

Misspecified Forecasts and Myopia in an Estimated New Keynesian Model*

Ina Hajdini[†]

November 7, 2020

Abstract

The paper considers a New Keynesian model in which agents form expectations based on a combination of misspecified forecasts and myopia. The proposed expectations formation process is tested against Rational Expectations (RE), as well other assumptions about expectations, with inflation forecasting data from the U.S. Survey of Professional Forecasters. The paper then derives the general equilibrium solution consistent with the proposed expectations formation process and estimates the model with likelihood-based Bayesian methods. The paper yields three novel results: (i) Data strongly prefer the combination of autoregressive misspecified forecasting rules and myopia over other alternatives, including RE or RE subject to information rigidities; (ii) The best fitting expectations formation process for both households and firms is characterized by high degrees of myopia and simple AR(1) forecasting rules; (iii) Despite the absence of real rigidities typically found necessary for New Keynesian models with RE, the estimated model with autoregressive forecasts and myopia generates substantial internal persistence and amplification to exogenous shocks.

JEL Classification: C11; C53; D84; E13; E30; E50; E52; E70

Keywords: Misspecified Forecasts; Myopia; Survey of Professional Forecasters; Bayesian Estimation; Internal Propagation.

*I am deeply grateful to my advisors, Marco Airaudo and André Kurmann, for their advice and constructive criticism throughout this project. I thank Englin Atalay, Thomas Eisenbach, Bart Hobijn, Ryota Iijima, Lars Ljungqvist, Thomas Lubik, Tristan Potter, Benjamin Pugsley, Ricardo Serrano-Padial and Fabian Winkler for comments and suggestions. All errors are my own.

[†]School of Economics, LeBow College of Business, Drexel University, 3220 Market Street, Philadelphia, PA 19104-2875. Email: ih65@drexel.edu.

1 Introduction

The Rational Expectations (RE) assumption in Macroeconomics postulates that agents understand the true underlying model of the economy and consequently have full knowledge of the equilibrium probability distribution of economic variables. This assumption is the workhorse of modern Macroeconomics and has brought much discipline and important insights. However, it contradicts with ample evidence that agents, due to cognitive limitations or information acquiring costs, often resort to simple non-model based forecasting rules (misspecified forecasts) *and* do not appropriately take into account future payoffs/quantities (myopia).¹ To date, the literature has not incorporated both departures from RE in an equilibrium macro framework and has not formally tested them with macroeconomic data.

The present paper addresses this gap in the literature and makes its first contribution by *jointly* introducing misspecified forecasting rules and myopia in a New Keynesian framework. The second contribution is to derive the Consistent Expectations equilibrium for the inflation process and estimate its testable implications with forecasting data from the Survey of Professional Forecasters (SPF) in the U.S. The third contribution is to develop the full general equilibrium solution, while allowing agents to perpetually learn about the equilibrium, and estimate it on U.S. macroeconomic data with likelihood-based Bayesian methods.² The key novel result of the paper is that both the regression analysis on forecasting data and the likelihood-based Bayesian estimation of the full New Keynesian model prefer the specification in which agents use a combination of misspecified forecasting rules and myopia over other alternatives.

Agents are assumed to be homogenous, but endowed with imperfect common knowledge about each other's economic problems, beliefs and shocks. Therefore, agents do not understand their uniformity and consequently, are not aware of the true model of the economy. As a result, they form forecasts about the endogenous variables based on misspecified perceived laws of motion, i.e., rules that are structurally different from the minimum state variable solution implied by RE. In particular, motivated by evidence in the literature (see footnote 1), I assume that the perceived laws of motion are of an autoregressive nature.

To model myopia, I build on the idea of judgmental adjustment to forecasting rules in

¹See for instance, Tversky and Kahneman (1973, 1974), Adam (2007), Hommes (2013, 2019), Petersen (2015), Malmendier and Nagel (2016), Ganong and Noel (2019), among others. See *Related literature* for more details.

²The literature has long shown that agents tend to focus mostly on recent observations, that is rely on perpetual or constant gain learning. For instance, Fuerster et. al. (2010) argue that "actual people's forecasts place too much weight on recent changes," Malmendier and Nagel (2016) find significant micro evidence in favor of constant-gain learning, Tversky and Kahneman (1973, 1974) provide theoretical considerations.

Bullard et. al. (2008).³ Agents receive *belief shocks*, which represent news about qualitative events that might happen some time in the future, and judgmentally adjust their forecasts to the extent they are sensitive to uncertainty induced by such disturbances.^{4,5} Specifically, I assume that agents, depending on their sensitivity to belief shocks, incorporate such information into their forecasting and furthermore, their vision about the future ‘blurs’ or equivalently they excess discount forecasts about future payoffs/quantities, similar to Gabaix (2020). For instance, if agents are highly sensitive to belief shocks, current optimal decision will be based primarily on forecasts about the near-future manifesting high degrees of myopia.⁶ This mechanism defines the degree of myopia as a continuous parameter that can be estimated along other parameters. So, the current paper offers a tractable definition of myopia that is empirically amenable, relative to for instance, truncating the horizon of expectations in the spirit of finite-horizon learning.⁷

The parameters of the autoregressive forecasting rules are pinned down by the solution concept of Consistent Expectations (CE) equilibrium, as defined in Hommes and Sorger (1998) and Hommes and Zhu (2014). A first-order CE equilibrium arises when the perceived unconditional mean and first-order autocorrelation coefficient/matrix of the endogenous variable(s) coincides with the same moments as implied by the data generating process, i.e. actual law of motion, of the endogenous variable(s).

To assess the empirical relevance of the proposed expectations formation process independently from the details of the model, I follow the strategy in Coibion and Gorodnichenko (2015). The resulting regression has testable implications that differentiate the paper’s expectations formation process from RE with or without information rigidities, cognitive discounting

³An increasing body of literature find evidence in favor of judgmental adjustment to forecasting rules. See for instance, McNees (1990), Bunn and Wright (1991), Lawrence et. al. (2006), Eroglou and Croxton (2010), Moritz et. al. (2014), Petropoulos et. al. (2018).

⁴Examples include professionals’ opinions about a recession coming up in the future, future monetary policy paths or trade policies, news about the global spread of COVID-19, etc.

⁵The purpose of belief shocks - referred to as “add-factors” or “news” in Bullard et. al. (2008) - in the present paper is to induce myopic behavior of the private sector, whereas in Bullard et. al. (2008) is studying the occurrence of self-fulfilling fluctuations. Importantly, Bullard et. al. (2008) assume that the “add-factors” *always* affect expectations on a one-to-one basis, whereas I allow for flexible sensitivity to belief shocks.

⁶For instance, one could think of the U.S. stock price index sharp decline in March 2020 as an example of an extremely myopic response of investors to the seemingly surprising news of COVID-19 spread. Here is how Kessler (2020, March 8) describes it for the *Wall Street Journal*:

[...] for every stock there’s this invisible arrow pointing to some moment, some time in the future, that markets focus on to evaluate it. That time horizon isn’t published anywhere; it’s something you feel, an equity duration. Typically, it’s 12 to 18 months. [...] it’s now a coronavirus roller-coaster market. Why? Because uncertainty can collapse time horizons to months or even tomorrow.

⁷See for example Branch et. al. (2012).

as in Gabaix (2020), well-specified forecasting rules and myopia, and misspecified forecasting rules without myopia.⁸ In particular, the combination of autoregressive forecasting rules and myopia predicts that, differently from the alternatives mentioned above, the mean ex-post forecast errors should significantly depend on both the ex-ante revision of the average forecast across forecasters and the second lag of the variable being forecasted. Estimation of the implied regression on quarterly inflation SPF forecasting data from 1968:Q4 to 2014:Q2 finds that the mean inflation forecast revisions and the second lag of inflation are the only significant predictors of the mean ex-post inflation forecast errors.⁹ Consequently, the proposed expectations formation process with misspecified forecasting rules and myopia is preferred over the aforementioned expectations alternatives. In particular it is favored over RE with information rigidities as considered by Coibion and Gorodnichenko (2015), which generally posit that the mean ex-post inflation forecast errors should significantly depend on mean inflation forecast revisions only.

Evidence in favor of the proposed expectations formation process for inflation represents a natural motivation to introduce the same process to households as well, and consider a full New Keynesian model. Bayesian estimation of the model on U.S. macroeconomic data from 1966:Q1 to 2018:Q3 yields the following key results. First, consistent with forecasting data evidence, macroeconomic data strongly prefer the model whose expectations formation process is defined by a combination of autoregressive forecasts and myopia over RE. Second, the best-fitting expectations formation process is characterized by high degrees of myopia - forecasts of *at most* 3 years ahead matter for current optimal decisions - and simple AR(1) forecast rules are favored over more elaborate VAR-based forecasts. Third, the estimated high degree of myopia generates substantial internal persistence and amplification to exogenous shocks. In particular, the impulse responses of the output gap to demand and monetary shocks in the presence of misspecified autoregressive forecasts and myopia, differently from RE, are hump-shaped. Despite absence of real rigidities typically found necessary for New Keynesian models with RE to replicate business cycle features, misspecified forecasting rules and myopia strengthen the internal propagative features of the model. Specifically, autoregressive forecasts induce excess persistence, while myopia generates excess volatility, implying that both departures from RE are necessary to deliver realistic business cycle features in an otherwise standard small-scale

⁸RE models with information rigidities include the sticky-information models as in Mankiw and Reis (2002) and Reis (2006), and the imperfect information models as in Woodford (2001), Sims (2003), and Maćkowiak and Wiederholt (2009).

⁹Moreover, presence of myopia delivers a hump-shaped correlation function between forecast errors and revisions, similar to what is observed in inflation SPF data.

New Keynesian model.

The estimation of the full New Keynesian model reveals the evolution of the estimated beliefs over time. For example over the last two decades, beliefs about annualized inflation have been well-anchored at an average of 2% and the perceived first-order autocorrelation of inflation has been steadily declining. Moreover, the estimation implies that supply-side belief shocks explain about 15% of inflation volatility, whereas demand-side belief disturbances have no significant explanatory power on the volatility of the output gap and inflation.

Related literature

The present paper contributes with additional evidence to a rich body of literature showing that simple forecasting processes are preferred by data over RE (e.g., Tversky and Kahneman (1973, 1974), Adam (2007), Hommes (2013, 2019), Greenwood and Shleifer (2014), Petersen (2015), and Malmendier and Nagel (2016)).¹⁰ The paper also relates to a series of papers that discuss the analytical implications of misspecified forecasting rules, as in Hommes and Sorger (1998), Fuster et. al. (2010, 2012), Hommes and Zhu (2014), among others. In particular, the paper relies on the solution concept of first-order Consistent Expectations equilibrium, developed by Hommes and Sorger (1998) and Hommes and Zhu (2014).

The paper shares a common idea with Gabaix (2020) about myopia being excess discounting of the future, induced by some form of uncertainty.¹¹ However, differently from the present paper, in Gabaix (2020) forecasts are based on structurally well-specified rules and as mentioned earlier, forecasting data evidence presented in this paper stands in favor of a combination of misspecified forecasting rules and myopia. Evidence of myopia presented in the current work further contributes to recent developments in the empirical literature in favor of myopic agents (see for instance, Ganong and Noel (2019) who show that the only model that could rationalize household behavior given a predictable decrease in income in the data was one with myopic/short-sighted agents.).

¹⁰Experimental evidence in Adam (2007), Hommes (2013, 2019) and Petersen (2015), among others, shows that agents are commonly not model-based rational and that they tend to use simple forecasting rules. Using Michigan Consumer Survey micro data on inflation expectations, Malmendier and Nagel (2016) evidence shows that expectations are history dependent not model-based rational. From a psychological standpoint, Tversky and Kahneman (1973, 1974) argue that when trying to solve complex problems people tend to employ a limited set of heuristics. Moreover, simpler processes generate on average smaller out-of-sample forecasting errors compared to AR(p) for $p > 1$ or VARs, especially for inflation series (see for example, Atkeson and Ohanian (2001) and Stock and Watson (2007)).

¹¹In Gabaix (2020) extra-discounting could be an implication of agents receiving noisy signals about future realizations of endogenous variables (see Gabaix (2014) as well for details on microfoundations). In the present paper, excess discounting is linked to the private sector's sensitivity toward uncertainty induced by belief shocks.

The present work relates to a growing body of literature that provides evidence in favor of judgmental adjustments to forecasts, see e.g., McNees (1990), Bunn and Wright (1991), Lawrence et. al. (2006), Erouglou and Croxton (2010), Moritz et. al. (2014), Petropoulos et. al. (2018). The present paper assumes that agents myopically adjust their forecasts in response to uncertainty induced by belief shocks, and estimates the degree of myopia, or equivalently the degree of myopic adjustment with macroeconomic data.¹²

This work also shares common insights with the literature that imperfect common knowledge can explain observed persistence better than its RE counterpart (see e.g. Milani (2006, 2007), Slobodyan and Wouters (2012a), Hommes et. al. (2019)). The novelty of the present paper, however, is the finding that imperfect common knowledge outperforms RE in terms of empirical fit and it has propagative effects in the economy relative to RE, when it is accompanied by a sufficiently high degree of myopia.¹³ Relatedly, while a model set in a RE framework with a rich set of frictions as in Smets and Wouters (2003, 2007) can fit data pretty well, this paper shows that a combination of autoregressive misspecified forecasts and myopia is powerful in replicating business cycle fluctuations characteristics, with a minimalistic set of mechanical frictions. Even though this paper is fundamentally distinctive from Angeletos and Huo (2020), the empirical evidence presented here stands in favor of their analytical result that myopia and “anchoring of the current outcome to the past outcome” can be a substitute for mechanical persistence.¹⁴

The paper is also related to that body of literature that estimates general equilibrium New Keynesian models free of the RE assumption as in for instance, Del Negro and Eusepi (2011), Slobodyan and Wouters (2012a, 2012b), Ormeño and Molnár (2015), Rychalovska (2016), Cole and Milani (2017), Gaus and Gibbs (2018). Differently, the present paper estimates the model conditional on the novel combination of autoregressive misspecified forecasts and myopia. Finally, the full New Keynesian model in the present paper embeds other structures commonly used in the literature. The particular case when agents are extremely sensitive to belief shocks and completely disregard expectations beyond next period is promoted as Euler equation learning by Evans and Honkapohja (2001). The other extreme prevails when agents

¹²Relatedly, Milani (2011) estimates a New Keynesian model with expectation shocks, which differently from the present paper, are simply added to the original forecasting rule.

¹³In fact, the economy constrained to no myopia resembles to a certain extent the one under RE - consistent with recent laboratory experimental findings in Evans et. al. (2019) showing that short-horizon forecasts are characterized by more substantial deviations from RE than long-horizon.

¹⁴Angeletos and Huo (2020) prove the equivalence between a RE model with incomplete information and another RE one with myopia along with “anchoring of the current income to the past outcome, as if there was habit.” In contrast, in this paper, backward-looking components are an attribute of backward-looking misspecified forecasting rules due to imperfect common knowledge, whereas myopia is realized through a judgmental adjustment process to misspecified forecasting rules.

have no sensitivity towards belief shocks, also known as infinite-horizon learning by Preston (2005).¹⁵

The rest of the paper is organized as follows. Section 2 describes the expectations formation process in a New Keynesian pricing problem. Section 3 derives testable implications and provides evidence from inflation SPF forecasting data. Section 4 nests the expectations formation process in a small-scale full New Keynesian model and presents the main Bayesian estimation results accompanied with a series of implications. Section 5 concludes.

2 Misspecified forecasts and myopia

This section integrates misspecified autoregressive forecasting rules and myopia in a New Keynesian pricing problem and then solves for the CE equilibrium. The rationale for focusing on the pricing problem is that the literature has remained consistently focused on testing implications of different expectations formation processes on inflation forecasting data, providing thus useful comparisons to the literature. However, as shown in Section 4, the proposed expectations formation process is easily nested in other dynamic problems, such as that of the household.

2.1 New Keynesian pricing

Following Woodford (2003) among others, I assume there is a continuum of household-owned monopolistically competitive firms that face the same economic problems, are subject to the same exogenous shocks, and share the same beliefs about the future. Due to imperfect common knowledge, firms are not aware of their uniformity. The pricing problem is subject to Calvo price stickiness: each period the price cannot be adjusted with some constant probability α . Each firm chooses the optimal price that will maximize the present value of current and expected future real profits such that the demand for its goods is satisfied, and then hire the optimal amount of labor hours that will minimize production costs. The log-linearized first-order condition of each firm's pricing problem is¹⁶

$$\hat{p}_t^* = \tilde{\mathbb{E}}_t \sum_{h=0}^{\infty} (\alpha\beta)^h (\omega \hat{m}_{t+h} + \alpha\beta \hat{\pi}_{t+h+1}) \quad (1)$$

¹⁵See Eusepi and Preston (2018) as well for a review.

¹⁶See Appendix A for more details.

where $\hat{p}_t^* = \log(P_t^*/P_t)$ is the log-linear optimal price in deviation from the aggregate price \hat{P}_t ; $\tilde{\mathbb{E}}_t$ is a generic subjective expectations operator that satisfies the law of iterative expectations and standard probability rules; $\hat{m}c_t$ is the marginal cost; $\hat{\pi}_t$ is inflation; β is a discount factor; ω is a function of underlying parameters. In this partial equilibrium setting, I assume that the marginal cost is exogenous and evolves according to

$$\hat{m}c_t = \rho \hat{m}c_{t-1} + \sigma_\varepsilon \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, 1) \quad (2)$$

Each firm receives the same ε_t in the beginning of period t , however as mentioned earlier, due to imperfect knowledge they are not aware they are subject to the same marginal cost.

2.2 Misspecified forecasts

Firms understand the process of the exogenous disturbances they are subject to, hence they correctly forecast the marginal cost, i.e.,

$$\tilde{\mathbb{E}}_t^* \hat{m}c_{t+h} = \rho^h \hat{m}c_t \quad (3)$$

Given the extent of limited common knowledge, the best firms can do is forecast the future by relying on the past. Specifically, I assume that the perceived law of motion for inflation follows an AR(1) process,

$$\hat{\pi}_t = \delta + \gamma(\hat{\pi}_{t-1} - \delta) + \epsilon_t \quad (4)$$

where $\delta \in \mathbb{R}$ is the perceived unconditional mean, $\gamma \in (-1, 1)$ is the perceived unconditional first-order autocorrelation and ϵ_t is perceived to follow a white noise process. The value of ϵ_t is unknown when firms forecast future inflation, therefore

$$\tilde{\mathbb{E}}_t^* \hat{\pi}_{t+h} = \delta(1 - \gamma^{h+1}) + \gamma^{h+1} \hat{\pi}_{t-1} \quad (5)$$

2.3 Myopia

Firms receive belief shocks that represent news about qualitative events that might happen in the near-future and are judged to significantly influence future realizations of inflation and marginal cost. When forming expectations about a variable of choice, firms use information conveyed by both the misspecified forecasts defined in (3) - (5) and belief shocks. Building on Bullard et. al. (2008), upon receiving a belief shock firms incorporate it as an “add-factor” to

their inflation misspecified forecast.¹⁷ The main difference and *novelty* of the present paper relative to Bullard et. al. (2008) is that additionally, firms myopically adjust their forecast, i.e., firms' vision about the future 'blurs', to the extent they are sensitive to the uncertainty induced by the belief shock.¹⁸

Firms are assumed to be exposed to the same news about the future, and as in Bullard et. al. (2008), such news are assumed to be i.i.d. and uncorrelated with other shocks in the economy and the misspecified forecast.¹⁹ So, each firm receives the same i.i.d. belief shock about inflation $\varrho_{\pi_t} \sim \mathcal{N}(0, \sigma_{\varrho_{\pi}}^2)$, judged to have a significant effect on $\hat{\pi}_{t+h}$ for $h \geq 1$. I assume that in the wake of belief shocks about inflation, firms judgmentally adjust their forecasts and form expectations as follows:

$$\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} = b \textcolor{red}{n}^{h-1} (\tilde{\mathbb{E}}_t^* \hat{\pi}_{t+h} + (1 - \textcolor{red}{n}) \varrho_{\pi_t}) \quad (6)$$

where $\textcolor{red}{n} \in (0, 1]$ denotes each firm's degree of myopic adjustment to their forecast and $b = \left(\frac{1-\alpha \textcolor{red}{n}}{1-\alpha}\right)$ (see below for rationale). The degree of myopia $\textcolor{red}{n} \in (0, 1]$ relates to the uncertainty induced by belief shocks as follows

$$n = \frac{1}{1 + s \sigma_{\varrho_{\pi}}} \quad (7)$$

where $s \geq 0$ measures each firm's *sensitivity* to the standard deviation of belief shocks distribution $\sigma_{\varrho_{\pi}} > 0$. When firms face belief shocks drawn from a non-degenerate distribution *and* have strictly positive *sensitivity* to such disturbances, they myopically adjust their forecast to the extent they are sensitive to such shocks as in (6).²⁰

Regarding b , when firms excess discount they make optimal pricing decisions as if they cannot reset the price with probability $\alpha \textcolor{red}{n} < \alpha$, or equivalently they behave as if the average duration of their current price is $\frac{1}{1-\alpha \textcolor{red}{n}} < \frac{1}{1-\alpha}$. Then, as firms become more short-sighted due to excess discounting, they reweigh upward the horizons that continue to matter by a factor $b = \frac{\frac{1}{1-\alpha}}{\frac{1}{1-\alpha \textcolor{red}{n}}} = \left(\frac{1-\alpha \textcolor{red}{n}}{1-\alpha}\right) \geq 1$.²¹

¹⁷Bullard et. al. (2008) refer to such disturbances as “news” or “judgmental add-factors”.

¹⁸In additional difference with Bullard et. al. (2008), the purpose of belief shocks in the present paper is to induce myopic behavior of the private sector, whereas in Bullard et. al. (2008) is studying the occurrence of self-fulfilling fluctuations.

¹⁹Bullard et. al. (2008) assume that news carry new information about the future qualitative events and new information about recent qualitative events that still have an impact on the economy, while I assume they represent only the former.

²⁰In the extreme case of infinitely large sensitivity to ϱ_{π_t} ($s \rightarrow \infty$), $n \rightarrow 0$ and the myopic adjustment will be so large that the forecast beyond period $t + 1$ will be disregarded. On the contrary, insensitivity to belief shocks ($s = 0$) implies that expectations are pinned down by the forecast only.

²¹Moreover, in the general equilibrium model in Section 4, reweighing upward by a factor that is consistent with the excess discounting and equivalently agents becoming more short-sighted, is preferred by macroeconomic

Applying the same idea to belief shocks about the marginal cost, the first-order condition becomes²²

$$\hat{p}_t^* = b\tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\alpha\beta\mathbf{n})^h (\omega\hat{m}c_{t+h} + \beta\hat{\pi}_{t+h+1}) + \frac{b(1-\mathbf{n})}{(1-\alpha\beta\mathbf{n})} \varrho_{p_t} \quad (8)$$

where ϱ_{p_t} is a linear combination of inflation and marginal cost belief shocks.

2.4 Consistent Expectations equilibrium

Aggregating equation (8), the aggregate pricing scheme is described by

$$\hat{\pi}_t = \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\alpha\beta\mathbf{n})^h (\kappa\hat{m}c_{t+h} + \beta(1-\alpha\mathbf{n})\hat{\pi}_{t+h+1}) + \chi\varrho_{p_t} \quad (9)$$

where $\kappa = \omega(1-\alpha\mathbf{n})/\alpha$ and $\chi = \frac{(1-\mathbf{n})(1-\alpha)}{\alpha(1-\alpha\beta\mathbf{n})}$. Equation (9) does not reduce to the standard Phillips curve, $\hat{\pi}_t = \kappa\hat{m}c_t + \beta\tilde{\mathbb{E}}_t^*\hat{\pi}_{t+1} + \chi\varrho_{p_t}$, *unless* the forecast ($\tilde{\mathbb{E}}_t^*$) is a well-specified rule, i.e., the structure of the forecasting rule is the same as the minimum state variable solution under.²³ Under RE, firms know they are all homogenous, hence can use their own optimal condition in (8) to form expectations about inflation and the infinite-horizon New Keynesian Phillips curve reduces to the 1-step ahead standard Phillips curve.²⁴ However, with imperfect common knowledge firms cannot use their individual optimal condition to make future inferences about aggregate variables, therefore (9) would not reduce to the standard Phillips curve.

Substituting for the misspecified forecasts (3) and (5) in (9) delivers the actual law of motion for inflation:

$$\hat{\pi}_t = \beta\delta(1-\alpha\mathbf{n}) \left(\frac{1}{1-\alpha\beta} - \frac{\gamma^2}{1-\alpha\beta\mathbf{n}\gamma} \right) + \frac{\kappa}{1-\alpha\beta\rho\mathbf{n}} \hat{m}c_t + \frac{\beta(1-\alpha\mathbf{n})\gamma^2}{1-\alpha\beta\mathbf{n}\gamma} \hat{\pi}_{t-1} + \chi\varrho_{p_t} \quad (10)$$

data.

²²For simplicity purposes only, I assume that the degree of myopic adjustment is the same for inflation and marginal cost, i.e., the same n applies. This assumption does not alter the main results of the paper and it can be easily relaxed.

²³See Preston (2005) as well.

