First Thought, Best Thought? Estimating Initial Beliefs in a Bayesian DSGE Model with Adaptive Learning

Jacob Thompson

July 2025

Abstract

In the following, I estimate a small-scale New Keynesian DSGE model in which agents use an adaptive learning scheme as described in (Evans and Honkapohja 2001) to form expectations of future economic variables. A nascent literature explores the empirical importance of adaptive learning in DSGE models and often finds that models with adaptive learning fare better, as measured by marginal data density, than their rational expectations analogues. We add to this literature by providing the first systematic Bayesian evaluation of the choice of initial beliefs with which to endow agents in the model. I first review the Adaptive Learning framework in DSGE modeling, after which I review Bayesian estimation of DSGE models so that the reader can understand the technical challenges that adaptive learning poses specifically to the modeler. Finally, I present the empirical results, including parameter estimates and marginal data densities for different initialization choices.

1 Introduction and Literature Review

The hypothesis of Rational Expectations remains a default assumption in empirical macroeconomic modeling, despite persistent objections raised throughout the literature. These objections include the facially dubious assumptions that agents have the perfect knowledge of the structure of the of the economy and its parameters and that agents can coordinate with each other towards a unique rational expectations equilibrium. Additionally, such models frequently require sources of mechanical persistence with arguably suspect microfoundations in order to provide satisfactory data fit. I estimate one such model presently under both rational expectations and an alternative expectations formation hypotheses.

I first review the important theoretical foundations of adaptive learning and the stability properties of DSGE models with learning. After reviewing the foundations of adaptive learning in DSGE models, I review much of the important empirical literature that studies adaptive learning. I will find that, along several dimensions, DSGE models that relax the assumption of Rational Expectations, including but especially those that embed an adaptive learning rule, improve the fitment of the data to the model. Earlier work, including (Milani 2007) or (Slobodyan and Wouters 2012b) find that mechanical persistence parameters become arguably superfluous when relaxing the assumption of rational expectations.

My paper expands principally upon the work of (Milani 2007) in three dimensions. First, to the best of my knowledge, this paper presents the first Bayesian estimation of agents' initial beliefs, in a model with mechanical persistence, by sampling over an explicit prior distribution of agents' initial beliefs. Second, I estimate models with differing information sets available to agents and find differences in the model fit. I depart from (Milani 2007), (Milani 2014), and (Slobodyan and Wouters 2012a) in omitting Markov-Chain Monte Carlo (MCMC) estimation of DSGE models in favor of a Sequential Monte Carlo (SMC) estimate, as described in (Herbst and Schorfheide 2013). The SMC method sports numerous advantages for the purpose of estimating DSGE models, especially those models with ill-behaved, fairly non-normal posterior densities like those of models that feature adaptive learning.

1.1 The Theory of Adaptive Learning

To motivate my empirical study study of the dynamics of DSGE models with adaptive learning, I review the important theoretical results from the Adaptive Learning literature. Theoretical work in this area deals extensively with the asymptotic properties of DSGE models with adaptive learning. For any dynamic equilibrium solution to the DSGE model, that equilibrium is asymptotically stable under learning if that equilibrium satisfies the "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. Thus the rational expectations solution to the model is the fixed-point of this T-map. The rational expectations equilibrium is considered "expectationally stable" if around the fixed point of the T-map there exists a neighborhood of beliefs wherein the differential equation

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

is asymptotically stable. One algorithm studied often in the literature is decreasing-gain least squares, wherein agents give equal weight to all previous observations and the effect of new data disappears in the limit as $t \to \infty$. Under decreasing gain least squares, agents' beliefs, captured in ϕ_t and the second moment matrix Σ_t are updated according to the following formula:

$$\phi_t = \phi_{t-1} + \frac{1}{t}\Sigma_t^{-1} X_t'(Z_t - \phi_{t-1}' X_t)'$$
$$\Sigma_t = \Sigma_{t-1} + \frac{1}{t}(X_t X_t' - \Sigma_{t-1})$$

where Z_t is the vector of variables observed agents at time t and $X_t = (1, Z_{t-1})'$. (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. A "projection facility" is a behavioral rule that keeps agents beliefs inside a compact set around the fixed point of the T-map, which is the rational expectations equilibrium. This asymptotic property would seem to imply that, in the limit as $t \to \infty$, initial beliefs should not matter to the distribution of beliefs across t.

Despite these attractive asymptotic properties of adaptive learning algorithms, I am presently interested in the short- and medium-run dynamics of the macroeconomy, and a growing empirical and computational literature has documented the importance of these initial conditions. (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. I will evaluate these choices using Bayesian methods, which I now review briefly

1.2 Bayesian Estimation of DSGE models

The benchmark standard for introducing SMC methods as they apply to DSGE models is (Herbst and Schorfheide 2016), who provide a code supplement that estimates a small-scale, purely forward looking DSGE model using this method. In obtaining my own results via SMC, I minimally modified the code provided by the authors of (Herbst and Schorfheide 2016) by replacing the given likelihood and prior density functions with my own.

Below I review the Kalman Filter based likelihod function and the challenges presented by Adaptive Learning. For this I rely principally upon (Zivot 2006), (Ljungqvist and Sargent 2012), and (Hamilton 1994).

Without loss of generality, suppose one has some DSGE model that is described by the equations $\alpha_t = A\alpha_{t-1} + B\alpha^e_{t+1} + C\varepsilon_t$. A unique solution to this model, if it exists, will be a vector autoregressive process $\alpha_t = T\alpha_{t-1} + c_t + R\eta_t$, where α_t is the observed or unobserved state variable. A state space model for an N-dimensional time series y_t consists of a measurement equation relating the observed data to an m-dimensional state vector α_t, and a Markovian transition equation that describes the evolution of the state vector over time. The measurement equation has the form

$$y_t = Z_t\alpha_t + d_t + \varepsilon_t, \quad t = 1,\dots,T,$$

where Z_t is an $N \times m$ matrix, d_t is an $N \times 1$ vector and ε_t is an $N \times 1$ error vector such that

$$\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, H_t).$$

The transition equation for the state vector α_t is the first order Markov process

$$\alpha_t = T_t\alpha_{t-1} + c_t + R_t\eta_t, \quad t = 1,\dots,T,$$

where T_t is an $m \times m$ transition matrix, c_t is an $m \times 1$ vector, R_t is a $m \times g$ matrix, and η_t is a $g \times 1$ error vector satisfying

$$\eta_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, Q_t).$$

The state space representation is completed by specifying the behavior of the initial state

$$\alpha_0 \sim \mathcal{N}(a_0, P_0),$$

$$E[\varepsilon_t a_0'] = 0,\ E[\eta_t a_0'] = 0 \quad \text{for } t = 1,\dots,T.$$

The matrices Z_t, d_t, H_t, T_t, c_t, R_t and Q_t are called the system matrices, and contain non-random elements. If the system matrices are time-invariant then one may compute the initial state and its

variance numerically using a Schur Decomposition. I used the function *lyapunov_symm* developed by (Adjemian 2023) and used by the *DYNARE* package. I use this instead of the kronecker product vectorization and inversion technique described in (Zivot 2006) and used by (Herbst and Schorfheide 2016) as it significantly speeds computation of the likelihood function for higher-dimensional DSGE models.

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

$$\begin{aligned} y_t &= Z_t\alpha_t + d_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, H_t) \\ \alpha_t &= T_t\alpha_{t-1} + c_t + R_t\eta_t, & \eta_t &\sim \mathcal{N}(0, Q_t) \end{aligned}$$

where y_t is an $N \times 1$ vector of observable variables and α_t an $m \times 1$ vector of (possibly) unobserved states. For a sequence of system matrices $\{Z_t, d_t, H_t, T_t, c_t, R_t, Q_t\}_{t=1}^{t=T}$, let $a_t = E(\alpha_t|y^t)$ be the optimal forecast of the state given information through time t, and $P_t = E((a_t - \alpha_t)(a_t - \alpha_t)'|y^t)$ 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:

$$\begin{aligned} a_{t|t-1} &= T_t a_{t-1} + c_t \\ P_{t|t-1} &= Z_t P_{t-1} Z_t' + R_t Q_t R_t' \end{aligned}$$

From these one can derive 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 $E(v_t v_t')$

$$\begin{aligned} y_{t|t-1} &= Z_t a_{t|t-1} + d_t \\ v_t &= y_t - y_{t|t-1} \\ E(v_t v_t') = F_t &= Z_t P_{t|t-1} Z_t' + H_t \end{aligned}$$

The updating equations allows one to update a_t and P_t:

$$a_t = a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} v_t \tag{1}$$

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1} \tag{2}$$

Which, when the system is linear and the innovations Gaussian, allows one to compute the prediction-error-decomposition of the likelihood function:

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

I review the Kalman filter to explain two challenges with this class of models, first the choice of information set with which to endow agents when they are making forecasts at time t, and second the more general question of how to compute the likelihood function when agents use adaptive learning to form expectations. Regarding the first challenge, I study two such information sets, one in which agents know all variables up to $t-1$ and one in which they know all variables up to $t-1$ and the time t i.i.d shocks. Knowledge of the i.i.d shocks does not directly affect the updating equation for a_t in 2, but only affects the updating step indirectly by changing $E(v_t v_t')$ by changing $P_{t|t-1}$ by changing R_t. It is at least possible, in principle, to incorporate knowledge of the i.i.d shocks directly into the updating equation for a_t by deriving a solution to the model in the form of

$$y_t = ay_{t-1} + bw_t \tag{3}$$
$$w_t = cw_{t-1} + d\varepsilon_t \tag{4}$$

through an eigenvalue decomposition solution to the DSGE model rather than the usual Schur decomposition method. However, an eigenvalue decomposition in the form of 4 may not be possible for a model with expectations of varibles at multiple points in the future, such as $t, t+1, t+2$.

One challenge for the DSGE modeler is that these system matrices, specifically T_t are not time-invariant but are instead functions of time-varying beliefs that agents hold. As shown in (Hamilton 1994), this does not by itself preclude computation of the likelihood function as long as those matrices are determinstic functions of prior states, which they are in the case of AL models. Another challenge I frequently encountered is the existence of singular F_t matrices. I was unable to eliminate the production of any such singular matrices, but imposition of a projection facility in which beliefs lied within an open subset of the unit circle seemed to reduce to triviality the incidence of such singular matrices.

Once I have a likelihood function and a prior density function, I can then estimate the posterior density.

The most popular sampler in the DSGE literature, but one which I do not use, is the Metropolis Hastings Random Walk ("MHRW" henceforth). The Sequential Monte Carlo method, I hope to show, offers numerous advantages for the task of estimating DSGE models, especially those that feature expectations formed via an adaptive learning mechanism.

