Expectations and the empirical fit of DSGE models

Eric Gaus[*] Christopher G. Gibbs[†]

September 2020

Abstract

This paper studies the improvement in empirical fit of dynamic stochastic general equilibrium (DSGE) models that assume adaptive learning in lieu of rational expectations (RE). The literature finds that estimated DSGE models with adaptive learning generate near universal improvements in fit, while inference on structural parameters is mostly unchanged. Improvements are attributed to the increased persistence generated by backward-looking expectations. We show, however, that improvements often result from altered cross-equation restrictions and not additional persistence assumptions. Nested comparisons of Euler-equation and infinite-horizon adaptive learning both significantly improve upon RE but only the latter's improvements are due to expectation formation. Bounded rationality assumptions offer an intuitive way to improve both in-sample and out-of-sample DSGE model fit. But our results suggest that learning models best-capture persistent deviation in beliefs from fundamentals rather than temporary deviations at business cycle frequencies.

JEL Classifications: E31; E32; D84; D83; C13

Key Words: Expectations; Adaptive learning; DSGE; Estimation.

[*]Gaus: Moody's Analytics (e-mail: egaus@gmail.com). Declarations of interest: none.

[†]Gibbs: University of Sydney. (e-mail: cgg1127@gmail.com). Declarations of interest: none. We would like to thank for comments and discussion Mariano Kulish, Bruce Preston, Stefano Eusepi, George Evans, Bruce McGough. Peter Ireland, James Morley, and Glenn Otto; Benaya Lie for research assistant support; seminar participants at the University of Queensland, Monash University, University of Melbourne, SUNY Binghamton, Fordham University, University of Birmingham, and the University of Sydney. We would also like to thank conference participants and organizers at the 2016 Expectations in Dynamic Macro Models Conference, the Society for Nonlinear Dynamics and Econometrics 2017 meetings, and the Society for Economic Measurement 2017 meetings. All errors are the fault of the authors.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models often struggle to endogenously reproduce the persistence observed in actual macroeconomic data. This has led many researchers to consider additional frictions, preference assumptions, or ad hoc adjustments within the standard rational expectations (RE) framework to increase persistence.[1] However, since evidence from forecasting surveys (see Coibion and Gorodnichenko 2015) and laboratory experiments (see Hommes 2013) often suggest deviations from rationality, many researchers have sought to generate this persistence directly through expectations by deviating from RE. For example, Milani (2006, 2007), Eusepi and Preston (2011), Del Negro and Eusepi (2011), Slobodyan and Wouters (2012a,b), Rychalovska (2016), Ormeño and Molnár (2015), Eusepi and Preston (2018), and Cole and Milani (2019) all consider adaptive learning with a constant gain and find a number of desirable properties including near universal improvements in model in-sample fit, the ability to capture survey forecasts of macroeconomic aggregates (Milani 2011; Ormeño and Molnár 2015; Cole and Milani 2019), or a lessened reliance in some cases on habit persistence or indexation in order to generate persistence (Milani 2006, 2007).

In this paper, we investigate the exact mechanisms that generate improvements in in-sample fit in estimated DSGE models using one of the most frequently studied bounded rationality modeling strategy: adaptive learning with a constant gain, also known as constant gain learning (CGL).[2] Specifically, we investigate whether improvements in fit in these models are always evidence for the theory of adaptive learning (belief updating) or if they are a consequence of other misspecification. One reason to think that it may be the latter is a striking consistency among the estimation results of many boundedly rational DSGE comparisons to RE in the literature with similarities that point towards misspecification rather than belief updating as the root cause. In particular, we note three stylized facts that emerge from these estimation studies:

1. Model fit significantly improves under adaptive learning for almost any specification

[1] The lack of persistence generated by DSGE models under RE is well-known. These modifications include ad hoc corrections such as adding lags of the endogenous variables to the structural equations (Galı and Gertler 1999 and Ireland (2004)), changes to preference such as habit persistence (Fuhrer 2000), changes to inflation setting by firm through inflation indexation (Cogley and Sbordone 2008), or adding information problems such as rule-of-thumb behavior (Amato and Laubach 2003), or rational inattention/sticky information (Mankiw and Reis 2002; Ball et al. 2005). Armed with subset of these modifications, Del Negro et al. (2007) declare that the New Keynesian model fits the data well enough to be used for policy evaluation.

[2] Though we have confined this paper to DSGE models and constant gain learning, other modeling frameworks might exhibit the same patterns. For example, Chow (1989) examines present value models and rejects rational expectations in favor of adaptive expectations.

of expectations considered.

2. The inference on the structural parameters of the model is mostly unchanged compared to inference under RE.

3. The gain parameters, which correspond to the persistence in expectations, are estimated to be relatively small.

To illustrate, Table 1 reports the parameter estimates from two of the most widely cited studies on estimated New Keynesian models with adaptive learning: Milani (2007) and Slobodyan and Wouters (2012b). Milani investigates adaptive learning in a small scale DSGE model, while Slobodyan and Wouters investigate learning in the medium scale DSGE model of Smets and Wouters (2007). The table reports some key parameter estimates from the two studies to allow a comparison between estimation results obtained under RE to those obtained under adaptive learning. We restrict attention to the case where the agents' perceived law of motion takes the functional form of minimum state variable (MSV) solution, although, similar results are often seen for other perceived law of motion specifications.

The model under constant gain learning exhibits a significant improvement in in-sample fit as measured by marginal log-likelihood relative to RE. Second, the difference between the key parameters that describe monetary policy and the exogenous shocks are small.[3] In fact, nearly all of the parameters estimates under CGL remain within the highest posterior density (HPD) interval of their RE counterparts. Finally, the gain parameters that govern how agents update beliefs in the learning algorithm are estimated to be relatively small.

The second stylized fact is often interpreted as evidence that persistence at a business cycle frequency explains the observed improvement in fit because small changes in the structural parameters are interpreted as the model fitting the data in the same way as under RE. However, the third fact - small estimated values for the gain parameters - complicates this interpretation. Small gain parameters such as these can imply extreme persistence in the learning process that goes far beyond business cycle frequencies.

For example, Figure 1 shows the time path under learning of the CGL-MSV case of Slobodyan and Wouters (2012b) for all beliefs initialized at steady state with the exception of inflation. For inflation, we assume that in the first quarter of 1948, the agents assume that steady state inflation rate is 0.5% higher than the actual value. As

[3]There is a transcription error between the working paper and published versions of Slobodyan and Wouters (2012b), where the 5% column of the HPD intervals is reported in the mean column. This explains the discrepancies between this table and the published version.

Table 1: Stylized facts of estimated NK models with CGL

Slobodyan and Wouters JEDC 2012

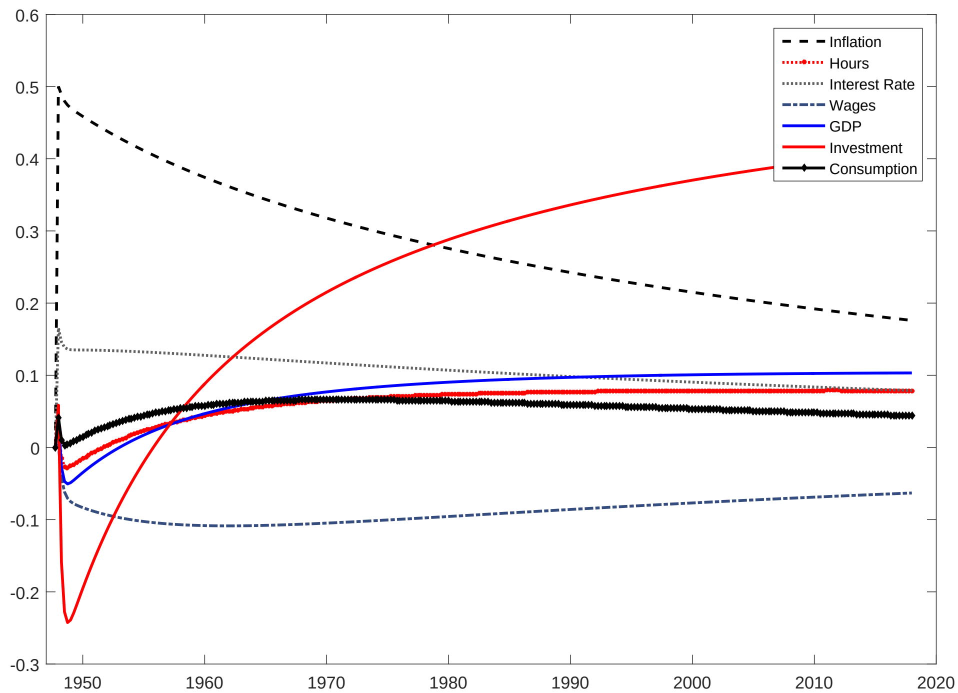
	RE	CGL-MSV	Difference
Monetary policy & habits			
MP inflation	2.04	1.91	0.13*
	[1.75, 2.33]	[1.58, 2.22	
MP output	0.09	0.13	-0.04*
	[0.05, 0.13]	[0.07, 0.18]	
MP output growth	0.22	0.19	0.03*
	[0.18, 0.27]	[0.15, 0.24]	
MP smoothing	0.81	0.84	-0.03*
	[0.77, 0.85]	[0.80, 0.88]	
Habits	0.71	0.80	-0.09
	[0.64, 0.78]	[0.75, 0.84]	
AR parameters			
Productivity	0.96	0.96	0.00*
	[0.94, 0.98]	[0.94, 0.99]	
Risk premium	0.22	0.23	-0.01*
	[0.08, 0.36]	[0.13, 0.32]	
Gov. spending	0.98	0.96	0.02*
	[0.96, 0.99]	[0.96, 0.99]	
Investment	0.71	0.45	0.26
	[0.62, 0.81]	[0.33, 0.56]	
MP shock	0.15	0.15	0.00*
	[0.04, 0.24]	[0.05, 0.26]	
Price mark-up	0.89	0.93	-0.04*
	[0.81. 0.97]	[0.88, 0.97]	
Wage mark-up	0.97	0.97	0.00*
	[0.95, 0.99]	[0.95, 0.99]	
St. Dev. Shocks			
Productivity	0.46	0.47	-0.01*
	[0.41, 0.51]	[0.42, 0.52]	
Risk premium	0.24	0.25	-0.01*
	[0.20, 0.28]	[0.22, 0.28]	
Gov. spending	0.53	0.53	0.00*
	[0.48, 0.58]	[0.48, 0.58]	
Investment	0.45	0.61	-0.16
	[0.37, 0.53]	[0.53, 0.68]	
MP shock	0.24	0.24	0.00*
	[0.22, 0.27]	[0.21, 0.26	
Price mark-up	0.14	0.14	0.00*
	[0.11, 0.17]	[0.12, 0.16]	
Wage mark-up	0.24	0.23	0.01*
	[0.21, 0.28]	[0.20, 0.26]	
Gains			
	-	0.017	
		[0.006,0.021]	
Marginal Likelihood			
	-922.75	-910.97	

Milani JME 2007

	RE	CGL - MSV	Difference
Monetary policy & habits			
MP inflation	1.433	1.484	-0.051*
	[1.06, 1.81]	[1.08, 1.90]	
MP output gap	0.792	0.801	-0.009*
	[0.425, 1.165]	[0.433, 1.18]	
MP smoothing	0.89	0.914	-0.024*
	[0.849,0.93]	[0.875, 0.947]	
Habits	0.911	0.117	0.794
	[0.717, 0.998]	[0.006, 0.289]	
AR Parameters			
Demand shock	0.87	0.845	0.025*
	[0.8,0.93]	[0.776, 0.908]	
Price mark-up	0.02	0.854	-0.834
	[0.0005, 0.07]	[0.778, 0.93]	
St. Dev Shocks			
MP Shock	0.933	0.86	0.073*
	[0.84, 1.04]	[0.777, 0.953]	
Demand shock	1.067	1.67	-0.603
	[0.89, 1.22]	[1.47, 1.91]	
Price mark-up	1.146	1.15	-0.004*
	[1.027, 1.27]	[1.02, 1.31]	
Gains			
		0.018	
		[0.0133,0.0231]	
Marginal Likelihood			
	-765.45	-759.08	

Notes: This table only reports a subset of the parameter estimates. Similar patterns are observed for the remaining parameters. The columns labeled "Difference" show the simple difference between the RE and CGL parameter estimates with asterisks denoting when the changes fall inside of the 95% HPD intervals of the RE estimate. The published version of Slobodyan and Wouters (2012b) does not report the HPD intervals for some estimates. We obtained these values from a working paper version dated 2009.

Figure 1: Small Gain Example



Notes: Return to steady state under constant gain learning in the Smets and Wouters model using Slobodyan and Wouters (2012b) posterior mode estimates (Column 2 in Figure 1).

the figure shows, the effects of such beliefs would still be felt today at the estimated value of the gain.[4] Chevillon and Mavroeidis (2017) recently highlight this feature of CGL to show that when gains are small, relative to sample size, that learning may actually generate long memory in the endogenous variables. This points to CGL as potentially explaining long run persistent expectation driven movements in macroeconomic data but not movements at typical business cycle frequencies.[5]

We show that one explanation for the three stylized facts lies in the separation of the estimation of structural parameters from the estimation of beliefs in CGL models.

[4]For the simulation, we set the variance-covariance matrix of the least squares algorithm to that of the relevant variances of each variable obtained under RE and we hold these values fixed.

[5]It is well-known that the choice of initial beliefs can have a significant impact on estimation results. Both Berardi and Galimberti (2017) and Slobodyan and Wouters (2012b) explore the effect of initial beliefs on model fit but neither study makes the connection between these effects and the overall role or lack thereof of time-variation in expectations for small gains.

Under RE, the structure of the model and beliefs are tightly linked. These linkages imply nonlinear cross-equation restrictions that enforce, for example, sign and zero restrictions on the model's predictions of the covariance and autocovariances of the observable data. Relaxing RE by assuming that beliefs are not tied to the structure of the model can significantly alter these restrictions allowing the model to better fit the data without increasing the number of freely estimated parameters or introducing new mechanisms to generate persistence. Furthermore, we show that this is a property of any linearized DSGE model.

To illustrate the point, we estimate a New Keynesian model following Ireland (2004) under five different expectations assumptions: RE, two common variants of CGL, and a restricted case of each learning model that prevents any time-variation in expectations, which we call fixed beliefs (FB). The fixed belief cases allow us to relax the RE assumption without introducing new freely estimated parameters or persistence through expectation. The two variants of CGL are Euler-equation CGL (EE-CGL) following Evans and Honkapohja (2001) and infinite-horizon CGL (IH-CGL) following Preston (2005). We perfectly nest all five expectational assumptions by using the MSV solution of the model for the perceived law of motion under CGL and by calibrating its initial value to that obtained from a full sample estimation of the model under RE. Therefore, if the RE model is the true model, then all five expectation assumptions yield the same in-sample fit and parameter estimates.

We compare the five different models' in-sample fit, real-time out-of-sample fit, and following McCallum (2001) by their variances and autocovariance functions. We find that the four bounded rationality cases largely generate results consistent with the three stylized facts. Comparing the two adaptive learning specifications to their fixed belief counterparts, we find no significant differences between the EE-CGL and EE-FB cases. This indicates that the relaxation of the RE restrictions alone explain the majority of the observed improvement in fit. We also show that similar results hold in the Smets and Wouters model. For the IH cases, however, we do find marginal increases in fit relative to the fixed belief case, substantive additional persistence generated via agents' beliefs, and different autocorrelation predictions for key endogenous variables when agents are learning, which indicates that the learning assumption materially adds to the predictions of the model beyond relaxing the RE restrictions. Therefore, only in the IH case do we conclude that learning makes a positive contribution towards fitting the data.

There are four takeaways from this exercise. First, the results actually provide support for considering bounded rationality assumptions since deviating from rationality clearly improves in-sample and out-of-sample model fit. Second, CGL appears best suited to explain long run drifts in beliefs rather than persistence at a business cycle frequency. This finding also explains why assuming simple forecasting rules as PLMs that specifically include lags of endogenous variables fit the data better. The lags needed to generate short run persistence are explicitly assumed in this case.[6] Third, our results demonstrate that comparisons between bounded rationality strategies and RE should be done with care. Model fit does not immediately imply evidence for a specific expectation assumption. Comparisons may say more about the misspecification of the model under RE.[7] Finally, it is important to stress that these conclusions are independent of issues relating to identification. Altering of cross-equation restrictions through different expectations assumptions changes the set of possible values a parameter may take on, which is distinct from whether the specific parameter is identifiable to the econometrician given the information available. In the exercises we consider, all parameters of interest are identified under the RE, FB, and CGL specifications but their implied relationship to the data through the model may differ greatly.

In the next section, we present examples of how bounded rationality strategies increase fit by altering cross-equation restrictions. In Section 3, we introduce a parsimonious DSGE model and explore the reduced form mappings implied under the five different expectations assumptions. In Section 4, we estimate the model under the different expectations assumptions and compare the results along the aforementioned dimensions. In Section 5, we explore EE-FB in a medium scale DSGE model and revisit the stylized facts. Section 6 concludes.

2 The effect of relaxing RE

To illustrate the effect of relaxing the RE assumption in a DSGE model, consider the following univariate example:

$$x_t = \alpha + \beta E_t(x_{t+1}) + w_t \tag{1}$$

where $w_t = \rho w_{t-1} + \varepsilon_t$ and $\varepsilon_t \sim N(0, \sigma)$. The model has the parameters α, β, ρ, and σ, which we refer to throughout the paper as the *structural parameters*. Under RE, the

[6]See, for example, Slobodyan and Wouters (2012a).

