gamma = \gamma^*$, whereas the red dashed curve plots $\gamma = \max(b, \rho)$. The rest of parameters are set as in Figure 1.

and as myopia vanishes, that is, as n approaches 1, $K_4 > 0, M_4 < 0$. This is consistent with what Figure 1 visualizes along the blue curve. What is important to note, however, is that for sufficiently low myopia, one can match all three empirical facts about forecast errors if γ is sufficiently close to γ^*, but it does not have to be exactly equal to γ^*.

Similarly, Figure 2 visualizes the results in the parametric space of (ρ, γ) for $n = 1$ in panel (a) and $n = 0.95$ in panel (b); whereas panels (c) and (d) visualize the results in the parametric space of (α, γ) absent of myopia for different forecast horizons. The blue and red dashed curves plot $\gamma = \gamma^*$ and $\gamma = \max(b, \rho)$, respectively. Panel (a) shows that, absent of myopia, the three facts can be matched for almost any value of ρ as long as γ is high enough. Panel (a) shows that a CE equilibrium can constrain the parametric space for which under-reaction to forecast revisions and over-reaction to current realizations occur. For example, for $\rho < 0.5$, γ^* is much lower than the values of unconstrained γ that are necessary to match the empirical facts. On the other hand, as myopia increases (panel (b)), the limitations of the parametric space that would match the empirical facts are less severe. Finally, panels (c) and (d) similarly show that not imposing a CE equilibrium would widen the range of price stickiness for which the three empirical facts are matched. In general, panels (a) and (d) highlight that both CE equilibrium and myopia can shrink the parametric space for which the three empirical facts are matched. However, it is important to emphasize that even when γ is considered a "free" parameter, it has to be somewhat close to values implied by the CE equilibrium for the expectations formation process to match the evidence on forecast errors.

A.4 Infinite Horizon Optimal Rules

In this section, I analyze whether the infinite horizon (IH) structure of the optimal decision rules matters for the forecasting error results and how. Recall that the ALM for inflation is currently given by $\hat{\pi}_t = a\hat{mc}_t + b\hat{\pi}_{t-1}$, where $a = \frac{\kappa}{1-\alpha\beta\rho n}$ and $b = \frac{\beta n(1-\alpha)}{1-\alpha\beta n\gamma^*}(\gamma^*)^2$. Without IH, the structure of the ALM remains intact but parameters a and b change to $\bar{a}$ and $\bar{b}$, respectively, that is,

$$\hat{\pi}_t = \bar{a}\hat{mc}_t + \bar{b}\hat{\pi}_{t-1} \tag{A.9}$$

where $\bar{a} = \kappa$ and $\bar{b} = \beta n(\gamma^*)^2$. Following the proof of Proposition 1, one can easily show that when IH is relaxed, an equilibrium γ^* always exists and it is unique. Moreover, I note that $\bar{b} > b$, which in turn implies that the equilibrium persistence of inflation would be higher when IH optimal rules are relaxed. When IH is relaxed, however, it is not possible to study the effects that a change in the price stickiness would have on the equilibrium persistence of inflation. The effects of myopia on γ^* would be the same as in the case of IH (see Corollary 1).

The implications of relaxing IH would be similar to the ones arising from decreasing price stickiness examined in Figure 2 in the main text. Specifically, with higher γ^* the region (in the (n, ρ) parametric space) where $K_4 > 0$ and $M_4 < 0$ would shift downwards.

Figure 3 compares the regions of delayed over-shooting, over-reaction to current realizations, and under-reaction to forecast revisions when IH is imposed versus not. First, relative to IH, the area of delayed over-shooting expands to accommodate more myopia. Second, compared with IH, the region of parameters for which there is under-reaction to forecast revisions and over-reaction to current realizations shifts downward.

A.5 Information Set Timing

The assumption is that firms do not observe inflation at the time of forecast, therefore, forecasts about inflation are based on inflation in period $t-1$. In this section, I discuss the implications

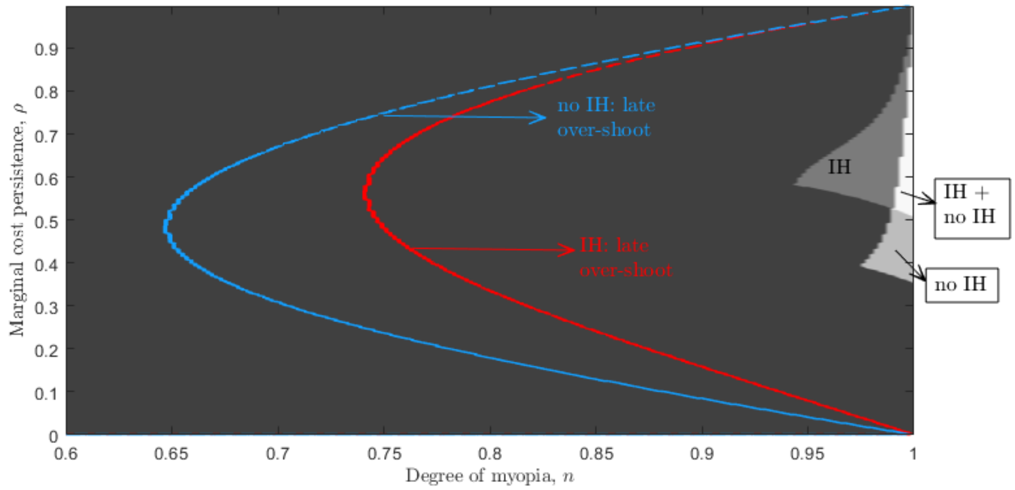


Figure 3: Regions of delayed over-shooting, under-reaction to ex-ante forecast revisions, and over-reaction to current realizations. Infinite horizon: delayed over-shooting holds to the right of the red curve; under-reaction to ex-ante forecast revisions and over-reaction to current realizations hold in the white and medium dark gray regions. No infinite horizon: delayed over-shooting occurs to the right of the blue curve; under-reaction to ex-ante forecast revisions and over-reaction to current realizations occur in the white and light gray regions. Parameterization is as in Figure 1.

that lifting this timing assumption has for Propositions 1-3. To keep things simple, I abstract from myopia.

In the particular New Keynesian pricing model considered in Section 2 in the main text, $t-1$ timing matters for Propositions 1-3. In absence of other shocks or lags of inflation in the optimal pricing rule, having inflation expectations depending on inflation in t implies that the actual law of motion for inflation would depend *only* on the marginal cost. Along the CE equilibrium, $\gamma^* = \rho$, and therefore, the CE equilibrium coincides with the RE one. However, that is only due to not accounting for other frictions, such as indexation to past inflation.

To see this, let's consider the optimal pricing rule when there is indexation to past inflation as assumed in the general equilibrium model:

$$\hat{p}^*_{jt} = \widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h \left[(1-\alpha\beta)\hat{mc}_{j,t+h} + \alpha\beta(\hat{\pi}_{t+h+1} - \rho_\pi \hat{\pi}_{t+h-1})\right] \tag{A.10}$$

where $\hat{p}^*_{jt} = \frac{\alpha}{1-\alpha}(\hat{\pi}_t - \rho_\pi \hat{\pi}_{t-1})$.[1] Assuming that the marginal cost is exogenously given, that expectations are formed based on inflation in period t, and abstracting from myopia for simplicity, the actual law of motion can be written as

$$\hat{\pi}_t = a\hat{mc}_t + b\hat{\pi}_{t-1} \tag{A.11}$$

where $a = \frac{\kappa(1-\alpha\beta\gamma)}{(1-\alpha\beta\rho)((1-\alpha\beta\gamma)-\beta(1-\alpha)(\gamma-\rho_\pi\alpha\beta))}$ and $b = \frac{\beta\rho_\pi(1-\alpha\beta\gamma)}{(1-\alpha\beta\gamma)-\beta(1-\alpha)(\gamma-\rho_\pi\alpha\beta)}$. A CE equilibrium occurs when $\gamma = F(\gamma) = \frac{b+\rho}{1+b\rho}$ and, following the result of Proposition 2 in Hommes and Zhu (2014), the equilibrium is unique.

[1] See derivations in Section C.

Along the CE equilibrium, $\gamma^* = (b+\rho)/(1+b\rho)$. Then, it is relatively straightforward to show that the following conditions need to be satisfied for delayed over-shooting and simultaneous over- and under-reaction in forecast errors:

- Late over-shooting: $\max(\rho^h, b^h) < \gamma^h < (b^{h+1} - \rho^{h+1})/(b-\rho)$
- Under-reaction to revisions and over-reaction to current inflation: $\frac{C(h)}{C(0)} < \gamma^h < \frac{C(h)-\gamma C(h+1)}{C(0)-\gamma C(1)}$

where $C(h) = \mathbb{E}(\hat{\pi}_{t+h}\hat{\pi}_t)$. Figure 4 visualizes the parametric space (ρ, γ) for which the three empirical facts are matched, together with the γ along the CE equilibrium. The figure clearly shows that forming forecasts based on information in period t can match the three empirical facts.

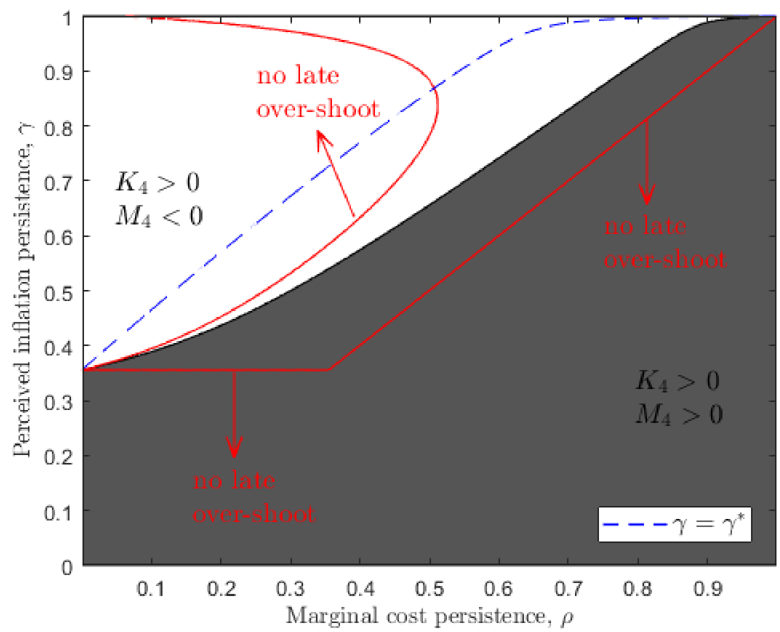


Figure 4: Regions of delayed over-shooting, under-reaction to ex-ante forecast revisions, and over-reaction to current realizations. Region within the red borders: delayed over-shooting; white region: $K_4 > 0$ and $M_4 < 0$; dark gray region: $K_4 > 0$ and $M_4 > 0$. The dashed blue curve indicates $\gamma = \gamma^*$. I set $\rho_\pi = 0.5$ and $n = 1$. The rest of parameters are set as in Figure 1.

B AR(1) Forecasting Rules and Delayed Over-shooting

This section discusses the conditions that need to be satisfied in order for AR(1) forecasting rules to generate delayed over-shooting. While delayed over-shooting can occur with more sophisticated PLMs as well, AR(1) processes are relatively easier to work with and deliver the desired outcome in terms of late over-shooting. To fix ideas, consider a general actual law of motion for the vector of endogenous variables S_t:

$$S_t = \boldsymbol{C} + \boldsymbol{A}S_{t-1} + \boldsymbol{B}\mathcal{E}_t \tag{B.1}$$

where $\boldsymbol{A}$ is a matrix with eigenvalues within the unit circle with at least one positive eigenvalue. Further, let the PLM about variable $S_{it} \in S_t$ be described by

$$S_{it} = \delta + \gamma(S_{i,t-1} - \delta) + \psi_t, \;\; \psi_t \sim WN \tag{B.2}$$

where $\gamma \in [0, 1)$. Provided that the response of $S_{i,t+k}$ to some innovation ε_{jt} in $\mathcal{E}_t$ does not switch sign for any k, Proposition 1 shows that γ exceeding the largest eigenvalue of $\boldsymbol{A}$ is a sufficient condition for forecasters to over-react with a delay.

Proposition 1 *Consider the actual law of motion in (B.1) and the PLM for some variable S_{it} in S_t in (B.2). Suppose that the sign of the response of $S_{i,t+k}$ to some innovation ε_{jt} in $\mathcal{E}_t$ is preserved for any $k \geq 1$. Then, γ being higher than the largest eigenvalue of $\boldsymbol{A}$ guarantees delayed over-shooting in the response of the h-period ahead forecast errors of S_i.*

Proof. See Section D.7. ■

For example, suppose that the actual law of motion is an AR(2) process with parameters $0 < \rho_1, \rho_2 < 1$, $\boldsymbol{A} = \begin{bmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{bmatrix}$. The largest eigenvalue of $\boldsymbol{A}$ is $\frac{\rho_1 + \sqrt{\rho_1^2 + 4\rho_2}}{2}$. It follows from Proposition 1 that $\gamma > \frac{\rho_1 + \sqrt{\rho_1^2 + 4\rho_2}}{2}$ is sufficient to have delayed over-shooting.

C DSGE Model

C.1 Non-linear Model

Households. There is a continuum of identical households, $i \in [0, 1]$, that consume from a set of differentiated goods, supply labor, and invest in riskless one-period bonds. First, households solve for the optimal allocation of consumption across differentiated goods, produced by monopolistically competitive firms $j \in [0, 1]$, i.e.,

$$\min_{c_{it}(j)} \int_{j=0}^{1} P_{jt} c_{it}(j) dj$$

s.t.

$$c_{it} = \left[\int_{j=0}^{1} c_{it}(j)^{\frac{\zeta-1}{\zeta}} dj \right]^{\frac{\zeta}{\zeta-1}} \tag{C.1}$$

and

$$P_t = \left[\int_{j=0}^{1} P_{jt}^{1-\zeta} \right]^{\frac{1}{1-\zeta}} \tag{C.2}$$

where ζ is the elasticity of substitution among the differentiated goods. The corresponding Lagrangian is

$$\mathcal{L}_{it} = \min_{c_{it}(j)} \int_{j=0}^{1} P_{jt} c_{it}(j) dj + \chi_{it} \left(c_{it} - \left[\int_{j=0}^{1} (c_{it}(j))^{\frac{\zeta-1}{\zeta}} dj \right]^{\frac{\zeta}{\zeta-1}} \right)$$

where χ_{it} is the Lagrangian multiplier for the Dixit-Stiglitz consumption aggregator in (C.1). The first-order condition is

$$c_{it}(j) = \left(\frac{\chi_{it}}{P_{jt}}\right)^{\zeta} c_{it} \tag{C.3}$$

Plugging the expression for $c_{it}(j)$ above into (C.1) and rearranging terms,

$$\chi_{it} = \left[\int_{j=0}^{1} P_{jt}^{1-\zeta} dj\right]^{\frac{1}{1-\zeta}}$$

This implies further that

$$c_{it}(j) = \left(\frac{P_{jt}}{P_t}\right)^{-\zeta} c_{it} \tag{C.4}$$

Equation (C.4) defines the optimal demand of the i^{th} household for the j^{th} good. The intertemporal problem for the household is to

$$\max_{c_{it}, H_{it}, B_{it}} \widetilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} \beta^h \left(ln(c_{i,t+h} - \eta c_{i,t+h-1}) - \psi \frac{H_{i,t+h}^{1+\varphi}}{1+\varphi} \right)$$

with budget constraint satisfying

$$\xi_{t-1} R_{t-1} B_{i,t-1} = B_{it} - W_t H_{it} - \int_{j=0}^{1} D_{it}(j) dj + \int_{j=0}^{1} P_{jt} c_{it}(j) dj$$

where H_{it} is labor supply; R_{t-1} gross return on nominal bond choice, $B_{i,t-1}$; W_t nominal wage; $D_{it}(j)$ nominal dividends from the j^{th} firm; and ξ_{t-1} is a shock to the return on bonds. Households internalize their optimal demand for good j into their intertemporal maximization problem, therefore

$$\int_{j=0}^{1} P_{jt} c_{it}(j) dj = P_t c_{it}$$

The budget constraint can be rewritten in real terms as

$$\xi_{t-1} R_{t-1} \frac{b_{i,t-1}}{\pi_t} = b_{it} - w_t H_{it} - d_{it} + c_{it} \tag{C.5}$$

where $w_t = W_t/P_t$ is the real wage; $b_{it} = B_{it}/P_t$ denotes real bond holdings; and $d_{it} = D_{it}/P_t$ with $\int_{j=0}^{1} D_{it}(j) dj = D_{it}$ denotes and real dividends. The first-order conditions (FOC) with respect to consumption, bonds, and hours, respectively, are

$$(c_{it} - \eta c_{i,t-1})^{-1} - \beta\eta \widetilde{\mathbb{E}}_{it} (c_{i,t+1} - \eta c_{it})^{-1} = \lambda_{it} \tag{C.6}$$

$$\lambda_{it} = \beta \widetilde{\mathbb{E}}_{it} R_t \frac{\lambda_{i,t+1}}{\pi_{t+1}} \xi_t \tag{C.7}$$

$$\psi H_{it}^{\varphi} = \lambda_{it} w_t \tag{C.8}$$

where $w_t = \frac{W_t}{P_t}$ is the real wage.

