Misspecified Forecasts and Myopia in an Estimated New Keynesian Model*

Ina Hajdini†

August 31, 2021

Abstract

The paper considers a New Keynesian framework 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), among others, 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; (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 that, consistently with recent evidence, deliver delayed over-reaction of forecasters; (iii) Frictions, such as habit in consumption, that are typically necessary for New Keynesian models with RE, are significantly less important in the estimated model with autoregressive forecasts and myopia because it 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 Enghin Atalay, Marco Del Negro, Keshav Dogra, Thomas Eisenbach, Xavier Gabaix, Bart Hobijn, Ryota Iijima, Lars Ljungqvist, Thomas Lubik, Karel Mertens, Tristan Potter, Benjamin Pugsley, Robert Rich, Raphael Schoenle, Ricardo Serrano-Padial, seminar participants at Aix-Marseille School of Economics, Copenhagen Business School, Cornivus University of Budapest, Duke Kunshan University, Federal Reserve Bank of Cleveland, Federal Reserve Bank of Dallas, Federal Reserve Bank of New York, Federal Reserve Board, Örebro University, Özyeğin University, Universitat Autonoma de Barcelona, University of Ottawa, and Vanguard for comments and suggestions. All errors are my own.

†Federal Reserve Bank of Cleveland. Email: Ina.Hajdini@clev.frb.org. The views expressed herein are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

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. 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 cognitive discounting in Gabaix (2020), where the private sector has a hard time understanding events that are far in the future. As agents try

[1]See 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.

[2]The literature has long shown that agents tend to focus mostly on recent observations, that is rely on perpetual or constant gain learning. For instance, Fuester 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.

to form expectations about the far future, they shrink their autoregressive forecasts toward the steady state of the economy. Consecutively, expectations about aggregate variables many periods from the present will shrink significantly more relative to a case of no myopia. As in Gabaix (2020), the private sector is globally patient with respect to steady-state variables, but is myopic with respect to deviations from the steady state. In addition to Gabaix (2020), I assume that as agents become less capable to predict future deviations from the steady state, the *shorter term* autoregressive forecasts become more volatile (more responsive to external stimuli) relative to when agents' vision of the future is not impaired (or equivalently when there is no myopia).

Once myopia is added to the original autoregressive forecasts, the parameters of the 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 asses the empirical relevance of the proposed expectations formation process independently from the details of the model, I first follow a strategy similar to Coibion and Gorodnichenko (2015). I derive a number of testable implications along the CE equilibrium path that differentiate the proposed expectations formation process from RE, well-specified forecasting rules and myopia, and misspecified forecasting rules without myopia.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.RE imply that forecasting errors cannot be predictable with information available at time t, whereas well-specified forecasting rules and myopia imply that forecasting errors are predictable by the ex-ante forecasting revisions and the first lag of the variable being forecasted. None of these two alternatives is supported by the empirical analysis on SPF data.

Angeletos et al. (2020) have presented a new fact regarding forecasters' behavior about inflation: there is initial under-reaction followed by over-shooting. The authors show that there should be over-extrapolation and sufficient noisy information for this fact to be replicated theoretically. The present paper derives a sufficient and necessary condition for which delayed over-reaction holds true under a combination of misspecified forecasts and myopia. Specifically, delayed over-shooting is guaranteed to occur if there is sufficiently high *endogenous* over-extrapolation, but in contrast to the requirement in Angeletos et al. (2020) it can also occur if there is *endogenous* under-extrapolation. The latter is due to the assumption that as agents become shorter-sighted, the volatility of shorter-term forecasts increases. On the other hand, misspecified forecasts absent of myopia would deliver late over-shooting only if there is over-extrapolation, whereas myopia alone implies under-reaction of forecasters at any point in time. Therefore, the proposed expectations formation process with misspecified forecasting rules and myopia is preferred over the aforementioned expectations alternatives.

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 with habit in consumption and inflation indexation, similar to the one in Milani (2006). Motivated by the work of Preston (2005) and Eusepi and Preston (2018), the implied optimal consumption and pricing rules of households and firms are of an infinite horizon nature and in the presence of misspecified forecasting rules, they cannot be reduced to the standard 1-step ahead Euler equation and Phillips curve, respectively. However, the paper shows that when agents use well-specified forecasting rules as in Gabaix (2020), the implied Euler equation and Phillips curve coincide with the ones in Gabaix (2020).

Bayesian estimation of the model on U.S. macroeconomic data from 1968:Q4 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 as a combination of autoregressive forecasts and myopia over the other aforementioned alternatives. Second, the best-fitting expectations formation process is characterized by a significant degree of myopia and simple AR(1) forecast rules. More elaborate VAR-based forecasts do not enhance the model's fit and do not provide any additional useful information to agents in terms of their beliefs. Third, compared to RE frictions such as habit in consumption, are significantly less important or even unnecessary in the presence of myopia combined with AR(1) forecasts, because such a combination strengthens the internal propagative features of the model: autoregressive forecasts induce excess persistence and myopia generates excess volatility. A combination of myopia and misspecified forecasts will often deliver impulse response functions to demand, cost-push and monetary shocks that are more persistent and volatile (especially for inflation), relative to a case of no myopia or RE. Finally, almost all the full model-implied impulse response functions of one-period ahead forecasting errors to demand, cost-push and monetary shocks are consistent with the empirical evidence of delayed over-shooting.

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.

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)).[3] 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 future deviations from the steady state. 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.).

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

[3]Experimental 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)).

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.[4] 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.[5]

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 have no sensitivity towards belief shocks, also known as infinite-horizon learning by Preston (2005).[6]

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.

[4] In 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.

[5] Angeletos 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 autoregressive misspecified forecasting rules due to imperfect common knowledge, whereas myopia is realized through a judgmental adjustment process to misspecified forecasting rules.

[6] See Eusepi and Preston (2018) as well for a review.

2 Misspecified Forecasts and Myopia

This section integrates a combination of misspecified autoregressive forecasting rules and myopia in a New Keynesian pricing problem and solves for the CE equilibrium path. This section builds the foundation for deriving testable implications for inflation forecasting data in the succeeding section. The rationale for focusing on the pricing problem instead of the household's is that since testing implications of various expectations assumptions on inflation forecasting data has become the benchmark in the literature, it will be easier to compare the present paper's expectations process with others in the literature.

2.1 New Keynesian Pricing

Following Woodford (2003), 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 is[7]

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

where $\hat{p}_t^* = log(P_t^*/P_t)$ is the log-linear optimal price in deviation from the aggregate price $\hat{P}_t$; $\widetilde{\mathbb{E}}_t$ is a generic subjective expectations operator that satisfies the law of iterative expectations and standard probability rules; $\hat{mc}_t$ is the marginal cost; $\hat{\pi}_t$ is inflation; $\hat{\mu}_t \sim iid\mathcal{N}(0, \sigma_\mu^2)$ is a cost-push shock; 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

$$\hat{mc}_t = \rho\hat{mc}_{t-1} + \sigma_\varepsilon\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0, 1) \tag{2}$$

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

[7] See Appendix A for more details.

cost.

2.2 Myopia

To model myopia, I build on the idea of cognitive discounting of Gabaix (2020), where the expected value of future inflation is discounted by a cognitive discount factor, or degree of myopia, $n \in (0, 1]$. In addition to Gabaix (2020), I assume that as firms become less capable to predict future deviations of inflation from the steady-state, shorter term discounted "original" forecasts become more volatile. This assumption will be captured by an upward reweighing factor $w_f(n) \geq 1$, such that $w_f(1) = 1$, that depends negatively on the degree of myopia n, but not on the forecast horizon. With $w_f > 1$, firms reduce forecasts to steady-state values at a delayed pace relative to a case of $w_f = 1$, and such a delay will be more pronounced for shorter-term forecasts.

Let $\widetilde{\mathbb{E}}_t^\star$ be the expectations operator absent of myopia. Then, the myopically adjusted forecast about future inflation, marginal cost, and cost-push shock in deviation from their steady-state values are, respectively,

$$\widetilde{\mathbb{E}}_t \begin{bmatrix} \hat{\pi}_{t+h} \\ \hat{mc}_{t+h} \\ \hat{\mu}_{t+h} \end{bmatrix} = w_f n^h \widetilde{\mathbb{E}}_t^\star \begin{bmatrix} \hat{\pi}_{t+h} \\ \hat{mc}_{t+h} \\ \hat{\mu}_{t+h} \end{bmatrix} \tag{3}$$

For the firms pricing problem, I assume that $w_f = \frac{1-\alpha n}{1-\alpha}$. When firms excess discount future deviations of inflation and marginal cost from their steady states, they make optimal pricing decisions *as if* they cannot reset the price with probability $\alpha n < \alpha$, or equivalently they behave *as if* the average duration of their current price is $\frac{1}{1-\alpha n} < \frac{1}{1-\alpha}$. Then, as firms become shorter-sighted due to excess discounting, the importance of each of the future periods will increase by a factor of $w_f = \frac{\text{average duration of current price w/o myopia}}{\text{average duration of current price w/ myopia}}$. So, $w_f = \frac{\frac{1}{1-\alpha}}{\frac{1}{1-\alpha n}} = \left(\frac{1-\alpha n}{1-\alpha}\right)$.

The optimal pricing decision in (1) can be written as,

$$\hat{p}_t^* = w_f \widetilde{\mathbb{E}}_t^\star \sum_{h=0}^{\infty} (\alpha\beta n)^h \left(\omega \hat{mc}_{t+h} + \beta n \hat{\pi}_{t+h+1} + \hat{\mu}_{t+h}\right) \tag{4}$$

2.3 Misspecified forecasts

Consider the expectations operator absent of myopia, $\widetilde{\mathbb{E}}_t^\star$. Firms are assumed to understand the process of the exogenous disturbances they are subject to; therefore absent of myopia, they correctly forecast the marginal cost and the cost-push shock, i.e.,

$$\widetilde{\mathbb{E}}_t^{\star} \hat{mc}_{t+h} = \rho^h \hat{mc}_t \tag{5}$$

$$\widetilde{\mathbb{E}}_t^{\star} \hat{\mu}_{t+h} = 0 \tag{6}$$

Firms are endowed with imperfect common knowledge, therefore each individual firm does not understand that i) every other firm in the economy faces the same optimal pricing rule as in (1), ii) has the same beliefs about future realizations of inflation and shocks, and iii) is subject to the same exogenous shocks. Leveraging on a large body of evidence showing that economic agents form forecasts based on simple autoregressive rules (see for instance, Adam (2007), Hommes and Zhu (2014), Malmendier and Nagel (2016), among others), I assume that the perceived law of motion for inflation follows an AR(1) process,

$$\hat{\pi}_t = \delta + \gamma(\hat{\pi}_{t-1} - \delta) + \epsilon_t \tag{7}$$

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

$$\widetilde{\mathbb{E}}_t^{\star} \hat{\pi}_{t+h} = \delta(1 - \gamma^{h+1}) + \gamma^{h+1} \hat{\pi}_{t-1} \tag{8}$$

As shown in section 2.4, parameters δ and γ will be pinned down using the solution concept of a Stochastic Consistent Expectations Equilibrium.