[7]This finding is consistent with other evidence from the DSGE-VAR literature (Cole and Milani 2019).

model has a MSV solution of the form

$$x_t = \mathfrak{a} + \mathfrak{b} w_t \tag{2}$$
$$w_t = \rho w_{t-1} + \epsilon_t, \tag{3}$$

which is characterised by the model's *reduced form parameters*: $\mathfrak{a}$ and $\mathfrak{b}$. The reduced form parameters are nonlinear combinations of the underlying structural parameters, where $\mathfrak{a} = \alpha/(1-\beta)$ and $\mathfrak{b} = 1/(1-\rho\beta)$. In addition, the structural parameters have restrictions imposed by theory. For example, $|\beta| < 1$ is required for determinacy and $-1 < \rho < 1$ is required for stationarity.

Typically, an econometrician is interested in obtaining estimates of the structural parameters while only observing x_t. The reduced form is a state space model and estimates of the structural parameters and the unobserved process ω_t may be obtained using likelihood-based techniques with the Kalman filter. Identification of the individual structural parameters under RE or learning requires calibrating some subset of parameters (such as β in this case). This setup leads to the following relationships between the reduced form parameters ($\mathfrak{a}$ and $\mathfrak{b}$), which if freely estimated would reflect the underlying correlations in the data, and the non-calibrated structural parameters of interest:

$$\alpha|_{RE} = \mathfrak{a}\,(1-\beta) \tag{4}$$
$$\rho|_{RE} = \frac{\mathfrak{b}-1}{\beta\mathfrak{b}} \tag{5}$$

From here it is clear that the structural parameters may be nonlinear combinations of the reduced form parameters, where bounds placed on the structural parameters by theory may limit the possible value that the reduced form parameters may take and hence the properties of the data the structural model can replicate.

Now consider the same model under adaptive learning of the MSV solution. Under adaptive learning, the agents take the same view point as an econometrician. The agents' perceived law of motion for the economy is given by Equations (2) and (3), but $\mathfrak{a}$ and $\mathfrak{b}$ are assumed to be unknown to the agents. Agents estimate $\mathfrak{a}$ and $\mathfrak{b}$ via recursive least squares with a constant gain:

$$\begin{pmatrix} \mathbf{a}_t \\ \mathbf{b}_t \end{pmatrix} = \begin{pmatrix} \mathbf{a}_{t-1} \\ \mathbf{b}_{t-1} \end{pmatrix} + \gamma \mathbf{R}_t^{-1} \begin{pmatrix} 1 \\ w_t \end{pmatrix} \left((1 \quad w_t) \begin{pmatrix} \mathbf{a}_{t-1} \\ \mathbf{b}_{t-1} \end{pmatrix} - x_t \right) \tag{6}$$
$$\mathbf{R}_t = \mathbf{R}_{t-1} + \gamma \left(w_t' w_t - \mathbf{R}_{t-1} \right), \tag{7}$$

where $0 < \gamma < 1$; $\mathbf{a}_0$, $\mathbf{b}_0$, and $\mathbf{R}_0$ are appropriate initial values of the recursion, and $\mathbf{R}_t$ is the estimated variance-covariance matrix. Substituting these beliefs into Equation (1), the actual law of motion for the economy is given by

$$
\begin{aligned}
x_t &= \mathfrak{a}_{t-1} + \mathfrak{b}_{t-1} w_t && (8) \\
w_t &= \rho w_{t-1} + \epsilon_t && (9)
\end{aligned}
$$

plus Equations (6) and (7), where $\mathfrak{a}_{t-1} = \alpha + \beta \mathbf{a}_{t-1}$ and $\mathfrak{b}_{t-1} = (\beta \mathbf{b}_{t-1} \rho + 1)$. The parameters of interest are now α, β, ρ, σ, and γ. The model again may be estimated using likelihood based techniques just like RE.

From the econometrician's perspective, adaptive learning allows for time-variation in the reduced form parameters of the model, which may capture more complicated dynamics present in the data. But it also changes the mapping from the reduced form parameters to the key structural parameters given by Equations (4) and (5). To see how, consider the case where no time-variation in beliefs is allowed by setting $\gamma = 0$, which fixes $\mathfrak{a}_t = \mathbf{a}_0 = a^{FB}$ and $\mathfrak{b}_t = \mathbf{b}_0 = b^{FB}$ to their initial values. This implies the following relationship between the structural parameters and the reduced form

$$
\begin{aligned}
\alpha|_{FB} &= -\beta a^{FB} + \mathfrak{a} && (10) \\
\rho|_{FB} &= \frac{\mathfrak{b} - 1}{\beta b^{FB}}. && (11)
\end{aligned}
$$

Both structural parameters of interest are now linear in the reduce form parameters.

If one compares Equations (4) and (5) to (10) and (11), then conditional on a^{FB} and b^{FB} being set to their RE values, the two cases are equivalent and the structural parameters are the same. However, if these parameters are set differently, or there is misspecification under RE so that the model does not well-capture the data, then the fixed belief case allows both the structural parameters and the reduced form parameters to take on different values.

For example, consider the case where x_t is US inflation. Stock and Watson (2007) show that inflation is well-describes by an IMA(1,1) process. If this process is the true data generating process, then x_t evolves as

$$
\begin{aligned}
x_t &= \tau_t + \epsilon_{x,t} \\
\tau_t &= \tau_{t-1} + \eta_t,
\end{aligned}
$$

where $\epsilon_{x,t}$ and η_t are the exogenous shocks. Here, the RE, FB, and CGL models are all misspecified. If the econometrician allows for measurement error, then the reduced form state space model given (2) and (3) would still nest this data generating process with $\mathfrak{a} = 0$, $\mathfrak{b} = 1$, and $\rho = 1$. But these values for $\mathfrak{a}$, $\mathfrak{b}$, and ρ are infeasible under the restrictions imposed by the RE structural models.[8]

In practice, the misspecification creates a tradeoff for fitting the proposed structural models to the data between matching persistence of x_t and matching the observed in-sample variance of x_t. Assuming the in-sample variance of x_t is sufficiently small for illustrative purposes, we can see how this tradeoff would work by rewriting Equations (2) and (3) as

$$x_t = (1-\rho)\mathfrak{a} + \rho x_{t-1} + \mathfrak{b}\varepsilon_t. \tag{12}$$

Under RE, the closer the estimated value of ρ is to its true value of 1, the larger is $\mathfrak{b} = 1/(1-\rho\beta)$. Therefore, higher values of ρ allow the model to better match the persistence at the cost of inflating the variance of the shock and vice versa.

2.1 Empirical implications in a multivariate setting

Now suppose we are interested in estimating the structural parameters of an endowment economy with a monetary authority that adjusts the nominal interest rate in response to changes in inflation. We can characterize inflation using the monetary policy rule and the Fisher equation:

$$i_t = \phi_\pi \pi_t + \epsilon_t \tag{13}$$

$$i_t = E_t \pi_{t+1} + r_t, \tag{14}$$

where r_t is the exogenous real rate of return that follows an AR(1) process, $r_t = \theta r_{t-1} + \eta_t$, and ϵ_t is an i.i.d monetary policy shock. The inflation process can be written in a similar form as (1),

$$\pi_t = \frac{1}{\phi_\pi} E_t \pi_{t+1} + \frac{1}{\phi_\pi} r_t, \tag{15}$$

[8]For example, using Equation (5), $\mathfrak{b} = 1$ implies ρ must be equal to zero.

which implies the same reduce form for π_t under RE and FB

$$\pi_t = \mathfrak{a} + \mathfrak{b} r_t. \tag{16}$$

In this case, both the interest rate and inflation are observable, which means that regardless of the expectation assumption, the parameter ϕ_π is pinned down by the monetary policy rule (13) and the contemporaneous correlation of the interest rates and inflation. The relationship between the remaining structural parameters and the reduced form, however, are different. Under rational expectations, the reduced form is $\mathfrak{a} = 0$ and $\mathfrak{b} = 1/(\phi_\pi - \theta)$, while under fixed beliefs it is $\mathfrak{a} = a^{FB}/\phi_\pi$ and $\mathfrak{b} = \frac{1+b^{FB}\theta}{\phi_\pi}$. For both expectations assumptions, there is a different relationship between $\mathfrak{b}$ and the persistence parameter θ. Depending on how b^{FB} is chosen and the actual correlations in the data, $\hat{\mathfrak{b}}$, the implied persistence of the two models may differ.

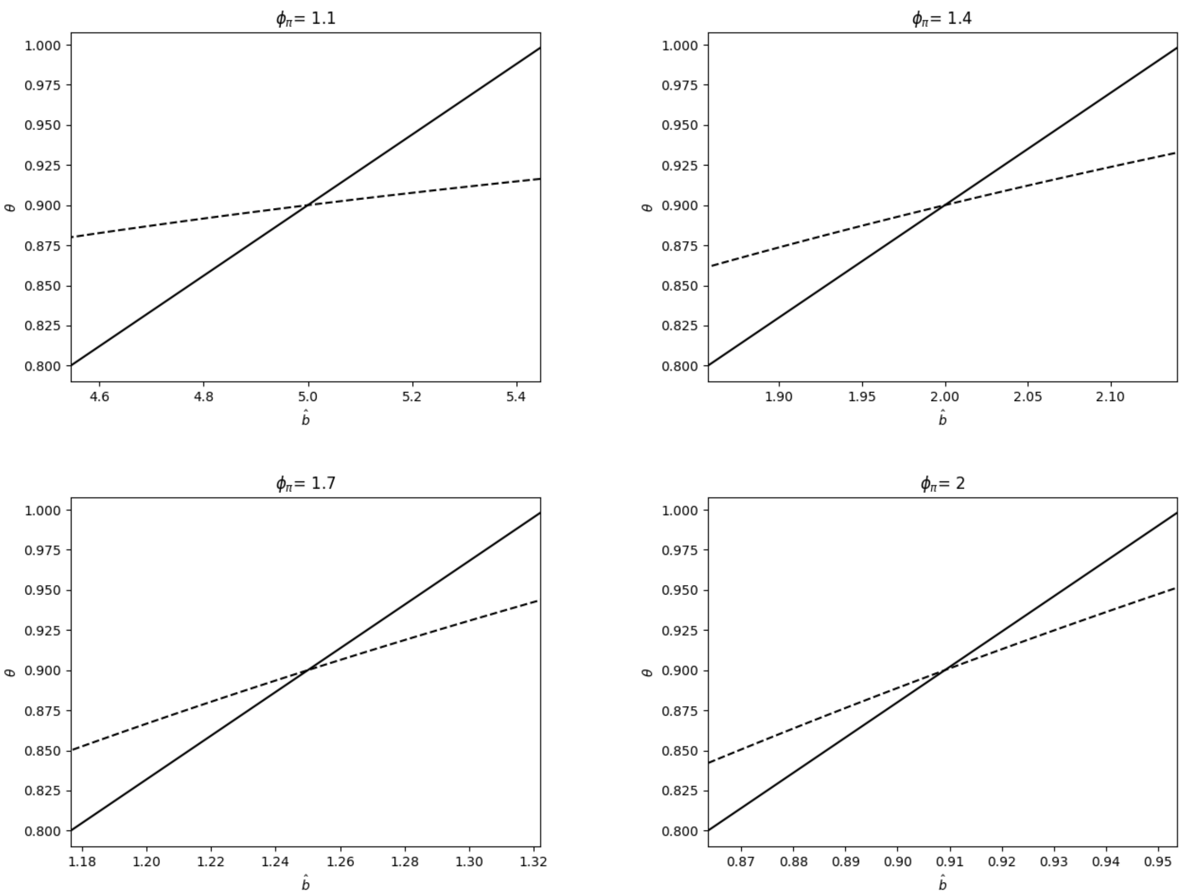
Figure 2 shows what happens to the range of permissible values of θ when we vary the value of $\hat{\mathfrak{b}}$, where b^{FB} is set to an RE solution with $\theta = 0.9$, which mimics the practice of initializing beliefs to RE estimates from a pre-sample. For the same range of $\hat{\mathfrak{b}}$, there is a wider range of possible structural parameters under FB. The FB formulation adds no additional persistence as would be the case under learning. But, given the same data, the estimated persistence may be different, while other structural parameters remain the same.

This illustrates a mechanism that may explain improvement in fit in models that deviate from RE. The interconnectedness of structural parameters through the non-linear cross-equation restrictions is significantly altered under non-rational expectations. As a consequence, such models may fit data better irrespective of the economic theory which motivates the use of the bounded rationality assumption. The fact that these changes are nonlinear means that what appear to be small changes in the structural parameters from the econometrician's perspective, may imply significant changes in the model's predictions for the observable data, which explains why fit can improve when parameter estimates appear to remain largely the same.

2.2 The general case

We can generalize this insight to any first-order approximated DSGE model. Consider the linearized structural equations of a DSGE model, which we write as

Figure 2: Reduced Form to Structural Form Range



Notes: Parameter $\theta = 0.9$ for the FB belief calibration. Solid line is FB, dashed line is RE.

$$\mathbf{y}_t = \mathbf{\Gamma} + \mathbf{A}\mathbf{y}_{t-1} + \mathbf{B}\mathbb{E}_t y_{t+1} + \mathbf{D}\mathbf{v}_t + \mathbf{K}\varepsilon_t \tag{17}$$

$$\mathbf{v}_t = \mathbf{R}\mathbf{v}_{t-1} + \mathbf{u}_t, \tag{18}$$

where $\mathbf{y}_t$ is a vector endogenous variables and $\mathbf{v}_t$ is vector of autoregressive exogenous shocks. The minimum state variable RE solution of the model takes the following form

$$\mathbf{y}_t = \mathbb{C} + \mathbb{F}\mathbf{y}_{t-1} + \mathbb{Q}\mathbf{v}_t + \mathbb{K}\varepsilon_t. \tag{19}$$

Collect the parameters of interest into the vector $\mathbf{\Theta}$. The RE restrictions can be summarized by the mapping $F : \mathbf{\Theta} \to \{\mathbb{C}, \mathbb{F}, \mathbb{Q}\}$, which governs how changes in the elements of $\mathbf{\Theta}$ change $\mathbb{C}$, $\mathbb{F}$, or $\mathbb{Q}$. In general, the mapping is highly nonlinear and often has no closed form solution. The MSV solution has a state space representation and inference on $\mathbf{\Theta}$ may be obtained by maximum likelihood. To this aim, let

$$\mathbf{\Theta}^{MLE} = \text{argmax}_{\mathbf{\Theta}} L(\mathbf{\Theta}|\mathbf{y}^{obs}, F),$$

where L is the likelihood function, $\mathbf{y}^{obs}$ is observable data, and F is the aforementioned mapping.

Now consider the case of boundedly rational agents with fixed beliefs. As in the simple example, suppose that beliefs are of the form of Equation (19), where $\mathbf{C}^{\mathbf{FB}}_{|\mathbf{\Theta}^{\mathbf{MLE}}}$, $\mathbf{F}^{\mathbf{FB}}_{|\mathbf{\Theta}^{\mathbf{MLE}}}$, and $\mathbf{Q}^{\mathbf{FB}}_{|\mathbf{\Theta}^{\mathbf{MLE}}}$ are formed using the ML estimates obtained from the RE model. Under this assumption, the data generating process for the economy takes the following form

$$\begin{aligned}\mathbf{y}_t = {} & \underbrace{(\mathbf{I} - \mathbf{B}\mathbf{F}^{FB})^{-1}(\mathbf{\Gamma} + \mathbf{B}\mathbf{C}^{\mathbf{FB}})}_{\mathbb{C}} + \underbrace{(\mathbf{I} - \mathbf{B}\mathbf{F}^{FB})^{-1}\mathbf{A}}_{\mathbb{F}}\,\mathbf{y}_{t-1} \\ & + \underbrace{(\mathbf{I} - \mathbf{B}\mathbf{F}^{FB})^{-1}(\mathbf{B}\mathbf{Q}^{\mathbf{FB}}\mathbf{R} + \mathbf{D})}_{\mathbb{Q}}\,\mathbf{v}_t + \mathbb{K}\varepsilon_t,\end{aligned} \tag{20}$$

where Equation (19) is substituted in for $\mathbb{E}_t y_{t+1}$. Note the following regarding Equation (20):

1. Equation (20) retains the same reduced form as Equation (19).

2. Conditioning on $\mathbf{C}^{\mathbf{FB}}$, $\mathbf{F}^{\mathbf{FB}}$, and $\mathbf{Q}^{\mathbf{FB}}$, Equation (20) is described by the same structural parameters as Equation (19), which we collect in the vector Ξ.

For the econometrician, the mapping of interest is now $G : \boldsymbol{\Xi} \times \boldsymbol{\Theta}^{MLE} \to \{\mathbb{C}, \mathbb{F}, \mathbb{Q}\}$. The question here is whether $\boldsymbol{\Xi}^{MLE}$ from Equation (20) is equal to $\boldsymbol{\Theta}^{MLE}$ from Equation (19) and whether

$$\max_{\boldsymbol{\Theta}} \mathbb{E}L(\boldsymbol{\Theta}|\mathbf{y}^{obs}, F) = \max_{\boldsymbol{\Xi}} \mathbb{E}L^*(\boldsymbol{\Xi}|\mathbf{y}^{obs}, \mathbf{G}, \boldsymbol{\Theta}^{MLE}).$$

Consider the case where the true data generating process is projected onto an unrestricted reduced form that nests the RE solution

$$\mathbf{y}_t = \mathbb{C}^T + \mathbb{F}^T \mathbf{y}_{t-1} + \mathbb{Q}^T \mathbf{v}_t + \mathbb{K}^T \varepsilon_t, \tag{21}$$

where in the event that Equation (19) is correctly specified, it is equivalent to the true data generating process.