Before describing the algorithm, it will be helpful to recount the basic Importance Sampling algorithm:

Algorithm 1 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:

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

The challenge to basic importance sampling as it applies to DSGE models is constructing a proposal density $g(\theta)$. SMC methods allow for the sequential construction of this proposal density. I recount the SMC algorithm here from (Herbst and Schorfheide 2016). Letting $\{\rho_n\}_{n=1}^{N_\phi}$ be an ex ante provided 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 2 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

$$\tilde{w}_n^i = [p(Y|\theta_{n-1}^i)]^{\phi_n - \phi_{n-1}}$$
$$\tilde{W}_n^i = \frac{\tilde{w}_n^i W_{n-1}^i}{\frac{1}{N}\sum_{i=1}^N \tilde{w}_n^i W_{n-1}^i}, \quad i = 1, \ldots, N.$$

- **Selection.**
  - Case (i): If $\rho_n = 1$, resample the particles via multinomial resampling. Let $\{\hat{\theta}\}_i^N$ 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 $\hat{\theta}_n^i = \theta_{n-1}^i$ and $W_n^i = \tilde{W}_n^i$, $i = 1, \ldots, N$.
- **Mutation.** Propagate the particles $\{\hat{\theta}^i, W_n^i\}$ via N_{MH} steps of a MH algorithm with transition density $\theta_n^i \sim K_n(\theta_n|\hat{\theta}_n^i; \zeta_n)$ and stationary distribution $\pi_n(\theta)$.

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

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

Estimating the model using Sequential Monte Carlo methods gives far more precise and consistent estimates of both the marginal posterior densities of each parameter and the marginal data density. This owes largely to the fact that the accuracy of an SMC posterior simulator scales with the number of particles used in that simulation, as (Herbst and Schorfheide 2013) show in reporting the mean and variance of several dozen SMC estimates of a forward looking DSGE model. As SMC particles are mutated in parallel, the speed of the algorithm scales approximately linearly with the number of processing cores available to the researcher. This enables the researcher to take advantage of a high-performance computing cluster, such as an Amazon Web Services Elastic Cloud Compute instance, to estimate a DSGE precisely, quickly, and consistently. Another advantage SMC methods have over MCMC methods is the behavior of SMC moments in the presence of multi-modal posterior densities. MCMC methods can and frequently do very easily get "stuck" in and around a local maximum away from the global maximum. While the probability of sampling from each local maximum approaches one as the number of draws approaches infinity, this probability can grow very slowly, far too slowly for even the most powerful workstation to adequately sample the posterior in a reasonable amount of time. Sequential monte carlo methods on the other hand effectively have thousands of starting points spread out through the prior of the DSGE parameters. I plot the marginal posterior densities of each of the estimated parameters to show that the SMC is able to effectively sample from multiple local maxima. One final advantage of the SMC sampler is computation of the marginal likelihood. Consider that

$$\widetilde{w}_n^i = \left[p(Y|\theta_{n-1}^i)\right]^{(\phi_n-\phi_{n-1})}$$

$$\frac{1}{N}\sum_{i=1}^{N}\widetilde{w}_n^i W_{n-1}^i \approx \int \left[p(Y|\theta)\right]^{(\phi_n-\phi_{n-1})} \frac{p^{\phi_{n-1}}(Y|\theta)p(\theta)}{\int p^{\phi_{n-1}}(Y|\theta)p(\theta)\,d\theta}\,d\theta$$

$$= \frac{\int p(Y|\theta)^{\phi_n}p(\theta)\,d\theta}{\int p(Y|\theta)^{\phi_{n-1}}p(\theta)\,d\theta}$$

Which implies

$$\prod_{n=1}^{N_\phi}\left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{w}_n^i W_{n-1}^i\right) \approx \int p(Y|\theta)p(\theta)\,d\theta$$

Thus with the SMC sampler I can compute the marginal likelihood sequentially. This is especially useful when computation of the MHME proves intractable due to numerical underflow errors.

When estimating a DSGE model via SMC methods, a researcher must choose the number of particles, the number of stages, and the tempering schedule for the algorithm. The number of particles N increases the overall accuracy of the Monte Carlo approximation, but also increases the number of likelihood evaluations and therefore the computational burden. Increasing the number of stages will reduce the distance between bridge distributions and therefore reduce the cost of maintaining uniformity of particle weights. The cost of increasing the number of stages is that each stage demands more likelihood evaluations. Letting n denote the $n-th$ stage, N_φ denote the number of stages, the sequence $\{\varphi_n\}_{n=0}^{N_\varphi}$ describes the tempering schedule of the algorithm. The parameter λ controls the shape of the tempering schedule

$$\varphi_n = \left(\frac{n}{N_\varphi}\right)^{\lambda}.$$

A large value of λ implies that the bridge distributions will be very similar (and close to the prior) at the very early stages of the algorithm and very different near the end of the algorithm. In the DSGE model applications, (Herbst and Schorfheide 2016) find that value of $\lambda = 2$ to be optimal because, generally speaking, for $\lambda < 2$, information from the likelihood function dominates the prior density too quickly and only a few particles survive the correction and selection steps. For $\lambda > 2$ the bridge distributions become redundant and the algorithm computes the likelihood function too many times unnecessarily.

1.3 Prior Estimation of DSGE models with learning

(Milani 2007), upon whom I expand in this paper, provides an example of an estimated DSGE model with constant-gain least squares learning. In it, the author estimates a small-scale New Keynesian DSGE model in which habit formation in consumption and inflation indexation in price-setting is nested in the following equations governing the output gap and inflation

$$
\begin{aligned}
\tilde{x}_t &= \hat{E}_t \tilde{x}_{t+1} - (1-\beta\eta)\sigma(i_t - \hat{E}_t \pi_{t+1} - r_t^n) \\
\tilde{\pi}_t &= \xi_p[w x_t + [(1-\eta\beta)\sigma]^{-1}\tilde{x}_t] + \beta \hat{E}_t \tilde{\pi}_{t+1} + u_t \\
i_t &= \rho i_{t-1} + (1-\rho)[\phi_\pi \pi_t + \phi_x x_t] + \varepsilon_t \\
r_t^n &= \phi^r r_{t-1}^n + v_t^r \\
u_t &= \phi^u u_{t-1} + v_t^u \\
\tilde{x}_t &\equiv (x_t - \eta x_{t-1}) - \beta\eta \hat{E}(x_{t+1} - \eta x_t) \\
\tilde{\pi}_t &\equiv \pi_t - \gamma \pi_{t-1}
\end{aligned}
$$

The variables x_t, π_t, i_t are the United States output gap, inflation rate, and federal funds target respectively. These are directly observed by the econometrician while $\tilde{x}_t, \tilde{\pi}_t, r_t^n, u_t$ are unobserved state variables. Milani finds that the improvement in data fit from adaptive learning over rational expectations is substantial, with a Bayes factor of 584 in favor of the model with learning. One of the more interesting results obtained by the author is the (near) elimination of indexation and habit persistence as a source of macroeconomic persistence. Under rational expectations, the mean estimate of $\gamma = 0.885$ and the estimate for $\eta = .911$. Under adaptive learning, however, both parameters nearly disappear, with the mean estimate for η being .117 and γ being .03. Milani then estimates the same model, but with the imposition of no mechanical persistence, including habit formation and inflation indexation, and finds that the Bayes factor in favor of the model with learning over the model with rational expectations is over two million. While I seem to have replicated the improved empirical fit of adaptive learning over rational expectations for the just-described model, I struggled to replicate the near-disappearance of mechanical persistence parameters.

A related outcome is presented in (Slobodyan and Wouters 2012a), who examine a model resembling (Smets and Wouters 2007) with the difference that agents form expectations using compact (under-parameterized) forecasting models. The adaptive learning approach in the paper employs a more generalized Kalman filter updating mechanism. In their paper, agents treat the model coefficients themselves as hidden states to be discovered via a Kalman Filter. In their work, the authors estimate the model under both rational expectations and adaptive learning frameworks. One important empirical result (Slobodyan and Wouters 2012a) find deals with the presistence of the wage and markup shock processes. The wage and price markup shocks under both Rational Expectations (RE) and Adaptive Learning (AL) are assumed to follow ARMA(1,1) (autoregressive moving average) processes. The mean estimates for the AR(1) and MA(1) components of the wage markup process are 0.96 and 0.88, respectively, while for the price markup, the AR(1) and MA(1) components are 0.85 and 0.7, respectively. However, under Adaptive Learning, the mean estimates for the AR(1) and MA(1) components of the wage process change to 0.53 and 0.43, respectively, and to 0.28 and 0.48 for the price markup. Moreover, the 90% confidence intervals for these estimates do not overlap, with the exception of the price markup MA(1) component. It is important to note, though, that the parameters describing wage and price stickiness remain present, and there is significant overlap in the reported confidence bounds between rational expectations and adaptive learning approaches.

In addition to removing various mechanical persistence sources, the fluctuating nature of expectations enables researchers to replicate macroeconomic time series featuring time-varying volatil-

ity, even when the actual shocks are homoskedastic by design. This implies that expectational shocks account for the great inflation and subsequent great moderation in US macroeconomic data. This finding is in stark contrast to (Cogley, Primiceri, and Sargent 2010), who attribute the primary cause of this volatility change to an exogenous shift in monetary policy. The incorporation of time-varying beliefs, capable of producing time-varying volatility, significantly contributes to the enhanced data-fit of numerous adaptive learning models compared to their rational expectations counterparts.

A crucial aspect of adaptive learning models is the initialization of learning dynamics. The authors consider several alternative approaches as a robustness check. In the baseline model, initial beliefs are derived from variable coefficients implied by the rational expectations solution. One alternative keeps the $\Sigma, V, \beta_{1|0}$ matrices constant while estimating the rest of the model. Another approach involves estimating the model with static agents' beliefs and then using those beliefs as initial values for a subsequent estimation.

The authors find that the choice of initialization does not significantly impact the model's performance. Adaptive learning consistently improves the marginal likelihood compared to rational expectations, regardless of the initial beliefs used. I expand upon this literature presently in two directions. First, I provide the first systematic Bayesian estimation of initial beliefs with explicit prior distributions upon said initial beliefs, which then gives researchers another dimension along which to compare model-fit. Second, I provide the first SMC estimate of such a model while prior estimates of DSGE models with learning have used MCMC methods that can, plausibly, suffer greatly from irregular posteriors. I turn now to the model itself whose parameters I shall estimate.

2 The Model

The exact model I intend to estimate is a benchmark New Keynesian model with habit persistence and inflation indexation derived in (Woodford 2003), and estimated with adaptive learning by (Milani 2007). It may be helpful to recount the derivation of the linearized model from the optimization problems facing firms, households, and the monetary authority.

2.1 Optimal Consumption

I assume first a continuum of households distributed uniformly on the interval $[0,1]$. Each i^{th} household maximizes a discounted sum of future in-period utilities of the form:

$$E_t \sum_{T=t}^{\infty} \beta^{(T-t)} \left\{ U\left(C_T^i - \eta C_{T-1}^i : \zeta_T\right) - \int_0^1 v\left(h_T^i(j) : \zeta_T\right) dj \right\}.$$

wherein $\beta \in (0,1)$ is the household's exponential discount factor, C_T^i is an index of the household's consumption of each of the differentiated, time-t supplied goods. $h_T^i(j)$ is the amount of household labor supplied for the production of each $j-th$ good. ζ_T is a vector of exogenous aggregate preference shocks. The parameter $0 \leq \eta \leq 1$ captures the degree of habit formation in consumption. Within-period marginal utility then is positive with respect to the deviation of C_T^i from the previous level of consumption ηC_{T-1}^i, and is negative with respect to the quantity of labor supplied. $U(\cdot;\zeta)$ is increasing and concave for in ζ while $v(\cdot;\zeta)$ is concave and increasing for each ζ. I assume

away the money-in-utility feature by working in the limiting case of a cashless economy. E_t represents the expectations operator and, in this exact case, denotes rational expectations. I relax this assumption of rational expectations when estimating the model.

