Essays in Behavioral Macroeconomics

by

Jacob Thompson

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Economics

Dissertation Committee:

Bruce McGough, Chair

Jeremy Piger, Core Member

David Evans, Core Member

Robert Ready, Institutional Representative

University of Oregon

Spring 2024

DISSERTATION ABSTRACT

Jacob Thompson

Doctor of Philosophy in Economics

Title: Essays in Behavioral Macroeconomics

This dissertation investigates a class of DSGE models with bounded rationality where agents use recursively updated forecasts to form expectations of future variables The two chapters explore the implications of the model builder's choice of initial forecasting model with which to endow agents. Each chapter estimates a different New Keynesian DSGE model, varying this initial model and finds that this has substantial impacts on parameter estimates as well as the ability of the model to fit macroeconomic data series.

Chapter 1 estimates a small scale, purely forward-looking DSGE model but relaxes the assumption of rational expectations. In so doing, it outlines the computational challenges of estimating such a model and the solutions thereto. It also introduces the reader to a new class of Bayesian posterior sampler called Sequential Monte Carlo which has key advantages over Markov Chain Monte Carlo samplers for the estimation of models with Adaptive Learning. I find two notable results: first, I find that one can greatly improve the ability of the model to explain the data by training agents' initial forecasting model on pre-sample data. Second, I find that, for this particular DSGE model, the estimated slope of the Phillips Curve is significantly greater than under Rational Expectations.

Chapter 2 estimates a small-scale DSGE model with habit persistence in household consumption and inflation indexation by price-setting firm, thereby inducing mechanical persistence in both the output and inflation processes. This chapter shows that the improved data-fit from training sample based initial beliefs is robust to the inclusion of mechanical lags. It also shows how initial forecasting models trained on pre-sample data cause the DSGE model to exhibit impulse response functions that show the "price puzzle" despite the additional restrictions of the DSGE model, and what restrictions to impose to avoid this outcome.

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

Figure | Page

LIST OF TABLES

CHAPTER 1

BAYESIAN COMPARISON OF INITIAL BELIEFS IN A FORWARD-LOOKING NEW KEYNESIAN DSGE MODEL

1.1 Introduction and Literature Review

Owing to the apparent failure of Neo-Keynesian Macroeconomics in the 1970s and the rise of New Classical Macroeconomics in the 1980s, macroeconomic modelers have paid much closer attention to how firms', workers', and consumers' expectations of the future affect the economy. As to how agents form such expectations, the Rational Expectations Hypothesis provides one very natural baseline assumption. In the context of macroeconomic modeling, this hypothesis posits that agents will form expectations so as to match the same expectations that the model itself would yield. While New Classical Macroeconomics had some early empirical success in appearing to replicate important features of American business cycles, as argued famously by Kydland and Prescott (1982), the New Keynesian framework has ascended to preeminence among micro-founded macroeconomic models, as documented in the extensive review by Fernandez-Villaverde (2009). Throughout these successive paradigm shifts macroeconomic modelers have retained the assumption of rational expectations. The Rational Expectations Hypothesis faces withering criticism, however, including heroic informational assumptions on the part of the modeled agents. Because of this, a nascent empirical literature has developed which follows the spirit of New Classical Economics and models the economy as a dynamic equilibrium based on optimizing behavior, but relaxes the assumption of rational expectations.

This chapter extends the empirical literature by estimating the benchmark model of An and Schorfheide (2007), but under an alternative expectations formation hypothesis. In my study, I relax the assumption of rational expectations and let agents use reduced-form econometric models with incomplete information to form expectations. These models are updated recursively as agents observe new data according to an Adaptive Learning algorithm. Because the the agents use a recursive updating rule, the researcher must ultimately choose with which beliefs to initialize agents' learning algorithm. My principal exercise investigates three choices of initial beliefs: equilibrium-based initial beliefs, training-sample based initial beliefs, and jointly estimated initial beliefs. I investigate how these choices affect the estimates of structural parameters as well as the model's ability to fit the data, as measured by the estimated marginal data density, a common measure for ranking model fit in Bayesian econometrics.

I begin by reviewing the literature on solution and estimation methods of linear DSGE models, with a special emphasis on Bayesian techniques, which have taken center stage in contemporary research. I also include a review of the important results from the Adaptive Learning literature, including some of the important asymptotic properties of DSGE models with adaptive learning. I also review the important empirical research, notably Milani (2007) and Slobodyan and Wouters (2012) that estimates well-known DSGE models augmented to use expectations formed through adaptive learning. These studies demonstrate many empirical improvements in relaxing the rational expectations hypothesis in favor of expectations formed through Adaptive Learning.

After reviewing the relevant literature, I outline the unique computational and econometric challenges presented by estimating a model with adaptive learning. These challenges arise from the adaptive learning updating scheme, which is non-linear, and incorporating it into an estimated linear dynamical system. I detail how to incorporate these non-linear updating rules into the linear system so as to make the model amenable to estimation via Bayesian methods.

Finally, I present the results of the estimation exercise in the form of parameter estimates and estimated marginal data densities. I show that for this particular model, jointly estimated initial beliefs demonstrate a marked improvement in marginal data density over every other model, and that all learning models perform significantly better than the rational expectations baseline.

By providing a rigorous analysis of a benchmark DSGE model with adaptive learning, this chapter both extends the empirical macroeconomics literature and provides a toolkit for macroeconomic modelers to use when solving, simulating, and estimating their own DSGE models with adaptive learning.

Bayesian Estimation of DSGE Models

I first review solution and estimation techniques for DSGE models, with a special emphasis on those that relate to estimation of such models with adaptive learning. Estimation of DSGE modeling currently uses likelihood-based methods rather than moment-matching or older calibration methods. Fernández-Villaverde and Guerrón-Quintana (2020) provides a few of the reasons for this outcome, one of which is that the ever-growing complexity of DSGE models has created far too many moments for modelers to match. A numerical likelihood function, by contrast, captures all the information in the relevant data. Likelihood based estimation methods fall under either Frequentist or Bayesian categories. Of these, Maximum likelihood estimation constitutes the most common Frequentist based estimation technique. However, for many reasons outlined by Fernandez-Villaverde (2009), some of which I shall recount, DSGE models are most commonly estimated via Bayesian methods instead of maximum likelihood. He observes that, first, the data with which DSGE models are estimated are frequently very sparse quarterly data that date back to the 1960s at the earliest and so have a relatively small sample size compared to the sort of data sets that applied micro economists would commonly use. This small sample size means that the asymptotic distributions used for inference in maximum likelihood estimation are usually poor approximations to the true limit of the maximum of the likelihood function. Second, most DSGE models have a great many structural parameters to be estimated by the researcher; even the simplest models can have up to a dozen estimated parameters. This results in a state of affairs where multiple clusters of parameter choices can give rise to the same data series, and therefore a likelihood function of a DSGE model can have many local maxima. For these reasons, Bayesian estimation is much less computationally intensive than maximum likelihood estimation. Thus I recount the explanation offered in Fernandez-Villaverde (2009), along with Herbst and Schorfheide (2016) at a fairly high level, the basics of Bayesian econometrics

Let $y^T = \{y_t\}_{t=1}^T \in \mathbb{R}^{n\times T}$ be a time series of data, with $n \in \mathbb{N}$ variables for $T \in \mathbb{N}$ periods, with which an econometrician wants to estimate their DSGE model which has $k \in \mathbb{N}$ parameters. A Bayesian model consists of three objects

1. A parameter set, $\Theta \in \mathbb{R}^k$,
2. A likelihood function: $p(y^T|\theta) : \mathbb{R}^{n\times T} \times \Theta \to \mathbb{R}$,
3. A prior density function $\pi(\theta) : \mathbb{R}^{n\times T} \times \Theta \to \mathbb{R}$.

From these I obtain the fundamental object of interest, a posterior distribution of the model parameters,

$$\pi(\theta_i|y^T) = \frac{p(y^T|\theta_i)\pi(\theta_i)}{\int p(y^T|\theta)\pi(\theta)d\theta},$$

which defines for every $\theta_i \in \Theta$ the probability of this parameter vector given the data series y^T.

Bayesian inference requires only these three objects. Prior densities are defined analytically, and so the challenge for the researcher is to compute the likelihood function and to sample from the posterior distribution.

For (virtually) no DSGE models do analytic likelihood functions exist and thus, for any parameter vector θ_i, the likelihood value $p(y^T|\theta_i)$ for any DSGE model can only be computed numerically. This requires the tools of filtering theory. As the model I estimate in this chapter is a linear model with gaussian shocks, I limit my attention to the Kalman filter, although there exists a growing empirical literature using non-linear particle filters for estimating DSGE models, an introduction to which can be found in Herbst and Schorfheide (2016). The assumption of gaussian innovations and linear dynamics certainly does not come without cost, however, which I should admit. First, as noted in Justiniano and Primiceri (2008), such assumptions do not allow for the possibility of time-varying volatility in the shocks that affect the economy, which many authors have found explains certain key features of US macroeconomic data. Second, linear dynamics imposes certainty equivalence upon the agents in the model and eliminates the possibility of precautionary saving. While this is a problem in richer DSGE models such as that of Smets and Wouters (2007), this is not a problem in the model I estimate presently as, in equilibrium, all output is consumed.

Recounting the explanation offered in Fernandez-Villaverde et al. (2015), I present how a generic filter provides the likelihood function for a dynamic model with a markov structure, which is described by the following measurement and transition equations:

$$\begin{aligned} y_t &= \Psi(s_t, t; \theta) + u_t, u_t \sim F_u(\cdot; \theta), \\ s_t &= \Phi(s_{t-1}, \epsilon_t; \theta), \epsilon_t \sim F_\epsilon(\cdot; \theta). \end{aligned} \tag{1.1}$$

The first equation, the measurement equation, states how the time series y_t relates to the state s_t. This is the measurement equation. The second equation, the transition equation states how the possibly unobserved state s_t evolves given s_{t-1} and i.i.d. shocks ε_t. Since one has a shock term $u_t \sim F_u(\cdot; \theta)$ included in the transition function Φ, the function Φ along with the distribution F_u generate a density function $p(s_t|s_{t-1})$ This is because the Φ function depends only on the current state. Letting the prior distribution be $p(s_0) = p(s_0|Y_{1:0})$, filter-based computation of the likelihood function proceeds as follows

---

Algorithm 1 Generic Filtering Algorithm

1: Forecasting t given $t-1$
   - Transition equation: $p(s_t|Y_{1:t-1}) = \int p(s_t|s_{t-1}, Y_{1:t-1})p(s_{t-1}|Y_{1:t-1})ds_{t-1}$
   - Measurement equation: $p(y_t|Y_{1:t-1}) = \int p(y_t|s_t, Y_{1:t-1})p(s_t|Y_{1:t-1})ds_t$

2: Updating: once y_t becomes available, update the distribution of states s_t
   - $p(s_t|Y_{1:T}) = p(s_t|y_t, Y_{1:t-1}) = \frac{p(y_t|s_t,Y_{1:t-1})p(s_t|Y_{1:t-1})}{p(y_t|Y_{1:t-1})}$

Computing the integrals in the measurement and the transition equations, then, presents the main challenge for the DSGE modeler. I limit my attention in this chapter and the next to models with linear dynamics and Gaussian i.i.d. shocks. This greatly simplifies the task of computing these integrals because the entirety of a normal distribution, including those associated integrals, is described by its mean and variance, and this generalizes to a multivariate normal distribution as well. Thus to compute the likelihood function, the modeler need only keep track of a sequence of means and variances of a sequence of distributions rather than the whole distribution of states.

To obtain these means and variances, then, one must obtain the measurement and transition equations for the DSGE model. As I use linear models in both chapters, both the transition and measurement equations are matrix equations:

$$
\begin{aligned}
y_t &= A + Bs_t + Hu_t, u_t \sim F_u(\cdot;\theta), \\
s_t &= T_c(t,\theta) + T_1(t,\theta)s_{t-1} + T_0(\theta)\epsilon_t, \epsilon_t \sim F_\epsilon(\cdot;\theta).
\end{aligned}
$$

The transition matrices T_c, T_1, T_0 cannot be obtained analytically and must be computed numerically. For many linear rational expectations models this can be done numerically using a Schur decomposition, as described first in Sims (2002), which is also detailed in a highly accessible manner in Herbst and Schorfheide (2016). The measurement equation, on the other hand, is defined analytically by the modeler themselves. I recount now briefly the Schur decomposition method to solve linear rational expectations models.

Consider some linearized DSGE model with a vector of variables x_t and some vector of forward-looking expectational terms $E_t x_{t+1}$. I seek to represent the model as a VAR(1) in the $n \times 1$ state vector $s_t = (x_t, E_t x_{t+1})'$. I proceed first by writing the (unsolved) model in the following form:

$$\Gamma_0(\theta)s_t = \Gamma_1(\theta)s_{t-1} + \Psi(\theta)\epsilon_t + \Pi(\theta)\eta_t, \tag{1.2}$$

where ϵ_t is a vector of i.i.d innovations and η_t is a vector of forecasting errors, $x_t - E_{t-1}x_t$. The system matrices $\Gamma_0, \Gamma_1, \Psi, \Pi$ are explicit functions of the model parameters θ.

For many DSGE models, including the ones I study in the present and next chapters, a solution exists only if a set of transversality conditions are met, and these transversality conditions met if and only if the associated law of motion for s_t is non-explosive. For some parameter draw θ_i, the DSGE model is said to be determinate if there exists a unique stable solution and indeterminate if there exist multiple stable solutions. When drawing from the posterior distribution of parameters $p(\theta_i|Y)$, I discard all parameter draws except for those with unique, stable solutions. To find a unique stable solution to this system, if one exists, I transform the above system through a generalized complex Schur decomposition, (also known as a "QZ decomposition"). This decomposition seeks to find $n \times n$ matrices Q, Z, Λ, Ω such that the following matrix equations hold hold:

$$\begin{aligned} Q'\Lambda Z' &= \Gamma_0, \\ Q'\Omega Z' &= \Gamma_1, \\ QQ' = ZZ' &= I, \end{aligned}$$

wherein Λ and Ω are upper-triangular matrices. Letting $w_t = Z's_t$, I then pre-multiply (1.2) by Q to obtain the following:

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ 0 & \Omega_{22} \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} (\Psi\epsilon_t + \Pi\eta_t).$$

Assuming that the system is ordered such that the $m \times 1$ vector $w_{2,t}$ is purely explosive, and that $0 \leq m \leq n$, the second matrix equation in the above can be re-written thusly:

$$w_{2,t} = \Lambda_{22}^{-1}\Omega_{22}w_{2,t-1} + \Lambda_{22}^{-1}Q_2(\Psi\epsilon_t + \Pi\eta_t).$$

Under the following conditions: that $w_{2,0} = 0$, that one can find an η_t for every possible ϵ_t that offsets the effect of ϵ_t upon $w_{2,t}$, or equivalently,

$$Q_2\Psi\epsilon_t + Q_t\Pi\eta_t = 0,$$

there exists a non-explosive solution to the system of linear rational expectations difference equations in (1.2).

Such a solution is unique if the number of forecasting errors equals the number of explosive elements in w_t. This is equivalent to the conditions for uniqueness and existence in Blanchard and Kahn (1980). If such a unique solution exists, then the sequence of forecasting errors that assure stability is:

$$\eta_t = -(Q_2\Pi)^{-1}Q_2\Psi\epsilon_t.$$

With a solution for the path of forecasting errors η_t, one can obtain the path of non-explosive solutions from $s_t = Zw_t$. If the solution is unique, one can now represent the path of s_t as a VAR(1):

$$s_t = T_c + T_1 s_{t-1} + T_0\epsilon_t.$$

This constitutes the transition equation for our filter from 1.1. Having obtained the transition equation, I now review how to compute the likelihood function.

The likelihood function for a DSGE model is computed through a prediction-error decomposition. For this I rely principally upon Zivot (2006), Ljungqvist and Sargent (2012), and Hamilton (1994). Suppose the DSGE model has a transition equation $s_t = T_c + T_1 s_{t-1} + T_0\epsilon_t, \epsilon_t \sim N(0, H_t)$ as derived above from the Schur decomposition. Suppose further that the underlying state s_t relates to an observed data series y_t through the measurement equation $y_t = A + Bs_t + Cu_t, u_t \sim N(0, Q_t)$, and finally suppose that the system matrices $T_c, T_1, T_0, A, B, C, H_t, Q_t$ are time-invariant so that one may compute numerically the initial state s_0 and its variance P_0. This numerical computation may be done by a kronecker product vectorization technique as described in Zivot (2006) or much more quickly using a Schur decomposition as described in Villemot (n.d.).

The Kalman Filter gives a sequence of predictive distributions of the unobserved and observed variables according to the equations below

$$
\begin{aligned}
y_t &= A + Bs_t + u_t, & u_t &\sim N(0, H_t),\\
s_t &= T_1\alpha_{t-1} + T_c + T_0\varepsilon_t, & \varepsilon_t &\sim N(0, Q_t),
\end{aligned}
$$

where y_t is an $N \times 1$ vector of observable variables and s_t an $m \times 1$ vector of (possibly) unobserved states. For a sequence of system matrices $\{A, B, H_t, T_1, T_c, T_0, Q_t\}_{t=1}^{t=T}$, let $a_t = E(s_t|y^t)$ be the optimal forecast of the state given information through time t, and $P_t = E((a_t - s_t)(a_t - s_t)'|y^t)$ be the variance of that optimal forecast. The Kalman filter consists of prediction and updating equations.

The prediction equations describe $E(a_t|y^{t-1}) = a_{t|t-1}$ and $E(P_t|y^{t-1}) = P_{t|t-1}$. Those prediction equations are:

$$
\begin{aligned}
a_{t|t-1} &= T_1 a_{t-1} + T_c,\\
P_{t|t-1} &= BP_{t-1}B' + T_0 Q_t T_0'.
\end{aligned}
$$

From these prediction equations, one can compute the optimal predictor of y_t based on the information set $y^{t-1} \equiv \{y_t\}_0^{t-1}$, the prediction error v_t and the prediction error variance $E(v_t v_t')$

$$
\begin{aligned}
y_{t|t-1} &= A + Ba_{t|t-1},\\
v_t &= y_t - y_{t|t-1},\\
E(v_t v_t') &= F_t = BP_{t|t-1}B' + H_t.
\end{aligned}
$$

The updating equations allow one to update a_t and P_t:

$$
\begin{aligned}
a_t &= a_{t|t-1} + P_{t|t-1}B'F_t^{-1}v_t,\\
P_t &= P_{t|t-1} - P_{t|t-1}B'F_t^{-1}BP_{t|t-1},
\end{aligned}
$$

which, when the system is linear and the innovations Gaussian, allows us to derive the prediction-error-decomposition of the likelihood function:

$$
\ln(L(\theta|y)) = -\frac{NT}{2} ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \ln(|F_t(\theta)|) - \frac{1}{2}\sum_{t=1}^{T} v_t(\theta)' F_t^{-1}(\theta) v_t(\theta).
$$

Once the researcher has obtained the likelihood function, they must choose an appropriate algorithm to simulate the posterior distribution of the model parameters in order to estimate the the model. One very popular class of models still in use today is Markov Chain Monte Carlo models, such as the Metropolis Hastings Random Walk ("MHRW" henceforth), first detailed in Metropolis et al. (1953). The MHRW algorithm works as follows:

Algorithm 2 MHRW Chain Algorithm

1: *Initialization*
- Choose initial parameter θ_0

2: *Iterate:* For each element i in the MHRW chain:
- Draw θ' from a proposal density $p(\theta'|\theta_{i-1})$
- Compute the acceptance probability: $\min\left(1, \frac{p(\theta'|y)p(\theta')}{p(\theta_{i-1}|y)p(\theta_{i-1})}\right)$

For approximately normal posterior distributions, a common choice of proposal density is a multivariate normal distribution whose variance-covariance matrix is equal to the negative inverse-Hessian matrix of the posterior at its mode, times a tuning constant to target an acceptance rate, usually between 20% and 40%. This requires said matrix to be a positive definite one, however for many DSGE models, including the models studied in this and the next chapter, both under Adaptive Learning and under Rational Expectations, for some datasets, this Hessian matrix may not be positive definite. This lack of a positive definite matrix at the mode of the posterior can reduce greatly the efficiency of an MHRW sampler and create inconsistent monte carlo averages from one MHRW run to another. This inability of numerical optimizers to find a positive definite inverse Hessian matrix is one of the reasons I opt in both chapters to estimate the models with a Sequential Monte Carlo sampler rather than a Markov Chain Monte Carlo sampler. I shall now describe, recounting the explanation offered by Herbst and Schorfheide (2013) the Sequential Monte Carlo method as it applies to estimation of DSGE models.

SMC samplers seek to simulate the posterior distribution by assigning probability weights, or importance weights, to parameter vectors sampled from the posterior distribution.

Algorithm 3 Importance Sampling

1: For $i = 1$ to N, draw $\theta_i \sim g(\theta)$ and compute the unnormalized importance weights $w_i = w(\theta_i) = \frac{f(\theta_i)}{g(\theta_i)}$. $f(\theta)$ is usually the product of the prior and likelihood densities while $g(\theta)$ is a proposal density

2: Compute the normalized importance weights:

$$W_i = \frac{w_i}{\frac{1}{N}\sum_{i=1}^{N} w_i}.$$

3: An approximation of $E_\pi[h(\theta)]$ is given by:

$$\bar{h}_N = \frac{1}{N}\sum_{i=1}^{N} W_i h(\theta_i).$$

One is thus left with the challenge of constructing a proposal density. SMC samplers allow one to construct this proposal density sequentially. Letting $\{\rho_n\}_{n=1}^{N_\phi}$ be some sequence of zeros and ones that determine whether particles are resampled in the selection step and let $\{\zeta_n\}_{n=1}^{N_\phi}$ be a sequence of tuning parameters for the Markov transition density in the mutation step.

Algorithm 4 Generic SMC with Likelihood Tempering

1: **Initialization.** ($\phi_0 = 0$). Draw the initial particles from the prior: $\theta_1^i \overset{\text{i.i.d.}}{\sim} p(\theta)$ and $W_1^i = 1$, $i = 1, \ldots, N$.