Theorem 1: *Given Equation (21) and the mappings* $F : \boldsymbol{\Theta} \to \{\mathbb{C}, \mathbb{F}, \mathbb{Q}\}$ *and* $G : \boldsymbol{\Xi} \times \boldsymbol{\Theta}^{MLE} \to \{\mathbb{C}, \mathbb{F}, \mathbb{Q}\}$ *such that* $\boldsymbol{\Theta}^{MLE} = argmax_{\boldsymbol{\Theta}} L(\boldsymbol{\Theta}|\mathbf{y}^{obs}, F)$,

a. if $\boldsymbol{\Theta}^* = F^{-1}(\{\mathbb{C}^T, \mathbb{F}^T, \mathbb{Q}^T\})$ *exists and is in the feasible parameter space, i.e. satisfies determinacy or other theoretically imposed restrictions, then*

$$\boldsymbol{\Theta}^* = argmax_{\boldsymbol{\Theta}} \mathbb{E}L(\boldsymbol{\Theta}|\mathbf{y}^{obs}, F) = argmax_{\boldsymbol{\Xi}} \mathbb{E}L^*(\boldsymbol{\Xi}|\mathbf{y}^{obs}, G, \boldsymbol{\Theta}^{MLE})$$

and

$$\max_{\boldsymbol{\Theta}} \mathbb{E}L(\boldsymbol{\Theta}|\mathbf{y}^{obs}, F) = \max_{\boldsymbol{\Xi}} \mathbb{E}L^*(\boldsymbol{\Xi}|\mathbf{y}^{obs}, G, \boldsymbol{\Theta}^{MLE})$$

b. if $\boldsymbol{\Theta}^* = F^{-1}(\{\mathbb{C}^T, \mathbb{F}^T, \mathbb{Q}^T\})$ *does not exist or, exists but is not in the feasible parameter space, then*

$$argmax_{\boldsymbol{\Theta}} \mathbb{E}L(\boldsymbol{\Theta}|\mathbf{y}^{obs}, F) \neq argmax_{\boldsymbol{\Xi}} \mathbb{E}L^*(\boldsymbol{\Xi}|\mathbf{y}^{obs}, G, \boldsymbol{\Theta}^{MLE})$$

and

$$\max_{\boldsymbol{\Theta}} \mathbb{E}L(\boldsymbol{\Theta}|\mathbf{y}^{obs}, F) \leq \max_{\boldsymbol{\Xi}} \mathbb{E}L^*(\boldsymbol{\Xi}|\mathbf{y}^{obs}, G, \boldsymbol{\Theta}^{MLE})$$

Proof. For part a, if $\boldsymbol{\Theta} = \boldsymbol{\Xi}$, then by construction Equation (20) is equal to Equation

(19), $\mathbf{\Xi}^{MLE}$ from Equation (20) is equal to $\mathbf{\Theta}^{MLE}$ from Equation (19), and

$$\max_{\mathbf{\Theta}} \mathbb{E}L(\mathbf{\Theta}|\mathbf{y}^{obs}, F) = \max_{\mathbf{\Xi}} \mathbb{E}L^*(\mathbf{\Xi}|\mathbf{y}^{obs}, \mathbf{G}, \mathbf{\Theta}^{MLE}).$$

For part b, the non-equivalence of the maximum likelihood coefficient estimates is straightforward. We obtain the inequality in likelihoods by noticing that the range of mapping F is, by construction, a subset of the range of G because $F(\mathbf{\Theta}) \coloneqq G(\mathbf{\Theta}, \mathbf{\Theta})$. □

The relaxation considered here is the starting point for most models of bounded rationality. Therefore, without adding extra structural parameters that are freely estimated or allowing for time-varying beliefs, the model becomes less restricted, while maintaining the same functional form, i.e. Equation (21), Equation (20), and Equation (19) continue to nest one another. This, of course, does not imply that further modification such as assuming time-variation in $\mathbb{C}$, $\mathbb{F}$, and $\mathbb{Q}$ as in adaptive learning will not better fit the data, but it highlights another mechanism through which boundedly rational models may improve in-sample fit without any time-variation in the reduced form parameters. In the remainder of the paper, we specifically ask what degree is this mechanism is empirically relevant.

3 Expectations and the reduced form

We consider five different assumptions for expectations that share a common reduced form:

$$\mathbf{y}_t = \mathbb{C}_t + \mathbb{F}_t y_{t-1} + \mathbb{Q}_t \mathbf{v}_t + \mathbb{K}\varepsilon_t, \tag{22}$$

Each assumption imposes different restrictions on $\mathbb{C}_t$, $\mathbb{F}_t$, and $\mathbb{Q}_t$. In this section, we put forward a tractable model that allows for analytical derivation of the mapping from the structural parameters to the reduced form so that changes implied by different expectations assumptions can be studied in detail. Besides the RE approach, we consider two variations of adaptive learning and their respective FB counterparts.

The first adaptive learning approach we consider follows Evans and Honkapohja (2001) by assuming Euler-equation constant gain learning.[9] In EE-CGL agents are only asked to form one-step-ahead forecasts and they ignore any implications of their forecasts for longer horizons. The appeal of this approach is that if the chosen forecasting process nests RE, then under well-known regularity conditions (E-stability) the agents' beliefs

[9] This is also the approach most often taken in the behavioral heterogeneous expectation literature as in Hommes (2013) and the references therein.

will stay in the neighborhood of the RE solution. Therefore, it allows for potentially complicated but bounded beliefs around the natural RE benchmark in a wide range of models.

The second approach we consider is infinite-horizon learning following Preston (2005), which asks agents to consider the implications of their beliefs on their entire decision problem. In many cases, this means solving out an agent's decision rule, given beliefs today, into the infinite future. The advantage of this approach is that decisions conditional on beliefs are consistent with the micro-foundations of the model at every point in time, which is not always true in the EE specifications. Like its EE counterpart, though, under well-known regularity conditions beliefs under constant gain learning will depart from those predicted under RE but in the long run tend towards the RE solution.

The two FB counterparts for the IH and EE specifications can be thought of as restricted cases, where the gain parameter in the learning algorithm is set to zero and not estimated. But it is important to distinguish these cases as standalone expectation assumptions because algebraically they are similar to other bounded rationality assumptions considered in the literature. For example, when expectations are assumed to be generated by fixed parameter VARs or other simple time series models as considered in Cornea-Madeira et al. (2017) or Cole and Milani (2019). Therefore, the impact on fit that these strategies have relative to RE are informative beyond MSV adaptive learning exercises.

3.1 The model

We study a parsimonious version of the New Keynesian model proposed by Ireland (2004). The microfoundations of the model are given in Appendix A. Following Preston (2005), the key equations under an unspecified expectations operator $\hat{\mathbb{E}}_t$ are

$$\begin{aligned} x_t &= \bar{r}(1-\beta)^{-1} - \omega\hat{a}_t \\ &\quad + \hat{\mathbb{E}}_t \sum_{T=t}^{\infty} \beta^{T-t}\left[(1-\beta)(x_{T+1} + \omega\rho_a\hat{a}_T) - (i_T - \hat{\pi}_{T+1}) - (\rho_a - 1)\hat{a}_T\right] && (23) \\ \pi_t &= \frac{1-\beta}{1-\lambda_1\beta}\bar{\pi} + \psi x_t - e_t + \lambda_1\beta\hat{\mathbb{E}}_t \sum_{T=t}^{\infty}(\lambda_1\beta)^{T-t}\left(\frac{1-\lambda_1}{\lambda_1}\pi_{T+1} + \psi x_{T+1} - e_{T+1}\right) && (24) \\ i_t &= \bar{r} + \bar{\pi} + \theta_\pi(\pi_t - \bar{\pi}) + \theta_x x_t + \epsilon_{i,t} && (25) \\ g_t &= \hat{y}_t - \hat{y}_{t-1} + \bar{g} + \epsilon_{z,t} && (26) \\ x_t &= \hat{y}_t - \omega a_t && (27) \\ a_t &= \rho_a a_{t-1} + \epsilon_{a,t} && (28) \\ e_t &= \rho_e e_{t-1} + \epsilon_{e,t}, && (29) \end{aligned}$$

where x_t is the output gap, π_t is inflation, i_t is the nominal interest rate, g_t is the growth rate of output, y_t is the stochastically detrended level of output, a_t is a preference shock, e_t is a cost push shock, $\epsilon_{i,t}$ is a monetary policy shock, and $\epsilon_{z,t}$ is the detrended TFP shock. The individual variables are in log terms with their log steady state values written out explicitly such that at steady state $x_t = \bar{x} = 0$, $\pi_t = \bar{\pi}$, $i_t = \bar{r} + \bar{\pi}$, and $g_t = \bar{g}$.

We choose this model because it has no internal propagation mechanisms other than expectations. In fact, the minimum state variable RE solution does not depend on any lagged endogenous variables. This has the added benefit of allowing us to estimate the model without a projection facility, which is usually required to keep expectations from becoming explosive during numerical optimization of the likelihood function. Gaus and Ramamurthy (2012) note that the choice of projection facility can have a significant effect on estimation outcomes, which may complicate a comparison with RE. In addition, the determinacy and E-stability conditions of the model perfectly coincide in all cases.[10] This allows us to impose identical restrictions on the parameter space for all estimated versions of the model.

3.1.1 The state space

The model under unspecified expectations has the following state space representation

[10] The determinacy and E-stability condition for the model is the Taylor principle: $\theta_\pi > \frac{(\beta-1)\theta_x+\psi}{\psi}$.

$$
\begin{aligned}
\mathbf{y}_t^{obs} &= \mathbf{H}X_t \\
\mathbf{X}_t &= \mathbf{J}_{t-1} + \mathbf{M}_{t-1}\mathbf{X}_{t-1} + \mathbf{N}_{t-1}\varepsilon_t
\end{aligned}
$$

where $\mathbf{J}_{t-1} = \mathbf{\Omega}_{t-1}^{-1}\mu_{t-1}$, $\mathbf{M}_{t-1} = \mathbf{\Omega}_{t-1}^{-1}\mathbf{\Psi}$, $\mathbf{N}_{t-1} = \mathbf{\Omega}_{t-1}^{-1}\zeta$,

$$
\mu_{t-1} = \begin{pmatrix} \mathbb{C}_t \\ \bar{r} + (1-\theta_\pi)\bar{\pi} \\ \bar{g} \\ \mathbf{0_{3\times 1}} \end{pmatrix}, \quad \mathbf{\Omega}_{t-1} = \begin{pmatrix} \mathbf{I_{2\times 2}} & & \mathbf{0_{2\times 3}} & & & -\mathbb{Q}_t & \\ -\theta_x & -\theta_\pi & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & -\omega & 0 \\ & & \mathbf{0_{2\times 5}} & & & \mathbf{I_{2\times 2}} & \end{pmatrix}
$$

$$
\Psi = \begin{pmatrix} & & & \mathbf{0_{3\times 7}} & & & \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_a & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho_e \end{pmatrix}, \quad \zeta = \begin{pmatrix} \mathbf{0_{1\times 2}} & -1 & \mathbf{0_{1\times 4}} \\ & \mathbf{0_{1\times 7}} & \\ \mathbf{0_{1\times 2}} & 1 & \mathbf{0_{1\times 4}} \\ \mathbf{0_{1\times 3}} & 1 & \mathbf{0_{1\times 3}} \\ & \mathbf{0_{1\times 7}} & \\ \mathbf{0_{2\times 5}} & & \mathbf{I_{2\times 2}} \end{pmatrix},
$$

$\mathbf{y}_t^{obs} = (x_t, \pi_t, i_t, g_t)'$, and $\mathbf{X}_t = (x_t, \pi_t, i_t, g_t, y_t, a_t, e_t)'$. The expectation assumption chosen by the researcher directly restrict the possible values of $\mathbb{C}_t$ and $\mathbb{Q}_t$, which become nonlinear functions of the structural parameters of the model. In what follows, we show how the different expectation assumption imply different relationships between $\mathbb{C}_t$ and $\mathbb{Q}_t$ and the structural parameters of interest.

3.1.2 Rational expectations

It is straightforward to show that imposing RE on Equations (23) and (24) allows them to be collapsed to a more familiar form:

$$
\begin{aligned}
x_t &= \bar{r} + \mathbb{E}_t^{RE} x_{t+1} - (i_t - \tilde{\mathbb{E}}_t^{RE}\pi_{t+1}) + (1-\omega)(1-\rho_a)a_t && (30) \\
\pi_t &= (1-\beta)\bar{\pi} + \beta\tilde{\mathbb{E}}_t^{RE}\pi_{t+1} + \psi x_t - e_t. && (31)
\end{aligned}
$$

Substituting in Equation (25), we can map the model into the general form of Equations (17) and (18)

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{\Gamma} + \mathbf{A}\mathbf{y}_{t-1} + \mathbf{B}\tilde{\mathbb{E}}_t y_{t+1} + \mathbf{D}\mathbf{v}_t + \mathbf{K}\varepsilon_t \\
\mathbf{v}_t &= \rho \mathbf{v}_{t-1} + \mathbf{u}_t,
\end{aligned}$$

where $\mathbf{y}_t = (x_t,\ \pi_t)'$, $\mathbf{v}_t = (a_t, e_t)'$, $\mathbf{u}_t = (\epsilon_{a,t}, \epsilon_{e,t})'$, $\varepsilon_t = (\epsilon_{i,t}, 0)'$,

$$\mathbf{\Gamma} = m \begin{pmatrix} \bar{\pi}(\theta_\pi \beta - 1) \\ -\bar{\pi}((\theta_x(\beta - 1) - 1 + \beta + \psi - \theta_\pi \psi) \end{pmatrix}, \ \mathbf{B} = m \begin{pmatrix} 1 & 1 - \theta_\pi \beta \\ \psi & \beta(1 + \theta_x) + \psi \end{pmatrix},$$

$$\mathbf{D} = m \begin{pmatrix} (\rho_a - 1)(\omega - 1) & \theta_\pi \\ (\rho_a - 1)(\omega - 1)\omega & -1 - \theta_x \end{pmatrix}, \ \rho = \begin{pmatrix} \rho_a & 0 \\ 0 & \rho_e \end{pmatrix}$$

$m = (1 + \theta_x + \theta_\pi \psi)^{-1}$, $\mathbf{A} = \mathbf{0}_{2\times 2}$, and $\mathbf{K} = -\mathbf{I}_2$. Using the reduced form of Equation (22) and the method of undetermined coefficient, we can solve analytically for the RE solution in terms of the structural parameters: $\mathbb{C}^{RE} = (0, \bar{\pi})'$,

$$\mathbb{Q}^{RE} = \begin{pmatrix} -\frac{(\rho_a-1)(\beta\rho_a-1)(\omega-1)}{1+\theta_x-\theta_x\beta\rho_a+\beta\rho_a^2+\theta_\pi\psi-\rho_a(1+\beta+\psi)} & \frac{\theta_\pi-\rho_e}{1+\theta_x-\theta_x\beta\rho_e+\beta\rho_e^2+\theta_\pi\psi-\rho_e(1+\beta+\psi)} \\ \frac{(\rho_a-1)\psi(\omega-1)}{1+\theta_x-\theta_x\beta\rho_a+\beta\rho_a^2+\theta_\pi\psi-\rho_a(1+\beta+\psi)} & \frac{\rho_e-1-\theta_x}{1+\theta_x-\theta_x\beta\rho_e+\beta\rho_e^2+\theta_\pi\psi-\rho_e(1+\beta+\psi)} \end{pmatrix}, \qquad (32)$$

$\mathbb{F}^{RE} = \mathbf{0}_{2\times 2}$, and $\mathbb{K}^{RE} = -\mathbf{I}$. This is of course the explicit mapping $F : \mathbf{\Theta} \to \{\mathbb{C}, \mathbb{F}, \mathbb{Q}\}$ discussed in Section 2.

There are two notable features of this mapping. First, RE places restrictions on the intercept term $\mathbb{C}^{RE}$ forcing them to be consistent with the steady state of the model, which will not be the case under the other expectation assumptions. Second, the mapping from the structural parameters to the reduced form of $\mathbb{Q}^{RE}$ is highly nonlinear. The mapping to each element of $\mathbb{Q}^{RE}$ is an eight degree polynomial in β, ψ, θ_π, θ_x, ρ_a, ρ_e, and ω. The nonlinearity of this mapping is important because it means that relatively small changes in the values of structural parameters can have large effects on the reduced form and vice versa.