2.4 Consistent Expectations Equilibrium

Aggregating equation (4), the aggregate pricing scheme is described by

$$\hat{\pi}_t = \widetilde{\mathbb{E}}_t^{\star} \sum_{h=0}^{\infty} (\alpha\beta n)^h \left(\kappa \hat{mc}_{t+h} + \beta n (1 - \alpha n) \hat{\pi}_{t+h+1}\right) + \hat{u}_t \tag{9}$$

where $\kappa = \omega(1 - \alpha n)/\alpha$ and $\hat{u}_t = \frac{1-\alpha n}{\alpha} \hat{\mu}_t$.

<u>*Remark 1:*</u> Equation (9) does not reduce to the Phillips curve of the behavioral New Keynesian model as in Gabaix (2020), $\hat{\pi}_t = \kappa \hat{mc}_t + \beta M^f \widetilde{\mathbb{E}}_t^{\star} \hat{\pi}_{t+1} + \hat{u}_t$ with $M^f \leq 1$, *unless* the forecast ($\widetilde{\mathbb{E}}_t^{*}$) is a well-specified rule, i.e., the structure of the forecasting rule is the same as the minimum

state variable solution under RE.[8]

<u>Remark 2:</u> Equation (9) reduces to the standard Phillips curve, $\hat{\pi}_t = \kappa \hat{mc}_t + \beta \widetilde{\mathbb{E}}_t^{\star} \hat{\pi}_{t+1} + \hat{u}_t$, *only if* the forecast $(\widetilde{\mathbb{E}}_t^{*})$ is a well-specified rule *and* there is no myopia ($n = 1$).

When firms use model-based forecasting rules, i.e. $\widetilde{\mathbb{E}}$ is a well-specified expectations operator, they know they are all homogenous, hence can use their own optimal condition in (4) to form expectations about inflation and the infinite-horizon New Keynesian Phillips curve reduces to the 1-step ahead (behavioral or standard) Phillips curve. 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 (5) and (8) in (9) delivers the actual law of motion for inflation:

$$\hat{\pi}_t = \beta n \delta (1 - \alpha n) \left(\frac{1}{1 - \alpha\beta} - \frac{\gamma^2}{1 - \alpha\beta n \gamma} \right) + \frac{\kappa}{1 - \alpha\beta\rho n} \hat{mc}_t + \frac{\beta n (1 - \alpha n)\gamma^2}{1 - \alpha\beta n\gamma} \hat{\pi}_{t-1} + \hat{u}_t \tag{10}$$

Firms believe (7) is a valid perceived law of motion for inflation if and 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:*

[8] Specifically, let $\widetilde{\mathbb{E}}_t^{\star} = \mathbb{E}_t$ be the RE operator. Then,

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

Hence,

$$\hat{\pi}_t = \kappa \hat{mc}_t + \beta n (1 - \alpha n) \mathbb{E}_t \hat{\pi}_{t+1} + \underbrace{\mathbb{E}_t \sum_{h=1}^{\infty} (\alpha\beta n)^h \left(\kappa \hat{mc}_{t+h} + \beta n (1 - \alpha n) \hat{\pi}_{t+h+1} \right)}_{\alpha\beta n \mathbb{E}_t \hat{\pi}_{t+1}} + \hat{\mu}_t = \kappa \hat{mc}_t + \beta M^f \mathbb{E}_t \hat{\pi}_{t+1} + \hat{u}_t$$

where $M^f = n(1 + \alpha(1 - n))$; $M^f < 1$ for any $n \in (0, 1)$, and $M^f = 1$ if and only if $n = 1$.

1. *The probability measure $\mathcal{P}$ is a non-degenerate 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).[9] So, the forecast and actual law of motion of inflation along the CE equilibrium path, are, respectively

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

$$\hat{\pi}_t = \frac{\kappa}{1-\alpha\beta\rho n}\hat{mc}_t + \frac{\beta n(1-\alpha n)(\gamma^*)^2}{1-\alpha\beta n(\gamma^*)}\hat{\pi}_{t-1} + \hat{u}_t \tag{12}$$

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.,

$$\hat{\pi}_t = \frac{\kappa}{1-\beta\rho n}\hat{mc}_t + \hat{u}_t \tag{13}$$

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 derive a number of implications for forecasting errors and test them with inflation SPF forecasting data.

3.1 Forecast errors as a function of revisions and inflation lags

Consider period t and $t-1$ expectations about future inflation as specified in equation (3), respectively,

$$\widetilde{\mathbb{E}}_t\hat{\pi}_{t+h} = w_f n^h \widetilde{\mathbb{E}}_t^{\star}\hat{\pi}_{t+h} \tag{14}$$

$$\widetilde{\mathbb{E}}_{t-1}\hat{\pi}_{t+h} = w_f n^{h+1} \widetilde{\mathbb{E}}_{t-1}^{\star}\hat{\pi}_{t+h} \tag{15}$$

Let $\widetilde{\mathbb{E}}_t^{\star}\hat{\pi}_{t+h} = \hat{\pi}_{t+h} - v_{t,t+h}$, 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 delivers[10]

[9] γ^* can be found numerically since an analytical solution is almost always impossible.
[10] See Appendix.

$$\underbrace{\hat{\pi}_{t+h} - \widetilde{\mathbb{E}}_t \hat{\pi}_{t+h}}_{\text{forecasting error}} = \frac{1 - w_f n^h}{w_f n^h} \underbrace{\left(\widetilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}\right)}_{\text{forecasting revision}} + n(1 - w_f n^h) \widetilde{\mathbb{E}}^{\star}_{t-1} \hat{\pi}_{t+h} + v_{t,t+h} \tag{16}$$

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.

When $\widetilde{\mathbb{E}}^{\star}_t$ is a consistent expectations operator as described in section 2.3, $\widetilde{\mathbb{E}}^{\star}_{t-1} \hat{\pi}_{t+h}$ 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

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

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

Similarly, when $\widetilde{\mathbb{E}}^{\star}_t$ denotes well-specified forecasts, $\widetilde{\mathbb{E}}^{\star}_{t-1} \hat{\pi}_{t+h}$ depends on inflation in period $t-1$.[11] So, the appropriate equation at any forecasting horizon, h, is

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

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

To test for implications 1a) and 1b), I use quarterly inflation SPF data from 1968:Q4 to 2014:Q2 and focus on the year-to-year inflation forecasts, i.e., $h = 3$ (similar to Coibion and Gorodnichenko (2015)).

Misspecified forecasts & myopia. 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 $v_{t,t+3} = \mathcal{F}(\hat{\pi}_{t-3}, \hat{\pi}_{t-2}, \hat{\pi}_{t-1}, \varepsilon_{t-1}, ..., \varepsilon_{t+3})$, where $\mathcal{F}$ is a linear function.[12] Then, $\mathbb{E}(v_{t,t+3}\hat{\pi}_{t-2}) \neq 0$

[11] $\widetilde{\mathbb{E}}^{\star}_{t-1} \hat{\pi}_{t+h}$ denotes expectations under no myopia, i.e., $\widetilde{\mathbb{E}}^{\star}_{t-1} \hat{\pi}_{t+h} = \widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}$ for $n = 1$. Hence, $\widetilde{\mathbb{E}}^{\star}_{t-1} \hat{\pi}_{t+h} = \frac{\kappa \rho^{h+1}}{1-\beta\rho} \hat{mc}_{t-1}$, while with no myopia $\hat{\pi}_{t-1} = \frac{\kappa}{1-\beta\rho} \hat{mc}_{t-1} \Rightarrow \hat{mc}_{t-1} = \frac{1-\beta\rho}{\kappa} \hat{\pi}_{t-1}$.

[12] See Appendix B for details.

	(1)	(2)	(3)
$\hat{\pi}_{t+3,t} - \widetilde{\mathbb{E}}_t \hat{\pi}_{t+3}$			
Panel A: OLS			
$\widetilde{\mathbb{E}}_t \hat{\pi}_{t+3} - \widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+3}$	1.193**	1.141**	1.141***
	(0.497)	(0.458)	(0.420)
$\hat{\pi}_{t-1}$		0.021	-0.001
		(0.050)	(0.073)
$\hat{\pi}_{t-2}$			0.099*
			(0.051)
$\hat{\pi}_{t-3}$			-0.081
			(0.055)
Constant	0.002	-0.074	-0.072
	(0.144)	(0.174)	(0.186)
R-squared	0.195	0.197	0.210
Panel B: IV			
$\hat{\pi}_{t+3,t} - \widetilde{\mathbb{E}}_t \hat{\pi}_{t+3}$	1.907***	2.095**	1.396**
	(0.605)	(0.878)	(0.688)
$\hat{\pi}_{t-1}$		-0.050	-0.059
		(0.077)	(0.132)
$\hat{\pi}_{t-2}$			0.131*
			(0.071)
$\hat{\pi}_{t-3}$			-0.051
			(0.082)
Constant	0.004	0.190	-0.106
	(0.084)	(0.263)	(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 various implications for forecasting errors.**

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

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

where $\hat{\pi}_{t+3,t}$ is the average inflation rate over the current and next three quarters. Real-time inflation data is used since final historical data might embed re-definitions and re-classifications.[13]

Well-specified forecasts & myopia. When the forecast rule is well-specified (as in Gabaix (2020)), the forecast revision might still be correlated with the error term, hence instrumental variable estimation should be used to estimate

$$\hat{\pi}_{t+3} - \widetilde{\mathbb{E}}_t \hat{\pi}_{t+3} = c_3 + K_3 \left(\widetilde{\mathbb{E}}_t \hat{\pi}_{t+3} - \widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+3} \right) + \zeta_{1,3} \hat{\pi}_{t-1} + v_{t,t+h} \tag{20}$$

Table 1, Panel A and B, columns (2) - (3), present the OLS and IV estimates of regressions in (19) - (20), 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%. As mentioned earlier, the true estimate of $\zeta_{2,3}$ is even larger than the OLS 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.

3.1.1 Delayed over-shooting

Angeletos et al. (2020) have brought forward evidence in favor of forecasters' delayed over-shooting for future unemployment and inflation. The authors show that in order to replicate this fact, forecasters need to over-extrapolate and be endowed with noisy information. Proposition 1 shows that a combination of myopia and misspecified forecasts delivers delayed over-shooting if there is sufficient endogenous over-extrapolation, *but* it can also deliver late over-reaction even if there is under-extrapolation, i.e., $\gamma^* < \rho$.[14] As shown below, the latter can happen in the presence of an upward reweighing factor $w_f \geq 1$.

Proposition 1 *Let $\mathbb{I}_k$ be the impulse response function (IRF) of the forecasting error at horizon*

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

[14] An important difference from Angeletos et al. (2020) is that over- or under-extrapolation in the present paper is endogenously determined through the consistent expectations equilibrium solution concept.