²⁴Specifically,

$$\mathbb{E}_t\hat{\pi}_{t+1} = \kappa\mathbb{E}_t\hat{m}c_{t+1} + \alpha\mathbf{n}\beta^2(1-\alpha\mathbf{n})\mathbb{E}_t\hat{\pi}_{t+2} + \mathbb{E}_t \sum_{h=2}^{\infty} (\alpha\beta\mathbf{n})^h (\kappa\hat{m}c_{t+h} + \beta(1-\alpha\mathbf{n})\hat{\pi}_{t+h+1})$$

Hence,

$$\hat{\pi}_t = \kappa\hat{m}c_t + \beta(1-\alpha\mathbf{n})\mathbb{E}_t\hat{\pi}_{t+1} + \underbrace{\mathbb{E}_t \sum_{h=1}^{\infty} (\alpha\beta\mathbf{n})^h (\kappa\hat{m}c_{t+h} + \beta(1-\alpha\mathbf{n})\hat{\pi}_{t+h+1})}_{\alpha\beta\mathbf{n}\mathbb{E}_t\hat{\pi}_{t+1}} + \chi\varrho_{p_t} = \kappa\hat{m}c_t + \beta\mathbb{E}_t\hat{\pi}_{t+1} + \chi\varrho_{p_t}$$

Firms trust (4) to be a valid perceived law of motion for inflation only if its parameters, which represent the perceived unconditional mean (δ) and first-order autocorrelation (γ), are consistent with the same moments from the data generating process, i.e., the actual law of motion for inflation. Coefficients δ and γ in equilibrium are pinned down through the solution concept of Stochastic Consistent Expectations Equilibrium, as defined by Hommes and Zhu (2014):

Definition 1 *A triple $(\mathcal{P}, \delta, \gamma)$, where \mathcal{P} is a probability measure, and δ and γ are real numbers with $\gamma \in (-1, 1)$, is called a first-order Stochastic Consistent Expectations Equilibrium if the following three conditions are satisfied:*

1. *The probability measure \mathcal{P} is a nondegenerate invariant measure for the stochastic difference equation (10);*
2. *The stationary stochastic process defined by (10) with the invariant measure \mathcal{P} has unconditional mean δ , that is, $\mathbb{E}\mathcal{P}(\pi) = \pi d\mathcal{P}(\pi) = \delta$;*
3. *The stationary stochastic process defined by (10) with the invariant measure \mathcal{P} has unconditional first-order autocorrelation coefficient γ*

Therefore, $\delta^* = 0$ and γ^* is the fixed point of $Corr(\hat{\pi}_t, \hat{\pi}_{t-1})$ in equation (10).²⁵ So, the forecast and actual law of motion of inflation along the CE equilibrium path, are, respectively

$$\tilde{\mathbb{E}}_t^* \hat{\pi}_{t+h} = (\gamma^*)^{h+1} \hat{\pi}_{t-1} \quad (11)$$

$$\hat{\pi}_t = \frac{\kappa}{1 - \alpha\beta\rho n} \hat{m}c_t + \frac{\beta(1 - \alpha n)(\gamma^*)^2}{1 - \alpha\beta n(\gamma^*)} \hat{\pi}_{t-1} + \frac{(1 - \alpha)(1 - n)}{\alpha(1 - \alpha\beta n)} \varrho_{p_t} \quad (12)$$

Note that the CE solution differs structurally from the RE one, which describes inflation as a linear function of the exogenous shocks only, i.e.,

$$\hat{\pi}_t = \frac{\kappa}{1 - \beta\rho} \hat{m}c_t + \frac{(1 - \alpha)(1 - n)}{\alpha(1 - \alpha\beta n)} \varrho_{p_t} \quad (13)$$

Moreover, myopia has a broader impact on the inflation dynamics along the CE equilibrium path relative to the RE alternative: in (12) the parameter n influences the effects of marginal cost, past inflation and belief shocks on inflation, whereas in (13) n alters only the effect of belief shocks on inflation.

²⁵ γ^* can be found numerically since an analytical solution is almost always impossible.

3 Forecasting data evidence

To assess the empirical relevance of the proposed expectations formation process in Section 2 independently from the details of the model, I follow the strategy in Coibion and Gorodnichenko (2015). I then assess the implied regression with inflation SPF forecasting data and rely on details of the model to generate unbiased estimates.

3.1 Testable implications

Consider period t and $t - 1$ expectations about future inflation as specified in equation (6), respectively,

$$\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} = bn^{h-1} \tilde{\mathbb{E}}_t^* \hat{\pi}_{t+h} + b(1-n)n^{h-1} \varrho_{p_t} \quad (14)$$

$$\tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} = bn^h \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} + b(1-n)n^h \varrho_{p_{t-1}} \quad (15)$$

Let $\tilde{\mathbb{E}}_t^* \hat{\pi}_{t+h} = \hat{\pi}_{t+h} - v_{t,t+h}$, where $v_{t,t+h}$ is the forecasting error term when there is no myopia ($n = 1$). Subtracting equation (15) from (14) and rearranging terms delivers²⁶

$$\underbrace{\hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h}}_{\text{forecasting error}} = \frac{1 - bn^{h-1}}{bn^{h-1}} \underbrace{\left(\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} \right)}_{\text{forecasting revision}} + n(1 - bn^{h-1}) \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} + error_t \quad (16)$$

where $error_t = v_{t,t+h} - (1-n)\varrho_{p_t} + n(1-n)(1 - bn^{h-1})\varrho_{p_{t-1}}$. Equation (16), which is derived solely from the expectations formation process, posits that the forecasting error for any horizon h depends on both the forecasting revision and the forecast in period $t-1$. In absence of myopia, $n = 1$, neither the forecasting revision nor the forecast in period $t-1$ would have predictive power over the forecast error.

²⁶Subtracting equation (15) from (14) and setting $\tilde{\mathbb{E}}_t^* \hat{\pi}_{t+h} = \hat{\pi}_{t+h} - v_{t,t+h}$,

$$\begin{aligned} \tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} &= bn^{h-1}(\hat{\pi}_{t+h} - v_{t,t+h}) - bn^h \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} + bn^{h-1}(1-n)(\varrho_{p_t} - n\varrho_{p_{t-1}}) \\ &= bn^{h-1}(\hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h}) + bn^{h-1} \tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - bn^{h-1} v_{t,t+h} - bn^h \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} \\ &\quad + bn^{h-1}(1-n)(\varrho_{p_t} - n\varrho_{p_{t-1}}) - bn^{h-1}(\tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) \\ &= bn^{h-1}(\hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h}) + bn^{h-1}(\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) + bn^{h-1} \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} - bn^h \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} \\ &\quad - bn^{h-1} v_{t,t+h} + bn^{h-1}(1-n)(\varrho_{p_t} - n\varrho_{p_{t-1}}) \end{aligned}$$

Hence,

$$\begin{aligned} \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h} &= \frac{1 - bn^{h-1}}{bn^{h-1}} (\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) - (\tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} - n \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h}) - (1-n)(\varrho_{p_t} - n\varrho_{p_{t-1}}) + v_{t,t+h} \\ &= \frac{1 - bn^{h-1}}{bn^{h-1}} (\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) + n(1 - bn^{h-1}) \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} + \underbrace{v_{t,t+h} - (1-n)(\varrho_{p_t} - n(1 - bn^{h-1})\varrho_{p_{t-1}})}_{error_t} \end{aligned}$$

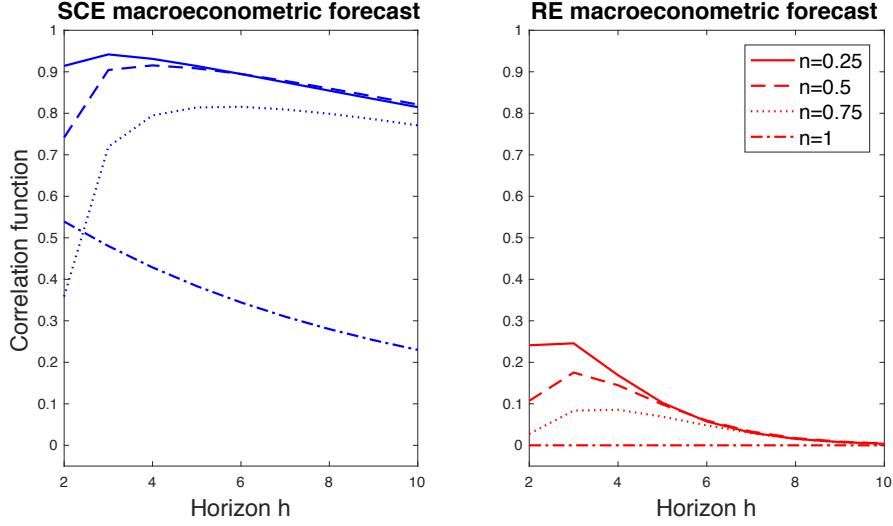


Figure 1: **Model implied correlation function between forecast errors and revisions across forecasting horizons.** Phillips curve parameterization: $\alpha = 0.48$, $\beta = 0.99$, $\rho = 0.7$, $\sigma_\varepsilon = \sigma_{\varrho_p} = 0.1$, $\kappa = 1$.

When $\tilde{\mathbb{E}}_t^*$ is a consistent expectations operator as described in section 2.2, $\tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h}$ is a linear function of inflation in period $t-2$. So, the appropriate relation between the forecasting error and revision at any forecasting horizon, h , is described by

$$\hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h} = c_h + K_h \left(\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} \right) + \zeta_{2,h} \hat{\pi}_{t-2} + error_t \quad (17)$$

- **Implication 1a):** For myopia and SCE to be supported by forecasting data, estimation should deliver significant $\hat{K}_h, \hat{\zeta}_{2,h} \neq 0$.

Similarly, when $\tilde{\mathbb{E}}_t^*$ is a RE operator, $\tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h}$ depends on inflation in period $t-1$.²⁷ So, the appropriate equation at any forecasting horizon, h , is

$$\hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h} = c_h + K_h \left(\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} \right) + \zeta_{1,h} \hat{\pi}_{t-1} + error_t \quad (18)$$

- **Implication 1b):** For myopia and RE to be supported by forecasting data, estimation should deliver significant $\hat{K}_h, \hat{\zeta}_{1,h} \neq 0$.

A nice feature of the expectations formation process of the paper is that moments, such as the correlation between the forecasting errors and revisions, vary with the horizon. In particular, as shown in Figure 1, for standard parameterizations of the Phillips curve the

²⁷ $\tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h}$ denotes expectations under no myopia, i.e., $\tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} = \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}$ for $n = 1$. Hence, $\tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} = \frac{\kappa \rho^{h+1}}{1-\beta \rho} \hat{m} c_{t-1}$, while with no myopia $\hat{\pi}_{t-1} = \frac{\kappa}{1-\beta \rho} \hat{m} c_{t-1} \Rightarrow \hat{m} c_{t-1} = \frac{1-\beta \rho}{\kappa} \hat{\pi}_{t-1}$.

correlation function between inflation forecasting errors and revisions is hump-shaped whenever $n < 1$, i.e., there is some degree of myopia. Moreover, the hump-shaped correlation function is amplified in the presence of misspecified beliefs (left panel). Under misspecified forecasts and no myopia, the correlation between error and revision is non-zero and monotonically decreasing.²⁸ On the contrary, if the forecast is well-specified and there is no myopia, we should not observe any correlation between the forecasting error and revision.

- **Implication 2:** For myopia to be supported by forecasting data, one should observe a hump-shaped correlation function between forecasting errors and revisions similar to Figure 1.

3.2 Evidence

Following Coibion and Gorodnichenko (2015), I use quarterly inflation SPF data from 1968:Q4 to 2014:Q2 and focus on the year-to-year inflation forecasts, i.e., $h = 3$.

CE. The error term in (17) is correlated with the forecasting revision, hence one has to use instruments to get an unbiased estimate for K_3 . The error term is also correlated with $\hat{\pi}_{t-2}$, because $\mathbb{E}(v_{t,t+3}\hat{\pi}_{t-2}) \neq 0$. To account for the latter bias, I take advantage of the actual law of motion in (10). One can write $v_{t,t+3}$ as $v_{t,t+3} = \mathcal{F}(\hat{\pi}_{t-3}, \hat{\pi}_{t-2}, \hat{\pi}_{t-1}, \varepsilon_{t-1}, \dots, \varepsilon_{t+3})$, where \mathcal{F} is a linear function.²⁹ Then, $\mathbb{E}(v_{t,t+3}\hat{\pi}_{t-2}) \neq 0$ because $\mathbb{E}(\hat{\pi}_{t-1}\hat{\pi}_{t-2}) \neq 0$, $\mathbb{E}(\hat{\pi}_{t-2}^2) \neq 0$ and $\mathbb{E}(\hat{\pi}_{t-3}\hat{\pi}_{t-2}) \neq 0$. Hence, adding $\hat{\pi}_{t-1}$ and $\hat{\pi}_{t-3}$ as regressors will eliminate the omitted variable bias. One can then show that after adding $\hat{\pi}_{t-1}$ and $\hat{\pi}_{t-3}$ as regressors the true estimate of $\zeta_{2,3}$ is larger than the OLS estimator. To mute the omitted variable bias in the OLS estimator of $\zeta_{2,3}$, I will consider the following regression

$$\hat{\pi}_{t+3} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+3} = c_3 + K_3 \left(\tilde{\mathbb{E}}_t \hat{\pi}_{t+3} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+3} \right) + \sum_{i=1}^3 \zeta_{i,3} \hat{\pi}_{t-i} + error_t \quad (19)$$

where $\hat{\pi}_{t+3,t}$ is the average inflation rate over the current and next three quarters. Real-time

²⁸In equation (17), $\hat{K}_h = \hat{\zeta}_{2,h} = 0$ should be interpreted as evidence for lack of myopia. As explained in the subsequent section, the error term is correlated with forecasting revision, so $\hat{K}_h = 0$ does not imply that the correlation between the forecasting error and revision is 0. Particularly, let $FE_{t,t+h}$ and $FR_{t,t+h}$ denote the forecasting error in period t and forecasting revision between period $t-1$ and t for inflation in $t+h$, $h \geq 1$. Then,

$$\mathbb{E}(FE_{t,t+h}, FR_{t,t+h}) = K_h Var(FR_{t,t+h}) + \underbrace{\mathbb{E}(FR_{t,t+h} error_t)}_{\neq 0}$$

So even if $\hat{K}_h = 0$ there should be some correlation between the forecasting revision and error.

²⁹See Appendix B for details.

	(1)	(2)	(3)
$\hat{\pi}_{t+3,t} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+3}$			
<i>Panel A: OLS</i>			
$\tilde{\mathbb{E}}_t \hat{\pi}_{t+3} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+3}$	1.193** (0.497)	1.141** (0.458)	1.141*** (0.420)
$\hat{\pi}_{t-1}$		0.021 (0.050)	-0.001 (0.073)
$\hat{\pi}_{t-2}$			0.099* (0.051)
$\hat{\pi}_{t-3}$			-0.081 (0.055)
Constant	0.002 (0.144)	-0.074 (0.174)	-0.072 (0.186)
R-squared	0.195	0.197	0.210
<i>Panel B: IV</i>			
$\hat{\pi}_{t+3,t} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+3}$	1.907*** (0.605)	2.095** (0.878)	1.396** (0.688)
$\hat{\pi}_{t-1}$		-0.050 (0.077)	-0.059 (0.132)
$\hat{\pi}_{t-2}$			0.131* (0.071)
$\hat{\pi}_{t-3}$			-0.051 (0.082)
Constant	0.004 (0.084)	0.190 (0.263)	-0.106 (0.177)
R-squared	0.123	0.093	0.201
F-stat	14.07	9.238	9.536
Observations	172	172	172

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 1: **Estimates of equation (19).**

inflation data is used since final historical data might embed redefinitions and reclassifications.³⁰

RE. When the forecast is well-specified, the forecast revision might still be correlated with the error term, hence instrumental variable estimation should be the . Moreover, the error

³⁰Real-time data is extracted from Real-Time Research Data Center, currently administered by the Federal Reserve Bank of Philadelphia.

term is uncorrelated with $\hat{\pi}_{t-1}$, hence the following regression is considered

$$\hat{\pi}_{t+3} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+3} = c_3 + K_3 \left(\tilde{\mathbb{E}}_t \hat{\pi}_{t+3} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+3} \right) + \zeta_{1,3} \hat{\pi}_{t-1} + error_t \quad (20)$$

Table 1, Panel A and B, columns (2) - (3), present estimates of regressions in (20) - (19) by OLS and instrumental variables (IV), respectively. In panel B, columns (1) - (2), the past two lags of oil price changes are used as instruments for forecasting revisions whereas in column (3) the second lag of Fernald's TFP series is used as an instrument as well. The estimate of K_3 remains significantly positive across all specifications. Column (3) provides the estimator for $\zeta_{2,3}$: $\hat{\zeta}_{2,3} \approx 0.1$ and it is significant at 10%. Note that the true estimate of $\zeta_{2,3}$ is even larger than the one reported in Table 1 (see Appendix B for more details). In none of the regressions is $\hat{\zeta}_{1,3}$ significantly different from 0, rejecting thus the null that the forecasting rule is well-specified.

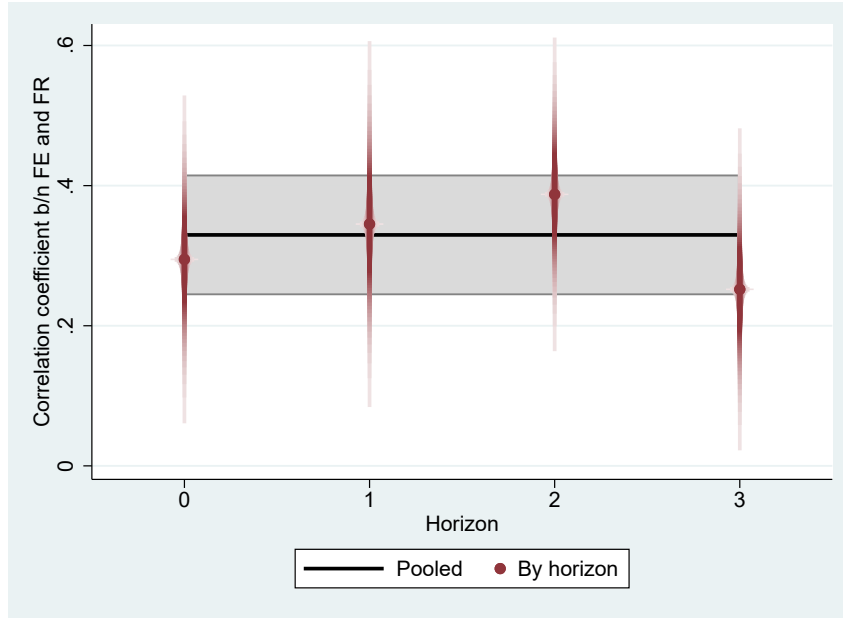


Figure 2: **SPF data implied correlation function between forecast errors and revisions across forecasting horizons.**

Figure 2 presents the correlation function between forecasting errors and revisions in SPF data. The correlation coefficients and their 95% confidence intervals were obtained through bootstrapping methods. The grey rectangle and black line depict the average and confidence interval of the correlation value for pooled data across horizons. It is obvious that the pattern resembles very much the hump-shaped function implied by the Phillips curve model with myopia in Figure 1.

3.3 Relation to other expectation processes in the literature

Equation (18) coincides with one of the main regressions in Coibion and Gorodnichenko (2015). The authors find that once forecast revisions are controlled for, the first lag of inflation becomes highly insignificant - see Table 1, columns (1) - (2) for a replication of their results. Such a result can be interpreted as evidence favoring informational frictions as in the sticky-information model of Reis (2006), noisy-information models of Woodford (2001), Sims (2003), Maćkowiak and Wiederholt (2009). However, equation (19) is a generalized version of the main regression considered in Coibion and Gorodnichenko (2015). Importantly, estimates of (19) in column (3) of Table reftab: main_FE_SCE_3 imply that forecasting data prefer misspecified autoregressive forecasts and myopia is preferred over RE with information rigidities.

Moreover, Coibion and Gorodnichenko (2015) interpret insignificance of $\hat{\pi}_{t-1}$ as lack of evidence in favor of structural departures from rationality, as considered in this paper. Results in Table 1 show that indeed that is the case when there is no myopia, however, in the presence of myopia inflation forecasting data stand in favor of structural departures from RE in the form of myopia and misspecified forecasts.

Equation (18) is implied by myopia (cognitive discounting) in Gabaix (2020) as well,³¹

$$\hat{\pi}_{t+h} - \underbrace{\tilde{\mathbb{E}}_t \hat{\pi}_{t+h}}_{K_h} = \underbrace{\left(\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} \right)}_{\zeta_{1,h}} + \underbrace{ng^{h+1}(1-n^h)}_{\zeta_{1,h}} \hat{\pi}_{t-1} + v_{t,t+h}$$

where $n \in [0, 1]$ is the extra-discounting parameter, $g \neq 0$ is a model parameter (see footnote below), and $v_{t,t+h}$ is the error term under RE consisting of innovation terms from period t to $t+h$.³² For the myopia mechanism in Gabaix (2020) to be validated by SPF forecasting data, it must be that *both* forecasting revisions and the first lag of inflation have significant predictive power over forecasting errors. Table 1 clearly shows that the forecasting revision is

³¹The implied actual law of motion for inflation under RE is $\hat{\pi}_t = g\hat{m}c_t = g\hat{\pi}_{t-1} + \varepsilon_t$, where $g = \frac{\kappa}{1-\beta\rho}$. However, agents perceive inflation evolving as $\hat{\pi}_t = ng\hat{\pi}_{t-1} + \varepsilon_t$. Hence, subjective expectations about inflation in period $t+h$ are $\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} = (ng)^h \hat{\pi}_t = n^h \tilde{\mathbb{E}}_t \hat{\pi}_{t+h}$, where $\tilde{\mathbb{E}}_t$ is the RE operator. $\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} = \hat{\pi}_{t+h} - v_{t,t+h}$ and $v_{t,t+h}$ is a linear function of innovation terms $\varepsilon_t, \varepsilon_{t+1}, \dots, \varepsilon_{t+h}$. So, $\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} = n^h(\hat{\pi}_{t+h} - v_{t,t+h})$

$$\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} = n^h(\hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h}) + n^h \tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} - n^h v_{t,t+h}$$

Rearranging terms and setting $\tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} = (ng)^{h+1} \hat{\pi}_{t-1}$, we have

$$\hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h} = \frac{1-n^h}{n^h} \left(\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} \right) + ng^{h+1}(1-n^h) \hat{\pi}_{t-1} + v_{t,t+h}$$

³²The error term might be correlated with the forecasting revision, hence instrumental variables should be used to deliver an unbiased estimate for K_h .

significantly positive, whereas $\hat{\pi}_{t-1}$ is not.

4 General equilibrium model

This section nests the proposed expectations formation process into the baseline New Keynesian DSGE model of Woodford (2003), with the only difference being that agents *learn* to use misspecified forecasting rules. Bayesian estimation of the model on U.S. aggregate data seeks to reveal i) the preferred forecasting process, ii) the estimated value of the degree of myopia; iii) the relative role of misspecified forecasting rules and myopia at the macroeconomic level.

Households and firms comprehend the exogenous shocks processes of the technology and mark-up shock processes, hence they correctly forecast future shocks. Since all households and firms are homogenous, they are subject to the same shocks. Nevertheless, due to imperfect common knowledge, they are unaware of this fact and do not use it to infer about the rest of the endogenous variables in the future. Agents are assumed to not know the Taylor rule governing interest rates - even though many choices of the central bank become public information through frequent announcements, that does not mean that people understand the rules guiding these monetary decisions. Finally, for sake of simplicity, I assume that the degree of myopic adjustment to forecasts is the same across sectors and belief shocks.³³ The model is fairly standard, hence I delegate all details to Appendix A.

4.1 Basics

Households. There is a continuum of identical households, $i \in [0, 1]$, who are unaware of each other's homogeneity. They consume from a set of differentiated goods, supply labor, and invest in riskless one-period bonds.³⁴ The consumption bundle of each household over the set of differentiated goods, $j \in [0, 1]$, is determined by the Dixit-Stiglitz aggregator with time-varying elasticity of substitution. Each period, the household receives labor income and dividends from the monopolistically competitive firms, and she maximizes expected lifetime utility with respect to consumption, labor supply and bonds, subject to her budget constraint.

Additionally, each household receives two belief shocks about qualitative events that might happen next period and affect their expectations about future income \hat{y}_{t+h} and $\hat{\Phi}_{t+h} = \frac{1}{\sigma} \left(\hat{R}_{t+h} - \hat{\pi}_{t+h+1} - \xi \right)$. The shocks, namely ϱ_{y_t} and ϱ_{Φ_t} , are i.i.d. normally distributed with mean 0 and variance $\sigma_{\varrho_y}^2$ and $\sigma_{\varrho_\Phi}^2$, respectively; they are uncorrelated with each other or other shocks in the economy.

³³This assumption can be easily relaxed. More details provided by the author upon request.

³⁴Bonds are assumed to be in zero net-supply.

Applying the myopic adjustment process, adjusted for the household problem, the optimal consumption rule for each household is

$$\begin{aligned}\hat{c}_t = & b_c \hat{b}_{t-1} + (1 - \beta) \hat{y}_t + \beta \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\beta \mathbf{n})^h \left((1 - \beta \mathbf{n}) \hat{y}_{t+h+1} - \frac{1}{\sigma} \left((\hat{R}_{t+h} - \hat{\pi}_{t+h+1}) - (\hat{\xi}_{t+h} - \hat{\xi}_{t+h+1}) \right) \right) \\ & + \frac{\beta(1 - \mathbf{n})}{1 - \beta \mathbf{n}} \varrho_{c_t}\end{aligned}\quad (21)$$

where b_c is a function of model's parameters and steady-state values; β is the discount factor; \hat{c}_t is real consumption; \hat{b}_{t-1} denotes past bond choice in real terms; \hat{y}_t is real income (function of the real wage and dividends); \hat{R}_t gross return on the nominal bond choice, $\hat{\xi}_t$ is a preference shock; $\varrho_{c_t} = (1 - \beta \mathbf{n}) \varrho_{y_t} - \varrho_{\Phi_t}$.

Firms. The firms side problem is similar to what is described in Section 2, with the difference that the marginal cost now is endogenous and the elasticity of substitution among firms is time-varying. The latter implies that the optimal pricing problem will be subject to a mark-up shock. After applying myopic adjustment to forecasts, the first-order condition of the pricing problem for each firm, log-linearized around steady-state values, is

$$\hat{P}_t^* - \hat{P}_t = b \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\alpha \beta \mathbf{n})^h (\omega \hat{m}_{c_{t+h}} + \alpha \beta \hat{\pi}_{t+h+1} + \hat{\mu}_{t+h}) + \frac{b(1 - \mathbf{n})}{1 - \alpha \beta \mathbf{n}} \varrho_{p_t} \quad (22)$$

where $\hat{\mu}_t$ denotes a mark-up shock.