3.1.3 Euler-equation learning and fixed beliefs

The next expectation assumptions we consider are EE-CGL and EE-FB. To implement these strategies, we start with the same structural equation as under RE given by Equations (17) and (18). We assume that the agents perceived law of motion (PLM) takes the same functional form of Equation (22)

$$\mathbf{y}_t = \mathbf{C}^{EE} + \mathbf{Q}^{EE}\mathbf{v}_t + \varepsilon_t. \tag{33}$$

The agents estimate their belief parameters using past data by a constant gain recursive least squares algorithm

$$\begin{aligned} \mathbf{\Phi}_t &= \mathbf{\Phi}_{t-1} + \gamma \mathbf{S}_t^{-1}\mathbf{z}_{t-1}(\mathbf{y}_t - 1 - \mathbf{z}'_{t-1}\mathbf{\Phi}_{t-1}) & (34)\\ \mathbf{S}_t &= \mathbf{S}_{t-1} + \gamma(\mathbf{z}_{t-1}\mathbf{z}'_{t-1} - \mathbf{S}_{t-1}), & (35) \end{aligned}$$

where $\mathbf{z}_t$ is vector of data, $\mathbf{\Phi}_t$ is a vector of regression coefficients, $\mathbf{S}_t$ is the estimated variance-covariance matrix, and γ is a matrix of gain parameters that govern the weight placed on new information.

Expectations under EE-CGL at time t are given by[11]

$$\mathbb{E}_t^{EE}\mathbf{y}_{t+1} = \mathbf{C}_{t-1}^{EE} + \mathbf{Q}_{t-1}^{EE}\mathbf{R}\mathbf{v}_t.. \tag{36}$$

Substituting Equation (36) in for the beliefs in Equation (17) yields the following mapping to the reduced form equations:

$$\mathbb{C}_t^{EE} = \mathbf{\Gamma} + \mathbf{B}\mathbf{C}_{t-1}^{EE} \quad = \quad m \begin{pmatrix} C_{t-1}^{EE,11} + (C_{t-1}^{EE,21} - \bar{\pi})(1 - \beta\theta_\pi) \\ C_{t-1}^{EE,11}\psi + C_{t-1}^{EE,21}(\beta(1+\theta_x) + \psi) - \bar{\pi}(\theta_x(\beta - 1) + \psi(1 - \theta_\pi) + \beta - 1) \end{pmatrix} \tag{37}$$

and

$$\begin{aligned} \mathbb{Q}_t^{EE} &= \mathbf{B}\mathbf{Q}_{t-1}^{EE}\rho + \mathbf{D} & (38)\\ &= m \begin{pmatrix} 1 - \omega + \rho_a(Q_{t-1}^{EE,11} + Q_{t-1}^{EE,21}(1 - \theta_\pi\beta) + \omega - 1) & \theta_\pi + (Q_{t-1}^{EE,12} + Q_{t-1}^{EE,22} - \theta_\pi Q_{t-1}^{EE,22}\beta)\rho_e \\ Q_{t-1}^{EE,21}\rho_a(\beta(1+\theta_x) + \psi) + \psi(1 - \omega + \rho_a(Q_{t-1}^{EE,11} + \omega - 1)) & \theta_x(Q_{t-1}^{EE,22}\beta\rho_e - 1) + Q_{t-1}^{EE,12}\rho_e\psi + Q_{t-1}^{EE,22}\rho_e(\beta + \psi) - 1, \end{pmatrix} \end{aligned}$$

where $C_{t-1}^{EE,ij}$ and $Q_{t-1}^{EE,ij}$ represent the i^{th} row and j^{th} column elements of $\mathbf{C}_{t-1}^{EE}$ and $\mathbf{Q}_{t-1}^{EE}$, respectively.

Comparing $\mathbb{C}^{RE}$ to $\mathbb{C}_t^{EE}$, the restriction placed on the reduced form under RE are loosened under EE-CGL in two ways. First, the reduced forms intercepts are no longer

[11] As is common in the literature, we assume agents know $\mathbf{R}$.

explicitly tied to the steady states of the model. They depend on other structural parameters and beliefs, which frees them to vary over time. Second, comparing $\mathbb{Q}^{RE}$ to $\mathbb{Q}_t^{EE}$, the degree of nonlinearity of the mapping from structural parameters to the reduced form has decreased. The reduction in nonlinearity allows for changes in structural parameters to have smaller effects on the other parameters, which allows these parameters to take on new values without having as significant an impact elsewhere in the model.

For example, consider the limit of the first element of $\mathbb{Q}_t^{EE}$ compared $\mathbb{Q}^{RE}$ as $\rho_a \to 1$. For the RE case, the coefficient goes to zero. Therefore, as the preference shock, a_t, becomes more persistent, its affect on the output gap approaches zero. This of course offsets the persistence by effectively removing the shock from the model. In the EE case, however, no such restriction is imposed. Here as ρ_a goes to one, $\mathbb{Q}_t^{EE,11}$ goes to $m(\mathbf{Q}_{t-1}^{EE,11} + \mathbf{Q}_{t-1}^{EE,21}(1-\theta_\pi\beta))$, which allows a more persistent preference shock to continue to affect the output gap so long as $\mathbf{Q}_{t-1}^{EE,11}$ and $\mathbf{Q}_{t-1}^{EE,21}$ are nonzero.

The EE-FB case implies the same reduced form relationships. The only difference between this case and EE-CGL is that time variation of $\mathbb{C}$ and $\mathbb{Q}$ is ruled out by assumption.

3.1.4 Infinite-horizon learning fixed beliefs

The infinite-horizon learning solution uses the full forward-looking decision rules given by Equation (23) and (24). This requires agents to forecast x_t and π_t as well as i_t for $T = t+1, ..., \infty$. In contrast, there is no need to forecast interest rates under EE-CGL. Therefore, to preserve our nested structure, we assume that agents know the coefficients of the Taylor rule, which allows them to construct their expectation of i_t using their forecast x_t and π_t from the same PLM assumed for the EE case (Equation 33). With this assumption, agents' expectations are computed as

$$\mathbb{E}_t^{IH} \sum_{T=t}^{\infty} \beta^{T-t} \mathbf{y}_{T+1} = \mathbf{C}_{t-1}^{IH}(1-\beta)^{-1} + \mathbf{Q}_{t-1}^{IH}(\mathbf{I} - \beta\rho)^{-t}\rho\mathbf{v}_t \tag{39}$$

and

$$\mathbb{E}_t^{IH} \sum_{T=t}^{\infty} (\lambda_1\beta)^{T-t} \mathbf{y}_{T+1} = \mathbf{C}_{t-1}^{IH}(1-\lambda_1\beta)^{-1} + \mathbf{Q}_{t-1}^{IH}(\mathbf{I} - \lambda_1\beta\rho)^{-t}\rho\mathbf{v}_t, \tag{40}$$

where $\mathbf{C}_{t-1}^{IH}$ and $\mathbf{Q}_{t-1}^{IH}$ are the coefficients of Equation (33) estimated using the constant gain recursive least squares algorithm discussed previously. Substituting these expectations into Equation (23) and (24), the reduced form matrices take the following form $\mathbb{C}_t^{IH} = \mathbf{\Gamma}_{IH} + \mathbf{B}_{IH}\mathbf{C}_{t-1}^{IH}$, where

$$\mathbf{\Gamma}_{IH} = m\begin{pmatrix} \frac{\bar{\pi}(\theta_\pi-1)}{(1-\beta)} - \frac{(1-\beta)\bar{\pi}\theta_\pi}{(1-\beta\lambda)} \\ \frac{(1-\beta)\bar{\pi}(\theta_x+1)}{(1-\beta\lambda)} + \frac{\bar{\pi}(\theta_\pi-1)\psi}{(1-\beta)} \end{pmatrix},$$

$$\mathbf{B_{IH}} = m\begin{pmatrix} \frac{1-\beta(\theta_x+1)}{(1-\beta)} - \frac{\beta\theta_\pi\lambda\psi}{(1-\beta\lambda)} & \frac{1-\beta\theta_\pi}{(1-\beta)} - \frac{\beta\theta_\pi(1-\lambda)}{(1-\beta\lambda)} \\ \frac{\beta\lambda\psi(\theta_x+1)}{(1-\beta\lambda)} + \frac{\psi(1-\beta(\theta_x+1))}{(1-\beta)} & \frac{\beta(1-\lambda)(\theta_x+1)}{(1-\beta\lambda)} + \frac{\psi(1-\beta\theta_\pi)}{(1-\beta)} \end{pmatrix},$$

and $\mathbb{Q}_t^{IH} = \mathbf{B}'_{1,IH}\mathbf{Q}_{t-1}^{IH}\left(\mathbf{I}-\beta\rho\right)^{-1}\rho + \mathbf{B}'_{2,IH}\mathbf{Q}_{t-1}^{IH}\left(\mathbf{I}-\beta\lambda_1\rho\right)^{-1}\rho + \mathbf{D}$, where

$$\mathbf{B}'_{1,IH} = m\begin{pmatrix} 1-\beta(1+\theta_x) & 1-\beta\theta_\pi \\ \psi\left(1-\beta\left(\theta_x+1\right)\right) & \psi\left(1-\beta\theta_\pi\right) \end{pmatrix}, \text{ and}$$

$$\mathbf{B}'_{2,IH} = m\begin{pmatrix} -\beta\theta_\pi\lambda\psi & -\beta\theta_\pi(1-\lambda) \\ \beta\lambda\psi\left(\theta_x+1\right) & \beta(1-\lambda_1)\left(\theta_x+1\right) \end{pmatrix}.$$

The IH-FB case is obtained by again setting $\gamma = 0$.

The infinite-horizon specifications case dramatically alters the mapping from structural parameters to the reduced form, while still nesting the RE solutions. The implied restrictions turn out to allow a wider range of possible reduced form parameterizations than is feasible under either RE or the EE specifications. We illustrate this numerically in the next section.

3.2 Understanding how RE restrictions affect fit

We use a numerical exercise to illustrate the practical implications of the nonlinear relationship between the structural and reduced form parameters. The idea of this exercise is to explore the range of the reduced form parameter, $\mathbb{Q}$, implied for a range of key structural parameters under different expectation assumptions.

We use the EE-FB and IH-FB assumptions to calculate the implied value of $\mathbb{Q}$ for the following ranges of the structural parameters: $1.4 < \theta_\pi < 1.6$, $0.1 < \theta_x < 0.3$, $.75 < \rho_a < .95$, and $.65 < \rho_a < .85$. We calibrate the remaining parameters to $\beta = 0.995$, $\psi = 0.1$, $\lambda = 0.93$, and $\omega = 0.06$ and set the fixed beliefs to the RE values implied by the midpoint of the respective ranges. We then ask what values of θ_π, θ_x, ρ_a, and ρ_e are necessary to generate the same $\mathbb{Q}$ under RE. To be precise, we let $\mathbf{\Xi} = (\theta_\pi, \theta_x, \rho_a, \rho_e)'$. and calculate $G(\mathbf{\Xi}, \bar{\mathbf{\Theta}})$ such that $\bar{\mathbf{\Theta}} = (1.5, 0.2, 0.85, 0.75)'$ for the aforementioned ranges of the elements of $\mathbf{\Xi}$. We then calculate $\mathbf{\Theta}^{RE} = F^{-1}(G(\mathbf{\Xi}, \bar{\mathbf{\Theta}}))$ and compare $\mathbf{\Theta}^{RE}$ to $\mathbf{\Xi}$.

Figure 3 shows the comparison. Panels A and B show the reduced form values of

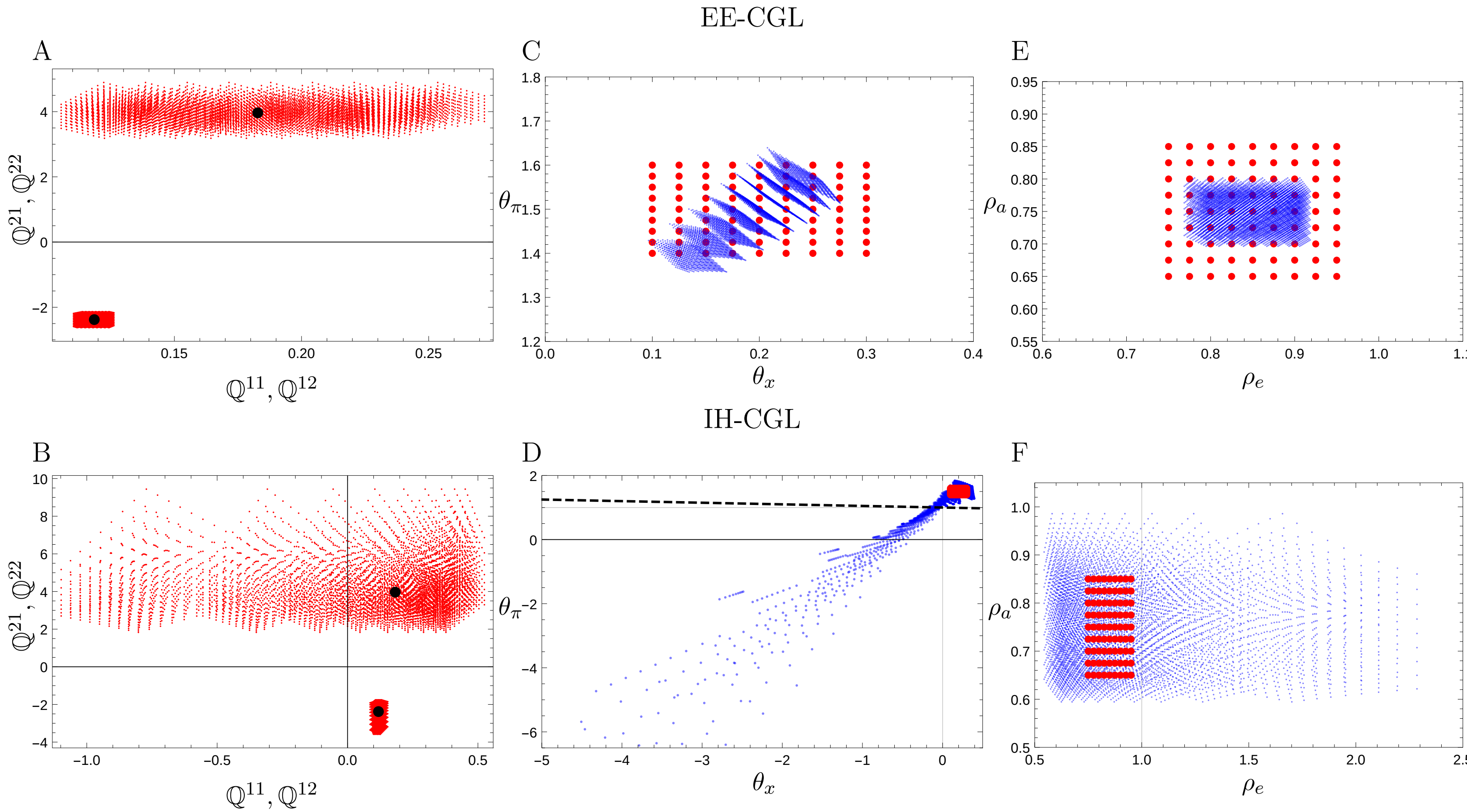
$\mathbb{Q}^{EE}$and $\mathbb{Q}^{IH}$ for the chosen grid, where the large black dots represents the values of $\mathbb{Q}^{RE}$ implied by $\bar{\boldsymbol{\Theta}}$ and $\mathbb{Q}^{ij}$ denote the ij^{th} element of the $\mathbb{Q}$ matrix. Panels C, D, E, and F show the grid of points for $\boldsymbol{\Xi}$ in red that is used to calculate $\mathbb{Q}^{EE}$ and $\mathbb{Q}^{IH}$, respectively, and the implied RE values of $\boldsymbol{\Theta}^{RE}$ that give rise to the same reduced form values in blue.

The EE and IH cases reveal two different ways in which the RE restrictions can be relaxed. In the EE case, the reduction in nonlinearity means that same reduced form parameter values can be explained by a wider range of Taylor rule parameters and AR coefficients. This means that there is a less of a trade-off when fitting different components of monetary policy and the shocks simultaneously. The persistence parameters can move over a much larger range without significantly affecting the value of the Taylor rule parameters. The RE case more tightly links these quantities together. Therefore, small adjustments to one parameter to fit some aspect of the data has more significant spillovers onto the other parameters in the model.

The IH case, however, is different. It explains a significantly larger reduced form space than is possible under RE for the same parameter values. The space is so large that the RE solution is incapable of covering the same area with parameter values that satisfy the Taylor principle or stationarity. Of course, it is possible that these ranges of the parameters space are not empirically relevant. However, as we will see in the next section, that is not what we find.

The key takeaway from this exercise is that it is not necessary for the structural parameters to change in a substantial way to fundamentally alter the way the model fits the data. The nonlinearity of the mappings makes it so that small changes may imply a fundamentally different fit.

Figure 3: The mapping from structural parameters to the reduced form



Notes: The $\mathbb{Q}^{ij}$'s are the reduced form values of the matrix that multiplies the exogenous shocks in the model. The range of points shown in panels A and B are all the possible values that elements of $\mathbb{Q}$ may take for the range of structural parameters shown in red in the remaining panels. The blue points in the remaining panels show the value that structural parameters would need to take under RE to replicate the same range of $\mathbb{Q}^{ij}$'s shown in panels A and B, respectively. In other words, the blue and red dots in panels C and E produce the same range of reduced form values for $\mathbb{Q}$ in panel A and likewise for panels D, F, and B. The black dots in panels A and B show the RE reduced form value implied by $\bar{\boldsymbol{\Theta}} = (1.5, 0.2, 0.85, 0.75)'$, the midpoint of the chosen ranges. The black dashed line in panel D denotes the determinacy condition for the model. Points below the line correspond to indeterminate solutions under RE.