2: **Recursion.** For $n = 1, \ldots, N_\phi$,

- **Correction.** Reweight the particles from stage $n-1$ by defining the incremental weights

$$\tilde{w}_n^i = p(Y|\theta_{n-1}^i)^{\phi_n - \phi_{n-1}} \quad (5.4)$$

$$\tilde{W}_n^i = \frac{\tilde{w}_n^i W_{n-1}^i}{\frac{1}{N}\sum_{i=1}^N \tilde{w}_n^i W_{n-1}^i}, \quad i = 1, \ldots, N. \quad (5.5)$$

$$\tilde{h}_{n,N} = \frac{1}{N}\sum_{i=1}^N \tilde{W}_n^i h(\theta_{n-1}^i). \quad (5.6)$$

- **Selection.**
  - Case (i): If $\rho_n = 1$, resample the particles via multinomial resampling. Let $\{\widehat{\theta}\}_i^N$ denote N i.i.d. draws from a multinomial distribution characterized by support points and weights $\{\theta_{n-1}^i, \tilde{W}_n^i\}_i^N$ and set $W_n^i = 1$.
  - Case (ii): If $\rho_n = 0$, let $\widehat{\theta}_n^i = \theta_{n-1}^i$ and $W_n^i = \tilde{W}_n^i$, $i = 1, \ldots, N$. An approximation of $E_{\pi_n}(h(\theta))$ is given by

$$h_{n,N} = \frac{1}{N}\sum_{i=1}^N W_n^i h(\widehat{\theta}_n^i). \quad (5.7)$$

- **Mutation.** Propagate the particles $\{\widehat{\theta}^i, W_n^i\}$ via N_{MH} steps of a MH algorithm with transition density $\theta_n^i \sim K_n(\theta_n|\widehat{\theta}_n^i; \zeta_n)$ and stationary distribution $\pi_n(\theta)$.

$$\bar{h}_{n,N} = \frac{1}{N}\sum_{i=1}^N h(\theta_n^i)W_n^i. \quad (5.8)$$

3: For $n = N_\phi(\phi_{N_\phi} = 1)$, the final importance sampling approximation of $E_\pi(h(\theta))$ is given by:

$$\bar{h}_{N_\phi,N} = \sum_{i=1}^N h(\theta_{N_\phi}^i)W_{N_\phi}^i. \quad (5.9)$$

Herbst and Schorfheide (2016) Use this algorithm to estimate the same model under Rational Expectations which I estimate under Adaptive Learning. I choose this method for one additional reason and that is the parallelizability of the algorithm and therefore the ability to make use of high performance computing. The Metropolis Hastings Random Walk algorithm is a markov process which samples a sequence of correlated draws from the posterior distribution. Because of this, one cannot allocate different parts of the algorithm to different processing units on a CPU or GPU; one cannot reduce the time needed to complete an MHRW simulation by increasing the number of CPU or GPU cores devoted to the task. By contrast, because the SMC method draws multiple particles and evaluates the likelihood function for each independently of the other, the task of likelihood evaluation can be divided up among different processing units. Thus one can reduce the time needed to complete an SMC simulation by increasing the number of CPU cores used. The models estimated in this dissertation were obtained using an Amazon Web Services Elastic Cloud Compute instance using 96 physical CPU cores.

The methods described thus far are amenable to estimating models under rational expectations with linear dynamics and gaussian innovations and time-invariant transition dynamics. Introducing an adaptive learning behavioral rule can confound estimation, however, because the transition matrices evolve in a manner which is nonlinear. However, as shown in Hamilton, 1994, for the transition and measurement equations $y_t = A + Bs_t + Cv_t, s_t = T_c + T_1(t)s_{t-1} + T_0\varepsilon_t$, as long as the transition matrix $T_1(t)$ is a deterministic function of lagged values of y_t, which is indeed the case for the models studied in this dissertation.

I turn now to review just some papers which estimate such models that have agents who use adaptive learning algorithms to form expectations. It is this literature to which I contribute by estimating a prototypical DSGE model and evaluating systematically the impact of the choice of agents' initial beliefs.

Prior Estimated DSGE models with learning

Primiceri (2006) Provides a New Keynesian DSGE model with adaptive learning, estimated by maximum likelihood, to explain the curious rise and persistently high inflation during the 1960s. In this model, the learning agent is the monetary authority, while private agents in the model are a mix of rational agents and agents who use adaptive expectations. The monetary authority is assumed to have incomplete information about the economy including the persistence of inflation and the natural rate of unemployment, but knows the structure of the economy. One important implication the authors draw is that the persistently high inflation experienced by the US in the 1960s and 1970s is unlikely going forward as the model suggests that the great inflation resulted from the unlikely confluence of simultaneous mis-measurement of both the natural rate of unemployment and the persistence of inflation.

Milani (2007) Provides one of the earliest Bayesian estimations of a DSGE model with Adaptive Learning, and the model which I also estimate in the next chapter. A persistent problem in the DSGE literature had been the failure to match the apparent persistence of inflation and output in response to monetary policy shocks under rational expectations. Some modelers had incorporated various sources of mechanical persistence to address this. Some papers, notably including Fuhrer (2000), showed that habit persistence in household consumption preferences is required to reproduce characteristic "hump-shaped" impulse response functions that estimated reduced-form vector-autoregressions seem to produce. Further still, New Keynesian models under Rational Expectations require often exhibit significant improvement in model fit from adding such mechanical persistence, as measured by the estimated marginal data density. Milani estimates a five-equation New Keynesian model with mechanical persistence arising from inflation indexation and from consumption habit-persistence. Milani shows that such assumptions may, in fact, be superfluous if one instead allows for agents in the model to have imperfect information about the economy and to form expectations using models that are updated using constant-gain recursive least-squares. When estimated under learning, the model of Giannoni and Woodford (2004) does not support habit formation or inflation indexation.

Milani (2014) Extends this framework to a DSGE model wherein the gain parameter, that is the speed at which agents update their beliefs in response to new information, varies endogenously. The mechanism is according to which this gain changes is similar to one outlined in Marcet and Nicolini (2003), where gain switches between a constant and a geometrically decreasing sequence. Under such an updating scheme, the canonical New Keynesian model is able to generate, endogenously, time-varying volatility. The ability to recreate time-varying volatility is necessary for a DSGE model to adequately explain the "Great Moderation" from about 1985 to 2007. Milani's model contrasts with models such as that of Justiniano and Primiceri (2008) that generate time varying volatility by allowing the variance of i.i.d shocks to vary stochastically. The linear model estimated by Milani also avoids the additional computational burden of particle-filter based likelihood functions that is required for models that are not driven by gaussian shocks.

Estimated models with adaptive learning have not been limited to small-scale models, but also to the much richer medium-scale DSGE models such as that of Christiano et al. (2005) and Smets and Wouters (2003) that have been able to match the marginal data densities of reduced-form VAR models. Slobodyan and Wouters (2012) estimates a modified version of Smets and Wouters (2007) that omits the endogenous flexible-price equilibria. In the model estimated by Slobodyan and Wouters, agents form expectations using small (under-parameterized) forecasting models. Adaptive learning in the paper takes the form of a Kalman filter updating scheme. Letting $\beta_{t|t}$ be agents coefficients' and $P_{t|t}$ be the variance of those coefficients, beliefs are updated according to the following equations:

$$\beta_{t|t} = B_{t|t-1} + P_{t|t-1}X_{t-1}(\Sigma + X'_{t-1}P_{t|t-1}X_{t-1})^{-1} \times (y_t^f - X'_{t-1}\beta_{t|t-1}),$$
$$P_{t|t} = P_{t|t-1} - P_{t|t-1}X_{t-1}(\Sigma + X'_{t-1}P_{t|t-1}X_{t-1})^{-1} \times X'_{t-1}P_{t|t-1}.$$

In it, the authors estimate the model both under rational expectations and adaptive learning. The authors find some notable differences in important parameter estimates between the adaptive learning and the rational expectations specifications. The wage and price markup shocks in both RE and AL are assumed to follow ARMA(1,1) (autoregressive moving average) processes. The mean estimate for the AR(1) and MA(1) terms for the wage markup process are .96 and .88 respectively, while the AR(1) and MA(1) terms for the price markup are .85 and .7 respectively. Under Adaptive Learning, however, the mean estimates for the AR(1) and MA(1) terms for the wage process are .53 and .43 respectively, and .28 and .48 for the price markup. Further, except for the price markup MA(1) term, the 90% confidence intervals are non-overlapping. Not all sources of mechanical persistence disappear under the adaptive learning specification, however, as the parameters describing wage and price stickiness do not disappear, and that there is substantial overlap in the reported confidence bounds between rational expectations and adaptive learning.

In addition to reducing or eliminating some estimates of mechanical persistence, the time-varying nature of expectations allows the authors to reproduce macroeconomic time series with time-varying volatility even though model has homoskedastic shocks, implying that endogenous expectational changes explain the great inflation and subsequent great moderation in US macroeconomic data. This result contrasts sharply with Cogley et al. (2010), who identify an exogenous change in monetary policy as the main explanation for this change in volatility. This introduction of time-varying beliefs, able to generate time-varying volatility, is a major reason for the improved data-fit of many adaptive learning models over their rational expectations analogues.

As my primary subject of investigation is initial beliefs, I note also that the Slobodyan and Wouters include a discussion on the initialization of learning dynamics as a robustness check for their results. Under the baseline model, the initial beliefs are assumed to be derived from the variable coefficients implied by the rational expectations solution. One other way the authors initialize the agents beliefs is to keep the $\Sigma, V, \beta_{1|0}$ matrices constant and to estimate the rest of the model. In yet another experiment, the authors estimate the model wherein agents' beliefs do not change, and then use those estimated beliefs as the initial beliefs when estimating the model's structural parameters. Under such initializations, the authors find that the adaptive learning still improves greatly the estimated model likelihood compared to that same model estimated under rational expectations, and that this result is not sensitive to the researcher's choice of agents' initial beliefs.

A newer approach for improving economic forecasting is the DSGE-VAR approach of Negro and Schorfheide (2006). The authors propose an approach where the DSGE model is nested within a more flexible VAR framework, allowing for systematic relaxation of the cross-equation restrictions imposed by the DSGE model. This DSGE-VAR model serves as a tool to assess the empirical fit and forecasting performance of the DSGE model, comparing its restrictions against the data-driven flexibility of VARs. The authors apply their methodology to evaluate a version of the Smets and Wouters (2003) model, examining the extent to which the DSGE model restrictions need to be relaxed to optimize the fit of the DSGE-VAR model. They find that certain ad hoc modifications and the introduction of frictions, such as price stickiness and wage indexation, improve the empirical fit of the DSGE model but caution against overreliance on such mechanisms due to their potential non-invariance to policy experiments.

Cole and Milani (2019) apply such a framework to a small-scale New Keynesian model but within the Adaptive Learning framework rather than the Rational Expectations framework. The paper demonstrates within their model that while DSGE restrictions are valuable when the model does not have to match observed expectations, such as those that are gleaned from consumer or professional survey data, imposition of such expectations necessitates a departure from DSGE restrictions, indicating a misspecification in the model, and that the misspecification in the model lies in the choice of expectations formation mechanism. In exploring alternatives, the authors examine models incorporating extrapolative, heterogeneous expectations and find that these can somewhat reconcile the New Keynesian model with the observed expectations. The authors find that the DSGE-VAR approach, with adjusted prior restrictions informed by the DSGE model, offers a better fit to the data than unrestricted VAR models.

The Framework of Adaptive Learning

It will prove valuable to review the general framework of adaptive learning within the context of a DSGE model, as this will provide the reader a clear picture of what I intend to investigate as the "initial beliefs" of agents in the model. Exploring the framework of Adaptive Learning also provides readers another view at what it means to actually "solve" a DSGE model under Rational Expectations beyond my earlier description of the Schur decomposition technique. I turn now to a critically important part of the AL framework, the expectational stability or "E-stability" principle.

Let $T(\phi)$ be a function that maps from agents' subjective beliefs about economic dynamics to actual economic dynamics. In the models which I estimate, this T-map arises from substituting expectations formed through the adaptive learning algorithm into the difference equations implied by the DSGE model. The rational expectations solution to this T-map, then, is the fixed point, where the agents' beliefs about the economy match the true dynamics of the economy. For linear DSGE models, the fixed point of this T-map is considered "expectationally stable" if around this fixed point there exists a neighborhood of beliefs wherein the differential equation

$$\frac{d\phi}{d\tau} = T(\phi) - \phi$$

is asymptotically stable. This asymptotic stability is equivalent to having eigenvalues strictly inside the unit circle.

For the models I study in this and the next chapter, agents use linear forecasting models known as Vector Autoregressions that are updated using the algorithm of constant-gain recursive least squares. Supposing that agents forecast a vector of variables Z_t using regressors X_t so that their linear model is $Z_t = \phi_t' X_t$ The formulae for recursive least-squares are given below:

$$\phi_t = \phi_{t-1} + \gamma_t \Sigma_t^{-1} X_t' (Z_t - \phi_{t-1}' X_t)', \tag{1.3}$$

$$\Sigma_t = \Sigma_{t-1} + \gamma_t(X_t X_t' - \Sigma_{t-1}), \tag{1.4}$$

where Σ_t is $E(X_t X_t')$ and sometimes equivalent to the second moment matrix.

The sequence $\{\gamma_t\}_0^\infty$ is a weakly decreasing sequence of real numbers. For recursive least squares, which is asymptotically equivalent to ordinary least squares, this sequence is $1/t$. This gives all observations equal weight, so that the effect of each subsequent observation disappears as $t \to \infty$. In constant gain least squares, the scheme which I use in both chapters, γ_t is a constant, $\bar{\gamma}$ for all periods and is estimated along with the structural parameters. This parameter, $ga\bar{m}ma$ called the "learning gain" parameter, describes the weight that agents place on the most recent observations when updating their beliefs. Since this weight is non-decreasing, this causes agents to place more weight on more recent observations than observations in the more distant past. The motivation often given for this is that this allows agents to update their forecasting models more quickly in response to regime switches or parameter shifts in the underlying economy, updating that would not be permitted if the economy parameters change after an arbitrarily long time.

In the models which I estimate in the present chapters, agents are using only lagged endogenous variables to forecast the endogenous variables, so $X_t = (1, Z_{t-1})'$.

Throughout both chapters, Agents use "Euler Equation Learning," which is described in Evans and Honkapohja (2001). This setup uses the linear model equations from the DSGE model but substitutes into the DSGE model the expectations formed through linear forecasting models for the assumed rational expectations. Described more formally, suppose the economy evolves according to some DSGE model:

$$x_t = Ax_{t-1} + BE_t x_{t+1} + C\varepsilon_t$$

In such a setup, agents would arrive at period t with model coefficients ϕ_t and second moments Σ_t based on information set $X_t = \{x_i\}_0^{t-1}$. Then the i.i.d shocks ε_t arrive and x_t evolves according to

$$x_t = Ax_{t-1} + B\phi_t^2 X_t + C\varepsilon_t$$

At which point agents their model coefficients to ϕ_{t+1} and second moments Σ_{t+1} based on x_t

In some models, it is necessary to rely upon a "projection facility" to provide stable dynamics. A projection facility is a behavioral rule that forces agents' beliefs to lie within some compact set around the rational expectations equilibrium. Marcet and Sargent (1989) show that agents' beliefs formed and updated through a recursive least squares algorithm will, if the rational expectations equilibrium is expectationally stable and agents employ a suitable projection facility, converge with probability one to the beliefs implied by the rational expectations equilibrium.

This property would seem to imply that agents' initial beliefs should not matter to model dynamics, and therefore should not affect model estimates. However, this is only true asymptotically as $t \to \infty$. Since the models I estimate are estimated based on quarterly data, I can only simulate such models for a few hundred quarters at most, and within such a short time horizon initial beliefs can influence model dynamics. Carceles-Poveda and Giannitsarou (2007) examines recursive least squares, stochastic gradient learning, and other learning algorithms and documents the importance to explaining short-run variation of the right initialization choice. The scheme agents use in the models that I estimate is known as "Euler Equation Learning". In such a setup, model variables propagate according to the equations of the DSGE model, with the twist that the agents' expectations formed through the linear forecasting model are substituted for the expectational terms in the DSGE rather than the rational expectation. Such a setup can theoretically nest the rational expectations solution if agents have the true model coefficients and observe all contemporaneous shocks when forming such expectations. In such a setup, the expectations that agents form using their linear models are equivalent to the rational expectations.

A more To embed Adaptive Learning into the model, I use a scheme described in Evans and Honkapohja (2001) as "Euler Equation Learning." Such a scheme uses the same Euler equations derived from the Rational Expectations solution to the model, but instead substitutes the expectations formed through small forecasting models for the rational expectations. The adaptive learning algorithm I use in the present chapter is constant gain recursive least squares, whose formulae I recount below, which is one among several adaptive learning schemes in the literature including shadow-price learning, infinite-horizong learning, or kalman-filter learning.

As these coefficients and second-moment matrices are updated recursively, it is therefore ultimately a choice of the modeler themselves what to use as the initial beliefs ϕ_0, Σ_0 when simulating and estimating a DSGE model with adaptive learning. I explain now the three choices I use in both this and the next chapter

The first choice of initial beliefs is the rational expectations solution-implied beliefs. For each draw θ_i from the posterior distribution $p(\theta|Y)$ for which there there exists a unique RE solution, such a solution can be cast into a VAR in the model variables: $s_t = T_c + T_1 s_{t-1} + T_0 \varepsilon_t$. ϕ_0 then would be the coefficients from T_c, T_1, T_0, and which elements of those matrices are used would depend on the variables agents are assumed to use as regressors when forming forecasts of future variables. If agents are assumed only to observe three endogenous variables, then ϕ_0 would include only the first three entries of T_c and the upper-right 3×3 matrix of T_1. The initial second-moment matrix Σ_0 is obtained via a Schur Decomposition, just as I use to obtain the second moments of a DSGE model. One important implication of this initialization choice is that the initial beliefs are a strict function of the structural parameters θ_i. In the case of training sample beliefs, initial beliefs are fixed, while in the case of joint estimation, initial beliefs are allowed to vary across θ_i.

The second choice of initial beliefs uses a pre-sample time-series of data on endogenous variables to estimate a VAR model of the set of forecasted variables. I obtain point estimates using maximum likelihood, and once the VAR is estimated, I set ϕ_0 equal to the MLE coefficients and use a Schur decomposition to solve for Σ_0.

The final choice of initial beliefs treats each unique element of ϕ_0 and Σ_0 as a model parameter to be estimated along with the structural parameters of the model θ_i. This means that each element of ϕ_0 must be estimated, along with the upper-triangular elements of Σ_0 since $E(X_t X_t')$ is necessarily a symmetric matrix. This requires the researcher to specify a prior distribution for each of these elements. As I am comparing model performance according to the estimated marginal data density, I want to be sure that improvements in this estimate are due to improvements in the likelihood and not due to arbitrarily tight prior distributions. For this reason, I univariate normal priors with mean zero and standard deviations of 1 for elements of ϕ_0. For the upper triangular elements of Σ_0 I use these same priors. For the diagonal elements of Σ_0 I use a uniform prior bounded between 0 and 2. After several estimation attempts, this prior seemed to be the best prior that did not result in particle degeneracy due to too many particles having zero likelihood.

1.2 The Model

The model I estimate follows that of An and Schorfheide (2007). An important assumption of this model, which is relaxed in the next chapter, is that the sole sources of mechanical persistence come from the autoregressive shock processes and the autoregressive coefficient in monetary policy. Many other DSGE models, such as that of Giannoni and Woodford (2004), which I estimate in the next chapter under adaptive learning, allow for mechanical persistence in the inflation and output equations. I will recount now the model's microfoundations, as presented in a very compact manner in Herbst and Schorfheide (2016). The model consists of households, a monetary authority, a fiscal authority, intermediate-goods producing firms, and final-goods producing firms. I turn first to the dynamic optimization problem faced by price-setting firms in a monopolistically competitive setting.

Firms

The model assumes that production of goods takes place in two stages, first by monopolistically competitive firms that produce intermediate goods, and then by perfectly competitive firms producing a single final good for both household and government consumption. These price-taking final good producing firms combine a continuum of intermediate goods indexed by $j \in (0,)$ using the CES production technology:

$$Y_t = \left(\int_0^1 Y_t(j)^{1-\nu} dj \right)^{\frac{1}{1-\nu}}.$$

Final goods producers are perfectly competitive, and so the input prices $P_t(j)$ and the price of the final good P_t are given. Revenue from the sale of the final good is $P_t Y_t$ while costs are $\int_0^1 P_t(j) Y_t(j) dj$. The first order condition for maximizing profits, which are:

$$\Pi_t = P_t \left(\int_0^1 Y_t(j)^{1-\nu} dj \right)^{\frac{1}{1-\nu}} - \int_0^1 P_t(j) Y_t(j) dj,$$

implies that the demand for the j^{th} intermediate good is:

$$Y_t(j) = \left(\frac{P_t(j)}{P_t} \right)^{-\frac{1}{\nu}} Y_t.$$

The parameter $1/v$ thus captures the elasticity of substitution between each of the intermediate goods in the firm's production technology and can be interpreted as capturing the degree of monopoly power by firms in the economy, with a higher v meaning a lower degree of monopoly power and greater degree of competition between firms. The model assumes away entry costs, and that therefore producers enter the market until profits equal zero. This zero-profit condition implies the following relationship between intermediate goods prices and the price of the final good:

$$P_t = \left(\int_0^1 P_t(j)^{\nu} dj \right)^{\frac{1}{\nu}}.$$

The j^{th} intermediate good is produced by a monopolist using the following production technology

$$Y_t(j) = A_t N_t(j).$$

A_t is an exogenous productivity process that does not vary across firms while $N_t(j)$ is the labor input of the j^{th} firm. Labor is the only input to production and is hired at the perfectly competitive wage rate W_t.

Nominal price stickiness is incorporated into the model via quadratic price adjustment costs a la Rotemberg (1982) according to the following cost schedule:

$$AC_t(j) = \frac{\phi}{2} \left(\frac{P_t(j)}{P_{t-1}(j)} - \pi \right)^2 Y_t(j).$$

In this setup, ϕ determines the price rigidity in the economy while π determines steady state price inflation of the final good. In the intermediate goods market, the j^{th} firm chooses its labor input $N_t(j)$ and its price $P_t(j)$ to maximize the expected present discounted value of future profits:

$$\sum_{s=0}^{\infty} \left(\beta^s Q_{t+s|t} \left(\frac{P_{t+s}(j)}{P_{t+s}} Y_{t+s}(j) - W_{t+s} N_{t+s}(j) - AC_{t+s}(j) \right) \right),$$

where $Q_{t+s|t}$ is the time t value of a unit of the final consumption good in period $t+s$ to the household, which is treated as exogenous to the firm since the final good is sold in a perfectly competitive market.