Monetary Policy. The central bank controls nominal interest rates through a standard Taylor rule that reacts to deviations of inflation from its target $\bar{\pi}$, and deviations of output gap x_t from its steady-state, while smoothing the interest rate path with some degree $\rho_r \in [0, 1)$. The log-linearized policy rule is

$$\hat{R}_t = \rho_r \hat{R}_{t-1} + (1 - \rho_r) \phi_{\pi} \hat{\pi}_t + (1 - \rho_r) \phi_x \hat{x}_t + \sigma_v \varepsilon_t^v, \quad \varepsilon_t^v \sim \mathcal{N}(0, 1) \quad (23)$$

Aggregate Economy. Imposing market clearing conditions in (21) - (22) and rewriting equations in terms of the output gap delivers the final aggregate demand and supply, respectively

$$\hat{x}_t = \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\beta \mathbf{n})^h \left((1 - \beta \mathbf{n}) \hat{x}_{t+h+1} - \frac{1}{\sigma} \left(\hat{R}_{t+h} - \hat{\pi}_{t+h+1} \right) + \frac{1}{\sigma} \hat{e}_{t+h} \right) + \frac{1 - \mathbf{n}}{1 - \beta} \varrho_{c_t} \quad (24)$$

$$\hat{\pi}_t = \tilde{\mathbb{E}}_t \sum_{h=0}^{\infty} (\alpha\beta n)^h (\kappa \hat{x}_{t+h} + \beta (1 - \alpha n) \hat{\pi}_{t+h+1} + \hat{u}_{t+h}) + \frac{(1 - n)(1 - \alpha n)}{\alpha(1 - \alpha\beta n)} \varrho_{p_t} \quad (25)$$

where $\kappa = \omega(\varphi + \sigma)$. Variables \hat{e}_t and $\hat{u}_t = \omega\hat{\mu}_t$ are, respectively, the natural interest rate (demand shock) and the cost-push shock, both assumed to follow AR(1) processes,

$$\hat{e}_t = \rho_e \hat{e}_{t-1} + \sigma_e \varepsilon_t^e, \varepsilon_t^e \sim \mathcal{N}(0, 1) \quad (26)$$

$$\hat{u}_t = \rho_u \hat{u}_{t-1} + \sigma_u \varepsilon_t^u, \varepsilon_t^u \sim \mathcal{N}(0, 1) \quad (27)$$

Let $\Theta = \{\alpha, \beta, n, \sigma, \kappa, \phi_\pi, \phi_x, \rho_r, \rho_e, \rho_u, \sigma_e, \sigma_u, \sigma_v, \sigma_{\varrho_c}, \sigma_{\varrho_p}\}$. Then the model can be compactly written in matrix form as

$$A_0(\Theta)S_t = A_1S_{t-1} + \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (F(\Theta))^h A_2(\Theta)S_{t+h+1} + B(\Theta)\mathcal{E}_t \quad (28)$$

where $S_t = [\hat{x}_t \ \hat{\pi}_t \ \hat{R}_t \ \hat{e}_t \ \hat{u}_t]'$ is the state vector; $\mathcal{E}_t = [\varepsilon_t^e \ \varepsilon_t^u \ \varepsilon_t^v \ \varrho_{c_t} \ \varrho_{p_t}]'$ is the exogenous shocks vector; A_0, A_1, A_2, B and F are coefficient matrices. The aggregate economy in (28) is sufficiently flexible to nest three model specifications. i) $n \rightarrow 0$: This model specification exhibits the highest degree of myopia, i.e., agents are extremely sensitive to belief shocks. ii) $n = 1$: In this case, agents are insensitive to belief shocks and they take into account an infinite-horizon of expectations to make optimal decisions, exhibiting no myopia at all.³⁵ iii) $n \in (0, 1)$: This is the novel and benchmark model specification of the paper, where a realistic value for n is provided through Bayesian inference in the Section 4.3.

4.2 SAC-learning

Households and firms *learn* to use autoregressive forecasting rules to form expectations about future endogenous variables, i.e., output gap, inflation and nominal interest rates, nested in $S_{1:3,t}$

$$S_{1:3,t} = \boldsymbol{\delta}_{t-1} + \boldsymbol{\gamma}_{t-1}(S_{1:3,t-1} - \boldsymbol{\delta}_{t-1}) + \boldsymbol{\epsilon}_t \quad (29)$$

where $\boldsymbol{\delta}_{t-1}$ is the mean of $S_{1:3,0:t-1}$ series; $\boldsymbol{\gamma}_{t-1}$ represents the first-order correlation matrix between $S_{1:3,0:t-2}$ and $S_{1:3,1:t-1}$ series; $\boldsymbol{\epsilon}_t$ is a white noise process. The formulation in (29)

³⁵Versions of the model in (28) with $n = 0$ have been used by Evans and Honkapohja (2001, 2009), Bullard and Mitra (2002) - among others - to study implications of adaptive learning; Bullard et. al. (2008) to study the occurrence and properties of exuberance equilibria; Hommes et. al. (2019) to investigate the occurrence and properties of stochastic consistent expectations equilibria. Preston (2005) and Milani (2006) have used the economy in (28) with $n = 1$ to investigate implications of adaptive learning in an infinite horizon learning setting (see Eusepi and Preston (2018) as well for an extensive review).

nests commonly used forecasting rules, such as AR(1) and VAR(1) processes, for which I will estimate the model. The forecast of $S_{1:3,t+h}$ conditional on information about $S_{1:3,t-1}$, available in the beginning of period t is

$$\widetilde{\mathbb{E}}_t^* S_{1:3,t+h} = \boldsymbol{\delta}_{t-1} + (\boldsymbol{\gamma}_{t-1})^{h+1} (S_{1:3,t-1} - \boldsymbol{\delta}_{t-1}) \quad (30)$$

Households and firms update their forecasting rules using sample autocorrelation coefficient (SAC) learning. This procedure has been first introduced in economics by Hommes and Sorger (1998) and it relies on the Yule-Walker equations combined with sample estimates of autocorrelation coefficients: $\boldsymbol{\delta}_t$ and $\boldsymbol{\gamma}_t$ are recursively updated according to

$$\begin{aligned} \boldsymbol{\delta}_t &= \boldsymbol{\delta}_{t-1} + \iota (S_{1:3,t} - \boldsymbol{\delta}_{t-1}) \\ \boldsymbol{\gamma}_t &= \boldsymbol{\gamma}_{t-1} + \iota ((S_{1:3,t} - \boldsymbol{\delta}_{t-1})(S_{1:3,t-1} - \boldsymbol{\delta}_{t-1})' - \boldsymbol{\gamma}_{t-1}(S_{1:3,t} - \boldsymbol{\delta}_{t-1})(S_{1:3,t-1} - \boldsymbol{\delta}_{t-1})') \boldsymbol{\eta}_t^{-1} \\ \boldsymbol{\eta}_t &= \boldsymbol{\eta}_{t-1} + \iota ((S_{1:3,t} - \boldsymbol{\delta}_{t-1})(S_{1:3,t} - \boldsymbol{\delta}_{t-1})' - \boldsymbol{\eta}_{t-1}) \end{aligned} \quad (31)$$

where $\boldsymbol{\eta}_t$ is the second moment matrix, and ι is the gain parameter that nests the two types of learning. With constant gain learning, $\iota = \bar{\iota}$ is a constant parameter and it mimics a situation where a rolling window of data with length approximately equal to $\frac{1}{\bar{\iota}}$ is used to revise moments. With decreasing gain learning, on the other hand, $\iota = \frac{1}{t+1}$ and all available historical data is used to update. The former approach is preferred because it has been universally found to improve empirical fit and the literature has shown that agents focus on recent observations when updating forecasting rules.^{36,37}

4.3 Bayesian estimation

The state-space representation of the model is

$$S_t = C_0(\Theta, \boldsymbol{\gamma}_{t-1}) \boldsymbol{\Delta}_{t-1} + C_1(\Theta, \boldsymbol{\gamma}_{t-1}) S_{t-1} + C_2(\Theta) \mathcal{E}_t \quad (32)$$

$$Y_t - \bar{Y} = P S_t \quad (33)$$

³⁶See for example, Del Negro and Eusepi (2011), Ormeño and Molnár (2015), Rychalovska (2016), Cole and Milani (2017), Gaus and Gibbs (2018), among many others.

³⁷For instance, Fuester et. al. (2010) claim that “actual people’s forecasts place too much weight on recent changes.” Malmendier and Nagel (2016) find significant micro evidence in favor of constant-gain learning. See Tversky and Kahneman (1973, 1974) as well for theoretical considerations. Additionally, the evolution of beliefs under decreasing gain learning depends on the length of data, whereas constant gain learning is robust to it.

together with the dynamic system in (31), where $\Delta_t = [\delta'_t \quad \mathbf{0}_{1 \times 2}]'$; $Y_t = [x_t^{obs} \quad \pi_t^{obs} \quad R_t^{obs}]'$ is the vector of observables; P is a matrix mapping model variables to the observables; \bar{Y} is a vector containing the observables' mean. I use quarterly data on real GDP, real potential GDP as reported by the U.S. Congressional Budget Office, GDP deflator and Fed funds rate from 1966 to 2018, extracted from the Federal Reserve Economic Data (FRED). The output gap is measured as the log difference between the real GDP and potential GDP.³⁸ I refer the reader to Appendix C.1 for more details on data preparation. The state-space form of the model in (31)-(33) is a Gaussian system, hence I evaluate the likelihood function using the Kalman filter. The posterior distribution then is calculated as

$$p(\Theta \mid Y_{1:T}) \propto p(Y_{1:T} \mid \Theta)p(\Theta) \quad (34)$$

where $p(Y_{1:T} \mid \Theta)$ is the data likelihood and $p(\Theta)$ the prior distribution of parameters. I use the Metropolis-Hastings algorithm to generate draws from the posterior distribution. I generate two blocks with 360000 draws each and discard the first 60000 draws. I evaluate moments of pre-sample data from 1960 to 1965 and use them as the initial learning parameters, δ_0 and γ_0 , for the Kalman filter procedure.³⁹

I fix the discount factor $\beta = 0.99$ and assume that belief shocks are drawn from the same normal distribution with mean 0 and standard deviation σ_{ϱ_p} . For the rest of parameters, I set priors commonly used in the literature, as in for instance, Milani (2006), Smets and Wouters (2007), Herbst and Schorfheide (2015), to mention a few. Priors are given in Table 2. The Calvo parameter, α , is following a beta prior with mean 0.5 and standard deviation 0.1. The prior for n follows a less informative beta distribution with mean 0.5 and standard deviation of 0.2. The inverse intertemporal elasticity of substitution coefficient, σ , follows a gamma distribution with mean 2. The Phillips curve elasticity with respect to current output, follows a beta prior with mean 0.3 and standard deviation 0.15. Policy reaction coefficients towards deviations of inflation and output from their steady-state values are normally distributed with mean 1.5 and 0.5, respectively, and share the same standard deviation of 0.25. The autocorrelation of

³⁸Bayesian inference when the HP-filtered series of output are utilized as a measure of potential output produces similar results. Estimates are provided by the author upon request.

³⁹In terms of beliefs initiation, forecasting rules that rely on past aggregate variables *only* have a slight advantage over rules that include shocks as well. Since beliefs are tied to moments from data, the natural choice is to match initial beliefs to pre-sample data moments. This makes the estimation vulnerability to initial beliefs - commonly faced in models with forecasting rules that depend on shocks - disappear. To give an idea of the different approaches used to generate initial beliefs when the forecasting process depends on shocks, Milani (2006) estimates initial conditions on pre-sample data; Milani (2007) treats initial beliefs as parameters and estimates them along with the model's structural parameters; Slobodyan and Wouters (2012a, 2012b) initiate beliefs at the implied moments of the RE solution, apart from the other two aforementioned methods.

Parameters		pdf	mean	standard deviation
Calvo parameter	α	\mathcal{B}	0.5	0.1
Negligence to belief shocks	n	\mathcal{B}	0.5	0.2
Inv. IES	σ	\mathcal{G}	2	0.5
Phillips curve elast.	κ	\mathcal{B}	0.3	0.15
Feedback to x	ϕ_x	\mathcal{N}	0.5	0.25
Feedback to π	ϕ_π	\mathcal{N}	1.5	0.25
Interest rate smooth	ρ_r	\mathcal{B}	0.5	0.2
Autocorr. e	ρ_e	\mathcal{B}	0.5	0.2
Autocorr. u	ρ_u	\mathcal{B}	0.5	0.2
Std. ε^e	σ_e	\mathcal{IG}	0.1	2
Std. ε^u	σ_u	\mathcal{IG}	0.1	2
Std. ε^v	σ_v	\mathcal{IG}	0.1	2
Std. ϱ	$\sigma_{\varrho p}$	\mathcal{IG}	0.1	2
Gain parameter	$\bar{\iota}$	\mathcal{G}	0.035	0.015

Table 2: **Priors**

all shocks and interest rate smoothing follow a beta distribution with mean 0.5 and standard deviation 0.2. The standard deviation of all shocks follows an inverse gamma distribution with mean 0.1 and standard deviation 2. Finally, the learning gain parameter follows a gamma distribution prior with mean 0.035 and standard deviation 0.015.

4.3.1 AR(1) versus VAR(1)

To investigate the model performance and behavior of beliefs when agents use VAR(1) versus AR(1) forecasting rules, I estimate the benchmark SAC-learning model specification with $\mathbf{n} \in (\mathbf{0}, \mathbf{1})$ under constant gain learning.⁴⁰

	AR(1)	VAR(1)
Laplace Approximation	-260.7*	-268.5
Bayes Factor	$e^{7.8}$	1
Modified Harmonic Mean	-260.1	-268.1
Bayes Factor	e^8	1

Table 3: **Log Marginal Likelihood**. The asterisk denotes strong evidence in favor of the model with AR(1) forecasting rules.

Table 3 reports the log marginal data likelihood evaluated using the Laplace approximation at the posterior mode and modified harmonic mean method as in Geweke (1999).⁴¹ Under

⁴⁰Results on the other two specifications, namely $\mathbf{n} = \mathbf{0}$ and $\mathbf{n} = \mathbf{1}$, are available by the author upon request.

⁴¹Bayes Theorem implies that $p(Y_t) = \int p(\Theta | Y_t)p(\Theta)d\Theta$, which is impossible to compute analytically. The

constant gain learning, the model with AR(1) forecasting rules delivers higher estimates of the log marginal data likelihood, and thus explains data better relative to VAR(1) forecasting. The odds in favor of AR(1) forecasting processes relative to VAR(1) are to the factor of more than e^7 . According to Kass and Raftery (1995), a factor magnitude whose natural log is higher than 3 denotes strong evidence in favor of the model with superior fit, in this case being the one with AR(1) forecasting rules.

Moreover, as exhibited in Figure 3, when agents engage in constant gain learning of a VAR(1) forecasting process, the perceived first-order correlation between any two distinct aggregate variables is estimated to fluctuate around 0, whereas the perceived first-order autocorrelation of each aggregate fluctuates around a strictly positive value. Overall, simple AR(1) forecasting processes fit the data better than a VAR(1) and using the latter is not more informative than AR(1).⁴²

4.3.2 Posterior distribution

In what follows, I focus on the results implied by the SAC-learning model with AR(1) forecasting rules. Table 4 reports characteristics of the posterior distribution under RE and constant gain SAC-learning with $\mathbf{n} = \mathbf{0}$ and $\mathbf{n} = \mathbf{1}$.⁴³ The model when agents learn to use misspecified forecasting rules and exhibit relatively high degrees of myopia fit the data significantly better than RE. Estimates of log marginal data densities obtained through the Laplace approximation and Modified Harmonic Mean method are significantly higher for the decreasing gain SAC-learning model specifications with high degrees of myopia. Values in parenthesis report Laplace approximation of the marginal data density estimated at the posterior mode is

$$p(Y_t) \approx \frac{p(Y_t | \Theta^*)p(\Theta^*)}{(2\pi)^{-N/2}|H_{\Theta\Theta}(\Theta^*)|^{1/2}}$$

where Θ^* is the posterior mode, N is the number of parameters being estimated, and $H_{\Theta\Theta}(\Theta^*)$ is the Hessian of the likelihood function with respect to Θ , evaluated at $\Theta = \Theta^*$. The Modified Harmonic Mean (MHM) method of Geweke (1999) is evaluated using the posterior distribution draws,

$$p(Y_t) \approx \left[\frac{1}{M - M_0} \sum_{m=M_0+1}^M \frac{f(\Theta^m)}{p(Y_t|\Theta^m)p(\Theta^m)} \right]^{-1}$$

where M is the total number of draws, M_0 is the number of discarded draws, and $f(\cdot)$ is the density of a truncated normal distribution.

⁴²Such an outcome is even more emphasized when agents engage in decreasing gain learning of VAR(1) forecasts. In this case, beliefs about the first-order correlation between any two distinct aggregate variables converges towards 0 over time, implying that γ_t converges towards a diagonal matrix. The plot of beliefs evolution with decreasing gain learning of VAR(1) forecasting rules is exhibited in Appendix C.3.

⁴³Posterior distributions are well-behaved, with no bimodal behavior. I rely on the method proposed by Brooks and Gelman (1998) to analyze convergence statistics. Histogram figures and convergence statistics for estimated models with AR(1) forecasting rules are exposed in Appendix C.2.1 and C.2.2.

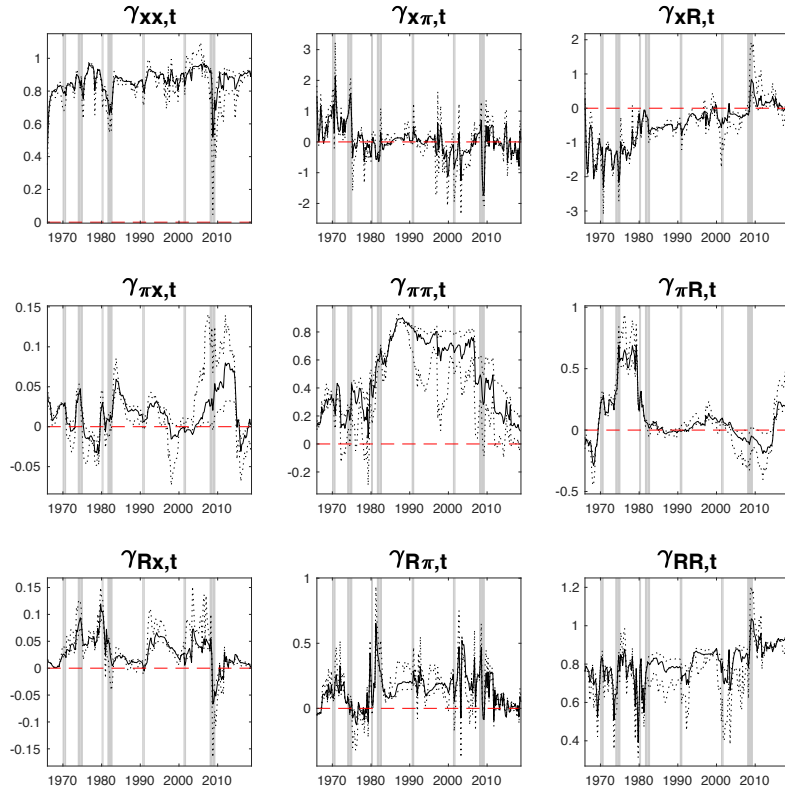


Figure 3: **Evolution of the VAR(1) forecast coefficients in the benchmark constant gain SAC-learning model.** The black and dotted curves plot implied beliefs for structural parameters set at their estimated posterior mode and 90% highest posterior density, respectively. Grey areas indicate recessionary periods as reported by the National Bureau of Economic Research.

the Bayes factor value of the model specification relative to RE: the log of the Bayes factor for the SAC-learning model with $\mathbf{n} \in (\mathbf{0}, \mathbf{1})$ is higher than 3.⁴⁴ On the other hand, the benchmark SAC-learning model with AR(1) forecasting rules and $\mathbf{n} = \mathbf{1}$, i.e., no myopia, fits data worse than RE, and as a result worse than SAC-learning with myopia.

The posterior estimate of the parameter capturing the degree of myopia, n , is significantly different from 1, showing evidence in favor of largely myopic agents in the U.S. economy. The posterior mean of n is around 0.32, meaning that on average current expectations about beyond 12 quarters ahead are practically disregarded in present optimal decisions. Recalling the relationship between n and the standard deviation of belief shocks in (7) that $n = \frac{1}{1+s\sigma_\theta}$, we can also compute a measure for the sensitivity to belief shocks, s . In particular, at the posterior mode estimates of n and σ_θ , s would be 14.8.

In the benchmark SAC-learning model, one can separately identify the Calvo parameter α and Phillips curve elasticity κ . The posterior mean of the Calvo parameter is estimated to be 0.49 for the benchmark SAC-learning model and 0.56 for the model with $\mathbf{n} = \mathbf{1}$. The implied expected price duration for the benchmark and no myopia specifications is, respectively, 5.9 and 6.8 months on average. This is in accordance with findings in Bils and Klenow (2004) that for most goods, prices change on average once every six to nine months, suggesting that $\alpha \in [0.5, 0.67]$. The Phillips curve elasticity is more than four times higher when agents are endowed with misspecified beliefs compared to RE ($\kappa \approx 0.04$ under SAC-learning, but it is only 0.01 under RE). Policy parameters are generally robust across specifications, with the posterior mode estimates of policy reaction to output gap being between 0.2 and 0.4, reaction to inflation around 1.4 - 1.7 and interest rate smooth parameter being close to 0.9.

The less myopic the economy is, the higher the degree of relative risk aversion: the inverse intertemporal elasticity of substitution is estimated to be only 2.38 for the benchmark specification and it increases to 6.25 when agents are extremely forward-looking. Experimental studies in the literature have identified a similar link between risk aversion and horizon. For instance, Noussair and Wu (2006) and Coble and Lusk (2010) conduct experiments where subjects are presented with a list of lotteries. The former find that the number of subjects with higher risk aversion for the present is higher than the number of subjects with lower risk aversion for the present. Similarly, the latter paper finds that the average risk aversion decreases with the temporal distance of risk.

The posterior mode values for the learning gain parameter $\bar{\tau}$ vary between 0.04 and 0.05.

⁴⁴Kass and Raftery (1995) state that a factor magnitude whose natural log is higher than 3 denotes strong evidence in favor of the model with superior fit.

This implies that a rolling window of 20 - 25 quarters (5 - 6 years) is used to update the forecasting process.

Shocks are significantly less persistent under SAC-learning than RE.⁴⁵ The novel empirical outcome of the paper is that myopia plays a *central* role in controlling how much of the observables' inertia is due to subjective expectations. The posterior mode estimate of the demand shock autocorrelation in the benchmark model is 0.74 and it is significantly lower than under RE or its estimate of 0.99 in the SAC-learning model with $\mathbf{n} = \mathbf{1}$. The posterior mode estimate of the cost-push shock in the benchmark model is 0.86 and it is (almost) significantly lower than under RE, but significantly higher than its estimate of 0.49 in the SAC-learning model with $\mathbf{n} = \mathbf{1}$.

On the other hand, because estimates of the first-order autocorrelation of the demand shock are decreasing with the degree of myopia, the opposite happens with the standard deviation of the innovation related to the shock: the posterior mode estimate for the innovation to the demand shock is 1.5 for the benchmark specification and 0.09 when there is no myopia. Similarly, the posterior mode estimate of the innovation related to the cost-push shock for the benchmark specification is 0.08, but 0.2 when $\mathbf{n} = \mathbf{1}$. The standard deviation of the monetary shock is generally robust across models at an estimated posterior mode of 0.2, while the same estimate for the standard deviation of belief shocks is around 0.15.

Higher estimates of the standard deviations in the benchmark SAC-learning model relative to RE or the specification with no myopia should not be interpreted as shocks being less propagated in the benchmark SAC-learning in terms of volatility. To see this I re-estimate the posterior mode of the RE model and SAC-learning specification with no myopia, while fixing the shocks' persistence to their benchmark AR(1)-SAC-learning posterior mode estimates in Table 4. As shown in Table 5, the implied variance of the demand shock \hat{e}_t for the benchmark SAC-learning model is the lowest. On the other hand, the implied variance of the cost-push shock \hat{u}_t for the benchmark SAC-learning model is higher than RE but much lower than in the SAC-learning specification with no myopia. Overall, as the economy becomes more shortsighted, misspecified forecasts become a more powerful source of propagation.

In the present paper, excess discounting starts in period $t + 2$, whereas “cognitive discounting” in Gabaix (2020) starts in period $t + 1$. The implication of the latter is an overly mitigation of general equilibrium effects. As a final exercise, I compare posterior mode estimates of the

⁴⁵This confirms previous findings in the literature, such as Milani (2006), Milani (2007), Slobodyan and Wouters (2012a, 2012b), Hommes et. al. (2019) that deviations from RE can partly account for observed persistence.

benchmark SAC-learning model with Gabaix (2020) in Table 6.

	RE				SAC-learning							
					$n \in (0, 1)$				$n = 1$			
Parameters	mode	mean	5%	95%	mode	mean	5%	95%	mode	mean	5%	95%
Inverse IES (σ)	4.41	4.30	3.70	5.00	2.38	2.58	1.88	3.43	6.25	6.78	5.75	7.96
Calvo (α)	-	-	-	-	0.49	0.5	0.32	0.67	0.56	0.54	0.37	0.72
Myopia (n)	-	-	-	-	0.31	0.32	0.1	0.57	-	-	-	-
Phil. curve elast. (κ)	0.006	0.008	0.002	0.013	0.04	0.04	0.02	0.06	0.03	0.03	0.02	0.04
Feedback to x (ϕ_x)	0.30	0.32	0.21	0.43	0.4	0.44	0.27	0.66	0.48	0.41	0.22	0.66
Feedback to π (ϕ_π)	1.42	1.45	1.19	1.70	1.44	1.45	1.13	1.79	1.03	1.45	1.13	1.79
Interest rate smooth (ρ_R)	0.86	0.86	0.83	0.89	0.91	0.92	0.89	0.95	0.93	0.92	0.88	0.95
Corr AD shock (ρ_e)	0.94	0.93	0.90	0.96	0.74	0.75	0.65	0.84	0.99	0.98	0.97	0.99
Corr AS shock (ρ_u)	0.87	0.86	0.82	0.91	0.86	0.79	0.61	0.91	0.43	0.48	0.36	0.59
Std. AD shock (σ^e)	0.43	0.47	0.33	0.60	1.38	1.46	0.95	2.06	0.09	0.15	0.09	0.23
Std. AS shock (σ^u)	0.04	0.04	0.03	0.05	0.08	0.1	0.07	0.16	0.22	0.2	0.16	0.24
Std. MP shock (σ^v)	0.21	0.21	0.20	0.23	0.21	0.21	0.19	0.23	0.21	0.21	0.19	0.23
Std. Belief (σ^e)	-	-	-	-	0.15	0.16	0.09	0.26	-	-	-	-
Learning gain (\bar{l})	-	-	-	-	0.04	0.05	0.03	0.08	0.05	0.04	0.03	0.06
Log marginal data density												
Laplace approximation	-299.50				-260.70*				-306.00			
	(1)				$(e^{38.8})$				$(e^{-6.5})$			
Modified Harmonic Mean	-299.59				-260.10*				-305.40			
	(1)				$(e^{39.5})$				$(e^{-5.9})$			

Table 4: **Posterior distribution for RE and benchmark SAC-learning.** Values in parentheses denote Bayes factor of the model relative to RE. The asterisk denotes strong evidence in favor of the model relative to RE.