4 Taking the model to the data

In this section, we estimate the five versions of the model using maximum likelihood and compare the in-sample fit, the out-of-sample fit, the predicted impulse responses for the structural shocks, and the implied moments. For estimation, we use US data from 1984q1 through 2008q3. As observables, we use the CBO measure of the output gap, the GDP deflator measure of inflation, the three-month Treasury bill rate, and growth rate of real GDP, which are each expressed in quarterly rates.

To avoid known issues with weak identification, we calibrate some of the structural parameters. We set $\beta = 0.995$ and $\bar{\pi} = 0.005$, which implies a steady state nominal interest rate of 4% in annualized terms.[12] We set the slope of the Phillips curve, ψ, to 0.1 following Ireland (2004), which in the Calvo pricing framework would correspond to the average firm adjusting its price roughly once a year. Finally, we calibrate the mean growth rate of output, $\bar{g}$, to the average growth rate observed over the estimation period.[13]

The remaining parameters $\{\theta_\pi, \theta_x, \omega, \rho_a, \rho_e, \sigma_a, \sigma_e, \sigma_i, \sigma_g\}$ are estimated.[14] In the CGL models, we also estimate the gain parameters. We allow there to be separate gains for output gap and inflation (γ_x, γ_π) to permit differing amounts of learning to contribute to each variable. The learning beliefs are initialized at the RE estimates from the full-sample. The FB beliefs are set to those same values. Therefore, all five models perfectly nest the RE results. If the RE model is the true data generating process, then all five models will provide the same inference.

[12]This puts the model slightly at odds with the data over our sample period, which has a mean inflation rate of around 2.5% and a mean interest rate of around 4.8%. However, the mean output gap in our sample is almost -0.75%, which makes it unclear whether the mean values of inflation and the interest rates actually reflect steady-state values. If $\bar{\pi}$ is freely estimated we find values ranging between 1% and 2%.

[13]We find that our results with respect to model fit are not sensitive to these calibrations choices or to freely estimating all of the parameters. However, freely estimating all parameters does result in significant difference in parameters estimates across the different model specifications, some of which reflects weak identification.

[14]To construct confidence intervals for the structural parameter estimates, we use a method similar to the one proposed by Stock and Watson (1998) for time-varying parameter models. We employ this method because the constrained optimization routine we use to maximize the likelihood functions provides unreliable numerical estimates of the Hessian matrix. The confidence intervals are constructed for each parameter by selecting a grid surrounding the ML estimate of interest. We then re-maximize the log-likelihood function by searching over all other parameters, while holding the parameter of interest fixed at one point on the grid. The maximized log-likelihood value obtained at that point is then compared to the original maximum log-likelihood value using a likelihood ratio test. The set of points on the grid that yield maximized log-likelihood values that fail to reject the null hypothesis that the true parameter vector lies in the restricted parameter space at the 5% level constitutes the 95% confidence interval.

Table 2: ML estimates

	RE	EE-FB	EE-CGL	IH-FB	IH-CGL
θ_π	1.577	1.123	1.114	1.578	1.591
	[1.31 , 1.97]	[0.99 , 1.61]	[0.99 , 1.59]	[1.32 , 1.94]	[1.45 , 1.73]
θ_x	0.166	0.208	0.211	0.160	0.140
	[0.08 , 0.29]	[0.11 , 0.36]	[0.11 , 0.37]	[0.08 , 0.26]	[0.07 , 0.23]
ω	0.000	0.000	0.000	0.066	0.000
	[0.00 , 0.02]	[0.00 , 0.01]	[0.00* , 0.03]	[0.00 , 0.48]	[0.00 , 0.00]
ρ_a	0.933	0.926	0.912	0.973	0.995
	[0.87 , 1.00]	[0.86 , 0.99]	[0.85 , 0.98]	[0.95 , 0.99]	[0.99 , 0.99]
ρ_e	0.967	0.917	0.927	0.684	0.543
	[0.87 , 1.00]	[0.80 , 1.00]	[0.81 , 0.99]	[0.50 , 0.86]	[0.32 , 0.78]
$\sigma_a \times 100$	1.671	1.633	1.560	0.555	0.110
	[0.06 , 0.68]	[1.36 , 2.04]	[1.23 , 1.99]	[0.19 , 1.38]	[0.01 , 1.46]
$\sigma_e \times 100$	0.071	0.072	0.073	0.166	0.229
	[0.06 , 0.10]	[0.06 , 0.09]	[0.07 , 0.08]	[0.11 , 0.22]	[0.18 , 0.27]
$\sigma_i \times 100$	0.558	0.558	0.561	0.560	0.557
	[0.46 , 0.68]	[0.46 , 0.73]	[0.46 , 0.72]	[0.47 , 0.70]	[0.52 , 0.61]
$\sigma_g \times 100$	0.179	0.180	0.180	0.175	0.179
	[0.16 , 0.21]	[0.16 , 0.22]	[0.16 , 0.21]	[0.15 , 0.21]	[0.16 , 0.19]
γ_x	-	-	0.011	-	0.0008
			[0.00 , 0.06]		[0.00 , 0.011]
γ_π	-	-	0.000	-	0.0017
			[0.00 , 0.02]		[0.00 , 0.0023]
Log Likelihood	1974.49	1982.60	1983.14	1986.84	1988.17
AIC	-3930.98	-3947.20	-3944.28	-3955.68	-3954.34
LR Statistic (rel. RE)		16.22	17.30	24.70	27.36
LR Statistic (rel. FB)			1.08		2.66

Notes: ML estimates for the Ireland model under the five different assumptions for expectations. 95% confidence intervals for the estimates are shown in brackets below the point estimates. The forecasting function in the FB cases and the initial beliefs in the CGL cases are set to the RE solution implied by the estimates in the first column. LR refers to the likelihood ratio.

4.1 In-sample fit comparison

Table 2 shows the estimation results for the five different model specifications. The parameter estimates mostly reflect the three stylized facts. The four bounded rationality strategies each improve the in-sample fit with respect to RE, the parameter estimates are fairly similar across the different specifications with a few exceptions, and the gains are estimated to be small. There is more movement in parameter estimates here than is observed in the examples discussed in Section 1, but the confidence intervals for θ_x, θ_π, ω, ρ_a, and ρ_e all nearly overlap and of course these estimates do not rely on priors.[15]

Within the EE specifications, we observe no significant difference between the FB and CGL cases. Both result in nearly identical parameter estimates and fit the data more or

[15] The likelihood surface is very flat for θ_π in the EE specification moving towards the boundary of the Taylor principle constraint. In monte carlo simulations, there is a pile-up problem in both the EE and IH specifications, where a small percentage of estimates end up on this boundary despite the true value being well away. We have not investigated whether this is a feature of other New Keynesian models estimated under learning.

less equally well. A similar result is obtained for the IH specifications. Although, the improvement in fit between FB and CGL is more than twice as large as that observed for the EE case. In addition, the overall fit of the IH version of the model is significantly better than both EE specifications.

The relatively small and insignificant improvements in-sample fit between the FB and CGL cases demonstrate that the introduction of time-varying parameters does not account for the majority of improvement in the in-sample fit of the model. Most of the increase in fit in both cases is obtained when the estimation of the structural parameters are separated from beliefs (the FB cases), which allows the model to more flexibly fit the data.

Table 3 quantifies the increased flexibility of the model under FB and CGL by reporting the elements of $\mathbb{Q}$ for each case (for CGL we show the values implied at the initial belief) at their estimated values reported in Table 2 and the values of θ_π, θ_x, ρ_a, and ρ_e that would imply the same reduced form under RE.[16] Consistent with the numerical exploration in Section 3.2, the EE-FB and EE-CGL cases generate a modest loosening of the cross-equation restrictions. The reduced form $\mathbb{Q}$ is modestly different from the values estimated under RE. While the implied RE parameters that reproduce the same $\mathbb{Q}$ remain within the feasible parameter space and even lie within the confidence intervals of the RE estimates. The IH-FB and IH-CGL cases, on the other hand, imply drastically different reduced forms from RE and the EE specifications. To reproduce the same $\mathbb{Q}$ under RE requires parameter values that are infeasible and which would imply explosive dynamics.

To further the point, we can quantify the severity of these restrictions by directly estimating $\mathbb{Q}$ along with the full complement of the other parameters.[17] This represents the fully unrestricted case. We obtain a log likelihood value of 1,991 and

$$\hat{\mathbb{Q}} = \begin{pmatrix} -0.910 & 2.131 \\ -0.008 & -1.441 \end{pmatrix}. \tag{41}$$

Therefore, even the IH case remains somewhat restricted.[18] However, it is the only case

[16]Because we calibrate $\bar{\pi}$, the implied values of $\mathbb{C}$ are roughly equivalent across the five cases and not shown here.

[17]For this exercise, we do not impose any structure on $\mathbb{Q}$ and allow its four elements to be freely estimated along with the other parameters. $\mathbb{C}$ retains the same restrictions as those imposed under RE. The remaining parameter estimates are almost identical to those obtained under IH-FB in this case with the exception of θ_π. Its value falls to $\theta_\pi = 1.01$.

[18]Although, IH dominates the unrestricted case in terms of AIC because of there are four extra parameters in this case.

Table 3: Implied reduced form estimates of $\mathbb{Q}$

	$\mathbb{Q}^{11}$	$\mathbb{Q}^{21}$	$\mathbb{Q}^{12}$	$\mathbb{Q}^{22}$	$\mathbf{\Theta}^{RE} = (\theta_x, \theta_\pi, \rho_a, \rho_e)'$
RE	0.060	0.083	8.899	-2.901	[1.577, 0.166, 0.933, 0.967]
EE-FB	0.091	0.085	7.272	-2.921	[1.735, 0.242, 0.897, 0.911]
EE-CGL	0.091	0.085	7.126	-2.881	[1.736, 0.241, 0.897, 0.905]
IH-FB	-0.858	0.023	2.367	-1.355	[15.625, 8.132, 4.754, 0.439]
IH-CGL	-3.251	0.081	1.840	-0.937	[12.578, 5.469, 5.039, 0.130]

Notes: Reduced form values implied by the estimated parameters given in Table 2. The final column replicates the exercise conducted in Section 3.2 and shows the value of the RE structural parameters, which would be consistent with the same reduced form values.

that can come close to capturing the unconstrained values.

4.2 Out-of-sample fit

Next, we compare the different models' real-time out-of-sample forecast accuracy for the four observable variables to see whether improved in-sample fit translates into actual forecasting power. We conduct a recursive real-time forecasting exercise, where we use multiple vintages of data to simulate the information set that would have been available to a forecaster at each point in time. We use the real-time data set provided by the Philadelphia Federal Reserve for real-time data on GDP Deflator, GDP growth, and the three-month Treasury bill. For a real-time measure of the output gap, we use the Fed's Green Book nowcast, which we assume that agents observe with a one-quarter lag.

Our full sample period runs 1984q1 - 2010q1. We construct forecast recursively starting in 1991q1 and ending 2008q3 at four different horizons: the nowcast, one quarter ahead, four quarters ahead, and six quarters ahead. The nowcast is included because in real-time, GDP growth and the GDP Deflator measure of inflation are observed with a one-quarter lag.[19] We also extract from the models a real-time estimates of inflation expectations.

To evaluate forecast accuracy, we compare the inflation and GDP forecasts to the second release values available in the real-time data set. For the output gap, we use the most recent vintage of the CBO measure of the variable. And for the inflation expectations, we use the SPF mean nowcast of GDP Deflator inflation for one-step-ahead expectations and the two-year ahead forecast of the same variable for long run inflation expectations. Inference on improvements in forecast accuracy are obtained using the Diebold and Mariano (1995) test statistic (DM) with the Harvey et al. (1997) small

[19] Since the interest rate is observed contemporaneously, we do not report a nowcast for this variable.

sample and forecast horizon correction.[20]

Table 4 shows the out-of-sample results for the five cases plus results from a random walk forecast for each variable. The random walk forecasts are included to show how well the models do in absolute terms. The top row for each variable gives the root mean squared forecast error (RMSFE) of the RE forecasts at the four different horizons. The remaining rows give the relative RMSFE of the forecasts compared to RE with the DM test statistic in parentheses below. Values below one represent an improvement in forecast accuracy relative to the RE forecast.

The EE-FB and EE-CGL forecasts show significantly greater accuracy with forecasting interest rates at all horizons relative to RE, marginal improvements for forecasting inflation, and little to no improvements for forecasting real variables. We find that there is also almost no difference in forecast accuracy between the EE-FB and EE-CGL cases. Therefore, the forecasting results mirror the in-sample results. Moving from RE to an EE specification generates some improvements in fit. However, the majority of the improvements are generated by the EE-FB case. The addition of learning adds little in the way of forecasting power.

The IH-FB and IH-CGL forecasts show significant increases in forecast accuracy relative to RE for nominal variables and no improvement in real variables. The IH-FB and IH-CGL cases also perform roughly equally well across all variables and horizons. Learning does not appear to materially add to the forecasting power over and above what occurs in the FB case. The IH cases, however, do on average forecast qualitatively better than the EE specifications, which indicates that the observed improvement in the in-sample fit does somewhat translate into out-of-sample forecasting power. Comparing the model to the random walk forecast, all of the considered specifications do surprisingly well. The four boundedly rational cases perform on average no worse than the random walk in most cases.

Figure 4 shows the real-time inflation expectation estimates versus the SPF, while Table 5 reports the RMSFE comparison. The steady state of inflation is calibrated so that the long run inflation expectation under RE and FB is by construction the same. Any difference in the short run expectations between RE and FB is driven by different estimates of the unobserved shocks. Keeping with theory, we find that the EE specifications produce short run expectations that more closely approximate the SPF, while the IH specification produces better long run expectations. Although, neither model produces compelling estimates that should be taken too seriously.

[20] This test statistic is found to work well on tests of real-time forecasts by Clark and McCracken (2009) and Clark and McCracken (2011).

Table 4: Real-time forecast results

Inflation — Annualized RMSFE

	t (Nowcast)	t+1	t+4	t+6
RE	1.09	1.01	1.08	0.95
RMSFE Relative to RE				
RW	1.04	0.84**	0.91	1.05
	(3.29)	(-1.99)	(-1.07)	(0.62)
EE-FB	0.98**	0.97	0.98*	0.97
	(-1.90)	(-1.15)	(-1.32)	(-1.05)
IH-FB	0.92***	0.85***	0.88***	0.84***
	(-3.20)	(-2.98)	(-2.44)	(-2.70)
EE-CGL	1.00	0.98	1.03	1.04
	(-0.26)	(-0.75)	(1.36)	(1.17)
IH-CGL	0.93***	0.88***	0.92**	0.89**
	(-2.58)	(-2.88)	(-1.82)	(-1.94)

GDP — Annualized RMSFE

	t (Nowcast)	t+1	t+4	t+6
RE	2.33	2.18	2.07	1.69
RMSFE Relative to RE				
RW	0.98	1.00	1.16	1.30
	(-0.20)	(-0.01)	(1.72)	(2.87)
EE-FB	1.13	1.01	0.99	0.99
	(0.96)	(0.16)	(-0.31)	(-0.54)
IH-FB	1.07	0.99	1.02	1.03
	(0.75)	(-0.27)	(0.87)	(1.29)
EE-CGL	1.03	0.98	1.00	1.00
	(0.36)	(-0.48)	(0.00)	(0.22)
IH-CGL	1.09	0.99	1.01	1.02
	(0.83)	(-0.20)	(0.49)	(0.76)

Interest Rates — Annualized RMSFE

	t (Nowcast)	t+1	t+4	t+6
RE	-	1.64	1.88	2.09
RMSFE Relative to RE				
RW	-	0.28***	0.81**	0.97
		(-4.75)	(-1.76)	(-0.30)
EE-FB	-	0.81***	0.89***	0.93**
		(-3.30)	(-2.18)	(-1.73)
IH-FB	-	0.70***	0.83***	0.91*
		(-4.11)	(-2.42)	(-1.37)
EE-CGL	-	0.83***	0.89***	0.94*
		(-3.01)	(-2.21)	(-1.52)
IH-CGL	-	0.73***	0.83***	0.90**
		(-3.96)	(-2.47)	(-1.70)

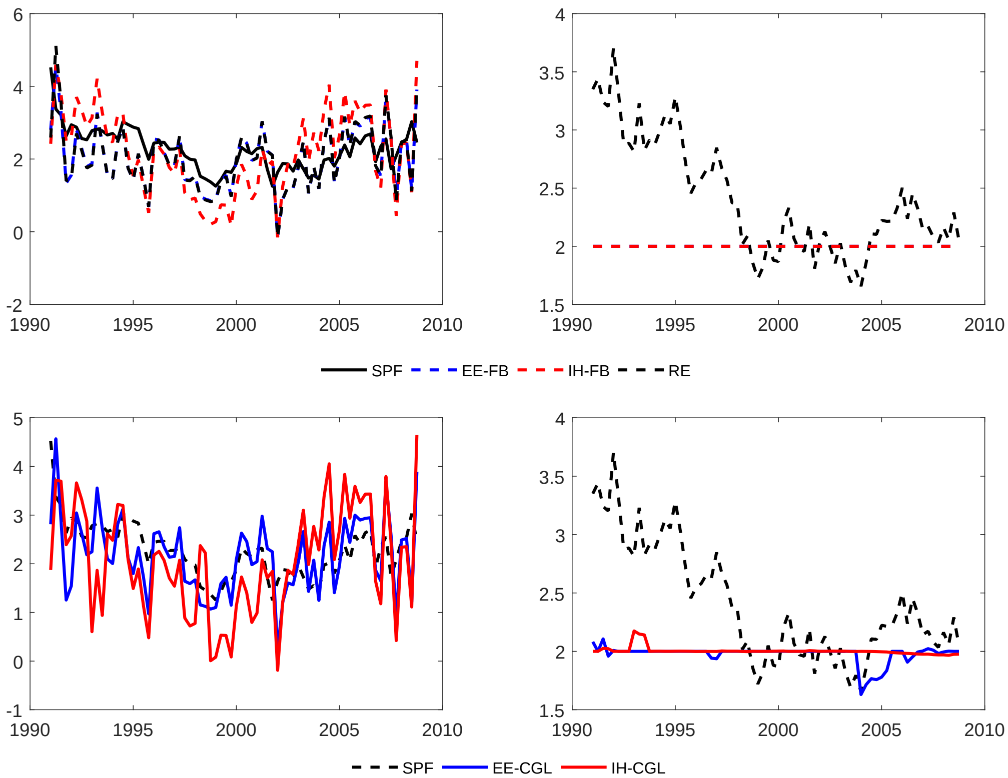
Output Gap — Annualized RMSFE

	t (Nowcast)	t+1	t+4	t+6
RE	1.42	1.56	2.20	2.48
RMSFE Relative to RE				
RW	1.03	1.05	1.06	1.08
	(4.16)	(3.51)	(4.66)	(4.31)
EE-FB	1.09	1.05	1.00	0.99
	(2.63)	(1.61)	(-0.03)	(-0.70)
IH-FB	1.08	1.05	1.02	1.02
	(2.48)	(1.64)	(1.02)	(1.38)
EE-CGL	1.03	0.97	0.95***	0.95***
	(1.08)	(-1.25)	(-2.56)	(-3.33)
IH-CGL	1.08	1.05	1.01	1.01
	(2.27)	(1.36)	(0.51)	(0.32)

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

Notes: Diebold and Mariano test statistics are reported in parenthesis. We only place asterisks on cases where a significant improvement is obtained.