The consumption and price indices are of the Dixit-Stiglitz CES form:

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

wherein $\theta > 1$ measures the elasticity of substitution between differentiated goods. In the optimum, consumption of the $j-th$ good is $c_t^i(j) = C_t^i(p_t(j)/P_t)^{-\theta}$. The model assumes the existence and completeness of markets for Arrow-Debreau securities so that all households face an identical inter-temporal budget constraint and insure fully against idiosyncratic risks. I also assume that the government runs a balanced budget.

Under habit formation, the first-order conditions for optimal consumption imply that:

$$\lambda_t = U_c(C_T - \eta C_{T-1} : \zeta_T) - \beta\eta E_t\left[U_c(C_{T+1} - \eta C_T; \zeta_{T+1})\right]$$

where $\eta > 0$ implies that the marginal utility of period-t income is not equal to the marginal utility of period-t consumption. The period-t lagrange multiplier still satisfies the equality

$$\lambda_t = \beta E_t\left[\lambda_{t+1}(1+i_t)P_t/P_{t+1}\right]$$

Here i_t denotes the risk-free, one-period nominal interest rate. From these two conditions one may derive the first-order approximation of the agent's Euler Equation:

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

wherein

$$\tilde{C}_t = \hat{C}_t - \eta\hat{C}_{t-1} - \beta\eta\mathbb{E}_t\left[\hat{C}_{t+1} - \eta\hat{C}_t\right]$$

. The intertemporal elasticity of substitution is captured by the parameter $\sigma \equiv \frac{U_c}{\bar{C}U_{cc}} > 0$ sans habit formation, $g_t \equiv \frac{\sigma U_{c\zeta}\zeta_t}{U_c}$ captures exogenous preference shocks, and the circumflex operator $\hat{}$ denotes log-deviations from the steady state values. In equilibrium, $C_t = Y_t$, which I can re-express in terms of the output gap to derive my first linearized state variable

$$\tilde{x}_t = \mathbb{E}_t[\tilde{x}_{t+1}] - (1-\beta\eta)\sigma[i_t - \mathbb{E}_t[\pi_{t+1}] - r_t^n]$$

where

$$\tilde{x}_t \equiv x_t - \eta x_{t-1} - \beta\eta\mathbb{E}_t[x_{t+1} - \eta x_t]$$

where $r_t^n \equiv ((1-\eta\beta)\sigma)^{-1}\left[Y_{t+1}^n - g_{t+1} - (Y_t^n - g_t)\right]$ describes the flexible-price real rate of interest.

I now turn to the equation describing the evolution of inflation as a result of the optimal price setting problem for a monopolistically competitive firm

2.2 Optimal Price Setting

The model assumes a continuum of monopolistically competitive firms who adjust prices as in (Calvo 1983) wherein some fraction $0 < 1-\alpha < 1$ of firms are allowed to adjust their price $p_j(t)$. The $1-\alpha$ fraction of firms adjusts their prices according to the rule

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

which describes how firms index their prices to past inflation. The parameter $0 \leq \gamma \leq 1$ measures the degree of indexation. The $i-th$ firm monopolistically supplies the $i-th$ good according to the production technology $y_t(i) = A_t f(h_t(i))$. A_t is an exogenous AR(1) technology shock and $h_t(i)$ is the labor input. $f(\cdot)$ is an increasing and concave function. The firm's stock of capital (and by extension the entire economy's stock of capital) is assumed to be fixed so that labor remains the only variable factor of production. Firms face a common demand curve $y_t(i) = Y_t \left(\frac{p_t(i)}{P_t}\right)^{-\theta}$ for their own product, and aggregate output $Y_t = \left[\int_0^1 y_t(i)^{\frac{\theta-1}{\theta}} di\right]^{\frac{\theta}{\theta-1}}$ and P_t is the aggregate price index. The firm takes as given the aggregate price and output level but monopolistically adjusts its own price and output. This Dixit-Stiglitz aggregator over a continuum of firms yields the valuable property that firms behave monopolistically but that their own pricing and production decisions do not influence the aggregate output or price levels, reducing the dimensionality of the state-space. A finite number of firms or households would require agents keep track of the reaction functions of every other firm and household when choosing its own prices and output levels. All firms are assumed to face identical decision problems and, when allowed to adjust their prices, set a common $p_t(i) = p_t^*$. Thus the aggregate price level follows the process:

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

Firms are assumed to maximize the present-discounted sum of future profits

$$\mathbb{E}_t \left\{ \sum_{T=t}^{\infty} \alpha^{T-t} Q_{t,T} \left[\Pi_T \left(p^* t(i) \left(\frac{P_{T-1}}{P_{t-1}}\right)^{\gamma}\right)\right]\right\}$$

where $Q_{t,T} = \beta^{T-t}\left(\frac{P_t}{P_T}\right)\left(\frac{\lambda_T}{\lambda_t}\right)$ is a stochastic discount factor and $\Pi_T(\cdot)$ denotes period-T nominal profits. Nominal profits are:

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

As households are indexed by j and intermediate-goods producing firms indexed by i, $w_t(i)$ thus represents the wage for labor supplied in the production of the $i-th$ good. Firms discount future profits at rate α because they expect that at any time t they have a probability equal to α that they can set the optimal price. Thus the firm chooses a sequence of optimal prices $\{p_t^*(i)\}$ to maximize the within-t profits given $\{Y_T, P_T, w_T(j), A_T, Q_{t,T}\}$ for $T \geq t, j \in [0,1]$

Log-linearizing this first order condition gives a sequence of $\hat{p^*}_t$:

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

$\hat{p}^*(i) = \log\left(\frac{p_t^*}{P_t}\right)$, $\quad \omega = \frac{v_{yy}\overline{Y}}{v_y}$ is the elasticity of the marginal dis-utility of producing output with respect to an increase in output.

I thus obtain the following law of motion for inflation:

$$\tilde{\pi}_t = \xi_p[\omega x_t + [(1-\eta\beta)\sigma]^{-1}\tilde{x}_t] + \beta\mathbb{E}_t\tilde{\pi}_{t+1} + u_t$$

wherein

$$\begin{aligned}
\tilde{\pi}_t &\equiv \pi_t - \gamma\pi_{t-1} \\
\tilde{x}_t &\equiv (x_t - \eta x_{t-1}) - \beta\eta\mathbb{E}(x_{t+1} - \eta x_t) \\
\xi_p &= \frac{(1-\alpha)(1-\alpha\beta)}{\alpha(1+\omega\theta)}
\end{aligned}$$

$u_t \equiv \frac{\xi_p v_{y,\zeta}}{v_y}\zeta_t$ is an exogenous aggregate supply shock. Thus x_t is the output gap, which is the difference between actual output and the hypothetical flexible-price-equilibrium output. I use the output gap as estimated by FRED, defined by the data series 100*(Real Gross Domestic Product-Real Potential Gross Domestic Product)/Real Potential Gross Domestic Product.

I close the model by imposing a Taylor Rule monetary policy

$$i_t = \rho i_{t-1} + (1-\rho)(\psi_\pi\pi_t + \psi_x x_t) + \varepsilon_t$$

I have thus arrived at the system of equations for the evolution of the endogenous state variables in my economy, plus two exogenous processes:

$\tilde{\pi}_t = \xi_p[\omega\tilde{x}_t + [(1-\eta\beta)\sigma]^{-1}\tilde{x}_t] + \beta\hat{\mathbb{E}}_t\tilde{\pi}_{t+1} + u_t$ (New Keynesian Phillips Curve)

$\tilde{x}_t = \hat{\mathbb{E}}_t\tilde{x}_{t+1} - (1-\beta\eta)\sigma[i_t - \hat{\mathbb{E}}_t\pi_{t+1} - r_t^n]$ (New Keynesian IS Curve)

$i_t = \rho i_{t-1} + (1-\rho)(\psi_\pi\tilde{\pi}_t + \psi_x\tilde{x}_t) + \varepsilon_t$ (Taylor Rule for Monetary Policy)

$r_t^n = \phi^r r_{t-1}^n + v_t^r$ (Natural Interest Rate process)

$u_t = \phi^u u_{t-1} + v_t^u$ (Productivity shock process)

$\tilde{\pi}_t \equiv \pi_t - \gamma\pi_{t-1}$ (Inflation Indexation)

$\tilde{x}_t \equiv (x_t - \eta x_{t-1}) - \beta\eta\hat{\mathbb{E}}_t(x_{t+1} - \eta x_t)$ (Habit Persistence)

where π_t measures inflation, x_t the output gap, i_t the federal funds rate, u_t a supply shock, r_t^n a natural interest rate shock, and ε_t a monetary policy shock. (Woodford and Walsh 2005) Provides a derivation of this very standard NK model. Expectational terms in the above model reflect the probability distribution that agents place over the space of possible values taken on by the endogenous variables. Under rational expectations, these expectations reflect the model-implied equilibrium distribution. Many researchers find this assumption highly implausible, as it assumes that agents know the true structure of the entire economy when even the best economists can never know the true structure of the economy. Other problematic assumptions, including the

ability for agents to coordinate among each other upon a single rational expectation, are explained further in (Evans 2019) and (Evans and Honkapohja 2001). I should note differences in agents' information set between when they form expectations and throughout the rest of the time series. When deriving the REE model correlations, agents are assumed to observe all exogenous shocks while in the estimated model agents are assumed to only to observe the endogenous variables.

2.3 Integrating learning within the model

Throughout this paper, agents will be assumed to have a "Perceived Law of Motion", (PLM) according to which agents perceive the economy as a three-equation VAR:

$$Z_t = a + BZ_{t-1} + C\varepsilon_t$$

where $Z_t \equiv (1, x_t, \pi_t, i_t, r_t^n, u_t)'$ is a 6×1 vector of endogenous variables, a a 5×1 vector of constants, η_t a 3×1 vector of i.i.d. shocks, and b a 5×3 matrix of VAR coefficients. One can use substitution then to find the agents expectation of endogenous variables, $E_t Z_{t+1}$, which is

$$\begin{aligned} E_t Z_{t+1} &= E_t(a + B(a + BZ_t + \varepsilon_t)) \\ &= a + Ba + B^2 Z_{t-1} + BC\varepsilon_t \end{aligned}$$

In the case where agents use a limited information set, they do not perceive the contemporaneous shocks ε_t but only the lagged endogenous variables $(x_t, \pi_t, i_t, r_t^n, u_t)$, in which case expectations are described below:

$$\begin{aligned} E_t Z_{t+1} &= E_t(a + B(a + BZ_t)) \\ &= a + Ba + B^2 Z_{t-1} \end{aligned}$$

Letting $X_t \equiv (1, Z_{t-1})'$, $\hat{\phi}_t = (a_t, b_t, c_t)'$, these coefficients are updated according to the following recursive formula:

$$\begin{aligned} \hat{\phi}_t &= \hat{\phi}_{t-1} + \gamma_t R_t^{-1} X_t (y_t - \hat{\phi}_{t-1}' X_t)' \\ R_t &= R_{t-1} + \gamma_t (X_t X_t' - R_{t-1}) \end{aligned}$$

where R_t is the second-moment matrix of the endogenous variables plus a constant. The researcher has a choice of γ_t in constructing the model. 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 $\gamma_t = \bar{\gamma}$. 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.