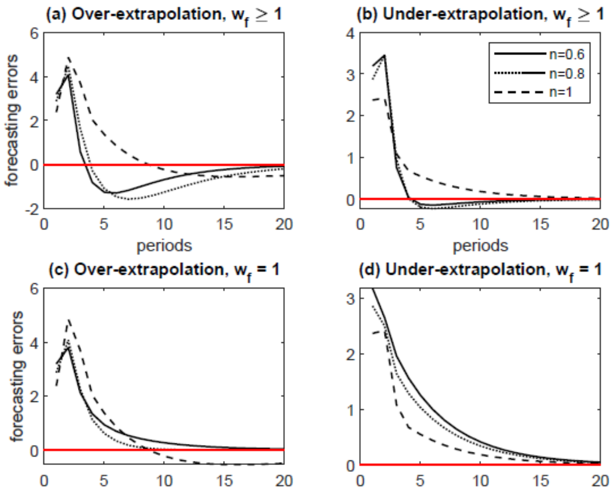


Figure 1: **Evolution of the IRF in cases of endogenous over- and under-extrapolation with $w_f > 1$ (upper panels) and $w_f = 1$ (lower panels).** Parameterization: $\kappa = 1$, $\rho = 0.8$, $\alpha = 0.73$, $\sigma_\varepsilon = 1$, $\sigma_u = 0$ for upper panels and $\sigma_u = 2.1$ for the lower panels. The implied equilibrium first-order autocorrelation coefficients of inflation are i) left upper panel: $\gamma^* = 0.924$ and $\gamma^* = 0.958$ for $n = 0.6$ and $n = 0.8$, respectively; ii) left lower panel: $\gamma^* = 0.859$ and $\gamma^* = 0.901$ for $n = 0.6$ and $n = 0.8$, respectively; iii) right upper panel: $\gamma^* = 0.753$ and $\gamma^* = 0.779$ for $n = 0.6$ and $n = 0.8$, respectively; iv) right lower panel: $\gamma^* = 0.295$ and $\gamma^* 0.396$ for $n = 0.6$ and $n = 0.8$, respectively.

$k \in \{0, 1, 2, ...\}$ *w.r.t. a one-time shock* $\varepsilon_t > 0$, *i.e.,*

$$\mathbb{I}_k = \frac{\partial(\hat{\pi}_{t+k} - \widetilde{\mathbb{E}}^{\star}_{t+k-1}\hat{\pi}_{t+1})}{\partial \varepsilon_t}$$

Then, there will be delayed over-shooting if and only if $\rho^2 < w_f n(\gamma^*)^2$. *Moreover*

- *if* $w_f = 1$ *for all* $n \in (0, 1]$, *there will be delayed over-shooting* only *if there is sufficiently high endogenous over-extrapolation* ($\gamma^* >> \rho$)*;*

- *if* $w_f > 1$, *there can be delayed over-shooting* even *if there is endogenous under-extrapolation* ($\gamma^* < \rho$)*.*

Proof. See Appendix B.1. ■

One can show that if the variance of the i.i.d shock $\hat{u}_t$ is 0, then there will always be endogenous over-extrapolation. More generally however, if the variance of $\hat{u}_t$ is different from 0, there could be both over- and under-extrapolation, in which case delayed over-shooting is guaranteed for a sufficiently high upward reweighing factor. To appreciate the value of the upward reweighing factor, Figure 1 exhibits examples of the IRF under endogenous over- and under-extrapolation with $w_f > 1$ (upper panels) and $w_f = 1$ (lower panels) along the CE equilibrium.

On the other hand, under FIRE the IRF would be 0 for any k. In the case of myopia and well-specified forecasting rules, forecasters would exhibit *only* under-reaction for any k. In the presence of consistent misspecified forecasts absent of myopia ($w_f = n = 1$), delayed overshooting would materialize only if there would be sufficient over-extrapolation. As discussed above, a similar implication would hold true if we turned off the upward reweighing factor ($w_f = 1$ for any $n \in (0, 1]$).

3.2 Taking Stock

The analytical and empirical analysis in the previous sections revealed

- supporting evidence in favor of a combination of misspecified autoregressive forecasting rules and myopia;
- no supporting evidence of well-specified forecasting rules and myopia;
- low likelihood of misspecified autoregressive forecasting rules without myopia;

The present paper documents novel evidence in favor of a combination of misspecified, yet consistent, forecasting rules and myopia. On the other hand, for myopia in Gabaix (2020) - combined with well-specified forecasts - 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. Estimates in Table 1 clearly show that the forecasting revision is significantly positive, whereas $\hat{\pi}_{t-1}$ is not. Moreover, as shown in Proposition 1, for misspecified forecasting rules to be supported absent of myopia, there should be *sufficient* over-extrapolation, which does not have to be the case generally. Subsequently, it would be harder to match empirical findings with the theoretical implications of misspecified autoregressive forecasts without myopia.

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).

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 misspecified forecasts.

4 General Equilibrium Model

This section nests the proposed expectations formation process into the otherwise baseline New Keynesian DSGE model with habit formation in consumption and inflation indexation, similar to Milani (2006). Bayesian estimation of the model on U.S. aggregate data seeks to mainly 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. 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.[15] 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 the deviation of current consumption from a stock of internal habits in consumption, labor supply and bonds, subject to her budget constraint. The problem each household faces is

$$\max_{c_{it},H_{it},B_{it}} \widetilde{\mathbb{E}}_{it} \sum_{h=0}^{\infty} \beta^h \xi_{t+h} \left(\frac{(c_{it} - \eta c_{i,t-1})^{1-\sigma}}{1-\sigma} - \psi \frac{H_{it}^{1+\varphi}}{1+\varphi} \right) \tag{21}$$

s.t.

$$\frac{R_{t-1}}{\pi_t} b_{i,t-1} = b_{it} - w_t H_{it} - d_{it} + c_{it} \tag{22}$$

where β is the discount factor; $0 \leq \eta \leq 1$ measures the degree of habit in consumption; σ is the inverse intertemporal elasticity of substitution; $\widetilde{\mathbb{E}}_{it}$ is a generic subjective expectations operator that satisfies the law of iterative expectations and standard probability rules; ξ_{it} is a preference shock; H_{it} is labor supply; R_{t-1} is the gross return on the real bond choice $b_{i,t-1}$; w_t is the real wage; d_{it} denote real dividends from firms.

[15]Bonds are assumed to be in zero net-supply.

Solving the household's optimization problem, applying the myopic adjustment process, and imposing market clearing conditions delivers

$$\begin{aligned}\hat{\tilde{x}}_t &= n\frac{(1-\beta n)(1-\beta+\beta(1-\eta))}{1-\beta}\widetilde{\mathbb{E}}_t^\star \hat{\tilde{x}}_{t+1} \\ &+ \widetilde{\mathbb{E}}_t^\star \sum_{h=0}^{\infty}(\beta n)^h \left(\beta n^2 (1-\beta n)(1-\eta)\hat{\tilde{x}}_{t+h+2} - \frac{1-\beta\eta}{\widetilde{\sigma}}\left(\hat{R}_{t+h} - \hat{\pi}_{t+h+1} - \hat{e}_{t+h}\right)\right)\end{aligned} \tag{23}$$

where $\widetilde{\sigma} = \frac{\sigma(1-\beta)}{1-\beta n}$, and

$$\hat{\tilde{x}}_t = \hat{x}_t - \eta \hat{x}_{t-1} \tag{24}$$

with $\hat{x}_t$ being the output gap. The variable $\hat{e}_t$ is a demand shock following an AR(1) process

$$\hat{e}_t = \rho_e \hat{e}_{t-1} + \sigma_e \varepsilon_t^e,\ \varepsilon_t^e \sim \mathcal{N}(0,1) \tag{25}$$

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 that the share of firms that are not allowed to optimize their price in period t can still adjust according to the indexation rule (see Christiano et al. (2005)):

$$\hat{p}_{jt} = \hat{p}_{jt-1} + \rho_\pi \hat{\pi}_{t-1} \tag{26}$$

where $0 \leq \rho_\pi \leq 1$ measures the degree of indexation to past inflation. Each firm j maximizes expected present discounted value of future profits $\Pi_{j,t+h}$ that are a function of $\left(p_{jt}^* \left(\frac{P_{t+h-1}}{P_{t-1}}\right)^{\rho_\pi}\right)$:

$$\max_{p_{jt}^*} \widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{t,t+h} \Pi_{j,t+h} \left(p_{jt}^* \left(\frac{P_{t+h-1}}{P_{t-1}}\right)^{\rho_\pi}\right) \tag{27}$$

s.t.

$$P_t = \left[\alpha \left(P_{t-1}\left(\frac{P_{t-1}}{P_{t-2}}^{\rho_\pi}\right)\right)^{1-\zeta} + (1-\alpha)(p_t^*)^{1-\zeta}\right]^{\frac{1}{1-\zeta}} \tag{28}$$

where $Q_{t,t+h}$ is a stochastic discount factor and ζ is the elasticity of substitution among goods produced by firms. Solving the firm's optimization problem, applying the myopic adjustment process, and imposing market clearing conditions, one derives the aggregate supply as follows

$$
\begin{aligned}
\hat{\hat{\pi}}_t &= \kappa\left(\omega\hat{x}_t + \frac{\widetilde{\sigma}}{1-\eta\beta}\hat{\hat{x}}_t\right) \\
&+ \widetilde{\mathbb{E}}^{\star}_t \sum_{h=1}^{\infty}(\alpha\beta n)^h \left(\kappa\omega\hat{x}_{t+h} + \frac{\kappa\widetilde{\sigma}\beta\eta(\alpha-\eta)}{\alpha(1-\eta\beta)}\hat{\hat{x}}_{t+h} + \frac{1-\alpha n}{\alpha}\hat{\hat{\pi}}_{t+h}\right) + \frac{1}{1-\alpha\beta\rho n}\hat{u}_t
\end{aligned} \tag{29}
$$

where $\kappa = \frac{\sigma(1-\alpha n)(1-\alpha)}{\widetilde{\sigma}\alpha(1+\varphi\varsigma)}$; $\omega = \frac{\varphi\widetilde{\sigma}}{\sigma}$; and

$$
\hat{\hat{\pi}}_t = \hat{\pi}_t - \rho_\pi \hat{\pi}_{t-1} \tag{30}
$$

The variable $\hat{u}_t$ is a cost-push shock assumed to follow an AR(1) processes,

$$
\hat{u}_t = \rho_u \hat{u}_{t-1} + \sigma_u \varepsilon^u_t,\ \varepsilon^u_t \sim \mathcal{N}(0,1) \tag{31}
$$

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

$$
\hat{R}_t = \rho_r \hat{R}_{t-1} + (1-\rho_r)\phi_\pi \hat{\pi}_t + (1-\rho_r)\phi_x \hat{x}_t + \sigma_v \varepsilon^v_t,\ \varepsilon^v_t \sim \mathcal{N}(0,1) \tag{32}
$$

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

$$
A_0(\Theta)S_t = A_1 S_{t-1} + \widetilde{\mathbb{E}}^{\star}_t \sum_{h=0}^{\infty}(F(\Theta))^h A_2(\Theta) S_{t+h+1} + B(\Theta)\mathcal{E}_t \tag{33}
$$

where $S_t = \begin{bmatrix} \hat{x}_t & \hat{\pi}_t & \hat{R}_t & \hat{e}_t & \hat{u}_t \end{bmatrix}'$ is the state vector; $\mathcal{E}_t = \begin{bmatrix} \varepsilon^e_t & \varepsilon^u_t & \varepsilon^v_t \end{bmatrix}'$ is the exogenous shocks vector; A_0, A_1, A_2, B and F are coefficient matrices. The aggregate economy in (33) is sufficiently flexible to nest three model specifications. i) $\boldsymbol{n = 0}$: This model specification exhibits the highest degree of myopia, i.e., optimal decisions - in deviations from the steady state - do not depend on the anticipated future state of the economy; ii) $\boldsymbol{n = 1}$: In this case, agents take into account an infinite-horizon of autoregressive forecasts about deviations from the steady state to make optimal decisions, exhibiting no myopia at all.[16] iii) $\boldsymbol{n \in (0,1)}$: This

[16] Preston (2005) and Milani (2006) have used the economy in (33) 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).