Figure 4: Inflation Expectations vs SPF



Notes: Comparison of the real-time model implied inflation expectations at a short and long horizon compared to the SPF.

The EE-CGL model produces significantly better inflation expectations estimates than EE-FB, however, this is completely driven by a persistent level shift in the EE-CGL estimates. The EE-FB and EE-CGL inflation expectations actually have a correlation of 0.97, which suggests that it is a low-frequency drift that is driving fit. The correlation between IH-FB and IH-CGL is 0.89 and the correlation between EE-CGL and IH-CGL is 0.71. Overall, none of the short run inflation expectations estimates have a correlation exceeding 0.55 with the mean SPF forecast.

The caveat, of course, is that by design we have fixed initial beliefs to their REE values. If we did not impose this restriction, both EE and IH specifications ability to match inflation expectations greatly improves as documented by Slobodyan and Wouters (2012a) and Cole and Milani (2019). The improvement though comes from matching the persistent differences between survey expectations and the actual data, which again highlight CGL's ability to capture long run dynamics as opposed to short-run business cycle dynamics.

Table 5: Inflation Expectations vs SPF

	RE	Relative to RE			
	RMSFE	EE-FB	IH-FB	EE-CGL	IH-CGL
SR	0.802	0.960***	1.231	0.834***	1.327
		(-1.849)	(2.893)	(-4.316)	(3.364)
LR	0.638	1.000	1.000	1.004	0.986**
		-	-	(0.497)	(-1.680)

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

Notes: Diebold and Mariano test statistics are reported in parenthesis. We only place asterisks on the cases were significant improvements are obtained.

4.3 Impulse responses

We now turn to exploring the persistence implied by the five models directly. Figure 5, 6, and 7 show the estimated impulse responses for the monetary policy ($\epsilon_{i,t}$), the preference ($\epsilon_{a,t}$), and the cost push ($\epsilon_{e,t}$) shocks, respectively. We set the size of the shock to the RE estimated values for one standard deviation and we allow beliefs to update in order to propagate the shocks.

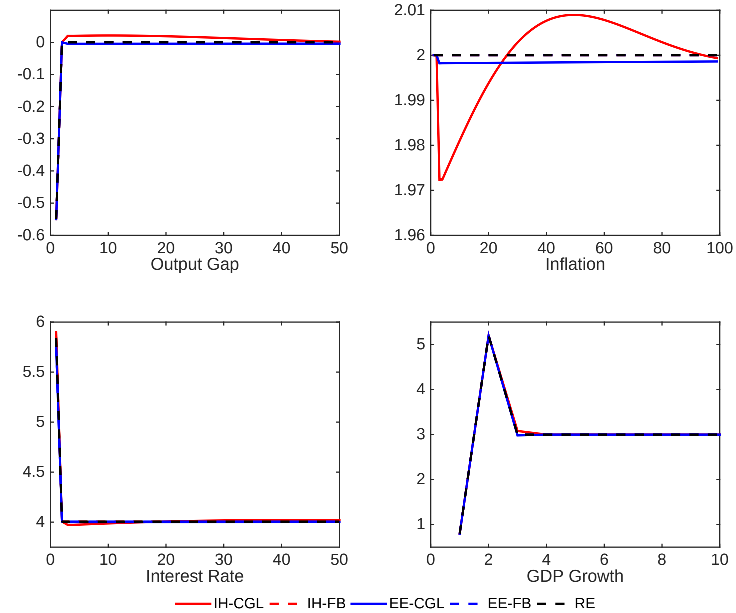
The monetary policy shock provides the clearest picture of the role that CGL can plays in adding persistence. There is no exogenous persistence for the monetary policy shocks in this model. Any persistence from a monetary policy shock must be generated through expectations. Therefore, the RE and FB cases, by construction, can only respond in the period the shock is realized. In Figure 5, we see that the shock is propagated to a degree by both EE-CGL and IH-CGL. The propagation is much greater in the IH case and even delivers a hump shaped response. The duration, however, goes far beyond typical business cycle frequencies with the shock generating effects that last roughly 25 years.

Figure 6 shows the impulse response for the preference shocks. The EE-FB and EE-CGL impulse responses are identical in this case. Learning does not appear to add any persistence. In contrast, the IH-FB and IH-CGL response are quite different. The IH-FB response is similar to EE and RE cases, while IH-CGL exhibits extreme persistence.

Figure 7 shows the cost-push shock. There is little difference between FB and CGL for either EE or IH specifications. Beliefs do not move much in response to cost-push shocks under either specification.

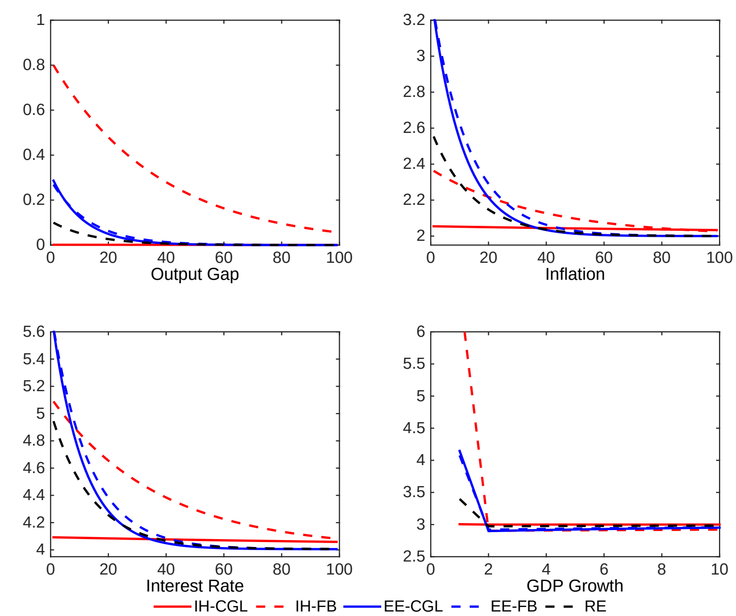
The net takeaway from the impulse responses is that CGL does not add much, if any, persistence in the EE specification. Only the IH specification implies any additional

Figure 5: Monetary policy shock



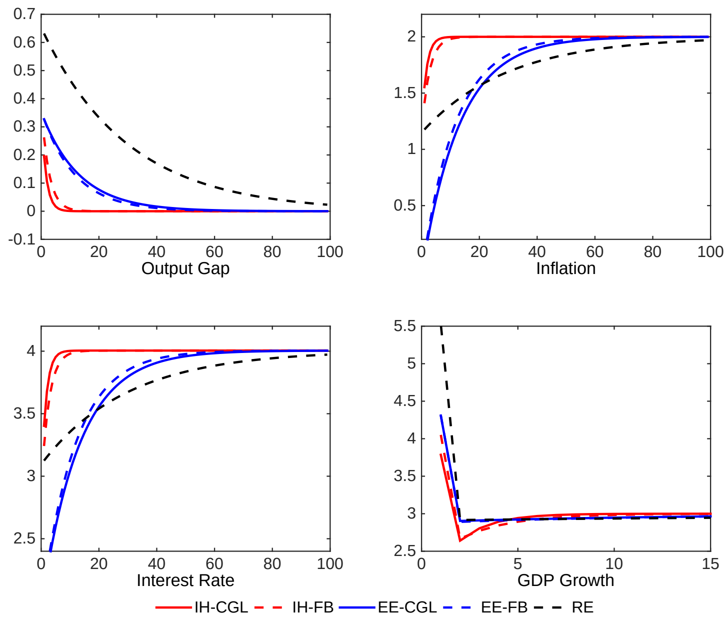
Notes: The shock is same for all models. The shock size is set to one standard deviation using the RE estimate for that value reported in Table 2.

Figure 6: Preference shock



Notes: The shock is same for all models. The shock size is set to one standard deviation using the RE estimate for that value reported in Table 2.

Figure 7: Cost-push shock



Notes: The shock is same for all models. The shock size is set to one standard deviation using the RE estimate for that value reported in Table 2.

persistence of shocks through expectations. However, the persistence is beyond what most modelers intend to capture with respect to the duration of shocks over the business cycle.

4.4 Model moments

Finally, we compare the estimated models' implied standard deviations and autocorrelation for the four observable endogenous variables to one another and to the actual data. Table 6 reports the actual and estimated model implied standard deviations for the output gap, inflation, the three-month treasury bill rate, and real GDP growth, which are expressed in annualized terms. The four bounded rationality strategies all imply variances that are qualitatively closer to the actual data than RE. The estimated RE model predicts too much volatility in each variable. The majority of the improvements here, though, occur in the FB cases and then carry over into the CGL cases. Therefore, once again, it appears that loosening the cross-equation restrictions brings the model closer to the data, while learning adds only marginal improvements.

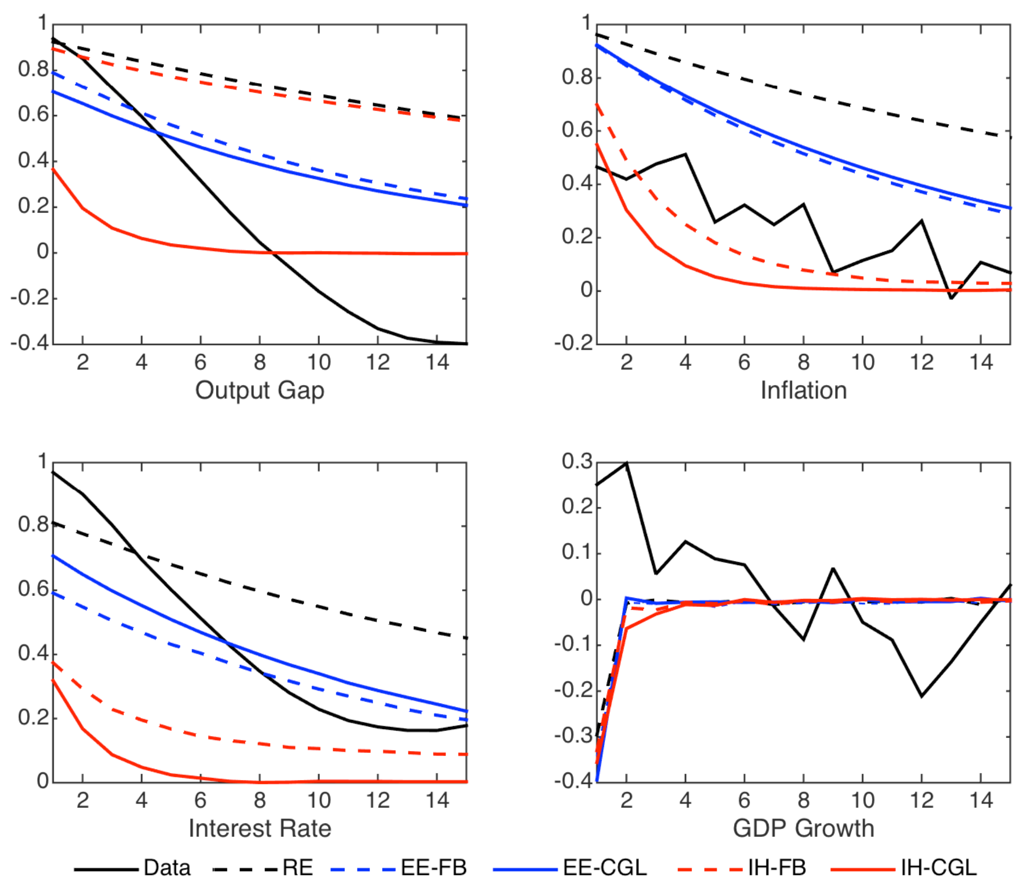
Figure 8 shows the autocorrelation functions implied by the estimated models compared to the actual data. Overall, the FB and CGL cases capture a wider range of correlations than the RE model is capable of, which is more in-line with the data. How-

Table 6: Actual and estimated model implied standard deviation of observable variables

Source	Output Gap	Inflation	Interest Rates	GDP Growth
Data	1.40	0.97	2.13	2.04
RE	2.63	3.68	4.79	4.14
EE-FB	1.48	2.58	2.93	3.92
EE-CGL	1.17	2.94	3.64	3.63
IH-FB	2.23	1.26	2.66	4.08
IH-CGL	0.93	1.66	2.94	4.26

Notes: Actual and estimated model implied standard deviations for the four observable data series used in estimation. The results reflect the estimated values provided in Table 2.

Figure 8: Autocorrelation functions



Notes: Actual and estimated model implied autocorrelation functions. The figures reflect the estimated values reported in Table 2.

ever, none of the models is able to fully approximate the autocorrelation present in the data.

As in previous comparisons along the other dimensions, the EE-FB and EE-CGL models predict nearly identical autocorrelations functions across the four variables. This indicates that learning does not add much above the loosening of restriction that is shared with the fixed belief case. But, there are significant differences between the IH-FB and the IH-CGL specifications. These cases predict fairly distinct autocorrelation functions despite having nearly identical parameterizations, which indicates that learning is playing a role in generating different dynamics in the simulated data.

4.5 Discussion

The common finding across the four considered dimensions is that the FB case moves the model significantly closer to capturing the data relative to RE. The additional assumption of CGL does not add much under the EE specification and makes only a modest contribution under the IH specification. This should be somewhat surprising given the fact that the FB specifications are conditioned on the RE estimates. It is arguably the smallest deviation from rationality that one can consider, yet it provides significant improvements in model fit across a range of dimensions.

We argue that an FB-type case is the appropriate benchmark to assess a bounded rationality expectation assumption. This case allows for the possibility that RE is the correct assumption while imposing different restrictions on the structural parameters. Not all expectations assumption will nest RE as in the cases considered here. But it should be possible to construct specifications that come close to nesting the RE predictions for most models, which would allow a researcher to distinguish which assumptions are supported by the data and which are not.

The fact that IH-CGL model delivers the best all-around performance of any of the specifications considered is a comforting finding for the DSGE research program. Although IH learning is squarely a bounded rationality strategy, it preserves the underlying microeconomic foundations of the model with respect to the agent's decision problem. Therefore, misspecification of how expectations are formed may be the key assumption putting the New Keynesian model at odds with the data, which makes bounded rationality strategies, such as infinite-horizon learning, a promising approach to reconcile these models with the data.[21]

[21] This conclusion is also supported by recent a DSGE-VAR study of bounded rationality models by Cole and Milani (2019).

5 Stylized facts revisited

In Section 2.2, we showed that any estimated first-order approximated DSGE model allows for similar pathologies as those highlighted in the previous section. However, the extent to which these issue matter will depend on the details of the specific model under consideration. The basic New Keynesian model we study here is at the center of almost all policy-relevant DSGE models. Therefore, it is likely that larger models will inherit these issues.

One way to quickly assess whether RE is a significant source of misspecification is to estimate a Fixed Belief case using the thought experiment we have explored throughout this paper. To illustrate, we explore an EE-FB case in the model of Smets and Wouters (2007). Recall that the Smets and Wouters model under adaptive learning, detailed in Slobodyan and Wouters (2012b), exhibited all three stylized facts. Therefore, we ask to what extent can an EE-FB case generate similar results as EE-CGL.