	SAC-learning, $n \in (0, 1)$	SAC-learning, $n = 1$	RE
Variance of \hat{e}_t	4.18	6.64	4.24
Variance of \hat{u}_t	0.02	0.07	0.01

Table 5: **Implied variance of the demand and cost-push shocks for the RE and AR(1)-SAC-learning models.** Models are re-estimated with ρ_e and ρ_u fixed at their posterior modes in the benchmark AR(1)-SAC-learning model specification.

	n	ρ_e	ρ_u	σ_e	σ_u	σ_v
SAC-learning, $n \in (0, 1)$	0.31	0.74	0.86	1.38	0.08	0.21
Gabaix (2020)	0.54	0.94	0.88	1.25	0.14	0.21

Table 6: **Posterior mode for parameters of interest.**

4.3.3 AR(1) forecasts and belief shocks

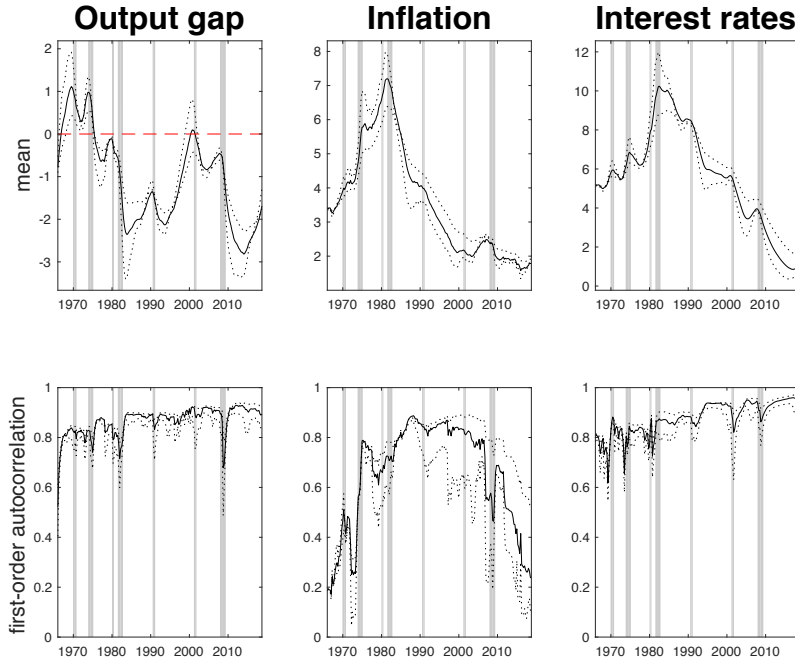


Figure 4: **Evolution of the AR(1) forecast coefficients in the benchmark SAC-learning model.** The black and dotted curves plot implied beliefs for structural parameters set at their estimated posterior mode and 90% highest posterior density, respectively. Grey areas indicate recessionary periods as reported by the National Bureau of Economic Research.

Figure 4 plots the historical evolution of the misspecified forecast coefficients in the benchmark SAC-learning model with parameters set at their posterior mode and 90% highest posterior density values. During recessions, output gap falls, causing a decrease of its mean over the recent periods and a break of the persistent pattern prior to the contraction. As shown

in Figure 4, recessionary periods, indicated by the shaded grey areas, have been historically associated with a decrease in the perceived mean and first-order autocorrelation of the output gap.

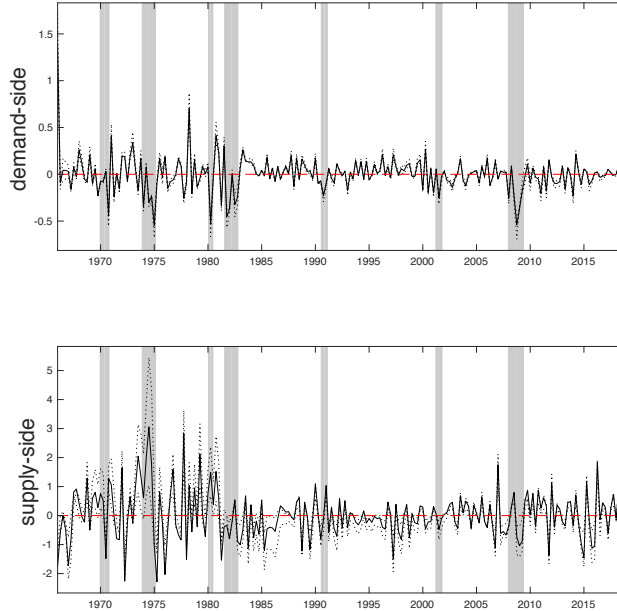


Figure 5: **Evolution of belief shocks in the benchmark SAC-learning model.** The black and dotted curves plot implied disturbances for structural parameters set at their estimated posterior mode and 90% highest posterior density, respectively. Grey areas indicate recessionary periods as reported by the National Bureau of Economic Research.

On the other hand, there seems to be a shift in the way agents perceive moments of inflation and interest rates during recessions, in early ‘80s. Before the early ‘80s, recessions have been associated with increasing beliefs about the mean and first-order autocorrelation of inflation and nominal rates. On the contrary, during and after the Great Moderation, economic turmoils are characterized by a decrease in beliefs about the mean and first-order autocorrelation of inflation and nominal interest rates. Therefore, the well-documented contrast between the U.S. macroeconomy during the ‘70s and the Great Moderation period is similarly mirrored in agents’ forecasts about inflation and nominal interest rates.⁴⁶ Another interesting observation from Figure 4 is that the implied beliefs about the annualized mean of inflation over the last decade have been particularly steady at 2%.

Figure 5 plots the evolution of the filtered demand- and supply-side belief shocks over time. Both types of belief shocks are more volatile in the first two decades of the timeline. The supply-side belief disturbances are in general of higher magnitude relative to the demand-sided

⁴⁶See for instance, Bianchi (2013) and references therein, for a discussion on the differences between the two periods.

shocks. Recessions are historically associated with negative belief shocks on the demand-side of the economy. Moreover, up until the early '80s economic contractions are characterized by positive supply-side shocks, whereas afterwards a clear link between recessions and the sign of belief shocks received by firms is not obvious.

4.3.4 Impulse response functions

Computing the impulse response functions (IRF) under SAC-learning is slightly more complicated than RE because the response of aggregates to any shock depends on the initial beliefs held at the period when the shock happens and beliefs respond to shocks as well. To make the IRF comparable across time periods and models, I assume that the economy prior to the shock is at its steady-state.⁴⁷ Figure 6 plots the 3-dimensional IRF of output gap and inflation to a one standard deviation demand, cost-push and monetary shock of the benchmark model calibrated at the estimated posterior mode.⁴⁸

Figure 6 shows that the response of aggregates to various fundamental shocks highly depends on the status of perceived persistence prior to the shock. Inflation has become less responsive to demand and monetary shocks, whereas output gap has been reacting less to cost-push shocks over the last decade.⁴⁹ Figure 7 then projects the three dimensional IRF on the [response - periods of response] plane, with the red curve plotting the average response of aggregates.⁵⁰ Apart from the relatively large variation in responses due to beliefs, the hump-shape response of output to most shocks is remarkable. In particular, Figure 7 emphasizes that even though there are no mechanical frictions, such as habit in consumption or labor market rigidities, and there is no persistence in monetary or belief shocks, the response of output gap and inflation replicates characteristics of a business cycle. Therefore, subjective beliefs together with myopia substitute for mechanical frictions and can thus explain business cycle fluctuations using simpler structural models.

⁴⁷Consecutively, prior beliefs about the mean of the aggregates are set to 0, whereas prior beliefs about the perceived first-order autocorrelation each period are set to their implied value by the model.

⁴⁸The 3-dimensional IRF to belief shocks are exposed in Appendix C.2.3.

⁴⁹I refer the reader to Appendix C.2.3 for projections of the 3-dimensional IRF on the [response - time] plane to get an idea of the change in the response magnitude of aggregates to shocks over the years. An interesting observation of Figure 19 is that during the Great Recession the response of the economy to external stimuli is especially shrunk relative to other periods. This is solely due to remarkably low perceptions about the output gap, inflation and nominal rates during those quarters, relative to the rest of the timeline.

⁵⁰The impact of a shock on each aggregate is the same across time periods because the prior belief about the mean of the aggregate prior to the shock is set to 0.

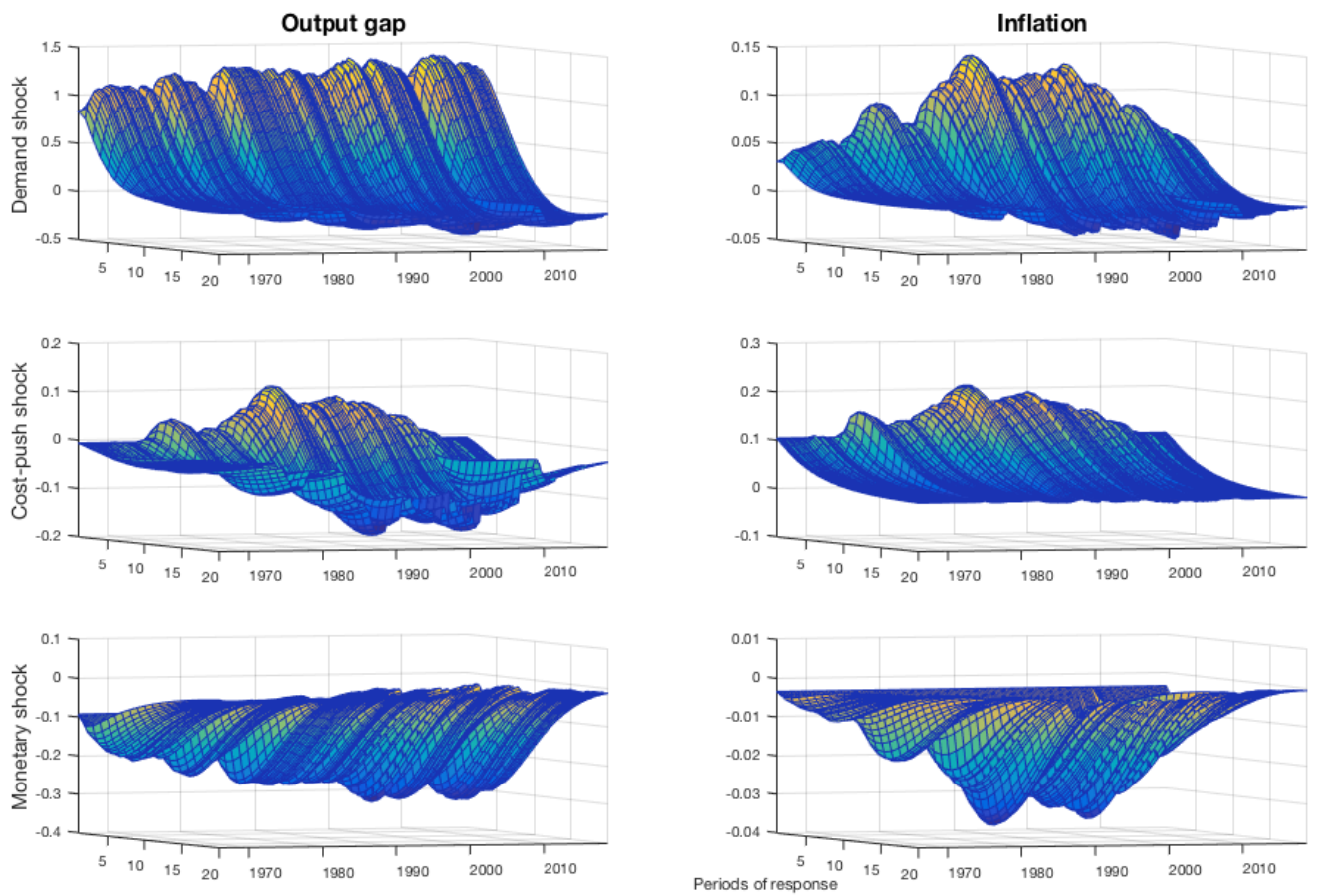


Figure 6: 3-Dimensional impulse response functions to a one standard deviation positive demand, cost-push and monetary shock for the benchmark SAC-learning model.

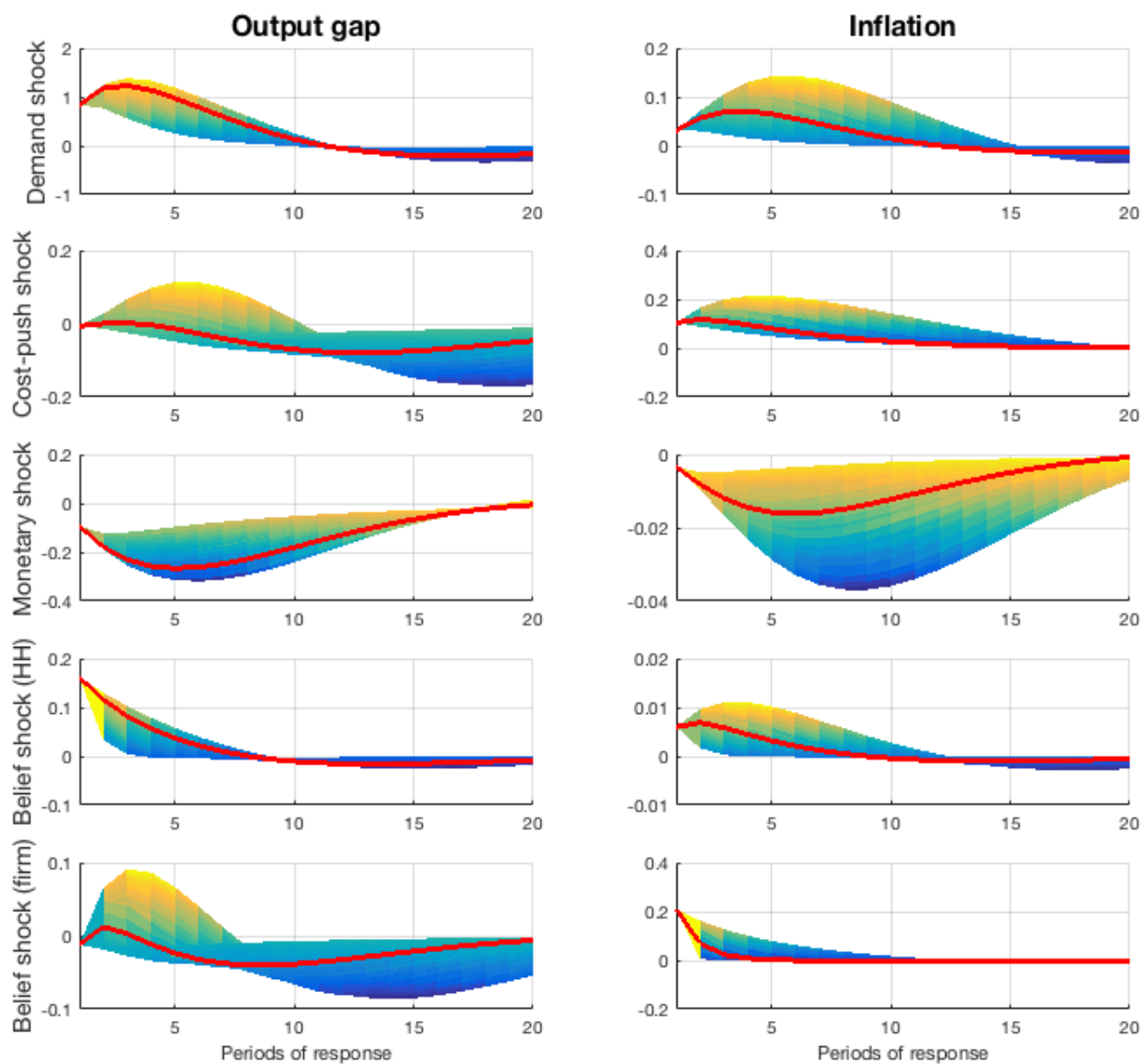


Figure 7: **Impulse response functions to a one standard deviation positive demand, cost-push, monetary and belief shocks for the benchmark model.** Red curve: average response.

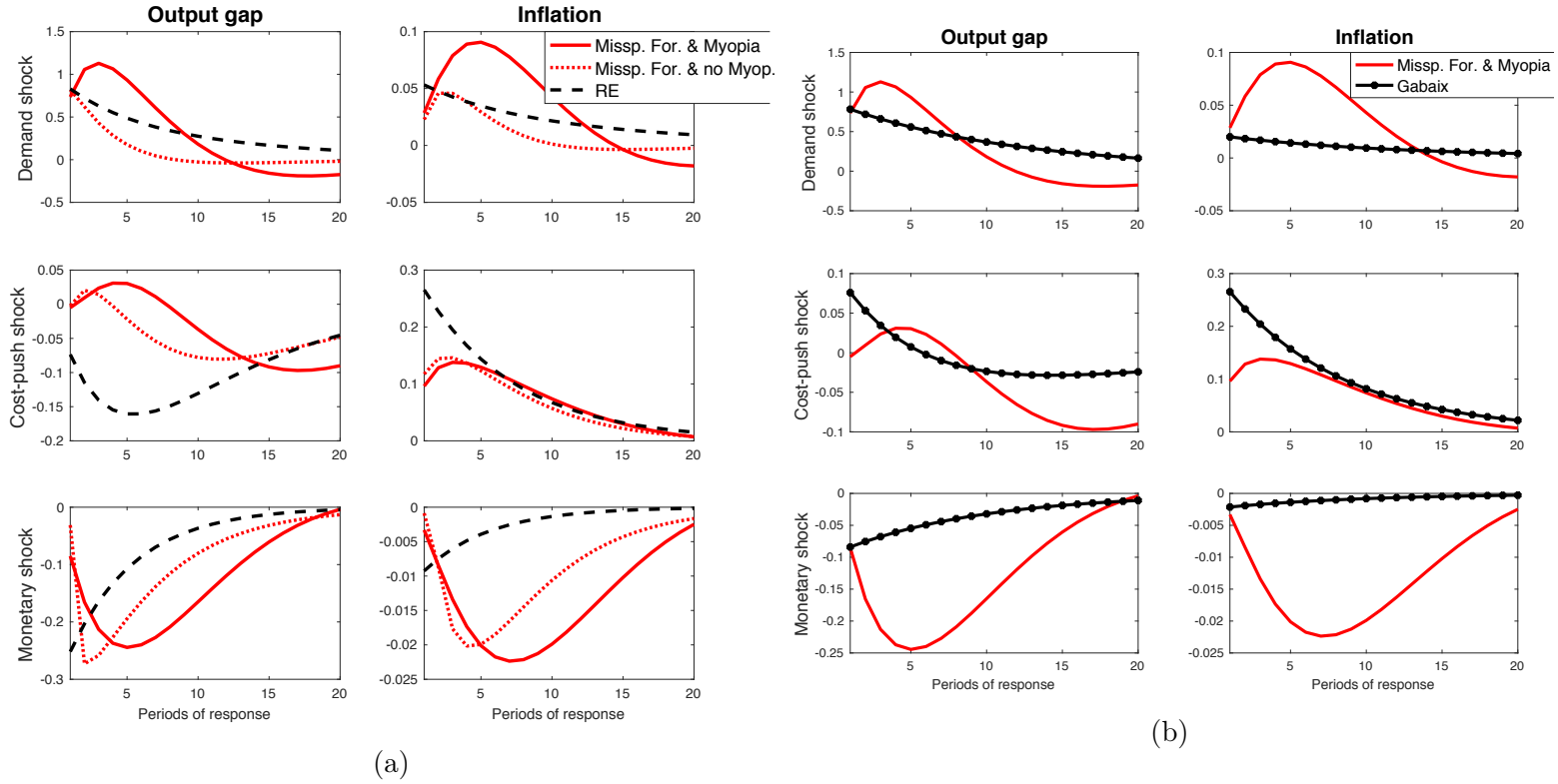


Figure 8: **Average impulse response functions to a one standard deviation positive demand, cost-push and monetary shock.** Red curve: average response under the benchmark SAC-learning model. Red dashed curve: average response under SAC-learning with $n = 1$. Black dotted curve: response under RE. Black circled curve: myopia as in Gabaix (2020).

To understand the relative importance of misspecified beliefs and myopia in propagating shocks and replicating features of business cycles, I calibrate the shocks' processes for all specifications at the posterior mode estimates of the benchmark SAC-learning model, while the rest of the parameters for each model are set at their respective posterior mode as reported in Table 4. Figure 8, left panel a), plots the average impulse responses of output gap and inflation for RE, SAC-learning with and without myopia. While the model with no myopia can generate mildly hump-shape responses of output gap when the economy is subject to a cost-push or monetary shock, that is not the case when the economy is hit by a demand shock. In fact, the output gap IRF of the SAC-learning model with no myopia is very much alike to the one under RE. Moreover, for almost all shocks, aggregates in the benchmark SAC-learning model return to their steady-state values later compared to the specification with no myopia or RE. Misspecified beliefs propagate shocks and induce excess persistence relative to RE, however, their amplification effect is stronger the more myopic agents are. While the IRF to demand and monetary shocks for the benchmark SAC-learning model are internally propagated relative to the other models, the propagative effects are less pronounced when the economy is hit by a cost-push shock. The right panel b) in the same figure compares the IRF under SAC-learning and myopia versus Gabaix (2020). The latter generates no hump-shaped response and internal propagation properties are even weaker than those under RE.⁵¹ Overall, the IRF point out to the idea that models with backward-looking components, which in this paper arise due to backward-looking forecasting rules, and myopia become a substitute for mechanical persistence, recently emphasized by Angeletos and Huo (2020).

4.3.5 Aggregate variables (pseudo) decomposition

To quantify the contribution of belief shocks, relative to the other disturbances, in generating output and inflation volatility over time, I compute the historical aggregate (pseudo) decomposition of output gap and inflation into shocks across different expectation formation specifications, and then perform a variance (pseudo) decomposition exercise. The decompositions are described as pseudo because they do not take into account the non-linear effects shocks have on the perceived first-order autocorrelation, γ_t . However, the exercises deliver approximations of the true decompositions.

Let the contribution of a shock ε_t on S_t at time t be S_t^h , where $i\varepsilon_t \in \{\varepsilon_t^e, \varepsilon_t^u, \varepsilon_t^v, \varrho_{c_t}, \varrho_{p_t}\}$. S_t^ε

⁵¹A more relevant comparison would be one where excess discounting is as in Gabaix (2020), but with SAC-learning of misspecified forecasting rules. In that case the IRF improve in terms of business cycle replication, yet they cannot outperform the IRF implied by the benchmark SAC-learning model of the paper. Results provided by the author upon request.

then is measured as follows

$$\begin{aligned}
S_t^\varepsilon &= C_0(n, \gamma_{t-1}) \begin{bmatrix} \delta_{t-1}^\varepsilon \\ \mathbf{0}_{2 \times 1} \end{bmatrix} + C_1(n, \gamma_{t-1}) S_{t-1}^\varepsilon + C_2(n) \mathbb{I}_\varepsilon \varepsilon_t \\
\delta_t^\varepsilon &= \delta_{t-1}^\varepsilon + \bar{v} (S_{1:3,t-1}^\varepsilon - \delta_{t-1}^h)
\end{aligned} \tag{35}$$

where $\mathbb{I}_\varepsilon = 1$ is an indicator vector whose element in the position of shock ε_t is 1 and all the other elements are 0. The contribution of the shock ε_t on the overall variance of the aggregates is computed as

$$V_\varepsilon = \frac{Var(S_{1:T}^\varepsilon)}{Var(S_{1:T})} \tag{36}$$

Models are calibrated at their posterior mode estimates as reported in Table 4 and ???. Figure 9 presents the historical decomposition of aggregate variables into shocks: panels to the left expose the decomposition under RE, while the ones to the right show the decomposition of the benchmark SAC-learning model. There are two stark differences between the two. First, effects of monetary policy shocks on output and inflation are much smaller under RE compared to SAC-learning. Under RE, agents understand the policy rule, and specifically the nature of the policy shock, whereas under SAC-learning, even though the policy shock has no persistence, subjective beliefs with high degrees of myopia propagate it over time. Similarly, demand shocks as well have a more persistent effect on both output and inflation under SAC-learning. Second, supply-side belief shocks have been driving inflation up until mid ‘80s, switching to a negative effect afterwards.

Table 7 shows that demand shocks explain most of the volatility in output gap, while cost-push shocks explain most of inflation variance across all models. Interestingly, in the benchmark SAC-learning model belief shocks received by firms explain more than 15% of inflation volatility. On the other hand, demand-side belief shocks explain less aggregate volatility relative to supply-side belief shocks. One explanation for this could be that, as shown in Figure 5, the implied demand-side belief shocks are much less volatile than supply-side belief shocks.

	Shocks	Output gap	Inflation	Interest rates
SAC-learning, benchmark	demand	80.12	20.57	50.35
	cost-push	1.50	50.63	13.72
	monetary	18.14	13.64	34.37
	belief (household)	0.02	0.00	0.01
	belief (firm)	0.21	15.16	1.55
SAC-learning, $n = 1$	demand	59.44	35.57	62.98
	cost-push	9.04	45.69	16.14
	monetary	31.52	18.74	20.88
RE	demand	86.71	6.60	42.31
	cost-push	6.79	93.29	28.38
	monetary	6.50	0.10	29.31

Table 7: **Variance Decomposition in %**

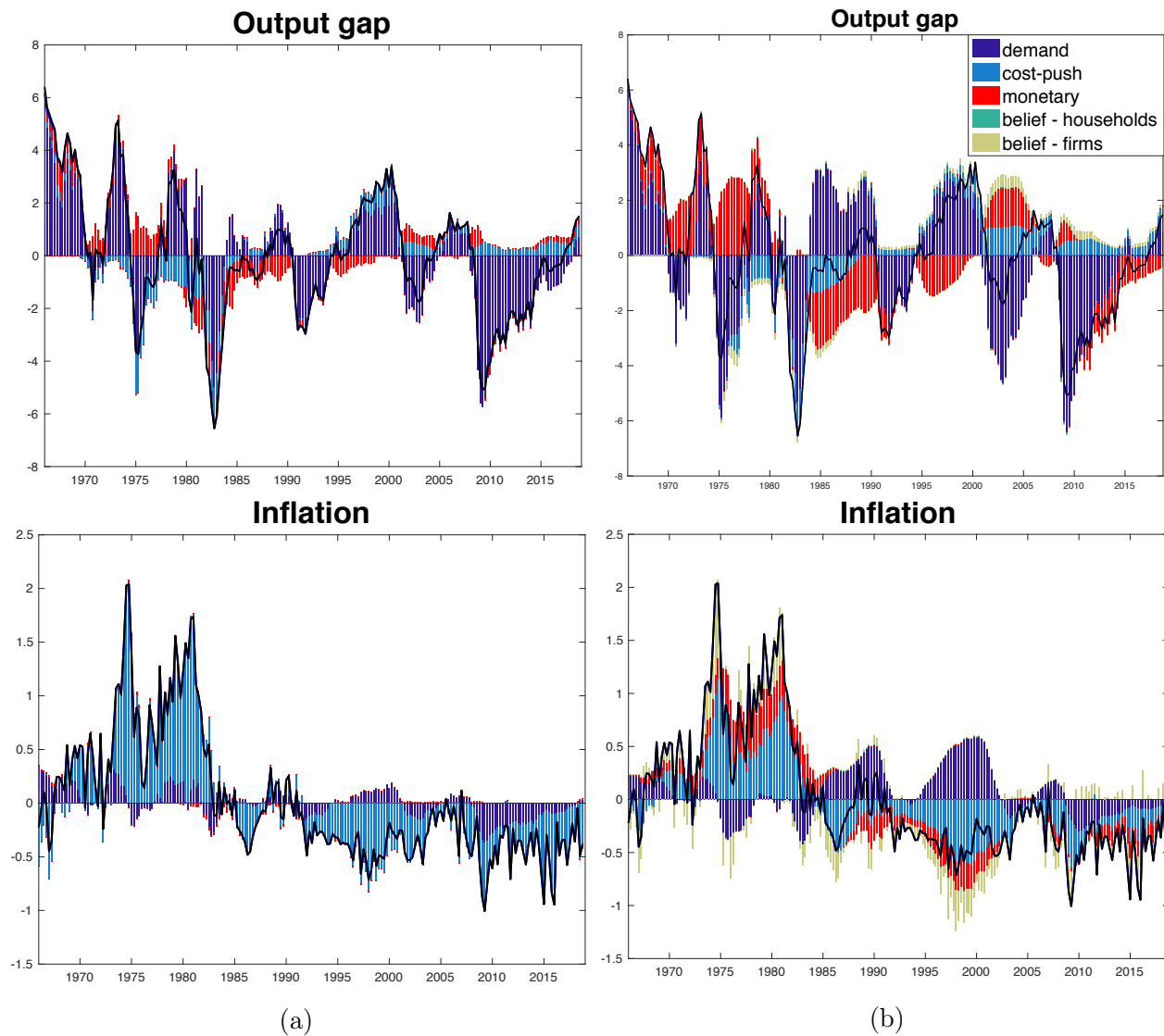


Figure 9: **Historical decomposition of output gap and inflation.** Left panels in (a): decomposition under RE. Right panels in (b): decomposition under benchmark SAC-learning. Black curve plots actual data.