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 $S_{1:3,t}$

$$S_{1:3,t} = \boldsymbol{\delta}_{t-1} + \boldsymbol{\gamma}_{t-1}(S_{1:3,t-1} - \boldsymbol{\delta}_{t-1}) + \boldsymbol{\epsilon}_t \tag{34}$$

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 (34) 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

$$\widetilde{\mathbb{E}}_t^{\star} S_{1:3,t+h} = \boldsymbol{\delta}_{t-1} + (\boldsymbol{\gamma}_{t-1})^{h+1}(S_{1:3,t-1} - \boldsymbol{\delta}_{t-1}) \tag{35}$$

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

$$\begin{aligned}
\boldsymbol{\delta}_t &= \boldsymbol{\delta}_{t-1} + \iota(S_{1:3,t} - \boldsymbol{\delta}_{t-1}) \\
\boldsymbol{\gamma}_t &= \boldsymbol{\gamma}_{t-1} + \iota\left((S_{1:3,t} - \boldsymbol{\delta}_{t-1})(S_{1:3,t-1} - \boldsymbol{\delta}_{t-1})' - \boldsymbol{\gamma}_{t-1}(S_{1:3,t} - \boldsymbol{\delta}_{t-1})(S_{1:3,t} - \boldsymbol{\delta}_{t-1})'\right)\boldsymbol{\eta}_t^{-1} \\
\boldsymbol{\eta}_t &= \boldsymbol{\eta}_{t-1} + \iota\left((S_{1:3,t} - \boldsymbol{\delta}_{t-1})(S_{1:3,t} - \boldsymbol{\delta}_{t-1})' - \boldsymbol{\eta}_{t-1}\right)
\end{aligned} \tag{36}$$

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 $\frac{1}{\iota}$ 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.[17,18]

4.3 Bayesian estimation

The state-space representation of the model is

$$S_t = C_0(\Theta, \boldsymbol{\gamma}_{t-1})\boldsymbol{\Delta}_{t-1} + C_1(\Theta, \boldsymbol{\gamma}_{t-1})S_{t-1} + C_2(\Theta)\mathcal{E}_t \tag{37}$$

$$Y_t - \bar{Y} = PS_t \tag{38}$$

together with the dynamic system in (36), where $\boldsymbol{\Delta}_t = \begin{bmatrix}\boldsymbol{\delta}_t' & \mathbf{0}_{1x2}\end{bmatrix}'$; $Y_t = \begin{bmatrix}x_t^{obs} & \pi_t^{obs} & R_t^{obs}\end{bmatrix}'$ 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 1968 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.[19] I refer the reader to Appendix C.1 for more details on data preparation. The state-space form of the model in (36)-(38) is a Gaussian system, hence I evaluate the likelihood function using the Kalman filter. The posterior distribution then is calculated as

$$p(\Theta \mid Y_{1:T}) \propto p(Y_{1:T} \mid \Theta)p(\Theta) \tag{39}$$

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, $\boldsymbol{\delta}_0$ and $\boldsymbol{\gamma}_0$, for the Kalman filter procedure.[20]

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

[18] For 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.

[19] Bayesian 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.

[20] In 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

Parameters		pdf	mean	standard deviation
Calvo parameter	α	$\mathcal{B}$	0.5	0.2
Degree of myopia	n	$\mathcal{U}$	0.5	$1/\sqrt{12}$
Adjusted Inv. IES	σ	$\mathcal{G}$	2	0.5
Phillips curve elast.	κ	$\mathcal{B}$	0.3	0.15
Habit in consumption	η	$\mathcal{U}$	0.5	$1/\sqrt{12}$
Inflation indexation	ρ_π	$\mathcal{U}$	0.5	$1/\sqrt{12}$
Elasticity mc	ω	$\mathcal{N}$	0.8975	0.4
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	$\mathcal{U}$	0.5	$1/\sqrt{12}$
Autocorr. u	ρ_u	$\mathcal{U}$	0.5	$1/\sqrt{12}$
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
Gain parameter	$\bar{\iota}$	$\mathcal{G}$	0.035	0.015

Table 2: **Priors**

I fix the discount factor $\beta = 0.99$. For most of the 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 0.2. The prior for n follows a non-informative uniform distribution with mean 0.5 and standard deviation of $1/\sqrt{12}$. 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. Habit in consumption and inflation indexation parameters follow a uniform distribution with mean 0.5. Elasticity of inflation w.r.t. marginal cost follows a normal prior with mean 0.8975 and standard deviation 0.4. 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 interest rate smoothing parameter follows a beta distribution with mean 0.5 and standard deviation 0.2. The autocorrelation of all shocks 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.

beliefs at the implied moments of the RE solution, apart from the other two aforementioned methods.

4.3.1 Posterior distribution

Table 3 reports characteristics of the posterior distribution under RE and constant gain SAC-learning with AR(1) forecasting rules and $\boldsymbol{n} \in (\mathbf{0}, \mathbf{1})$ or $\boldsymbol{n} = \mathbf{1}$.[21] The model when agents learn to use simple autoregressive misspecified forecasting rules and exhibit significantly high degrees of myopia and fit the data significantly better than RE. The log marginal data likelihood is evaluated using the modified harmonic mean method in Geweke (1999).[22] Values in parenthesis report the Bayes factor value of the model specification relative to RE: the log of the Bayes factor for the SAC-learning model with $\boldsymbol{n} \in (\mathbf{0}, \mathbf{1})$ is higher than 3. 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 SAC-learning with myopia. On the other hand, the benchmark SAC-learning model with AR(1) forecasting rules and $\boldsymbol{n} = \mathbf{1}$, i.e., no myopia, fits data worse than RE, and as a result worse than SAC-learning with myopia.

The posterior mean 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 0.424, meaning that on average current expectations about beyond 10 quarters ahead are practically disregarded in present optimal decisions. In the 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 0.58 for the benchmark SAC-learning model and 0.94 for the model with $\boldsymbol{n} = \mathbf{1}$. Note that in the specification with no myopia, there needs to be much more price stickiness relative to the case with significant degrees of myopia. The implied expected price duration for the benchmark and no myopia specifications is, respectively, 7.1 and 50 months on average. The former 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 $\alpha \in [0.5, 0.67]$, while the latter implies too much price stickiness in the economy.

The RE model requires high degrees of habit in consumption and inflation indexation to fit aggregate macro data, i.e., the posterior mean of η and ρ_π respectively are 0.926 and 0.922.

[21] Posterior distributions are well-behaved, with no bimodal behavior. I rely on the method proposed by Brooks and Gelman (1998) to analyze convergence statistics.

[22] Bayes Theorem implies that $p(Y_t) = \int p(Y_t \mid \Theta)p(\Theta)d\Theta$, which is impossible to compute analytically. 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 $f(.)$ is the density of a truncated normal distribution.

The estimated posterior mean of the cost-push shock under RE is almost 0, so one might argue that some degree of inflation indexation is necessary. On the other hand, the benchmark SAC-learning model with myopia and no myopia require almost no habit in consumption ($\eta = 0.01$ and 0.001 at the posterior mean, respectively). However, the SAC-learning model with myopia requires quite some degree of inflation indexation at the posterior mean ($\rho_\pi = 0.881$), but under inflation indexation there is no need for a persistent cost-push shock ($\rho_u = 0.027$). The model with no myopia, on the contrary, requires no inflation indexation ($\rho_\pi = 0.022$), but it needs quite a persistent cost-push shock ($\rho_u = 0.851$). The Phillips curve elasticity is higher when agents are endowed with misspecified beliefs compared to RE ($\kappa = 0.009$ and $\kappa = 0.002$ under SAC-learning with myopia and no myopia, respectively, but it is only 0.001 under RE). The parameter $\widetilde{\sigma}$ at the posterior mean is estimated to be only 2.4 for the benchmark specification and it increases to 5.14 when agents are extremely forward-looking. Policy parameters are generally robust across specifications, with the posterior mode estimates of policy reaction to output gap being between 0.3 and 0.4, reaction to inflation around 1.5 and interest rate smooth parameter being close to 0.9. Likewise, the posterior mean of the marginal cost elasticity is robust at around 0.8 - 0.9 across the different specifications.

The posterior mean values for the learning gain parameter $\bar{\iota}$ is 0.005 when $n = 1$ and 0.05 for $n \in (0, 1)$, which further implies that a rolling window of 25 quarters is used to update the forecasting process when there is myopia.

Shocks under benchmark SAC-learning with myopia are more persistent than RE, but less persistent than in the specification without myopia. 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 autoregressive forecasting rules. The posterior mean estimate of the demand shock autocorrelation in the benchmark model is 0.86 and it is significantly lower than its estimate of 0.96 in the SAC-learning model with $\boldsymbol{n = 1}$. The posterior mean estimate of the cost-push shock in the benchmark model is 0.03 and it is significantly lower than under RE and SAC-learning model with $\boldsymbol{n = 1}$ of 0.85.

Shock innovations are significantly less volatile under SAC-learning than RE. 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 1.12 for the benchmark specification and 0.19 when there is no myopia. Similarly, the posterior mode estimate of the innovation related to the cost-push shock for the benchmark specification is 0.18, but 0.06 when $\boldsymbol{n} = \mathbf{1}$. The standard deviation of the monetary shock is generally robust across models at an estimated posterior mode of 0.2.

Overall, myopia and autoregressive misspecified forecasts substitute for other frictions, such as habit in consumption and price indexation, while being consistent with forecasting data evidence.