Firms. There is a continuum of household-owned monopolistically competitive firms, $j \in [0,1]$, that optimize with respect to price and labor demand. The production technology of each firm is

$$y_{jt} = Z_t h_{jt}^{a_h} \tag{C.9}$$

where Z_t and h_{jt} are a technology shock and labor demand, respectively, and $0 < a_h \leq 1$. The technology shock, Z_t, has a deterministic growth rate $\Upsilon > 1$ and can be written as $Z_t = z_t \Upsilon^t$, where z_t follows a stationary process.

The price optimization problem is subject to Calvo price stickiness as in Section A.1. Differently from Section A.1, if firms cannot optimize the price they can still adjust prices according to

$$P_{j,t+h} = P_{j,t+h-1}(\pi_{t+h-1})^{\rho_\pi} = P_{jt}\left(\frac{P_{t+h-1}}{P_{t-1}}\right)^{\rho_\pi} \tag{C.10}$$

where $0 \leq \rho_\pi < 1$. Given the price aggregator in (C.2) and the nominal rigidities firms face, we have

$$P_t = \left[\alpha\left(P_{t-1}\left(\frac{P_{t-1}}{P_{t-2}}\right)^{\rho_\pi}\right)^{1-\zeta} + (1-\alpha)(P_t^*)^{1-\zeta}\right]^{\frac{1}{1-\zeta}} \tag{C.11}$$

Each firm chooses the optimal price that will maximize the present discounted value of real profits such that the demand for its good is satisfied, and then hire the optimal amount of labor hours that will minimize production costs. Using backward induction, I solve the cost minimization problem first,

$$\mathcal{L}_{jt} = \min_{h_{jt}} w_t h_{jt} + mc_{jt}(y_{jt} - Z_t h_{jt}^{a_h}) \tag{C.12}$$

where mc_{jt} is the real marginal cost of production. The FOC with respect to labor reads

$$mc_{jt} = \frac{w_t}{a_h Z_t h_{jt}^{a_h - 1}} \tag{C.13}$$

The intermediate firms' problem is

$$\max_{P_{jt}^*} \widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{t,t+h}\left(\frac{P_{jt}^*}{P_{t+h}}\left(\frac{P_{t+h-1}}{P_{t-1}}\right)^{\rho_\pi} y_{j,t+h} - w_{t+h}h_{j,t+h}\right) \tag{C.14}$$

where $Q_{t,t+h} = \frac{\lambda_{t+h}}{\lambda_t}$ denotes the stochastic discount factor. Aggregating $c_{it}(j)$ across households in (C.3), we have that the demand faced by the j^{th} firm in period $(t+h)$ is

$$y_{j,t+h} = \left(\frac{P_{jt}^*}{P_{t+h}}\right)^{-\zeta} y_{t+h} \tag{C.15}$$

Substituting for $y_{j,t+h}$ and w_{t+h}, the pricing problem becomes

$$\max_{P_{jt}^*} \widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{t,t+h} y_{t+h}\left(\left(\frac{P_{jt}^*}{P_{t+h}}\right)^{1-\zeta}\left(\frac{P_{t+h-1}}{P_{t-1}}\right)^{\rho_\pi(1-\zeta)} - a_h mc_{j,t+h}\left(\frac{P_{jt}^*}{P_{t+h}}\right)^{-\zeta}\left(\frac{P_{t+h-1}}{P_{t-1}}\right)^{-\rho_\pi\zeta}\right) \tag{C.16}$$

The first-order condition with respect to P^*_{jt} reads

$$\widetilde{\mathbb{E}}_{jt}\sum_{h=0}^{\infty}(\alpha\beta)^h Q_{t,t+h}\pi_{t-1,t+h-1}^{-\rho_\pi\zeta}P_{t+h}^{\zeta-1}y_{t+h}\left(a_h\zeta mc_{j,t+h}P_{t+h}-(\zeta-1)P^*_{jt}\pi^{\rho_\pi}_{t-1,t+h-1}\right)=0 \tag{C.17}$$

Monetary Policy. The central bank controls nominal interest rates through a Taylor rule that reacts to inflation and output gap deviations from their steady-state values, with some interest rate smoothing, i.e.,

$$\frac{R_t}{\bar{R}}=\left(\frac{R_{t-1}}{\bar{R}}\right)^{\rho_r}\left(\frac{\pi_t}{\bar{\pi}}\right)^{(1-\rho_r)\phi_\pi}\left(\frac{y_t/y_{t-1}}{\Upsilon}\right)^{(1-\rho_r)\phi_y}e^{\sigma_v\varepsilon^v_t},\ \varepsilon^v_t\sim\mathcal{N}(0,1) \tag{C.18}$$

where $\bar{\pi}$ denotes the inflation target and $\rho_r\in[0,1)$.

Rescaling. In order to induce stationarity, output, consumption, wages, and bond holdings are detrended using Υ^t, that is, $\widetilde{y}_{jt}=y_{jt}/\Upsilon^t$, $\widetilde{y}_t=y_t/\Upsilon^t$, $\widetilde{c}_{it}=c_{it}/\Upsilon^t$, $\widetilde{w}_t=w_t/\Upsilon^t$, $\widetilde{b}_{it}=b_{it}/\Upsilon^t$. The Lagrange multipliers are rescaled as $\widetilde{\lambda}_{it}=\lambda_{it}\Upsilon^t$ and $\widetilde{\lambda}_t=\lambda_t\Upsilon^t$. The stochastic discount factor is re-written as $Q_{t,t+h}=\frac{\widetilde{Q}_{t,t+h}}{\Upsilon^h}$ where $\widetilde{Q}_{t,t+h}=\frac{\widetilde{\lambda}_{t+h}}{\widetilde{\lambda}_t}$. The new system of equations is given by

$$\left(\widetilde{c}_{it}-\frac{\eta}{\Upsilon}\widetilde{c}_{i,t-1}\right)^{-1}-\frac{\beta\eta}{\Upsilon}\widetilde{\mathbb{E}}_{it}\left(\widetilde{c}_{i,t+1}-\frac{\eta}{\Upsilon}\widetilde{c}_{it}\right)^{-1}=\widetilde{\lambda}_{it} \tag{C.19}$$

$$\widetilde{\lambda}_{it}=\frac{\beta}{\Upsilon}\widetilde{\mathbb{E}}_{it}R_t\frac{\widetilde{\lambda}_{i,t+1}}{\pi_{t+1}}\xi_t \tag{C.20}$$

$$\psi H^{\varphi}_{it}=\widetilde{\lambda}_{it}\widetilde{w}_t \tag{C.21}$$

$$\frac{R_{t-1}\widetilde{b}_{i,t-1}}{\Upsilon\pi_t}=\widetilde{b}_{it}-\widetilde{w}_tH_{it}-\widetilde{d}_{it}+\widetilde{c}_{it} \tag{C.22}$$

$$mc_{jt}=\frac{\widetilde{w}_t}{a_hz_th^{a_h-1}_{jt}} \tag{C.23}$$

$$\widetilde{y}_{jt}=\left(\frac{P^*_{jt}}{P_t}\right)^{-\zeta}\widetilde{y}_t \tag{C.24}$$

$$\widetilde{y}_{jt}=z_th^{a_h}_{jt} \tag{C.25}$$

$$\widetilde{\mathbb{E}}_{jt}\sum_{h=0}^{\infty}(\alpha\beta)^h\widetilde{Q}_{t,t+h}\pi_{t-1,t+h-1}^{-\rho_\pi\zeta}P_{t+h}^{\zeta-1}\widetilde{y}_{t+h}\left(a_h\zeta mc_{j,t+h}P_{t+h}-(\zeta-1)P^*_{jt}\pi^{\rho_\pi}_{t-1,t+h-1}\right)=0 \tag{C.26}$$

$$\frac{R_t}{\bar{R}}=\left(\frac{R_{t-1}}{\bar{R}}\right)^{\rho_r}\left(\frac{\pi_t}{\bar{\pi}}\right)^{(1-\rho_r)\phi_\pi}\left(\frac{\widetilde{y}_t}{\widetilde{y}_{t-1}}\right)^{(1-\rho_r)\phi_y}e^{\sigma_v\varepsilon^v_t},\ \varepsilon^v_t\sim\mathcal{N}(0,1) \tag{C.27}$$

Steady-state Equilibrium. I calculate steady-state values, given $\bar{\xi}=1$, $\bar{z}=1$, $\bar{v}=1$, and $\bar{\pi}=1$.

$$\beta\bar{R}=\Upsilon \tag{C.28}$$

$$\bar{\lambda}=\frac{1-\beta\eta/\Upsilon}{\bar{c}(1-\eta/\Upsilon)} \tag{C.29}$$

$$\bar{w} = \frac{\psi(1-\eta/\Upsilon)}{1-\beta\eta/\Upsilon}\bar{H}^{\varphi}\bar{c} \tag{C.30}$$

$$\bar{d} = \bar{c} - \frac{1-\beta}{\beta}\bar{b} - \bar{w}\bar{H} \tag{C.31}$$

$$\bar{y} = \bar{h}^{a_h} \tag{C.32}$$

$$\bar{mc} = \frac{\zeta - 1}{a_h\zeta} \tag{C.33}$$

where $\bar{b} = \frac{\bar{B}}{\bar{P}}$ and $\bar{d} = \frac{\bar{D}}{\bar{P}}$ denote steady-state bond holdings and dividends in real terms.

C.2 Log-linearized Model

I now log-linearize the rescaled model around steady-state values, and denote log-linearized variables with a hat on top.

Households. Log-linearizing (C.19) and (C.20) around steady states generates

$$\hat{cc}_{it} = \widetilde{\mathbb{E}}_{it}\hat{cc}_{i,t+1} - \left(1 - \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_{it}(\hat{R}_t - \hat{\pi}_{t+1} + \hat{\xi}_t) \tag{C.34}$$

where $\hat{cc}_{it} = \hat{c}_{it} - \frac{\eta}{\Upsilon}\hat{c}_{i,t-1} - \frac{\beta\eta}{\Upsilon}\widetilde{\mathbb{E}}_{it}\left(\hat{c}_{i,t+1} - \frac{\eta}{\Upsilon}\hat{c}_{it}\right)$. One can make inferences about $\widetilde{\mathbb{E}}_{it}\hat{cc}_{i,t+1}$ by iterating the Euler equation above, i.e.,

$$\hat{cc}_{i,t+1} = \widetilde{\mathbb{E}}_{i,t+1}\hat{cc}_{i,t+2} - \left(1 - \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_{i,t+1}(\hat{R}_{t+1} - \hat{\pi}_{t+2} + \hat{\xi}_{t+1})$$

So,

$$\begin{aligned}
\widetilde{\mathbb{E}}_{it}\hat{cc}_{i,t+1} &= \widetilde{\mathbb{E}}_{it}\widetilde{\mathbb{E}}_{i,t+1}\hat{cc}_{i,t+2} - \left(1 - \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_{it}\widetilde{\mathbb{E}}_{i,t+1}(\hat{R}_{t+1} - \hat{\pi}_{t+2} + \hat{\xi}_{t+1}) \\
&= \widetilde{\mathbb{E}}_{it}\hat{cc}_{i,t+2} - \left(1 - \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_{it}(\hat{R}_{t+1} - \hat{\pi}_{t+2} + \hat{\xi}_{t+1})
\end{aligned}$$

where the second equality is an application of the law of iterative expectations. Plugging expectations into the log-linear individual Euler equation, we get

$$\hat{cc}_{it} = \widetilde{\mathbb{E}}_{it}\hat{cc}_{i,t+2} - \left(1 - \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{1}(\hat{R}_{t+h} - \hat{\pi}_{t+h+1} + \hat{\xi}_{t+h})$$

Similarly, the h-periods-ahead forwardly iterated Euler equation reads

$$\hat{cc}_{it} = \widetilde{\mathbb{E}}_{it}\hat{cc}_{i,t+h} - \left(1 - \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_{it}\sum_{l=0}^{h-1}(\hat{R}_{t+l} - \hat{\pi}_{t+l+1} + \hat{\xi}_{t+l}) \tag{C.35}$$

It is worth noting that if households knew that everyone is subject to the same preference shocks, and that they all have the same preferences over consumption and labor, then they would know that in the infinite future, consumption is expected to be at its steady state, implying that

$\lim_{h\to\infty} \widetilde{\mathbb{E}}_{it}\hat{c}c_{i,t+h} = 0$. This would further imply that households would use the one-step-ahead Euler equation, as under RE. However, households have imperfect knowledge about the rest of the population, and one needs to combine (C.35) with the infinitely forward iterated household's budget constraint in (C.22):

$$
\begin{aligned}
\widetilde{b}_{i,t-1} &= \frac{\Upsilon\pi_t}{R_{t-1}}(\widetilde{b}_{it} - \widetilde{w}_t H_{it} - \widetilde{w}_t H_{it} + \widetilde{c}_{it}) \\
&= \Upsilon^2 \widetilde{\mathbb{E}}_{it}\left[\frac{\pi_{t,t+1}}{R_{t-1,t}}\left(\widetilde{b}_{i,t+1} - \sum_{h=0}^{1}(\widetilde{w}_{t+h}H_{i,t+h} + \widetilde{d}_{i,t+h} - \widetilde{c}_{i,t+h})\right)\right] \\
&= \ldots \\
&= \lim_{h\to\infty}\widetilde{\mathbb{E}}_{it}\left(\frac{\Upsilon^{h+1}\pi_{t,t+h+1}}{R_{t-1,t+h}}\widetilde{b}_{i,t+h+1}\right) - \widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\frac{\Upsilon^{h+1}\pi_{t,t+h+1}}{R_{t-1,t+h}}(\widetilde{w}_{t+h}H_{i,t+h} + \widetilde{d}_{i,t+h} - \widetilde{c}_{i,t+h}) \\
&= \widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\frac{\Upsilon^{h+1}\pi_{t,t+h+1}}{R_{t-1,t+h}}(\widetilde{c}_{i,t+h} - \widetilde{w}_{t+h}H_{i,t+h} - \widetilde{d}_{i,t+h})
\end{aligned}
$$

where $R_{t-1,t+h} = \prod_{l=t-1}^{t+h} R_l$ and $\pi_{t,t+h+1} = \prod_{l=t}^{t+h+1}\pi_l$. To get the last equality I impose the appropriate no-Ponzi constraint, i.e., $\lim_{h\to\infty}\widetilde{\mathbb{E}}_{it}\left(\frac{\Upsilon^{h+1}\pi_{t,t+h+1}}{R_{t-1,t+h}}\widetilde{b}_{i,t+h+1}\right) = 0$. The log-linearized version of the iterated budget constraint is:

$$
\begin{aligned}
\bar{b}\hat{b}_{i,t-1} = {}& \widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\Upsilon^{h+1}\frac{\bar{\pi}_{t,t+h+1}}{\bar{R}_{t-1,t+h}}\bar{c}(\hat{\pi}_{t,t+h+1} - \hat{R}_{t-1,t+h} + \hat{c}_{i,t+h}) \\
& - \widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\Upsilon^{h+1}\frac{\bar{\pi}_{t,t+h+1}}{\bar{R}_{t-1,t+h}}\bar{w}\bar{H}(\hat{\pi}_{t,t+h+1} - \hat{R}_{t-1,t+h} + \hat{w}_{t+h} + \hat{H}_{i,t+h}) \\
& - \widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\Upsilon^{h+1}\frac{\bar{\pi}_{t,t+h+1}}{\bar{R}_{t-1,t+h}}\bar{d}(\hat{\pi}_{t,t+h+1} - \hat{R}_{t-1,t+h} + \hat{d}_{i,t+h})
\end{aligned}
$$

Using (C.29), $\frac{\bar{\pi}_{t,t+h+1}}{\bar{R}_{t-1,t+h}} = \frac{\bar{\pi}^{h+1}}{\bar{R}^{h+1}} = \left(\frac{\beta}{\Upsilon}\right)^{h+1}$. Substituting for $\frac{\bar{\pi}_{t,t+h+1}}{\bar{R}_{t-1,t+h}}$, one gets

$$
\begin{aligned}
\bar{b}\hat{b}_{i,t-1} = {}& \bar{c}\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}\hat{c}_{i,t+h} - \widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}\left[\bar{w}\bar{H}(\hat{w}_{t+h} + \hat{H}_{i,t+h}) + \bar{d}\hat{d}_{i,t+h}\right] \\
& + (\bar{c} - \bar{w}\bar{H} - \bar{d})\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}(\hat{\pi}_{t,t+h+1} - \hat{R}_{t-1,t+h})
\end{aligned} \tag{C.36}
$$