Table 7 underestimates the true contribution of belief shocks in aggregate volatility, because it does not explicitly measure the effect that belief shocks have on the economy through altering the structure of the expectations formation process. For instance, it is because of a strong response of the private sector to belief shocks that the contribution of the demand shock on output gap in the benchmark model is around 40% higher compared to the model with no myopia.

5 Concluding remarks

The present paper contributes to the literature by simultaneously incorporating two of the most conspicuous deviations from the RE assumption - misspecified forecasts and myopia in a unified New Keynesian model framework that is amenable to macroeconomic data. The second contribution is to incorporate these departures from RE in a partial equilibrium pricing problem, derive testable implications, and estimates them with inflation forecasting data from the SPF. The third contribution is to embed the same departures in a full New Keynesian model, derive the general equilibrium under sample autocorrelation coefficient learning and estimate the model using Bayesian methods. The paper underscores four novel results. First, both estimation approaches strongly prefer the model in which agents use misspecified forecasts and myopia over the RE alternative. Second, partial equilibrium analysis with SPF data implies that the proposed expectations formation process is also preferred over RE with information rigidity. Third, the best fitting expectations formation process for both households and firms is characterized by high degrees of myopia and simple AR(1) forecasts as opposed to VAR(1). Fourth, the estimated high degree of myopia - in the presence of misspecified forecasts - generates substantial internal persistence and amplification to exogenous shocks.

The current paper lays solid grounds in service to future research. I am currently working on extending the framework to allow for reversible regime switches in monetary policy's responsiveness to inflation and output gap, and volatility of exogenous shocks. The limited knowledge of agents extends to regime shifts as well. This is in stark difference with regime shift studies under RE that pose the strong assumption that agents become instantly aware of any regime switches. The goal is to quantify the relative importance of policy and shocks' volatility regime shifts in a New Keynesian model to account for the Great Inflation, Volcker Disinflation and the Great Moderation in the U.S., as well as entertain the novel idea of a shift in the degree of myopia over time. The estimation will shed new light on (i) why it was so costly for the Federal Reserve to reduce inflation in the 1980s, and (ii) why inflation did not

fluctuate much during the Great Recession.

Appendix

A DSGE Model

A.1 Structural Model

Households. There is a continuum of identical households, $i \in [0, 1]$, who are unaware of each other's homogeneity. They 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.

$$c_{it} = \left[\int_{j=0}^1 c_{it}(j)^{\frac{\zeta_t-1}{\zeta_t}} dj \right]^{\frac{\zeta_t}{\zeta_t-1}} \quad (\text{A.1})$$

$$P_t = \left[\int_{j=0}^1 P_{jt}^{1-\zeta_t} \right]^{\frac{1}{1-\zeta_t}} \quad (\text{A.2})$$

where ζ_t is the time-varying elasticity of substitution. The corresponding Lagrangian is

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

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

$$c_{it}(j) = \left(\frac{\chi_{it}}{P_{jt}} \right)^{\zeta_t} c_{it} \quad (\text{A.3})$$

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

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

This implies further that

$$c_{it}(j) = \left(\frac{P_{jt}}{P_t} \right)^{-\zeta_t} c_{it} \quad (\text{A.4})$$

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

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

with budget constraint satisfying

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

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; ξ_{it} a preference shock. 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 as

$$R_{t-1}B_{i,t-1} = B_{it} - W_t H_{it} - D_{it} + P_t c_{it} \quad (\text{A.5})$$

where $\int_{j=0}^1 D_{it}(j) dj = D_{it}$. The first-order conditions (FOC) with respect to consumption, bonds and hours, respectively, are

$$\xi_{it} c_{it}^{-\sigma} = \lambda_{it} P_t \quad (\text{A.6})$$

$$\lambda_{it} = \beta \tilde{\mathbb{E}}_{it} R_t \lambda_{i,t+1} \quad (\text{A.7})$$

$$\psi \xi_{it} H_{it}^{\varphi} = \lambda_{it} W_t \quad (\text{A.8})$$

Combining (A.6) with (A.7) and (A.8),

$$c_{it}^{-\sigma} = \beta \tilde{\mathbb{E}}_{it} R_t \frac{\xi_{i,t+1} c_{i,t+1}^{-\sigma}}{\xi_t \pi_{t+1}} \quad (\text{A.9})$$

$$\psi H_{it}^{\varphi} = c_{it}^{-\sigma} w_t \quad (\text{A.10})$$

where $\pi_{t+1} = \frac{P_{t+1}}{P_t}$ denotes inflation next period and $w_t = \frac{W_t}{P_t}$ is the real wage.

Firms. There is a continuum of household-owned monopolistically firms, $j \in [0, 1]$, who face the same economic problems and share the same beliefs about the future. However, like households, firms are not aware of their uniformity. Firms optimize with respect to price and

labor demand. The production technology of each firm is

$$y_{jt} = z_{jt} h_{jt}^\psi \quad (\text{A.11})$$

where z_{jt} and h_{jt} are the technology shock and labor demand, respectively; $\psi \in (0, 1]$ to guarantee that the production function has diminishing returns with respect to labor. The price optimization problem is subject to Calvo price stickiness: the price cannot be adjusted with some constant probability α . Each firm chooses the optimal price that will maximize the present value of current and expected future 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,

$$\mathcal{L}_{jt} = \min_{h_{jt}} w_t h_{jt} + mc_{jt}(y_{jt} - z_{jt} h_{jt}^\psi) \quad (\text{A.12})$$

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

$$mc_{jt} = \frac{w_t}{z_{jt}^\psi} \frac{1}{h_{jt}^{\psi-1}} \quad (\text{A.13})$$

Each firm maximizes the present value of current and expected future real profits with respect to price

$$\max_{P_{jt}^*} \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{j,t+h} \left(\frac{P_{jt}^*}{P_t} y_{j,t+h} - w_{t+h} h_{j,t+h} \right) \quad (\text{A.14})$$

where Q_{jt} is a generic stochastic discount factor of the j^{th} firm. From (A.11) and (A.13),

$$w_{t+h} h_{j,t+h} = \omega mc_{j,t+h} y_{j,t+h}$$

Substituting for y_{jt} and h_{jt} , using (A.11) and (A.13), the problem becomes

$$\max_{P_{jt}^*} \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{j,t+h} y_{j,t+h} \left(\frac{P_{jt}^*}{P_t} \left(\frac{P_{jt}^*}{P_{t+h}} \right)^{-\zeta_{t+h}} - \omega mc_{j,t+h} \left(\frac{P_{jt}^*}{P_{t+h}} \right)^{-\zeta_{t+h}} \right) \quad (\text{A.15})$$

Following a line of logic similar to Woodford (2003), the j^{th} firm understands from (A.13) that any deviation of its marginal cost mc_{jt} from the average mc_t should be somehow driven by deviations of its own optimal price and technology shock from the average price level and technology shock, respectively. In other words, it understands that

$$\frac{mc_{jt}}{mc_t} = f \left(\frac{P_{jt}^*}{P_t}, \frac{z_{jt}}{z_t} \right) \quad (\text{A.16})$$

where $f(\cdot)$ is decreasing in (P_{jt}^*/P_t) and (z_{jt}/z_t) . All else equal, when the j^{th} firm sets a higher price than the average, the demand for its good will lower, decreasing labor demand and marginal costs relative to the average. Deviations of the individual technology shock from the average will be irrelevant at the aggregate level since all firms are subject to the same shock.

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

$$\omega \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{t+h} \zeta_{t+h} \left(\frac{P_{jt}^*}{P_t} \right)^{-\zeta_{t+h}-1} \pi_{t,t+h}^{\zeta_{t+h}} mc_{j,t+h} y_{t+h} = \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{t+h} (\zeta_{t+h}-1) \left(\frac{P_{jt}^*}{P_t} \right)^{-\zeta_{t+h}} \pi_{t,t+h}^{\zeta_{t+h}-1} y_{t+h} \quad (\text{A.17})$$

where $\pi_{t,t+h} = \frac{P_{t+h}}{P_t}$.

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

$$\frac{R_t}{\bar{R}} = \left(\frac{R_{t-1}}{\bar{R}} \right)^{\rho_r} \left(\frac{\pi_t}{\bar{\pi}} \right)^{(1-\rho_r)\phi_\pi} \left(\frac{x_t}{\bar{x}} \right)^{(1-\rho_r)\phi_x} e^{\sigma_v \varepsilon_t^v}, \quad \varepsilon_t^v \sim \mathcal{N}(0, 1) \quad (\text{A.18})$$

where x_t is the output gap; $\bar{\pi}$ and \bar{x} denote the inflation target and output gap steady-state value, respectively; $\rho_r \in [0, 1)$.

Steady-states. I calculate steady-state values from (A.5) - (A.10), (A.17), (A.18):

$$\bar{v} = 1 \quad (\text{A.19})$$

$$\bar{\pi} = \beta \bar{R} = 1 \quad (\text{A.20})$$

$$\bar{w} = \psi(\bar{H})^\varphi (\bar{C})^\sigma \quad (\text{A.21})$$

$$\bar{d} = \bar{C} - \frac{1-\beta}{\beta} \bar{b} - \bar{w} \bar{H} \quad (\text{A.22})$$

$$\bar{y} = \bar{z} \bar{h}^\psi \quad (\text{A.23})$$

$$\bar{mc} = \frac{\bar{\zeta} - 1}{\bar{\zeta} \psi} \frac{\bar{P}^*}{\bar{P}} = \frac{\bar{\zeta} - 1}{\bar{\zeta} \psi} \quad (\text{A.24})$$

where $b_t = \frac{B_t}{P_t}$ and $d_t = \frac{D_t}{P_t}$ denote bond holding and dividends in real terms.

A.2 Log-linearized Model

Each variable with a hat sign on top, is log-linearized around its steady-state.

Households. Log-linearizing (A.9) around steady-states generates

$$\hat{c}_{it} = \tilde{\mathbb{E}}_{it}\hat{c}_{i,t+1} - \frac{1}{\sigma}\tilde{\mathbb{E}}_{it}(\hat{R}_t - \hat{\pi}_{t+1}) + \frac{1}{\sigma}\tilde{\mathbb{E}}_{it}(\hat{\xi}_{it} - \hat{\xi}_{i,t+1}) \quad (\text{A.25})$$

One can make inferences about $\tilde{\mathbb{E}}_{it}\hat{c}_{i,t+1}$ by iterating the linearized consumption rule above, i.e.,

$$\hat{c}_{i,t+1} = \tilde{\mathbb{E}}_{i,t+1}\hat{c}_{i,t+2} - \frac{1}{\sigma}\tilde{\mathbb{E}}_{i,t+1}(\hat{R}_{t+1} - \hat{\pi}_{t+2}) + \frac{1}{\sigma}\tilde{\mathbb{E}}_{i,t+1}(\hat{\xi}_{i,t+1} - \hat{\xi}_{i,t+2})$$

So,

$$\begin{aligned} \tilde{\mathbb{E}}_{it}\hat{c}_{i,t+1} &= \tilde{\mathbb{E}}_{it}\tilde{\mathbb{E}}_{i,t+1}\hat{c}_{i,t+2} - \frac{1}{\sigma}\tilde{\mathbb{E}}_{it}\tilde{\mathbb{E}}_{i,t+1}(\hat{R}_{t+1} - \hat{\pi}_{t+2}) - \frac{1}{\sigma}\tilde{\mathbb{E}}_{it}\tilde{\mathbb{E}}_{i,t+1}(\hat{\xi}_{i,t+1} - \hat{\xi}_{i,t+2}) \\ &= \tilde{\mathbb{E}}_{it}\hat{c}_{i,t+2} - \frac{1}{\sigma}\tilde{\mathbb{E}}_{it}(\hat{R}_{t+1} - \hat{\pi}_{t+2}) - \frac{1}{\sigma}\tilde{\mathbb{E}}_{it}(\hat{\xi}_{i,t+1} - \hat{\xi}_{i,t+2}) \end{aligned}$$

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

$$\hat{c}_{it} = \tilde{\mathbb{E}}_{it}\hat{c}_{i,t+2} - \frac{1}{\sigma}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{t+1}(\hat{R}_{t+h} - \hat{\pi}_{t+h+1}) + \frac{1}{\sigma}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{t+1}(\hat{\xi}_{i,t+h} - \hat{\xi}_{i,t+h+1})$$

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

$$\hat{c}_{it} = \tilde{\mathbb{E}}_{it}\hat{c}_{i,t+h} - \frac{1}{\sigma}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{t+k-1}(\hat{R}_{t+h} - \hat{\pi}_{t+h+1}) + \frac{1}{\sigma}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{t+k-1}(\hat{\xi}_{i,t+h} - \hat{\xi}_{i,t+h+1}) \quad (\text{A.26})$$

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 $\lim_{h \rightarrow \infty} \tilde{\mathbb{E}}_{it}\hat{c}_{i,t+h} = 0$. This would further imply that households would use the 1-step ahead Euler equation, as under RE. However, households have imperfect knowledge about the rest of the population, and one needs to combine (A.26) with the infinitely forward

iterated household's budget constraint in (A.5):

$$\begin{aligned}
B_{i,t-1} &= \frac{B_{it}}{R_{t-1}} - \frac{W_t H_{it}}{R_{t-1}} - \frac{D_{it}}{R_{t-1}} + \frac{P_t c_{it}}{R_{t-1}} \\
&= \tilde{\mathbb{E}}_{it} R R_{t-1,t} B_{i,t+1} - \tilde{\mathbb{E}}_{it} \sum_{h=0}^{t+1} R R_{t-1,t+h} (W_{t+h} H_{i,t+h} + D_{i,t+h}) + \tilde{\mathbb{E}}_{it} \sum_{h=0}^{t+1} R R_{t-1,t+h} P_{t+h} c_{i,t+h} \\
&= \dots \\
&= \lim_{\tau \rightarrow \infty} \tilde{\mathbb{E}}_{it} R R_{t-1,t+h} B_{i,t+h+1} - \tilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} R R_{t-1,t+h} (W_{t+h} H_{i,t+h} + D_{i,t+h}) + \tilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} R R_{t-1,t+h} P_{t+h} c_{i,t+h} \\
&= \tilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} F_{t-1,t+h} P_{t+h} c_{i,t+h} - \tilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} R R_{t-1,t+h} (W_{t+h} H_{i,t+h} + D_{i,t+h})
\end{aligned}$$

where $R R_{t-1,t+h} = \prod_{m=t-1}^{t+h} \frac{1}{R_m}$. To get the last equality I impose the appropriate no-Ponzi constraint, i.e., $\lim_{h \rightarrow \infty} \tilde{\mathbb{E}}_{it} R R_{t-1,t+h} B_{i,t+h+1} = 0$. To write everything in real terms, I divide by P_{t-1} and get

$$b_{i,t-1} = \tilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} R R_{t-1,t+h} \pi_{t-1,t+h} c_{i,t+h} - \tilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} R R_{t-1,t+h} \pi_{t-1,T} (w_{t+h} H_{i,t+h} + d_{i,t+h}) \quad (\text{A.27})$$

The log-linearized version of the iterated budget constraint is:

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

Using (A.20), $\bar{R} R_{t-1,t+h} \bar{\pi}_{t-1,t+h} = \frac{\bar{\pi}^{h+1}}{\bar{R}^{h+1}} = \beta^{h+1}$. Substituting for $\bar{R} R_{t-1,t+h} \bar{\pi}_{t-1,t+h}$ and optimal labor supply, the final log-linearized iterated budget constraint is

$$\begin{aligned}
\bar{b} \hat{b}_{i,t-1} &= \left(\bar{c} + \bar{w} \bar{H} \frac{\sigma}{\varphi} \right) \tilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} \beta^{h+1} \hat{c}_{i,t+h} - \bar{w} \bar{H} \frac{1+\varphi}{\varphi} \tilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} \beta^{h+1} \hat{w}_{t+h} - \bar{d} \tilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} \beta^{h+1} \hat{d}_{i,t+h} \\
&\quad + (\bar{c} - \bar{w} \bar{H} - \bar{d}) \tilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} \beta^{h+1} (\hat{R} R_{t-1,t+h} + \hat{\pi}_{t-1,t+h})
\end{aligned} \quad (\text{A.28})$$

From (A.26), one can isolate $\tilde{\mathbb{E}}_{it} \hat{c}_{i,t+h}$ and substitute for expectations of individual future con-

sumption into (A.28):

$$\begin{aligned}\bar{b}\hat{b}_{i,t-1} &= \frac{\beta(\bar{c} + \bar{w}\bar{H}\frac{\sigma}{\varphi})}{1-\beta}\hat{c}_{it} + \frac{\beta(\bar{c} + \frac{\sigma}{\varphi})}{\sigma(1-\beta)}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}((\hat{R}_{t+h} - \hat{\pi}_{i,\tau+1}) - (\hat{\xi}_{i,t+h} - \hat{\xi}_{i,\tau+1})) \\ &\quad - \bar{w}\bar{H}\frac{1+\varphi}{\varphi}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}\hat{w}_{t+h} - \bar{d}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}\hat{d}_{i,t+h} \\ &\quad + (\bar{c} - \bar{w}\bar{H} - \bar{d})\tilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^{h+1}(\hat{R}R_{t-1,t+h} + \hat{\pi}_{t-1,t+h})\end{aligned}$$

Isolating \hat{c}_{it} , one retrieves the individual demand as

$$\begin{aligned}\hat{c}_{it} &= \frac{\bar{b}(1-\beta)}{\beta(\bar{c} + \bar{w}\bar{H}\frac{\sigma}{\varphi})}\hat{b}_t + \frac{\bar{w}\bar{H}(1-\beta)(1+\varphi)}{\varphi(\bar{c} + \bar{w}\bar{H}\frac{\sigma}{\varphi})}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^h\hat{w}_{t+h} + \frac{\bar{d}(1-\beta)}{\bar{c} + \bar{w}\bar{H}\frac{\sigma}{\varphi}}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^h\hat{d}_{i,t+h} \\ &\quad - \frac{\beta}{\sigma}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^h(\hat{R}_{t+h} - \hat{\pi}_{t+h+1}) - \frac{\beta}{\sigma}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^h(\hat{\xi}_{i,t+h} - \hat{\xi}_{i,t+h+1}) \\ &\quad - \frac{(1-\beta)(\bar{c} - \bar{w}\bar{H} - \bar{d})}{\bar{c} + \bar{w}\bar{H}\frac{\sigma}{\varphi}}\tilde{\mathbb{E}}_{it}\sum_{h=0}^{\infty}\beta^h(\hat{R}R_{t-1,t+h} + \hat{\pi}_{t-1,t+h})\end{aligned}$$

Let $b_c = \frac{\bar{b}}{\beta(\bar{c} + \bar{w}\bar{H}\frac{\sigma}{\varphi})}$, $w_c = \frac{\bar{w}\bar{H}(1+\varphi)}{\varphi(\bar{c} + \bar{w}\bar{H}\frac{\sigma}{\varphi})}$, $d_c = \frac{\bar{d}}{\bar{c} + \bar{w}\bar{H}\frac{\sigma}{\varphi}}$ and $RR_c = \frac{(\bar{c} - \bar{w}\bar{H} - \bar{d})}{\bar{c} + \bar{w}\bar{H}\frac{\sigma}{\varphi}}$. Households are homogeneous, hence the demand of each household is described by

$$\hat{c}_t = b_c\hat{b}_{t-1} + (1-\beta)\tilde{\mathbb{E}}_t\sum_{t+h=t}^{\infty}\beta^h\hat{y}_{t+h} - \beta\tilde{\mathbb{E}}_t\sum_{t+h=t}^{\infty}\beta^h\hat{\Phi}_{t+h} \quad (\text{A.29})$$

where $\hat{y}_t = w_c\hat{w}_t + d_c\hat{d}_t$ and $\hat{\Phi}_t = \frac{1}{\sigma}((\hat{R}_t - \hat{\pi}_{t+1}) - (\hat{\xi}_t - \hat{\xi}_{t+1})) - RR_c(\hat{R}R_{t-1,t-1} + \hat{\pi}_{t-1,t})$. The household's labor supply and budget constraint are log-linearized into, respectively:

$$\varphi\hat{H}_t = -\sigma\hat{c}_t + \hat{w}_t \quad (\text{A.30})$$

$$\bar{R}\bar{b}(\hat{R}_{t-1} + \hat{b}_{t-1}) = \bar{b}(\hat{b}_t + \hat{\pi}_t) - \bar{w}\bar{H}(\hat{w}_t + \hat{H}_t) - \bar{d}\hat{d}_t + \bar{c}\hat{c}_t \quad (\text{A.31})$$

Firms. Log-linearizing firms' optimal condition for labor demand, we get,

$$\hat{m}c_{jt} = \hat{w}_t - \hat{z}_{jt} + (1-\psi)\hat{h}_{jt} \quad (\text{A.32})$$

Log-linearizing the firm's condition for the optimal price choice in (A.17),

$$\begin{aligned}
\omega \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h \left(\bar{Q} \bar{\zeta} \left(\frac{\bar{P}^*}{\bar{P}} \right)^{-\bar{\zeta}-1} \bar{\pi}_{t,t+h}^{\bar{\zeta}} \bar{m} c_{j,t+h} \bar{y} \right) & \left[\hat{Q}_{j,t+h} + \hat{\zeta}_{t+h} + \hat{m} c_{j,t+h} + \hat{y}_{j,t+h} + \bar{\zeta} \hat{\pi}_{t,t+h} \right. \\
& \left. - (\bar{\zeta} + 1)(\hat{P}_{jt}^* - \hat{P}_t) + \left(\ln(\bar{\pi}_{t,t+h}) \right. \right. \\
& \left. \left. - \ln \left(\frac{\bar{P}^*}{\bar{P}} \right) \right) \hat{\zeta}_{t+h} \right] = \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h \left(\bar{Q} \left(\frac{\bar{P}^*}{\bar{P}} \right)^{-\bar{\zeta}} \bar{\pi}_{t,t+h}^{\bar{\zeta}-1} \bar{y} \right) \left[(\bar{\zeta} - 1)(\hat{Q}_{j,t+h} + \hat{y}_{j,t+h}) + \bar{\zeta} \hat{\zeta} \right. \\
& \left. + (\bar{\zeta} - 1)^2 \hat{\pi}_{t,t+h} - \bar{\zeta}(\bar{\zeta} - 1)(\hat{P}_{jt}^* - \hat{P}_t) + \bar{\zeta}(\bar{\zeta} - 1) \left(\ln(\bar{\pi}_{t,t+h}) - \ln \left(\frac{\bar{P}^*}{\bar{P}} \right) \right) \hat{\zeta}_{t+h} \right] \quad (\text{A.33})
\end{aligned}$$

Using $\psi \bar{Q} \bar{\zeta} \left(\frac{\bar{P}^*}{\bar{P}} \right)^{-\bar{\zeta}-1} \bar{\pi}_{t,t+h}^{\bar{\zeta}} \bar{m} c_{j,t+h} \bar{y} = \bar{Q} \left(\frac{\bar{P}^*}{\bar{P}} \right)^{-\bar{\zeta}} \bar{\pi}_{t,t+h}^{\bar{\zeta}-1} \bar{y} (\bar{\zeta} - 1)$, $\bar{\pi} = 1$, $\left(\frac{\bar{P}^*}{\bar{P}} \right) = 1$ and rearranging terms,

$$\hat{P}_{jt}^* - \hat{P}_t = (1 - \alpha\beta) \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h \left(\hat{m} c_{j,t+h} + \hat{\pi}_{t,t+h} - \frac{1}{\bar{\zeta} - 1} \hat{\zeta}_{t+h} \right)$$

Log-linearizing (A.16) around the symmetric steady-state values,

$$\hat{m} c_{jt} - \hat{m} c_t = f_p(\hat{P}_{jt}^* - \hat{P}_t) + f_z(\hat{z}_{jt} - \hat{z}_t) \quad (\text{A.34})$$

with $f_p < 0$ is the partial derivative of $f(\cdot)$ with respect to (P_{jt}^*/P_t) and $f_z < 0$ is the partial derivative of $f(\cdot)$ with respect to (z_{jt}/z_t) , both evaluated at the steady-state. Note that f_p will be pinned down by market clearing conditions, whereas deviations of $m c_{jt}$ from the average due to technology differences do not matter at the aggregate level, since all firms are subject to the same shocks. Let $\mu_t = \frac{\zeta_t}{\bar{\zeta}-1}$ denote a mark-up shock, governed by

$$\ln \mu_t = (1 - \rho_\mu) \ln \bar{\mu} + \rho_\mu \ln \mu_{t-1} + \sigma_\mu \varepsilon_t^\mu, \quad \varepsilon_t^\mu \sim \mathcal{N}(0, 1)$$

Then, $\hat{\mu}_t = -\frac{1}{\bar{\zeta}-1} \hat{\zeta}_t$. Using (A.34) to substitute for $\hat{m} c_{jt}$ in firm's optimal price rule and that $P_{j,t+h}^* = P_{jt}^*$ for $h \geq 1$,

$$\begin{aligned}
\hat{P}_{jt}^* - \hat{P}_t &= (1 - \alpha\beta) \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h (\hat{m} c_{j,t+h} + \hat{\pi}_{t,t+h} + \hat{\mu}_{t+h}) \\
&= (1 - \alpha\beta) \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h \left(f_p(\hat{P}_{j,t+h}^* - \hat{P}_{t+h}) + \hat{m} c_{t+h} + \hat{\pi}_{t,t+h} - f_z(\hat{z}_{j,t+h} - \hat{z}_{t+h}) + \hat{\mu}_{t+h} \right) \\
&= (1 - \alpha\beta) \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h \left(f_p(\hat{P}_{jt}^* - \hat{P}_t - \hat{\pi}_{t,t+h}) + \hat{m} c_{t+h} + \hat{\pi}_{t,t+h} - f_z(\hat{z}_{j,t+h} - \hat{z}_{t+h}) + \hat{\mu}_{t+h} \right) \\
&= f_p(\hat{P}_{jt}^* - \hat{P}_t) + (1 - \alpha\beta) \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h (\hat{m} c_{t+h} + (1 - f_p) \hat{\pi}_{t,t+h} - f_z(\hat{z}_{j,t+h} - \hat{z}_{t+h}) + \hat{\mu}_{t+h})
\end{aligned}$$

Therefore,

$$\begin{aligned}\hat{P}_{jt}^* - \hat{P}_t &= (1 - \alpha\beta)\tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h \left(\frac{1}{1 - f_p} (\hat{m}c_{t+h} + \hat{\mu}_{t+h} - f_z(\hat{z}_{j,t+h} - \hat{z}_{t+h})) + \hat{\pi}_{t,t+h} \right) \\ &= \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h \left(\left(\frac{1 - \alpha\beta}{1 - f_p} \right) (\hat{m}c_{t+h} + \hat{\mu}_{t+h} - f_z(\hat{z}_{j,t+h} - \hat{z}_{t+h})) + \alpha\beta\hat{\pi}_{t+h+1} \right)\end{aligned}$$

where the second equality follows from

$$\tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h \hat{\pi}_{t,t+h} = \frac{\alpha\beta}{1 - \alpha\beta} \tilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h \hat{\pi}_{t+h+1}$$

Firms are subject to the same economic problem, beliefs and shocks, therefore, the final optimal pricing rule for each firm is

$$\hat{P}_t^* - \hat{P}_t = \tilde{\mathbb{E}}_t \sum_{h=0}^{\infty} (\alpha\beta)^h \left(\left(\frac{1 - \alpha\beta}{1 - f_p} \right) (\hat{m}c_{t+h} + \hat{\mu}_{t+h}) + \alpha\beta\hat{\pi}_{t+h+1} \right) \quad (\text{A.35})$$

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

$$\hat{R}_t = \rho_r \hat{R}_{t-1} + (1 - \rho_r)\phi_\pi \hat{\pi}_t + (1 - \rho_r)\phi_x \hat{x}_t + \sigma_v \varepsilon_t \quad (\text{A.36})$$

A.3 expectations formation process

The expectations formation process for each variable consists of a misspecified forecasting rule and belief shocks.