	RE			**SAC-learning**					
				$\boldsymbol{n \in (0, 1)}$			$\boldsymbol{n = 1}$		
Parameters	mean	5%	95%	mean	5%	95%	mean	5%	95%
σ	2.216	1.309	3.063	2.408	1.840	3.064	5.136	4.502	5.676
α	0.496	0.171	0.826	0.576	0.220	0.866	0.935	0.775	0.994
n	-	-	-	0.424	0.230	0.612	-	-	-
κ	0.001	0.000	0.002	0.009	0.004	0.014	0.002	0.001	0.004
η	0.926	0.877	0.982	0.011	0.002	0.025	0.001	0.000	0.004
ρ_π	0.922	0.866	0.987	0.881	0.773	0.976	0.022	0.001	0.064
ω	0.869	0.233	1.463	0.791	0.187	1.436	0.829	0.382	1.395
ϕ_x	0.367	0.225	0.513	0.435	0.266	0.651	0.395	0.240	0.584
ϕ_π	1.438	1.176	1.712	1.447	1.138	1.774	1.527	1.243	1.849
ρ_r	0.888	0.855	0.921	0.916	0.884	0.945	0.915	0.885	0.943
ρ_e	0.623	0.504	0.735	0.855	0.763	0.927	0.964	0.934	0.990
ρ_u	0.022	0.000	0.050	0.027	0.002	0.079	0.851	0.784	0.913
σ_e	6.144	2.770	10.000	1.185	1.050	1.317	0.185	0.080	0.306
σ_u	0.266	0.242	0.288	0.179	0.067	0.307	0.057	0.034	0.094
σ_v	0.209	0.192	0.226	0.208	0.192	0.226	0.208	0.192	0.226
$\bar{\iota}$	-	-	-	0.050	0.026	0.084	0.005	0.002	0.011
Log marg. data dens.									
Modified Harmonic Mean	-276.54			-267.75*			-303.18		
Bayes factor	(1)			($e^{8.79}$)			($e^{-26.64}$)		

Table 3: **Posterior distribution for RE and 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.

Figure 2 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 2, 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. On the other hand, there seems to be a shift in the way agents perceive moments of inflation

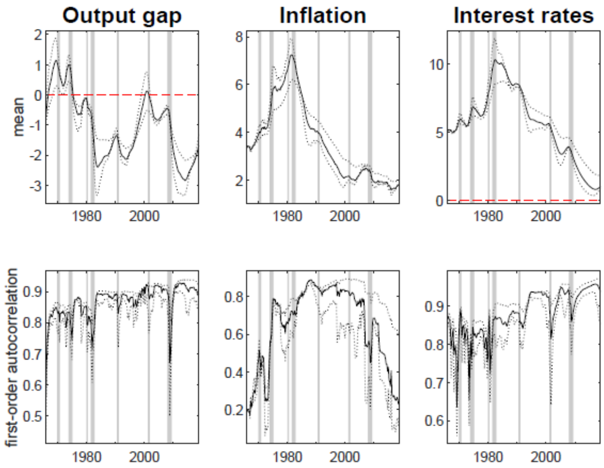


Figure 2: **Evolution of the AR(1) forecast coefficients in the benchmark SAC-learning model.** The black and dotted curves plot implied beliefs for structural parameters set at their estimated posterior mode and 90% highest posterior density, respectively. Grey areas indicate recessionary periods as reported by the National Bureau of Economic Research.

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.[23] Another interesting observation from Figure 2 is that the implied beliefs about the annualized mean of inflation over the last decade have been particularly steady at 2%.

4.3.2 VAR(1) forecasting rules

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

Table 4 exhibits posterior distribution of parameters when agents use VAR(1) forecasting rules instead of univariate ones. Estimates of parameters are extremely similar to the ones under AR(1) forecasts, and the log marginal data density is almost identical to the same measure under SAC-learning with AR(1) 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

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

Parameters	mean	5%	95%
σ	2.3409	1.6659	3.2078
α	0.5664	0.2194	0.8435
n	0.1859	0.0062	0.5311
κ	0.0081	0.0034	0.0138
η	0.0698	0.0018	0.2292
ρ_π	0.8898	0.8147	0.9625
ω	0.6208	0.0985	1.1938
ϕ_x	0.4184	0.2603	0.618
ϕ_π	1.4821	1.1899	1.7921
ρ_r	0.9142	0.8831	0.9431
ρ_e	0.8918	0.8243	0.9467
ρ_u	0.0242	0.0013	0.0718
σ_e	1.6667	0.9139	2.385
σ_u	0.1433	0.0485	0.2638
σ_v	0.2089	0.1913	0.2269
$\bar{\iota}$	0.0478	0.0224	0.0767
Log marg. data dens.			
Modified Harmonic Mean		-267.09	

Table 4: **Posterior distribution for SAC-learning with VAR(1) forecasting rules**.

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. Therefore, using more elaborate forecasting rules, such as VAR(1), will not add any useful information to households and firms, and it will not enhance the model's data fit. Moving forward then I remain focused on AR(1) forecasting rules.

4.3.3 Impulse response functions

Computing the 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.[24] Figure 4 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 mean.[25]

[24]Consecutively, 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.

[25]The 3-dimensional IRF to belief shocks are exposed in Appendix C.2.

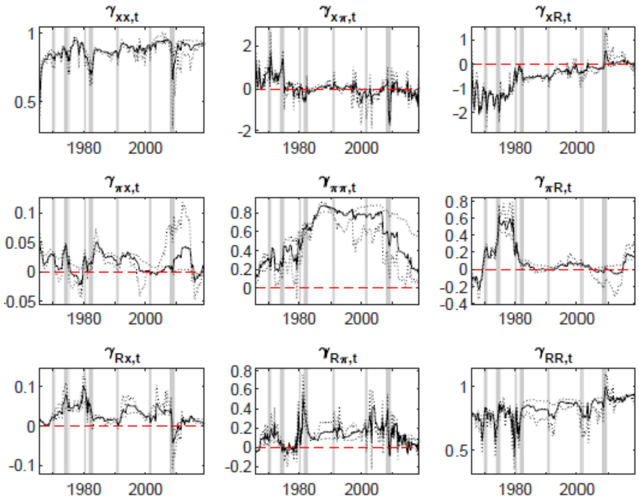


Figure 3: **Evolution of the VAR(1) forecast coefficients in the benchmark constant gain SAC-learning model.** The black and dotted curves plot implied beliefs for structural parameters set at their estimated posterior mode and 90% highest posterior density, respectively. Grey areas indicate recessionary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

Figure 4 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 and monetary shocks, whereas output gap has been reacting less to cost-push shocks over the last decade.[26] Figure 5 then projects the three dimensional IRF on the [response - periods of response] plane, with the red curve plotting the average response of aggregates.[27] 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 5 emphasizes that even though there is no habit in consumption and there is no persistence in cost-push and monetary 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.

[26] I refer the reader to Appendix C.2 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 8 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.

[27] The 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.

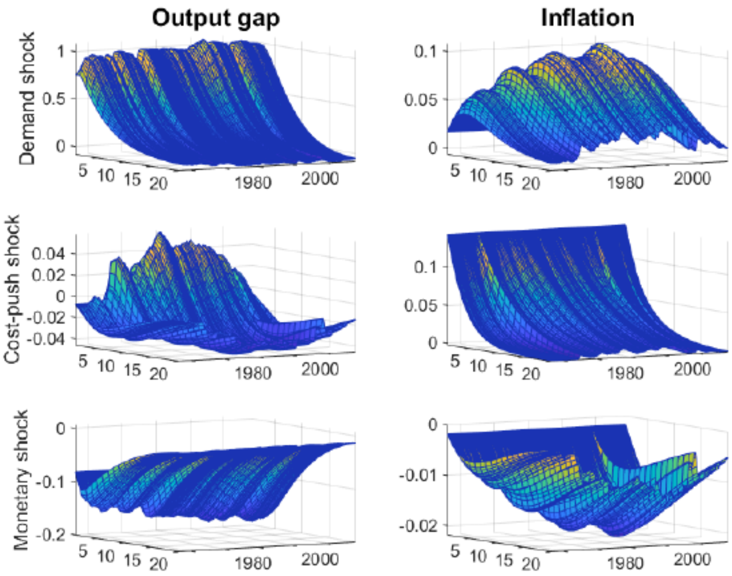


Figure 4: **3-Dimensional impulse response functions to a one standard deviation positive demand, cost-push and monetary shock for the benchmark SAC-learning model.**

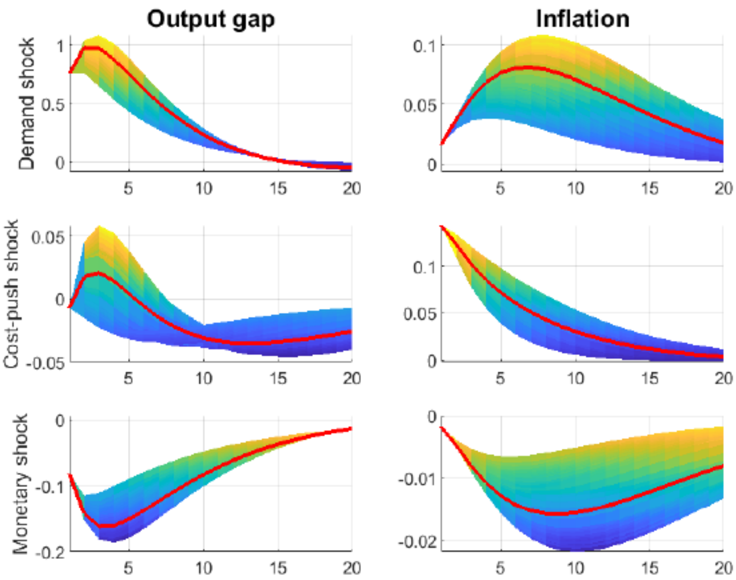
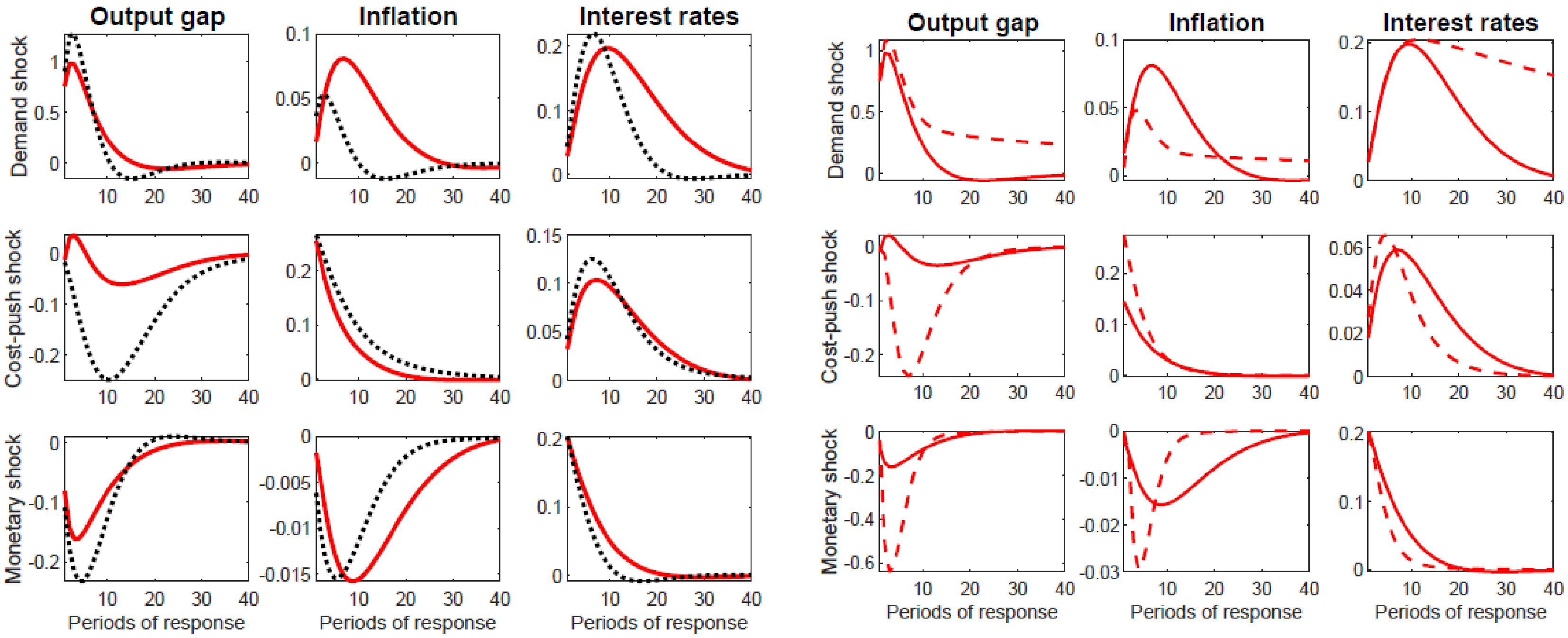