2.3.1 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 paper, 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. Further, such informational assumptions upon the monetary authority would be inconsistent with the overall model, which assumes that private agents do not have information about current endogenous variables when forming expectations.

In the model I estimate, agents form expectations of $x_t, \pi_t, i_t, u_t, r_t^n$ 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

$$s_t = a + Bs_{t-1} + C\varepsilon_t$$

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:

$$\begin{aligned}
E_t s_t &= a + Bs_{t-1} + C\varepsilon_t \\
E_t s_{t+1} &= a + B(a + Bs_{t-1} + C\varepsilon_t) = a + Ba + B^2 s_{t-1} + BC\varepsilon_t \\
E_t s_{t+2} &= a + B(a + Ba + B^2 s_{t-1} + BC\varepsilon_t) = a + Ba + B^2 a + B^3 s_{t-1} + B^2 C\varepsilon_t
\end{aligned}$$

This provides a very direct way of solving for the VAR(1) form of the system. Recall the original form of the DSGE model: $s_t = P + Qs_{t+1}^e + Rs_{t-1} + S\varepsilon_t$. From the above system describing the expectations, one can substitute for the matrices P, Q, S with $a + Ba, B^2, B^2 C$ respectively so that the DSGE model reduces directly to:

$$s_t = a + Ba + (B^2 + R)s_{t-1} + (BC + S)\varepsilon_t$$

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, $y_t = [x_t, \hat{\pi}_t, \hat{i}_t]'$, is 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 $\hat{\pi}_t, \hat{i}_t$ are divided by four to yield the state variables π_t, i_t.

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.

3 Results

3.1 Priors on Parameters

As I am attempting to show the impact of initial beliefs by themselves 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 likelihood. Instead, I assume that agents begin life with a simple VAR model of the following form:

$$\begin{bmatrix} x_t \\ \pi_t \\ i_t \\ r_t^n \\ u_t \end{bmatrix} = a + \Phi \begin{bmatrix} x_{t-1} \\ \pi_{t-1} \\ i_{t-1} \\ r_{t-1}^n \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_i & 0 & 0 \\ 0 & \sigma_{rn} & 0 \\ 0 & 0 & \sigma_u \end{bmatrix} \right). \tag{5}$$

To describe our priors for joint estimation, I review the agents' forecasting model. Agents in the model use a vector autoregression of the following form:

$$\underset{(5\times1)}{y_t} = \underset{(5\times1)}{a} + \underset{(5\times5)}{\boldsymbol{B}}\, \underset{(5\times1)}{y_{t-1}} + \underset{(5\times3)}{\boldsymbol{C}}\, \underset{(3\times1)}{\varepsilon_t} \tag{6}$$

Recall likewise the formulae for recursive least squares:

$$\phi_t = \phi_{t-1} + \frac{1}{t}\Sigma_t^{-1} X_t' (Z_t - \phi_{t-1}' X_t)'$$
$$\Sigma_t = \Sigma_{t-1} + \frac{1}{t}(X_t X_t' - \Sigma_{t-1})$$

In my setup, then, $Z_t = (1, y_{t-1}', \varepsilon_t')'$ while $\phi_t = \begin{bmatrix} a & \boldsymbol{B} & \boldsymbol{C} \end{bmatrix}$. For any set of deep parameters for which the Rational Expectations model is determinate, there exists a unique ϕ_{RE} that describes the dynamics of the model. To form the priors on each element, I collect all the parameter draws from the SMC simulated posterior for the Rational Expectations model. I then save every associated $\phi_{RE,i}$ and reshape the elements of interest into a 26×1 vector, and I am left with a $26 \times n$ array, with n being the number of particles drawn during the SMC estimation of the Rational Expectations model. I then fit a multivariate normal distribution to the 26×1 vector. Since I want to sample over a large space of initial beliefs, and since I do not want to arbitrarily increase the posterior density by adding to log prior density, I set each element of the multivariate normal distribution's Σ matrix equal to the maximum of the sample variance and one. This multivariate normal distribution serves as the prior distribution of the estimated beliefs.

3.2 The Data

The data used for estimating the model are from 1982:q1 to 2002:q4. These include the output gap, inflation, and the effective federal funds rate. I depart from (Milani 2007) and (Milani 2014) and use de-meaned values for the Federal Funds Rate and the rate of inflation. I do this because the latent variables generated by the model are zero-mean processes ex hypothesi, but for obvious institutional reasons the rate of inflation and the federal funds rate are not zero-mean processes.

3.3 Rational Expectations Baseline

For each model I estimate, I generate five independent sequential monte carlo distributions. The number of particles and parameter-blocks varied from model to model owing to the number of parameters estimated. For the Rational Expectations model, the Adaptive learning models with equilibrium-based initial beliefs, and the adaptive learning models with training-sample based beliefs I employed 5,000 particles and 300 stages with three randomized parameter blocks in the

Metropolis-Hastings step. For both models with jointly-estimated initial beliefs, I doubled the number of particles to 10,000 and doubled the number of parameter blocks to six. I obtained estimates similar to results found in (Milani 2007). Under Rational Expectations I find fairly high degrees of mechanical persistence in the form of a high value of inflation indexation and a high value of habit persistence, reported in table 2.

3.4 Equilibrium-based Initial beliefs

In all models, agents have a VAR(1) perceived law of motion: $x_t = a + Bx_{t-1} + C\varepsilon_t$. The information set describes which elements of x_t and ε_t are observed by agents when generating forecasts $E_t x_{t+n}, n \in \mathbb{Z}$. Agents update via recursive least squares the linear model $x_t = \phi' z_t$, where z_t is the set of all variables observed by agents at time t that are presumed to affect x_t, and possibly a constant.

For any determinate Rational Expectations model there also exists a unique VAR(1) representation $x_t = a(\theta) + b(\theta)x_{t-1} + c(\theta)\varepsilon_t$. Initializing beliefs around the Rational Expectations solution means to take these matrices $a(\theta), b(\theta), c(\theta)$ and substitute them into agents' perceived law of motion $x_t = a + bx_{t-1} + c\varepsilon_t$. Since the VAR(1) implied by the rational expectations solution is stationary, there also exists a unique $\Sigma(\theta) = E(x_t x_t')$.

Consider a stationary VAR(1) process:

$$x_t = \Phi x_{t-1} + R\varepsilon_t,$$

where x_t and x_{t-1} are $n \times 1$ vectors, Φ is an $n \times n$ matrix of coefficients, R is an $n \times n$ matrix, and ε_t is an $n \times 1$ vector of white noise with covariance matrix Q, ($Q = E[\varepsilon_t \varepsilon_t']$).

I want to find the variance-covariance matrix Σ of the process, where $\Sigma = E[(x_t - E[x_t])(x_t - E[x_t])']$. Since the process is stationary, $E[x_t] = E[x_{t-1}] = \mu$, so $\mu = \Phi\mu + 0$, and $x_t = \Phi x_{t-1} + R\varepsilon_t - \mu$.

The equation defining the variance-covariance matrix Σ can be written as:

$$\begin{aligned}
\Sigma &= E[(x_t - \mu)(x_t - \mu)'] \\
&= E[((\Phi x_{t-1} + R\varepsilon_t) - \mu)((\Phi x_{t-1} + R\varepsilon_t) - \mu)'] \\
&= E[(\Phi(x_{t-1} - \mu) + R\varepsilon_t)(\Phi(x_{t-1} - \mu) + R\varepsilon_t)'] \\
&= \Phi E[(x_{t-1} - \mu)(x_{t-1} - \mu)']\Phi' + RE[\varepsilon_t \varepsilon_t']R' \\
&= \Phi\Sigma\Phi' + RQR'.
\end{aligned}$$

The matrix Σ can be solved for numerically using a "doubling algorithm" or through a Schur decomposition. For a limited-information set, this Σ can be the initial second-moment matrix. For the full-information set, I need to find the covariance of the endogenous variables x_t and the vector of innovations ε_t. One can show algebraically that this is equal to $(I - \Phi)^{-1} R\Sigma_\varepsilon$ and that the second moment matrix implied by the model is

$$R_0 = \begin{bmatrix} \Sigma & (I - \Phi)^{-1} R)\Sigma_\varepsilon \\ ((I - \Phi)^{-1} R\Sigma_\varepsilon)' & \Sigma_\varepsilon \end{bmatrix}.$$

where $\Sigma_\varepsilon = E(\varepsilon_t \varepsilon_t')$.

I report parameter estimates from the for both models with equilibrium initials in tables 3 and 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. This 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.

I report marginal data densities in 9 and 10. Under both datasets, the model with equilibrium-based initials performs at least as well as the rational expectations baseline.

3.5 Training Sample initial beliefs

A common strategy for initializing agents' beliefs is to estimate a model based upon pre-sample data. This strategy is employed 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:

$$x_{t+1} = \Phi x_t + R w_t$$

- Observation equation:

$$y_t = H x_t.$$

In my case the x_t is a (partially) unobserved state vector consisting of the output gap, inflation, the effective federal funds rate, a natural interest rate shock process and a productivity process. The w_t is a vector of i.i.d shocks that directly affect the federal funds rate, the natural interest rate, and the productivity process respectively. The vector y_t consists of the three observable variables, the output gap, inflation, and the federal funds rate.

I estimate via Maximum Likelihood the values $\Phi_{1,1}, \Phi_{1,2}, \ldots$ and the variances of the shocks $\varepsilon_{1,t} \ldots$. I assume that the R matrix is a block diagonal matrix with two zeros and three ones. I make this assumption because, for a white noise process ε_t, it is not possible to separately identify σ_ε^2 from the square root of a constant by which ε_t is multiplied. I also aim to exploit the restrictions implied by the DSGE model itself, and one of the restrictions is a diagonal R matrix. Once I have my parameter vector that maximizes the likelihood function of this model, I then use the implied moment matrices $E(x_t x_t')$ to initialize Φ_0, Σ_0 that I update through constant gain recursive-least-squares adaptive learning with a gain value of .01. The data used in this adaptive learning algorithm are the Kalman-filtered states of the five-equation model. Ideally the gain value would be estimated along with the rest of the model, but it does not affect the likelihood function of the reduced-form SSM. I had attempted to allow the gain value in this original reduced-form model to vary along with the model, but it did not yield any improvements in model likelihood, so I only report those estimates with a gain value of .01.

The structural parameter estimates are reported in 5 and 6, and the estimated marginal data density is reported in 9 and 10. As measured by the marginal likelihood, initializing beliefs via a training sample results can result in slightly improved estimated marginal data density, for the 1982-2002 time series, but not for the 1961-2006 dataset.

I wish to check the robustness of this result to different data sets for the training sample. For that reason I re-estimated the model but only used the previous 30 quarters of macroeconomic data. I wish to re-estimate in this fashion for those instances where a researcher may only have a very limited time series on which to train agents' initial beliefs. The results reported in 9 are the result of this pared-down training sample. As the marginal data densities show, the training sample model still gives significant improvement in model fit over the rational expectations baseline. Our findings, however, are clearly sensitive to the quality of pre-sample data used to train initial beliefs. I report as well estimates of the marginal data density from all seven models estimated using data from 1961-2006 in 10. For the training sample I used data from 1954:q3 to 1961:q1 I find that the Rational Expectations baseline model has a marginal data density of -839 and that only equilibrium-based initials, under both information sets, and jointly-estimated initial beliefs under the limited information perform better than this baseline model. Adaptive learning models with training-sample informed initial beliefs fare significantly worse under both information sets.