Next, recall that $\hat{c}c_{it} = \hat{c}_{it} - \frac{\eta}{\Upsilon}\hat{c}_{i,t-1} - \frac{\beta\eta}{\Upsilon}\widetilde{\mathbb{E}}_{it}(\hat{c}_{i,t+1} - \frac{\eta}{\Upsilon}\hat{c}_{it})$, from which it follows that $\hat{c}_{it} = \hat{c}c_{it} + \frac{\eta}{\Upsilon}\hat{c}_{i,t-1} + \frac{\beta\eta}{\Upsilon}\widetilde{\mathbb{E}}_{it}(\hat{c}_{i,t+1} - \frac{\eta}{\Upsilon}\hat{c}_{it})$. Substituting for $\hat{c}_{i,t+h}$ into (C.36), I rewrite the intertermporal budget constraint as

$$\bar{b}\hat{b}_{i,t-1} = \bar{c}\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}\left(\hat{c}c_{i,t+h} + \frac{\eta}{\Upsilon}\hat{c}_{i,t+h-1} + \frac{\beta\eta}{\Upsilon}\left(\hat{c}_{i,t+h+1} - \frac{\eta}{\Upsilon}\hat{c}_{i,t+h}\right)\right)$$
$$- \widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}\left[\bar{w}\bar{H}(\hat{w}_{t+h} + \hat{H}_{i,t+h}) + \bar{d}\hat{d}_{i,t+h}\right] \tag{C.37}$$
$$+ (\bar{c} - \bar{w}\bar{H} - \bar{d})\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}(\hat{\pi}_{t,t+h+1} - \hat{R}_{t-1,t+h})$$

From (C.35), one can isolate $\widetilde{\mathbb{E}}_{it}\hat{c}c_{i,t+h}$ and substitute for it into (C.36):

$$\bar{b}\hat{b}_{i,t-1} = \bar{c}\beta\left(\frac{1}{1-\beta}\hat{c}c_{it} + \widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h}\left(\frac{\eta}{\Upsilon}\hat{c}_{i,t+h-1} + \frac{\beta\eta}{\Upsilon}\left(\hat{c}_{i,t+h+1} - \frac{\eta}{\Upsilon}\hat{c}_{i,t+h}\right)\right)\right)$$
$$+ \beta\bar{c}\left(1 - \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}(\hat{R}_{t+h} - \hat{\pi}_{t+h+1} + \hat{\xi}_{t+h})$$
$$- \widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}\left[\bar{w}\bar{H}(\hat{w}_{t+h} + \hat{H}_{i,t+h}) + \bar{d}\hat{d}_{i,t+h}\right]$$
$$+ (\bar{c} - \bar{w}\bar{H} - \bar{d})\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}(\hat{\pi}_{t,t+h+1} - \hat{R}_{t-1,t+h})$$

Isolating $\hat{c}c_{it}$, one retrieves the individual demand in terms of $\hat{c}c_{it}$,

$$\hat{c}c_{it} = \frac{\bar{b}(1-\beta)}{\beta\bar{c}}\hat{b}_{i,t-1} - (1-\beta)\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h}\left(\frac{\eta}{\Upsilon}\hat{c}_{i,t+h-1} + \frac{\beta\eta}{\Upsilon}\left(\hat{c}_{i,t+h+1} - \frac{\eta}{\Upsilon}\hat{c}_{i,t+h}\right)\right)$$
$$+ \frac{1-\beta}{\bar{c}}\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h}\left[\bar{w}\bar{H}(\hat{w}_{t+h} + \hat{H}_{i,t+h}) + \bar{d}\hat{d}_{i,t+h}\right] - \beta\left(1 - \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h}(\hat{R}_{t+h} - \hat{\pi}_{t+h+1} + \hat{\xi}_{t+h})$$
$$- \frac{(1-\beta)(\bar{c} - \bar{w}\bar{H} - \bar{d})}{\bar{c}}\widetilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h}(\hat{\pi}_{t,t+h+1} - \hat{R}R_{t-1,t+h}) \tag{C.38}$$

Define $\hat{x}_t = \hat{y}_t - \frac{\eta}{\Upsilon}\hat{y}_{t-1}$. Then, aggregating equation (C.38), imposing market clearing conditions such that $\bar{c}\hat{c}_t = \bar{y}\hat{y}_t = \bar{w}\bar{H}(\hat{w}_t + \hat{H}_t) + \bar{d}\hat{d}_t$, $\hat{c}c_t = \hat{y}y_t$, $(\bar{c} - \bar{w}\bar{H} - \bar{d}) = 0$, and $\hat{b}_t = 0$ (since households are homogeneous) one gets

$$\hat{x}_t = \left(1 - \beta + \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_t\hat{x}_{t+1} + \widetilde{\mathbb{E}}_t\sum_{h=0}^{\infty}\beta^{h}\left[\beta(1-\beta)\left(1 - \frac{\eta}{\Upsilon}\right)\hat{x}_{t+h+2} - \left(1 - \frac{\beta\eta}{\Upsilon}\right)(\hat{R}_{t+h} - \hat{\pi}_{t+h+1}) + \hat{e}_{t+h}\right] \tag{C.39}$$

where $\hat{e}_t = -\left(1 - \frac{\beta\eta}{\Upsilon}\right)\hat{\xi}_t$ is such that

$$\hat{e}_t = \rho_e\hat{e}_{t-1} + \sigma_e\varepsilon_t^e,\ \varepsilon_t^e \sim \mathcal{N}(0,1) \tag{C.40}$$

Applying the myopic adjustment to (C.39), the aggregate demand is rewritten as

$$\hat{x}_t = n\left(1-\beta+\frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_t^{\star}\hat{x}_{t+1}+\widetilde{\mathbb{E}}_t^{\star}\sum_{h=0}^{\infty}(\beta n)^h\left[n^2\beta(1-\beta)\left(1-\frac{\eta}{\Upsilon}\right)\hat{x}_{t+h+2}-\left(1-\frac{\beta\eta}{\Upsilon}\right)(\hat{R}_{t+h}-\hat{\pi}_{t+h+1})+\hat{e}_{t+h}\right] \tag{C.41}$$

Finally, substituting $\hat{x}_t=\hat{y}_t-\frac{\eta}{\Upsilon}\hat{y}_{t-1}$ delivers

$$\begin{aligned}\hat{y}_t &= \frac{\eta}{\Upsilon+n\eta\upsilon}\hat{y}_{t-1}+n\frac{\upsilon\Upsilon-n\beta\eta(1-\beta)(1-\frac{\eta}{\Upsilon})}{\Upsilon+n\eta\upsilon}\widetilde{\mathbb{E}}_t^{\star}\hat{y}_{t+1}+\frac{\beta n^2(1-\beta)(1-\frac{\eta}{\Upsilon})(\Upsilon-n\beta\eta)}{\Upsilon+n\eta\upsilon}\widetilde{\mathbb{E}}_t^{\star}\sum_{h=0}^{\infty}(\beta n)^h\hat{y}_{t+h+2}\\ &\quad -\frac{\Upsilon-\beta\eta}{\Upsilon+n\eta\upsilon}\widetilde{\mathbb{E}}_t^{\star}\sum_{h=0}^{\infty}(\beta n)^h\left(\hat{R}_{t+h}-\hat{\pi}_{t+h+1}\right)+\frac{1}{(1-\beta n\rho_e)(\Upsilon+n\eta\upsilon)}\hat{e}_t\end{aligned} \tag{C.42}$$

where $\upsilon=1-\beta+\frac{\beta\eta}{\Upsilon}$.

Firms. Log-linearizing firms' optimal price condition, we get,

$$\hat{P}_{jt}^*-\hat{P}_t=\widetilde{\mathbb{E}}_{jt}\sum_{h=0}^{\infty}(\alpha\beta)^h\left[(1-\alpha\beta)\hat{mc}_{j,t+h}+\alpha\beta(\hat{\pi}_{t+h+1}-\rho_\pi\hat{\pi}_{t+h-1})\right] \tag{C.43}$$

Define $\hat{p}_{jt}^*=\hat{P}_{jt}^*-\hat{P}_t$. The marginal cost of the j^{th} firm is given by

$$\begin{aligned}\hat{mc}_{j,t+h} &= \hat{w}_{t+h}-\hat{z}_{t+h}+(1-a_h)\hat{h}_{j,t+h}\\ &= \hat{w}_{t+h}-\hat{z}_{t+h}+\frac{1-a_h}{a_h}\hat{h}_{j,t+h}\\ &= \hat{w}_{t+h}-\frac{1}{a_h}\hat{z}_{t+h}+\frac{1-a_h}{a_h}y_{t+h}-\zeta\frac{1-a_h}{a_h}(\hat{P}_{j,t+h}-\hat{P}_{t+h})\\ &= \hat{w}_{t+h}-\frac{1}{a_h}\hat{z}_{t+h}+\frac{1-a_h}{a_h}y_{t+h}-\zeta\frac{1-a_h}{a_h}\left(\hat{p}_j^*-(\hat{\pi}_{t,t+h}-\rho_\pi\hat{\pi}_{t-1,t+h-1})\right)\end{aligned} \tag{C.44}$$

Therefore,

$$\begin{aligned}(1-\alpha\beta)\widetilde{\mathbb{E}}_{jt}\sum_{h=0}^{\infty}(\alpha\beta)^h\hat{mc}_{j,t+h} &= (1-\alpha\beta)\widetilde{\mathbb{E}}_{jt}\sum_{h=0}^{\infty}(\alpha\beta)^h\left[\hat{w}_{t+h}+\frac{1}{a_h}\hat{z}_{t+h}+\frac{1-a_h}{a_h}\hat{y}_{t+h}\right]-\zeta\frac{1-a_h}{a_h}\hat{p}_j^*\\ &\quad+\alpha\beta\zeta\frac{1-a_h}{a_h}\widetilde{\mathbb{E}}_{jt}\sum_{h=0}^{\infty}(\alpha\beta)^h\left(\hat{\pi}_{t+h+1}-\rho_\pi\hat{\pi}_{t+h}\right))\end{aligned}$$

On the other hand, making use of the labor supply equation and the log-linearized expression for $\hat{\lambda}_t$, we have

$$\begin{aligned}\hat{w}_{t+h}+\frac{1}{a_h}\hat{z}_{t+h}+\frac{1-a_h}{a_h}\hat{y}_{t+h} &= \varphi\hat{h}_{t+h}-\hat{\lambda}_{t+h}+\frac{1}{a_h}\hat{z}_{t+h}+\frac{1-a_h}{a_h}\hat{y}_{t+h}\\ &= \frac{1-\varphi}{a_h}\hat{z}_{t+h}+\frac{1-a_h+\varphi}{a_h}\hat{y}_{t+h}-\left(\hat{x}_{t+h}-\frac{\beta\eta}{\Upsilon}\hat{x}_{t+h+1}\right)\end{aligned}$$

$$\left(1+\zeta\frac{1-a_h}{a_h}\right)\hat{p}^*_{jt} = (1-\alpha\beta)\widetilde{\mathbb{E}}_{jt}\sum_{h=0}^{\infty}(\alpha\beta)^h\left(\frac{1-\varphi}{a_h}\hat{z}_{t+h}+\frac{1-a_h+\varphi}{a_h}\hat{y}_{t+h}-\left(\hat{x}_{t+h}-\frac{\beta\eta}{\Upsilon}\hat{x}_{t+h+1}\right)\right)$$
$$+\left(1+\zeta\frac{1-a_h}{a_h}\right)\alpha\beta\widetilde{\mathbb{E}}_{jt}\sum_{h=0}^{\infty}(\alpha\beta)^h(\hat{\pi}_{t+h+1}-\rho_\pi\hat{\pi}_{t+h-1}) \tag{C.45}$$

From (C.10), $\hat{p}^*_{jt} = \hat{P}^*_{jt} - \hat{P}_t = \frac{\alpha}{1-\alpha}(\hat{\pi}_t - \rho_\pi\hat{\pi}_{t-1}) = \frac{\alpha}{1-\alpha}\hat{\tilde{\pi}}_t$. Since all firms face the same optimal pricing condition above, I drop the subscript j.

$$\hat{\tilde{\pi}}_t = \widetilde{\mathbb{E}}_{jt}\sum_{h=0}^{\infty}(\alpha\beta)^h\left[\kappa\left((1-\varphi)\hat{z}_{t+h}+\omega\hat{y}_{t+h}-\left(\hat{x}_{t+h}-\frac{\beta\eta}{\Upsilon}\hat{x}_{t+h+1}\right)\right)+\alpha\beta\hat{\tilde{\pi}}_{t+h+1}\right] \tag{C.46}$$

where $\kappa = \frac{a_h(1-\alpha)(1-\alpha\beta)}{\alpha(a_h+\zeta(1-a_h))}$, and $\omega = \frac{1+\varphi-a_h}{a_h}$ is the elasticity of the marginal disutility of producing output with respect to output. Substituting for $\hat{x}_t = \hat{y}_t - \frac{\eta}{\Upsilon}\hat{y}_{t-1}$, the aggregated optimal pricing rule can be written as

$$\hat{\tilde{\pi}}_t = \kappa\left(\omega\hat{x}_t+\frac{\Upsilon}{\Upsilon-\eta\beta}\hat{y}_t\right)+\widetilde{\mathbb{E}}_t\sum_{h=0}^{\infty}(\alpha\beta)^h\left(\kappa\alpha\beta\left(\omega\hat{y}_{t+h+1}+\frac{\beta(\alpha\Upsilon-\eta)}{\alpha(\Upsilon-\eta\beta)}\hat{y}_{t+h+1}\right)+\beta(1-\alpha)\hat{\tilde{\pi}}_{t+h+1}+\hat{u}_{t+h}\right) \tag{C.47}$$

where $\hat{u}_t = \frac{\kappa(1-\varphi)}{a_h}\hat{z}_t$ is a supply shock assumed to follow an AR(1) process

$$\hat{u}_t = \rho_u\hat{u}_{t-1}+\sigma_u\varepsilon^u_t,\ \varepsilon^u_t \sim \mathcal{N}(0,1) \tag{C.48}$$

Applying the myopic adjustment yields

$$\hat{\tilde{\pi}}_t = \kappa\left(\omega\hat{y}_t+\frac{\Upsilon}{\Upsilon-\eta\beta}\hat{x}_t\right)+\widetilde{\mathbb{E}}^\star_t\sum_{h=0}^{\infty}(\alpha\beta n)^h\left(\kappa\alpha\beta n\left(\omega\hat{y}_{t+h+1}+\frac{(\alpha\Upsilon-\eta)}{\alpha(\Upsilon-\eta\beta)}\hat{x}_{t+h+1}\right)\right)$$
$$+\widetilde{\mathbb{E}}^\star_t\sum_{h=0}^{\infty}(\alpha\beta n)^h\left(n\beta(1-\alpha)\hat{\tilde{\pi}}_{t+h+1}+\hat{u}_{t+h}\right) \tag{C.49}$$

Substituting for $\hat{\tilde{\pi}}_t = \hat{\pi}_t - \rho_\pi\hat{\pi}_{t-1}$,

$$\hat{\pi}_t = \frac{1}{1+n\beta\rho_\pi(1-\alpha)}\left(\rho_\pi\hat{\pi}_{t-1}-\frac{\kappa\tau\eta}{\Upsilon}\hat{y}_{t-1}\right)+\frac{\kappa\left(\omega+\tau\left(1-\frac{n\beta\eta}{\Upsilon}\left(\alpha-\frac{\eta}{\Upsilon}\right)\right)\right)}{1+n\beta\rho_\pi(1-\alpha)}\hat{y}_t$$
$$+\frac{n\beta(1-\alpha)(1-\alpha\beta n\rho_\pi)}{1+n\beta\rho_\pi(1-\alpha)}\widetilde{\mathbb{E}}^\star_t\sum_{h=0}^{\infty}(\alpha\beta n)^h\hat{\pi}_{t+h+1}+\frac{1}{(1-\alpha\beta n\rho_u)(1+n\beta\rho_\pi(1-\alpha))}\hat{u}_t \tag{C.50}$$
$$+\frac{n\beta\kappa}{1+n\beta\rho_\pi(1-\alpha)}\left(\alpha\omega+\tau\left(\alpha-\frac{\eta}{\Upsilon}\right)\left(1-\frac{\alpha\beta n\eta}{\Upsilon}\right)\right)\widetilde{\mathbb{E}}^\star_t\sum_{h=0}^{\infty}(\alpha\beta n)^h\sum_{h=0}^{\infty}\hat{y}_{t+h+1}$$

where $\tau = \frac{\Upsilon}{\Upsilon-\beta\eta}$.