For this exercise, we use the replication files provided by the *American Economic Review* for Smets and Wouters (2007), which estimate the model using Bayesian techniques with the software package Dynare. We start by replicating the benchmark case found in Table 1 and Table 1B of Smets and Wouters paper.[22] For the FB case, we construct the fixed belief using the posterior mode estimates from the RE estimation. We then estimate structural parameters of the FB model using the same priors as in the RE case.

Table 7 shows the replication and EE-FB results in the rightmost panel. For ease of comparison to our previous results, we only report a subset of the parameters estimates. The full set of parameters estimates and priors are given in Table 8 in the appendix. The EE-FB case clearly exhibits the same patterns as noted in the parsimonious model studied in Section 4. The parameter estimates are nearly unchanged relative to the RE estimates, yet the marginal likelihood is significantly improved.

To compare EE-FB with Slobodyan and Wouters (2012b), we present their estimation results for two of their reported EE case and their RE results. The first case, MSV-CGL with initial beliefs set to RE, represents an interesting alternative to our EE-FB case to assess the role that learning plays in improving fit. Here, they set the initial beliefs to those implied by RE using the current estimated parameter values. Therefore, in the first period, the model is equivalent to RE solution, where beliefs and structural parameters satisfy all RE restrictions. After the first period, beliefs are allowed to drift. They find

[22] Our results differ slightly from the results reported in Smets and Wouters (2007) for the RE case. However, since we use their official replications files, we have no reason to doubt that the observed differences are anything more than numerical imprecision, which arises from running the estimation on different versions of Dynare and Matlab.

that letting these beliefs drift does not add to fit. It is only when initial beliefs are allowed to be materially different from what is implied by the current estimates of the structural parameters that fit is found to improve.

The second MSV-CGL case reported uses optimized initial beliefs, which is a joint estimation of initial beliefs and CGL parameter estimates without imposing that the initial beliefs conform to any RE solution of the model. The estimation is done iteratively with the estimation of structural parameters taking the initial beliefs as given, which like our EE-FB case allows the initial beliefs to differ from the structural parameters at all times. Our EE-FB case produces an improvement in fit relative to RE that explains about a third of the improvement that is obtained under the optimized initial belief case. Although, the optimized initial beliefs are jointly estimated, so the degrees of freedom, in this case, are substantially higher than in the EE-FB case. In addition, the estimated gain remain small. It is likely that an EE-FB case with optimized beliefs would perform similarly.[23] Therefore, a significant proportion of the improvement in fit in the Smets and Wouters model for MSV learning can also be explained by a relaxation in the cross-equation restrictions imposed by RE.

6 Conclusion

This paper demonstrates that improvements in the in-sample fit of New Keynesian DSGE models under adaptive learning may speak more to the misspecification of the model under RE than to the veracity of the learning assumption that is being considered. In particular, we have shown that both Euler-equation and infinite-horizon learning generate significant improvements in in-sample fit and modest improvements in real-time out-of-sample forecast accuracy compared to RE. However, the actual assumption of learning only appears to meaningfully add to the model's predictions in the infinite-horizon case. The improvements under Euler-equation learning are instead explained by the relaxation of the RE restrictions and do not rely on backward-looking behavior by the agents. We conclude that constant gain learning appears to best capture longer-run movements in data that go beyond typical business cycle frequencies.

Our findings suggest that empirical comparisons between bounded rationality models and RE should be done with care. Significant improvements in model fit relative to RE do not necessarily provide evidence in favor of the alternative strategy since even the most

[23] There is a discrepancy in the marginal likelihood values, however, since our results match Smets and Wouters (2007) we believe this is due to numerical issues relating to difference in software versions across both Dynare and Matlab.

Table 7: Stylized facts revisited

	Slobodyan and Wouters (2012b)			Replication	
	RE	MSV-CGL (Init. REE)	MSV-CGL (Opt. Init.)	RE Replication	FB Replication (Init. REE)
Monetary policy and habits					
MP inflation	2.04	2.02	1.91	2.05	1.74
	[1.75, 2.33]	[1.74, 2.32]	[1.58, 2.22	[1.77, 2.35]	[1.41, 2.07]
MP output	0.09	0.08	0.13	0.09	0.11
	[0.05, 0.13]	[0.05, 0.12]	[0.07, 0.18]	[0.05, 0.13]	[0.05, 0.17]
MP output growth	0.22	0.22	0.19	0.22	0.19
	[0.18, 0.27]	[0.017, 0.26]	[0.15, 0.24]	[0.17, 0.26]	[0.15, 0.23]
MP smoothing	0.81	0.81	0.84	0.82	0.88
	[0.77, 0.85]	[0.77, 0.85]	[0.80, 0.88]	[0.78, 0.86]	[0.84, 0.92]
Habits	0.71	0.72	0.80	0.71	0.70
	[0.64, 0.78]	[0.65, 0.79]	[0.75, 0.84]	[0.64, 0.79]	[0.58, 0.83]
AR parameters					
Productivity	0.96	0.96	0.96	0.96	0.96
	[0.94, 0.98]	[0.94, 0.98]	[0.94, 0.99]	[0.94, 0.97]	[0.93, 0.99]
Risk premium	0.22	0.23	0.23	0.20	0.21
	[0.08, 0.36]	[0.07, 0.36]	[0.13, 0.32]	[0.07, 0.34]	[0.04, 0.36]
Gov. spending	0.98	0.97	0.96	0.97	0.97
	[0.96, 0.99]	[0.96, 0.99]	[0.96, 0.99]	[0.95, 0.98]	[0.95, 0.99]
Investment	0.71	0.72	0.45	0.74	0.71
	[0.62, 0.81]	[0.62, 0.82]	[0.33, 0.56]	[0.64, 0.84]	[0.57, 0.87]
MP shock	0.15	0.16	0.15	0.15	0.14
	[0.04, 0.24]	[0.05, 0.26]	[0.05, 0.26]	[0.04, 0.25]	[0.04, 0.23]
Price mark-up	0.89	0.89	0.93	0.90	0.90
	[0.81. 0.97]	[0.81, 0.97]	[0.88, 0.97]	[0.82, 0.98]	[0.83, 0.97]
Wage mark-up	0.97	0.97	0.97	0.97	0.95
	[0.95, 0.99]	[0.95, 0.99]	[0.95, 0.99]	[0.95, 0.99]	[0.92, 0.99]
St. Dev. Shocks					
Productivity	0.46	0.45	0.47	0.43	0.44
	[0.41, 0.51]	[0.41, 0.50]	[0.42, 0.52]	[0.39, 0.47]	[0.39, 0.49]
Risk premium	0.24	0.24	0.25	0.24	0.26
	[0.20, 0.28]	[0.20, 0.28]	[0.22, 0.28]	[0.21, 0.28]	[0.23, 0.29]
Gov. spending	0.53	0.53	0.53	0.54	0.54
	[0.48, 0.58]	[0.48, 0.58]	[0.48, 0.58]	[0,49, 0.59]	[0.49, 0.59]
Investment	0.45	0.45	0.61	0.45	0.64
	[0.37, 0.53]	[0.37, 0.53]	[0.53, 0.68]	[0.38, 0.53]	[0.58, 0.70]
MP shock	0.24	0.24	0.24	0.25	0.24
	[0.22, 0.27]	[0.22, 0.27]	[0.21, 0.26	[0.22, 0.27]	[0.21, 0.26]
Price mark-up	0.14	0.14	0.14	0.14	0.12
	[0.11, 0.17]	[0.11, 0.17]	[0.12, 0.16]	[0.11, 0.17]	[0.11, 0.14]
Wage mark-up	0.24	0.24	0.23	0.24	0.26
	[0.21, 0.28]	[0.20, 0.28]	[0.20, 0.26]	[0.20, 0.28]	[0.24, 0.29]
Gains	-	0.018	0.017	-	-
		[0.001, 0.034]	[0.006, 0.021]		
Marginal Likelihood	-922.8	-922.6	-911.0	-924.8	-920.3

Notes: This table reports the estimated values reported by Slobodyan and Wouters (2012b) compared with a replication of their results under RE and FB using the model of Smets and Wouters (2007).

parsimonious deviations from rationality can generate large improvements in fit. On the other hand, the dramatic improvements in the fit that can be obtained by deviating from rationality strongly support considering such approaches in empirical DSGE work. The underlying economic decisions that are captured by the DSGE framework remain perfectly intact within the infinite-horizon specification, for example, and we find that it fits the data best. Therefore, bounded rationality remains a promising way to reconcile DSGE models with data but should be approached cautiously.

Appendix A: The model

We describe the model and derive the households' and firms' decision rules following Preston (2005) that depend on expectations but not specifically on any explicit assumption for how expectations are formed. This gives us a general setting from which the consequences of different expectation assumption on the reduced form may be tracked systematically.

Households seek to maximize the following expected utility function

$$\tilde{\mathbb{E}}_{i,t} \sum_{t=0}^{\infty} \beta^t \left[a_t \ln(C_{i,t}) + \ln(M_{i,t}/P_t) - \eta^{-1} h_{i,t}^{\eta} \right] \tag{A1}$$

by choosing consumption, money holdings, labor supply, and by taking into account a preference shock a_t, where $\tilde{\mathbb{E}}_t$ represents a general expectations operator that is yet to be defined.[24] The representative household is faced with the budget constraint

$$M_{i,t-1} + B_{i,t-1} + T_t + W_t h_{i,t} + \Delta_{i,t} \geq P_t C_{i,t} + B_{i,t}/r_t + M_t, \tag{A2}$$

where M is nominal money balances, B is nominal bond, T is transfers, W is the nominal wage, and Δ is nominal profits the household receives from ownership of firms.

Production in the economy is separated into two sectors: a perfectly competitive finished goods sector and a monopolistically competitive intermediates goods sector. The finished goods sector uses a continuum of intermediates goods of prices P_j to construct the finished good. The production function is a CES constant returns to scale technology

$$\left(\int_0^1 Y_{j,t}^{(\theta_t - 1)/\theta_t} dj \right)^{\theta_t/(\theta_t - 1)} \geq Y_t, \tag{A3}$$

where θ_t is a cost push shock.[25] The finished-good-producing firms maximize profits

[24] It is assumed that $0 < \beta < 1$, $\eta \geq 1$, and $\ln(a_t) = \rho_a \ln(a_{t-1}) + \epsilon_{a,t}$.

[25] $\theta_t = (1 - \rho_\theta)\ln(\theta) + \rho_\theta \ln(\theta_{t-1}) + \epsilon_(\theta, t)$.

subject to demand for their good

$$Y_{j,t} = (P_{j,t}/P_t)^{-\theta_t} Y_t. \tag{A4}$$

The finished good price is given by

$$P_t = \left(\int_0^1 P_{j,t}^{1-\theta_t} dj \right)^{1/(1-\theta_t)} \tag{A5}$$

for all t. The intermediate-goods-producing firms hire $h_{j,t}$ units of labor to manufacture $Y_{j,t}$ units of outputs using

$$Z_t h_{j,t} \geq Y_{j,t}, \tag{A6}$$

where Z_t is an aggregate technology shock.[26] To introduce price stickiness, it is assumed that firms face an explicit cost to adjust nominal prices following Rotemberg (1982) that is measured in terms of finished goods

$$\frac{\phi}{2} \left(\frac{P_{j,t}}{\bar{\pi} P_{j,t-1}} - 1 \right)^2 Y_t. \tag{A7}$$

Lastly, the output gap is defined as the ratio between the actual and efficient levels of output

$$x_t = \left(\frac{1}{a_t} \right)^{1/\eta} \frac{Y_t}{Z_t}. \tag{A8}$$

Household decision rule

The first order conditions of the household's optimal decision are given by

$$\begin{aligned} a_t C_{i,t}^{-1} &= \beta r_t \tilde{\mathbb{E}}_{i,t} a_{t+1} C_{i,t+1}^{-1} \pi_{t+1}^{-} && \text{(A9)} \\ h_{i,t}^{\eta-1} &= a_t C_{i,t}^{-1} W_t P_t^{-1} && \text{(A10)} \\ B_{i,t-1} + W_t h_{i,t} + \Delta_{i,t} &= P_t C_{i,t} + B_t r_t^{-1}, && \text{(A11)} \end{aligned}$$

where we have eliminate the variables dealing with money and transfers in the budget constraint. Starting with budget constraint, we put it into real terms by dividing by the price level[27] and make it stationary using substitution to account for the unit root TFP

[26] $\ln(Z_t) = \ln(z) + \ln(Z_{t-1}) + \epsilon_{z,t}$.

[27] We assume $w_t = W_t/P_t$, $D_t = \Delta_t/P_t$, $b_t = B_t/P_t$.

process Z_t[28]

$$b_{i,t-1}\pi_t^{-1} + \omega_t Z_t h_{i,t} + d_{i,t} Z_t = c_{i,t} Z_t + b_{i,t} r_t^{-1}.$$

Then, rearranging and summing, we obtain the lifetime budget constraint

$$\sum_{T=t}^{\infty} \beta^{T-t} c_{i,T} Z_t = \sum_{T=t}^{\infty} \beta^{T-t} (w_T Z_T h_{i,T} + d_{i,T} Z_t),$$

which allows us to divide out Z_t. Then noting that $w_T h_{i,T} + d_{i,T} = y_{i,T}$, we have

$$\sum_{T=t}^{\infty} \beta^{T-t} \hat{c}_{i,T} = \sum_{T=t}^{\infty} \beta^{T-t} \hat{y}_{i,t}. \tag{A12}$$

We then log-linearize the stationary Euler-equation

$$\hat{c}_{i,t} = \tilde{\mathbb{E}}_{i,t} \hat{c}_{i,t+1} - (i_t - E_t \hat{\pi}_{t+1}) + (\rho_a - 1)\hat{a}_t$$

and solve it backwards recursively to get

$$\tilde{\mathbb{E}}_{i,t} \hat{c}_{i,T+1} = \hat{c}_{i,t} + \tilde{\mathbb{E}}_{i,t} \sum_{s=t}^{T} ((i_s - \hat{\pi}_{s+1}) - (\rho_a - 1)\hat{a}_t).$$

Then summing and discounting this expectation we get

$$\tilde{\mathbb{E}}_{i,t} \sum_{T=t}^{\infty} \beta^{T-t} \hat{c}_{i,T+1} = \frac{1}{1-\beta} \hat{c}_{i,t} + \frac{1}{1-\beta} \tilde{\mathbb{E}}_{i,t} \sum_{T=t}^{\infty} \beta^{T-t} ((i_T - \hat{\pi}_{T+1}) - (\rho_a - 1)\hat{a}_T). \tag{A13}$$

Combining Equation (A12) and (A13) yields the household's decision rule for consumption

$$\hat{c}_{i,t} = \tilde{\mathbb{E}}_{i,t} \sum_{T=t}^{\infty} \beta^{T-t} \left[(1-\beta) y_{j,T+1} - (i_T - \hat{\pi}_{T+1}) - (\rho_a - 1)\hat{a}_T \right].$$

Aggregating across households and using the log-linearized output gap, we obtain the aggregate IS curve

$$x_t = -\omega \hat{a}_t + \tilde{\mathbb{E}}_t \sum_{T=t}^{\infty} \beta^{T-t} \left[(1-\beta)(x_{T+1} + \omega \rho_a \hat{a}_T) - (i_T - \hat{\pi}_{T+1}) - (\rho_a - 1)\hat{a}_T \right], \tag{A14}$$

[28] We assume $\omega_t = W_t/(P_t Z_t)$, $d_t = \Delta_t/(P_t Z_t)$.

which is absent any assumption about how expectations are formed.