Households

A representative household derives utility from consumption C_t relative to a habit stock, which is approximated by the level of technology A_t and real money balances, M_t/P_t

$$U = \sum_{s=0}^{\infty} \beta^s \left(\frac{(C_{t+s}/A_{t+s})^{1-\tau} - 1}{1-\tau} + \chi_M \ln \left(\frac{M_{t+s}}{P_{t+s}} \right) - \chi_H H_{t+s} \right).$$

In the above, β is the discount factor, $1/\tau$ is the intertemporal elasticity of substitution, and χ_M and χ_H are scale factors that determine the steady state money balances and hours worked respectively, the latter of which is set equal to one. The household sells labor to the firm at a competitive wage rate W_t. The household has access to a domestic bond market where nominal government bonds B_t are traded that pay interest rate R_t. The household additionally receives profits D_t from the firms and must pay taxes T_t. Thus the firm has the following budget constraint:

$$P_tC_t + B_t + M_t + T_t = P_tW_tH_t + (R_{t-1}B_{t-1}) + (M_{t-1} + P_tD_t + P_tSC_t),$$

where SC_t is the net cash inflow from trading a full set of Arrow Debreau securities.

Monetary and Fiscal Policy

Monetary policy is described by an interest rate feedback rule of the form:

$$R_t = R_t^{*1-\rho_R} R_{t-1}^{\rho_R} e^{\epsilon_{R,t}},$$

where $\epsilon_{R,t}$ is a monetary policy shock and R_t^* is the nominal target rate

$$R_t^* = r\pi^* \left(\frac{\pi_t}{\pi^*}\right)^{\phi_1} \left(\frac{Y_t}{Y_t^*}\right)^{\phi_2}, \qquad (1.5)$$

where r is the steady state real interest rate that is a function of the model's parameters, which I define below in 1.8, π_t is the gross inflation rate defined as $\pi_t = P_t/P_{t-1}$, and π^* is the target inflation rate. Y_t^* is the level of output that would prevail in the absence of nominal rigidities.

The fiscal authority consumes some fraction ζ_t of aggregate output Y_t so that $G_t = \zeta_t Y_t$, and this fraction follows an exogenous process. The government levies a lump sum tax T_t to finance short term falls in government revenues. Thus the government's budget constraint is as follows:

$$P_tG_t + R_{t-1}B_{t-1} + M_{t-1} = T_t + B_t + M_t.$$

Exogenous (shock) processes

The economy is driven by three exogenous processes. Aggregate productivity evolves according to the following:

$$\ln A_t = \ln \gamma \ln A_{t-1} + \ln z_t, \quad \ln z_t = \rho_z \ln z_{t-1} + \epsilon_{zt}.$$

Government spending evolves according to the following process:

$$\ln g_t = (1 - \rho_g) \ln \bar{g} + \rho_g \ln g_{t-1} + \epsilon_{gt},$$

where $g_t = 1/(1 - \zeta_t)$, where ζ_t was defined as the fraction of aggregate output that was purchased by the government. Finally the monetary policy shock, $\epsilon_{R,t}$ is assumed to be i.i.d. All three shocks are assumed to be i.i.d. and uncorrelated.

Equilibria

Intermediate goods producing firms are assumed to make identical decisions, so that j subscripts are omitted. Thus the market-clearing conditions are given by:

$$Y_t = C_t + G_t + AC_t, \quad H_t = N_t.$$

Households have access to the full set of Arrow Debreu securities, and thus

$$Q_{t+s|t} = \left(\frac{C_{t+s}}{C_t}\right)^{-\tau} \left(\frac{A_t}{A_{t+s}}\right)^{1-\tau}.$$

Thus in equilibrium, households and firms are using the same stochastic discount factor. It can also be shown that output, consumption, interest rates, and inflation must satisfy the following optimality conditions:

$$1 = \beta \mathbb{E}_t \left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t}\right)^{\tau} \frac{A_t}{A_{t+1}} \frac{R_t}{\pi_{t+1}}, \tag{1.6}$$

$$1 = \phi(\pi_t - \pi) \left(\left(1 - \frac{1}{2v}\right) \pi_t + \frac{\pi}{2v}\right), \tag{1.7}$$

$$- \phi\beta\mathbb{E}_t \left(\left(\frac{C_{t+1}/A_{t+1}}{C_t/A_t}\right)^{-\tau} \frac{Y_{t+1}A_{t+1}}{Y_t/A_t} (\pi_{t+1} - \pi)\pi_{t+1}\right),$$

$$+ \frac{1}{v}\left(1 - \left(\frac{C_t}{A_t}\right)^{\tau}\right).$$

Equation 1.6 is the consumption euler equation that reflects the FOC with respect to government bonds. Equation 1.7 characterizes the profit maximizing choice of intermediate goods producing firms. The FOC depends on the wage rate W_t. In the absence of nominal rigidities, aggregate output is given by:

$$Y_t^* = (1 - v)^{1/\tau} A_t g_t,$$

which is also the target level of output that appears in the monetary policy rule of equation 1.5.

The model is approximated in a linear form using a first-order Taylor Series approximation around the steady state. Output and consupmption are detrended by A_t so $c_t = C_t/A_t, y_t = Y_t/A_t, y_t^* = Y_t^*/A_t$. The steady states are given by:

$$\begin{aligned} \pi = \pi^*, r = \frac{\gamma}{\beta}, R = r\pi^*, \\ c = (1-v)^{1/\tau)}, y = gc = y^*. \end{aligned} \tag{1.8}$$

Letting $\widehat{x}_t = \ln(x_t/x)$ be the percentage deviation of any variable x_t from its own steady state value, we can write the the laws of motion for each variable in terms of percent deviations from their respective steady states:

$$\begin{aligned}
1 &= \beta \mathbb{E}_t \left(e^{-\tau \widehat{c}_{t+1} + \tau \widehat{c}_t + \widehat{R}_t - \widehat{z}_{t+1} - \widehat{\pi}_{t+1}} \right), \\
0 &= \left(e^{\widehat{\pi}_t} - 1 \right) \left(\left(1 - \frac{1}{2v} \right) e^{\widehat{\pi}_t} + \frac{1}{2v} \right), \\
&\quad - \beta \mathbb{E}_t (e^{\widehat{\pi}_{t+1}} - 1) e^{-\tau \widehat{c}_{t+1} + \tau \widehat{c}_t + \widehat{y}_{t+1} - \widehat{y}_t + \widehat{\pi}_{t+1}}, \\
&\quad + \frac{1-v}{v\phi\pi^2} (1 - e^{\tau \widehat{c}_t}), \\
e^{\widehat{c}_t - \widehat{y}_t} &= e^{-\widehat{g}_t} - \frac{\phi\pi^2 g}{2} (e^{\widehat{\pi}_t} - 1)^2, \\
\widehat{R}_t &= \rho_R \widehat{R}_{t-1} + (1-\rho_R)\phi_1 \widehat{\pi}_t, \\
&\quad + (1-\rho_R)\phi_2 (\widehat{y}_t - \widehat{g}_t) + \epsilon_{R,t}, \\
\widehat{g}_{t+1} &= \rho_g \widehat{g}_{t-1} + \epsilon_{g,t}, \\
\widehat{z}_{t+1} &= \rho_z \widehat{z}_{t-1} + \epsilon_{z,t},
\end{aligned}$$

Log linearizing the first three equations above yields the following system of equations that determine the path of output, inflation, and interest rates in the model studied:

$$\begin{aligned}
\widehat{y}_t &= E_t(\widehat{y}_{t+1}) - \frac{1}{\tau} \left(\widehat{R}_t - E_t(\widehat{\pi}_{t+1}) \right) + \widehat{g}_t - E_t(\widehat{g}_{t+1}), \\
\widehat{\pi}_t &= \beta E_t(\widehat{\pi}_{t+1}) + \kappa(\widehat{y}_t - \widehat{g}_t), \\
\widehat{R}_t &= \rho_R \widehat{R}_{t-1} + (1-\rho_R)\psi_1 \widehat{\pi}_t + (1-\rho_R)\psi_2(\widehat{y}_t - \widehat{g}_t) + \varepsilon_{Rt},
\end{aligned}$$

$$\kappa = \tau \frac{1-\nu}{\nu \pi^2 \phi}, \tag{1.9}$$

Under Euler Equation learning, agents are assumed to form forecasts using a VAR of the form $x_t = A + Bx_{t-1} + C\varepsilon_t$ so that $\mathbb{E}_{t-1}x_{t+1} = A + BA + B^2 x_{t-1}$. Matrices A, B are populated with elements from ϕ_{t-1} which is updated according to the CGL formulae of 1.3 and 1.4. By substituting directly the formulae for expectations formed through adaptive learning, I obtain the transition equation for the state space model. Note that I assume that agents only have knowledge of endogenous variables $(\widehat{y}_{t-1}, \widehat{\pi}_{t-1}, \widehat{R}_{t-1})'$ and not $\widehat{g}_{t-1}, \widehat{z}_{t-1}$ or i.i.d. shocks. Giving agents a forecasting model in which they use lagged exogenous shock processes renders impossible updating the second moment matrix, and so I limit attention in this chapter to models where agents observe only $(\widehat{y}_{t-1}, \widehat{\pi}_{t-1}, \widehat{R}_{t-1})'$ when forming expectations. I expound upon this problem later when I compare my three initialization choices.

The Data

I now explain the data I use and how the data relate to the underlying state variables. The measurement equations I use are as follows

$$\begin{aligned} YGR_t &= \gamma^Q + 100(\widehat{y}_t - \widehat{y}_{t-1} + \widehat{z}_t), \\ INFL_t &= \pi^A + 400\widehat{\pi}_t, \\ INT_t &= \pi^A + r^A + 4\gamma^Q + 400\widehat{R}_t, \end{aligned}$$

YGR_t is the observed quarter-over-quarter growth rate in per capita real GDP, $INFL_t$ is the observed quarter-over-quarter inflation rate, and $INFL_t$ is the observed annualized *nominal* interest rate.

The parameters γ^Q, π^A, r^A are related to the steady states of the model as follows:

$$\begin{aligned} \gamma &= 1 + \frac{\gamma^Q}{100}, \\ \beta &= \frac{1}{1 + r^A/400}, \\ \pi &= 1 + \frac{\pi^A}{400}, \end{aligned}$$

These three parameters are structural parameters to be estimated from the posterior distribution. As ν and ϕ are not separately identifiable, the model is expressed in terms of κ as defined in 1.9. For estimation of the structural parameters, I follow An and Schorfheide (2007) and use quarterly data from 1982:2002. To construct per capita real output growth, I use

$$\text{Growth} = 100\left(\ln\left(\frac{GDP_t}{POP_t}\right) - \ln\left(\frac{GDP_{t-1}}{POP_{t-1}}\right)\right),$$

For GDP, I use the FRED series "GDPC1" and for population I take the quarterly average of the civilian non-institutional population, FRED series CNP16OV/ (BLS series LNS10000000). To construct annualized inflation I use

$$\text{inflation} = 400\ln\left(\frac{CPI_t}{CPI_{t-1}}\right),$$

using the FRED series "CPIAUCSL" for the CPI_t. For the nominal interest rate, I just use the federal funds rate, FRED series "FEDFUNDS."

To construct the pre-sample data series for the training sample initial beliefs, I use the the same procedure for annualized inflation and the federal funds rate, but use the FRED series A939RX0Q048SBEA, which is FRED's series for Real Gross Domestic Product per capita. I use 50 quarters of pre-sample data to train agents' VAR. I limit myself to the relatively short 1982-2002 data series as the computational burden grows significantly as the time series is expanded. Adding pre-1982 data may be of dubious value as well since it pre-dates the great moderation while data post-2002 approaches the housing bubble and the great recession, and thus each choice increases the risk of an unidentified regime switch. With a full description of the measurement equations and the data series on which I estimate the model, I turn now to an exposition of the initial belief choices.

1.3 Initial Beliefs

Equilibrium Based Initials

The equilibrium based initial beliefs centers the coefficients ϕ_0 and elements of Σ_0 from 1.3 and 1.4 around the solution obtained from the rational expectations solution. To obtain this solution, I use the Schur decomposition method described earlier, which gives the general solution $x_t = T_c + T_1 x_{t-1} + T_0\varepsilon_t$ where $x_t = (\widehat{y}_t, \widehat{\pi}_t, \widehat{R}_t, \widehat{g}_t, \widehat{z}_t)'$. Agents are assumed to only observe $(\widehat{y}_{t-1}, \widehat{\pi}_t, \widehat{R}_{t-1})'$ when forming expectations $\mathbb{E}(\widehat{y}_{t+1}, \widehat{\pi}_{t+1})$.

Thus I substitute elements from T_1 into B from the formula for agents' individual expectations $E_{t-1}x_{t+1} = A + BA + B^2 x_{t-1}$.

To obtain the Σ_0 matrix, I use the Schur decomposition as described in Villemot (n.d.). This gives the second moment matrix for the state variables of a DSGE model. I use this because the model variables $(\widehat{y}_{t-1}, \widehat{\pi}_{t-1}, \widehat{R}_{t-1})'$ are zero mean processes, and so
$\mathbb{E}(\widehat{y}_{t-1}, \widehat{\pi}_{t-1}, \widehat{R}_{t-1})(\widehat{y}_{t-1}, \widehat{\pi}_{t-1}, \widehat{R}_{t-1})' = cov(\widehat{y}_{t-1}, \widehat{\pi}_{t-1}, \widehat{R}_{t-1})$. I then substitute the last three rows and columns of this numerically computed matrix into my Σ_0 matrix. As I am allowing agents to regress endogenous variables on a constant, the first column and row are zeros save for the top-left element, which is 1.

I omit estimation of models wherein agents use a forecasting model that regresses $(\widehat{y}_t, \widehat{\pi}_t, \widehat{R}_t)'$ upon $(\widehat{y}_{t-1}, \widehat{\pi}_{t-1}, \widehat{R}_{t-1}, g_{t-1}, z_{t-1})'$. This is because in the rational expectations solution, the endogenous variables and the shock processes are collinear and, as a result, the Σ_0 matrix is non-invertible. This can be confirmed by computing the determinant of the second-moment matrix from the rational expectations solution and verifying that it is very close, to within a rounding error, to zero. This renders it impossible to update coefficients via recursive least squares as the updating equation inverts this second moment matrix, as shown in 1.4.

I depart from Milani (2007) in omitting a constant from agents' forecasting models as well, so that expectations of future economic variables depend only upon a combination of previous economic variables. I impose this assumption for two reasons; first, one can argue that such a model is more consistent with the informational assumptions placed upon the agents. If agents are assumed to know the structure of the economy but not its parameters, then agents would know that the true value of the constant term in the Rational Expectations solution to the DSGE model does not include any constant terms. Recall that the regressors that agents use in their updating equations are the kalman filtered estimates of $\widehat{y}_{t-1}, \widehat{\pi}_{t-1}, \widehat{R}_{t-1}$, which are determined by the transition equations which do not have constant terms, thus the constant parameters γ^Q, π^A, r^A would not map onto agents' constant terms in their updating equations. Second, omitting a constant term reduces the computational burden of estimating the model. Future variations of this study that include constant terms in agents' forecasting models may, however, prove fruitful.

This problem does not present when estimating initial beliefs jointly with structural parameters or when initializing beliefs via a training sample because such models do not necessarily produce collinearity in the transition equation in the way that the

DSGE model solution does. As I did not want to "stack the deck" in favor of other models, I omit from my final comparison models that include richer information sets.

Training Sample Based Initial Beliefs

The next choice of initial beliefs I employ are those which are formed through a training sample. This strategy is employed, among others, by Milani (2014). In my case, I use a method similar to the method explored in Berardi and Galimberti (2017). The first step of choosing initial beliefs is to maximize the likelihood function implied by a state-space model of the following form:

- State equation:
$$x_{t+1} = \Phi x_t + \Psi w_t,$$
- Observation equation:
$$y_t = H x_t.$$

When estimating my reduced-form VAR, I used identity matrices for H and Ψ to simplify estimation. The state vector x_t, then, consists of the output process, inflation process, and the nominal interest rate. Because the Federal Reserve has a positive inflation target, the observed inflation and interest rates cannot have zero unconditional means, and so I used the de-meaned values for both of these processes while using the ordinary values for GDP growth. Estimation was then performed by maximizing a likelihood function, which was computed numerically via prediction-error variance decomposition through a Kalman Filter. This gives an estimate $\widehat{\Phi}_{MLE}$, which I use as my ϕ_0 matrix, and the variances w_t. With this information I can also compute the second moment matrix Σ_0 with a Schur Decomposition. These matrices ϕ_0, Σ_0 do not vary across the posterior distribution of structural parameters, in contrast to both equilibrium-based initial beliefs and jointly-estimated initial beliefs.

Jointly Estimated Initial Beliefs

Joint estimation treats initial beliefs as parameters to be estimated along with the rest of the structural parameters of the model. A procedure like this is performed in Milani (2009), but the impact of this choice on the marginal data density is not provided. Thus, for there to be a posterior distribution of the initial beliefs, I must define prior distributions of the initial beliefs. As I am investigating whether joint estimation can improve the performance of a DSGE model, I seek to allow the data to drive the results of the model and use fairly diffuse priors. For each element of ϕ_0, I use a standard normal distribution. A diffuse distribution with a large standard deviation samples many particles that lead to unstable dynamics and can lead to particle degeneracy, which causes the SMC algorithm to crash. However, as I discuss in the section on Data-Fit Comparison, the marginal data density is not particularly sensitive to the standard deviation of this prior.

A possible candidate for a prior distribution that may improve upon the results obtained in both this chapter and the next for the second moment matrix Σ_0 might be a Wishart prior distribution. A Wishart prior distribution first draws a $p \times n$ matrix where each n^{th} column is drawn from a p-variate normal distribution. Then the $p \times n$ matrix G premultiplies its own transpose to generate a positive semi-definite matrix GG^T.

For ϕ_0, One could use a prior that generates random matrices with eigenvalues uniformly distributed within the unit circle using a straightforward combination of random orthogonal matrices and uniform eigenvalues. This method would generate an orthogonal matrix of a given size n and generate a diagonal matrix of eigenvalues, and then combine the matrix to generate a PSD matrix whose eigenvalues lie within the unit circle. This would assure that all draws of initial beliefs satisfy the assumed stability property

Berardi and Galimberti (2017) include a discussion on joint estimation of initial beliefs. In that paper, Berardi and Galimberti generate data using a two-equation New Keynesian Phillips Curve model and then use GMM techniques to estimate model parameters along with the agents' initial beliefs. The authors find that the accuracy of the estimation of initial beliefs deteriorates as the sample size increases, and that biases in the estimates of initial beliefs can actually have spillover effects onto the estimates of other parameters. It is important to note, however, that my own study differs fundamentally in that I use Bayesian estimation of initial beliefs and structural parameters. Therefore, this dependence between the estimates of initial beliefs and structural parameters is modeled explicitly by the simulated posterior.

1.4 Results

Priors

For structural parameters, I follow the priors used by Herbst and Schorfheide (2016) as I am seeking to find the effect of learning upon the data fit of the model, *ceteris paribus*. For the learning gain, I used a uniform prior distributed between 0 and .05 to match common findings in the DSGE learning literature, as well as to avoid particle degeneracy in my SMC algorithm when jointly estimating initial beliefs. Larger gain values both in estimation and in simulation often resulted in unstable dynamics. Moving from a diffuse to narrow prior for the gain parameter had no impact on model likelihood for equilibrium-based or training-sample based initial beliefs. The priors distributions are reported in A.1 and are graphed along with marginal posterior distributions in B.1 through B.4. r^A and π^A, which are functions of the the steady state natural interest rate and inflation rate, are gamma distributions and bounded below from one. γ^Q is the steady state growth rate and is not bounded but is centered at a positive value, .4. τ is the risk aversion parameter and has a gamma prior distribution with a shape parameter of 2 and a scale parameter of .5. κ, the slope of the Phillips Curve, has a uniform distribution bounded between 0 and 1. ψ_1 and ψ_2 are the inflation and output feedback rules for monetary policy, respectively, while ρ_R is the feedback rule on lagged interest rates for monetary policy. ρ_g, ρ_z have uniform priors to assure stability, while the variance of the shock processes have inverse gamma distributions. I use a uniform prior from 0 to .05 for the learning gain to use as diffuse a prior as possible to avoid particle degeneracy and to assure that the covariance matrix of my SMC algorithm stays positive definite. However, a more diffuse distribution likely could have been used had I imposed a projection facility that restricted agents' beliefs to a stable set of beliefs.

Parameter Results

Estimates of structural parameters are quite robust to the choice of initial beliefs or even to the choice of expectations formation mechanism, as can be seen in tables C.2 through A.5. One notable difference in parameter estimates, however, is the slope of the Phillips Curve, κ, between the Rational Expectations baseline and the equilibrium-based initials. The mean estimate under the baseline is .85 while the mean estimate under equilibrium-based initial beliefs is .49. The 95% confidence intervals are also nearly non-overlapping.

Table 1.1. Estimates of Phillips Curve Slope, κ

	Mean	5% Interval	95% Interval
Rational Expectations	0.85	0.62	0.99
Equilibrium Inits	0.49	0.34	0.67
Training Sample Inits	0.57	0.41	0.78
Jointly Estimated Inits	0.56	0.33	0.87

Data-Fit Comparison

Within the wider context of Bayesian econometrics, model selection is often done using the marginal data density, which is the likelihood function integrated over the prior density function. For a given model M, data series Y, and parameter distribution θ, the marginal data density for model M is defined as

$$p(Y|M) = \int p(Y|\theta, M)p(\theta|M)d\theta.$$

This integral cannot be evaluated analytically for a DSGE model, and so must be approximated via monte carlo methods. This can be done from the output of a Metropolis Hastings algorithm using either the method of Geweke (1998) or Chib and Jeliazkov (2001). The SMC algorithm used to estimate the models in the present chapters, however, provide a straightforward approximation from the unnormalized particle weights that, unlike the MCMC methods, do not require additional computations of the likelihood function. Herbst and Schorfheide (2016) show that the monte carlo product of unnormalized particle weights $\prod_{n=1}^{N_\phi}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{w}_n^i W_{n-1}^i\right)$ converges to the marginal data density under suitable regularity conditions. Unlike the method of Chib and Jeliazkov or Geweke, this does not require additional computations of the likelihood function. This estimate also yields the Bayesian analog to the classical Likelihood Ratio test called the "Bayes Factor" which compares the ratio of the absolute likelihoods.

I seek to find the initialization scheme that can maximize the marginal data density so as to best fit the DSGE model parameters to macroeconomic data. To this end, I ran 5 SMC estimations for each of my three initialization schemes. The average and standard deviation of the estimated marginal data densities are reported in table

A.6. I provide 5 SMC runs for each model due to the cost constraints of estimation, but also to assure that one would not get wildly different parameter estimates from one SMC run to the next due to coding errors or particle degeneracy.