Monetary policy. The log-linearized version of the policy rule is

$$\hat{R}_t = \rho_r\hat{R}_{t-1}+(1-\rho_r)\phi_\pi\hat{\pi}_t+(1-\rho_r)\phi_y(\hat{y}_t-\hat{y}_{t-1})+\sigma_v\varepsilon_t \tag{C.51}$$

Model in matrix form. The aggregate economy model in matrix form is described by

$$A_0(\Theta)S_t = A_{-1}(\Theta)S_{t-1} + A_1(\Theta)\widetilde{\mathbb{E}}_t^\star S_{t+1} + \widetilde{\mathbb{E}}_t^\star \sum_{h=0}^{\infty} F^h A_2(\Theta)S_{t+h+2} + B(\Theta)\mathcal{E}_t \tag{C.52}$$

where $S_t = \begin{bmatrix}\hat{y}_t & \hat{\pi}_t & \hat{R}_t & (\hat{y}_t - \hat{y}_{t-1}) & \hat{e}_t & \hat{u}_t\end{bmatrix}'$; $\mathcal{E}_t = \begin{bmatrix}\varepsilon_t^e & \varepsilon_t^u & \varepsilon_t^v\end{bmatrix}'$; F is a zero matrix, with only the first two diagonal entries equal to βn and $\alpha\beta n$, respectively; and $\Theta = \{\alpha, \beta, n, \kappa, \eta, \rho_\pi, \omega, \Upsilon, \phi_\pi, \phi_y, \rho_r, \rho_e, \rho_u, \sigma_e,$

SAC learning. The perceived law of motion (PLM) in matrix form can be written as

$$S_t = \underbrace{\boldsymbol{\Delta}_{t-1} + \boldsymbol{\Gamma}_{t-1}(S_{t-1} - \boldsymbol{\Delta}_{t-1})}_{\text{PLM for aggregate endo var's}} + \underbrace{HS_{t-1}}_{\text{PLM for the shocks}} + \boldsymbol{\epsilon}_t \tag{C.53}$$

where $\boldsymbol{\Delta}_t = \begin{bmatrix}\boldsymbol{\delta}_t' & \mathbf{0}_{1x2}\end{bmatrix}'$; $\boldsymbol{\Gamma}_t = \begin{bmatrix}\boldsymbol{\gamma}_t & \mathbf{0}_{4x2} \\ \mathbf{0}_{2x4} & \mathbf{0}_{2x2}\end{bmatrix}$; H is a diagonal matrix with diagonal equal to $\begin{bmatrix}\mathbf{0}_{1x4} & \rho_e & \rho_u\end{bmatrix}'$; $\boldsymbol{\epsilon}_t = \begin{bmatrix}\boldsymbol{\psi}_t' & \sigma_e\varepsilon_t^e & \sigma_u\varepsilon_t^u\end{bmatrix}'$. The forecast of the state vector $h \geq 1$ periods ahead is described by

$$\widetilde{\mathbb{E}}_t^\star S_{t+h} = \underbrace{\boldsymbol{\Delta}_{t-1} + \boldsymbol{\Gamma}_{t-1}^{h+1}(S_{t-1} - \boldsymbol{\Delta}_{t-1})}_{\text{forecast of endo variables}} + \underbrace{H^h S_t}_{\text{forecast of shocks}} \tag{C.54}$$

Plugging (C.54) into (C.52), one can re-write the model as:

$$\widetilde{A}_0(\Theta)S_t = \widetilde{A}_1(\Theta, \boldsymbol{\Gamma}_{t-1})\boldsymbol{\Delta}_{t-1} + \widetilde{A}_2(\Theta, \boldsymbol{\Gamma}_{t-1})S_{t-1} + B\mathcal{E}_t \tag{C.55}$$

where

$$\widetilde{A}_0 = A_0 - A_1 H - \left(\sum_{h=0}^{\infty} F^h A_2 H^h\right) H^2$$

$$\widetilde{A}_1 = A_1(I - \Gamma_{t-1}^2)\Delta_{t-1} + \left(\sum_{h=0}^{\infty} F^h A_2 (I - \boldsymbol{\Gamma}_{t-1}^{h+3})\right) \Delta_{t-1}$$

$$\widetilde{A}_2 = A_{-1} + A_1\boldsymbol{\Gamma}_{t-1}^2 + \left(\sum_{h=0}^{\infty} F^h A_2 \boldsymbol{\Gamma}_{t-1}^h\right) \boldsymbol{\Gamma}_{t-1}^3$$

The infinite sums are defined as follows

$$\sum_{h=0}^{\infty} F^h = (I - F)^{-1}$$

$$vec\left(\sum_{h=0}^{\infty} F^h A_2 H^h\right) = (I - H \otimes F)^{-1} A_2(:)$$

$$\begin{aligned} vec\left(\sum_{h=0}^{\infty} F^h A_2 \boldsymbol{\Gamma}_{t-1}^h\right) &= vec(A_2 + F A_2 \boldsymbol{\Gamma}_{t-1} + F^2 A_2 \boldsymbol{\Gamma}_{t-1}^2 + \ldots) \\ &= (I \otimes I + \boldsymbol{\Gamma}_{t-1}' \otimes F + (\boldsymbol{\Gamma}_{t-1}')^2 \otimes F^2 + \ldots) A_2(:) \\ &= (I - \boldsymbol{\Gamma}_{t-1}' \otimes F)^{-1} A_2(:) \end{aligned}$$

The last equality uses the Kronecker product property that $(\mathbf{\Gamma}'_{t-1} \otimes F)(\mathbf{\Gamma}'_{t-1} \otimes F) = (\mathbf{\Gamma}'_{t-1})^2 \otimes F^2$. Finally, the actual law of motion is given by

$$S_t = \boldsymbol{C}\mathbf{\Delta}_{t-1} + \boldsymbol{A}S_{t-1} + \boldsymbol{B}\mathcal{E}_t \tag{C.56}$$

where $\boldsymbol{A} = \widetilde{A}_0^{-1}\widetilde{A}_2$; $\boldsymbol{B} = \widetilde{A}_0^{-1}B$; and $\boldsymbol{C} = \widetilde{A}_0^{-1}\widetilde{A}_1$.

CE equilibrium. The actual law of motion is derived as above after substituting $\boldsymbol{\delta}_{t-1}$ with $\mathbf{0}_{4\times 1}$ and $\boldsymbol{\gamma}_{t-1}$ with $\boldsymbol{\gamma}^*$, implying that the actual law of motion is given by $S_t = \boldsymbol{A}S_{t-1} + \boldsymbol{B}\mathcal{E}_t$. The variance-covariance matrix of S_t in vec form is given by $V_0(:) = (I - \boldsymbol{A} \otimes \boldsymbol{A})^{-1} (\boldsymbol{B} \otimes \boldsymbol{B})\, I(:)$, and the first-order autocovariance of S_t is given by $V_1 = \boldsymbol{A}V_0$. For a given set of parameters Θ, the first-order autocorrelation function of $S_{1:4,t}$ is described by

$$F(\boldsymbol{\gamma}, \Theta) = diag(diag(V_1(1:4, 1:4))/diag(V_0(1:4, 1:4))) \tag{C.57}$$

To find $\boldsymbol{\gamma}^*$ for a given set of parameters Θ, one solves the non-linear fixed-point equation $\boldsymbol{\gamma} = F(\boldsymbol{\gamma}, \Theta)$. Solving this equation asks for numerical methods and I follow the algorithm proposed in Hommes et al. (2022), so that for a given set of model parameters Θ,

1. Set initial $\boldsymbol{\gamma}_0$.
2. At each iteration $k \geq 2000$, set $\boldsymbol{\gamma}_k = F(\boldsymbol{\gamma}_{k-1}, \Theta)$.
3. Stop if the Euclidean norm $||\boldsymbol{\gamma}_k - \boldsymbol{\gamma}_{k-1}|| < \epsilon$, for $\epsilon = 0.00001$, or if $k = 2000$, otherwise go back to step 2.

C.3 Aggregate Demand and Supply under Well-specified Forecasting Rules

In this subsection, I derive the equilibrium conditions when $\widetilde{\mathbb{E}}_t^\star$ is associated with well-specified forecasting rules. Consider the aggregate demand

$$\hat{x}_t = n\left(1 - \beta + \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_t^\star \hat{x}_{t+1} + \widetilde{\mathbb{E}}_t^\star \sum_{h=0}^{\infty} (\beta n)^h \left[n^2\beta(1-\beta)\left(1 - \frac{\eta}{\Upsilon}\right)\hat{x}_{t+h+2} - \left(1 - \frac{\beta\eta}{\Upsilon}\right)(\hat{R}_{t+h} - \hat{\pi}_{t+h+1}) + \hat{e}_{t+h}\right] \tag{C.58}$$

Then,

$$\begin{aligned}\widetilde{\mathbb{E}}_t^\star \hat{x}_{t+1} &= n\left(1 - \beta + \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_t^\star \hat{x}_{t+2} \\ &+ \widetilde{\mathbb{E}}_t^\star \sum_{h=0}^{\infty} (\beta n)^h \left[n^2\beta(1-\beta)\left(1 - \frac{\eta}{\Upsilon}\right)\hat{x}_{t+h+3} - \left(1 - \frac{\beta\eta}{\Upsilon}\right)(\hat{R}_{t+h+1} - \hat{\pi}_{t+h+2}) + \hat{e}_{t+h+1}\right]\end{aligned} \tag{C.59}$$

from where

$$\widetilde{\mathbb{E}}_t^\star \sum_{h=0}^{\infty} (\beta n)^h \left[n^2\beta(1-\beta)\left(1 - \frac{\eta}{\Upsilon}\right)\hat{x}_{t+h+3} - \left(1 - \frac{\beta\eta}{\Upsilon}\right)(\hat{R}_{t+h+1} - \hat{\pi}_{t+h+2}) + \hat{e}_{t+h+1}\right] = \widetilde{\mathbb{E}}_t^\star \hat{x}_{t+1} - n\left(1 - \beta + \frac{\beta\eta}{\Upsilon}\right)\widetilde{\mathbb{E}}_t^\star \hat{x}_t \tag{C.60}$$

Substituting for the expression in the left-hand side in the equation above into the original aggregate demand in (C.58) and setting $\widetilde{\mathbb{E}}_t^\star \equiv \mathbb{E}_t$, we have

$$\hat{x}_t = n\mathbb{E}_t\left((1+\beta\eta)\hat{x}_{t+1} - \frac{n\beta\eta}{\Upsilon}\hat{x}_{t+2}\right) - \left(1-\frac{\beta\eta}{\Upsilon}\right)(\hat{R}_t - \mathbb{E}_t\hat{\pi}_{t+1}) + \hat{e}_t \tag{C.61}$$

If $n = 1$, then the equation above coincides with the standard Euler equation derived under FIRE. Similarly, consider the aggregate supply,

$$\begin{aligned}\hat{\tilde{\pi}}_t = \kappa\left(\omega\hat{y}_t + \frac{\Upsilon}{\Upsilon-\eta\beta}\hat{x}_t\right) + \widetilde{\mathbb{E}}_t^\star\sum_{h=0}^{\infty}(\alpha\beta n)^h\left(\kappa\alpha\beta n\left(\omega\hat{y}_{t+h+1} + \frac{(\alpha\Upsilon-\eta)}{\alpha(\Upsilon-\eta\beta)}\hat{x}_{t+h+1}\right)\right)\\ + \widetilde{\mathbb{E}}_t^\star\sum_{h=0}^{\infty}(\alpha\beta n)^h\left(n\beta(1-\alpha)\hat{\tilde{\pi}}_{t+h+1} + \hat{u}_{t+h}\right)\end{aligned} \tag{C.62}$$

Hence,

$$\begin{aligned}\widetilde{\mathbb{E}}_t^\star\hat{\tilde{\pi}}_{t+1} = \kappa\widetilde{\mathbb{E}}_t^\star\left(\omega\hat{y}_{t+1} + \frac{\Upsilon}{\Upsilon-\eta\beta}\hat{x}_{t+1}\right) + \widetilde{\mathbb{E}}_t^\star\sum_{h=0}^{\infty}(\alpha\beta n)^h\left(\kappa\alpha\beta n\left(\omega\hat{y}_{t+h+2} + \frac{(\alpha\Upsilon-\eta)}{\alpha(\Upsilon-\eta\beta)}\hat{x}_{t+h+2}\right)\right)\\ + \widetilde{\mathbb{E}}_t^\star\sum_{h=0}^{\infty}(\alpha\beta n)^h\left(n\beta(1-\alpha)\hat{\tilde{\pi}}_{t+h+2} + \hat{u}_{t+h+1}\right)\end{aligned} \tag{C.63}$$

Isolating $\kappa\omega\widetilde{\mathbb{E}}_t^\star\sum_{h=0}^{\infty}(\alpha\beta n)^h\hat{y}_{t+h+1}$ from (C.63), substituting for it into (C.62), and setting $\widetilde{\mathbb{E}}_t^\star \equiv \mathbb{E}_t$, we have

$$\hat{\tilde{\pi}}_t = \kappa\omega\hat{y}_t + \frac{\kappa}{1-\frac{\eta\beta}{\Upsilon}}\mathbb{E}_t\left(\hat{x}_t - \frac{\beta n\eta}{\Upsilon}\hat{x}_{t+1}\right) + n\beta\mathbb{E}_t\hat{\tilde{\pi}}_{t+1} + \hat{u}_t \tag{C.64}$$

If $n = 1$, then the equation above coincides with the standard Phillips curve derived under FIRE.

C.4 Data

I use quarterly data from 1966:Q2 to 2018:Q4. All data are extracted from FRED and described as follows

$$y_t = 100ln\left(\frac{GDPC1_t}{POP_{index,t}}\right)$$

$$\Delta y_t^{obs} = y_t - y_{t-1}$$

$$\pi_t^{obs} = 100ln\left(\frac{GDPDEF_t}{GDPDEF_{t-1}}\right)$$

$$R_t^{obs} = \frac{Funds_t}{4}$$

where

- $GDPC1$ – Real GDP, Billions of Chained 2012 Dollars, Seasonally Adjusted Annual Rate (see U.S. Bureau of Economic Analysis (1966 - 2018c)).

- $POP_{index} = \frac{CNP16OV}{CNP16OV_{1992Q3}}$.
- $CNP16OV$ – Civilian non-institutional population, thousands, 16 years and above (see U.S. Bureau of Economic Analysis (1966 - 2018b)).
- $GDPDEF$ – GDP-Implicit Price Deflator, 2012 = 100, Seasonally Adjusted (see U.S. Bureau of Economic Analysis (1966 - 2018a)).
- $Funds$ – Federal funds rate, daily figure averages in percentages (see Board of Governors of the Federal Reserve System (US) (1966 - 2018)).

The model is estimated using output growth data, however, within the model, households and firms need to form expectations about de-trended output. This implies that beliefs about de-trended output need to be initiated when there is SAC learning. Using pre-sample data, as explained in the main text, I compute de-trended output as follows. I compute the mean output growth in pre-sample data, that is, $\bar{\Upsilon}_{presample} = 0.70$, implying $\Upsilon_{presample} = \bar{\Upsilon}_{presample}/100 + 1 = 1.007$ and $y_t^{detrend} = y_t - 100ln\left(\Upsilon^t_{presample}\right)$. Last, I compute first and second moments of demeaned $y_t^{detrend}$ to initiate beliefs about de-trended output in the model. In Section C.6, I consider the case when all beliefs are initiated from the CE equilibrium and show that the results remain largely unchanged.

As commonly done in the literature, I fix the discount factor $\beta = 0.99$. Following Milani (2006), I also fix $\kappa = \frac{(1-\alpha)(1-\alpha\beta)}{\alpha(a_h+\zeta(1-a_h))}$ to 0.0015, which is the value estimated in Giannoni and Woodford (2004) for the flexible wages case. I note that since I estimate both ω and α, the posterior of κ will likely not be well-identified and it will be driven by its prior distribution or its starting values. To show that setting $\kappa = 0.0015$ is reasonable, I estimate κ along other parameters for the model with SAC learning for both the mean and first-order autocorrelation and find that its posterior mode is equal to 0.0015 (see column (4) in Table 1). In this exercise, the prior distribution for κ is gamma with mean 0.015 and standard deviation 0.011 as in Milani (2007).

C.5 Additional Details on Bayesian Estimation

As mentioned in the main text, I use the Metropolis-Hastings algorithm to generate two blocks with 1,200,000 draws each and discard the first 200,000 draws from the posterior distribution. In terms of the initiation of beliefs under SAC learning, I evaluate moments of the pre-sample data from 1960 to 1965 and use them as the initial learning parameters, $\boldsymbol{\delta}_0$ and/or $\boldsymbol{\gamma}_0$, for the Kalman filter procedure.

For specifications that involve a CE equilibrium, I follow a penalty strategy to ensure that the estimation is performed along the CE equilibrium. For each parameter draw, I compute the CE equilibrium, following the iterative e-stability algorithm of Hommes et al. (2022), which discards all unstable equilibria. For instance, fewer than 0.1% of draws cannot generate an equilibrium during the estimation procedure in the presence of myopia. This share is quite minimal especially when compared with the objective of the Metropolis-Hastings algorithm that about 75% of all draws should be discarded. If a stable equilibrium exists, then I compute the value of the likelihood function. If a stable equilibrium does not exist, I penalize the likelihood to be extremely low so that the draw is discarded. Posterior distributions are generally well-behaved. I rely on the method proposed by Brooks and Gelman (1998) to analyze convergence statistics, shown below.

Figure 5 and 6 plot the evolution of posterior draws as well as the posterior distribution when myopia is combined with AR(1) forecasting rules under SAC learning for both the mean and first-order autocorrelation. The posterior distribution is generally smooth for all estimated parameters. Additionally, Figure 7 shows the evolution of a crucial convergence statistic. Convergence is assessed based on the Brooks and Gelman (1998) methodology. I first estimate the evolution of the mean across draws for each parameter for each one of the two chains of the Metropolis-Hastings and compute the variance of the two means over time, $\hat{B}$. Then, I estimate the evolution of the draws variance for each parameter for each one of the two chains of the Metropolis-Hastings and define $\hat{W}$ to be the mean of the two computed variance values. Convergence is achieved when the evolution of $(\hat{W} + \hat{B})/\hat{W}$ converges to 1. Figure 7 shows that the convergence statistic remains below 1.1, which is the benchmark for convergence, and it approaches 1 as the number of draws increases.