Firm decision rule

Firms maximize the present value of their companies

$$\tilde{\mathbb{E}}_{j,t} \sum_{T=t} Q_{t,T} P_T \Pi_{j,T}, \tag{A15}$$

where

$$\Pi_{j,T} = \left(\frac{P_{j,T}}{P_T} - \frac{MC_{j,T}}{P_T} \right) Y_{j,t} - \frac{\Phi}{2} \left(\frac{P_{j,T}}{\bar{\pi} P_{j,T-1}} - 1 \right)^2 Y_t \tag{A16}$$

and $Q_{t,T} = \beta^{T-t} \frac{a_T}{c_T}$. The first order condition of the firm's problem is

$$\begin{aligned} \Phi \left(\frac{P_{j,t}}{\bar{\pi} P_{j,t-1}} - 1 \right) \frac{1}{\bar{\pi} P_{j,t-1}} &= \tilde{\mathbb{E}}_{j,t} \left[\frac{Q_{t,t+1} Y_{t+1}}{Q_{t,t} Y_t} \Phi \left(\frac{P_{j,t+1}}{\bar{\pi} P_{j,t}} - 1 \right) \frac{P_{j,t+1}}{\bar{\pi} P_{j,t}^2} \right] \\ &+ \theta_t \left(\frac{P_{j,t}}{P_t} \right)^{-1-\theta_t} \left(\frac{w_t \theta_t - Z_t(\theta_t - 1) P_{j,t}}{P_t^2 Z_t} \right). \end{aligned} \tag{A17}$$

Because we assume $\bar{\pi} \geq 1$, the model does not have a steady state price level. Therefore, to make the model stationary, we define: $\hat{P}_{j,t+i} = P_{j,t+i}/P_{t+i}$ and $\pi_{t+i} = P_{t+i}/P_{t+i-1}$ for all i. Likewise, wages are made stationary by the following substitution $\omega_t = w_t/P_t Z_t$. Substituting in these definitions yields

$$\begin{aligned} \Phi \left(\frac{\hat{P}_{j,t} P_t}{\bar{\pi} \hat{P}_{j,t-1} P_{t-1}} - 1 \right) \frac{1}{\bar{\pi} \hat{P}_{j,t-1} P_{t-1}} &= \tilde{\mathbb{E}}_{j,t} \left[\frac{Q_{t,t+1} Y_{t+1}}{Q_{t,t} Y_t} \Phi \left(\frac{\hat{P}_{j,t+1} P_{t+1}}{\bar{\pi} \hat{P}_{j,t} P_t} - 1 \right) \frac{\hat{P}_{j,t+1} P_{t+1}}{\bar{\pi} \hat{P}_{j,t}^2 P_t^2} \right] \\ &+ \left(\hat{P}_{j,t} \right)^{-\theta_t} \left(\theta_t \omega_t + (1 - \theta_t) \hat{P}_{j,t} \right). \end{aligned}$$

Now, noting that market clearing implies $C_t = Y_t$, multiplying both sides by P_t, and using the definition of inflation we get

$$\begin{aligned} \Phi \left(\frac{\hat{P}_{j,t} \Pi_t}{\bar{\Pi} \hat{P}_{j,t-1}} - 1 \right) \frac{\Pi_t}{\bar{\pi} \hat{P}_{j,t-1}} &= \tilde{\mathbb{E}}_{j,t} \left[\frac{a_{t+1}}{a_t} \Phi \left(\frac{\hat{P}_{j,t+1} \Pi_{t+1}}{\bar{\pi} \hat{P}_{j,t}} - 1 \right) \frac{\hat{P}_{j,t+1} \Pi_{t+1}}{\bar{\pi} \hat{P}_{j,t}^2} \right] \\ &+ \left(\hat{P}_{j,t} \right)^{-\theta_t} \left(\theta_t \omega_t + (1 - \theta_t) \hat{P}_{j,t} \right). \end{aligned}$$

Log-linearizing this expression

$$-\Phi p_{j,t-1}+(\Phi(\beta+1)-1+\bar{\theta})p_{j,t}-\Phi\beta\tilde{\mathbb{E}}_{j,t}p_{j,t+1}=\Phi\beta\tilde{\mathbb{E}}_{j,t}\pi_{t+1}-\Phi\pi_t-\hat{\omega}_t(1-\bar{\theta})-\hat{\theta}_t \quad \text{(A18)}$$

and introducing lag polynomials, we can write this as

$$\left(\frac{1}{\beta}L^2-\frac{\Phi(\beta+1)-1+\bar{\theta}}{\Phi\beta}L+1\right)\tilde{\mathbb{E}}_{j,t}p_{j,t+1}=\frac{1}{\beta}\left(\pi_t-\beta E_t\pi_{t+1}\right)+\frac{1-\bar{\theta}}{\beta\Phi}\hat{\omega}_t+\frac{1}{\Phi\beta}\hat{\theta}_t. \quad \text{(A19)}$$

Factoring the lag polynomial and solving forward the unstable root yields

$$\begin{aligned}(1-\lambda_1 L)p_{j,t} &= \frac{-\lambda_2^{-1}L^{-1}}{1-\lambda_2^{-1}L^{-1}}\frac{L}{\beta}\left(\pi_t-\beta E_t\pi_{t+1}+\frac{1-\bar{\theta}}{\Phi}\hat{\omega}_t+\frac{1}{\Phi}e_t\right)\\ &= -\lambda_1\tilde{\mathbb{E}}_{j,t}\sum_{T=t}^{\infty}(\lambda_1\beta)^{T-t}\pi_T+\lambda_1\beta\tilde{\mathbb{E}}_{j,t}\sum_{T=t}^{\infty}(\lambda_1\beta)^{T-t}\pi_{T+1}\\ &\quad +\tilde{\mathbb{E}}_{j,t}\sum_{T=t}^{\infty}\beta^{T-t}(\psi x_T-e_T).\end{aligned}$$

where $(\bar{\theta}-1)\eta\Phi^{-1}\omega_t=\psi x_t$, $\Phi^{-1}\hat{\theta}_t=e_t$, $0<\lambda_1<1$, $\lambda_2>1$, $\lambda_2=\frac{1}{\lambda_1\beta}$, and $\lambda_1+\lambda_2=(\Phi\beta)^{-1}(\Phi(\beta+1)-1+\bar{\theta})$. Combining terms we have[29]

$$(1-\lambda_1 L)p_{j,t} = -\pi_t+\lambda_1\tilde{\mathbb{E}}_{j,t}\sum_{T=t}^{\infty}(\lambda_1\beta)^{T-t}\frac{1-\lambda_1}{\lambda_1}\pi_T+\psi x_T-e_T.$$

Finally, aggregating across firms yields the aggregate Phillips curve, which is free of any expectations assumptions

$$\pi_t=\lambda_1\tilde{\mathbb{E}}_t\sum_{T=t}^{\infty}(\lambda_1\beta)^{T-t}\left(\frac{1-\lambda_1}{\lambda_1}\pi_T+\psi x_T-e_T\right). \quad \text{(A20)}$$

Monetary policy

The model is closed with a standard contemporaneous Taylor rule

$$i_t=\bar{r}+\bar{\pi}+\theta_\pi(\pi_t-\bar{\pi})+\theta_x x_t+\epsilon_{i,t}, \quad \text{(A21)}$$

[29] noting that $-\lambda_1\pi_t-\lambda_1^2\beta\pi_{t+1}-\lambda_1^3\beta^2\pi_{t+2}...$ and $\lambda_1\beta\pi_{t+1}+(\lambda_1\beta)^2\pi_{t+2}+(\lambda_1\beta)^3\pi_{t+3}...$ and $-\lambda_1\pi_t+\lambda_1\beta(1-\lambda_1)\pi_{t+1}+(\lambda_1\beta)^2(1-\lambda_1)\pi_{t+2}+(\lambda_1\beta)^3(1-\lambda_1)\pi_{t+3}....$

where $\epsilon_{i,t}$ is an i.i.d. monetary policy shock.

Appendix B: Smets and Wouters replication

Table 8: Smets and Wouters' Model

	RE							FB					
	Prior Mean	Post. Mean	90% HPD Interval		Prior	P. St. Dev		Prior Mean	Post. Mean	90% HPD Interval		Prior	P. St. Dev
ρ_a (ρ_z)	0.5	0.955	0.936	0.974	beta	0.2	ρ_a (ρ_z)	0.5	0.956	0.927	0.986	beta	0.2
ρ_b (ρ_a)	0.5	0.204	0.070	0.337	beta	0.2	ρ_b (ρ_a)	0.5	0.205	0.036	0.358	beta	0.2
ρ_g	0.5	0.966	0.951	0.981	beta	0.2	ρ_g	0.5	0.967	0.946	0.989	beta	0.2
ρ_l	0.5	0.742	0.642	0.844	beta	0.2	ρ_l	0.5	0.714	0.565	0.863	beta	0.2
ρ_r (ρ_i)	0.5	0.148	0.044	0.249	beta	0.2	ρ_r (ρ_i)	0.5	0.141	0.041	0.234	beta	0.2
ρ_p (ρ_e)	0.5	0.895	0.817	0.977	beta	0.2	ρ_p (ρ_e)	0.5	0.897	0.829	0.968	beta	0.2
ρ_w	0.5	0.968	0.948	0.988	beta	0.2	ρ_w	0.5	0.950	0.915	0.987	beta	0.2
μ_p	0.5	0.715	0.547	0.882	beta	0.2	μ_p	0.5	0.699	0.605	0.791	beta	0.2
μ_w	0.5	0.830	0.728	0.938	beta	0.2	μ_w	0.5	0.829	0.769	0.890	beta	0.2
ϕ	4	5.382	3.688	7.102	norm	1.5	ϕ	4	4.657	3.147	6.113	norm	1.5
σ_c	1.5	1.381	1.163	1.593	norm	0.375	σ_c	1.5	1.006	0.421	1.561	norm	0.375
h	0.7	0.713	0.643	0.788	beta	0.1	h	0.7	0.704	0.579	0.833	beta	0.1
ξ_w	0.5	0.687	0.577	0.803	beta	0.1	ξ_w	0.5	0.684	0.590	0.778	beta	0.1
σ_l	2	1.532	0.598	2.457	norm	0.75	σ_l	2	1.887	0.829	2.894	norm	0.75
ξ_p	0.5	0.652	0.555	0.745	beta	0.1	ξ_p	0.5	0.596	0.500	0.682	beta	0.1
ι_w	0.5	0.568	0.358	0.774	beta	0.15	ι_w	0.5	0.601	0.391	0.817	beta	0.15
ι_p	0.5	0.241	0.096	0.382	beta	0.15	ι_p	0.5	0.384	0.192	0.589	beta	0.15
ψ	0.5	0.476	0.306	0.648	beta	0.15	ψ	0.5	0.418	0.219	0.612	beta	0.15
Ψ	1.25	1.703	1.577	1.825	norm	0.125	Ψ	1.25	1.652	1.530	1.775	norm	0.125
r_π (θ_π)	1.5	2.054	1.765	2.354	norm	0.25	r_π (θ_π)	1.5	1.740	1.406	2.071	norm	0.25
ρ	0.75	0.817	0.776	0.857	beta	0.1	ρ	0.75	0.880	0.843	0.918	beta	0.1
r_y (θ_x)	0.125	0.090	0.051	0.126	norm	0.05	r_y (θ_x)	0.125	0.112	0.053	0.171	norm	0.05
r_{Deltay}	0.125	0.217	0.173	0.262	norm	0.05	r_{Deltay}	0.125	0.190	0.145	0.234	norm	0.05
$\bar{pi}$	0.625	0.654	0.535	0.770	gamma	0.1	$\bar{pi}$	0.625	0.631	0.497	0.765	gamma	0.1
$100(\beta^{-1}-1)$	0.25	0.251	0.093	0.407	gamma	0.1	$100(\beta^{-1}-1)$	0.25	0.252	0.091	0.401	gamma	0.1
$\bar{l}$	0	0.291	-1.285	1.912	norm	2	$\bar{l}$	0	-0.066	-1.635	1.518	norm	2
$\bar{\gamma}$	0.4	0.442	0.415	0.470	norm	0.1	$\bar{\gamma}$	0.4	0.440	0.413	0.467	norm	0.1
$\bar{g}$	0.5	0.609	0.454	0.766	norm	0.25	$\bar{g}$	0.5	0.600	0.442	0.760	norm	0.25
α	0.3	0.292	0.218	0.362	norm	0.05	α	0.3	0.266	0.194	0.335	norm	0.05
σ_a (σ_z)	0.1	0.432	0.386	0.474	invg	2	σ_a (σ_z)	0.1	0.443	0.397	0.488	invg	2
σ_b (σ_a)	0.1	0.242	0.205	0.282	invg	2	σ_b (σ_a)	0.1	0.259	0.232	0.286	invg	2
σ_g	0.1	0.540	0.488	0.590	invg	2	σ_g	0.1	0.538	0.486	0.588	invg	2
σ_l	0.1	0.453	0.375	0.529	invg	2	σ_l	0.1	0.639	0.575	0.703	invg	2
σ_r (σ_i)	0.1	0.246	0.220	0.270	invg	2	σ_r (σ_i)	0.1	0.235	0.212	0.258	invg	2
σ_p (σ_e)	0.1	0.138	0.108	0.166	invg	2	σ_p (σ_e)	0.1	0.123	0.106	0.138	invg	2
σ_w	0.1	0.242	0.204	0.279	invg	2	σ_w	0.1	0.262	0.236	0.287	invg	2

Notes: Prior and posterior distributions for the model of Smets and Wouters (2007) estimated under RE and FB. The main labels correspond to parameters names in Smets and Wouters' paper. The parameters names in parenthesis show the corresponding parameter in the Ireland model.

References

Amato, Jeffery D and Thomas Laubach, "Rule-of-thumb behaviour and monetary policy," *European Economic Review*, 2003, *47* (5), 791–831.

Ball, Laurence, N Gregory Mankiw, and Ricardo Reis, "Monetary policy for inattentive economies," *Journal of Monetary Economics*, 2005, *52* (4), 703–725.

Berardi, Michele and Jaqueson K Galimberti, "Empirical calibration of adaptive learning," *Journal of Economic Behavior & Organization*, 2017.

Chevillon, Guillaume and Sophocles Mavroeidis, "Learning can generate long memory," *Journal of Econometrics*, 2017, *198* (1), 1–9.

Chow, Gregory C, "Rational Versus Adaptive Expectations in Present Value Models," *Review of Economics and Statistics*, 1989.

Clark, Todd E and Michael W McCracken, "Tests of equal predictive ability with real-time data," *Journal of Business & Economic Statistics*, 2009, *27* (4), 441–454.

_ and _ , "Advances in forecast evaluation," *working paper*, 2011.

Cogley, Timothy and Argia M Sbordone, "Trend inflation, indexation, and inflation persistence in the New Keynesian Phillips curve," *The American Economic Review*, 2008, *98* (5), 2101–2126.

Coibion, Olivier and Yuriy Gorodnichenko, "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, August 2015, *105* (8), 2644–2678.

Cole, Stephen J and Fabio Milani, "The misspecification of expectations in New Keynesian models: a DSGE-VAR approach," *Macroeconomic Dynamics*, 2019, *23* (3), 974–1007.

Cornea-Madeira, Adriana, Cars Hommes, and Domenico Massaro, "Behavioral heterogeneity in US inflation dynamics," *Journal of Business & Economic Statistics*, 2017, (just-accepted).

Diebold, Francis X and Robert S Mariano, "Comparing predictive accuracy," *Journal of Business & economic statistics*, 1995, *13* (3).

Eusepi, Stefano and Bruce Preston, "Expectations, learning, and business cycle fluctuations," *The American Economic Review*, 2011, *101* (6), 2844–2872.

_ and _ , "Fiscal foundations of inflation: imperfect knowledge," *American Economic Review*, 2018, *108* (9), 2551–89.

Evans, George W and Seppo Honkapohja, *Learning and expectations in macroeconomics*, Princeton: Princeton University Press, January 2001.

Fuhrer, Jeffrey C, "Habit formation in consumption and its implications for monetary-policy models," *American Economic Review*, 2000, pp. 367–390.

Galı, Jordi and Mark Gertler, "Inflation dynamics: A structural econometric analysis," *Journal of Monetary Economics*, 1999, *44* (2), 195–222.

Gaus, Eric and Srikanth Ramamurthy, "Estimation of Constant Gain Learning Models," *mimeo*, August 2012.

Harvey, David, Stephen Leybourne, and Paul Newbold, "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, 1997, *13*, 281–291.

Hommes, Cars, *Behavioral rationality and heterogeneous expectations in complex economic systems*, Cambridge University Press, 2013.

Ireland, Peter N, "Technology Shocks in the New Keynesian Model," *The Review of Economics and Statistics*, November 2004, *86* (4), 923–936.

Mankiw, N Gregory and Ricardo Reis, "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics*, 2002, pp. 1295–1328.

McCallum, Bennett T, "Analysis of the Monetary Transmission Mechanism: Methodological Issues," in "The Monetary Transmission Mechanism" 2001.

Milani, Fabio, "A Bayesian DSGE Model with Infinite-Horizon Learning: Do" Mechanical" Sources of Persistence Become Superfluous?," *International Journal Of Central Banking*, 2006.

_ , "Expectations, learning and macroeconomic persistence," *Journal of Monetary Economics*, 2007, *54* (7), 2065–2082.

_ , "Expectation shocks and learning as drivers of the business cycle," *The Economic Journal*, 2011, *121* (552), 379–401.

Negro, Marco Del and Stefano Eusepi, "Fitting observed inflation expectations," *Journal of Economic Dynamics and control*, 2011, *35* (12), 2105–2131.

_ , Frank Schorfheide, Frank Smets, and Rafael Wouters, "On the Fit of New Keynesian Models," *Journal of Business and Economic Statistics*, April 2007, *25* (2), 123–143.

Ormeño, Arturo and Krisztina Molnár, "Using survey data of inflation expectations in the estimation of learning and rational expectations models," *Journal of Money, Credit and Banking*, 2015, *47* (4), 673–699.

Preston, Bruce, "Learning about Monetary Policy Rules when Long-Horizon Expectations Matter," *International Journal of Central Banking*, 2005.

Rotemberg, Julio J, "Monopolistic price adjustment and aggregate output," *The Review of Economic Studies*, 1982, *49* (4), 517–531.

Rychalovska, Yuliya, "The implications of financial frictions and imperfect knowledge in the estimated DSGE model of the US economy," *Journal of Economic Dynamics and Control*, 2016, *73*, 259–282.

Slobodyan, Sergey and Raf Wouters, "Learning in a Medium-Scale DSGE Model with Expectations Based on Small Forecasting Models," *American Economic Journal: Macroeconomics*, April 2012, *4* (2), 65–101.

_ **and** _ , "Learning in an estimated medium-scale DSGE model," *Journal of Economic Dynamics and Control*, 2012, *36* (1), 26–46.

Smets, Frank and Rafael Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *The American Economic Review*, 2007, pp. 586–606.

Stock, James H and Mark W Watson, "Median unbiased estimation of coefficient variance in a time-varying parameter model," *Journal of the American Statistical Association*, 1998, *93* (441), 349–358.

_ **and** _ , "Why has US inflation become harder to forecast?," *Journal of Money, Credit and banking*, 2007, *39*, 3–33.