The marginal likelihoods show very promising improvements to in-sample data fit that can be gained from relaxing the assumption of rational expectations. Under the baseline model there is an average marginal likelihood of -336, but under the worst-performing learning model, that with equilibrium-based initial beliefs, the marginal likelihood averages -315. This implies a Bayes Factor of over 1.5 billion in favor of the learning model over the Rational Expectations model. The model with training sample based initial beliefs performs better still with an average estimated marginal data density of -311, and a Bayes Factor of 71 over the equilibrium-based initials. Finally, the model with jointly-estimated initial beliefs had an average estimated marginal likelihood of -309 for a Bayes Factor of 8.7 over the model with beliefs trained on pre-sample data, and a Bayes Factor of just over 1 trillion over the Rational Expectations baseline model. The performance of joint estimation offers considerable promise as it does not require the researcher to have access to pre-sample data. The joint estimation does not come without costs, however, as the marginal data density is less precisely estimated, with a higher standard deviation and, as revealed in the graphs of the marginal posterior densities, less precisely estimated structural parameters in some cases.

1.5 Summary and some Limitations

I have in this chapter estimated one New Keynesian DSGE model following An and Schorfheide (2007) and found striking results on the effect of initial beliefs upon in-sample data fit of the DSGE model studied. The measure I chose by which to evaluate this was the marginal data density, and on this measure the jointly-estimated initial beliefs performed the best. This might seem surprising at face value since asymptotic properties of adaptive learning models imply that initial beliefs should not matter to the dynamic properties of such models. However, as I am estimating and implicitly simulating models with only 80 quarters of data, these initial beliefs can affect the model dynamics considerably, and the differing ability of the models to fit the data show this.

I have throughout this chapter labored under a significant limitation, however, namely that my model has not integrated any of the numerous frictions such as consumption habit persistence, inflation indexation, or investment adjustment costs that DSGE modelers have integrated into the current generation of such models. The model of An and Schorfheide is ex hypothesi a purely forward looking model and, under Adaptive Learning, the only possible source of persistence in the inflation and output equations is through agents' forecasting processes. This limits the sort of features that one can integrate into the model, and as we saw earlier due to the collinearity in the DSGE solution, it also limits the kind of forecasting model agents can use for adaptive learning algorithms. In the next chapter, I study a New Keynesian model following Giannoni and Woodford (2004), which was estimated with agents that use Adaptive Learning by Milani (2007). I then evaluate systematically the initial beliefs and information sets and their effects on the data fit and parameter estimates of interest.

CHAPTER 2

BAYESIAN COMPARISON OF INITIAL BELIEFS IN A NEW KEYNESIAN PHILLIPS CURVE MODEL WITH MECHANICAL PERSISTENCE

2.1 Introduction and Literature Review

In this chapter, I estimate a small scale New Keynesian model with adaptive learning that has two additional sources of mechanical persistence so that inflation and output depend mechanically upon their lagged values, extending the seminal paper of Milani (2007). The primary subject of my inquiry is the impact of agents' initial beliefs upon parameter estimates and the upon ability of the models to fit macroeconomic data. The additional sources of persistence in the models I estimate presently consist of habit formation in household's consumption decisions and indexation in firms' pricing decisions. Early DSGE models had trouble fitting macroeconomic data because endogenous variables displayed far too much persistence relative to the i.i.d. fundamental shocks that were assumed to drive macroeconomic events. A cottage industry of sorts within the DSGE modeling literature arose that sought to find micro-founded sources of mechanical persistence to integrate into the DSGE framework. These additional sources of persistence improved the ability of DSGE models to fit the data by delaying the adjustment of model variables back to their "long run" values after experiencing exogenous shocks. I review some of the important proposed sources of mechanical persistence in the literature that have also contributed to the Adaptive Learning DSGE literature.

The seminal paper of Fuhrer (2000) showed the importance of embedding consumption habit persistence in household behavior in order to match DSGE models to the data. Earlier models assuming fully rational, neoclassical consumers treated consumption as "jump" variables that should follow a random walk that responds to new information regarding lifetime income, but aggregate data had displayed an apparent excessive smoothness for this model of consumer behavior. Fuhrer proposes then to modify the consumers' utility function in order to provide a fully micro-founded explanation for this persistence. The proposed utility function is of the form:

$$U_t = \frac{1}{1-\sigma}\left(\frac{C_t}{C_{t-1}^{\eta}}\right)^{1-\sigma}.$$

Fuhrer derives a household consumption function from this utility specification and compares shows that it can produce the hump-shaped response in consumption to income shocks, the very same impulse response displayed by estimated VAR models.

The model of Giannoni and Woodford (2004) includes a utility function with this sort of habit persistence for households, along with inflation indexation by firms. In the model, prices are adjusted according to the Calvo mechanism, where some fraction $0 < \alpha < 1$ of firms are able to set prices to maximize expected discounted profits. The rest of the firms choose prices according to the rule

$$\log p_t(i) = \log p_{t-1}(i) + \gamma \pi_{t-1},$$

where $0 \leq \gamma \leq 1$ measures the degree of indexation to the most recently available inflation measure. Under rational expectations, both my and Milani (2007) find very high estimated values of this parameter. One substantial difference between my own and Milani's results, however, is that Milani finds that the estimated value of both the inflation indexation parameter and the habit persistence parameter fall to nearly zero when the model is estimated under adaptive learning, while my estimates do so only under some choices of information set and initial beliefs, none of which were present in the original paper.

A much richer model developed by Christiano et al. (2005) incorporates the standard New Keynesian features of sticky prices and sticky wages, along with several other frictions. The model includes the habit persistence in household consumption as in Fuhrer, variable capital utilization, investment-specific costs, and a rule that businesses must borrow capital to finance their wage costs. The authors begin by estimating a reduced form VAR(4) based on data from 1965 to 1995. The variables included in the vector autoregression include real gross domestic product, real consumption, the GDP deflator, real investment, the real wage, labor productivity, the federal funds target rate, real profits , and the growth rate of the M2 component of the money supply. The ordering of the variables just listed contains the identifying assumption that real gross domestic product, real consumption, the GDP deflator, real investment, the real wage, labor productivity do not respond contemporaneously to a monetary policy shock while the federal funds target rate, real profits , and the growth rate of the M2 component of the money supply all do respond contemporaneously to such a shock. Impulse response functions show that an expansionary monetary policy shock increases output and increases inflation, and that both have hump-shaped responses to monetary shocks. So it would be very desirable of any DSGE model to display such a hump-shaped response of output and inflation to an expansionary monetary policy shock.

The model parameters are chosen via a combination of methods including calibration to match long-run steady state values, impulse response matching to the estimated VAR model, and GMM estimation. The authors find that sticky prices play only a limited role in explaining the persistence of output and inflation to monetary shocks. This finding is reached by showing that other parameter estimates and impulse responses differ little when estimating the model under the assumption of perfectly flexible prices. Inflation indexation, similarly, is found to play little role in explaining the hump-shaped response to policy shocks. Wage stickiness, variable capital utilization and investment adjustment costs, on the other hand, are estimated to play a crucial role in explaining aggregate fluctuations in the model, by the same criteria.

Since the study I undertake is ultimately an exercise in Bayesian econometrics, and since I aim to introduce the reader to important extensions to the DSGE framework, no such literature review would be complete without a review of the Smets and Wouters (2007) model. Smets and Wouters provide medium-scale DSGE model of the US macroeconomy that is estimated using Bayesian methods. Importantly, the authors provide the first such model that is able to match the in-sample data-fit, as measured by the marginal data density, of that provided by vector autoregression models.

The model is based on a time series of seven variables: real GDP, hours worked, consumption, investment, real wages, prices, and the short-term nominal interest rate. The frictions of interest incorporated into the model include sticky nominal price and wage, backward inflation indexation, habit formation in consumption and investment adjustment costs, variable capital utilization and fixed costs in production. The model dynamics are driven by seven i.i.d. shocks including a total factor productivity shock, a risk premium shock, investment-specific technology shock, wage and price markup shocks, and fiscal and monetary policy shocks.

Bayesian likelihood-based methods provide a natural framework for testing the empirical performance of each nominal friction. The authors do so by restricting each of the nominal frictions equal to zero, estimating the model, and then comparing the posterior mode of the rest of the parameters, along with the estimated marginal data density. A large change in the posterior mode or a statistically significant decrease in the marginal data density provides a straightforward measure of the empirical significance of each friction. With these evaluative criteria, the authors find, contra

Christiano et al. (2005), approximate parity between price and wage stickiness in empirical importance for fitting the model to the data. Inflation indexation is found to be relatively unimportant, and investment adjustment costs are found to be the most important.

The models surveyed so far date back prior to the great recession of 2008. A criticism of then-existing DSGE models that arose in the literature was the failure of such models to forecast the great recession, even after financial crisis had reached its tipping point, namely the collapse of the investment bank Lehman Brothers. This failure of DSGE forecasts to match the path of the macroeconomy is detailed in Wolters and Wieland (2012). A promising line of research has sought to integrate another source of persistence and volatility into DSGE models in the form of financial frictions. The model of Del Negro et al. (2015) combines the model of Smets and Wouters (2007) with a financial accelerator model of B. Bernanke et al. (1996). It will be helpful to first review the financial accelerator mechanism proposed by Bernanke, Gertler, and Gilchrist. The financial accelerator model is a model of credit markets with asymmetric information. In contrast to the result of Modigliani and Miller (1958) wherein the structure of a firm's financing will not affect its economic activity, in credit markets with asymmetric information, firms with low net equity face higher external borrowing costs. This implies that if a firm's net worth falls, its need for external borrowing rises simultaneously with higher external borrowing costs. These simultaneous events will cause a reduction in the firm's output. The authors formalize this by showing the firm's collateral-in-advance constraint in the following way.

Suppose the firm produces output in two periods, 0 and 1, and the firm buys variable input x_1 and uses fixed input K in period 0 to sell in period 1 at a price of q_1 according to production technology $a_1 f(x_1)$. The firm begins with cash flow from the previous period $a_0 f(x_0$ and debt burden from the previous period $r_0 b_0$. The firm's budget constraint on the purchase of the variable input x_1 is as follows:

$$x_1 = a_0 f(x_0) + b_1 - r_0 b_0.$$

Supposing then that unsecured lending is not possible so that all debt is fully secured, the borrowing constraint then is:

$$b_1 \leq (q_1/r_1)K,$$

which implies:

$$x_1 \leq a_0 f(x_0) + (q_1/r_1)K - r_0 b_0.$$

These equations show very straightforwardly that changes in the firm's net worth, by changing the value of K, will very directly affect the firm's output x_1. This highly simplified model underlies the intuition behind how a change in asset prices can have real effects; much like changes in wealth levels have real effects on household consumption behavior, credit constraints can can cause changes in firms net worth to affect their real behavior in the same manner. The authors integrate this financial accelerator model into a larger dynamic general equilibrium model of the economy in B. S. Bernanke et al. (1999). Del Negro, Giannoni, and Schorfheide integrate financial accelerator mechanism into the DSGE model of Smets and Wouters (2007) and, estimating parameters up until Fall of 2008, show that the model is able to predict the sharp contraction in output following the period of severe financial distress immediately following the collapse of Lehman Brothers.

At a micro-level, financial frictions are added to the model of Smets Wouters by assuming a pooling equilibrium between more risky and less risky firms in the market for credit. Banks take funds from households and lend them to firms and charge a spread over the riskless rate of return. This spread is affected by firms' net worth, just as in the Bernanke, Gertler and Gilchrist model of credit constrained firms.

The model is estimated based on nine time series, namely the original seven used to estimate the Smets and Wouters model along with 10-year inflation expectations, which are obtained from the Blue Chip Economic Indicators survey and the Survey of Professional Forecasters, and a time series measuring bond yield spreads between riskless and risky assets, for which the authors use the Baa Corporate Bond Yield spread over the 10-Year Treasury Note Yield at constant maturity. The authors show that the model, from Fall 2008 to 2013, is able to predict the path of GDP remarkably well compared to other DSGE models. Further, the DSGE model does not, contra most other DSGE models of the time, predict a sharp disinflation. This is due to how the authors chose to model the Zero Lower Bound condition, using the solution technique of Cagliarini and Kulish (2013) that allows one to solve linear rational expectations models with anticipated structural changes. The model's predictions fared significantly worse for marginal costs and interest rates, however, which the authors argue was due to adverse shocks and significantly accomodative monetary policy.

While surveying a small cross section of the varieties of mechanical persistence that have been used in DSGE modeling, I have limited my attention to those which retain the Rational Expectations Hypothesis, as very few such models, much less estimated models, relax this assumption as I do in the present study. One of the earliest and most cited examples of an estimated DSGE model with bounded rationality appears in Milani (2007), who estimates a small-scale New Keynesian DSGE model with both habit formation and inflation indexation. This DSGE model consists of five equations, plus two definitions:

$$\begin{aligned}
\tilde{x}_t &= E_t\tilde{x}_{t+1} - (1-\beta\eta)\sigma\left(i_t - E_t\pi_{t+1} - r_t^n\right), \\
\widehat{\pi}_t &= \xi_p\left(\omega x_t + ((1-\eta\beta)\sigma)^{-1}\widehat{x}_t\right) + \beta E_t\widehat{\pi}_{t+1} + u_t, \\
i_t &= \rho i_{t-1} + (1-\rho)(\phi_\pi\pi_t + \phi_x x_t) + \varepsilon_{i,t}, \\
r_t^n &= \phi_r r_{t-1}^n + v_t^{r^n}, \\
u_t &= \phi_u u_{t-1} + v_t^u, \\
\widehat{\pi}_t &\equiv \pi_t - \gamma\pi_{t-1}, \\
\tilde{x}_t &\equiv (x_t - \eta x_{t-1}) - \beta\eta\left(E_t(x_{t+1} - \eta x_t)\right),
\end{aligned}$$

where x_t is the output gap, π_t is the inflation rate, i_t is the nominal interest rate, r_t^n is the natural interest rate, and u_t is a cost-push shock process. It is this model which I also estimate in the present chapter and investigate agents initial beliefs and information sets, and their effects on the model's in-sample forecasting performance and impulse response functions. In contrast to the Rational Expectations Hypothesis, agents in the model use simple linear models in order to form expectations, which are updated according to an adaptive learning algorithm, as described in Evans and Honkapohja (2001), called constant gain recursive least squares, which I describe later when detailing my own estimation procedure. When estimating the model under Rational Expectations, Milani finds that both habit persistence and inflation indexation are important features in explaining features of business cycle data. However, when relaxing rational expectations, these sources of mechanical persistence become much less important and their estimated values, measured in their posterior mean values, fall significantly.

A similar result is obtained in Slobodyan and Wouters (2012), who estimate a version of the model Smets and Wouters (2007) but in which agents form expectations using small (under-parameterized) forecasting models. Adaptive learning in the paper takes the form of a Kalman filter updating scheme. In it, Slobodyan and Wouters estimate the model under both rational expectations and adaptive learning. Wage and price markup shocks in both RE and AL are assumed to follow ARMA(1,1) (autoregressive moving average) processes. The posterior mean estimate for the AR(1) and MA(1) terms for the wage markup process are .96 and .88, while the AR(1) and MA(1) terms for the price markup are .85 and .7. Under the Adaptive Learning specification the posterior mean estimate for the AR(1) and MA(1) terms for the wage process are .53 and .43 respectively, and .28 and .48 for the price markup. Further, save for the price markup MA(1) term, the 90% confidence intervals are non-overlapping. Estimates for the degree of price and wage stickiness remain statistically significant, however.

The authors also include a discussion on the initialization of learning dynamics as a robustness check. Under the baseline Adaptive learning model, initial beliefs are derived as a function of the structural parameters from the rational expectations solution. One other way the authors initialize the agents beliefs is to keep the $\Sigma, V, \beta_{1|0}$ matrices constant across the sampling of parameters and to estimate the rest of the structural parameters. In another specification, the authors estimate the model wherein agents' beliefs do not change, and then use those estimated beliefs as the initial beliefs when estimating the model again. Under such a specification, the authors find that an adaptive learning rule still improves appreciably the marginal likelihood compared to that same model estimated under the rational expectations hypothesis. The authors do not find that this particular result is sensitive to the researchers' choice of initial beliefs.

2.2 The Model

The model estimated here follows the basic New Keynesian model of Woodford (2003), which was estimated under Adaptive Learning by Milani (2007). The critically important difference between this model and the one estimated in the previous chapter is the presence of additional rigidities in the IS equation and the Phillips Curve equation. I recount the microfoundations of the model, in a manner similar to that of Milani (2007), to elucidate how these rigidities are incorporated into the model.

Households

The economy is populated by a continuum of households uniformly distributed on the $[0,1]$ interval who seek to maximize a sum of expected, discounted future utilities that take the form:

$$E_t\left(\sum_{T=t}^{\infty}\beta^{T-t}\left(U(C_T-\eta C_{T-1};\zeta_T)+v(h_T^i(j);\zeta_T)\,dj\right)\right).$$

In the above, $\beta\in(0,1)$ represents the household discount factor, C_T^i is an index of the household's consumption of each of the differentiated goods supplied in period t, $h_T^i(j)$ is the amount of labor supplied for the production of good j while ζ_T is a vector of exogenous aggregate preference shocks. The parameter $\eta\in(0,1)$ is the degree of habit formation and the source of mechanical persistence in the household consumption problem. $U(\cdot)$ is increasing and concave in both ζ and deviations of current consumption C_T from a stock of consumption C_{T-1}. E_t is the expectations operator, which can denote either rational expectations or those formed through adaptive learning. In setting up the microfoundations I follow Woodford (2003) and assume rational expectations but when estimating the moodel I substitute those expectations formed through adaptive learning. The consumption index is of the common Dixit-Stiglitz constant elasticity of substitution form:

$$C_t^i\equiv\left(\int_0^1 c_t^i(j)^{\frac{\theta-1}{\theta}}\,dj\right)^{\frac{\theta}{\theta-1}},$$

and the associated price index is of the form:

$$P_t\equiv\left(\int_0^1 p_t(j)^{1-\theta}\,dj\right)^{\frac{1}{1-\theta}},$$

where $\theta>1$ is the elasticity of substitution between differentiated goods and can be thought to represent the degree of monopoly power that differentiated firms have. Household i's optimal consumption of good j is:

$$c_t^i(j)=C_t^i(p_t(j)/P_t)^{-\theta},$$

where $p_t(j)$ is the price of differentiated good j at time t. To simplify computation of equilibria, financial markets are assumed to be complete and fiscal policy is assumed to be Ricardian.

Thus with habit formation, the first-order condition for consumption implies:

$$\lambda_t = U_c(C_T - \eta C_{T-1}; \zeta_T) - \beta E_t U_c(C_{T+1} - \eta C_T; \zeta_{T+1}). \tag{2.1}$$

Importantly, the marginal utility of additional inflation-adjusted income in period t, equal to the period t lagrange multiplier λ_t, is no longer equal to the marginal utility of consumption in that period. However the marginal utility of income still satisfies the following equality:

$$\lambda_t = \beta E_t \left(\lambda_{t+1} \left(1 + i_t\right) \frac{P_t}{P_{t+1}} \right), \tag{2.2}$$

where i_t denotes riskless one-period nominal interest rates. Here one can take a log linear approximation of the household's Euler equation and substitute 2.1 and 2.2 and derive:

$$\tilde{C}_t = E_t \tilde{C}_{t+1} - (1 - \beta\eta)\sigma \left(\widehat{i}_t - E_t \widehat{\pi}_{t+1} - \right) + g_t - E_t g_{t+1},$$

where $\tilde{C}_t$ is defined thusly:

$$\tilde{C}_t = \widehat{C}_t - \widehat{C}_{t-1} - \beta \left(E_t \widehat{C}_{t+1} - \widehat{C}_t \right),$$

and where the elasticity of intertemporal substitution of consumption is defined as $\sigma \equiv \frac{-U_c}{\bar{C} U_{CC}} > 0$. Exogenous preference shocks are given by $g_t \equiv \frac{\sigma U_{C\zeta} \zeta_t}{U_c}$ and the circumflex operator $\widehat{}$ on C_t, i_t, π_t denotes log-deviations of those variables from their steady state values. The model assumes away investment so that all output, in the aggregate, is consumed, so that $C_t = Y_t$, and the output gap is $x_t \equiv Y_t - Y_t^n$ where Y_t^n is the natural rate of output, or the rate of output sans sticky prices. This definition yields the following:

$$\tilde{x}_t = E_t \tilde{x}_{t+1} - (1 - \beta\eta)\sigma \left(i_t - E_t \pi_{t+1} - r_t^n \right),$$

where $\tilde{x}_t$ is defined thusly:

$$\tilde{x}_t \equiv (x_t - \eta x_{t-1}) - \beta\eta \left(E_t (x_{t+1} - \eta x_t) \right),$$

and the natural interest rate, or the flexible-price interest rate, is
$r_T^n \equiv ((1 - \eta\beta)\sigma)^{-1}((Y_{t+1}^n - g_{t+1}) - (Y_t^n - g_t))$.

Firms

The model assumes a continuum of monopolistically competitive firms with sticky prices in the manner of Calvo (1983). A fraction of firms, $\alpha \in (0,1)$ are allowed to change their prices optimally in a given period while the prices of the remaining firms are also adjusted but indexed to previous prices according to the following rule:

$$\log p_t(i) = \log p_{t-1}(i) + \gamma \pi_{t-1},$$

where γ measures the degree of inflation indexation. This particular rule was proposed by Christiano et al. (2005), but inflation indexation has been used in a variety of DSGE models.

Each monopolistically competitive firm i supplies its good using to the production technology $y_t(i) = A_t f(h_t(i))$ where A_t is an exogenous technology process, $h_t(i)$ is the labor input, and the production function f is increasing and concave. The capital stock is fixed across t so the only variable input to production is labor. Each i firm faces the same demand curve so $y_t(i) = Y_t \left(\frac{p_t(i)}{P_t}\right)$ for their differentiated product, where $Y_t = \left(\int_0^1 y_t(i)^{\frac{\theta-1}{\theta}di}\right)^{\frac{\theta}{\theta-1}}$ is the aggregate output and P_t is the aggregate price index, which the firm takes as a given. Both the price indices and output indices use the Dixit-Stiglitz CES aggregator because they allow for monopolistic competition without requiring individual firms to consider how other firms will react to their own pricing decisions, avoiding the curse of dimensionality. Since all firms face identical decision problems, they would set a common price p_t^* sans staggered price adjustment. From this it follows that the aggregate price index evolves according to the following law of motion:

$$P_t = \left(\alpha\left(P_{t-1}\left(\frac{P_{t-1}}{P_{t-2}}\right)^{\gamma}\right)^{1-\theta} + (1-\alpha)p_t^{*1-\theta}\right)^{\frac{1}{1-\theta}}.$$

What is left is to compute p_t^*. Firms set this price to maximize the expected discounted future sum of of profits:

$$\Pi_t(p) = E_t \sum_{T=t}^{\infty} \alpha^{T-t} Q_{t,T}\left(\Pi_T\left(p_t^*(i)\left(\frac{P_{T-1}}{P_{t-1}}\right)^{\gamma}\right)\right).$$

where $Q_{t,T} = \beta^{T-t}\frac{P_t}{P_T}\frac{\lambda_T}{\lambda_t}$ is the stochastic discount factor while $\Pi_T(\cdot)$ denotes the nominal profits in period-T, which are revenues minus wage costs:

$$\Pi_t(p) = p_t^*(i)\left(\frac{P_{T-1}}{P_{t-1}}\right)^{\gamma} Y_T \left(\frac{p_t^*(i)\left(\frac{P_{T-1}}{P_{t-1}}\right)^{\gamma}}{P_T}\right)^{-\theta} - w_t(i) f^{-1}\left(\frac{Y_T}{A_T}\left(\frac{p_t^*(i)\left(\frac{P_{T-1}}{P_{t-1}}\right)^{\gamma}}{P_T}\right)^{-\theta}\right).$$

Profits are discounted at the rate α since the optimal price chosen at date t can be expected to apply in period T with probability α^{T-t} at discount factor $Q_{t,T}$.