C.6 Robustness Checks

C.6.1 Estimation

I consider several robustness checks for the estimated posterior mode. First, I re-estimate the preferred specification with AR(1) forecasts and myopia under SAC learning when the initial beliefs are initiated at the CE equilibrium as reported in column (4) of Table 3 in the main text. Second, I estimate a number of model specifications when the degree of myopia is assumed to have a uniform prior distribution with mean 0.5 and standard deviation $1/\sqrt{12}$. Third, I estimate κ along other parameters for the model with SAC learning for both the mean and first-order autocorrelation. The prior distribution for κ in this exercise is gamma with mean 0.015 and standard deviation 0.011 as in Milani (2007). Table 1 reports the posterior mode and the Laplace marginal likelihood. The estimated posterior mode and marginal likelihood for all three additional

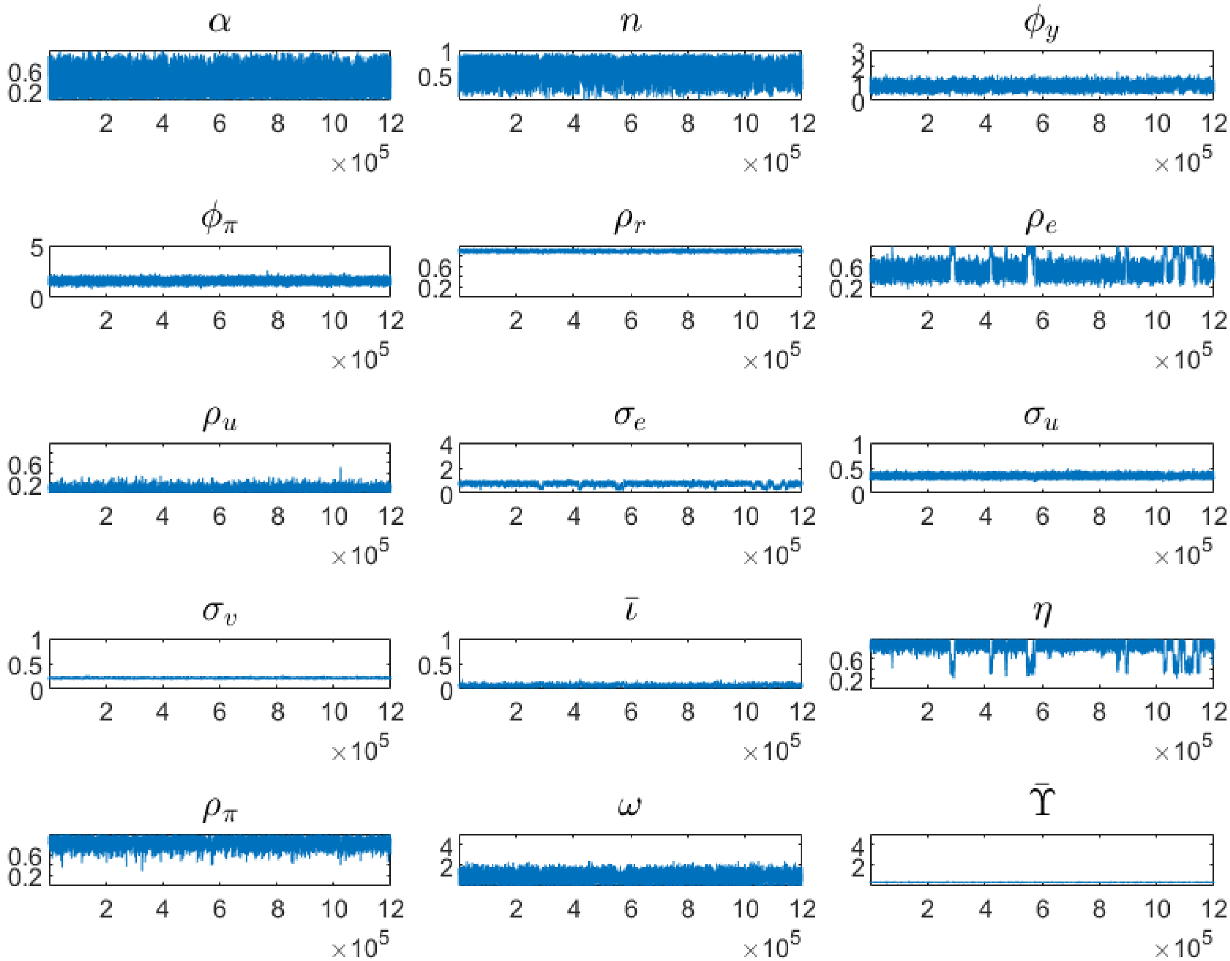


Figure 5: Evolution of posterior draws, when the expectations formation process is characterized by a combination of AR(1) misspecified forecasts and myopia under SAC learning for both the mean and first-order autocorrelation.

specifications remain pretty much unchanged relative to their respective benchmarks.

C.6.2 Generalized Impulse Response Functions

In this section, I present generalized IRFs (GIRFs) that are computed using the methodology proposed in Koop, Pesaran, and Potter (1996). Specifically, for a given shock, I draw $N = 1000$ innovations $\varepsilon \sim \mathcal{N}(0,1)$. For each innovation draw, I compute the response of output growth, output, inflation, and nominal interest rate for each model specification of interest. Clearly, for the models with SAC learning, beliefs evolve over time as well (the initial beliefs are set equal to the CE equilibrium values). Finally, I average across the N paths of responses for each model. Figure 8 exhibits the GIRFs for different model specifications. The implications are similar to the ones in Figure 7 in the main text.

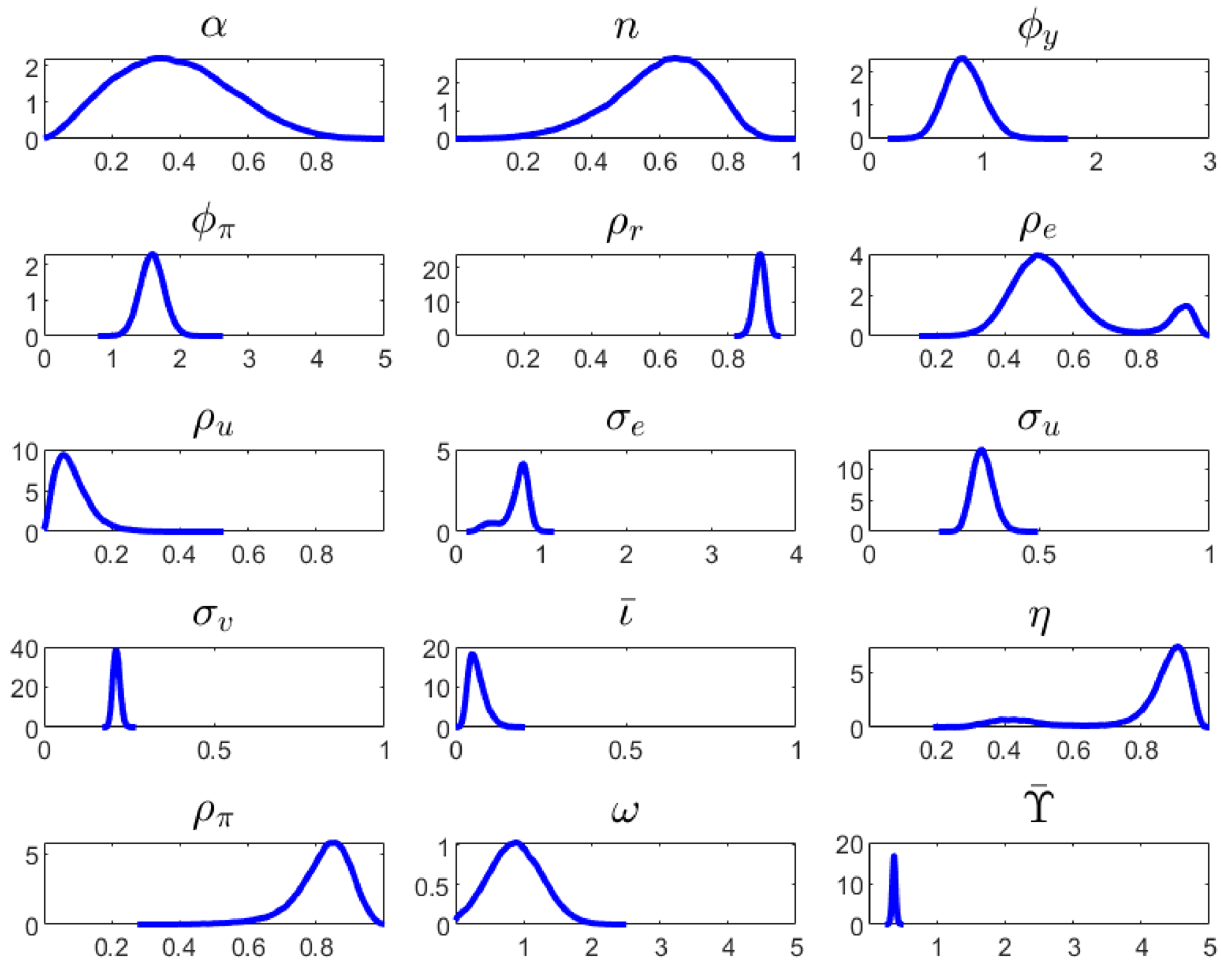


Figure 6: Posterior distribution of draws when the expectations formation process is characterized by a combination of AR(1) mis-specified forecasts and myopia under SAC learning for both the mean and first-order autocorrelation.

C.7 Other Model Specifications

In this section, I show estimation results for three additional model specifications. First, to investigate the model's performance when agents learn to use more sophisticated, yet mis-specified, forecasting rules, I estimate the model with VAR(1) forecasting rules with and without myopia under SAC learning (for both the mean and first-order autocorrelation) with constant gain parameter. Second, to understand whether imposing some discipline on the AR(1) mis-specified forecasting rules is important, I consider a specification where the mean of the AR(1) processes is set equal to the CE equilibrium, whereas the first-order autocorrelation coefficients are fixed over time and estimated along other parameters without imposing an equilibrium restriction on them. Third, I re-estimate the posterior mode of the model specifications with mis-specified and well-specified forecasting rules when the demand and supply shocks follow an ARMA process with moving average parameters μ_e and μ_u, respectively. The characteristics of the posterior distribu-

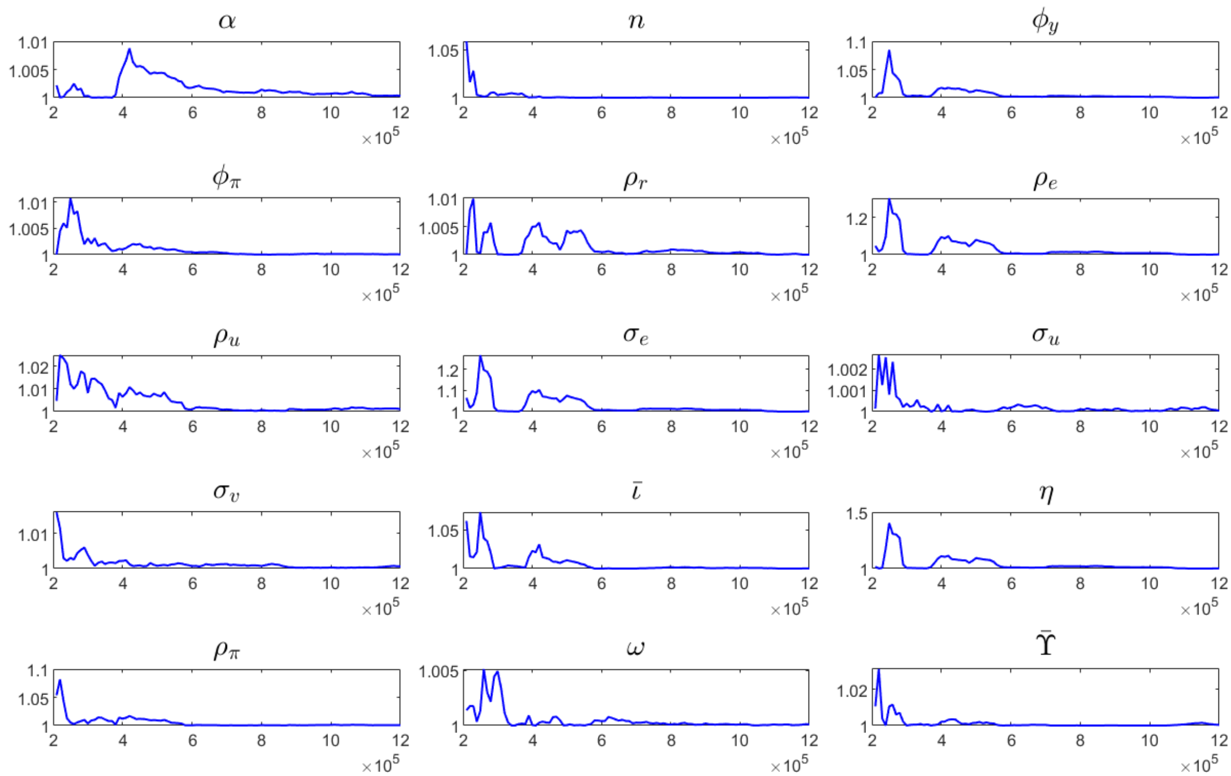


Figure 7: Evolution of the convergence statistic, $(\hat{B}+\hat{W})/\hat{W}$, when the expectations formation process is characterized by a combination of AR(1) mis-specified forecasts and myopia under SAC learning for both the mean and first-order autocorrelation.

tion of parameters for the first two specifications are exhibited in Table 2, whereas the posterior mode for the third specification is reported in Table 3.

Starting with Table 2, to judge model fit, I set the expectations formation process with well-specified forecasts and myopia, reported in column (2) of Table 3, to be the benchmark specification and compare the other two models relative to that benchmark. The values in parenthesis in the last row of Table 2 report the Bayes factor for the model specification relative to the benchmark. The model with VAR(1) forecasting rules and myopia performs very similarly to models with mis-specified forecasts under SAC learning reported in Table 4 in the main text. That is, in the presence of myopia VAR(1) forecasts perform better then the benchmark, but absent myopia they perform worse than the benchmark. Figure 9 plots the evolution of the estimated agents' beliefs when they engage in constant-gain learning of a VAR(1) forecasting process in the specification with myopia. The perceived first-order correlation between any two distinct aggregate variables seems to fluctuate around 0. Therefore, using more elaborate forecasting rules, such as VAR(1), will not add, on average, any significant information to households and firms in terms of forecasting, and it will not strongly enhance the model's fit of the data. I discuss the performance of VAR(1) rules for inflation forecasts in Section C.8.

	SAC: mean & autocorr.				SAC: mean		CEE	
	bench.	**beliefs**	$\mathcal{U}$	κ	**bench.**	$\mathcal{U}$	**bench.**	$\mathcal{U}$
Parameters	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Calvo parameter, α	0.44	0.46	0.45	0.44	0.40	0.42	0.52	0.51
Func. of price stick., κ	-	-	-	0.0015	-	-	-	-
Degree of myopia, n	0.56	0.56	0.60	0.56	0.71	0.75	0.68	0.73
Habit in consumption, η	0.89	0.89	0.89	0.89	0.88	0.89	0.89	0.88
Inflation indexation, ρ_π	0.87	0.86	0.87	0.87	0.76	0.76	0.86	0.85
Elasticity mc, ω	0.88	0.87	0.88	0.88	0.87	0.88	0.82	0.81
Deterministic growth, $\bar{\Upsilon}$	0.37	0.37	0.37	0.37	0.36	0.36	0.37	0.36
Feedback to growth, ϕ_y	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82
Feedback to inflation, ϕ_π	1.59	1.59	1.59	1.59	1.60	1.61	1.59	1.59
Interest rate smooth., ρ_r	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
Demand autocorr., ρ_e	0.47	0.44	0.48	0.47	0.37	0.36	0.36	0.35
Supply autocorr., ρ_u	0.05	0.05	0.05	0.05	0.04	0.05	0.04	0.04
Demand std., σ_e	0.80	0.82	0.80	0.80	0.86	0.86	0.86	0.87
Supply std., σ_u	0.33	0.33	0.33	0.33	0.34	0.34	0.33	0.34
Monetary std., σ_v	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
Learning gain, $\bar{\iota}$	0.04	0.04	0.05	0.04	0.05	0.05	-	-
Log marg. data dens.								
Laplace	-266.33	-267.53	-267.32	-269.06	-265.4	-265.85	-268.73	-268.46
Bayes factor	$(e^{0.00})$	$(e^{-1.2})$	$(e^{-0.99})$	$(e^{-2.73})$	$(e^{0.00})$	$(e^{-0.45})$	$(e^{0.00})$	$(e^{0.27})$

Table 1: Robustness checks at the posterior mode for various assumptions on the expectations formation process with myopia. The prior for myopia is uniform with mean 0.5 and standard deviation $1/\sqrt{12}$. Columns (1), (5), and (7) report the posterior mode of the same model specifications reported in columns (2) and (4) in Table 4 and column (4) in Table 3 in the main text, respectively. I refer to these model specifications as benchmarks. Column (2) sets initial beliefs equal to the CE equilibrium reported in column (4) of Table 3 in the main text. Column (4) estimates κ along other parameters; the prior is a gamma distribution with mean 0.015 and standard deviation 0.011. Columns (3), (6), and (8) report the posterior mode when the prior distribution for the degree of myopia is assumed to be uniform with mean 0.5 and standard deviation $1/\sqrt{12}$. Values in parentheses denote the Bayes factor of the model relative to the benchmark specification.