A.3.1 Myopia

Each household receives a belief shock about income and $\hat{\Phi}_t$, representing news about qualitative events in the future. The two shocks are i.i.d. and uncorrelated with each other or other shocks in the economy.

$$\begin{bmatrix} \varrho_{\Phi_t} \\ \varrho_{y_t} \end{bmatrix} \sim N(\mathbf{0}_{2,1}, \Sigma_{\varrho_{hh}}^2)$$

where $\Sigma_{\varrho_{hh}}$ is a diagonal matrix with diagonal entries $\sigma_{\varrho_{\Phi}}$ and σ_{ϱ_y} . Conditional on being sensitive to uncertainty induced by ϱ_{Φ_t} , each household's expectation about $\hat{\Phi}_{t+h}$, for $h \geq 1$, becomes a linear combination of ϱ_{Φ_t} and the macroeconomic forecast of $\hat{\Phi}_{t+h}$, $\tilde{\mathbb{E}}_t^* \hat{\Phi}_{t+h}$, as shown below

$$\tilde{\mathbb{E}}_t \hat{\Phi}_{t+h} = \textcolor{red}{n}^{h-1} \left(\tilde{\mathbb{E}}_t^* \hat{\Phi}_{t+h} + \underbrace{(1 - \textcolor{red}{n})}_{\text{sensitivity to } \varrho_{\Phi_t}} \varrho_{\Phi_t} \right) \quad (\text{A.37})$$

where $n = \frac{1}{1+s_\Phi\sigma_{\varrho_\Phi}^2} \in [0, 1]$ with $s_\Phi \geq 0$ controlling sensitivity to uncertainty induced by ϱ_Φ . Therefore, n guides the degree of myopic adjustment. For simplicity purposes only, I assume that the myopic adjustment across all belief shocks and agents is the same, i.e., the myopic adjustment is governed by the same n . This assumption can be easily relaxed. ‘

Similarly, subject to ϱ_{y_t} , each household’s expectations about income in period $h \geq 1$ is adjusted as follows:

$$\tilde{\mathbb{E}}_t \hat{y}_{t+h} = \textcolor{red}{n}^{h-1} \left(\frac{1 - \beta \textcolor{red}{n}}{1 - \beta} \right) \left(\tilde{\mathbb{E}}_t^* \hat{y}_{t+h} + \underbrace{(1 - \textcolor{red}{n})}_{\text{sensitivity to } \varrho_{y_t}} \varrho_{y_t} \right) \quad (\text{A.38})$$

In (A.29) each expected stream of income is multiplied by $(1 - \beta)$, which should be interpreted as the “weight” each of the income streams carries in forming the current optimal consumption decision, before the news shock was revealed. Now the household adjusts the “weight” to $(1 - \beta \textcolor{red}{n}) \geq (1 - \beta)$ as well, and the fraction $\frac{1 - \beta \textcolor{red}{n}}{1 - \beta}$ the “after-adjustment-weight” each income stream carries relative to non-adjusting.

Regarding firms expectations myopic adjustment, let $\hat{\Upsilon}_t = \omega \hat{m} c_t + \omega \hat{\mu}_t + \alpha \beta \hat{\pi}_{t+1}$. Each firm receives a belief shock, $\varrho_{p_t} \sim \mathcal{N}(0, \sigma_{\varrho_p}^2)$, judged to have significant influence on future realizations of $\hat{\Upsilon}_{t+h}$ for $h \geq 1$. The judgmental adjustment process to expectations about the future is a combination between ϱ_{p_t} and forecasts about future $\hat{\Upsilon}_{t+h}$, namely $\tilde{\mathbb{E}}_t^* \hat{\Upsilon}_{t+h}$, as described below

$$\tilde{\mathbb{E}}_t \hat{\Upsilon}_{t+h} = \textcolor{red}{n}^{h-1} \left(\frac{1 - \alpha \textcolor{red}{n}}{1 - \alpha} \right) \left(\tilde{\mathbb{E}}_t^* \hat{\Upsilon}_{t+h} + \underbrace{(1 - \textcolor{red}{n})}_{\text{sensitivity to } \varrho_{p_t}} \varrho_{p_t} \right) \quad (\text{A.39})$$

period, relative to the actual probability of resetting the optimal price next period.

A.3.2 forecasts

Agents are assumed to know the true processes of shocks, therefore they correctly forecast future exogenous disturbances. On the other hand, households and firms learn to use the same stationary autoregressive process to forecast future endogenous variables, such as output gap, inflation and nominal interest rates, nested in vector Z_t :

$$Z_t = \boldsymbol{\delta}_{t-1} + \boldsymbol{\gamma}_{t-1}(Z_{t-1} - \boldsymbol{\delta}_{t-1}) + \boldsymbol{\epsilon}_t \quad (\text{A.40})$$

where $\boldsymbol{\delta}_{t-1}$ is the mean of $Z_{0:t-1}$ series, $\boldsymbol{\gamma}_{t-1}$ represents the first-order correlation between $Z_{0:t-2}$ and $Z_{1:t-1}$ series, and $\boldsymbol{\epsilon}_t \sim WN$. The value of $\boldsymbol{\epsilon}_t$ is unknown when agents form expectations in

period t . For (A.40) to be stationary, the eigenvalues of γ_{t-1} have to always be within the unit circle. As soon as households and firms observe a new data point, they update the coefficients of (A.40) as follows

$$\begin{aligned}\delta_t &= \delta_{t-1} + \iota(Z_t - \delta_{t-1}) \\ \gamma_t &= \gamma_{t-1} + \iota((Z_t - \delta_{t-1})(Z_{t-1} - \delta_{t-1})' - \gamma_{t-1}(Z_t - \delta_{t-1})(Z_t - \delta_{t-1})') \eta_t^{-1} \\ \eta_t &= \eta_{t-1} + \iota((Z_t - \delta_{t-1})(Z_t - \delta_{t-1})' - \eta_{t-1})\end{aligned}\tag{A.41}$$

where η_t is the second moment vector/matrix. The forecast of future realization of Z_{t+h} , for $h \geq 1$, conditional on Z_{t-1} is

$$\tilde{\mathbb{E}}_t^* Z_{t+h} = \delta_{t-1} + \gamma_{t-1}^{h+1}(Z_{t-1} - \delta_{t-1})\tag{A.42}$$

where $\tilde{\mathbb{E}}_t^* Z_{t+h} := \tilde{\mathbb{E}}^*(Z_{t+h} \mid Z_{t-1})$ since ϵ_t in (A.40) is unknown when agent form expectations in period t .

A.4 Aggregate Economy

The final optimal consumption and price rules after incorporating (??) - (??) into (??) - (??) are, respectively

$$\begin{aligned}\hat{c}_t &= b_c \hat{b}_{t-1} + (1 - \beta) \hat{y}_t + \beta \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\beta \mathbf{n})^h \left((1 - \beta \mathbf{n}) \hat{y}_{t+h+1} - \frac{1}{\sigma} (\hat{R}_{t+h} - \hat{\pi}_{t+h+1} - (\hat{\xi}_{t+h} - \hat{\xi}_{t+h+1})) \right) \\ &+ \frac{\beta(1 - \mathbf{n})}{1 - \beta \mathbf{n}} \varrho_{c_t}\end{aligned}\tag{A.43}$$

$$\begin{aligned}\hat{P}_t^* - \hat{P}_t &= \left(\frac{1 - \alpha \mathbf{n}}{1 - \alpha} \right) \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\alpha \beta \mathbf{n})^h \left(\left(\frac{1 - \alpha \beta}{1 - f_p} \right) (\hat{m} c_{t+h} + \hat{\mu}_{t+h}) + \alpha \beta \hat{\pi}_{t+h+1} \right) \\ &+ \frac{(1 - \mathbf{n})(1 - \alpha \mathbf{n})}{(1 - \alpha)(1 - \alpha \beta \mathbf{n})} \varrho_{p_t}\end{aligned}\tag{A.44}$$

where $\varrho_{c_t} = (1 - \beta \mathbf{n}) \varrho_{y_t} - \varrho_{\Phi_t}$. Market clearing conditions imply that $\hat{y}_t = \hat{c}_t$, $\hat{h}_t = \hat{H}_t$ and $\hat{b}_t = 0$. Additionally, due to Calvo price stickiness, (A.2) becomes $P_t^{1-\zeta_t} = (1 - \alpha)(P_t^*)^{1-\zeta_t} + \alpha P_{t-1}^{1-\zeta_t}$. Dividing by $P_t^{1-\zeta_t}$ on both sides and log-linearizing,

$$\hat{P}_t^* - \hat{P}_t = \frac{\alpha}{1 - \alpha} \hat{\pi}_t\tag{A.45}$$

Integrating over the individual household and firm optimal price rule in (A.43) - (A.44) and imposing market clearing conditions, we get the aggregate demand and supply of the model

$$\hat{y}_t = \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\beta \mathbf{n})^h \left((1 - \beta \mathbf{n}) \hat{y}_{t+h+1} - \frac{1}{\sigma} (\hat{R}_{t+h} - \hat{\pi}_{t+h+1} - (\hat{\xi}_{t+h} - \hat{\xi}_{t+h+1})) \right) + \frac{1 - \mathbf{n}}{1 - \beta \mathbf{n}} \varrho_{c_t} \quad (\text{A.46})$$

$$\hat{\pi}_t = \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\alpha \beta \mathbf{n})^h (\omega (\hat{m}c_{t+h} + \hat{\mu}_{t+h}) + \beta (1 - \alpha \mathbf{n}) \hat{\pi}_{t+h+1}) + \frac{(1 - \mathbf{n})(1 - \alpha \mathbf{n})}{\alpha (1 - \alpha \beta \mathbf{n})} \varrho_{p_t} \quad (\text{A.47})$$

where $\omega = \frac{(1 - \alpha \beta \mathbf{n})(1 - \alpha \beta)}{\alpha (1 - f_p)}$. To define f_p , recall that it is equal to $\frac{\partial(\hat{m}c_{j_t} - \hat{m}c_t)}{\partial(\hat{P}_{j_t}^* - \hat{P}_t)}$. Recall that $\hat{m}c_{j_t} = \hat{w}_t - \hat{z}_{j_t} + (1 - \psi) \hat{h}_{j_t}$. On the other hand, the labor supply condition of the household implies $\hat{w}_t = \varphi \hat{h}_t + \sigma \hat{y}_t$, so in the labor market equilibrium, $\hat{w}_t = \varphi \hat{h}_{j_t} + \sigma \hat{y}_t$ should hold for the j^{th} firm. Moreover, $\hat{h}_{j_t} = \frac{1}{\psi} \hat{y}_{j_t} - \frac{1}{\psi} \hat{z}_{j_t}$. Therefore, $\hat{m}c_{j_t} = \sigma \hat{y}_t + \frac{\varphi + 1 - \psi}{\psi} \hat{y}_{j_t} - \frac{1 + \varphi}{\psi} \hat{z}_{j_t}$, and

$$\hat{m}c_{j_t} - \hat{m}c_t = \frac{\varphi + 1 - \psi}{\psi} (\hat{y}_{j_t} - \hat{y}_t) - \frac{1 + \varphi}{\psi} (\hat{z}_{j_t} - \hat{z}_t) = -\bar{\zeta} \frac{\varphi + 1 - \psi}{\psi} (\hat{P}_{j_t} - \hat{P}_t) - \frac{1 + \varphi}{\psi} (\hat{z}_{j_t} - \hat{z}_t)$$

$$\text{So, } f_p = \frac{\partial(\hat{m}c_{j_t} - \hat{m}c_t)}{\partial(\hat{P}_{j_t}^* - \hat{P}_t)} = -\bar{\zeta} \frac{\varphi + 1 - \psi}{\psi}.$$

The output gap is defined as $\hat{x}_t = \hat{y}_t - \hat{y}_t^f$, where \hat{y}_t^f is the flexible price output in equilibrium. Household labor supply condition implies that $\hat{w}_t = \frac{\sigma \psi + \varphi}{\psi} \hat{y}_t^f - \frac{\varphi}{\psi} \hat{z}_t$. Furthermore, $\hat{m}c_t = \hat{w}_t - \hat{z}_t + \frac{1 - \psi}{\psi} (\hat{y}_t^f - \hat{z}_t)$. Thus,

$$\hat{y}_t^f = \frac{\psi}{\omega(\sigma - 1) + 1 + \varphi} \hat{m}c_t^f + \frac{1 + \varphi}{\psi(\sigma - 1) + 1 + \varphi} \hat{z}_t \quad (\text{A.48})$$

Under flexible prices and constant mark-up, the marginal cost in deviation from its steady-state becomes $\hat{m}c_t^f = 0$. Finally, output gap is

$$\hat{x}_t = \hat{y}_t - \hat{y}_t^f = \frac{\psi}{\psi(\sigma - 1) + 1 + \varphi} \hat{m}c_t \quad (\text{A.49})$$

Therefore, the aggregate demand in terms of output gap is

$$\hat{x}_t = \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\beta \mathbf{n})^h \left((1 - \beta \mathbf{n}) \hat{x}_{t+h+1} - \frac{1}{\sigma} (\hat{R}_{t+h} - \hat{\pi}_{t+h+1} - \hat{e}_{t+h}) \right) + (1 - \mathbf{n}) \varrho_{c_t} \quad (\text{A.50})$$

where $\hat{e}_t = \sigma \left((\hat{y}_{t+1}^f - \frac{1}{\sigma} \hat{\xi}_{t+1}) - (\hat{y}_t^f - \frac{1}{\sigma} \hat{\xi}_t) \right)$ is the natural rate of interest and it is assumed to follow an AR(1) process

$$\ln e_t = \rho_e \ln e_{t-1} + \sigma_e \varepsilon_t^e, \quad \varepsilon_t^e \sim \mathcal{N}(0, 1) \quad (\text{A.51})$$

Letting $\hat{u}_t = \omega \hat{\mu}_t$ be the cost-push shock, the aggregate supply becomes

$$\hat{\pi}_t = \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} (\alpha \beta \textcolor{red}{n})^h (\kappa \hat{x}_{t+h} + \hat{u}_{t+h} + \beta(1 - \alpha \textcolor{red}{n}) \hat{\pi}_{t+h+1}) + \frac{(1 - \textcolor{red}{n})(1 - \alpha \textcolor{red}{n})}{\alpha(1 - \alpha \beta \textcolor{red}{n})} \varrho_{p_t} \quad (\text{A.52})$$

where $\kappa = \omega(\sigma + \varphi)$. Since $\hat{\mu}_t$ is governed by an AR(1) process, it follows that

$$\hat{u}_t = \rho_u \hat{u}_{t-1} + \sigma_u \varepsilon_t^u \text{ with } \varepsilon_t^u \sim \mathcal{N}(0, 1) \quad (\text{A.53})$$

Moreover,

$$\tilde{\mathbb{E}}_t^* \hat{e}_{t+h} = \rho_e^h \hat{e}_t$$

$$\tilde{\mathbb{E}}_t^* \hat{u}_{t+h} = \rho_u^h \hat{u}_t$$

Finally, the aggregate economy model in matrix form is described by

$$A_0(\Theta)S_t = A_1(\Theta)S_{t-1} + \tilde{\mathbb{E}}_t^* \sum_{h=0}^{\infty} F^{\tau-t-1} A_2(\Theta)S_{t+h} + B(\Theta)\mathcal{E}_t \quad (\text{A.54})$$

where $S_t = [\hat{x}_t \ \hat{\pi}_t \ \hat{R}_t \ \hat{e}_t \ \hat{u}_t]'$; $\mathcal{E}_t = [\varepsilon_t^e \ \varepsilon_t^u \ \varepsilon_t^v \ \varrho_{c_t} \ \varrho_{p_t}]'$; $\Theta = \{\alpha, \beta, n, \sigma, \kappa, \phi_\pi, \phi_x, \rho_r, \rho_e, \rho_u, \sigma_e, \sigma_u, \sigma_v\}$. F is a zero matrix, with only the first two diagonal entries equal to βn and $\alpha \beta n$, respectively. Using results from the previous subsection, the perceived law of motion (PLM) in matrix form can be written as

$$S_t = \underbrace{\Delta_{t-1} + \Gamma_{t-1}(S_{t-1} - \Delta_{t-1})}_{\text{PLM for aggregate endo var's}} + \underbrace{H S_{t-1}}_{\text{PLM for the shocks}} + \tilde{\epsilon}_t \quad (\text{A.55})$$

where $\delta_t = [\delta_t' \ \mathbf{0}_{1 \times 2}]'$; $\Gamma_t = \begin{bmatrix} \gamma_t & \mathbf{0}_{3 \times 2} \\ \mathbf{0}_{2 \times 3} & \mathbf{0}_{2 \times 2} \end{bmatrix}$; H is a diagonal matrix with diagonal equal to $[\mathbf{0}_{1 \times 3} \ \rho_e \ \rho_u]'$; $\tilde{\epsilon}_t = [\epsilon_t' \ \sigma_e \varepsilon_t^e \ \sigma_u \varepsilon_t^u]'$. The forecast of the state vector $h \geq 1$ periods ahead is described by

$$\tilde{\mathbb{E}}_t^* S_{t+h} = \underbrace{\Delta_{t-1} + \Gamma_{t-1}^{\tau-t+1}(S_{t-1} - \Delta_{t-1})}_{\text{forecast of endo var's}} + \underbrace{H^h S_t}_{\text{forecast of shocks}} \quad (\text{A.56})$$

Plugging (A.56) into (A.54), we get the actual law of motion:

$$\tilde{A}_0(\Theta)S_t = \tilde{A}_1(\Theta)\Delta_{t-1} + \tilde{A}_2(\Theta, \Gamma_{t-1})S_{t-1} + B\mathcal{E}_t \quad (\text{A.57})$$

where

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

$$\tilde{A}_1 = \sum_{h=0}^{\infty} F^h A_2 - \left(\sum_{h=0}^{\infty} F^h A_2 \mathbf{\Gamma}_{t-1}^h \right) \mathbf{\Gamma}_{t-1}^2$$

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

The infinite sums are defined as,

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

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

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

The last equality uses the Kronecker product property that $(\mathbf{\Gamma}'_{t-1} \otimes F)(\mathbf{\Gamma}'_{t-1} \otimes F) = (\mathbf{\Gamma}'_{t-1})^2 \otimes F^2$. We have thus defined $\tilde{A}_0(\Theta)$, $\tilde{A}_1(\Theta)$ and $\tilde{A}_2(\Theta, \mathbf{\Gamma}_{t-1})$.

B Testable Implications

The myopic adjustment of the forecast in period t and $t - 1$, respectively, is

$$\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} = bn^{h-1} \tilde{\mathbb{E}}_t^* \hat{\pi}_{t+h} + b(1-n)n^{h-1} \varrho_{p_t} \quad (\text{B.1})$$

$$\tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} = bn^h \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} + b(1-n)n^h \varrho_{p_{t-1}} \quad (\text{B.2})$$

Let $\tilde{\mathbb{E}}_t^* \hat{\pi}_{t+h} = \hat{\pi}_{t+h} - v_{t,t+h}$, where $v_{t,t+h}$ is the forecasting error term when there is no myopia ($n = 1$). Subtracting equation (B.2) from (B.1) and setting $\tilde{\mathbb{E}}_t^* \hat{\pi}_{t+h} = \hat{\pi}_{t+h} - v_{t,t+h}$,

$$\begin{aligned} \tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} &= bn^{h-1}(\hat{\pi}_{t+h} - v_{t,t+h}) - bn^{h-1} \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} + bn^{h-1}(1-n)(\varrho_{p_t} - n\varrho_{p_{t-1}}) \\ &= bn^{h-1}(\hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h}) + bn^{h-1} \tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - bn^{h-1} v_{t,t+h} - bn^{h-1} \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} \\ &\quad + bn^{h-1}(1-n)(\varrho_{p_t} - n\varrho_{p_{t-1}}) - bn^{h-1}(\tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) \\ &= bn^{h-1}(\hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h}) + bn^{h-1}(\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) + bn^{h-1} \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} - bn^{h-1} \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} \\ &\quad - bn^{h-1} v_{t,t+h} + bn^{h-1}(1-n)(\varrho_{p_t} - n\varrho_{p_{t-1}}) \end{aligned}$$

Hence,

$$\begin{aligned}
\hat{\pi}_{t+h} - \tilde{\mathbb{E}}_t \hat{\pi}_{t+h} &= \frac{1 - bn^{h-1}}{bn^{h-1}} (\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) - (\tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} - n \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h}) - (1-n)(\varrho_{p_t} - n\varrho_{p_{t-1}}) + v_{t,t+h} \\
&= \frac{1 - bn^{h-1}}{bn^{h-1}} (\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) + n(1 - bn^{h-1}) \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} + \underbrace{v_{t,t+h} - (1-n)(\varrho_{p_t} - n(1 - bn^{h-1})\varrho_{p_{t-1}})}_{error_t} \\
&= K_h (\tilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \tilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) + n(1 - bn^{h-1}) \tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h} + error_t
\end{aligned}$$

where $\tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+h}$ denotes expectations when there is no myopia, hence by construction it is uncorrelated with ϱ_{p_t} or $\varrho_{p_{t-1}}$. Let $h = 3$. Then, when the forecast is a SCE operator, $\tilde{\mathbb{E}}_{t-1}^* \hat{\pi}_{t+3}$ depends on $\hat{\pi}_{t-2}$. Moreover, the actual law of motion for inflation when $n = 1$ is

$$\hat{\pi}_t = \underbrace{\frac{\kappa}{1 - \alpha\beta\rho}}_{a_0} \hat{m}c_t + \underbrace{\frac{\beta(1 - \alpha)}{1 - \alpha\beta\gamma^*}(\gamma^*)^2}_{a_1} \hat{\pi}_{t-1}$$

$$\begin{aligned}
v_{t,t+3} &= \hat{\pi}_{t+3} - \tilde{\mathbb{E}}_t^* \hat{\pi}_{t+3} \\
&= a_0 \hat{m}c_{t+3} + a_1 \hat{\pi}_{t+2} - (\gamma^*)^5 \hat{\pi}_{t-1} \\
&= a_0 \sum_{i=0}^5 \rho^{5-i} a_1^i \hat{m}c_{t-2} + a_1^5 \hat{\pi}_{t-3} - (\gamma^*)^5 \hat{\pi}_{t-1} + f(\varepsilon_{t-1}, \dots, \varepsilon_{t+3}) \\
&= \frac{1}{\rho} \sum_{i=0}^5 \rho^{5-i} a_1^i (\hat{\pi}_{t-1} - a_1 \hat{\pi}_{t-2}) + a_1^5 \hat{\pi}_{t-3} - (\gamma^*)^5 \hat{\pi}_{t-1} + f(\varepsilon_{t-1}, \dots, \varepsilon_{t+3}) \\
&= \mathcal{F}(\hat{\pi}_{t-3}, \hat{\pi}_{t-2}, \hat{\pi}_{t-1}, \varepsilon_{t-1}, \dots, \varepsilon_{t+3})
\end{aligned}$$

Then, $\mathbb{E}(v_{t,t+3} \hat{\pi}_{t-2})$ is a linear function of $\mathbb{E}(\hat{\pi}_{t-1} \hat{\pi}_{t-2})$, $\mathbb{E}(\hat{\pi}_{t-2}^2)$ and $\mathbb{E}(\hat{\pi}_{t-3} \hat{\pi}_{t-2})$. After taking care of the omitted variable bias, we are left with $\mathbb{E}(v_{t,t+3} \hat{\pi}_{t-2}) = -\frac{a_1}{\rho} \sum_{i=0}^5 \rho^{5-i} a_1^i \mathbb{E}(\hat{\pi}_{t-2}^2) \leq 0$. Hence the true estimate of $\zeta_{2,3}$ after adding $\hat{\pi}_{t-1}$ and $\hat{\pi}_{t-3}$ as regressors will be even larger than the OLS estimator.