One can log-linearize this first order condition around the steady state solution to yield the following:

$$\widehat{p}_t(i) = E_t \sum_{T=t}^{\infty} (\alpha\beta)^{T-t} \left(\frac{1-\alpha\beta}{1+\omega\theta} \left(\omega \widehat{Y}_T - \widehat{\lambda}_T + \frac{v_{y\zeta}}{v_y}\zeta_T \right) + \alpha\beta(\widehat{\pi}_{T+1} - \gamma\widehat{\pi}_T) \right).$$

where $\widehat{p^*}_t \equiv \log(p_t^*/P_t)$ and $\omega \equiv v_{yy}\bar{Y}/v_y$ is the elasticity of marginal *dis*utility of producing output with respect to an increase in output. Log linearizing the law of motion for the aggregate price index gives $\widehat{p^*}_t = \frac{\alpha}{1-\alpha}(\widehat{\pi}_t - \gamma\widehat{\pi}_{t-1})$ which can be plugged into the previous expression for $\widehat{p^*_t}_t$ to yield the law of motion for inflation:

$$\widehat{\pi}_t = \xi_p \left(\omega x_t + ((1-\eta\beta)\sigma)^{-1}\widehat{x}_t \right) + \beta E_t \widehat{\pi}_{t+1} + u_t,$$

where

$$\widehat{\pi}_t \equiv \pi_t - \gamma\pi_{t-1}, \tag{2.3}$$
$$\widehat{x}_t \equiv (x_t - \eta x_{t-1}) - \beta\eta E(x_{t+1} - x_t),$$

$$\xi_p = \frac{(1-\alpha)(1-\alpha\beta)}{\alpha(1+\omega\theta)},$$

and where $u_t \equiv \frac{v_{y\zeta}}{v_y \xi_p}\zeta_t$ represents an aggregate supply shock. Here, x_t represents the deviation of the sticky price equilibrium output from the flexible price equilibrium output.

Closing the model, I assume that monetary policy follows a Taylor Rule

$$i_t = \rho i_{t-1} + (1-\rho)(\phi_\pi \pi_t + \phi_x x_t) + \varepsilon_{i,t},$$

where ρ, ϕ_π, ϕ_x are structural parameters to be estimated. For the purpose of simplifying estimation, I assume that these coefficients do not vary over time. For the natural interest rate and cost-push processes r_t^n and u_t respectively, I assume each follow univariate, $AR(1)$ processes:

$$
\begin{aligned}
r_t^n &= \phi_r r_{t-1}^n + v_t^{r_n}, & v_t^{r_n} &\sim N(0, \sigma_{r_n}), \\
u_t &= \phi_u u_{t-1} + v_t^u, & v_t^u &\sim N(0, \sigma_u).
\end{aligned}
$$

I am thus left with five linear equations and two definitions, which I will summarize here:

$$
\begin{aligned}
\tilde{x}_t &= E_t\tilde{x}_{t+1} - (1-\beta\eta)\sigma\left(i_t - E_t\pi_{t+1} - r_t^n\right), && (2.4) \\
\widehat{\pi}_t &= \xi_p\left(\omega x_t + ((1-\eta\beta)\sigma)^{-1}\widehat{x}_t\right) + \beta E_t\widehat{\pi}_{t+1} + u_t, && (2.5) \\
i_t &= \rho i_{t-1} + (1-\rho)(\phi_\pi \pi_t + \phi_x x_t) + \varepsilon_{i,t}, \\
r_t^n &= \phi_r r_{t-1}^n + v_t^{r_n}, \\
u_t &= \phi_u u_{t-1} + v_t^u, \\
\widehat{\pi}_t &\equiv \pi_t - \gamma\pi_{t-1}, \\
\tilde{x}_t &\equiv (x_t - \eta x_{t-1}) - \beta\eta\left(E_t(x_{t+1} - \eta x_t)\right).
\end{aligned}
$$

How Agents Learn to Forecast

To embed Adaptive Learning into the model, I use a scheme described in Evans and Honkapohja (2001) as "Euler Equation Learning." Such a scheme uses the same Euler equations derived from the Rational Expectations solution to the model, but instead substitutes the expectations formed through small forecasting models for the rational expectations. The adaptive learning algorithm I use in the present chapter is constant gain recursive least squares, whose formulae I recount below, which is one among several adaptive learning schemes in the literature including shadow-price learning, infinite-horizong learning, or kalman-filter learning.

Suppose agents regress a vector of variables y_t on X_t. In the cases studied in the present chapter, y_t is a 3×1 vector of time t endogenous variables while X_t can be a 4×1, 6×1 or 9×1 vector in the case of the minimal, limited, and full information sets respectively. Suppose further that agents have the linear model $y_t = \phi' X_t$. Let $\bar{g} > 0$ be some small constant. With these elements, the formulae for Constant Gain Recursive Least Squares are as follows:

$$\phi_t = \phi_{t-1} + \bar{g}\Sigma_t^{-1}X_t'(y_t - \phi_{t-1}'X_t)',$$
$$\Sigma_t = \Sigma_{t-1} + \bar{g}(X_tX_t' - \Sigma_{t-1}),$$

where Σ_t is $E(X_tX_t')$ referred to variously as the "second-moment matrix." As expectations depend mechanically upon previous values of endogenous variables, this embeds an additional source of persistence into the model, thereby fitting the model better to the data. A more natural choice might be to use decreasing-gain least squares rather than constant-gain least-squares. Decreasing-gain least squares learning, which is asymptotically equivalent to ordinary least squares, chooses $\gamma_t = t^{-1}$. Constant-gain least squares, by contrast, chooses a constant scalar for $\gamma_t = \bar{g}$. This learning structure places larger weight on more recent observations, and thus allows beliefs to adapt more quickly in the face of structural change. Further, as Branch and Evans (2007) note, since the volatility of endogenous variables differs, agents behaving optimally will use different values for each endogenous variable. For the sake of reducing computational burden, I omit this feature from my estimated models and assume that agents' gain parameter does not vary across the three endogenous variables. However, the model of Milani (2014) does feature varying gain parameters for different forecasted variables.

I depart from Milani (2007) in omitting a constant from agents' forecasting models as well, so that expectations of future economic variables depend only upon a combination of previous economic variables. I impose this assumption for two reasons; first, one can argue that such a model is more consistent with the informational assumptions placed upon the agents. If agents are assumed to know the structure of the economy but not its parameters, then agents would know that the true value of the constant term in the Rational Expectations solution to the DSGE model does not include any constant terms. Second, omitting a constant term reduces the computational burden of estimating the model. Future variations of this study that include constant terms in agents' forecasting models may, however, prove fruitful.

Timing of Expectations Formation

In standard estimations of the rational expectations models, expectations of time $t+1$ endogenous variables are formed in time t, that is, such expectations are realized simultaneously with the model's endogenous variables. One should note that in the models estimated in this chapter, expectations of time $t+1$ endogenous variables are formed at time $t-1$. That is to say, at time t agents enter the period with expectations formed during $t-1$ of endogenous variables to be realized at time $t+1$. These expectations then interact with the exogenously determined variables, u_t, g_t, to determine the time t endogenous variables.

The researcher further has a choice in deciding the timing of monetary policy, and how monetary policy relates to agents' expectations. Bullard and Mitra (2002) evaluate several such rules including *contemporaneous data* specifications, in which the monetary authority uses contemporaneous realizations of endogenous variables, *lagged data* specifications, in which time $t-1$ data is used to determine the time t interest rate target, *forward looking* specifications and finally *current expectations* based rules. The authors find that forward looking rules produce determinate rational expectations equilibria that, importantly, agents are able to learn through standard adaptive learning algorithms and that lagged data specifications often do not produce learnable, determinate equilibria.

In the model I estimate, agents form expectations of x_t, π_t based on current beliefs and up-to-date information on state variables and possibly information on contemporaneous shocks. Let s_t be the 5×1 vector of state variables, in which case agents PLM is

$$s_t = a + \Phi s_{t-1} + \Psi \varepsilon_t,$$

where s_t is an augmented state vector containing endogenous and exogenous variables while ε_t is a vector of i.i.d shocks with variance-covariance matrix Σ_ε Expectations of these variables at time $t, t+1, t+2$ can be computed by iterating forward this PLM thusly:

$$\begin{aligned} E_t s_t &= a + s_{t-1} + \Psi \varepsilon_t, \\ E_t s_{t+1} &= a + \Phi(a + \Phi s_{t-1} + \Psi \varepsilon_t) = a + \Phi a + \Phi^2 s_{t-1} + \Phi \Psi \varepsilon_t, \\ E_t s_{t+2} &= a + \Phi(a + \Phi a + \Phi^2 s_{t-1} + \Phi \Psi \varepsilon_t) = a + \Phi a + \Phi^2 a + \Phi^3 s_{t-1} + \Phi^2 \Psi \varepsilon_t. \end{aligned}$$

This provides a very direct way of solving for the VAR(1) form of the system. Recall the original form of the DSGE model: $s_t = P + Q s^e_{t+1} + R s_{t-1} + S \varepsilon_t$. From the above system describing the expectations, one can substitute for s^e_{t+1} for the matrix function $a + \Phi a + \Phi^2 s_{t-1} + \Phi \Psi \varepsilon_t$ to yield

$$s_t = P + Q(a + \Phi a + \Phi^2 s_{t-1} + \Phi\Psi\varepsilon_t) + Rs_{t-1} + S\varepsilon_t$$
$$= P + Qa + Q\Phi a + (Q\Phi^2 + R)s_{t-1} + (R + \Phi\Psi)\varepsilon_t,$$

which is itself the transition equation of the state space model whose likelihood function I will evaluate using a Kalman Filter. It will thus be useful to define clearly the transition and measurement equations for my state-space model.

Let $s_t = (x_t, \pi_t, i_t, r_t^n, u_t)'$ be the partially-observed state variables. These are the output gap, inflation rate, federal funds rate, natural interest rate, and a productivity shock process. The first three are observed directly while the last two are assumed to be observed by agents in the model. The observable vector of variables, $y_t = (x_t, \widehat{\pi}_t, \widehat{i}_t)'$, contains the output gap, taken from the Federal Reserve Bank of St. Louis, defined by the data series 100*(Real Gross Domestic Product-Real Potential Gross Domestic Product)/Real Potential Gross Domestic Product. The inflation rate is defined as the annualized log-difference in the Consumer Price Index for all urban consumers, or "CPIAUCSL", and the i_t is the annualized effective federal funds rate, defined by the FRED series "FEDFUNDS". Thus $\widehat{\pi}_t, \widehat{i}_t$ are divided by four to yield the state variables π_t, i_t.

The timing of expectations formation is important to define when simulating and estimating the adaptive learning algorithm. At time-t, agents arrive with their beliefs $\phi_t = (a_t, B_t, C_t)$ and they observe s_{t-1} and possibly ε_t. They then form expectations $E_t x_{t+n}$ for $n = 0, 1, 2$. After these expectations are formed, the endogenous variables arise according to the DSGE model. Once they observe the new state variables s_t, they update their beliefs according to constant-gain least-squares to ϕ_{t+1}, R_{t+1} and the process repeats. This implies that the transition matrix at time$-t$ is determined only by agents' beliefs. When computing the Kalman Filter, agents are assumed to observe the Kalman-filtered states.

In the previous chapter I limited agents to observing only lagged endogenous variables. This was due to collinearity in the DSGE solution for $(x_t, \pi_t, i_t, r_t^n, u_t)'$, resulting in a non-invertible second moment matrix for the DSGE model. One important consequence of including mechanical lags in the IS and Phillips Curve equations of the model, however, is the disappearance of this collinearity. When solving the model for $\eta, \gamma > 0$, the second moment matrix becomes invertible, but when forcing $\eta = \gamma = 0$, or for values very close to zero, the determinant of this second moment matrix becomes zero, and thus the matrix is non-invertible.

2.3 Results

Priors on Parameters

As I am attempting to show the impact of initial beliefs by themselves, *ceteris paribus*, on forecasting performance, I seek to match the common conventions in the DSGE literature when choosing prior distributions for my parameters. I use inverse gamma distributions for the variances of shock processes, partly to bound them from below zero. Each of my prior distributions for my three i.i.d shocks has a mean of one and a standard deviation of .5.

For the inflation indexation value, I used a uniform prior on zero to one. I use a tightly bound beta distribution for the discount rate, centered at .99 with a standard deviation of .01. For the elasticity of substitution of consumption I used a gamma prior with a mean of .125 and a standard deviation of .09. For the habit persistence parameter I used a uniform parameter from 0 to 1. For the feedback rule on inflation in the monetary authority's Taylor rule, I used a normal distribution centered at 1.5 with a standard deviation of .25. This was to assure that few draws fell outside the region of determinacy, as a feedback rule on inflation that is less than one often leads to indeterminacy. For the Taylor Rule feedback parameter on output, I used a normal distribution with a mean of .5 and a standard deviation of .25. The prior for the autoregressive parameter in the natural interest rate shock process is a uniform prior distributed from 0 to .97, as is the prior for the autoregressive productivity process. Finally for the gain parameter I used a beta distribution with a mean of .031 and a standard deviation of .022.

When jointly estimating initial beliefs, I do not estimate each element of the R_0 matrix, as this greatly increases the number of estimated parameters, and therefore can lead to inconsistent SMC estimates of model parameters and of the marginal data density. Instead, I assume that agents begin life with a simple VAR model of the following form:

$$\begin{bmatrix} x_t \\ \pi_t \\ i_t \\ r_t^n \\ u_t \end{bmatrix} = a + \Phi \begin{bmatrix} x_{t-1} \\ \pi_{t-1} \\ i_{t-1} \\ r_{t-1}^n \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_i & 0 & 0 \\ 0 & \sigma_{rn} & 0 \\ 0 & 0 & \sigma_u \end{bmatrix}\right).$$

The above describes additional restrictions that I use when estimating agents' initial beliefs, namely that agents are assumed to know that the economy is driven by three i.i.d shocks that drive monetary policy and two latent autoregressive processes, and that the autoregressive shock processes are affected directly only through their individual i.i.d. shocks. I estimate elements of the Φ matrix and elements $c_{11}, ..., c_{33}$ using normal distributions that I obtained using from the monte carlo posterior from the model estimated under Rational Expectations. The mean of this multivariate normal was obtained by stacking the elements of the $\Phi_{RE}(\theta_i)$ from the posterior $p(\theta_i|Y)$ from the estimated model under rational expectations, and its covariance matrix was the monte carlo covariance, with the diagonal elements set to the maximum of the monte carlo variance and one. For elements of $c_{11}, ..., c_{33}$, I used unit standard normal distributions for each element.

The Data

The model equations to be solved are given in 2.5 which yield the transition equation for the model. The measurement equation for the model to be estimated is a straightforward

$$y_t = \underset{3\times 3}{I} s_t,$$

where y_t is the data series available to the researcher while s_t is the latent vector containing the output gap process, the inflation process, and the nominal interest rate process.

For the output gap process, I used the log difference between the real gross domestic product per capita, as provided by FRED data series GDPC1, and real potential gross domestic product per capita as estimated by the Congressional Budget Office, as provided by FRED series GDPPOT. For the inflation process, I first constructed an inflation series by taking the log difference of the consumer price index, provided by FRED series CPIAUCSL, and then using the de-meaned values of that time series. For the nominal interest rate process, I used the de-meaned values of the federal funds rate, provided by FRED series FEDFUNDS. I use de-meaned values for these series because the Federal Reserve maintains a positive inflation target, and this is incompatible with a zero value for steady state observed inflation or nominal interest rate. As a useful approximation, I assume that the Federal Reserve has achieved its goal of keeping average inflation and interest rates around its long-term target. The data on which I estimate my models span 1961 to 2005. I omit data near the great recession to avoid risk of regime switches confounding my estimates. I use data from 1955 to 1961 to form a training sample on which to train agents' initial beliefs in models wherein agents use a reduced-form VAR to form initial beliefs.

Important Parameter Estimates

I report parameter estimates from the models with equilibrium-based initial beliefs in C.3 and C.4. Important to note are the estimates of the degree of habit persistence in household consumption, η, and the degree of inflation indexation by price-setting firms, γ. The choice of information set consistently affects the mean estimate of each parameter, however for both parameters the 95% confidence intervals overlap. Finally, the in-sample forecasting performance of each model, as measured by the estimated marginal data density, is somewhat higher for the model wherein agents use the full information set available to them. The high estimated values for inflation indexation and habit persistence would appear to contradict the findings of Milani (2007) who finds that habit formation and inflation indexation drop to nearly zero when agents are assumed to form expectations using VAR and MSV learning rules. My estimates of inflation indexation and habit persistence, however, do seem to comport with Cole and Milani (2019) who estimate the model of Giannoni and Woodford (2004) under several expectations formations mechanisms in addition to Rational Expectations, and find little change in η, γ. The significant difference between Milani (2007) uses a single-chained metropolis-hastings random walk estimate, while Cole and Milani (2019) reports estimates from a DSGE-VAR model.

Table 2.1. Estimates of Habit Persistence Parameter, η

	Mean	5% Interval	95% Interval
Rational Expectations	0.46	0.28	0.66
Equilibrium Initials, Full Info	0.51	0.33	0.73
Equilibrium Initials, Limited Info	0.79	0.65	0.89
Training Sample Initials, Full Info	0.70	0.46	0.79
Training Sample Initials, Limited Info	0.93	0.86	0.99
Jointly Estimated Initials, Full Info	0.78	0.70	0.86
Jointly Estimated Initials, Limited Info	0.77	0.64	0.98

Table 2.2. Estimates of Inflation Indexation Parameter, γ

	Mean	5% Interval	95% Interval
Rational Expectations	0.91	0.81	0.99
Equilibrium Initials, Full Info	0.91	0.82	0.99
Equilibrium Initials, Limited Info	0.81	0.59	0.97
Training Sample Initials, Full Info	0.80	0.61	0.97
Training Sample Initials, Limited Info	0.94	0.53	1.00
Jointly Estimated Initials, Full Info	0.48	0.33	0.66
Jointly Estimated Initials, Limited Info	0.72	0.44	0.98

I report estimated marginal data densities in C.14. The model with full information performs roughly as well if not somewhat better than the rational expectations baseline, while the model with limited information can perform better than the rational expectations baseline when there is available pre-sample data on which to train agents' initial beliefs. The model estimated under a minimal information set, where agents use only the lagged endogenous variables to forecast inflation and output, shows the most consistent improvement over the rational expectations baseline, outperforming in 3 of 4 initialization choices.

One important difference between the model with lags and the model without lags is the impact of learning upon the estimated slope of the Phillips Curve. The model of An and Schorfheide, 2007 has a very steep Phillips Curve, with a slightly flatter Phillips Curve under the model estimated with Adaptive Learning. This finding is not preserved in the model with lags, as the estimated Phillips Curve is almost perfectly flat.

The slope of the Phillips Curve in this model is not captured in a single parameter like the previous model, however. It is instead a function of multiple deep parameters, namely $\frac{\xi_p}{\sigma(1-\beta\eta)}$. From looking at the prior distribution in C.1, it might initially appear that this is due to a dogmatically tight prior restriction on the calvo parameter ξ_p, which has a standard deviation of only .011. However, as I have the slope of the Phillips curve as a function of the model parameters, I can construct the mean and standard deviation of the prior distribution via monte carlo methods. Drawing one million particles, I find the prior mean value of this slope to be roughly 1.3 and the prior standard deviation of this slope to be roughly 12. One can construct the marginal posterior density of the slope of the Phillips Curve from the monte carlo simulated joint posterior of the parameters θ fairly easily. I report the means and standard deviations of this slope for each initial belief choice and information set in table C.13. The only specification for which there is a significant change is the jointly estimated initial beliefs with the "limited" information set in C.13, which has an average estimated slope of 0.31. All other specifications have average estimated slope values of 0.11 or less.