3.6 Joint Estimation

Joint estimation of initial beliefs treats each element of the agents' initial beliefs as a parameter with its own prior distribution. The choice of prior distribution is left to the researcher, which I describe now.

At the outset of this project I sought priors that are informed by other estimation procedures. For every parameter draw θ_i, there is a unique auto-regressive transition matrix $\rho(\theta_i)$ and a unique second-moment matrix $\Sigma(\theta_i)$. Thus we can use the simulated posterior to back out the distribution of the elements of the ρ and Σ matrices.

Given this resource, I first sought to use kernel densities fitted to each unique element of the ρ, Σ matrices, but this proved far too computationally expensive to allow for estimation of a DSGE model in any reasonable amount of time. Since the Σ matrix is necessarily symmetric but the ρ matrix likely asymmetric, this forces the researcher to estimate fifteen extra parameters, if the agents use only the three endogenous variables in their perceived law of motion. If, instead, agents use all five state variables in their VAR-PLM, the researcher must estimate an additional forty extra parameters. Letting $N \in \mathbb{N}$ be the number of states assumed to be observed by agents, the number of extra parameters is equal to $N^2 + N(N+1)/2$ since I only require the upper-triangular elements of the Σ matrix.

My final joint estimation pares down significantly the dimensionality of the problem, but does assume some additional but, I argue, reasonable cognitive abilities of the agents. Rather than estimating the elements of agents Σ_0, Φ_0 matrices directly, I estimated a subjective PLM of the following form:

$$\begin{bmatrix} x_t \\ \pi_t \\ i_t \\ r_t^n \\ u_t \end{bmatrix} = \begin{bmatrix} b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\ b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ b_{32} & b_{33} & b_{34} & b_{35} & b_{36} \\ 0 & 0 & 0 & b_{44} & 0 \\ 0 & 0 & 0 & 0 & b_{55} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \pi_{t-1} \\ i_{t-1} \\ r_{t-1}^n \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t^{r^n} \\ \varepsilon_t^u \\ \varepsilon_t^i \end{bmatrix}$$

This PLM assumes that agents know the structure of the economy including at least some of restrictions placed upon possible values of b_t, c_t, namely those elements that cannot take on values different than zero. This is because the autoregressive shocks, if known to be autoregressive, will not be affected by past or present values of any variable except itself. I argue that this is a more reasonable assumption than allowing for the the elements of the b_0 and c_0 matrix to to take on any value as it comports more with the "Cognitive Consistency Principle" which I articulated earlier. The CCP enjoins researchers to assume that agents in their model are at least as intelligent or informed as the researchers themselves. The researcher, in my case, knows that the autoregressive shocks are autoregressive shocks but does not know the value of the autoregressive coefficient nor the standard deviation of its i.i.d shock. Imposition of such restrictions on the b_0, c_0 matrices endows agents with this same information; agents do not know the autoregressive coefficients nor the standard deviations but they do know that these are AR(1) processes.

One shortcoming of this procedure is that, for some parameter draws θ_i, $\Sigma_0(\theta_i)$ is not positive semi-definite, which is required for any covariance matrix. In future work, I aim to incorporate an Inverse-Wishart prior distribution over Σ_0 draws so as to eliminate any Σ_0 that is not PSD.

Under the full information set where agents are assumed to observe $\left[\varepsilon_t^{r^n}, \varepsilon_t^u, \varepsilon_t^i\right]'$, this requires the researcher to estimate 29 additional parameters, namely the 26 non-zero elements of the b, c matrices and the variances of the i.i.d shocks. Under the limited information set where agents are only assumed to observe $[x_t, \pi_t, i_t, r_t^n, u_t]'$, researchers need only estimate the 17 non-zero elements of the b matrix and the variances of the i.i.d shocks for a total of 20 additional parameters. The prior distribution I use is informed by previous SMC estimations of the model under rational expectations. Vectorizing ϕ_0 into a 26×1 allows me to use a multivariate normal distribution as my prior for the elements of the b, c matrices. I use inverse gamma distributions for agents' subjective σ_i^2.

Having derived how I estimate ϕ_0, I now need to derive the initial second moment matrix of regressors, or $\mathbb{E}(x_t, \pi_t, r_t^n, u_t)'(x_t, \pi_t, r_t^n, u_t)$ in the case of the limited information set and $\mathbb{E}(x_t, \pi_t, r_t^n, u_t, \varepsilon_t^{r^n}, \varepsilon_t^u, \varepsilon_t^i)'(x_t,$ This is fairly straightforward as I have already completed this exercise for initializing beliefs around the Rational Expectations solution:

$$R_0 = \begin{bmatrix} \Sigma_x & (I-b)^{-1}c)\Sigma_\varepsilon \\ ((I-b)^{-1}c\Sigma_\varepsilon)' & \Sigma_\varepsilon \end{bmatrix}.$$

where Σ_x is solved for using a Schur decomposition method. For the limited information set, this second moment matrix becomes

$$R_0 = \begin{bmatrix} \Sigma_x & (I-b)^{-1})\Sigma_\varepsilon \\ ((I-b)^{-1}\Sigma_\varepsilon)' & \Sigma_\varepsilon \end{bmatrix}.$$

Estimation results from Joint Estimation of initial beliefs, however, showed some promise compared to all other initialization schemes, for two reasons. First, as shown in 10, the model with

jointly estimated initial beliefs and a limited information set had the highest estimated marginal data density while also having a lower variance than three other estimated models. The two models with jointly estimated initial beliefs with limited and full information sets had Bayes factors of over 900,000 and over 3000 respectively over the Rational Expectations baseline. Second, posterior estimates of the model's deep parameters were fairly consistent between SMC runs, relative to the Rational Expectations baseline, as shown in figure 7.

3.7 Evolution of beliefs

For each of the estimated models with adaptive learning, I also plotted percentile charts that display the estimated evolution of agents' beliefs. Supposing agents have the PLM $x_t = a + bx_{t-1} + c\varepsilon_t$, the figures 8 through 13 show the percentiles of each element of $\hat{\phi}_t = [\hat{a}, \hat{b}, \hat{c}]$ in the case of the full information model or $\hat{\phi}_t = [\hat{a}, \hat{b}]$ in the case of the limited information model, where $\hat{y}$ denotes the difference between the estimated adaptive learning value and the rational expectations solution value. Of concern was the apparently low estimated learning gain for each model paired with, relative to (Milani 2007) and (Slobodyan and Wouters 2012b) and (Slobodyan and Wouters 2012a) a relatively high value for the mechanical persistence parameters η, γ. A low estimated learning gain indicates agents are not encountering large forecast errors and thus not updating their beliefs by a large magnitude. One explanation I sought to rebut was the possibility that my chosen projection facility was artificially constraining agents' beliefs within a region outside one warranted by the data.

For any parameter guess $\hat{\theta}_i$ one can plot, using the kalman filtered states, the estimated agents' beliefs and the difference between that and the rational expectations implied coefficients. A very flat series indicates little to no updating while a series with a high variance indicates substantial updating. Looking at the graph reveals that this problem of beliefs not updating due to the projection facility, a problem one might call "belief degeneracy," arises only in the case where beliefs are initialized with a training sample model with a full information set. This problem of belief degeneracy does not arise in any other initialization scheme with any other information set. That belief degeneracy would arise in this model should not surprise the reader, however, as the training sample initialization scheme forces initial beliefs to a single point in the possible space of initial beliefs, regardless of the other model parameters. This can force initial beliefs to stay far away from the rational expectations solution that previous theoretical work has shown is expectationally stable under a large set of assumptions, as described in (Evans and Honkapohja 2001).

4 Summary and Conclusions

I sought in this paper to evaluate three different choices for initializing agents' beliefs in a small-scale New Keynesian DSGE model; those choices included centering beliefs at the rational expectations equilibrium-implied coefficients and second-moments, using a training sample of data to estimate a reduced-form VAR model, and finally jointly estimating initial beliefs along with the structural parameters of the model. The results shown here may guide future research and inform the choice of initialization scheme that modelers choose in DSGE models with adaptive learning. I evaluated those choices according to the estimated marginal data density. I review the biggest advantages and disadvantages of each choice presently.

Equilibrium-based initialization was shown, in both the shorter twenty-year data set and the much longer forty-year data set, and with both information sets, to perform at least as well or better than the rational expectations baseline. However, under no information set or data set did equilibrium initials give conclusively the highest marginal data density. For both data sets, jointly estimated initial beliefs under a limited information set provided the highest estimated marginal data density, but with a somewhat higher variance of that estimate than equilibrium or training sample based initial beliefs. Finally, training sample based initial beliefs performed the worst out of the three, delivering marginal data densities lower than the Rational Expectations baseline under both information sets and both data sets on which to estimate the model. The significant change in marginal data density when moving from the small to the large data set reveals a weakness of initializing beliefs using training samples, namely that the quality of said beliefs depends on the quality of the training sample data itself. Having investigated all three initialization choices, I propose the following heuristic to guide future research in this area: For models wherein agents use relatively small forecasting models so that the number of additional estimated parameters is low, researchers should try to jointly estimate those initial beliefs even if they must impose a relatively impoverished forecasting model upon the agents. For models wherein agents must use larger forecasting models, such as in the model of (Smets and Wouters 2007) and a richer information set, researchers should instead initialize beliefs around the rational expectations equilibrium.

References

ADJEMIAN, S. (2023): "lyapunov_symm.m," `https://git.dynare.org/Dynare/dynare/-/blob/master/matlab/lyapunov_symm.m`.

BERARDI, M., AND J. K. GALIMBERTI (2017): "Smoothing-based Initialization for Learning-to-Forecast Algorithms," KOF Working papers 17-425, KOF Swiss Economic Institute, ETH Zurich.

BRANCH, W., AND G. EVANS (2007): "Model Uncertainty and Endogenous Volatility," *Review of Economic Dynamics*, 10(2), 207–237.

BULLARD, J., AND K. MITRA (2002): "Learning about monetary policy rules," *Journal of Monetary Economics*, 49(6), 1105 – 1129.

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

CARCELES-POVEDA, E., AND C. GIANNITSAROU (2007): "Adaptive learning in practice," *Journal of Economic Dynamics and Control*, 31(8), 2659 – 2697.

COGLEY, T., G. E. PRIMICERI, AND T. J. SARGENT (2010): "Inflation-Gap Persistence in the US," *American Economic Journal: Macroeconomics*, 2(1), 43–69.

COLE, S. J., AND F. MILANI (2019): "THE MISSPECIFICATION OF EXPECTATIONS IN NEW KEYNESIAN MODELS: A DSGE-VAR APPROACH," *Macroeconomic Dynamics*, 23(3), 974–1007.

EVANS, AND HONKAPOHJA (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press.

EVANS, G. W. (2019): "Adaptive Learning in Macroeconomics," https://pages.uoregon.edu/gevans/AdaptiveLearningBambergJune2019withFigs.pdf, University of Oregon and University of St. Andrews.