On the other hand, the model with unrestricted AR(1) forecasts and myopia or absent it fits macroeconomic data similarly to the benchmark, hence worse than the models with mis-specified forecasts under SAC learning and myopia. However, it is interesting that the presence of myopia does not improve model fit. Moreover, there are some stark differences in terms of estimation. For instance, the perceived first-order autocorrelation of inflation is significantly smaller than its equilibrium counterpart reported in columns (3) and (4) of Tables 3 and 4 in the main text. The estimated degree of myopia is smaller than what is found in the other model specifications. I discuss the performance of such unrestricted AR(1) rules for inflation forecasts in Section C.8.

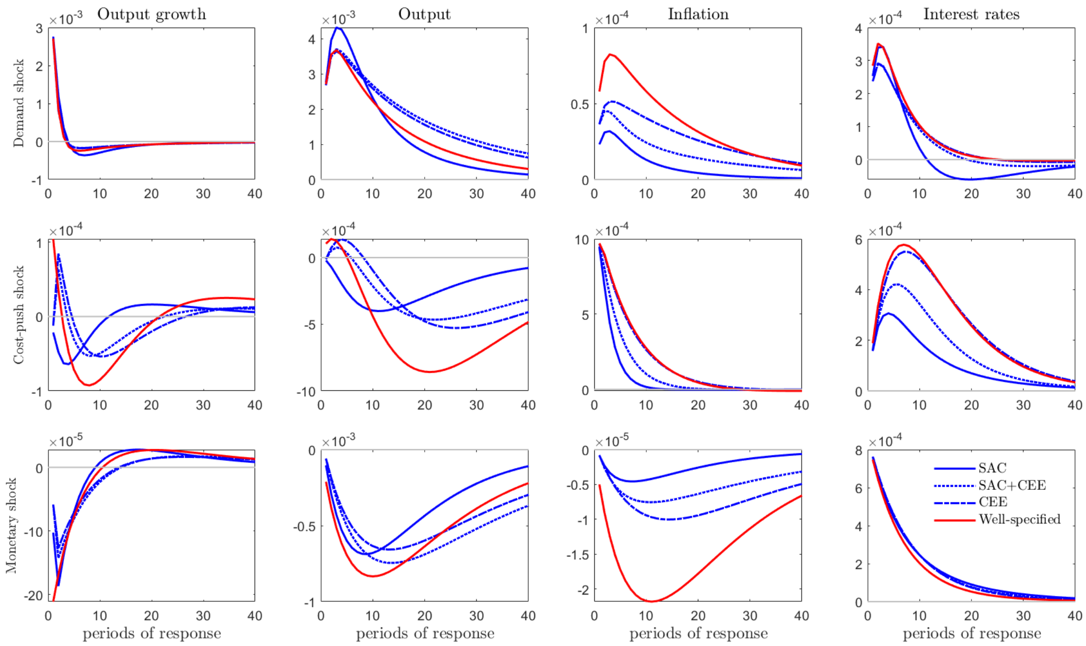


Figure 8: Average generalized impulse response functions to a one standard deviation positive demand, cost-push and monetary shock. SAC denotes the specification with SAC learning for both the mean and first-order autocorrelation (solid blue); SAC + CEE denotes the specification with SAC learning for the mean and with the first-order autocorrelation set at the CE equilibrium (dashed blue); CEE denotes the specification where both the mean and first-order autocorrelation are set at their CE equilibrium (dot dashed blue); Well-specified denotes the model with well-specified forecasts (solid red). Parameters are set at their estimated posterior mean as shown in Tables 3 and 4 in the main text.

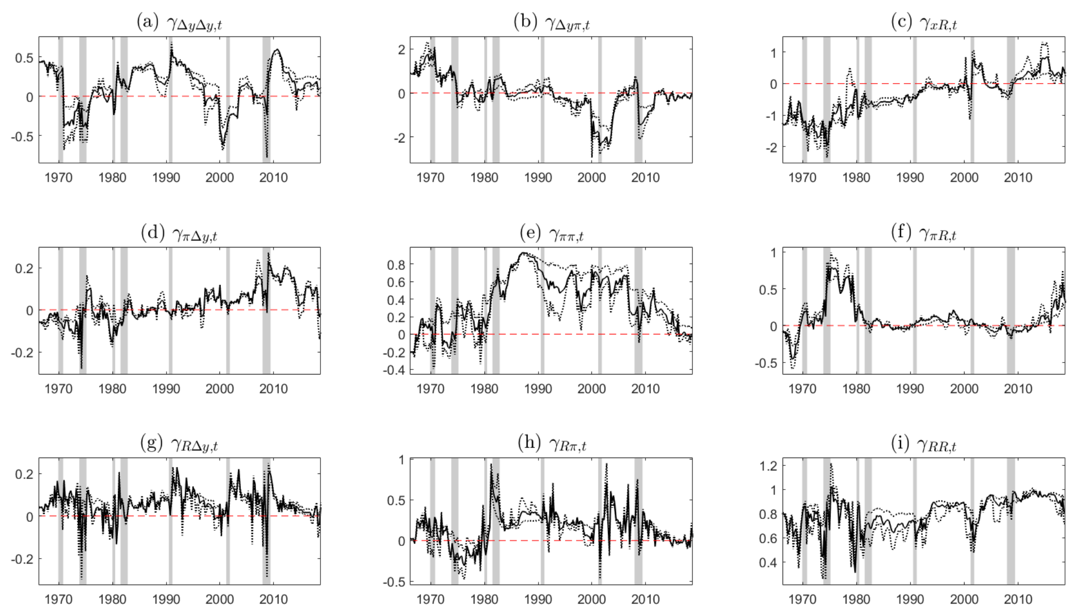


Figure 9: Evolution of the VAR(1) forecast coefficients in the model with SAC learning and myopia. The black and dotted curves plot implied beliefs for structural parameters set at their estimated posterior mean and 90 percent highest posterior density, respectively. Gray areas indicate recessionary periods as reported by the National Bureau of Economic Research. The dashed red lines indicate the x axis.

	VAR(1) Rules, SAC learning						Unrestricted AR(1) Rules					
	(1)			(2)			(3)			(4)		
	no myopia, $n = 1$			**myopia, $n \in (0, 1)$**			**no myopia, $n = 1$**			**myopia, $n \in (0, 1)$**		
Parameters	mean	5%	95%	mean	5%	95%	mean	5%	95%	mean	5%	95%
Calvo parameter, α	0.34	0.11	0.64	0.28	0.09	0.52	0.64	0.28	0.9	0.69	0.35	0.92
Degree of myopia, n	-	-	-	0.65	0.44	0.82	-	-	-	0.39	0.1	0.78
Habit in consumption, η	0.97	0.94	0.98	0.43	0.27	0.87	0.76	0.64	0.88	0.42	0.29	0.57
Inflation indexation, ρ_π	0.57	0.17	0.86	0.77	0.54	0.92	0.06	0.01	0.12	0.91	0.84	0.97
Elasticity mc, ω	0.88	0.28	1.54	0.91	0.28	1.54	0.87	0.26	1.5	0.87	0.24	1.51
Deterministic growth, Υ	0.38	0.29	0.46	0.4	0.36	0.44	0.35	0.31	0.38	0.38	0.34	0.42
Feedback to output growth, ϕ_y	0.82	0.57	1.09	0.98	0.71	1.24	0.85	0.59	1.13	0.98	0.73	1.25
Feedback to inflation, ϕ_π	1.61	1.34	1.9	1.61	1.33	1.88	1.59	1.31	1.88	1.6	1.33	1.88
Interest rate smoothing, ρ_r	0.89	0.87	0.92	0.88	0.86	0.91	0.89	0.87	0.92	0.89	0.86	0.91
Demand shock autocorr., ρ_e	0.88	0.8	0.95	0.9	0.65	0.97	0.38	0.25	0.52	0.94	0.89	0.98
Supply shock autocorr., ρ_u	0.21	0.06	0.46	0.12	0.03	0.27	0.86	0.81	0.91	0.06	0.02	0.13
Demand shock std., σ_e	0.19	0.09	0.32	0.38	0.22	0.66	0.74	0.55	0.94	0.54	0.24	0.74
Supply shock std., σ_u	0.33	0.22	0.42	0.34	0.29	0.39	0.13	0.06	0.22	0.29	0.26	0.32
Monetary shock std., σ_v	0.21	0.2	0.23	0.21	0.2	0.23	0.21	0.2	0.23	0.21	0.2	0.23
Learning gain, ι	0.12	0.09	0.15	0.09	0.06	0.12	-	-	-	-	-	-
$\gamma_{\Delta y}$	-	-	-	-	-	-	0.5	0.18	0.83	0.5	0.17	0.83
γ_π	-	-	-	-	-	-	0.2	0.06	0.37	0.4	0.14	0.67
γ_R	-	-	-	-	-	-	0.5	0.19	0.75	0.47	0.16	0.79
γ_y	-	-	-	-	-	-	0.96	0.93	0.99	0.46	0.15	0.8
Log marg. data dens.												
Modified Harmonic Mean	-277.45			-263.04			-269.27			-271.86		
Bayes factor	$(e^{-7.21})$			$(e^{7.20})$			$(e^{0.97})$			$(e^{-1.62})$		

Table 2: Posterior distribution of the model with VAR(1) mis-specified forecasting rules under SAC learning and AR(1) backward-looking rules with estimated parameters. Values in parentheses denote the Bayes factor of the model relative to the model in column (2) of Table 3 in the main text. The first-order autocorrelation coefficients for the specification with backward-looking AR(1) forecasts are assumed to have a beta prior distribution with mean 0.5 and standard deviation 0.2, whereas the mean is set to 0.

Finally, Table 3 shows that the models with well-specified forecasts make use of the additional persistence that the ARMA(1,1) processes for the shocks induce. Specifically, the estimate of μ_e at the posterior mode is about 2 to 3 times larger for the models with well-specified forecasts compared to the ones with mis-specified rules. Furthermore, the estimate of μ_u for the model with well-specified forecasts and myopia is twice as large as the model with mis-specified rules and myopia. Last, the model specifications with myopia continue to outperform the ones absent of it and the model with mis-specified rules and myopia fits data better than the model with well-specified rules without myopia.

	Well-specified rules		**AR(1), SAC: mean & autocorr.**	
	(1)	(2)	(3)	(4)
	no myopia	**myopia**	**no myopia**	**myopia**
Parameters				
Calvo parameter, α	-	-	0.88	0.40
Degree of myopia, n	-	0.74	-	0.59
Habit in consumption, η	0.99	0.94	0.98	0.89
Inflation indexation, ρ_π	0.92	0.03	0.92	0.85
Elasticity mc, ω	0.42	0.76	0.67	0.87
Deterministic growth, $\bar{\Upsilon}$	0.40	0.33	0.39	0.37
Feedback to output growth, ϕ_y	0.80	0.92	0.79	0.82
Feedback to inflation, ϕ_π	1.68	1.65	1.59	1.59
Interest rate smoothing, ρ_r	0.86	0.88	0.89	0.89
Demand shock autocorr., ρ_e	0.51	0.34	0.43	0.43
Supply shock autocorr., ρ_u	0.04	0.86	0.04	0.05
Demand shock std., σ_e	0.16	0.37	0.81	0.80
Supply shock std., σ_u	0.27	0.09	0.29	0.34
Monetary shock std., σ_v	0.21	0.21	0.21	0.21
Demand shock moving average, μ_e	0.26	0.15	0.08	0.07
Supply shock moving average, μ_u	0.04	0.10	0.03	0.04
Learning gain, $\bar{\iota}$	-	-	0.01	0.05
Log marg. data dens.				
Laplace	-294.30	-275.78	-301.61	-277.45
Bayes factor	($e^{-18.52}$)	($e^{0.00}$)	($e^{-25.83}$)	($e^{-1.67}$)

Table 3: Posterior mode when the demand and supply shocks follow an ARMA(1,1) process. Values in parentheses denote the Bayes factor of the model relative to the specification in column (2).

C.8 Additional Results on Forecast Behavior

This section repeats the analysis of Section 5 in the main text for various model specifications.

External validation. Figure 10 plots the mean-squared distance between model-implied annual forecast about inflation and a series of actual annual inflation forecasts. The top two panels plot the mean-squared distance of annual forecasts implied by models with myopia, whereas the bottom two panels plot the mean-squared distance of annual forecasts implied by models without myopia. The figure shows that in terms of matching both SPF and MSC annual inflation forecasts, mis-specified forecasts generally outperform well-specified forecasting rules, regardless of whether there is myopia or not. The model specification with unrestricted AR(1) rules performs much poorer than the other specifications in terms of matching SPF and MSC forecasting data: the estimates of K_4 and M_4 are both estimated to be positive and can be orders of magnitude larger than what is

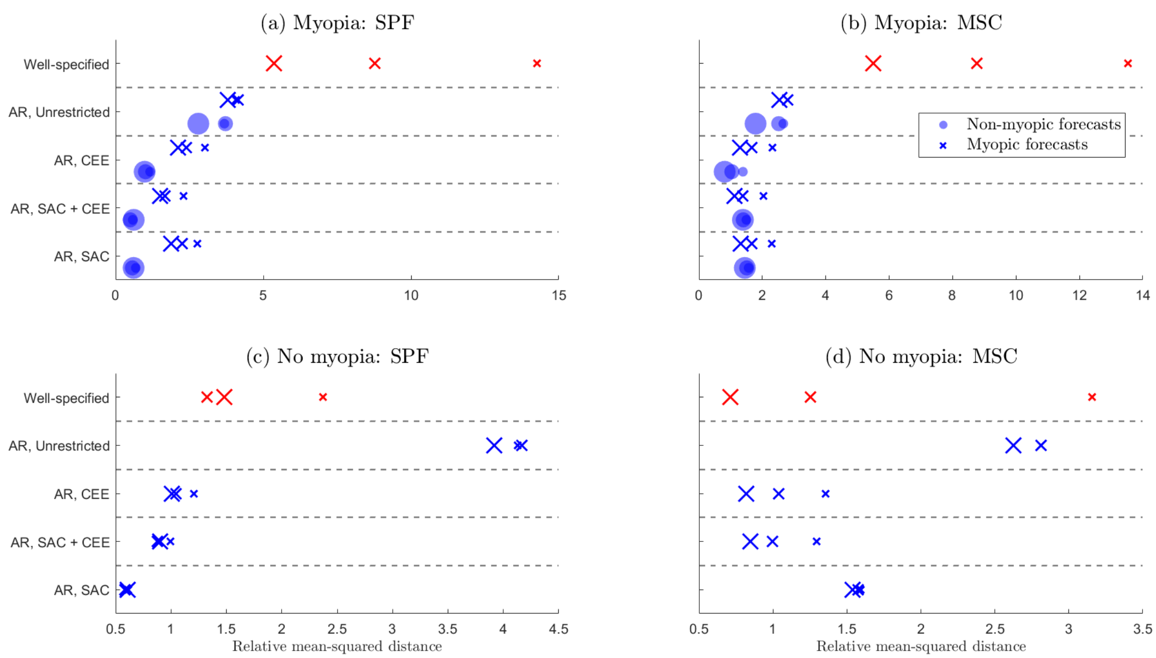
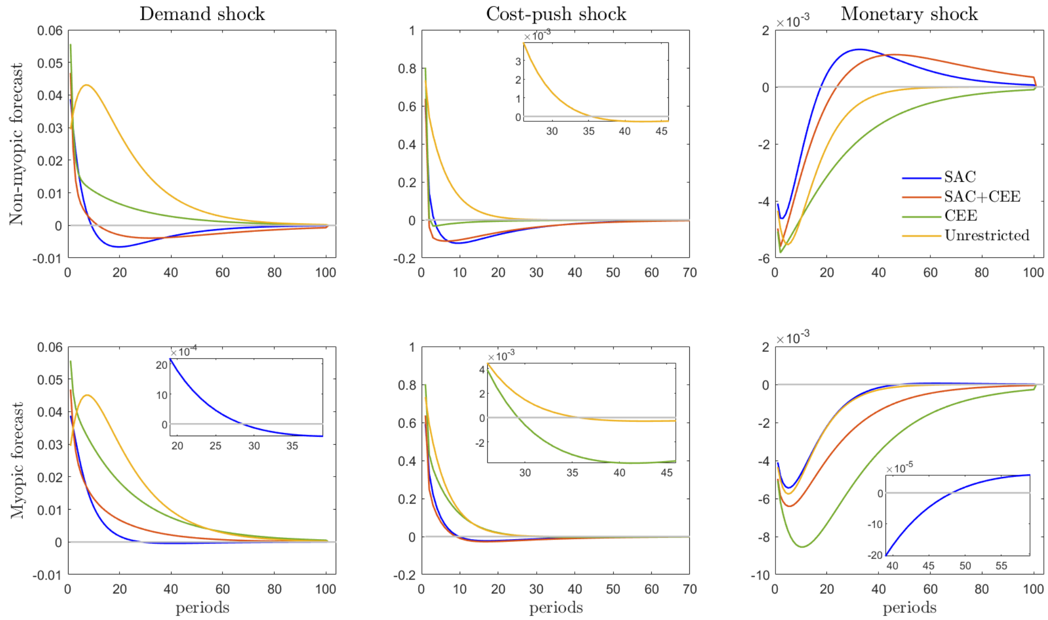


Figure 10: Mean-squared distance between model-implied and survey-based inflation forecasts. The top two panels plot the mean-squared distance of annual forecasts implied by of mis-specified forecasts and myopia specification. The bottom two panels plot the mean-squared distance of annual forecasts implied by the model with mis-specified forecasts and no myopia. The models denoted SAC, SAC + CEE, and CEE are the same as explained in Figure 8. Unrestricted is the specification where the mean is set at the CE equilibrium and the first-order autocorrelation is estimated along other parameters in the model; Well-specified is the model with well-specified forecasts. Circles denote non-myopic forecasts; crosses denote myopic forecasts. Model parameters are set at the posterior 5^{th} percentile (smallest circle/cross mark), mean (medium-sized circle/cross mark), and the 95^{th} percentile (largest circle/cross mark) of the estimated posterior distribution as documented in columns (3)-(4) of Table 3 in the main text, columns (1)-(4) in Table 4 in the main text, and columns (3)-(4) in Table 1 here.