Impulse Response Functions

A curiosity arises from differing initial beliefs in the form of differing impulse response functions, namely the recurrence of the "price puzzle" first observed in Sims (1986). It is observed widely in the time series literature that small VAR models with output, inflation, and nominal interest rates frequently generate impulse response functions with inflationary responses to contractionary monetary policy, an effect that runs entirely contrary to accepted monetary and business wisdom. Of note for the present study is that the price puzzle remains even after the additional restrictions imposed by the DSGE models, as the agents' initial beliefs work only through DSGE model equations. While causally identified models are certainly the most useful tools for policy analysis, agents in the DSGE model are assumed to use such models only for the purpose of forecasting future variables. Thus one would need some micro-founded motivation for agents to use causally identified models rather than optimal forecasting models when training initial beliefs on pre-sample data. One speculative reason might be to suppose that agents in the model seek forecasting models that are invariant to regime shifts in monetary policy which would confound predictions of the theretofore optimal forecasting model. Thus one may reasonably assume when training initial beliefs that agents use a minimal set of restrictions on their VAR, such as those offered by Estrella (2015). I do so for the model and then re-estimate the DSGE structural parameters with training-sample based initial beliefs, but instead of using an unrestricted VAR, I estimate a VAR of the following form:

$$\begin{aligned}\pi_t &= a_1\pi_{t-1} + a_2 x_{t-1} + \varepsilon_t,\\ x_t &= b_1 x_{t-1} + b_2(i_{t-1} - \pi_t) + v_t,\\ i_t &= c_1 i_{t-1} + c_2\pi_t + c_3 x_t + \mu_t,\end{aligned}$$

with the restrictions $a_1 > 0, a_2 > 0, b_1 > 0, b_2 < 0, c_i > 0$,

Letting $X_t = (\pi_t, x_t, i_t)'$ the model may be written as

$$
\begin{aligned}
A_0 X_t &= A_1 X_{t-1} + e_{t-1}, \\
X_t &= A_0^{-1} A_1 X_{t-1} + A_0^{-1} e_{t-1}, \\
X_t &= B X_{t-1} + A_0^{-1} e_{t-1},
\end{aligned} \tag{2.6}
$$

where A^{-1} is lower triangular and B has the restriction that $(1,3)$ element, the coefficient of inflation on lagged interest rates, is zero. Imposing these restrictions on the VAR used to train agents' initial beliefs gives impulse response functions that show a deflationary response to a negative monetary policy shock. It would stand to reason, then, that a model which has more "reasonable" impulse response functions would also fit the data better, and this prior belief is in fact borne out by the estimated marginal data density of a DSGE model whose agents' initial beliefs are based on a VAR with such restrictions. This model has a Bayes Factor of almost 20 above the Rational Expectations baseline and a similar Bayes factor above any of the equilibrium-based initialization schemes. Since this particular VAR assumes that three i.i.d. shocks that drive the three endogenous variables, it is not compatible with any of the other information sets and, for this reason, this exercise is done only with the minimal information set. I plot the impulse response function for the DSGE model with this initialization scheme in figure D.11

Data-Fit and Model Comparison

The other goal of this chapter is to find the method of initializing beliefs that best fits a DSGE model to the data. The framework of Bayesian econometrics offers a natural measure of this in the form of the marginal data density. To recap, the marginal data density is an average of the model's likelihood function over the space of possible parameter values, weighted by the prior density function. It provides the likelihood of observing the sequence of data $\{y_t\}_{t=0}^{t=T}$ given the model M_i. The ratio of the marginal data densities between two models M_i/M_j, called the Bayes Factor, is a Bayesian analog to the classical Likelihood Ratio test.

The marginal data density is an integral, namely $p(Y|M) = \int p(Y|\theta, M)p(\theta|M)d\theta$. This integral cannot be computed analytically for any model estimated presently, so I estimated it numerically using the output from my SMC algorithm. For each model estimated, I drew 5,000 particles with 300 stages, except for the jointly estimated initial beliefs. For those, I drew 10,000 particles, as models with more estimated parameters need a greater number of particles to obtain consistent estimates. I report the means and standard deviations of the log marginal data densities in C.14. While the highest average estimate belongs to the model with training sample based initial beliefs, this model also has the greatest variance, owing to one run that had an unrealistically high marginal likelihood of -145. Discarding that run would have resulted in an average estimated marginal data density of -896. The next highest marginal likelihood belonged to the model with jointly estimated initial beliefs and limited information. While the high standard deviation of estimated marginal likelihoods ought to worry practitioners, none of the SMC runs under jointly estimated initial beliefs had lower marginal likelihoods than any of the SMC runs under Rational Expectations. The Bayes Factor for the Jointly Estimated, Limited Information model over the Rational Expectations baseline is almost 15,000.

When jointly estimating agents' initial beliefs, I sought priors that were as diffuse as possible while allowing for the SMC algorithm to remain tractable without sampling too many particles from regions that lead to unstable learning dynamics, which then lead to particle degeneracy. Diffuse priors are desirable when ranking models as a diffuse prior will penalize a model with very low average likelihood by sampling over a large space with low likelihood, while models that explain data well will tend to have likelihood functions that drive the shape of the posterior and minimize the effect of the prior distribution.

Sensitivity of MDD to Priors

When jointly estimating agents' initial beliefs, the researcher must specify a prior distribution for those beliefs. For the results reported in table C.7 through C.11 I used, for the elements of the transition matrix, a multivariate normal distribution with an identity covariance matrix multiplied by five to generate a diffuse prior that did not generate too many transition matrices that caused non-invertible second-moment matrices or have eigenvalues with greater than unity in magnitude, as this would cause particle degeneracy in the SMC sampler. As can be seen in C.12, estimates for structural parameters are not particularly sensitive to the choice of prior for initial beliefs as the 95% confidence intervals for all parameters, including the learning gain, overlap between the two posteriors. On the other hand, the estimated marginal data density does exhibit some sensitivity to the choice of prior for the estimated initial beliefs, as using a narrow prior results in an average log likelihood of -826.0820 with a standard deviation of 2.6463 across five SMC runs. One likely reason for this outcome relates to the stability properties of adaptive learning algorithms, as initial beliefs that produce unstable dynamics or non-invertible second moment matrices are severely penalized by the likelihood function, and a diffuse prior samples over a large space that contains such unstable beliefs. A multivariate normal distribution centered around 0^N with a small standard deviation will sample most beliefs from inside the unit circle. One cannot, however, arbitrarily increase the marginal data density by narrowing the prior density of the initial beliefs. When re-estimating the model with a multivariate normal whose covariance matrix is an identity matrix scaled by 0.5, I obtain an average likelihood of -827.8112 with a standard deviation of 1.6225. This indicates that the marginal likelihood is not being arbitrarily raised by increasing the prior density.

2.4 Summary and Conclusions

In this dissertation I sought to investigate the importance of the choice of agents initial beliefs in a DSGE model with adaptive learning. In the first chapter I examined a purely forward-looking model without any mechanical persistence in the inflation or output equations. In that chapter I found that the choice of initial belief set had notable implications for the fitting of the model to the data and had important effects on parameter estimates that have exciting policy implications. I found that training initial beliefs on pre-sample data resulted in the best-fitting model, with a much higher marginal data density than any specification estimated.

In the second chapter I relaxed the rather stringent assumption of no mechanical persistence in the inflation or output processes and estimated a model with habit persistence and inflation indexation. The importance of agents' initial beliefs was studied in three main areas of interest, namely the estimates of structural parameters, the models' predicted response to monetary policy changes, and finally the ability of the models to match macroeconomic data series.

Both chapters showed consistent improvements in the ability of models with learning to match economic data over their rational expectations counterparts, and both chapters showed that it was indeed possible to improve further still upon this data-fit by an appropriate choice of initial beliefs. In both chapters, training beliefs based on pre-sample data provided the most reliable method of improving the fit of DSGE models to the data. This, however, is not always available to the DSGE model builder. In the absence of available or reliable pre-sample data, the model builder may wish to use joint estimation of initial beliefs, as this did seem to, under reasonable informational assumptions, improve the fit of DSGE models to the data over the equilibrium-based initial beliefs.

APPENDIX A

TABLES, CHAPTER 1

Table A.1. Prior Distribution for Structural Parameters

Name	Domain	Prior Density	Para (1)	Para (2)
Steady State Related Parameters θ_{ss}				
r^A	$\mathbb{R}+$	Gamma	0.50	0.50
π^A	$\mathbb{R}+$	Gamma	7.00	2.00
γ^Q	$\mathbb{R}$	Normal	0.40	0.20
Endogenous Propagation Parameters θ_{endog}				
τ	$\mathbb{R}+$	Gamma	2.00	0.50
κ	$[0, 1]$	Uniform	0.00	1.00
ψ_1	$\mathbb{R}+$	Gamma	1.5	0.25
ψ_2	$\mathbb{R}+$	Gamma	0.50	0.25
ρ_R	$[0, 1)$	Uniform	0.00	1.00
Exogenous Shock Parameters θ_{exog}				
ρ_G	$[0, 1)$	Uniform	0.00	1.00
ρ_Z	$[0, 1)$	Uniform	0.00	1.00
$100\sigma_R$	$\mathbb{R}+$	InvGamma	0.40	4.00
$100\sigma_g$	$\mathbb{R}+$	InvGamma	1.00	4.00
$100\sigma_z$	$\mathbb{R}+$	InvGamma	0.50	4.00
Adaptive Learning Parameters θ_{AL}				
gain	$[0, .0.05]$	Uniform	0.00	0.05
$\hat{\phi}_0$ elements	$[-10, 10]$	Uniform	-10.00	10.00
$\hat{\Sigma}_0$ diagonal elements	$[0, 5]$	Uniform	0.00	5.00
$\hat{\Sigma}_0$ off- diagonal elements	$[-10, 10]$	Uniform	-10.00	10.00

Table A.2. SMC Estimates, 2000 particles with 100 stages, Rational Expectations, 5 runs

Parameter	Mean	5% Interval	95% Interval
τ	2.27	1.57	3.19
κ	0.85	0.62	0.99
ψ_1	1.85	1.51	2.20
ψ_2	0.59	0.22	1.04
rA	0.46	0.05	1.05
π_A	3.40	2.76	4.02
γ_Q	0.59	0.37	0.79
ρ_R	0.76	0.70	0.82
ρ_g	0.98	0.95	1.00
ρ_z	0.92	0.88	0.96
σ_R	0.22	0.18	0.26
σ_g	0.65	0.57	0.75
σ_z	0.20	0.16	0.24

Table A.3. SMC Estimates, 2000 particles with 100 stages, Equilibrium Initials, minimal Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
τ	2.41	1.74	3.10
κ	0.49	0.34	0.67
ψ_1	1.47	1.13	1.87
ψ_2	0.33	0.15	0.58
rA	0.64	0.05	1.55
π_A	3.66	2.91	4.52
γ_Q	0.53	0.30	0.77
ρ_R	0.83	0.78	0.87
ρ_g	0.93	0.87	0.98
ρ_z	0.87	0.80	0.94
σ_R	0.18	0.15	0.20
σ_g	0.80	0.59	1.06
σ_z	0.38	0.32	0.43
gain	0.0189	0.0029	0.0393

Table A.4. SMC Estimates, 2000 particles with 100 stages, Training Sample Initials, minimal Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
τ	2.04	1.41	2.89
κ	0.57	0.41	0.78
ψ_1	1.45	1.12	1.86
ψ_2	0.55	0.20	1.04
rA	0.24	0.01	0.71
π_A	5.19	4.00	6.21
γ_Q	0.50	0.29	0.67
ρ_R	0.95	0.91	0.98
ρ_g	0.82	0.72	0.92
ρ_z	0.74	0.58	0.87
σ_R	0.18	0.15	0.20
σ_g	0.73	0.56	0.92
σ_z	0.39	0.32	0.47
gain	0.0297	0.0031	0.0559

Table A.5. SMC Estimates, 2000 particles with 100 stages, Jointly Estimated Initials, minimal Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
τ	1.95	1.41	2.54
κ	0.56	0.33	0.87
ψ_1	1.33	1.07	1.64
ψ_2	0.38	0.18	0.62
rA	0.42	0.04	1.00
π_A	4.54	3.69	5.44
γ_Q	0.54	0.39	0.69
ρ_R	0.91	0.86	0.96
ρ_g	0.83	0.64	0.95
ρ_z	0.72	0.55	0.85
σ_R	0.18	0.15	0.21
σ_g	0.73	0.51	1.03
σ_z	0.38	0.32	0.44
gain	0.0267	0.0031	0.0535

Table A.6. Mean and Standard Deviation of Natural Logarithms of the Marginal Likelihoods, 1982-2002 data

	Full Information	Limited Information
Rational Expectations	-336.3391 (1.3285)	N/A
Equilibrium Initials	0.0000 (0.0000)	-315.1337 (0.5108)
Training Sample Initials	0.0000 (0.0000)	-310.8610 (0.2914)
Jointly Estimated Initials	0.0000 (0.0000)	-308.6947 (1.5541)

APPENDIX B

CHAPTER 1 FIGURES

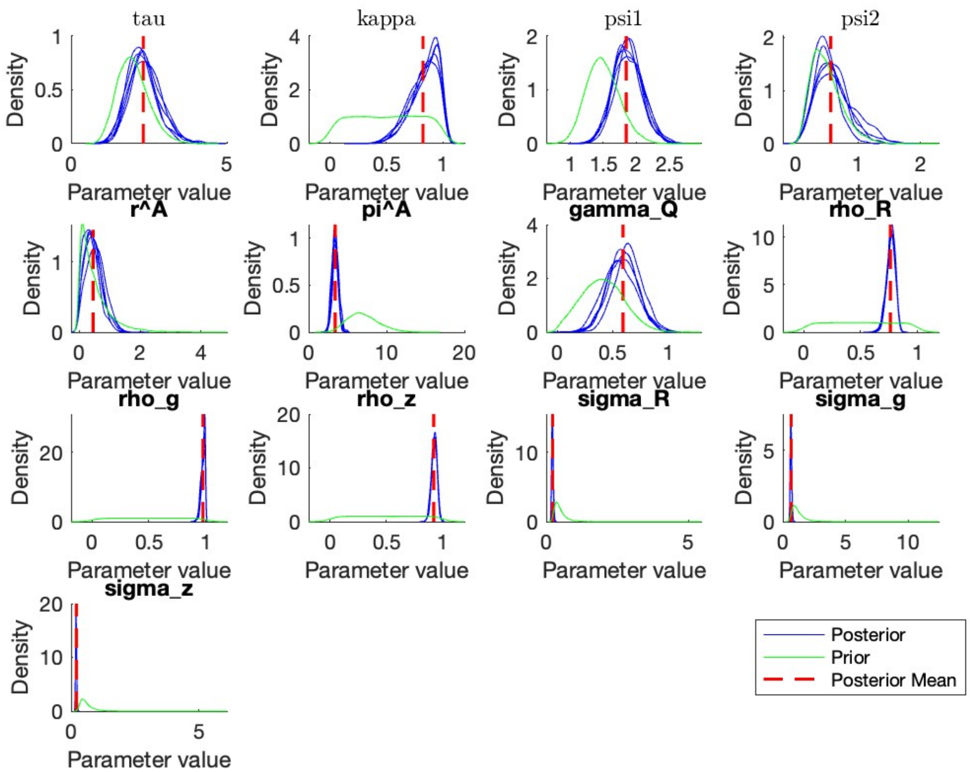

Figure B.1. Prior and Posterior distributions, Rational Expectations

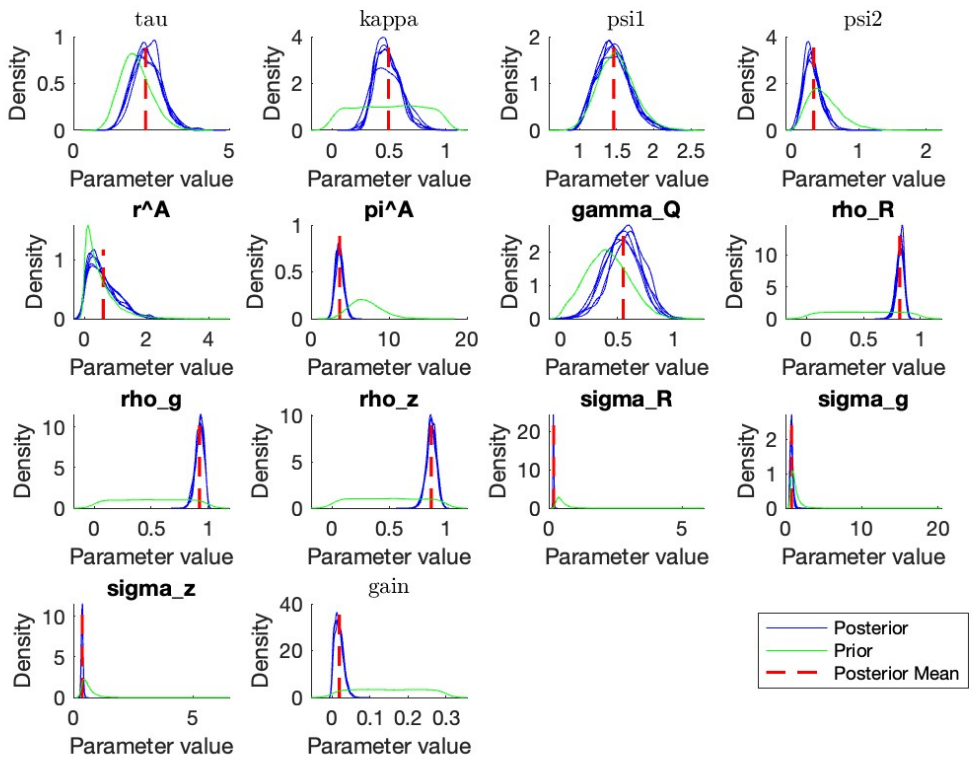

Figure B.2. Prior and Posterior distributions, Equilibrium Based Initial Beliefs

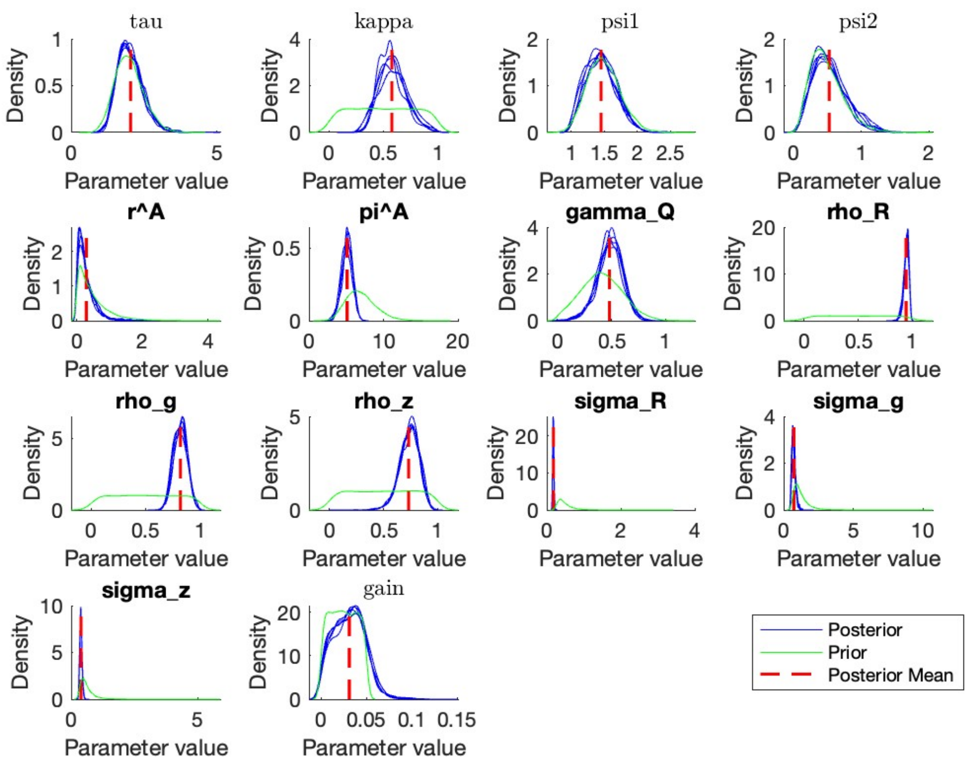

Figure B.3. Prior and Posterior distributions, Training Sample Initial Beliefs

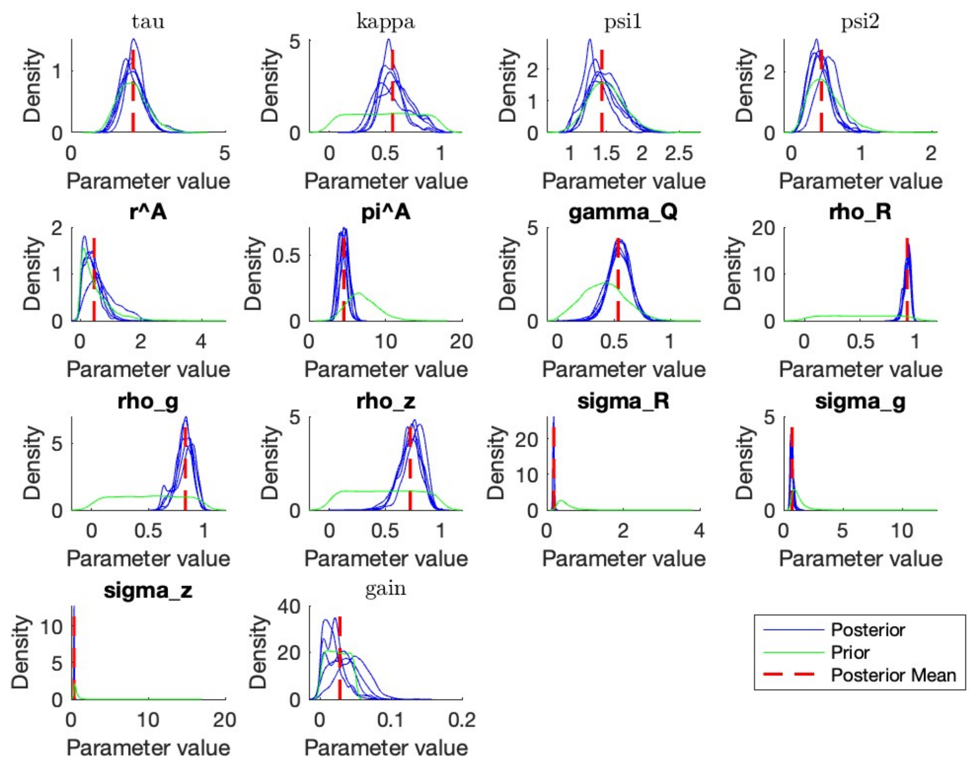

Figure B.4. Prior and Posterior distributions, Jointly Estimated Initial Beliefs

APPENDIX C

TABLES, CHAPTER 2

Table C.1. Prior Distributions for Model Parameters

Parameter	Description	Prior(mean, std)
η	Habit persistence	UNIFORM[0,1]
β	Discount factor	BETA[.99,.01]
σ	Intertemporal Elasticity of Substitution (IES)	GAMMA[0.125, 0.09]
γ	Inflation indexation	UNIFORM[0,1]
ξ_p	Phillips Curve slope	GAMMA[0.015, 0.011]
ω	Marginal Disutility of Work	NORMAL[0.8975, 0.4]
ρ	Taylor Rule Feedback on Interest	UNIFORM[0, 0.97]
ξ_π	Taylor Rule Feedback on Inflation	NORMAL[1.5, 0.25]
ξ_x	Taylor Rule Feedback on Output	NORMAL[0.5, 0.25]
ϕ_r	Natural Interest Rate Coefficient	UNIFORM[0, 0.97]
ϕ_u	Productivity Shock Coefficient	UNIFORM[0, 0.97]
σ_e	Monetary Policy Variance	INV_GAMMA[1, 0.5]
σ_r	Natural Interest Rate Variance	INV_GAMMA[1, 0.5]
σ_u	Productivity Variance	INV_GAMMA[1, 0.5]
$\bar{g}$	Learning Gain	BETA[.031, .022]

Table C.2. SMC Estimates, 5000 particles with 100 stages, Rational Expectations, 5 runs