C Estimation

C.1 Data

I use quarterly data from 1966 to 2018. All data is extracted from the FRED and described as follows

$$\begin{aligned}
y_t &= 100 \ln \left(\frac{GDPC1_t}{POP_{index,t}} \right) \\
y_t^{potential} &= 100 \ln \left(\frac{GDP POT_t}{POP_{index,t}} \right)
\end{aligned}$$

$$x_t^{obs} = y_t - y_t^{potential}$$

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

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

where

- *GDPC1* – Real GDP, Billions of Chained 2012 Dollars, Seasonally Adjusted Annual Rate.
- $POP_{index} = \frac{CNP160V}{CNP160V_{1992Q3}}$.
- *CNP160V* – Civilian non institutional population, thousands, 16 years and above.
- *GDPPOT* – Real potential GDP, Billions of Chained 2012 Dollars, as reported by the U.S. Congressional Budget Office.
- *GDPDEF* – GDP-Implicit Price Deflator, 2012 = 100, Seasonally Adjusted.
- *Funds* – Federal funds rate, daily figure averages in percentages.

C.2 Figures

C.2.1 Posterior distribution

Figures 10-13 exhibit posterior and prior distributions for estimated parameters for the SAC-learning specifications.

C.2.2 Convergence diagnostics

To plot convergence diagnostics in Figures 14-17, I compute the between blocks (\hat{B}) and within blocks variance (\hat{W}) as suggested in Brooks and Gelman (1998). Convergence is reached when \hat{B} converges to 0, i.e., when \hat{W} and $(\hat{W} + \hat{B})$ are very close to each other. Convergence is obtained for all parameters across all models.

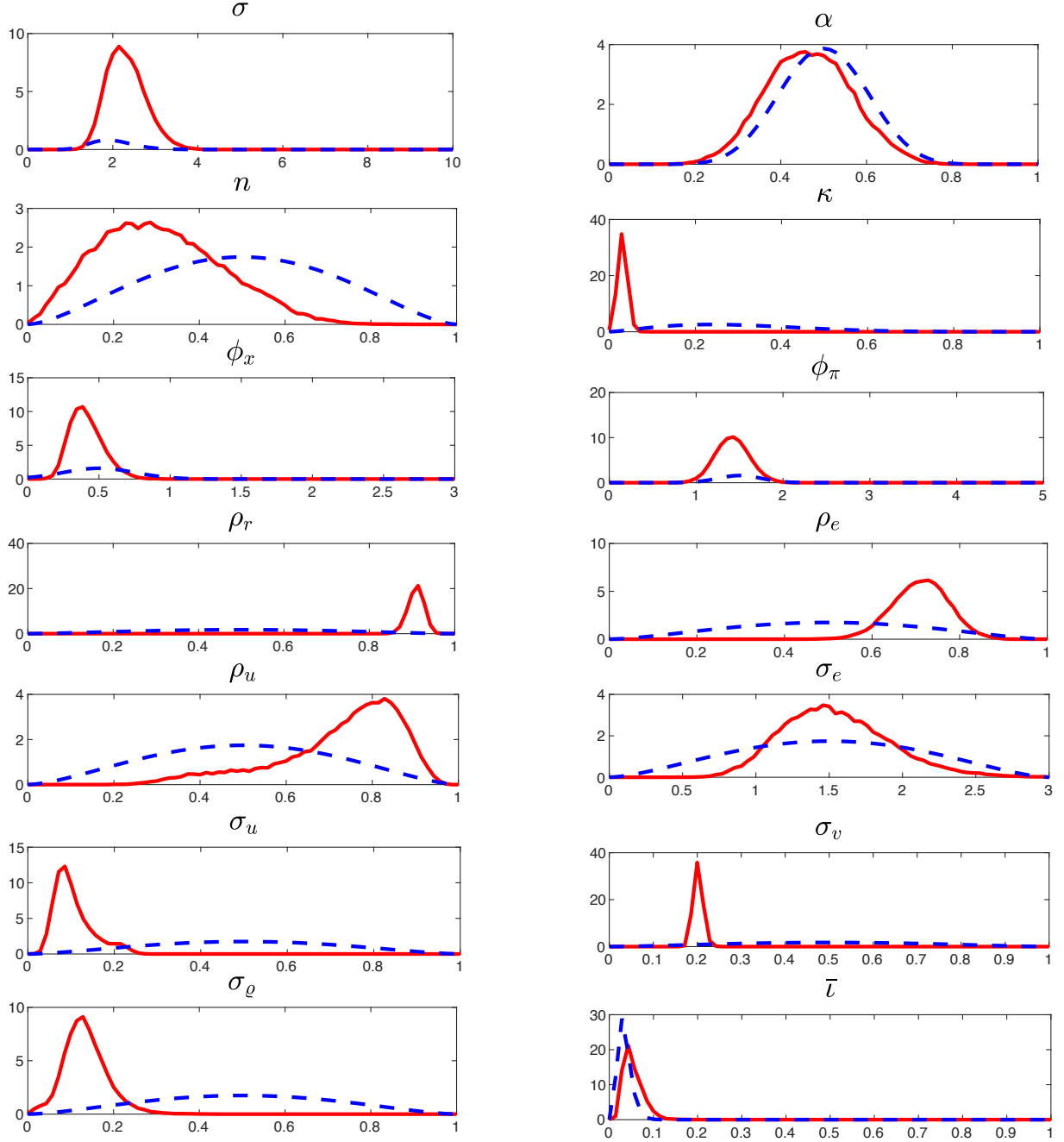


Figure 10: Posterior (red) and prior (dashed blue) distribution of the benchmark AR(1)-SAC-learning model.

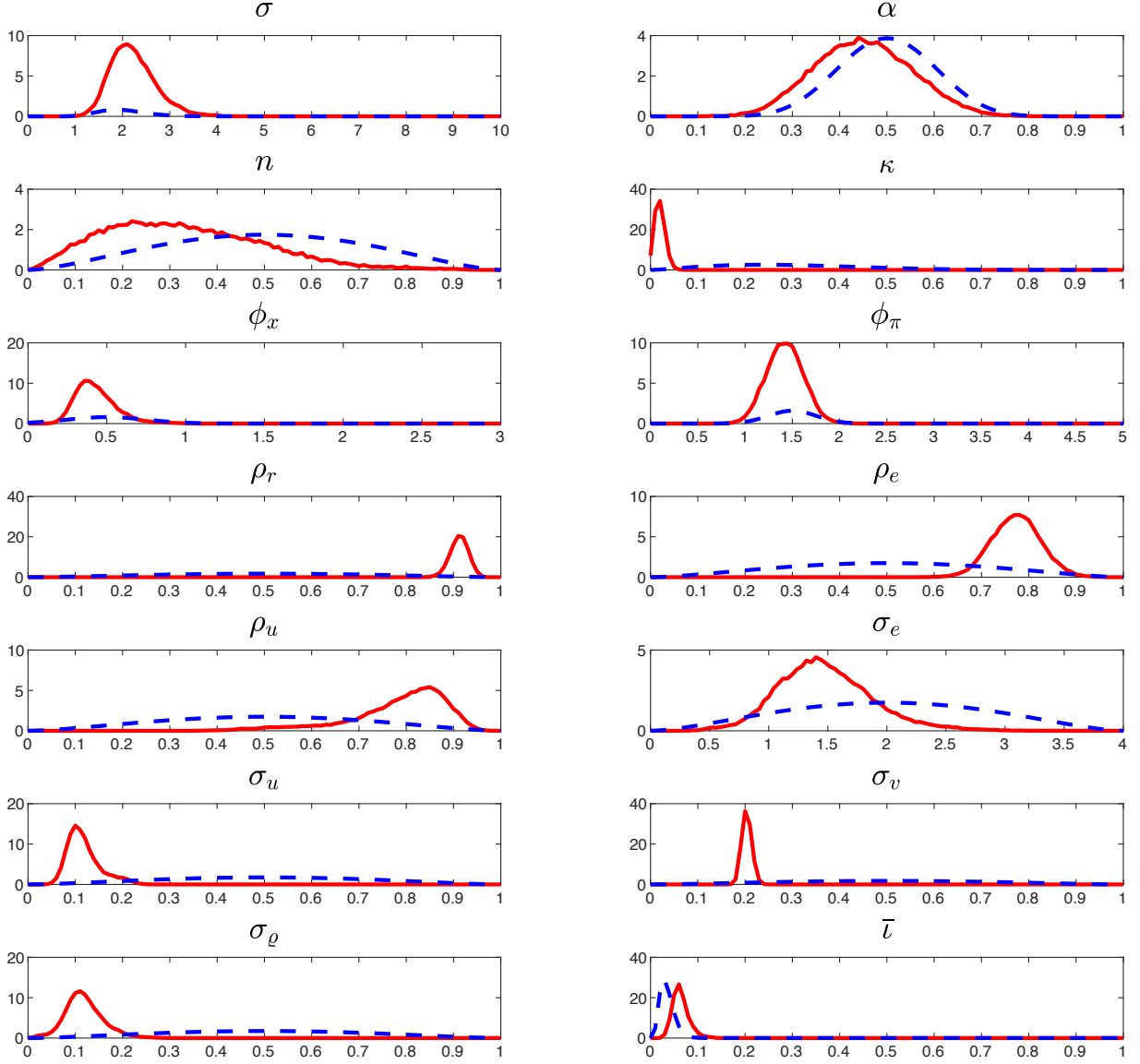


Figure 11: Posterior (red) and prior (dashed blue) distribution of the benchmark VAR(1)-SAC-learning model.

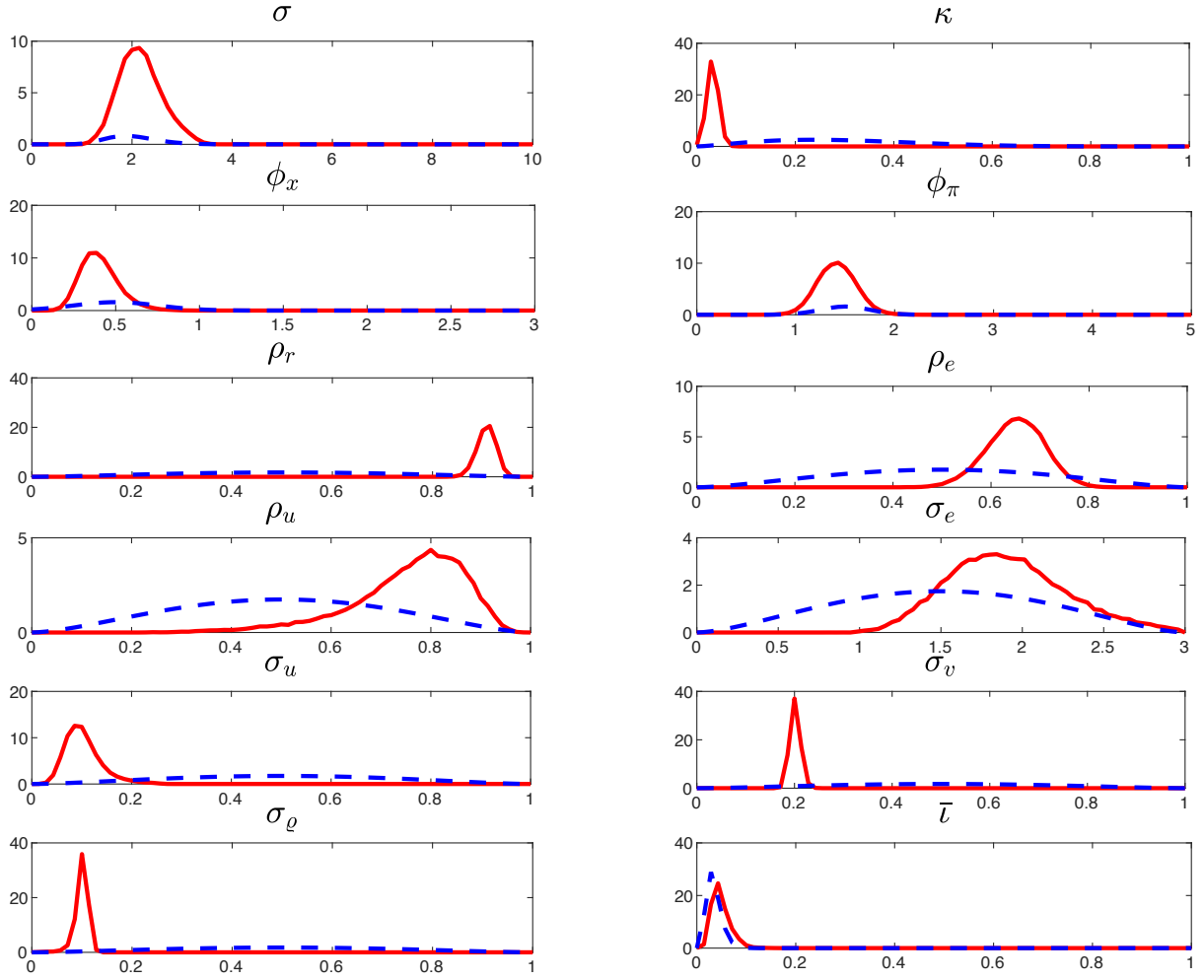


Figure 12: Posterior (red) and prior (dashed blue) distribution of the AR(1)-SAC-learning model with $n = 0$

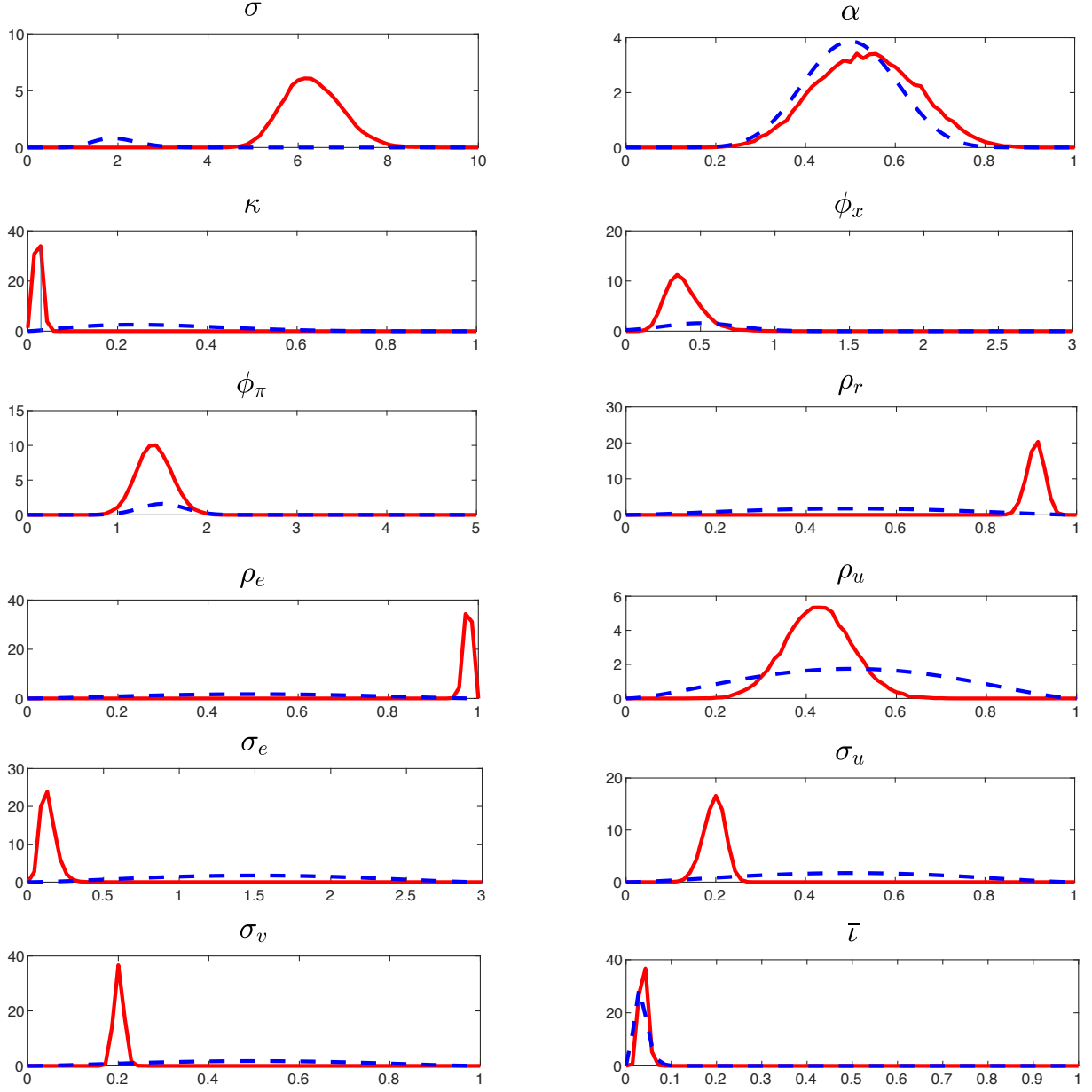


Figure 13: Posterior (red) and prior (dashed blue) distribution of the AR(1)-SAC-learning model with $n = 1$

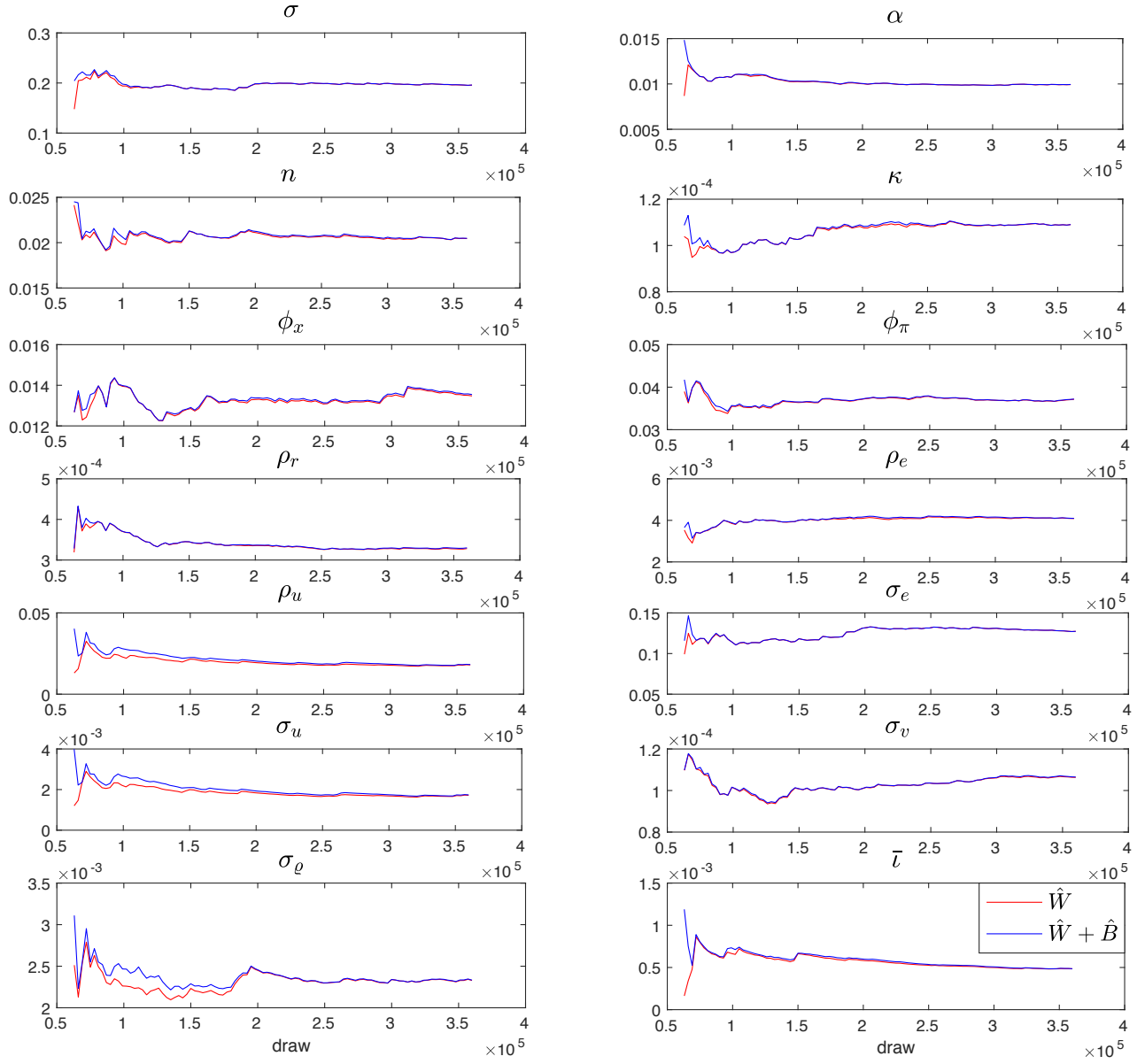


Figure 14: Convergence diagnostics of the benchmark AR(1)-SAC-learning model.

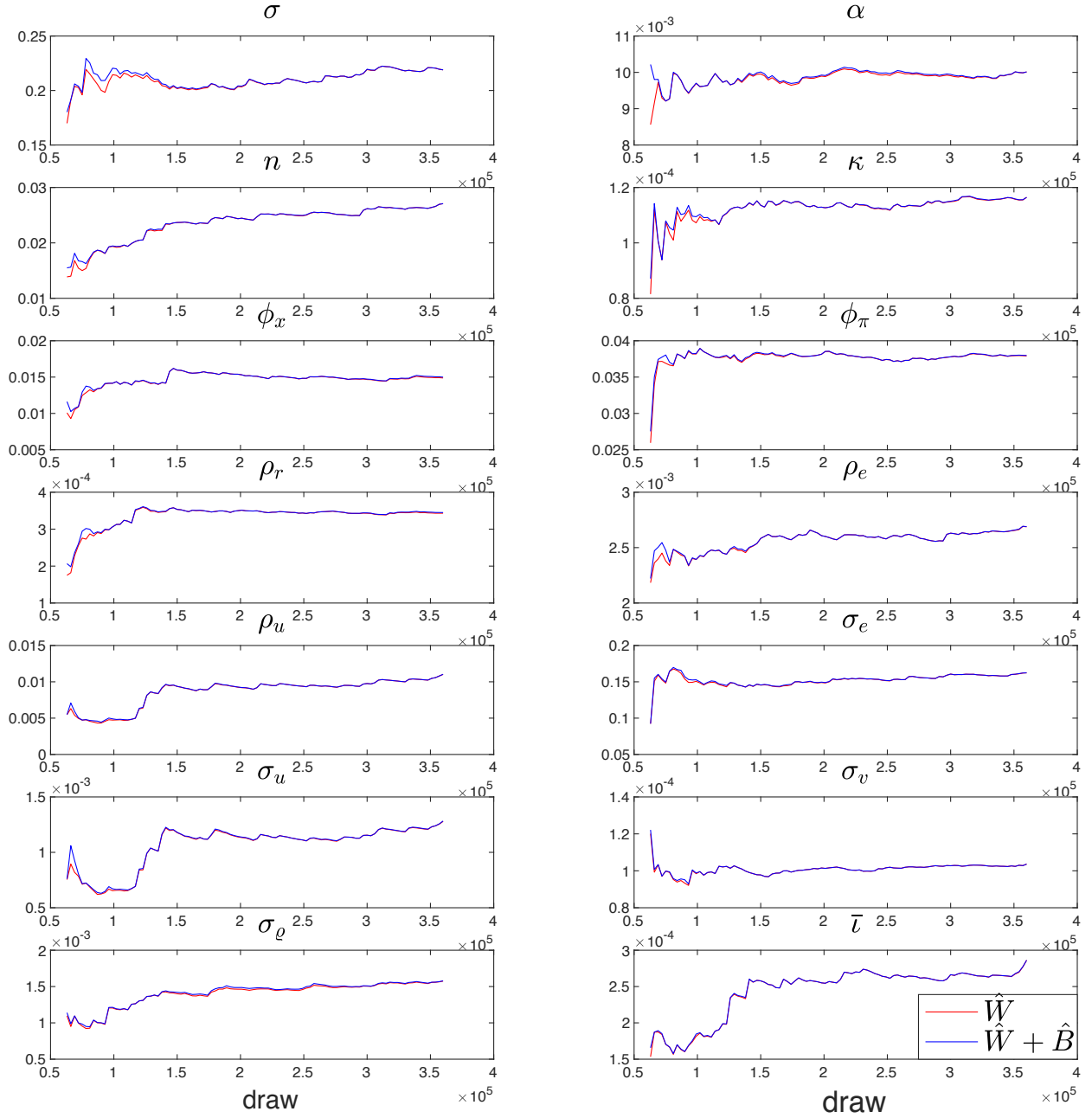


Figure 15: Convergence diagnostics of the benchmark VAR(1)-SAC-learning model.

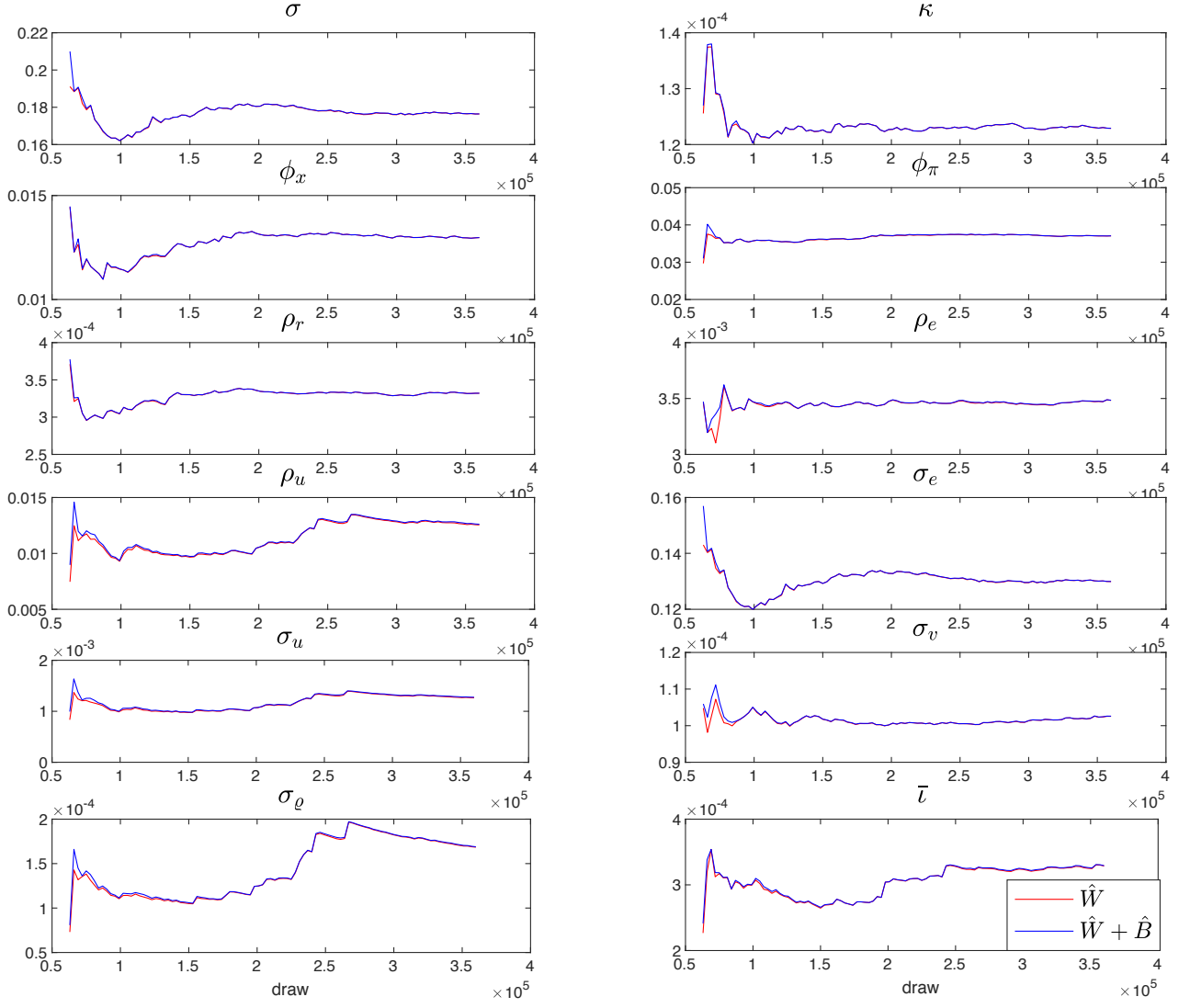


Figure 16: Convergence diagnostics of the AR(1)-SAC-learning model with $n = 0$.