GIANNONI, M., AND M. WOODFORD (2004): "Optimal Inflation-Targeting Rules," in *The Inflation-Targeting Debate*, NBER Chapters. National Bureau of Economic Research, Inc.

HAMILTON, J. D. (1994): "Chapter 50 State-space models," vol. 4 of *Handbook of Econometrics*, pp. 3039 – 3080. Elsevier.

HERBST, E. P., AND F. SCHORFHEIDE (2013): "Sequential Monte Carlo Sampling for DSGE Models," NBER Working Papers 19152, National Bureau of Economic Research, Inc.

——— (2016): *Bayesian Estimation of DSGE Models*, no. 10612 in Economics Books. Princeton University Press.

LJUNGQVIST, L., AND T. J. SARGENT (2012): *Recursive Macroeconomic Theory, Third Edition*, vol. 1 of *MIT Press Books*. The MIT Press.

MARCET, A., AND SARGENT (1989): "Convergence of least squares learning mechanisms in self-referential linear stochastic models," *Journal of Economic Theory*, 48(2), 337 – 368.

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

——— (2014): "Learning and time-varying macroeconomic volatility," *Journal of Economic Dynamics and Control*, 47(C), 94–114.

SLOBODYAN, S., AND R. WOUTERS (2012a): "Learning in a Medium-Scale DSGE Model with Expectations Based on Small Forecasting Models," *American Economic Journal: Macroeconomics*, 4(2), 65–101.

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

SMETS, F., AND R. WOUTERS (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97(3), 586–606.

WOODFORD, M. (2003): *Interest and prices*. Princeton Univ. Press, Princeton, NJ [u.a.].

WOODFORD, M., AND C. E. WALSH (2005): "Interest And Prices: Foundations Of A Theory Of Monetary Policy," *Macroeconomic Dynamics*, 9(3), 462–468.

ZIVOT, E. (2006): "State Space Models and the Kalman Filter," https://faculty.washington.edu/ezivot/econ584/notes/statespacemodels.pdf.

A Tables

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

Parameter	Mean	5% Interval	95% Interval
η	0.67	0.52	0.78
β	0.99	0.97	1.00
σ	0.26	0.16	0.34
γ	0.97	0.91	1.00
ξ_p	0.00	0.00	0.00
ω	0.75	0.19	1.25
ρ	0.95	0.92	0.97
ξ_π	1.56	1.30	1.84
ξ_x	0.43	0.23	0.61
ϕ_r	0.94	0.91	0.96
ϕ_u	0.04	0.00	0.12
σ_e	0.23	0.20	0.26
σ_r	1.15	0.84	1.49
σ_u	0.40	0.36	0.44

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

Parameter	Mean	5% Interval	95% Interval
η	0.54	0.33	0.72
β	0.99	0.97	1.00
σ	0.19	0.09	0.33
γ	0.96	0.88	1.00
ξ_p	0.00	0.00	0.00
ω	0.86	0.25	1.47
ρ	0.94	0.91	0.97
ξ_π	1.58	1.24	1.94
ξ_x	0.35	0.07	0.68
ϕ_r	0.95	0.90	0.97
ϕ_u	0.04	0.00	0.12
σ_e	0.23	0.20	0.26
σ_r	1.10	0.63	1.80
σ_u	0.42	0.36	0.48
$\bar{g}$	0.0218	0.0049	0.0484

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

Parameter	Mean	5% Interval	95% Interval
η	0.26	0.10	0.45
β	0.99	0.97	1.00
σ	0.29	0.16	0.45
γ	0.59	0.10	0.97
ξ_p	0.00	0.00	0.00
ω	0.81	0.19	1.48
ρ	0.93	0.88	0.97
ξ_π	1.65	1.27	2.00
ξ_x	0.28	0.04	0.62
ϕ_r	0.61	0.52	0.71
ϕ_u	0.12	0.01	0.36
σ_e	0.23	0.21	0.27
σ_r	3.01	1.75	4.94
σ_u	0.42	0.36	0.48
$\bar{g}$	0.0176	0.0088	0.0313

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

Parameter	Mean	5% Interval	95% Interval
η	0.40	0.14	0.61
β	0.98	0.96	1.00
σ	0.34	0.20	0.52
γ	0.23	0.02	0.53
ξ_p	0.00	0.00	0.01
ω	0.82	0.22	1.45
ρ	0.97	0.96	0.97
ξ_π	1.43	1.03	1.83
ξ_x	0.13	0.01	0.35
ϕ_r	0.82	0.66	0.95
ϕ_u	0.37	0.03	0.83
σ_e	0.23	0.21	0.27
σ_r	0.37	0.33	0.43
σ_u	0.32	0.27	0.37
$\bar{g}$	0.0061	0.0012	0.0153

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

Parameter	Mean	5% Interval	95% Interval
η	0.24	0.09	0.42
β	0.98	0.96	1.00
σ	0.24	0.18	0.33
γ	0.15	0.01	0.38
ξ_p	0.00	0.00	0.00
ω	0.80	0.23	1.38
ρ	0.97	0.96	0.97
ξ_π	1.42	1.05	1.81
ξ_x	0.15	0.02	0.34
ϕ_r	0.27	0.06	0.51
ϕ_u	0.22	0.03	0.45
σ_e	0.24	0.21	0.27
σ_r	0.62	0.50	0.75
σ_u	0.39	0.34	0.44
$\bar{g}$	0.0106	0.0025	0.0212

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

Parameter	Mean	5% Interval	95% Interval
η	0.28	0.13	0.42
β	1.00	0.99	1.00
σ	0.27	0.18	0.38
γ	0.93	0.77	1.00
ξ_p	0.00	0.00	0.01
ω	0.67	0.27	1.05
ρ	0.93	0.90	0.96
ξ_π	1.81	1.44	2.08
ξ_x	0.31	0.11	0.59
ϕ_r	0.87	0.76	0.94
ϕ_u	0.12	0.01	0.36
σ_e	0.24	0.21	0.28
σ_r	1.12	0.83	1.44
σ_u	0.53	0.41	0.74
$\bar{g}$	0.0145	0.0079	0.0256

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

Parameter	Mean	5% Interval	95% Interval
η	0.39	0.12	0.66
β	0.99	0.98	1.00
σ	0.60	0.37	0.87
γ	0.37	0.02	0.94
ξ_p	0.01	0.00	0.02
ω	0.80	0.28	1.35
ρ	0.93	0.89	0.96
ξ_π	1.65	1.29	2.01
ξ_x	0.31	0.07	0.61
ϕ_r	0.79	0.60	0.94
ϕ_u	0.67	0.17	0.90
σ_e	0.23	0.20	0.27
σ_r	1.48	0.84	2.35
σ_u	0.38	0.33	0.44
$\bar{g}$	0.0062	0.0011	0.0138

Table 9: Mean and Standard Deviation of Natural Logarithms of the Marginal Likelihoods, 1982-2002 data

	Full Information	Limited Information
Rational Expectations	-331.8948 (0.9613)	N/A
Equilibrium Initials	-329.5719 (0.9909)	-332.9946 (0.4705)
Training Sample Initials	-650.1220 (112.3248)	-351.2298 (7.3816)
Jointly Estimated Initials	-328.9922 (2.6250)	-326.1411 (0.9756)

Table 10: Mean and Standard Deviation of Natural Logarithms of the Marginal Likelihoods, 1961-2006 data

	Full Information	Limited Information
Rational Expectations	-839.2973 (3.4430)	N/A
Equilibrium Initials	-833.9175 (0.8594)	-838.5927 (0.7877)
Training Sample Initials	-2764.7828 (0.6802)	-859.9934 (0.3554)
Jointly Estimated Initials	-885.0740 (12.0577)	-833.8648 (1.7775)