Delayed over-shooting. Figure 11 plots average pseudo and generalized impulse response functions of annual inflation forecasting errors to demand, cost-push, and monetary shocks. I

(a) *Average pseudo IRFs*



(b) *Average generalized IRFs*

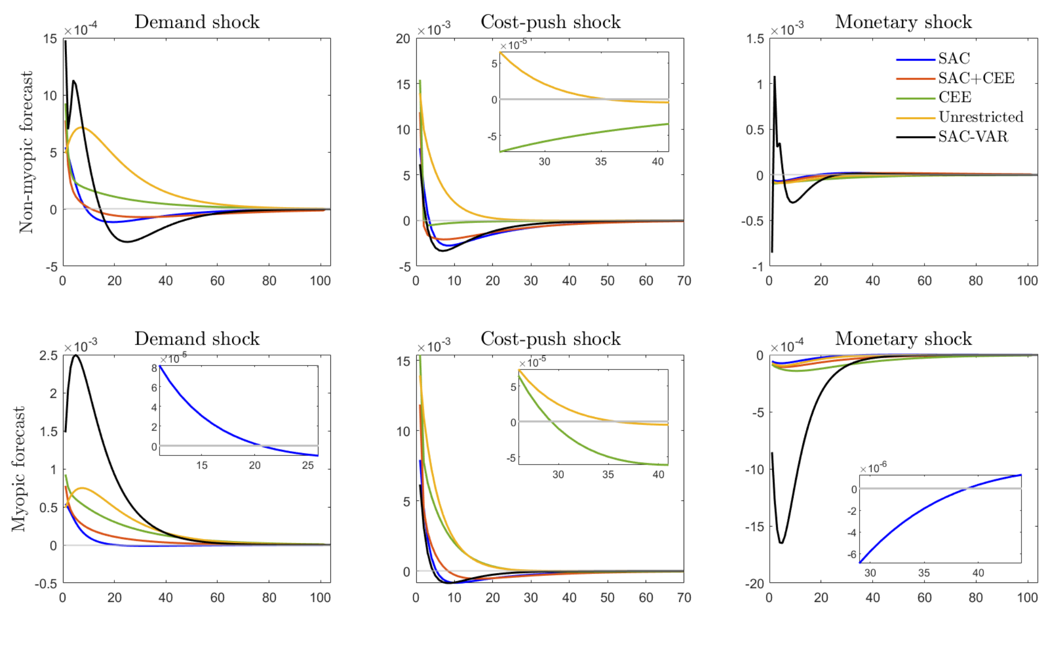


Figure 11: Average pseudo and generalized impulse response functions of annual inflation forecasting errors to a one standard deviation positive demand, cost-push, and monetary shock in models with mis-specified forecasts and myopia. I consider the same four models with mis-specified forecasts as in Figure 10; in addition, panel (b) plots the generalized impulse responses of forecast errors implied by the model with VAR(1) forecast rules and myopia. Model parameters are set equal to the estimated posterior mean as shown in Tables 3 and 4 in the main text, and Tables 1 and 2 here.

consider four model specifications with AR(1) mis-specified forecasts and myopia, and for each model specifications I consider myopic and non-myopic forecasts. The top three panels exhibit the response of non-myopic forecast errors, and the bottom three panels show the responses in the case of myopic forecasts. As Figure 11 shows, delayed over-shooting of forecast errors is quite robust across all specifications in the case of cost-push shocks, but some learning of the first-order autocorrelation coefficient is necessary to grant late over-shooting in the case of demand and monetary shocks. Importantly, as panel (b) shows, myopic VAR(1) forecasts are not always able to generate delayed over-shooting to shocks.

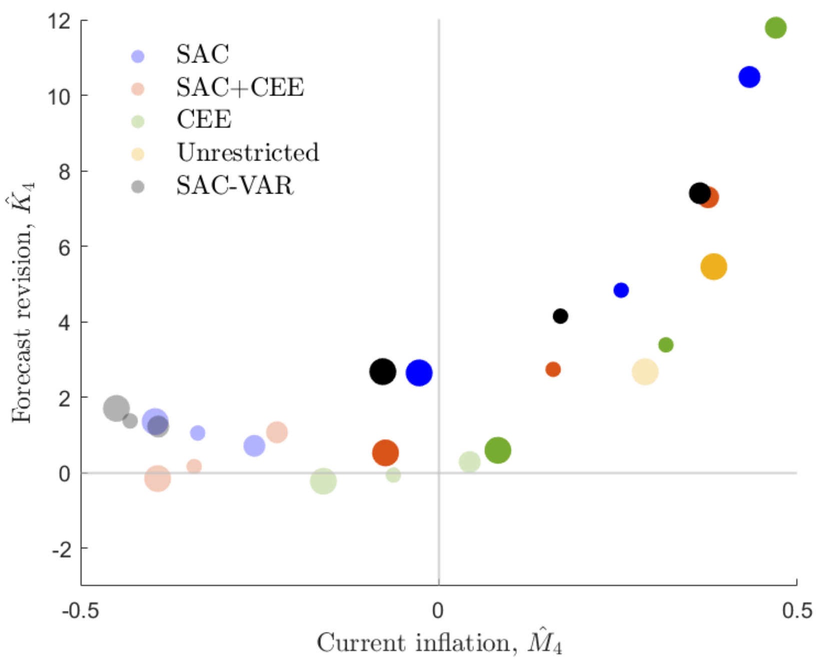


Figure 12: Estimates of regressions in (14) and (15) in the main text on model-based forecasting data for the same five models with mis-specified forecasts as in Figure 11. Model parameters are set at the posterior 5^{th} percentile (smallest circle), mean (medium-sized circle), and the 95^{th} percentile (largest circle) of the estimated posterior distribution as documented in column (4) of Table 3 in the main text, column (2) and (4) in Table 4 in the main text, column (4) in Table 1 here, and column (2) in Table 2 here. The transparent circles visualize the estimated $\hat{K}_4$ and $\hat{M}_4$ when forecasts are non-myopic. All regressions include a constant term.

Under-reaction to ex-ante forecast revisions, over-reaction to current inflation. I estimate the regressions in (14) and (15) with inflation forecasting data implied by the model with AR(1) mis-specified forecasting rules and myopia, as well as VAR(1) forecasts and myopia, over the 1968:Q3 through 2018:Q4 period. For each model specification, I consider myopic and non-myopic forecasts. Figure 12 summarizes the estimates of coefficients K_4 for the full sample and M_4 for the sample starting from 1982:Q3 through 2018:Q4. All specifications can generally produce under-reaction to ex-ante forecasting revisions, but only the models with some learning can generate over-reaction to current inflation simultaneously with under-reaction to forecast revisions.

Moreover, non-myopic forecasts outperform myopic forecasts in terms of matching over-reaction to current inflation. Finally, the AR(1) specification with unrestricted coefficients always yields positive values for K_4 and M_4 that can be orders of magnitude higher than the SPF evidence.

D Proofs

D.1 Proposition 1

Let the data-generating process for inflation be given by $\hat{\pi}_t = a\hat{mc}_t + b\hat{\pi}_{t-1}$, where $a = \frac{\kappa}{1-\alpha\beta\rho n}$ and $b = \frac{\beta n(1-\alpha)}{1-\alpha\beta n\gamma^*}(\gamma^*)^2$. Then, one can show that

$$F(\gamma) = \frac{\mathbb{E}(\hat{\pi}_t\hat{\pi}_{t-1})}{\mathbb{E}(\hat{\pi}_t^2)} = \frac{b+\rho}{1+\rho b} \tag{D.1}$$

For a CE equilibrium to exist, we must have that $F(\gamma) = \gamma$ for at least one value of $\gamma \in (0,1)$. Moreover, $F(\gamma)$ is an increasing function of γ, with $F(0) = \rho > 0$ and $F(1) = \frac{\beta n(1-\alpha)+\rho(1-\alpha\beta n)}{1-\alpha\beta n+\rho\beta n(1-\alpha)}$, where $\rho \leq F(1) < 1$. Therefore, $F(\gamma)$ crosses the 45° line at least once; that is, a CE equilibrium is guaranteed to exist. Since $F(\gamma) \geq \rho$, it follows that $\gamma^* \in [\rho, 1)$.

To show that the CE equilibrium is unique, I show that $F(\gamma)$ is convex whenever it intersects with the 45° line, i.e., whenever (D.1) holds. Note that $F(\gamma)$ is an increasing function of γ, such that $F(0) = \rho$ and $F(1) < 1$. Therefore, if multiple fixed points existed for $\gamma \in [0,1)$, it must be that at least one fixed point, $F(\gamma)$, is concave.

$$F''(\gamma)\mid_{\gamma=\gamma^*} = (1-\rho\gamma^*)\frac{b''(1+\rho b) - \rho(b')^2}{(1+\rho b)^2} \tag{D.2}$$

where $b' = \partial b/\partial\gamma$ and b'' denotes the second-order partial derivative of b w.r.t. γ. Therefore, $F''(\gamma = \gamma^*) > 0 \iff b'' > \frac{\rho(b')^2}{1+\rho b}$. One can show that

$$b'' = \frac{2\beta n(1-\alpha)}{(1-\alpha\beta n\gamma^*)^3} \tag{D.3}$$

Then,

$$\begin{aligned} b'' - \frac{\rho(b')^2}{1+\rho b} &= \frac{2\beta n(1-\alpha)}{(1-\alpha\beta n\gamma^*)^3} - \frac{\rho(\beta n\gamma^*(1-\alpha))^2(2-\alpha\beta n\gamma^*)^2}{(1-\alpha\beta n\gamma^*)^3(1-\alpha\beta n\gamma^*+\beta n\rho(1-\alpha)(\gamma^*)^2)} \\ &= \underbrace{\frac{\beta n\gamma^*(1-\alpha)}{(1-\alpha\beta n\gamma^*)^3(1-\alpha\beta n\gamma^*+\beta n\rho(1-\alpha)(\gamma^*)^2)}}_{(+)} \\ &\times \underbrace{\left(2(1-\alpha\beta n\gamma^*) + 2(\beta n\rho(1-\alpha)(\gamma^*)^2) - \rho\beta n(\gamma^*)^2(1-\alpha)(2-\alpha\beta n\gamma^*)^2\right)}_{G(\gamma)} \end{aligned} \tag{D.4}$$

Hence, the sign of $F''(\gamma=\gamma^*)$ is determined by the sign of $G(\gamma)$, which is always positive.

$$\begin{aligned} G(\gamma) &= 2(1-\alpha\beta n\gamma^*) + \beta n\rho(1-\alpha)(\gamma^*)^2(2-4+4\alpha\beta n\gamma^* - (\alpha\beta n\gamma^*)) \\ &= 2(1-\alpha\beta n\gamma^*)(1-\beta n\rho(1-\alpha)(\gamma^*)^2) + \alpha\beta^2 n^2\rho(1-\alpha)(\gamma^*)^3)(2-\alpha\beta n\gamma^*) \geq 0 \end{aligned} \tag{D.5}$$

D.2 Corollary 1

Consider $F(\gamma)$, with $F(\gamma)$ as defined in (D.1). Since the CE equilibrium is unique, γ^*, following a change in price stickiness or myopia, will change in the same direction as $F(\gamma)$. Taking the first-order partial derivative with respect to α of $F(\gamma)$ yields

$$\frac{\partial F(\gamma)}{\partial\alpha} = \frac{1-\rho^2}{(1+\rho b)^2}\underbrace{\frac{\partial b}{\partial\alpha}}_{(-)} < 0 \tag{D.6}$$

Similarly, taking the first-order partial derivative with respect to n of $F(\gamma)$ yields

$$\frac{\partial F(\gamma)}{\partial n} = \frac{1-\rho^2}{(1+\rho b)^2}\underbrace{\frac{\partial b}{\partial n}}_{(+)} > 0 \tag{D.7}$$

D.3 Proposition 2

The actual law of motion for inflation along the CE equilibrium is $\hat{\pi}_t = a\hat{mc}_t + b\hat{\pi}_{t-1}$, and the forecast about next period's inflation along the equilibrium path is $\widetilde{\mathbb{E}}_t\hat{\pi}_{t+1} = n(\gamma^*)^2\hat{\pi}_{t-1}$. Hence, the h-period-ahead forecasting error about inflation in period $(t+k)$, following a one-time shock ε_t in period t, is

$$\begin{aligned} \hat{\pi}_{t+k} - \widetilde{\mathbb{E}}_{t+k-h}\hat{\pi}_{t+k} &= a\hat{mc}_{t+k} + b\hat{\pi}_{t+k-1} - n^h(\gamma^*)^{h+1}\hat{\pi}_{t+k-h-1} \\ &= a\rho^k\varepsilon_t + ab(\rho^{k-1} + b\rho^{k-2} + \ldots + b^{k-1})\varepsilon_t - an^h(\gamma^*)^{h+1}(\rho^{k-h-1} + \ldots + b^{k-h-1})\varepsilon_t \\ &= a\rho^{k-h-1}\left(\rho^{h+1} + b\rho^h\left(1+\ldots+\left(\frac{b}{\rho}\right)^h + \ldots + \left(\frac{b}{\rho}\right)^{k-1}\right)\right) \\ &\quad - a\rho^{k-h-1}\left(n^h(\gamma^*)^{h+1}\left(1+\ldots+\left(\frac{b}{\rho}\right)^{k-h-1}\right)\right)\varepsilon_t \\ &= a\rho^{k-h-1}\left(\left(b^{h+1} - n^h(\gamma^*)^{h+1}\right)\sum_{j=0}^{k-h-1}\left(\frac{b}{\rho}\right)^j + \rho\left(\rho^h + b\rho\frac{\rho^h - b^h}{\rho - b}\right)\right)\varepsilon_t \end{aligned} \tag{D.8}$$

The effect of $\varepsilon_t > 0$ on the forecasting error for $k=0$ is positive; hence, forecasters under-react on impact. Moreover, $\lim_{k\to\infty}\rho^{k-h-1} = 0$, and therefore the forecasting error will eventually dissipate at some point in the future. The question remains whether, as $k\to\infty$, we approach the

0 forecasting errors from below (delayed over-shooting) or above. Given that $a > 0$ and $\rho > 0$, delayed over-shooting is guaranteed to occur if

$$\lim_{k\to\infty}\left(\left(b^{h+1}-n^h(\gamma^*)^{h+1}\right)\sum_{j=0}^{k-h-1}\left(\frac{b}{\rho}\right)^j+\rho\left(\rho^h+b\rho\frac{\rho^h-b^h}{\rho-b}\right)\right)<0 \tag{D.9}$$

One can easily show that $(b^{h+1} - n^h(\gamma^*)^{h+1}) < 0$. Then, if $b > \rho$, we have that

$$\lim_{k\to\infty}\left(b^{h+1}-n^h(\gamma^*)^{h+1}\right)\sum_{j=0}^{k-h-1}\left(\frac{b}{\rho}\right)^j=-\infty$$

so

$$\lim_{k\to\infty}\left(\left(b^{h+1}-n^h(\gamma^*)^{h+1}\right)\sum_{j=0}^{k-h-1}\left(\frac{b}{\rho}\right)^j+\rho\left(\rho^h+b\rho\frac{\rho^h-b^h}{\rho-b}\right)\right)=-\infty \tag{D.10}$$

On the other hand, if $b < \rho$, we have that $\lim_{k\to\infty}\left(b^{h+1}-n^h(\gamma^*)^{h+1}\right)\sum_{j=0}^{k-h-1}\left(\frac{b}{\rho}\right)^j=\frac{\rho(b^{h+1}-n^h(\gamma^*)^{h+1})}{\rho-b}$, so

$$\lim_{k\to\infty}\left(\left(b^{h+1}-n^h(\gamma^*)^{h+1}\right)\sum_{j=0}^{k-h-1}\left(\frac{b}{\rho}\right)^j+\rho\left(\rho^h+b\rho\frac{\rho^h-b^h}{\rho-b}\right)\right)=\frac{\rho(\rho^{h+1}-n^h(\gamma^*)^{h+1})}{\rho-b} \tag{D.11}$$

Hence, when $b < \rho$, delayed over-shooting is guaranteed to exist if $\rho^{h+1} < n^h(\gamma^*)^{h+1}$. Mis-specified, to show that the two conditions stated above are sufficient for late over-response, we have to show that if the forecast error response turns negative, it will never become positive. Showing this proves that if the forecast error impulse response approaches 0 from above in the limit as $k \to \infty$, it has never been negative before. Suppose there exists $k^* \geq 1$, such that for $k \geq k^*$,

$$\mathbb{I}_{k,h}=\frac{\partial(\hat{\pi}_{t+k}-\widetilde{\mathbb{E}}_{t+k-h}\hat{\pi}_{t+k})}{\partial\varepsilon_t}=a\rho^{k-2}\left(\left(b^2-n(\gamma^*)^2\right)\sum_{j=0}^{k-h-1}\left(\frac{b}{\rho}\right)^j+\rho(b+\rho)\right)<0 \tag{D.12}$$

Since $(b^{h+1} - n^h(\gamma^*)^{h+1}) < 0$, as k increases the impulse response of forecast errors becomes more negative, and the sign of $\mathbb{I}_{k,h}$ can never flip as k increases.