Figure 5: **Impulse response functions to a one standard deviation positive demand, cost-push, and monetary shocks for the benchmark model.** Red curve: average response.



(a) (b)

Figure 6: **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.

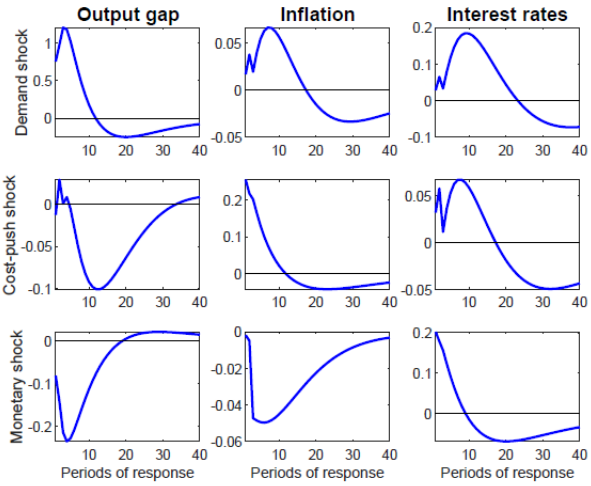


Figure 7: **Impulse response functions of one-period ahead forecasting errors to a one standard deviation positive demand, cost-push and monetary shock in the benchmark SAC-learning model.**

Figure 6 plots the average impulse responses of output gap, inflation, and nominal interest rates for RE, SAC-learning with and without myopia to a one standard deviation demand, cost-push and monetary shock. Parameters are set at their posterior mean as presented in Table 3. Due to the presence of habit in consumption and inflation indexation, all three model specifications deliver hump-shaped responses. Relative to its RE counterpart, the model under SAC-learning with myopia generates more persistent and often times more volatile response functions. This happens even though the posterior mean of habit in consumption is estimated to be practically absent for the latter model. Similarly, the SAC-learning specification with myopia generates significantly more persistent and volatile responses of inflation to demand and monetary shocks. How would these IRF compare to responses of myopia and well-specified forecasts? Cognitive discounting would simply tame the RE responses presented in Figure 6, panel (a).[28]

Finally, we will have a look at the implied average IRF of firms' forecasting errors for the benchmark SAC-learning model with myopia and other parameters set at their posterior mean. Figure 7 plots the implied IRF of one-period ahead forecasting errors to a one standard deviation demand, cost-push and monetary shock. Most of the IRFs replicate well the *initial under-reaction and delayed over-shooting*, consistent with evidence presented in Angeletos et al. (2020).

[28] A 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 relative to Gabaix (2020), yet they cannot outperform the IRF implied by the benchmark SAC-learning model of the paper. Results provided by the author upon request.

5 Concluding remarks

The present paper simultaneously incorporates 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 first part of the paper focuses on a partial equilibrium pricing problem, derives a number of testable implications, and estimates them with inflation forecasting data from the SPF. The second part of the paper embeds the same departures in a full New Keynesian model with habit in consumption and inflation indexation, derives the general equilibrium under sample autocorrelation coefficient learning, and estimates the model using Bayesian methods. The paper underscores three novel results. First, both estimation approaches strongly prefer the model in which agents use misspecified forecasts and myopia a number of other alternatives. Importantly, a combination of misspecified forecasting rules and myopia gives rise to initial under-reaction and delayed over-shooting of forecasters, consistent with recent evidence. Second, the best fitting expectations formation process for both households and firms is characterized by high degrees of myopia and simple AR(1) forecasts, and more elaborate VAR(1) rules do not add any useful information to private agents. Third, 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.

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

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

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

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

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

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

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

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

This implies further that

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

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

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

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

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

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

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

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

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

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

$$y_{jt} = z_{jt} h_{jt}^{a_h} \tag{A.9}$$

where z_{jt} and h_{jt} are the technology shock and labor demand, respectively, and $0 < a_h \leq 1$. The price optimization problem is subject to Calvo price stickiness: firms can adjust the price with some constant probability $0 \leq 1-\alpha \leq 1$. If the firm cannot optimize the price, they can still adjust their prices according to

$$logp_{it} = logp_{i,t-1} + \rho_\pi log\pi_{t-1}$$

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,

$$\mathcal{L}_{jt} = \min_{h_{jt}} w_t h_{jt} + mc_{jt}(y_{jt} - z_{jt} h_{jt}^{a_h}) \tag{A.10}$$

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

$$mc_{jt} = \frac{w_t}{a_h z_{jt} h_{jt}^{a_h - 1}} \tag{A.11}$$

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

$$\max_{P_{jt}^*} \widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{j,t+h} \left(\frac{P_{jt}^*}{P_t} y_{j,t+h} - w_{t+h} h_{j,t+h} \right) \tag{A.12}$$

where $Q_{j,t+h} = \beta^h \frac{P_t}{P_{t+h}} \frac{\lambda_{t+h}}{\lambda_t}$ is the stochastic discount factor of the j^{th} firm. From (A.9) and (A.11),

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

Substituting for y_{jt} and h_{jt}, using (A.9) and (A.11), the problem becomes

$$\max_{P_{jt}^*} \widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{j,t+h} y_{t+h} \left(\frac{P_{jt}^*}{P_t} \left(\frac{P_{jt}^*}{P_{t+h}} \right)^{-\zeta} \left(\frac{P_{t+h-1}}{P_{t-1}} \right)^{\rho_\pi(1-\zeta)} - \widetilde{\omega} mc_{j,t+h} \left(\frac{P_{jt}^*}{P_{t+h}} \right)^{-\zeta} \left(\frac{P_{t+h-1}}{P_{t-1}} \right)^{-\rho_\pi \zeta} \right) \tag{A.13}$$

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

$$\widetilde{\mathbb{E}}_{jt} \sum_{h=0}^{\infty} (\alpha\beta)^h Q_{t+h} \zeta \left(\frac{P_{jt}^*}{P_t} \right)^{-\zeta-1} \pi_{t,t+h}^{\zeta} \pi_{t-1,t+h-1}^{-\rho_\pi \zeta} y_{t+h} \left(\zeta \widetilde{\omega} mc_{j,t+h} \pi_{t,t+h} - (\zeta - 1) \left(\frac{P_{jt}^*}{P_t} \right) \pi_{t-1,t+h-1}^{\rho_\pi} \right) = 0 \tag{A.14}$$

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.,

$$\frac{R_t}{\bar{R}} = \left(\frac{R_{t-1}}{\bar{R}}\right)^{\rho_r} \left(\frac{\pi_t}{\bar{\pi}}\right)^{(1-\rho_r)\phi_\pi} \left(\frac{x_t}{\bar{x}}\right)^{(1-\rho_r)\phi_x} e^{\sigma_v \varepsilon_t^v},\ \varepsilon_t^v \sim \mathcal{N}(0,1) \tag{A.15}$$

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

A.2 Log-linearized Model

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

Households. Log-linearizing (??) around steady-states generates

$$\hat{\tilde{c}}_{it} = \widetilde{\mathbb{E}}_{it}\hat{\tilde{c}}_{i,t+1} - \frac{1-\beta\eta}{\sigma}\widetilde{\mathbb{E}}_{it}(\hat{R}_t - \hat{\pi}_{t+1}) + \frac{1}{\sigma}\widetilde{\mathbb{E}}_{it}\left(\hat{g}_{it} - \hat{g}_{i,t+1}\right) \tag{A.16}$$

where $\hat{\tilde{c}}_t = \hat{c}_{it} - \eta c_{i,t-1} - \beta\eta\widetilde{\mathbb{E}}_{it}(\hat{c}_{i,t+1} - \hat{c}_{it})$ and $\hat{g}_{it} = \hat{\xi}_{it} - \beta\eta\hat{\xi}_{i,t+1}$. Solving the equation backwards and substituting into the log-linearized intertemporal budget constraint (after substituting $\hat{c}_{it}$ by $\hat{\tilde{c}}_t + \eta\hat{c}_{i,t-1} + \beta\eta\widetilde{\mathbb{E}}_{it}(\hat{c}_{i,t+1} - \hat{c}_{it})$), dropping the subscript i since all households are the same, using $c_t = y_t$, and expressing everything in terms of the output gap $\hat{x}_t = \hat{y}_t - \hat{y}_t^n$ where $\hat{y}_t^n$ is the natural rate of output, yields the aggregate demand equation:

$$\hat{\tilde{x}}_t = (1-\beta+\beta(1-\eta))\widetilde{\mathbb{E}}_t\hat{\tilde{x}}_{t+1} + \widetilde{\mathbb{E}}_t\sum_{h=0}^{\infty}\beta^h\left(\beta(1-\beta)(1-\eta)\hat{x}_{t+h+2} - \frac{1-\beta\eta}{\sigma}\left(\hat{R}_{t+h} - \hat{\pi}_{t+h+1} - \hat{e}_{t+h}\right)\right) \tag{A.17}$$

$$\hat{\tilde{x}}_t = \hat{x}_t - \eta\hat{x}_{t-1} \tag{A.18}$$

and $\hat{e}_t = \frac{\sigma}{1-\beta\eta}\left((\hat{y}_{t+1}^n - \hat{g}_{t+1}) - (\hat{y}_t^n - \hat{g}_t)\right)$ is such that

$$\hat{e}_t = \rho_e\hat{e}_{t-1} + \sigma_e\varepsilon_t^e,\ \varepsilon_t^e \sim \mathcal{N}(0,1) \tag{A.19}$$