B Figures

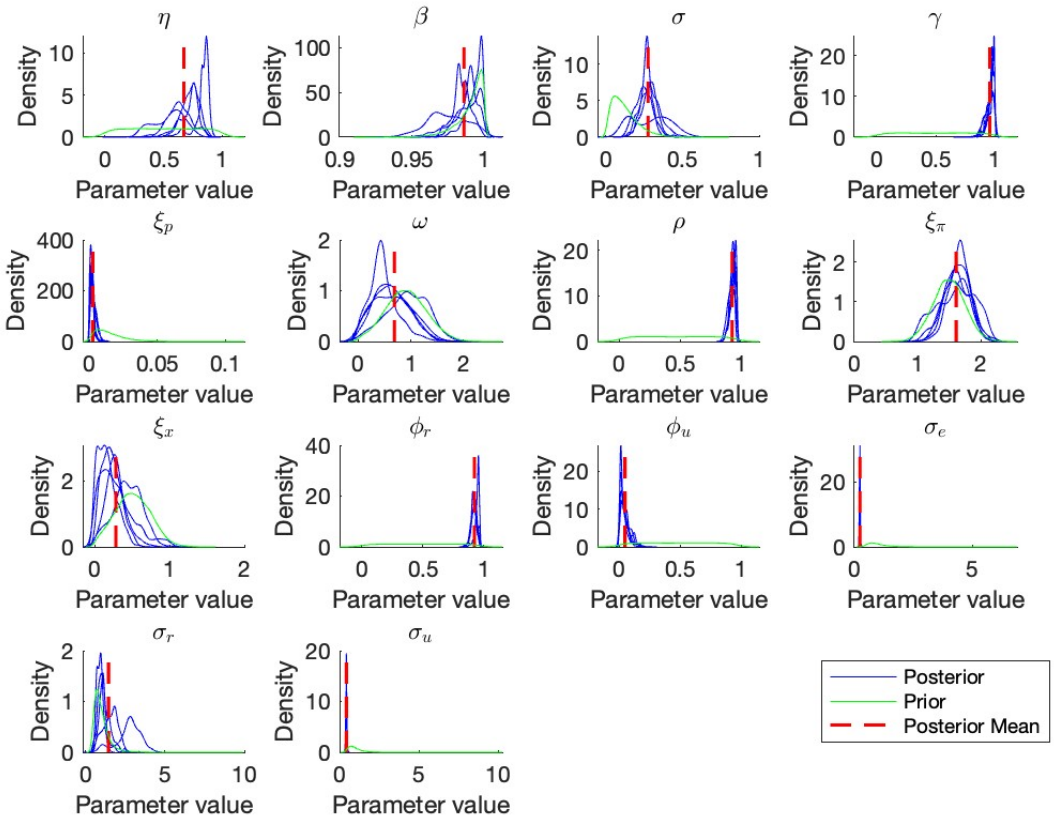

Figure 1: Marginal Posterior Density, SMC, Rational Expectations

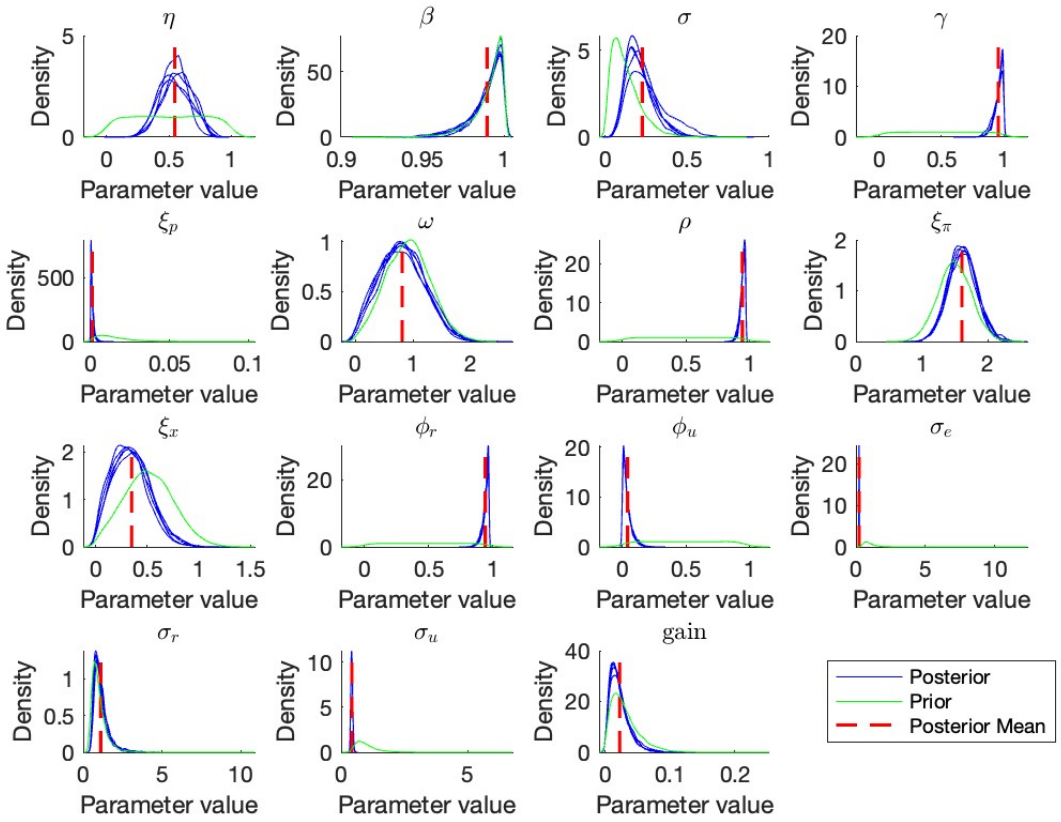

Figure 2: Marginal Posterior Density, SMC, Equilibrium Initials, Full Information

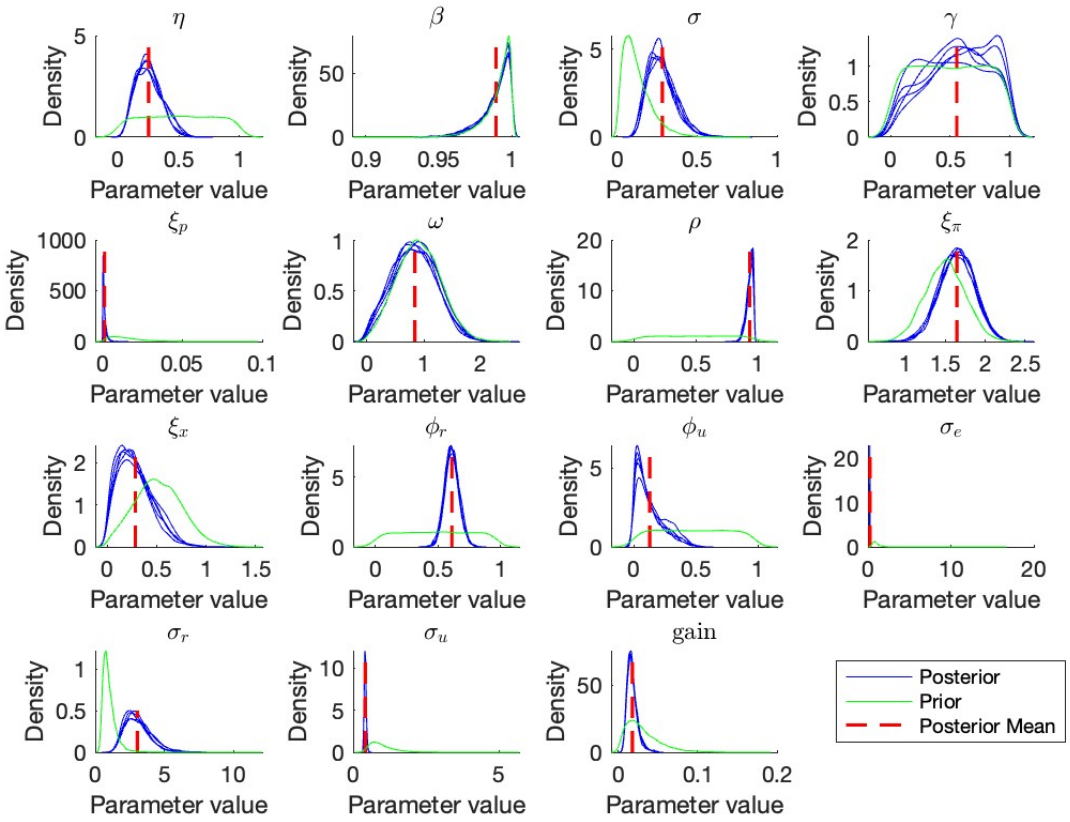

Figure 3: Marginal Posterior Density, SMC, Equilibrium Initials, Limited Information

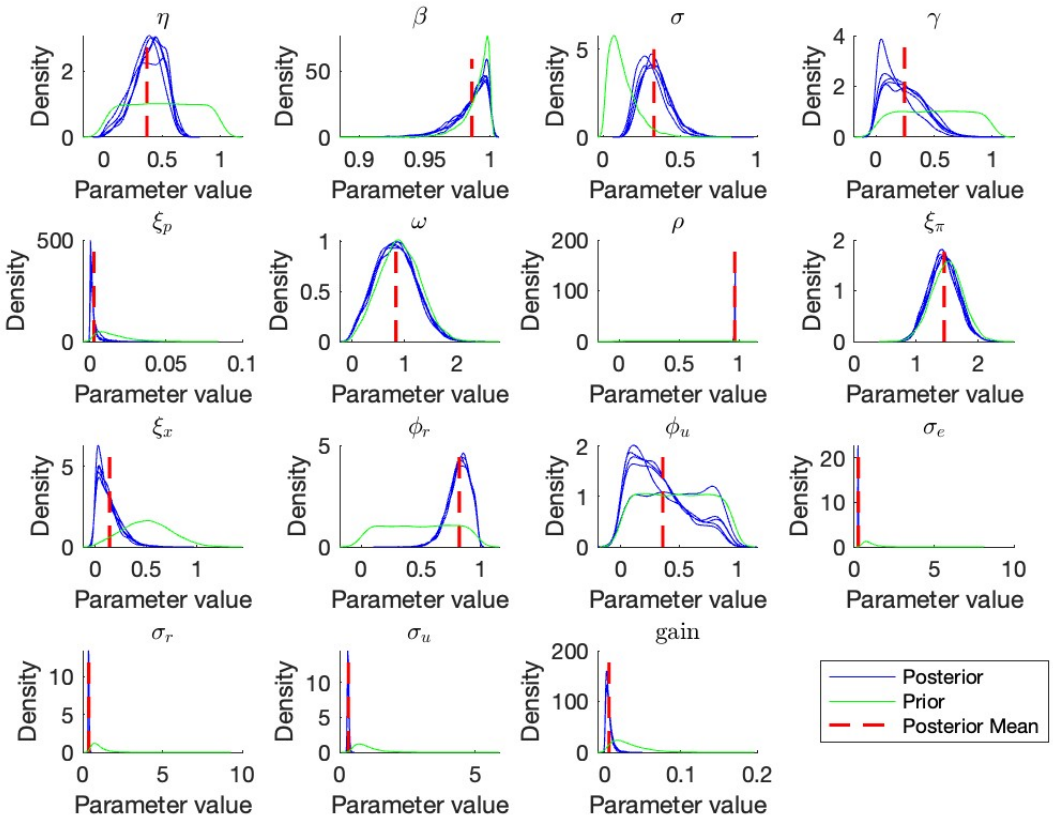

Figure 4: Marginal Posterior Density, SMC, Training Sample Initials, Full Information

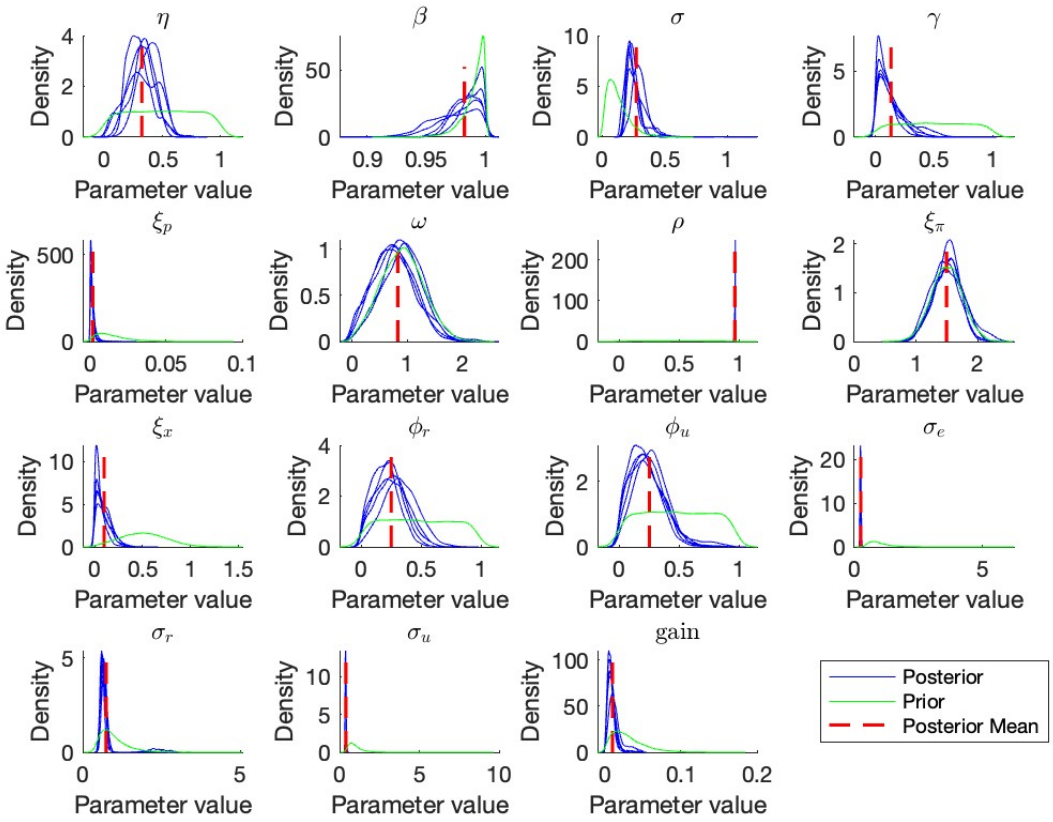

Figure 5: Marginal Posterior Density, SMC, Training Sample Initials, Limited Information