Parameter	Mean	5% Interval	95% Interval
η	0.46	0.28	0.66
β	0.99	0.97	1.00
σ	0.11	0.06	0.20
γ	0.91	0.81	0.99
ξ_p	0.00	0.00	0.00
ω	0.87	0.24	1.51
ρ	0.89	0.86	0.93
ξ_π	1.45	1.15	1.77
ξ_x	0.38	0.20	0.62
ϕ_r	0.85	0.73	0.93
ϕ_u	0.02	0.00	0.07
σ_e	0.26	0.24	0.28
σ_r	1.82	0.97	3.24
σ_u	0.44	0.39	0.48

Table C.3. SMC Estimates, 5000 particles with 300 stages, Equilibrium Initials, Full Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
η	0.51	0.33	0.73
β	0.99	0.97	1.00
σ	0.11	0.05	0.20
γ	0.91	0.82	0.99
ξ_p	0.00	0.00	0.00
ω	0.87	0.26	1.56
ρ	0.89	0.86	0.93
ξ_π	1.44	1.15	1.76
ξ_x	0.39	0.20	0.63
ϕ_r	0.81	0.68	0.91
ϕ_u	0.02	0.00	0.06
σ_e	0.26	0.23	0.28
σ_r	2.67	1.00	5.49
σ_u	0.44	0.39	0.48
gain	0.0266	0.0045	0.0623

Table C.4. SMC Estimates, 5000 particles with 300 stages, Equilibrium Initials, Limited Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
η	0.79	0.65	0.89
β	0.99	0.96	1.00
σ	0.33	0.18	0.54
γ	0.81	0.59	0.97
ξ_p	0.01	0.00	0.01
ω	0.80	0.18	1.44
ρ	0.91	0.87	0.94
ξ_π	1.49	1.16	1.83
ξ_x	0.38	0.18	0.64
ϕ_r	0.22	0.11	0.32
ϕ_u	0.03	0.00	0.08
σ_e	0.26	0.24	0.28
σ_r	12.20	6.85	20.06
σ_u	0.44	0.40	0.48
gain	0.0126	0.0067	0.0198

Table C.5. SMC Estimates, 5000 particles with 300 stages, Training Sample Initials, Full Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
η	0.70	0.46	0.79
β	0.98	0.95	1.00
σ	0.34	0.17	0.80
γ	0.80	0.61	0.97
ξ_p	0.01	0.00	0.02
ω	0.70	0.18	1.25
ρ	0.94	0.92	0.96
ξ_π	1.53	1.17	1.88
ξ_x	0.63	0.28	1.00
ϕ_r	0.86	0.82	0.89
ϕ_u	0.05	0.00	0.13
σ_e	0.26	0.24	0.28
σ_r	3.19	0.99	4.51
σ_u	0.46	0.39	0.53
gain	0.0522	0.0402	0.0598

Table C.6. SMC Estimates, 5000 particles with 300 stages, Training Sample Initials, Limited Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
η	0.93	0.86	0.99
β	0.98	0.95	1.00
σ	0.23	0.09	0.41
γ	0.94	0.53	1.00
ξ_p	0.00	0.00	0.00
ω	0.79	0.20	1.42
ρ	0.93	0.89	0.96
ξ_π	1.47	1.11	1.84
ξ_x	0.40	0.14	0.72
ϕ_r	0.43	0.29	0.59
ϕ_u	0.13	0.01	0.30
σ_e	0.26	0.24	0.28
σ_r	0.87	0.71	1.07
σ_u	0.43	0.39	0.47
gain	0.0088	0.0045	0.0137

Table C.7. SMC Estimates, 10000 particles with 500 stages, Jointly Estimated Initials, Full Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
β	0.99	0.99	1.00
σ	0.24	0.19	0.29
γ	0.48	0.33	0.66
ξ_p	0.01	0.00	0.01
ω	0.77	0.51	0.98
ρ	0.90	0.88	0.93
ξ_π	1.60	1.46	1.72
ξ_x	0.27	0.19	0.34
ϕ_r	0.89	0.83	0.93
ϕ_u	0.09	0.02	0.15
σ_e	0.26	0.24	0.28
σ_r	1.30	1.11	1.47
σ_u	0.52	0.42	0.58
gain	0.0071	0.0054	0.0089

Table C.8. SMC Estimates, 10000 particles with 500 stages, Jointly Estimated Initials, Limited Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
η	0.77	0.64	0.98
β	0.97	0.94	1.00
σ	0.26	0.14	0.39
γ	0.72	0.44	0.98
ξ_p	0.01	0.00	0.01
ω	1.19	0.68	1.70
ρ	0.93	0.89	0.96
ξ_π	1.49	1.27	1.69
ξ_x	0.46	0.19	0.75
ϕ_r	0.74	0.09	0.94
ϕ_u	0.04	0.00	0.11
σ_e	0.26	0.24	0.29
σ_r	1.00	0.65	1.51
σ_u	0.43	0.40	0.46
gain	0.0106	0.0035	0.0276

Table C.9. SMC Estimates, 5000 particles with 300 stages, Equilibrium Initials, Minimal Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
η	0.74	0.64	0.83
β	0.99	0.96	1.00
σ	0.29	0.16	0.43
γ	0.20	0.00	0.92
ξ_p	0.00	0.00	0.01
ω	0.79	0.19	1.41
ρ	0.92	0.88	0.94
ξ_π	1.50	1.19	1.82
ξ_x	0.41	0.20	0.64
ϕ_r	0.30	0.17	0.44
ϕ_u	0.63	0.01	0.86
σ_e	0.25	0.23	0.28
σ_r	10.76	6.28	17.11
σ_u	0.43	0.39	0.47
gain	0.0157	0.0099	0.0231

Table C.10. SMC Estimates, 5000 particles with 300 stages, Training Sample Initials, Minimal Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
η	0.66	0.44	0.83
β	0.99	0.96	1.00
σ	0.30	0.17	0.47
γ	0.82	0.58	0.98
ξ_p	0.01	0.00	0.01
ω	0.87	0.24	1.53
ρ	0.92	0.88	0.95
ξ_π	1.51	1.17	1.87
ξ_x	0.43	0.20	0.70
ϕ_r	0.32	0.16	0.49
ϕ_u	0.05	0.00	0.13
σ_e	0.26	0.23	0.28
σ_r	9.03	4.37	16.44
σ_u	0.42	0.39	0.46
gain	0.0071	0.0007	0.0191

Table C.11. SMC Estimates, 10000 particles with 500 stages, Jointly Estimated Initials, Minimal Information, 5 runs

Parameter	Mean	5% Interval	95% Interval
η	0.22	0.05	0.40
β	0.99	0.98	1.00
σ	0.37	0.25	0.52
γ	0.16	0.01	0.41
ξ_p	0.01	0.00	0.02
ω	0.79	0.28	1.29
ρ	0.92	0.90	0.95
ξ_π	1.60	1.29	1.91
ξ_x	0.49	0.29	0.70
ϕ_r	0.55	0.33	0.83
ϕ_u	0.79	0.67	0.88
σ_e	0.26	0.24	0.28
σ_r	2.81	1.97	3.81
σ_u	0.39	0.36	0.43
gain	0.0060	0.0013	0.0125

Table C.12. SMC Estimates, 10000 particles with 500 stages, Jointly Estimated Initials, Minimal Information, narrow prior 5 runs

Parameter	Mean	5% Interval	95% Interval
η	0.38	0.15	0.58
β	0.99	0.98	1.00
σ	0.30	0.20	0.42
γ	0.25	0.02	0.54
ξ_p	0.01	0.00	0.02
ω	0.84	0.27	1.42
ρ	0.90	0.87	0.94
ξ_π	1.40	1.12	1.69
ξ_x	0.36	0.18	0.57
ϕ_r	0.30	0.14	0.47
ϕ_u	0.81	0.70	0.91
σ_e	0.26	0.24	0.28
σ_r	4.36	2.80	6.15
σ_u	0.40	0.36	0.43
gain	0.0076	0.0025	0.0142

Table C.13. Monte Carlo Mean and Standard Deviation of Phillips Curve Slope

	Full Info	Limited Info	Minimum Info
Rational Expectations	0.04 (0.01)	N/A	N/A
Equilibrium Initials	0.04 (0.01)	0.07 (0.03)	0.07 (0.00)
Training Sample Initials	0.07 (0.03)	0.11 (0.00)	0.08 (0.00)
Jointly Estimated Initials	0.08 (0.04)	0.31 (0.25)	0.05 (0.01)

Table C.14. Mean and Standard Deviation of Natural Logarithms of the Marginal Likelihoods

	Full Info	Limited Info	Minimum Info
Rational Expectations	-834.92 (0.43)	N/A	N/A
Equilibrium Initials	-835.40 (0.75)	-841.13 (1.41)	-834.79 (0.84)
Training Sample Initials	-746.10 (336.06)	-831.02 (0.44)	-833.61 (0.60)
Jointly Estimated Initials	-831.9359 (11.8306)	-825.32 (5.36)	-836.97 (2.29)
Restricted VAR	N/A	N/A	-831.94 (0.42)

APPENDIX D

FIGURES, CHAPTER 2

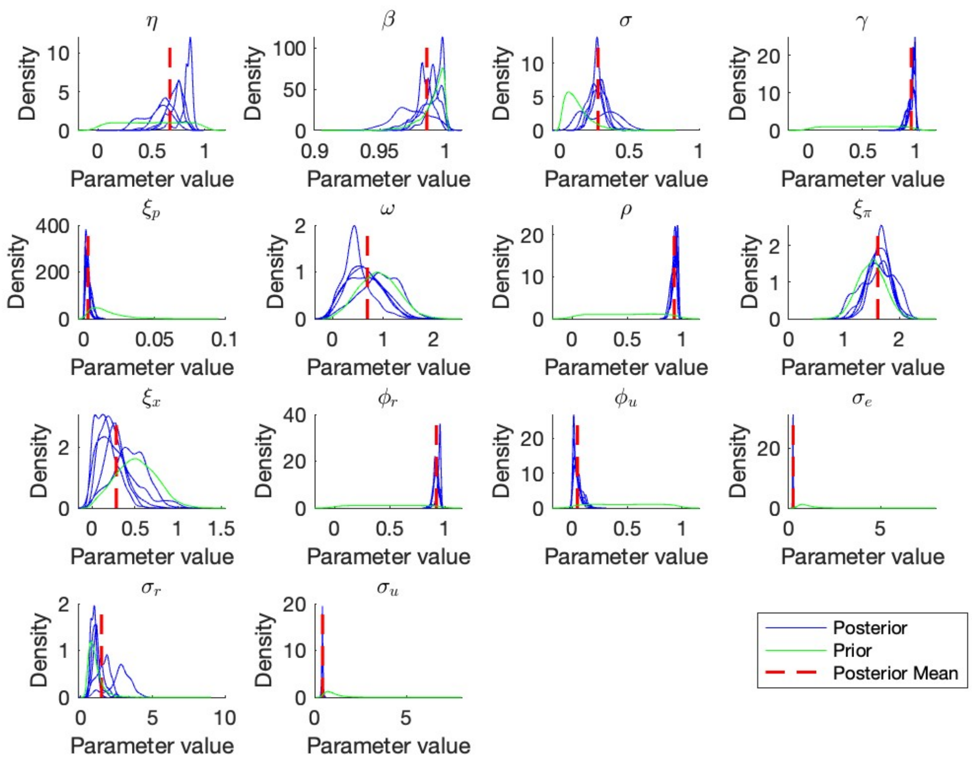

Figure D.1. Prior and Posterior distributions, Rational Expectations

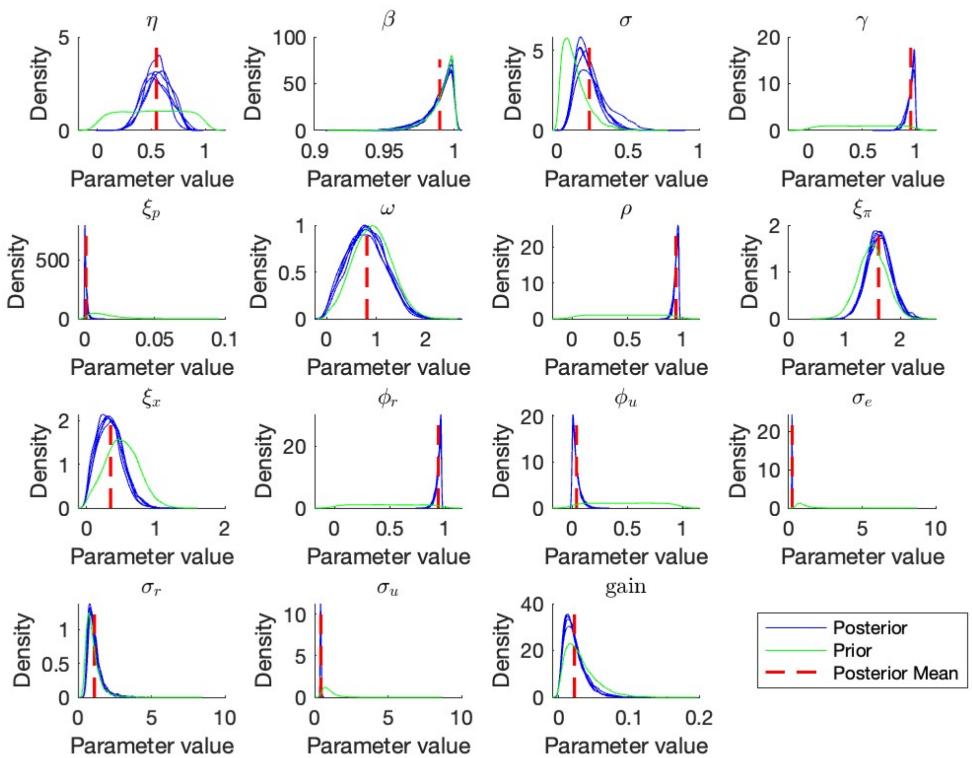

Figure D.2. Prior and Posterior distributions, Equilibrium Initials, Full Information Set

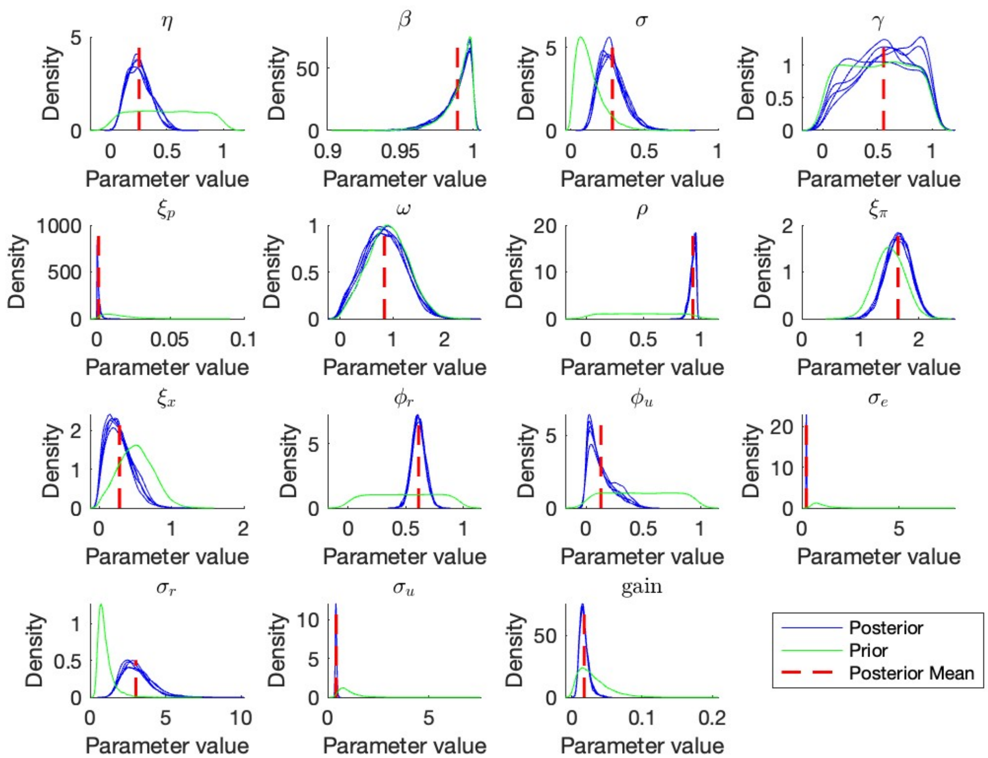

Figure D.3. Prior and Posterior distributions, Equilibrium Initials, Limited Information Set

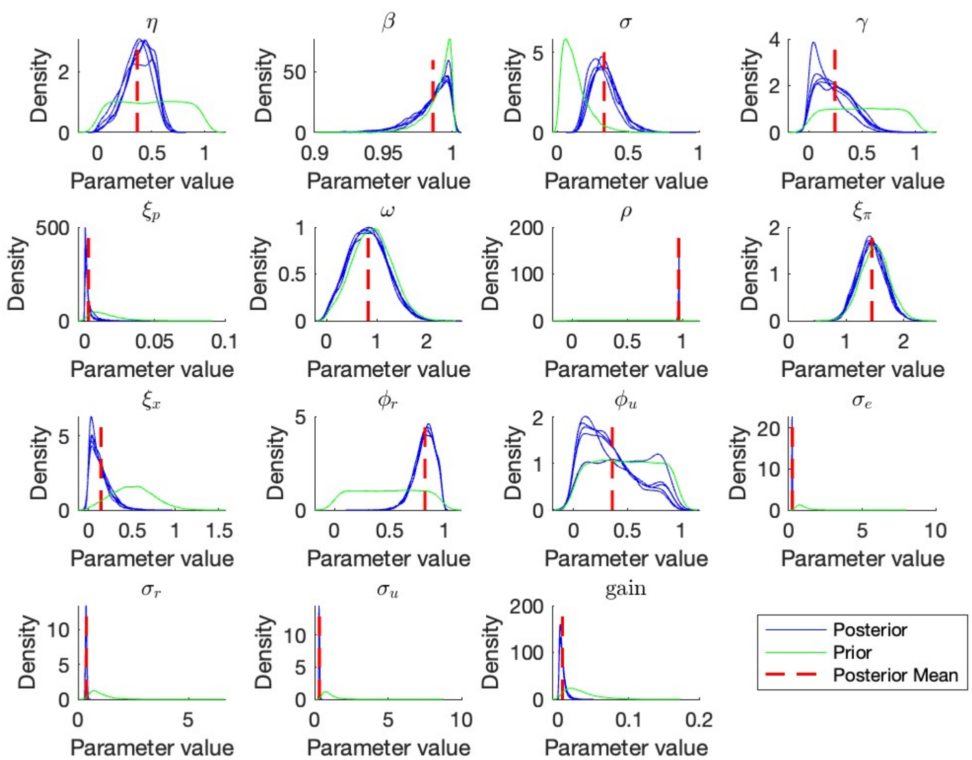

Figure D.4. Prior and Posterior distributions, VAR Initials, Full Information Set

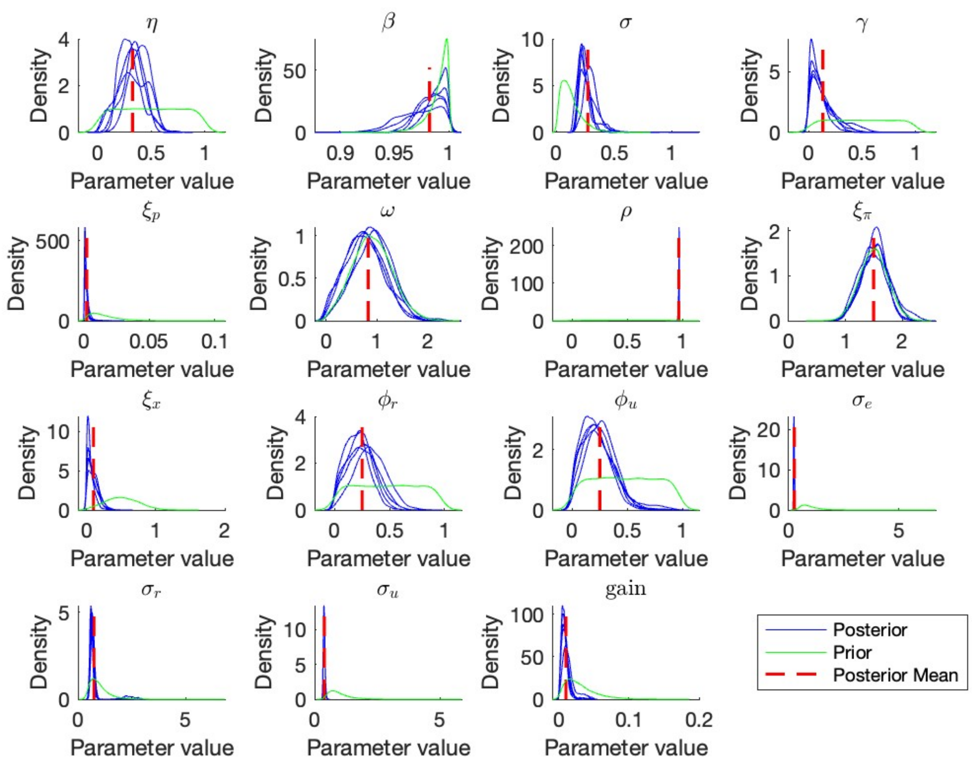

Figure D.5. Prior and Posterior distributions, VAR Initials, Limited Information Set

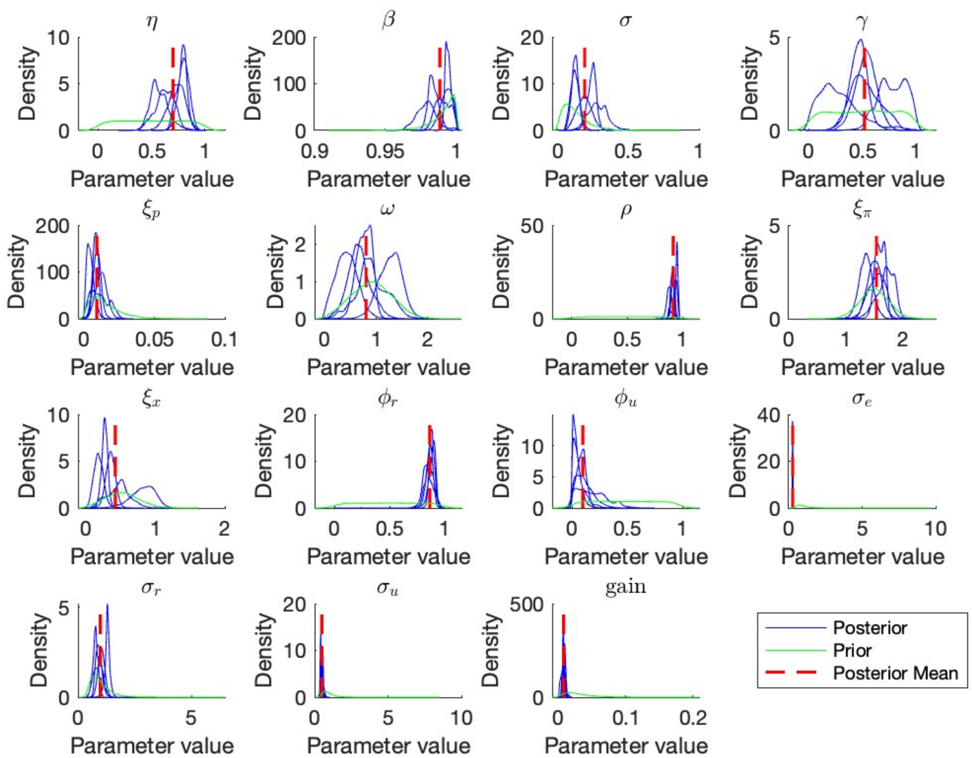

Figure D.6. Prior and Posterior distributions, Jointly Estimated Initials, Full Information Set