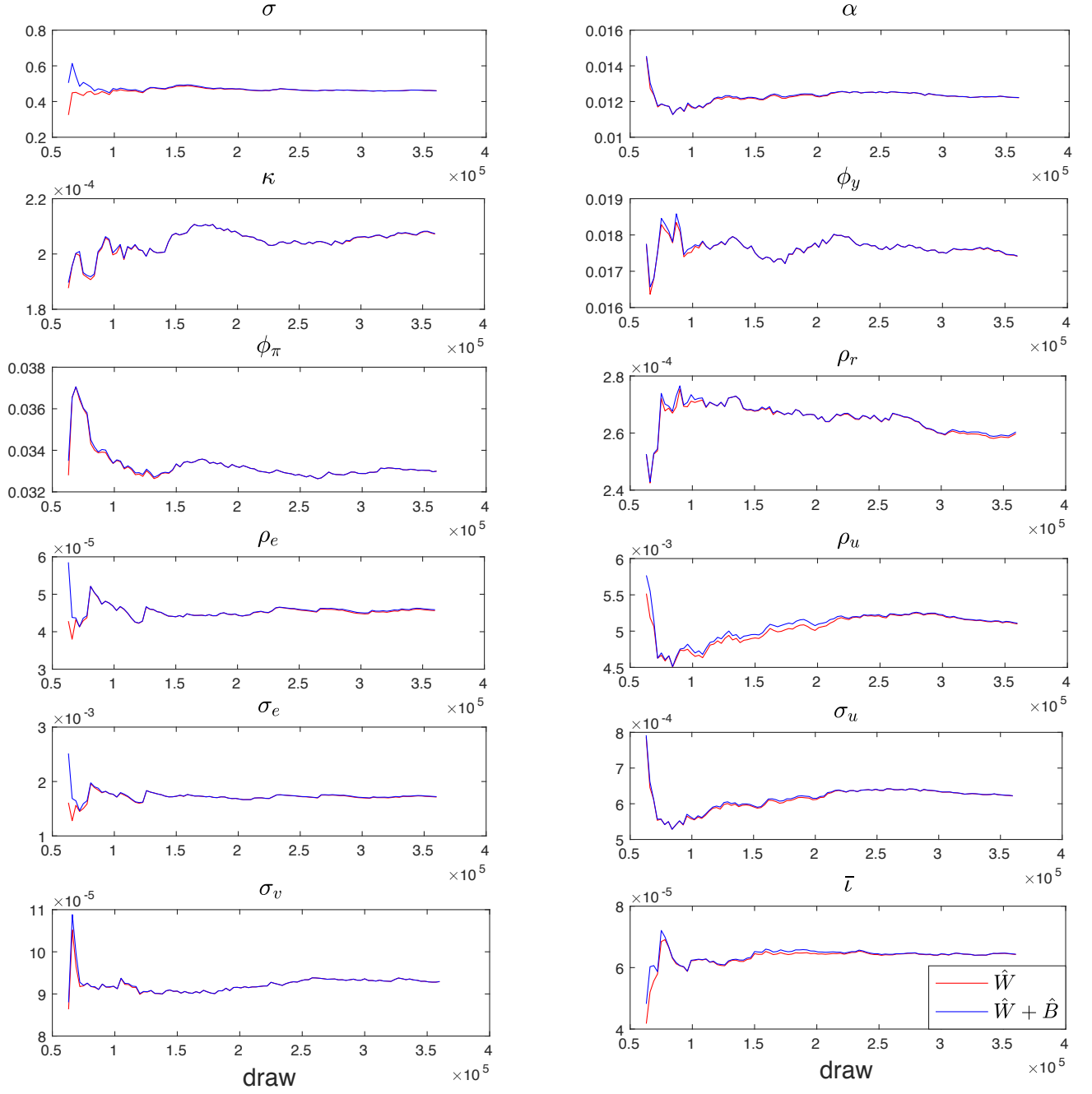


Figure 17: Convergence diagnostics of the AR(1)-SAC-learning model with $n = 1$.

C.2.3 More impulse response functions

Figure 18 visualizes the 3-dimensional response of the output gap and inflation to belief shocks, whereas 19 presents the 3-dimensional IRF, projected over the [response - time] plane over time for the benchmark SAC-learning model.

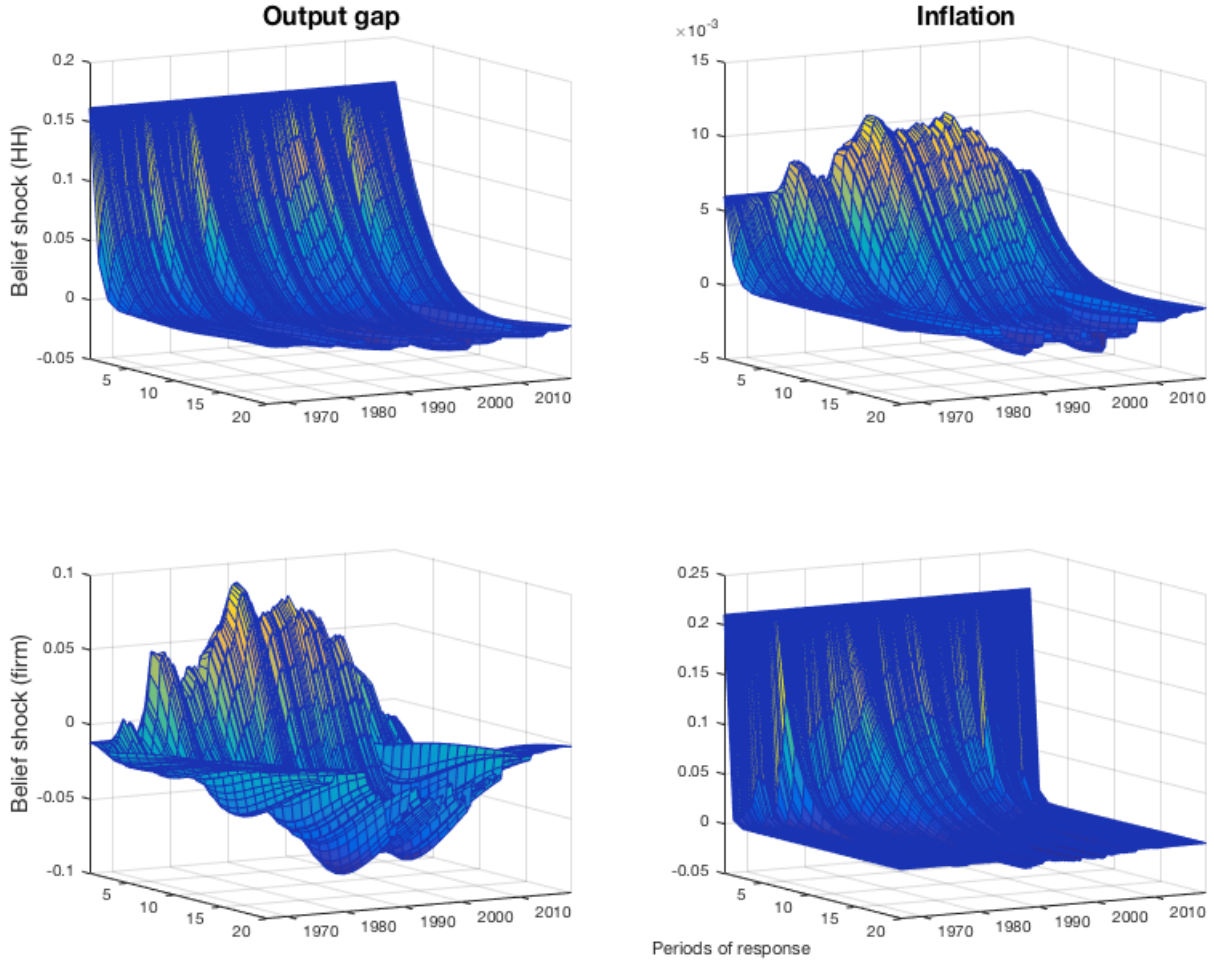


Figure 18: **3-Dimensional impulse response functions to a one standard deviation positive demand- and supply-side belief shock for the benchmark SAC-learning model.**

C.3 VAR(1)

In this subsection I present the historical evolution of beliefs about aggregates' moments when the forecasting rule has a VAR(1) structure under decreasing gain learning.

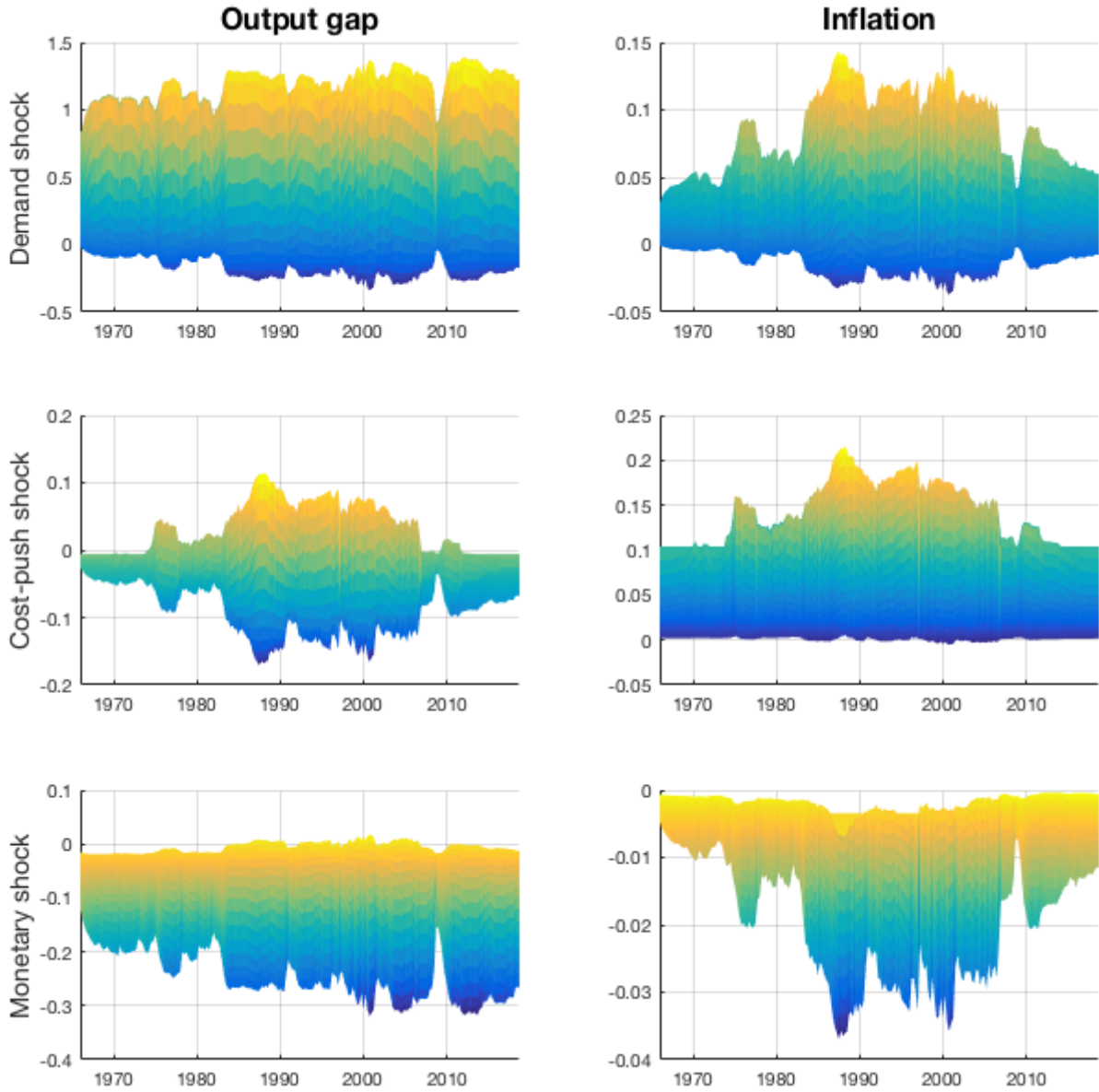


Figure 19: Impulse response functions to a one standard deviation demand, cost-push and monetary shock in the benchmark AR(1)-SAC-learning model.

References

- [1] Adam, K. (2007). Experimental Evidence of the Persistence of Output and Inflation. *The Economic Journal*, 117(520), 603-636.
- [2] Angeletos, G. M. and Z. Huo (2020). Myopia and Anchoring. Working paper.
- [3] Atkeson, A. and L. A. Ohanian (2001). Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1), 2-11.
- [4] Bianchi, F. (2013). Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics. *The Review of Economic Studies*, 80(2), 463-490.

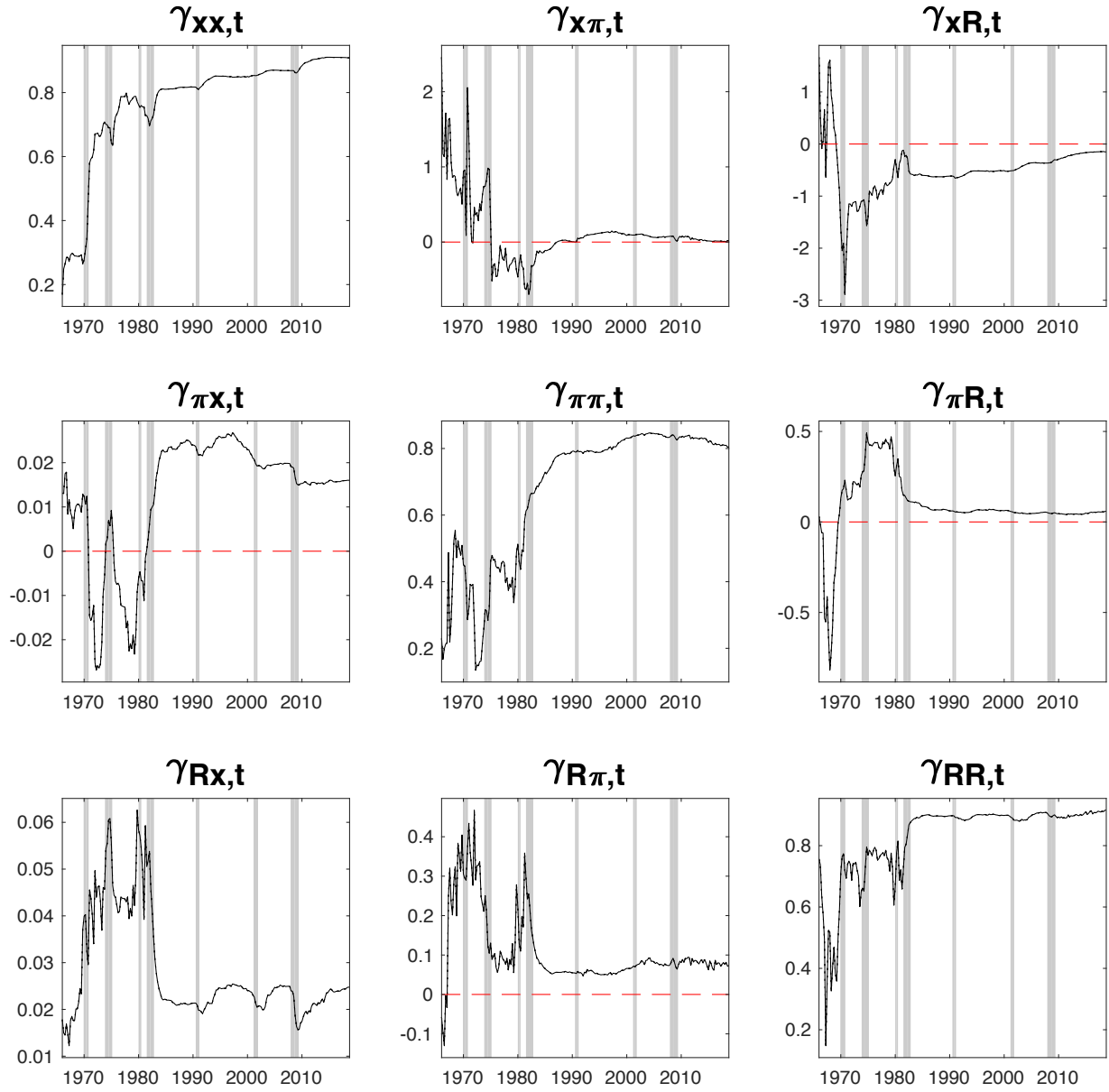


Figure 20: **Evolution of the VAR(1) forecast coefficients in the benchmark decreasing gain SAC-learning model.** Structural parameters set at their estimated posterior mode. Grey areas indicate recessionary periods as reported by the National Bureau of Economic Research.

- [5] Bills, M. and P. J. Klenow (2004). Some Evidence on the Importance of Sticky Prices. *Journal of Political Economy*, 112(5), 947-985.
- [6] Branch, W., G.W. Evans and B. McGough (2012). Finite Horizon Models. In *Macroeconomics at the Service of Public Policy*, ed. T.J. Sargent and J. Vilmunen, Oxford University Press.
- [7] Brooks, S.P. and A. Gelman (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434-455.
- [8] Bullard, J., G. W. Evans, S. Honkapohja (2008). Monetary Policy, Judgement, and Near-Rational Exuberance. *American Economic Review*, 98(3), 1163-77.
- [9] Bullard, J. and K. Mitra (2002). Learning About Monetary Policy Rules. *Journal of Monetary Economics*, 49(6), 1105-29.
- [10] Bunn, D. and G. Wright (1991). Interaction of Judgemental and Statistical Forecasting Methods: Issues and Analysis. *Management Science*, 37(5), 501-518.
- [11] Coble, K. and J. Lusk (2010). At the Nexus of Risk and Time Preferences: An Experimental Investigation. *Journal of Risk and Uncertainty*, 41, 67-79.
- [12] Coibion, O. and Y. Gorodnichenko (2015). Information Rigidity and the Expectations Formation Process: A simple Framework and New Facts. *American Economic Review*, 105(8), 2644-2678.
- [13] Cole S. J. and F. Milani (2017). The Misspecification of Expectations in New Keynesian Models: a DSGE-VAR approach, *Macroeconomic Dynamics*, 1-34.
- [14] Del Negro M. and S. Eusepi (2011). Fitting Observed Inflation Expectations. *Journal of Economic Dynamics and control*, 2011, 35(12), 2105-31.
- [15] Eroglu, C., and K. L. Croxton (2010). Biases in Judgmental Adjustments of Statistical Forecasts: The Role of Individual Differences. *International Journal of Forecasting*, 26, 116 -133.
- [16] Eusepi, S. and B. Preston (2018). The Science of Monetary Policy: An Imperfect Knowledge Oerspective. *Journal of Economic Literature*, 56(1), 3-59.
- [17] Evans, G.W. and S. Honkapohja (2001). Learning and Expectations in Macroeconomics. Princeton University Press, Princeton.
- [18] Evans, G.W. and S. Honkapohja (2009). Learning and Macroeconomics. *Annual Economic Review*, 1, 421-51.

- [19] Evans, G.W., C.H. Hommes, B. McGough, I. Salle (2019). Are Long-horizon Expectations (de-)Stabilizing? Theory and Experiments. *Bank of Canada*.
- [20] Fuster, A., D. Laibson, B. Mendel (2010). Natural Expectations and Macroeconomic Dynamics. *Journal of Economic Perspectives*, 24(4), 67-84.
- [21] Fuster, A., D. Laibson, B. Mendel (2012). Natural Expectations, Macroeconomic Dynamics and Asset Pricing. *NBER Macroeconomics Annual*, 26.
- [22] Gabaix, X. (2020). A Behavioral New Keynesian Model. *American Economic Review*, 110(8):2271-2327.
- [23] Ganong P. and P. J. Noel (2019). Consumer Spending During Unemployment: Positive and Normative Implications. *American Economic Review*, 109(7), 2383-2424.
- [24] Gaus, E. and C. G. Gibbs (2018). Expectations and the empirical fit of DSGE models. Forthcoming in *Studies in Nonlinear Dynamics and Econometrics*.
- [25] Geweke, J. (1999). Simulation Methods for Model Criticism and Robustness Analysis. In: Berger, J.O., Bernardo, J.M., David, A.P., Smith, A.F.M. (Eds.), *Bayesian Statistics*, 6, 275-299.
- [26] Greenwood, R. and A. Shleifer (2014). Expectations of Returns and Expected Returns. *Review of Financial Studies* 27(3), 714-46.
- [27] Herbst, E. and F. Schorfheide (2015). *Bayesian Estimation of DSGE Models*. Princeton University Press.
- [28] Hommes, C.H. and G. Sorger (1998). Consistent Expectations Equilibria. *Macroeconomic Dynamics*, 2(3), 287-321.
- [29] Hommes, C.H. (2013). *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*. Cambridge: Cambridge University Press.
- [30] Hommes, C.H. and M. Zhu (2014). Behavioral Learning Equilibria. *Journal of Economic Theory*, 150, 778-814.
- [31] Hommes, C.H. (2019). Behavioral and Experimental Macroeconomic and Policy Analysis: A Complex Systems Approach. Forthcoming in *Journal of Economic Literature*.
- [32] Hommes, C.H., K. Mavromatis, T. Ozden and M. Zhu (2019). Behavioral Learning Equilibria in the New Keynesian Model. De Nederlandsche Bank Working Paper, No. 654.

- [33] Kass, R. E. and A. E. Raftery (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773-795. University Press, 242-74.
- [34] Lawrence, M., Goodwin, P., O'Connor, M. and D. Onkal (2006). Judgmental Forecasting: A Review of Progress Over the Last 25 Years. *International Journal of Forecasting*, 22, 493 - 518.
- [35] Maćkowiak, B. and M. Wiederholt (2009). Optimal Sticky Prices under Rational Inattention. *American Economic Review*, 99(3), 769-803.
- [36] Malmendier, U. and S. Nagel (2016). Learning from Inflation Experiences. *Quarterly Journal of Economics*, 131(1), 53-87.
- [37] Mankiw, and Reis (2002). Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *Quarterly Journal of Economics*, 117(4), 1295-1328.
- [38] McNees, S. K. (1990). Man vs. Model? The Role of Judgment. *New England Economic Review*, July/August, 41 - 52.
- [39] Milani, F. (2005). A Bayesian DSGE Model with Infinite-Horizon Learning: Do “Mechanical” Sources of Persistence Become Superfluous? *International Journal of Central Banking*.
- [40] Milani, F. (2007). Expectations, Learning and Macroeconomic Persistence. *Journal of Monetary Economics*, 54(7), 2065-2082.
- [41] Milani, F. (2011). Expectation Shocks and Learning As Drivers of Business Cycle Fluctuations. *The Economic Journal*, 121(May), 379-401.
- [42] Moritz, B., E. Siemsen and M. Kremer (2014). Judgmental Forecasting: Cognitive Reflection and Decision Speed. *Production and Operations Management*, 23(7), 1146-1160.
- [43] Nimark, K. P. (2017). Dynamic Higher Order Expectations. *Cornell University mimeo*.
- [44] Noussair, C. and P. Wu (2010). Risk Tolerance in the Present and the Future: An Experimental Study. *Managerial and Decision Economics*, 27(6), 401-412.
- [45] Ormeño, A. and K. Molnár (2015). Using Survey Data of Inflation Expectations in the Estimation of Learning and Rational Expectations Models. *Journal of Money, Credit and Banking*, 47(4), 673-699.
- [46] Petersen, L. (2015). Do Expectations and Decisions Respond to Monetary Policy? *Journal of Economic Studies*, 42(6), 972-1004.

- [47] Petropoulos, F., N. Kourentzes, K. Nikolopoulos, E. Siemsen (2018). Judgmental Selection of Forecasting Models. *Journal of Operations Management*, 60, 34 - 46.
- [48] Preston, B. (2005). Learning About Monetary Policy Rules When Long-Horizon Expectations Matter. *International Journal of Central Banking*, 2, 81-126.
- [49] Reis, R. (2006). Inattentive Producers. *Review of Economic Studies*, 73(3), 793-821.
- [50] Rychalovska, Y. (2016). The Implications of Financial Frictions and Imperfect Knowledge in the Estimated DSGE Model of the US Economy. *Journal of Economic Dynamics and Control*, 73, 259-282.
- [51] Sims, C. A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics*, 50(3), 665-690.
- [52] Slobodyan, S. and R. Wouters (2012a). Learning in a Medium-scale DSGE Model with Expectations Based on Small Forecasting Models. *American Economic Journal: Macroeconomics*, 4(2), 65-101.
- [53] Slobodyan, S. and R. Wouters (2012b). Learning in an estimated medium-scale DSGE model. *Journal of Economic Dynamics and Control, Elsevier*, 36(1), 26-46.
- [54] Smets, F. and R. Wouters (2003). An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. *Journal of European Economic Association*, 1(5), 1123-1175.
- [55] Smets, F. and R. Wouters (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3), 586-606.
- [56] Stock J. H. and M. W. Watson (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit, and Banking*, 39, 3-34
- [57] Tversky, A. and D. Kahneman (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5(2), 207-232.
- [58] Tversky, A. and D. Kahneman (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-31.
- [59] Woodford, M. (2001). Imperfect Common Knowledge and the Effects of Monetary Policy. In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, ed. P. Aghion, R. Frydman, J. Stiglitz and M. Woodford, Princeton University Press, 25-58.
- [60] Woodford, M. (2003). Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton University Press.