D.4 Proposition 3

I first derive a number of important moments. Consider first the covariance between $\hat{\pi}_{t+h}$ and $\hat{\pi}_t$ for any $h > 0$:

$$Cov(h)=\mathbb{E}(\hat{\pi}_{t+h}\hat{\pi}_t)=a^2\left(\frac{\rho(\rho^h-b^h)}{(\rho-b)(1-\rho b)}+\frac{b^h(1+\rho b)}{(1-b^2)(1-\rho b)}\right)\mathbb{E}(\hat{mc}_t^2)$$

Next, I derive the covariance between the ex-post forecast errors, $FE_{t,t+h}$ and ex-ante forecast revisions, $FR_{t,t+h}$:

$$\begin{aligned}
\mathbb{E}(FE_{t,t+h}FR_{t,t+h}) &= n^h(\gamma^*)^{h+1}\left[a\rho^{h+1}(1-n\rho\gamma^*)\mathbb{E}(\hat{mc}_t\hat{\pi}_t) + b(Cov(h) - n\gamma^* Cov(h+1))\right] \\
&- n^h(\gamma^*)^{h+1}\left[n^h(\gamma^*)^{h+1}(\mathbb{E}(\hat{\pi}_t^2) - n\gamma^* Cov(1))\right] \\
&= \frac{a^2 n^h(\gamma^*)^{h+1}\rho^{h+1}(1-n\rho\gamma^*)}{1-\rho b}\mathbb{E}(\hat{mc}_t^2) \\
&+ \frac{a^2 n^h(\gamma^*)^{h+1}}{1-\rho b}\frac{b^{h+1}(1+\rho b)(1-nb\gamma^*) - n^h(\gamma^*)^{h+1}(1+\rho b - n\gamma^*(\rho+b))}{1-b^2}\mathbb{E}(\hat{mc}_t^2) \\
&+ \frac{a^2 n^h(\gamma^*)^{h+1}}{1-\rho b}\left[b\rho^h(1-n\rho\gamma^*)\sum_{j=0}^{h-1}\left(\frac{b}{\rho}\right)^j - n\rho\gamma^* b^{h+1}\right]\mathbb{E}(\hat{mc}_t^2)
\end{aligned} \tag{D.13}$$

On the other hand, one can show that the variance of forecast errors is given by

$$\mathbb{E}(FR_{t,t+h}^2) = a^2 n^{2h}(\gamma^*)^{2(h+1)}\frac{(1+n^2(\gamma^*)^2 - 2n(\gamma^*)^2)(1+\rho b)}{(1-\rho b)(1-b^2)}\mathbb{E}(\hat{mc}_t^2)$$

Finally, K_h is given by the covariance between forecast errors and forecast revisions and divided by the variance of forecast revisions, that is,

$$\begin{aligned}
K_h = &\underbrace{\frac{\rho^h(1-b^2)(1-n\rho\gamma^*)\left(\rho + b\sum_{j=0}^{h-1}\left(\frac{b}{\rho}\right)^j\right) + b^{h+1}(\rho(b-n\gamma^*) + 1 - nb\gamma^*)}{n^h(\gamma^*)^{h+1}(1+n^2(\gamma^*)^2 - 2n(\gamma^*)^2)(1+\rho b)}}_{(+)} \\
&\underbrace{-\frac{\rho(b-n\gamma^*) + 1 - nb\gamma^*}{(1+n^2(\gamma^*)^2 - 2n(\gamma^*)^2)(1+\rho b)}}_{(-)}
\end{aligned} \tag{D.14}$$

Now, I compute the covariance between forecast errors and inflation realized in period t,

$$\begin{aligned}
\mathbb{E}(FE_{t,t+h}\hat{\pi}_t) &= \frac{a^2\rho^h}{1-\rho b}\mathbb{E}(\hat{mc}_t^2) + bCov(h-1) - n^h(\gamma^*)^{h+1}Cov(1) \\
&= a^2\left[\frac{\rho(\rho^h - b^h)}{(\rho-b)(1-\rho b)} + \frac{b^h(1+\rho b)}{(1-\rho b)(1-b^2)} - \frac{n^h(\gamma^*)^{h+1}(b+\rho)}{(1-\rho b)(1-b^2)}\right]\mathbb{E}(\hat{mc}_t^2)
\end{aligned} \tag{D.15}$$

Dividing the expression above by the variance of inflation, I derive M_h:

$$M_h = \underbrace{\frac{\rho^h}{1+\rho b}\left[\sum_{j=0}^{h}\left(\frac{b}{\rho}\right)^j - b^2\sum_{j=0}^{h-2}\left(\frac{b}{\rho}\right)^j\right]}_{(+)}\underbrace{-\frac{n^h(\gamma^*)^{h+1}(b+\rho)}{1+\rho b}}_{(-)} \tag{D.16}$$

D.5 Corollary 2

First, from Proposition 2, it is trivial to see that delayed over-shooting is guaranteed to occur for any parameterization of the model.

Second, I re-write the condition for which $K_h > 0$ as follows

$$K_h = \rho(1-b^2)(1-\rho\gamma^*)\frac{\rho^{h+1}-b^{h+1}}{\rho-b} + (b^{h+1}-(\gamma^*)^{h+1})(1+\rho b-\gamma^*(b+\rho)) > 0$$

Simple re-arrangement gives rise to the following inequality,

$$\begin{aligned}(\gamma^*)^{h+1} &< b^{h+1} + \rho(1-b^2)(1-\rho\gamma^*)\frac{\rho^{h+1}-b^{h+1}}{(\rho-b)(1+\rho b-\gamma^*(b+\rho))} \\ &= b^{h+1} + \rho(1-b^2)(1-\rho^2)\frac{\rho^{h+1}-b^{h+1}}{(\rho-b)(1-\rho^2)(1-b^2)} \\ &= \rho^{h+1} + \rho^h b + ... + \rho b^h + b^{h+1} = \bar{\gamma}\end{aligned} \tag{D.17}$$

where for the second equality, I rely on the fact that along the CE equilibrium, $\gamma^* = \frac{b+\rho}{1+\rho b}$, as shown in Proposition 1.

Third, I re-write the condition for which $M_h < 0$ as follows

$$M_h = \frac{\rho^h}{1+\rho b}\left[\sum_{j=0}^{h}\left(\frac{b}{\rho}\right)^j - b^2\sum_{j=0}^{h-2}\left(\frac{b}{\rho}\right)^j\right] - \frac{(\gamma^*)^{h+1}(b+\rho)}{1+\rho b} < 0$$

from which it follows that

$$(\gamma^*)^{h+1} > \frac{\rho^{h+1}-b^{h+1}-\rho^2 b^2(\rho^{h-1}-b^{h-1})}{\rho^2-b^2} = \underline{\gamma}$$

D.6 Proposition 4

As shown in the main text, when myopia is combined with well-specified forecasting rules, the aggregated optimal pricing rule can be written as $\hat{\pi}_t = \kappa \hat{mc}_t + \beta n \mathbb{E}_t \hat{\pi}_{t+1}$. The solution for inflation then is given by $\hat{\pi}_t = a_0 \hat{mc}_t$, where $a_0 = \frac{\kappa}{1-\beta n\rho}$.

1. I show that $\mathbb{I}_{k,h} \geq 0$:

$$\mathbb{I}_{k,h} = a_0 \hat{mc}_{t+k} - a_0 \rho^h \hat{mc}_{t+k-h} = a_0(\rho^k - n\rho^k)\varepsilon_t \geq 0 \tag{D.18}$$

   for any $k \geq 0$. From here, it follows that if $n = 1$, i.e., if we impose well-specified forecasts absent myopia (FIRE), $\mathbb{I}_{k,h} = 0$ for any $k \geq 0$.

2. Next, I compute the covariance between forecast errors and forecast revisions:

$$\mathbb{E}\left((\hat{\pi}_{t+h} - n^h \mathbb{E}_t \hat{\pi}_{t+h})(\mathbb{E}_t \hat{\pi}_{t+h} - \mathbb{E}_{t-1}\hat{\pi}_{t+h})\right) = n^h \rho^{2h}(1-n\rho^2)(1-n^h)\mathbb{E}(\hat{\pi}_t^2)$$

   Dividing the expression above by the variance of forecast revisions delivers

$$K_h = \frac{(1-n^h)(1-n\rho^2)}{n^h(1+n^2\rho^2-2n\rho^2)} \geq 0$$

3. Finally, I compute the covariance between forecast errors and current inflation:

$$\mathbb{E}\left((\hat{\pi}_{t+h} - n^h \mathbb{E}_t \hat{\pi}_{t+h})\hat{\pi}_t\right) = \rho^h (1 - n^h)\mathbb{E}(\hat{\pi}_t^2)$$

Dividing the expression above by the variance of current inflation gives rise to

$$M_h = \rho^h (1 - n^h) \geq 0$$

D.7 Proposition 6

Suppose there is a one-time shock in period t. The h-period ahead forecast error in period $t+h$ is given by

$$S_{i,t+h} - \widetilde{\mathbb{E}}_t^{\star} S_{i,t+h} = \mathbf{1}_i \left[\left(\sum_{j=0}^{h} \boldsymbol{A}^j \right) \boldsymbol{C} + \boldsymbol{A}^h \boldsymbol{B} \mathcal{E}_t \right] - \delta(1 - \gamma^{h+1}) \tag{D.19}$$

where $\mathbf{1}_i$ is a row vector of the same length as S_t, with 1 in the i^{th} position and 0 anywhere else. Similarly, the forecast errors for any period $t+k$ where $k > h$ are described by

$$S_{i,t+k} - \widetilde{\mathbb{E}}_{t+k-h}^{\star} S_{i,t+k} = \mathbf{1}_i \left[\sum_{j=0}^{k} \boldsymbol{A}^j \boldsymbol{C} + \boldsymbol{A}^k \boldsymbol{B} \mathcal{E}_t - \gamma^{h+1} \left(\sum_{j=0}^{k-h-1} \boldsymbol{A}^j \boldsymbol{C} + \boldsymbol{A}^{k-h-1} \boldsymbol{B} \mathcal{E}_t \right) \right] - \delta(1 - \gamma^{h+1}) \tag{D.20}$$

Suppose that the response of forecast errors in period $t+1$ to a particular shock ε_{jt} in $\mathcal{E}_t$ is positive, i.e.,

$$\mathbb{I}_{1,h} = \frac{\partial \left(S_{i,t+h} - \widetilde{\mathbb{E}}_t^{\star} S_{i,t+h} \right)}{\partial \varepsilon_{jt}} = \mathbf{1}_i \boldsymbol{A}^h \boldsymbol{B} \mathbf{1}_j' = \mathbf{1}_i Q \Lambda^h Q^{-1} \boldsymbol{B} \mathbf{1}_j' > 0 \tag{D.21}$$

where $\mathbf{1}_j$ is a row vector of the same length as $\mathcal{E}_t$, with 1 in the j^{th} position and 0 anywhere else, and $\boldsymbol{A}$ is decomposed as $\boldsymbol{A} = Q\Lambda Q^{-1}$, where Λ is a diagonal matrix containing the eigenvalues of $\boldsymbol{A}$ and the column of Q contain the respective eigenvectors of $\boldsymbol{A}$.

$$\begin{aligned} \mathbb{I}_{k,h} &= \frac{\partial \left(S_{i,t+k} - \widetilde{\mathbb{E}}_{t+k-h}^{\star} S_{i,t+k} \right)}{\partial \varepsilon_{jt}} = \mathbf{1}_i (\boldsymbol{A}^k - \gamma^{h+1} \boldsymbol{A}^{k-h-1}) \boldsymbol{B} \mathbf{1}_j' \\ &= \mathbf{1}_i Q \Lambda^k Q^{-1} \boldsymbol{B} \mathbf{1}_j' - \gamma^{h+1} Q \Lambda^{k-h-1} Q^{-1} \boldsymbol{B} \mathbf{1}_j' \end{aligned} \tag{D.22}$$

I assume that the response of S_{it+k} to ε_{jt} preserves the sign for any $k \geq h$. Then, one has to find conditions for which $\lim_{k\to\infty} \mathbb{I}_{k,h} < 0$:

$$\lim_{k\to\infty} \mathbb{I}_{k,h} = \lim_{k\to\infty} \left[\underbrace{\left(Q \Lambda^{k-h-1} Q^{-1} \boldsymbol{B} \mathbf{1}_j' \right)}_{+} \left(\frac{\mathbf{1}_i Q \Lambda^k Q^{-1} \boldsymbol{B} h_j}{Q \Lambda^{k-h-1} Q^{-1} \boldsymbol{B} h_j} - \gamma^{h+1} \right) \right] < 0 \tag{D.23}$$

Let M be the length of S_t; q_m be the product between the element of Q located in row i and column m and the element in Q^{-1} located in row m and column j; and λ_1 be the largest eigenvalue of $\boldsymbol{A}$. Then, $\lim_{k\to\infty} \mathbb{I}_{k,h} < 0$ if and only if the following limit is negative:

$$\begin{aligned}
\lim_{k\to\infty}\left(\frac{\mathbf{1}_i Q\Lambda^k Q^{-1}\boldsymbol{B}\mathbf{1}_j'}{Q\Lambda^{k-h-1}Q^{-1}\boldsymbol{B}\mathbf{1}_j'} - \gamma^{h+1}\right) &= \lim_{k\to\infty}\left(\frac{\sum_{m=1}^{M} q_m\lambda_m^k}{\sum_{m=1}^{M} q_m\lambda_m^{k-h-1}} - \gamma^{h+1}\right) \\
&= \lambda_1^{h+1}\lim_{k\to\infty}\left(\frac{q_1 + \sum_{m=2}^{M} q_m(\lambda_m/\lambda_1)^k}{q_1 + \sum_{m=2}^{M} q_m(\lambda_m/\lambda_1)^{k-h-1}}\right) - \gamma^{h+1} \\
&= \lambda_1^{h+1} - \gamma^{h+1} < 0
\end{aligned} \tag{D.24}$$

If γ exceeds the highest eigenvalue of $\boldsymbol{A}$, then $\lim_{k\to\infty} \mathbb{I}_{k,h} < 0$ is guaranteed.

References

Board of Governors of the Federal Reserve System (US). 1966 - 2018. “Federal Funds Effective Rate [FEDFUNDS].” URL https://fred.stlouisfed.org/series/FEDFUNDS.

Brooks, Stephen P. and Andrew Gelman. 1998. “General Methods for Monitoring Convergence of Iterative Simulations.” *Journal of Computational and Graphical Statistics* 7 (4):434–455. URL https://doi.org/10.1080/10618600.1998.10474787.

Giannoni, Marc and Michael Woodford. 2004. “Optimal Inflation-Targeting Rules.” University of Chicago Press. URL https://www.nber.org/books-and-chapters/inflation-targeting-debate/optimal-inflation-targeting-rules.

Hommes, Cars H., Kostas Mavromatis, Tolga Ozden, and Mei Zhu. 2022. “Behavioral Learning Equilibria in the New Keynesian Model.” Tech. Rep. 42, Netherlands Central Bank, Research Department. URL https://doi.org/10.34989/swp-2022-42.

Hommes, Cars H. and Mei Zhu. 2014. “Behavioral Learning Equilibria.” *Journal of Economic Theory* 150:778–814. URL https://doi.org/10.1016/j.jet.2013.09.002.

Koop, Gary, M. Hashem Pesaran, and Simon M. Potter. 1996. “Impulse response analysis in nonlinear multivariate models.” *Journal of Econometrics* 74:119–147. URL https://doi.org/10.1016/0304-4076(95)01753-4.

Milani, Fabio. 2006. “A Bayesian DSGE Model with Infinite-Horizon Learning: Do 'Mechanical' Sources of Persistence Become Superfluous?” *International Journal of Central Banking* 2 (3):87–106.

———. 2007. “Expectations, Learning and Macroeconomic Persistence.” *Journal of Monetary Economics* 54 (7):2065–2082. URL https://doi.org/10.1016/j.jmoneco.2006.11.007.

U.S. Bureau of Economic Analysis. 1966 - 2018a. “Gross Domestic Product: Implicit Price Deflator [GDPDEF].” URL https://fred.stlouisfed.org/series/GDPDEF.

———. 1966 - 2018b. “Population Level [CNP16OV].” URL https://fred.stlouisfed.org/series/CNP16OV.

———. 1966 - 2018c. “Real Gross Domestic Product [GDPC1].” URL https://fred.stlouisfed.org/series/GDPC1.