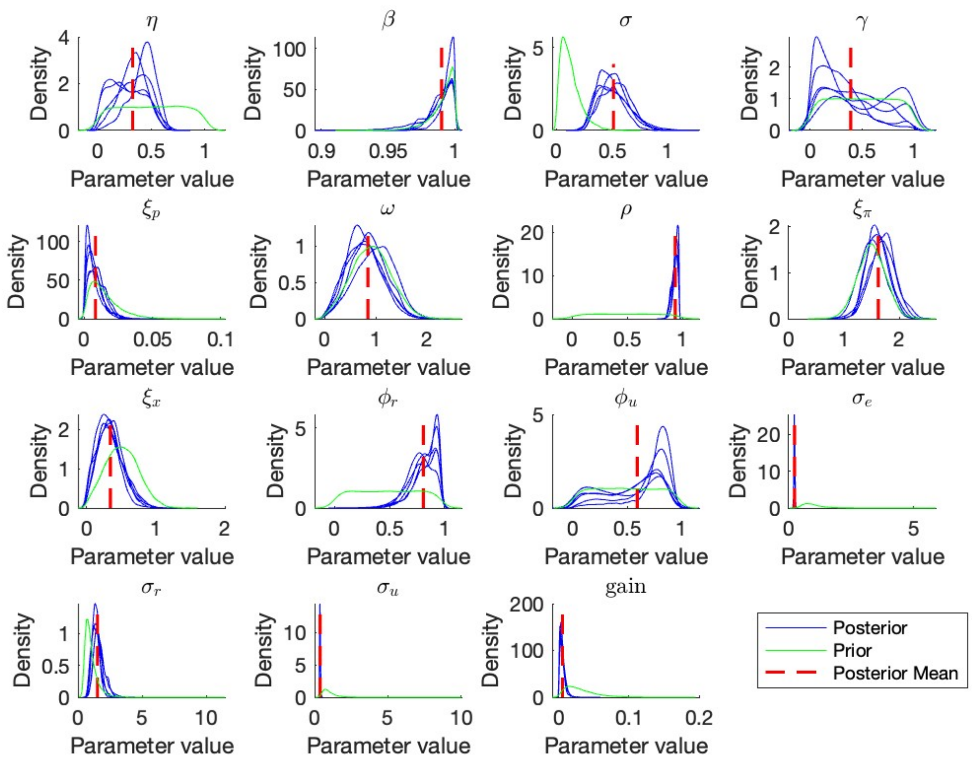

Figure D.7. Prior and Posterior distributions, Jointly Estimated Initials, Limited Information Set

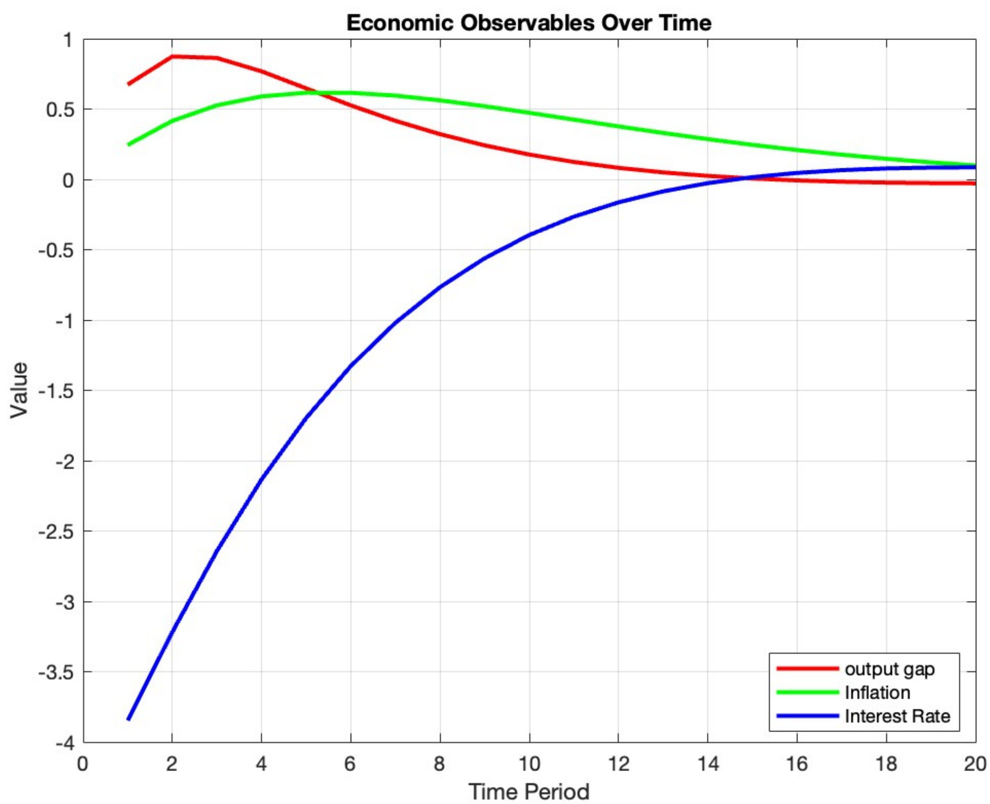

Figure D.8. Impulse Response to Expansionary Monetary Shock, Rational Expectations

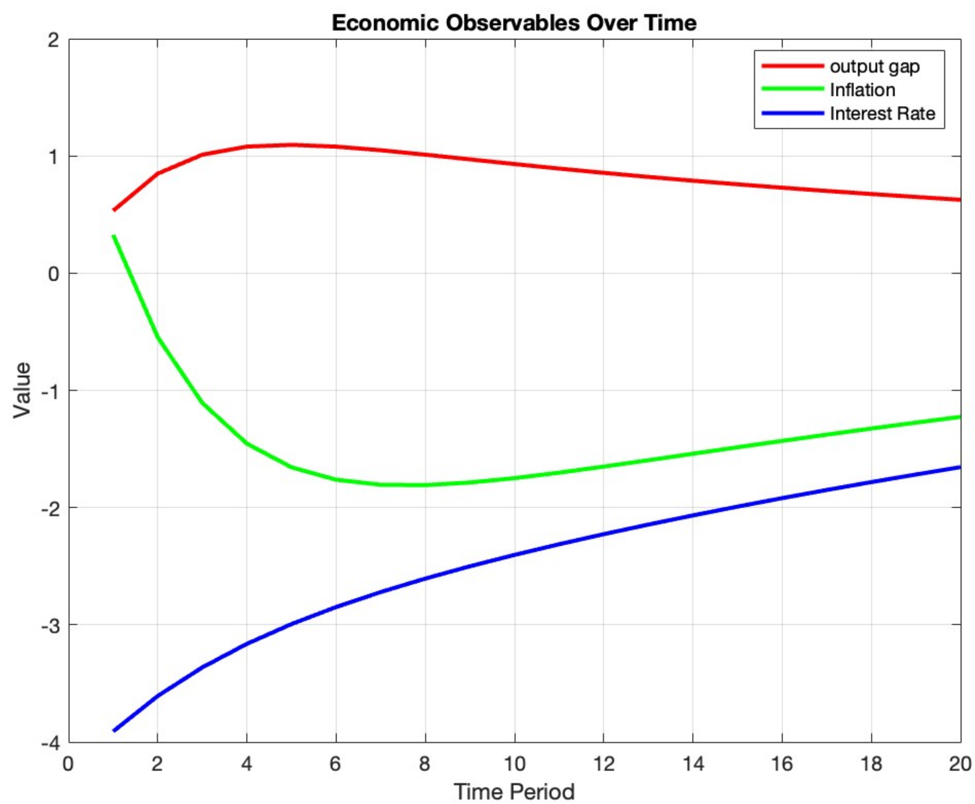

Figure D.9. Impulse Response to Expansionary Monetary Shock, Training Sample Initial beliefs, limited information

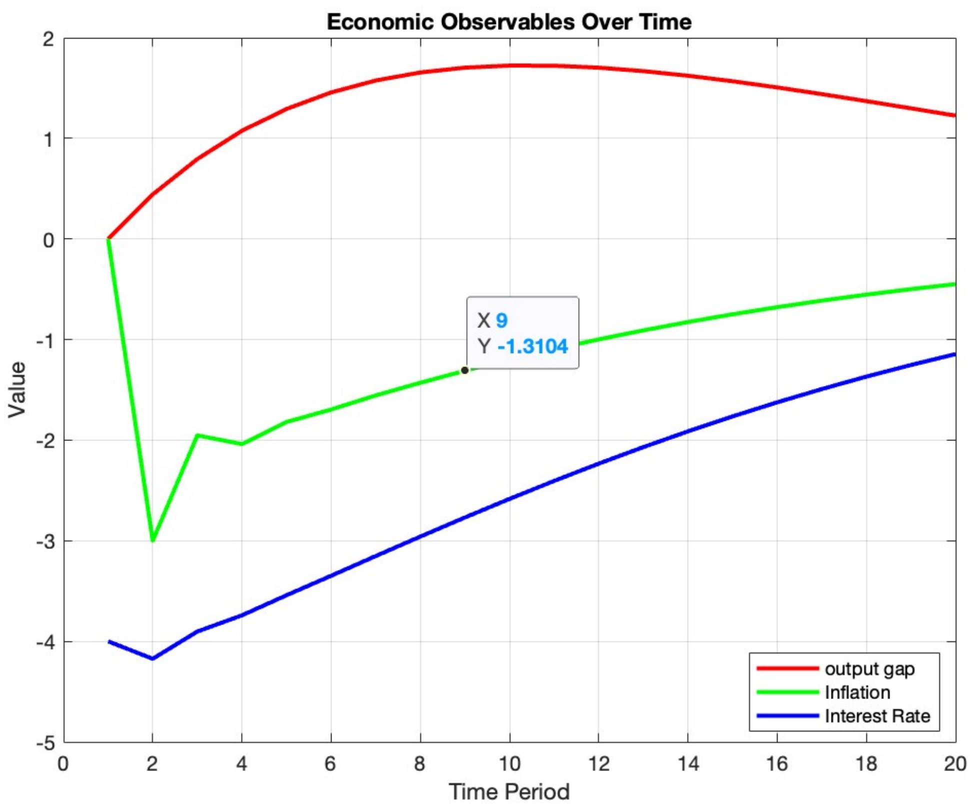

Figure D.10. Impulse Response to Expansionary Monetary Shock, Unrestricted VAR(1)

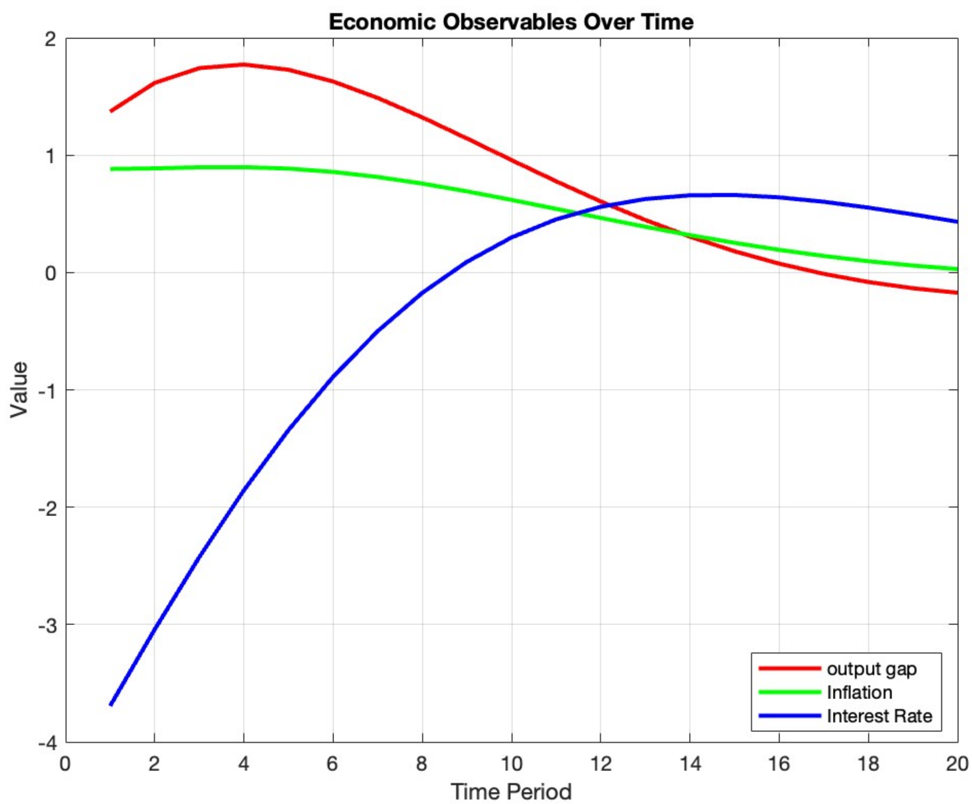

Figure D.11. Impulse Response to Expansionary Monetary Shock, Nearly Unrestricted VAR(1)

REFERENCES CITED

An, S., & Schorfheide, F., *Bayesian Analysis of DSGE Models*, Econometric Rev. **26** (2007), no. 2-4, 113–172, https://doi.org/10.1080/07474930701220071 https://doi.org/10.1080/07474930701220071.

Berardi, M., & Galimberti, J. K., (2017, January), *Smoothing-based Initialization for Learning-to-Forecast Algorithms* (KOF Working papers No. 17-425), KOF Swiss Economic Institute, ETH Zurich, https://doi.org/10.3929/ethz-a-010820132

Bernanke, B., Gertler, M., & Gilchrist, S., *The Financial Accelerator and the Flight to Quality*, The Review of Economics and Statistics **78** (1996), no. 1, 1–15.

Bernanke, B. S., Gertler, M., & Gilchrist, S., *Chapter 21 The financial accelerator in a quantitative business cycle framework*, Vol. 1, Handbook of Macroeconomics, Elsevier, 1999, 1341–1393, https://doi.org/https://doi.org/10.1016/S1574-0048(99)10034-X

Blanchard, O. J., & Kahn, C. M., *The Solution of Linear Difference Models under Rational Expectations*, Econometrica **48** (1980), no. 5, 1305–1311.

Branch, W., & Evans, G., *Model Uncertainty and Endogenous Volatility*, Review of Economic Dynamics **10** (2007), no. 2, 207–237.

Bullard, J., & Mitra, K., *Learning about monetary policy rules*, Journal of Monetary Economics **49** (2002), no. 6, 1105–1129, https://doi.org/https://doi.org/10.1016/S0304-3932(02)00144-7

Cagliarini, A., & Kulish, M., *Solving Linear Rational Expectations Models with Predictable Structural Changes*, The Review of Economics and Statistics **95** (2013), no. 1, 328–336.

Calvo, G. A., *Staggered prices in a utility-maximizing framework*, Journal of Monetary Economics **12** (3 1983), 383–398.

Carceles-Poveda, E., & Giannitsarou, C., *Adaptive learning in practice*, Journal of Economic Dynamics and Control **31** (2007), no. 8, 2659–2697, https://doi.org/https://doi.org/10.1016/j.jedc.2006.09.004

Chib, S., & Jeliazkov, I., *Marginal Likelihood From the Metropolis-Hastings Output*, J. Amer. Statist. Assoc. **96** (2001), 270–281.

Christiano, L. J., Eichenbaum, M., & Evans, *Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy*, Journal of Political Economy **113** (2005), no. 1, 1–45.

Cogley, T., Primiceri, G. E., & Sargent, T. J., *Inflation-Gap Persistence in the US*, American Economic Journal: Macroeconomics **2** (2010), no. 1, 43–69, https://doi.org/10.1257/mac.2.1.43

Cole, S. J., & Milani, F., *THE MISSPECIFICATION OF EXPECTATIONS IN NEW KEYNESIAN MODELS: A DSGE-VAR APPROACH*, Macroeconomic Dynamics **23** (2019), no. 3, 974–1007, https://doi.org/10.1017/S1365100517000104

Del Negro, M., Giannoni, M. P., & Schorfheide, F., *Inflation in the Great Recession and New Keynesian Models*, American Economic Journal: Macroeconomics **7** (2015), no. 1, 168–96, https://doi.org/10.1257/mac.20140097

Estrella, A., *THE PRICE PUZZLE AND VAR IDENTIFICATION*, Macroeconomic Dynamics **19** (2015), no. 8, 1880–1887.

Evans & Honkapohja, *Learning and Expectations in Macroeconomics*, Princeton University Press, 2001.

Fernandez-Villaverde, J., (2009), *The econometrics of dsge models* (NBER Working Papers No. 14677), National Bureau of Economic Research, Inc.

Fernandez-Villaverde, J., Rubio-Ramirez, J., & Schorfheide, F., (2015, December), *Solution and Estimation Methods for DSGE Models* (PIER Working Paper Archive No. 15-042), Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Fernández-Villaverde, J., & Guerrón-Quintana, P. A., (2020, August), *Estimating dsge models: Recent advances and future challenges* (Working Paper No. 27715), National Bureau of Economic Research, https://doi.org/10.3386/w27715

Fuhrer, J. C., *Habit Formation in Consumption and Its Implications for Monetary-Policy Models*, American Economic Review **90** (2000), no. 3, 367–390, https://doi.org/10.1257/aer.90.3.367

Geweke, J. F., (1998, January), *Using simulation methods for Bayesian econometric models: inference, development, and communication* (Staff Report No. 249), Federal Reserve Bank of Minneapolis.

Giannoni, M., & Woodford, M., *Optimal Inflation-Targeting Rules*, In The Inflation-Targeting Debate, NBER Chapters, National Bureau of Economic Research, Inc, 2004.

Hamilton, J. D., *Chapter 50 State-space models*, Vol. 4, Handbook of Econometrics, Elsevier, 1994, 3039–3080, https://doi.org/https://doi.org/10.1016/S1573-4412(05)80019-4

Herbst, E. P., & Schorfheide, F., (2013, June), *Sequential Monte Carlo Sampling for DSGE Models* (NBER Working Papers No. 19152), National Bureau of Economic Research, Inc.

———, *Bayesian Estimation of DSGE Models*, Economics Books **10612**, Princeton University Press, 2016.

Justiniano, A., & Primiceri, G. E., *The Time-Varying Volatility of Macroeconomic Fluctuations*, American Economic Review **98** (2008), no. 3, 604–41, https://doi.org/10.1257/aer.98.3.604

Kydland, F. E., & Prescott, E. C., *Time to Build and Aggregate Fluctuations*, Econometrica **50** (1982), no. 6, 1345–1370.

Ljungqvist, L., & Sargent, T. J., *Recursive Macroeconomic Theory*, The MIT Press, 2012.

Marcet, A., & Nicolini, J. P., *Recurrent Hyperinflations and Learning*, American Economic Review **93** (2003), no. 5, 1476–1498, https://doi.org/10.1257/000282803322655400

Marcet, A., & Sargent, *Convergence of least squares learning mechanisms in self-referential linear stochastic models*, J. Econom. Theory **48** (1989), no. 2, 337–368, https://doi.org/https://doi.org/10.1016/0022-0531(89)90032-X

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E., *Equation of State Calculations by Fast Computing Machines*, Journal of Chemical Physics **21** (1953), no. 6, 1087–1092.

Milani, F., *Expectations, learning and macroeconomic persistence*, Journal of Monetary Economics **54** (2007), no. 7, 2065–2082.

———, *Adaptive Learning and Macroeconomic Inertia in the Euro Area*, Journal of Common Market Studies **47** (2009), no. 3, 579–599, https://doi.org/10.1111/j.1468-5965.2009.

Milani, F., *Learning and time-varying macroeconomic volatility*, Journal of Economic Dynamics and Control **47** (2014), no. C, 94–114, https://doi.org/10.1016/j.jedc.2014.07.01

Modigliani, F., & Miller, M. H., *The Cost of Capital, Corporation Finance and the Theory of Investment*, The American Economic Review **48** (1958), no. 3, 261–297.

Negro, M. D., & Schorfheide, F., *How good is what you've got? DSGE-VAR as a toolkit for evaluating DSGE models*, Economic Review **91** (2006), no. Q 2, 21–37.

Primiceri, G. E., *Why Inflation Rose and Fell: Policy-Makers' Beliefs and U. S. Postwar Stabilization Policy**, The Quarterly Journal of Economics **121** (2006), no. 3, 867–901, https://doi.org/10.1162/qjec.121.3.867 https://academic.oup.com/qje/article-pdf/121/3/867/5408983/121-3-867.pdf.

Rotemberg, J. J., *Monopolistic Price Adjustment and Aggregate Output*, The Review of Economic Studies **49** (1982), no. 4, 517–531.

Sims, C. A., *Are forecasting models usable for policy analysis?*, Quarterly Review **10** (1986), no. Win, 2–16.

———, *Solving Linear Rational Expectations Models*, Computational Economics **20** (2002), no. 1-2, 1–20.

Slobodyan, S., & Wouters, R., *Learning in a Medium-Scale DSGE Model with Expectations Based on Small Forecasting Models*, American Economic Journal: Macroeconomics **4** (2012), no. 2, 65–101, https://doi.org/10.1257/mac.4.2.65

Smets, F., & Wouters, R., *An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area*, Journal of the European Economic Association **1** (2003), no. 5, 1123–1175.

Smets, F., & Wouters, R., *Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach*, American Economic Review **97** (2007), no. 3, 586–606, https://doi.org/10.1257/aer.97.3.586

Villemot, S., *Matlab/LYAPUNOV_SYMM.M.*

Wolters, M., & Wieland, V., *Macroeconomic model comparisons and Forecast Competitions*, 2012.

Woodford, M., *Interest and prices: foundations of a theory of monetary policy*, Princeton Univ. Press, Princeton, NJ [u.a.], 2003.

Zivot, E., *State Space Models and the Kalman Filter*, 2006.