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

$$\hat{P}_{jt}^* - \hat{P}_t = (1-\alpha\beta)\widetilde{\mathbb{E}}_{jt}\sum_{h=0}^{\infty}(\alpha\beta)^h\left(\frac{\widetilde{\omega}}{1+\zeta\widetilde{\omega}}\left(\widetilde{\omega}\hat{y}_{t+h} - \hat{\lambda}_{t+h} + \xi_{t+h}\right) + \alpha\beta(\hat{\pi}_{t+h+1} - \rho_\pi\hat{\pi}_{t-1})\right)$$

Then, the aggregate supply can be written as

$$\hat{\tilde{\pi}}_t = \widetilde{\kappa}\left(\widetilde{\omega}\hat{x}_t + \frac{\sigma}{1-\eta\beta}\hat{\tilde{x}}_t\right) + \widetilde{\mathbb{E}}_t\sum_{h=1}^{\infty}(\alpha\beta)^h\left(\widetilde{\kappa}\widetilde{\omega}\hat{x}_{t+h} + \frac{\widetilde{\kappa}\sigma\beta\eta(\alpha-\eta)}{1-\eta\beta}\hat{\tilde{x}}_{t+h} + \frac{1-\alpha}{\alpha}\hat{\tilde{\pi}}_{t+1} + \hat{u}_{t+h}\right) \tag{A.20}$$

where

$$\hat{u}_t = \rho_u \hat{u}_{t-1} + \sigma_u \varepsilon_t^u,\ \varepsilon_t^u \sim \mathcal{N}(0,1) \tag{A.21}$$

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

$$\hat{R}_t = \rho_r \hat{R}_{t-1} + (1-\rho_r)\phi_\pi \hat{\pi}_t + (1-\rho_r)\phi_x \hat{x}_t + \sigma_v \varepsilon_t \tag{A.22}$$

A.3 Myopic adjustment

Each household's expectations about any variable $\hat{Z}_{t+h} = \begin{bmatrix} \hat{\tilde{x}}_{t+h} & \hat{R}_{t+h} & \hat{\pi}_{t+h} \end{bmatrix}'$ for $h \geq 0$ are adjusted as follows:

$$\widetilde{\mathbb{E}}_t^h \hat{Z}_{t+h} = w_h n^h \widetilde{\mathbb{E}}_t^\star \hat{Z}_{t+h} \tag{A.23}$$

where $w_h = \frac{1-\beta n}{1-\beta} \geq 1$ is an upward reweighing factor. Suppose for a moment that $\eta = 0$, i.e., there is no habit in consumption. If $n = 1$, agents should consume a constant fraction $(1-\beta)$ of the expected future discounted wealth, given a constant real interest rate. If $n < 1$, agents behave as if, given a constant real interest rate, they consume a constant fraction $(1-\beta n) > (1-\beta)$ of the expected future discounted wealth. Hence, as households become shorter-sighted due to excess discounting, the importance of each of the future deviations from steady state values will increase by a factor of $w_h = \frac{\text{consumption share of expected future discounted wealth w/ myopia}}{\text{consumption share of expected future discounted wealth w/o myopia}} = \frac{1-\beta n}{1-\beta}$.

Firm's expectations about any variable $\hat{Z}_{t+h} = \begin{bmatrix} \hat{x}_{t+h} & \hat{\tilde{x}}_{t+h} & \hat{\tilde{\pi}}_{t+h} \end{bmatrix}'$ for $h \geq 0$ are adjusted as follows:

$$\widetilde{\mathbb{E}}_t^f \hat{Z}_{t+h} = w_f n^h \widetilde{\mathbb{E}}_t^\star \hat{Z}_{t+h} \tag{A.24}$$

where $w_f = \frac{1-\alpha n}{1-\alpha}$.

Finally, the aggregate demand and supply, respectively, become:

$$\begin{aligned} \hat{\tilde{x}}_t &= n\frac{(1-\beta n)(1-\beta+\beta(1-\eta))}{1-\beta}\widetilde{\mathbb{E}}_t^\star \hat{\tilde{x}}_{t+1} \\ &+ \widetilde{\mathbb{E}}_t^\star \sum_{h=0}^{\infty} (\beta n)^h \left(\beta n^2 (1-\beta n)(1-\eta)\hat{x}_{t+h+2} - \frac{1}{\widetilde{\sigma}}\left(\hat{R}_{t+h} - \hat{\pi}_{t+h+1} - \hat{e}_{t+h}\right)\right) \end{aligned} \tag{A.25}$$

where $\widetilde{\sigma} = \sigma\frac{1-\beta n}{1-\beta}$.

$$\begin{aligned} \hat{\tilde{\pi}}_t &= \kappa\left(\omega \hat{x}_t + \frac{\widetilde{\sigma}}{1-\eta\beta}\hat{\tilde{x}}_t\right) \\ &+ \widetilde{\mathbb{E}}_t^\star \sum_{h=1}^{\infty} (\alpha\beta n)^h \left(\kappa\omega \hat{x}_{t+h} + \frac{\kappa\widetilde{\sigma}\beta\eta(\alpha-\eta)}{\alpha(1-\eta\beta)}\hat{\tilde{x}}_{t+h} + \frac{1-\alpha n}{\alpha}\hat{\tilde{\pi}}_{t+h}\right) + \frac{1}{1-\alpha\beta\rho n}\hat{u}_t \end{aligned} \tag{A.26}$$

where $\kappa = \widetilde{\kappa}\frac{\sigma(1-\alpha n)}{\widetilde{\sigma}(1-\alpha)}$, $\omega = \frac{\sigma(1-\alpha n)}{\widetilde{\sigma}(1-\alpha)}$.

A.4 Model in matrix form

The aggregate economy model in matrix form is described by

$$A_0(\Theta)S_t = A_1(\Theta)S_{t-1} + \widetilde{\mathbb{E}}_t^\star \sum_{h=1}^{\infty} F^h A_2(\Theta)S_{t+h} + B(\Theta)\mathcal{E}_t \tag{A.27}$$

where $S_t = \begin{bmatrix}\hat{x}_t & \hat{\pi}_t & \hat{R}_t & \hat{e}_t & \hat{u}_t\end{bmatrix}'$; $\mathcal{E}_t = \begin{bmatrix}\varepsilon_t^e & \varepsilon_t^u & \varepsilon_t^v\end{bmatrix}'$; $\Theta = \{\alpha, \beta, n, \widetilde{\sigma}, \kappa, \phi_\pi, \phi_x, \rho_r, \rho_e, \rho_u, \sigma_e, \sigma_u, \sigma_v\}$, 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 asmo

$$S_t = \underbrace{\boldsymbol{\Delta}_{t-1} + \boldsymbol{\Gamma}_{t-1}(S_{t-1} - \boldsymbol{\Delta}_{t-1})}_{\text{PLM for aggregate endo var's}} + \underbrace{HS_{t-1}}_{\text{PLM for the shocks}} + \widetilde{\boldsymbol{\epsilon}}_t \tag{A.28}$$

where $\boldsymbol{\delta}_t = \begin{bmatrix}\boldsymbol{\delta}_t' & \mathbf{0}_{1x2}\end{bmatrix}'$; $\boldsymbol{\Gamma}_t = \begin{bmatrix}\boldsymbol{\gamma}_t & \mathbf{0}_{3x2} \\ \mathbf{0}_{2x3} & \mathbf{0}_{2x2}\end{bmatrix}$; H is a diagonal matrix with diagonal equal to $\begin{bmatrix}\mathbf{0}_{1x3} & \rho_e & \rho_u\end{bmatrix}'$; $\widetilde{\boldsymbol{\epsilon}}_t = \begin{bmatrix}\boldsymbol{\epsilon}_t' & \sigma_e\varepsilon_t^e & \sigma_u\varepsilon_t^u\end{bmatrix}'$. The forecast of the state vector $h \geq 1$ periods ahead is described by

$$\widetilde{\mathbb{E}}_t^\star S_{t+h} = \underbrace{\boldsymbol{\Delta}_{t-1} + \boldsymbol{\Gamma}_{t-1}^{\tau-t+1}(S_{t-1} - \boldsymbol{\Delta}_{t-1})}_{\text{forecast of endo var's}} + \underbrace{H^h S_t}_{\text{forecast of shocks}} \tag{A.29}$$

Plugging (A.29) into (A.27), we get the actual law of motion:

$$\widetilde{A}_0(\Theta)S_t = \widetilde{A}_1(\Theta)\boldsymbol{\Delta}_{t-1} + \widetilde{A}_2(\Theta, \boldsymbol{\Gamma}_{t-1})S_{t-1} + B\mathcal{E}_t \tag{A.30}$$

where

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

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

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

The infinite sums are defined as,

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

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

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

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

B Testable Implications

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

$$\widetilde{\mathbb{E}}_t \hat{\pi}_{t+h} = w_f n^h \widetilde{\mathbb{E}}_t^{\star} \hat{\pi}_{t+h} \tag{B.1}$$

$$\widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} = w_f n^{h+1} \widetilde{\mathbb{E}}_{t-1}^{\star} \hat{\pi}_{t+h} \tag{B.2}$$

Let $\widetilde{\mathbb{E}}_t^{\star} \hat{\pi}_{t+h} = \hat{\pi}_{t+h} - v_{t,t+h}$, 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 $\widetilde{\mathbb{E}}_t^{\star} \hat{\pi}_{t+h} = \hat{\pi}_{t+h} - v_{t,t+h}$,

$$\begin{aligned}
\widetilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} &= w_f n^h (\hat{\pi}_{t+h} - v_{t,t+h}) - w_f n^h \widetilde{\mathbb{E}}_{t-1}^{\star} \hat{\pi}_{t+h} \\
&= w_f n^h (\hat{\pi}_{t+h} - \widetilde{\mathbb{E}}_t \hat{\pi}_{t+h}) + w_f n^h \widetilde{\mathbb{E}}_t \hat{\pi}_{t+h} - w_f n^h v_{t,t+h} - w_f n^h \widetilde{\mathbb{E}}_{t-1}^{\star} \hat{\pi}_{t+h} - w_f n^h (\widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} - \\
&= w_f n^h (\hat{\pi}_{t+h} - \widetilde{\mathbb{E}}_t \hat{\pi}_{t+h}) + w_f n^h (\widetilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) \\
&+ w_f n^h \widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} - w_f n^h \widetilde{\mathbb{E}}_{t-1}^{\star} \hat{\pi}_{t+h} - w_f n^h v_{t,t+h}
\end{aligned}$$

Hence,

$$\begin{aligned}
\hat{\pi}_{t+h} - \widetilde{\mathbb{E}}_t \hat{\pi}_{t+h} &= \frac{1 - w_f n^h}{w_f n^h} (\widetilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) - (\widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h} - n \widetilde{\mathbb{E}}_{t-1}^{\star} \hat{\pi}_{t+h}) + v_{t,t+h} \\
&= \frac{1 - w_f n^h}{w_f n^h} (\widetilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) + n(1 - w_f n^h) \widetilde{\mathbb{E}}_{t-1}^{\star} \hat{\pi}_{t+h} + v_{t,t+h} \\
&= K_h (\widetilde{\mathbb{E}}_t \hat{\pi}_{t+h} - \widetilde{\mathbb{E}}_{t-1} \hat{\pi}_{t+h}) + n(1 - w_f n^h) \widetilde{\mathbb{E}}_{t-1}^{\star} \hat{\pi}_{t+h} + v_{t,t+h}
\end{aligned}$$

Let $h = 3$. Then, when the forecast is a SCE operator, $\widetilde{\mathbb{E}}^{\star}_{t-1}\hat{\pi}_{t+3}$ depends on $\hat{\pi}_{t-2}$. Moreover, the actual law of motion for inflation when $n = 1$ is

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

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

Then, $\mathbb{E}(v_{t,t+3}\hat{\pi}_{t-2})$ is a linear function of $\mathbb{E}(\hat{\pi}_{t-1}\hat{\pi}_{t-2})$, $\mathbb{E}(\hat{\pi}^2_{t-2})$ and $\mathbb{E}(\hat{\pi}_{t-3}\hat{\pi}_{t-2})$. After taking care of the omitted variable bias, we are left with $\mathbb{E}(v_{t,t+3}\hat{\pi}_{t-2}) = -\frac{a_1}{\rho}\sum_{i=0}^{5}\rho^{5-i}a_1^i\mathbb{E}(\hat{\pi}^2_{t-2}) \leq 0$. 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.

B.1 Testable Implications: Proofs

Proof of Proposition 1. The actual law of motion for inflation along the CE equilibrium is

$$\hat{\pi}_t = \mathbf{a}\hat{mc}_t + \mathbf{b}\hat{\pi}_{t-1} + u_t$$

Similarly, the forecast about inflation at forecasting horizon $h = 1$ along the equilibrium path is

$$\widetilde{\mathbb{E}}^{\star}_{t}\hat{\pi}_{t+1} = w_f n(\gamma^*)^2\hat{\pi}_{t-1}$$

Hence, the forecasting error in period $t + k$ following a positive one-time shock $\varepsilon_t > 0$ is

$$\hat{\pi}_{t+k} - \widetilde{\mathbb{E}}^{\star}_{t+k-1}\hat{\pi}_{t+k} = a\rho^{k-1}\left((b^2 - w_f n * (\gamma^*)^2)\sum_{j=0}^{k-2}\left(\frac{b}{\rho}\right) + \rho(b+\rho)\right)$$

The impact of ε_t is positive, hence there will be under-reaction of forecasters on impact. Moreover, $\lim_{k\to\infty}\rho^{k-1} = 0$ therefore the forecasting error will eventually dissipate at some point in the future. That combined with $a > 0$ implies that there will be delayed over-shooting if and only if

$$\lim_{k\to\infty}\left((\mathbf{b}^2 - w_f n(\gamma^*)^2)\sum_{j=0}^{k-2}\left(\frac{\mathbf{b}}{\rho}\right) + \rho(\mathbf{b}+\rho)\right) < 0$$

One can easily show that $(\mathbf{b}^2 - w_f n(\gamma^*)^2) < 0$. Then, if $\mathbf{b} > \rho$,

$$\lim_{k\to\infty}\left((b^2 - w_f n(\gamma^*)^2)\sum_{j=0}^{k-2}\left(\frac{\mathbf{b}}{\rho}\right) + \rho(\mathbf{b}+\rho)\right) = -\infty$$

On the other hand, if $\mathbf{b} < \rho$,

$$\lim_{k\to\infty}\left((\mathbf{b}^2 - w_f n(\gamma^*)^2)\sum_{j=0}^{k-2}\left(\frac{\mathbf{b}}{\rho}\right) + \rho(b+\rho)\right) = \frac{\rho(\rho^2 - w_f n(\gamma^*)^2)}{\rho - \mathbf{b}}$$

Hence, if $\mathbf{b} < \rho$, there will be delayed over-shooting if and only if $\rho^2 < w_f n(\gamma^*)^2$. The rest of the implications follow from here.

C Estimation

C.1 Data

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

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

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

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

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

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

where

- $GDPC1$ – Real GDP, Billions of Chained 2012 Dollars, Seasonally Adjusted Annual Rate.
- $POP_{index} = \frac{CNP160V}{CNP160V_{1992Q3}}$.
- $CNP160V$ – 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 More impulse response functions

Figure 8 presents the 3-dimensional IRF, projected over the [response - time] plane over time for the benchmark SAC-learning model.

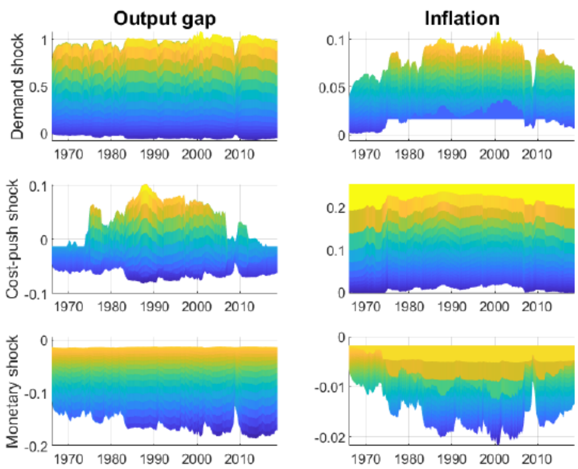


Figure 8: **Impulse response functions to a one standard deviation demand, cost-push and monetary shock in the benchmark AR(1)-SAC-learning model.**

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. *American Economic Review*, forthcoming.

[3] Angeletos, G. M., Z. Huo, K. A. Sastry (2020). Imperfect Macroeconomic Expectations: Evidence and Theory. *NBER Macro Annual*, forthcoming.

[4] 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.

[5] 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.

[6] Bils, M. and P. J. Klenow (2004). Some Evidence on the Importance of Sticky Prices. *Journal of Political Economy*, 112(5), 947-985.

[7] 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.

[8] 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.

[9] Bullard, J. and K. Mitra (2002). Learning About Monetary Policy Rules. *Journal of Monetary Economics*, 49(6), 1105-29.

[10] Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005). "Nominal rigidities and the dynamic effects of s shock to monetary policy." *Journal of Political Economy*, 113(1), 1-45.

[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] Eusepi, S. and B. Preston (2018). The Science of Monetary Policy: An Imperfect Knowledge Oerspective. *Journal of Economic Literature*, 56(1), 3-59.

[16] Evans, G.W. and S. Honkapohja (2001). Learning and Expectations in Macroeconomics. Princeton University Press, Princeton.

[17] Evans, G.W. and S. Honkapohja (2009). Learning and Macroeconomics. *Annual Economic Review*, 1, 421-51.

[18] Evans, G.W., C.H. Hommes, B. McGough, I. Salle (2019). Are Long-horizon Expectations (de-)Stabilizing? Theory and Experiments. *Bank of Canada.*

[19] Fuster, A., D. Laibson, B. Mendel (2010). Natural Expectations and Macroeconomic Dynamics. *Journal of Economic Perspectives*, 24(4), 67-84.

[20] Fuster, A., D. Laibson, B. Mendel (2012). Natural Expectations, Macroeconomic Dynamics and Asset Pricing. *NBER Macroeconomics Annual*, 26.

[21] Gabaix, X. (2020). A Behavioral New Keynesian Model. *American Economic Review*, 110(8):2271-2327.

[22] Ganong P. and P. J. Noel (2019). Consumer Spending During Unemployment: Positive and Normative Implications. *American Economic Review*, 109(7), 2383-2424.

[23] Gaus, E. and C. G. Gibbs (2018). Expectations and the empirical fit of DSGE models. Forthcoming in*Studies in Nonlinear Dynamics and Econometrics.*

[24] Geweke, J. (1999). Simulation Methods for Model Criticism and Robustness Analysis. In: Berger, J.O., Bernado, J.M., David, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, 6, 275-299.

[25] Greenwood, R. and A. Shleifer (2014). Expectations of Returns and Expected Returns. *Review of Financial Studies* 27(3), 714-46.

[26] Herbst, E. and F. Schorfheide (2015). Bayesian Estimation of DSGE Models. Princeton University Press.

[27] Hommes, C.H. and G. Sorger (1998). Consistent Expectations Equilibria. *Macroeconomic Dynamics*, 2(3), 287-321.

[28] Hommes, C.H. (2013). *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems.* Cambridge: Cambridge University Press.

[29] Hommes, C.H. and M. Zhu (2014). Behavioral Learning Equilibria. *Journal of Economic Theory*, 150, 778-814.

[30] Hommes, C.H. (2019). Behavioral and Experimental Macroeconomic and Policy Analysis: A Complex Systems Approach. Forthcoming in *Journal of Economic Literature.*

[31] 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.

[32] Kass, R. E. and A. E. Raftery (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773-795. University Press, 242-74.

[33] Maćkowiak, B. and M. Wiederholt (2009). Optimal Sticky Prices under Rational Inattention. *American Economic Review*, 99(3), 769-803.

[34] Malmendier, U. and S. Nagel (2016). Learning from Inflation Experiences. *Quarterly Journal of Economics*,131(1), 53-87.

[35] 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.

[36] Milani, F. (2005). A Bayesian DSGE Model with Infinite-Horizon Learning: Do "Mechanical" Sources of Persistence Become Superfluous? *International Journal of Central Banking.*

[37] Milani, F. (2007). Expectations, Learning and Macroeconomic Persistence. *Journal of Monetary Economics*, 54(7), 2065-2082.

[38] Milani, F. (2011). Expectation Shocks and Learning As Drivers of Business Cycle Fluctuations. *The Economic Journal*, 121(May), 379-401.

[39] Nimark, K. P. (2017). Dynamic Higher Order Expectations. *Cornell University mimeo.*

[40] 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.

[41] Petersen, L. (2015). Do Expectations and Decisions Respond to Monetary Policy? *Journal of Economic Studies*, 42(6), 972-1004.

[42] Petropoulos, F., N. Kourentzes, K. Nikolopoulos, E. Siemsen (2018). Judgmental Selection of Forecasting Models. *Journal of Operations Management*, 60, 34 - 46.

[43] Preston, B. (2005). Learning About Monetary Policy Rules When Long-Horizon Expectations Matter. *International Journal of Central Banking*, 2, 81-126.

[44] Reis, R. (2006). Inattentive Producers. *Review of Economic Studies*, 73(3), 793-821.

[45] 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.

[46] Sims, C. A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics*, 50(3), 665-690.

[47] 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.

[48] 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.

[49] 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.

[50] Smets, F. and R. Wouters (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3), 586-606.

[51] 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

[52] Tversky, A. and D. Kahneman (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5(2), 207-232.

[53] Tversky, A. and D. Kahneman (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-31.

[54] 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.

[55] Woodford, M. (2003). Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton University Press.