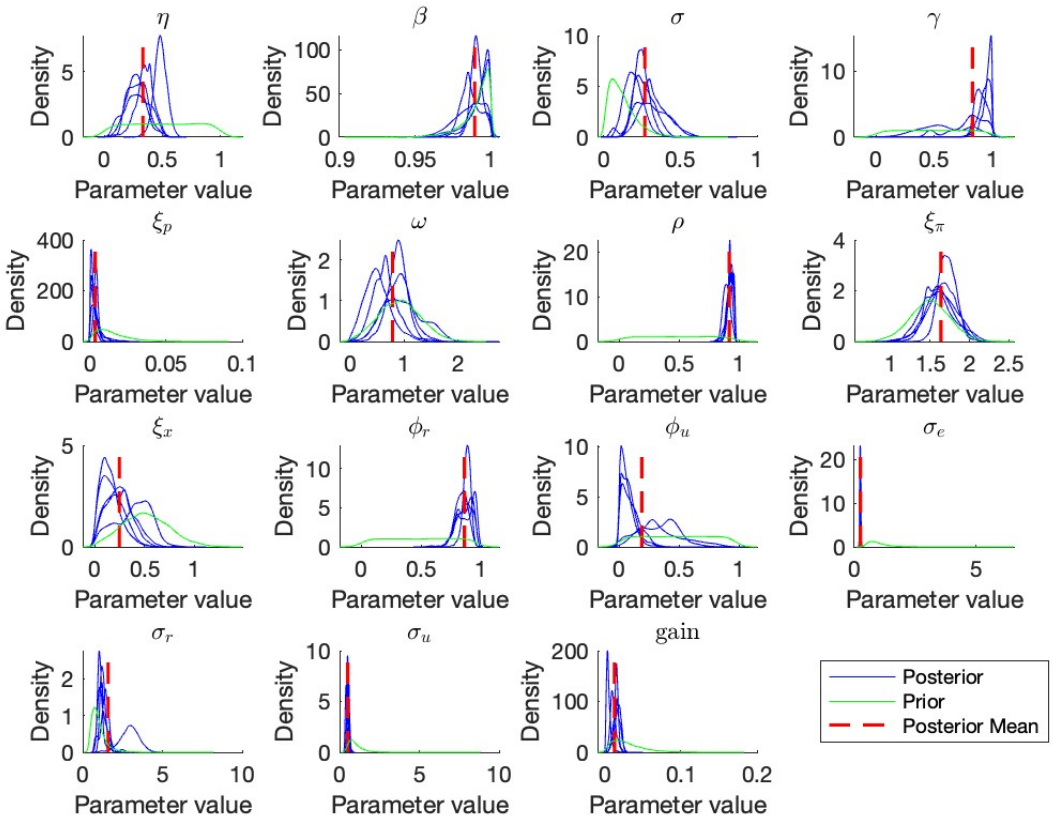

Figure 6: Marginal Posterior Density, SMC, Jointly Estimated Initials, Full Information

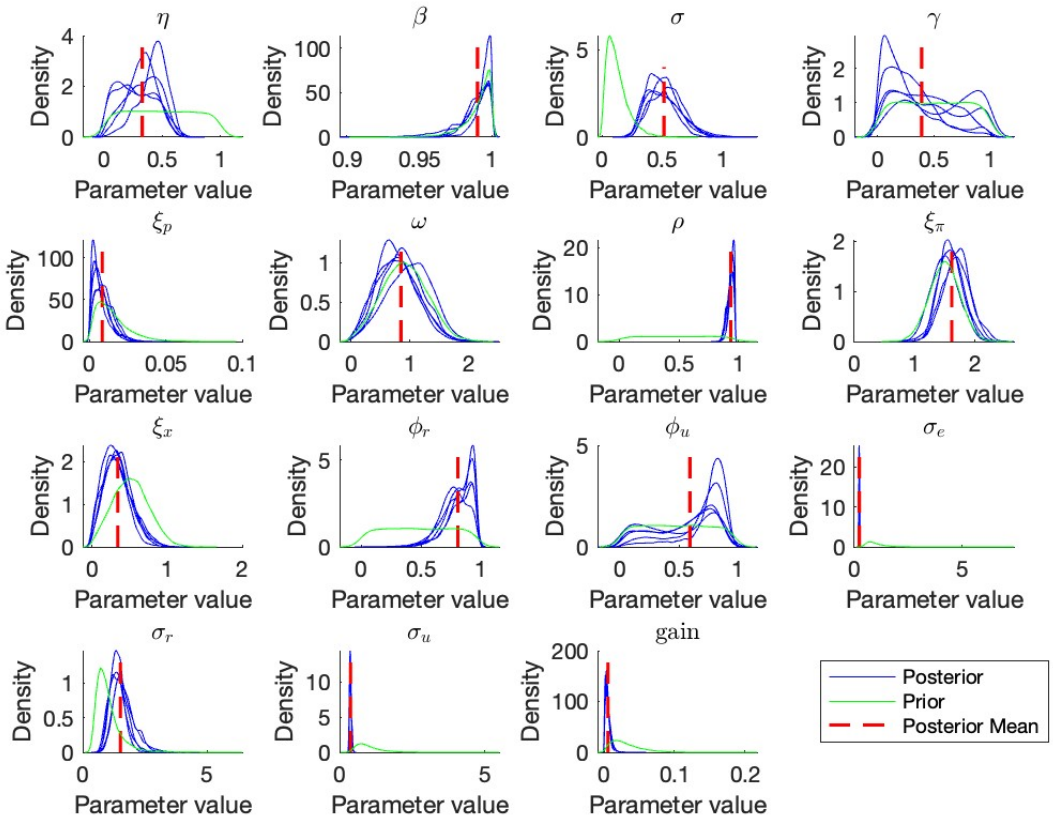

Figure 7: Marginal Posterior Density, SMC, Jointly Estimated Initials, Limited Information

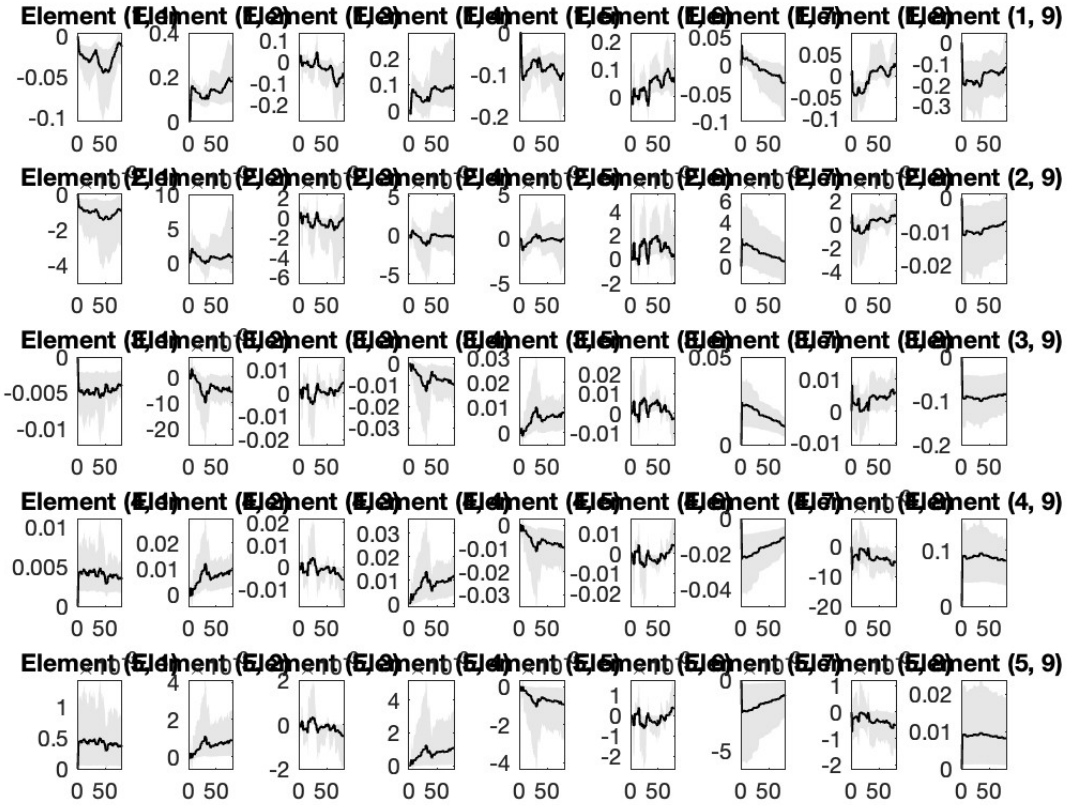

Figure 8: Beliefs Evolution, Equilibrium Inits, Full Info

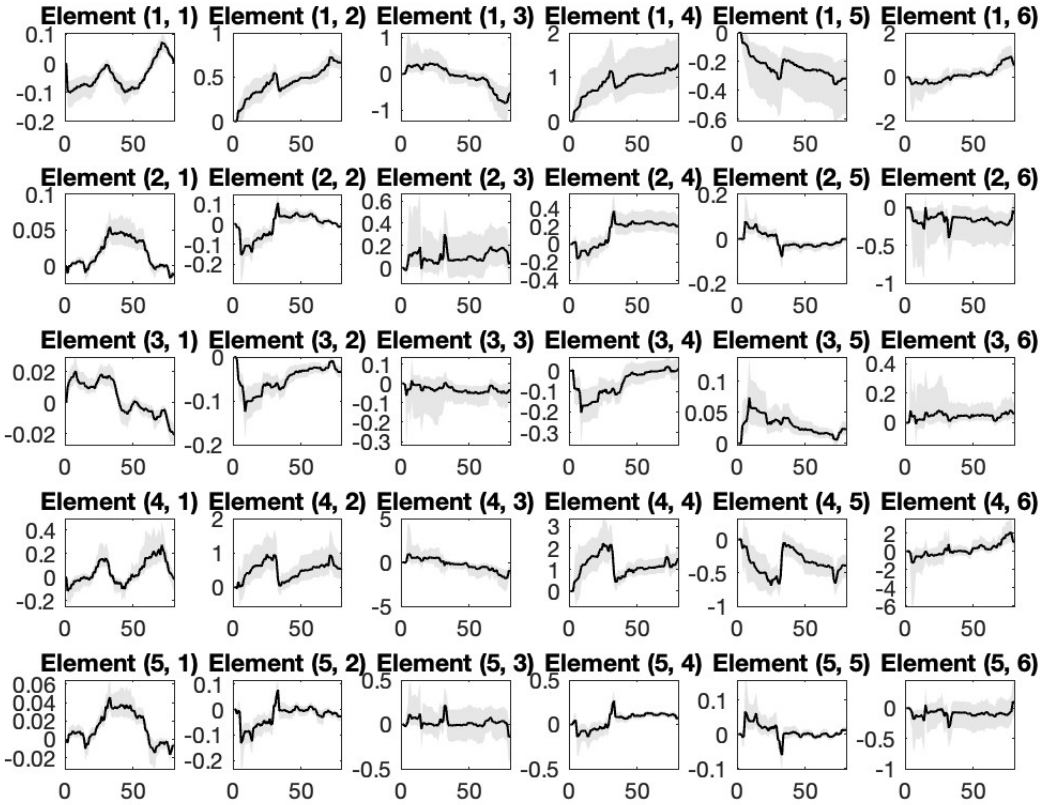

Figure 9: Beliefs Evolution, Equilibrium Inits, Limited Info

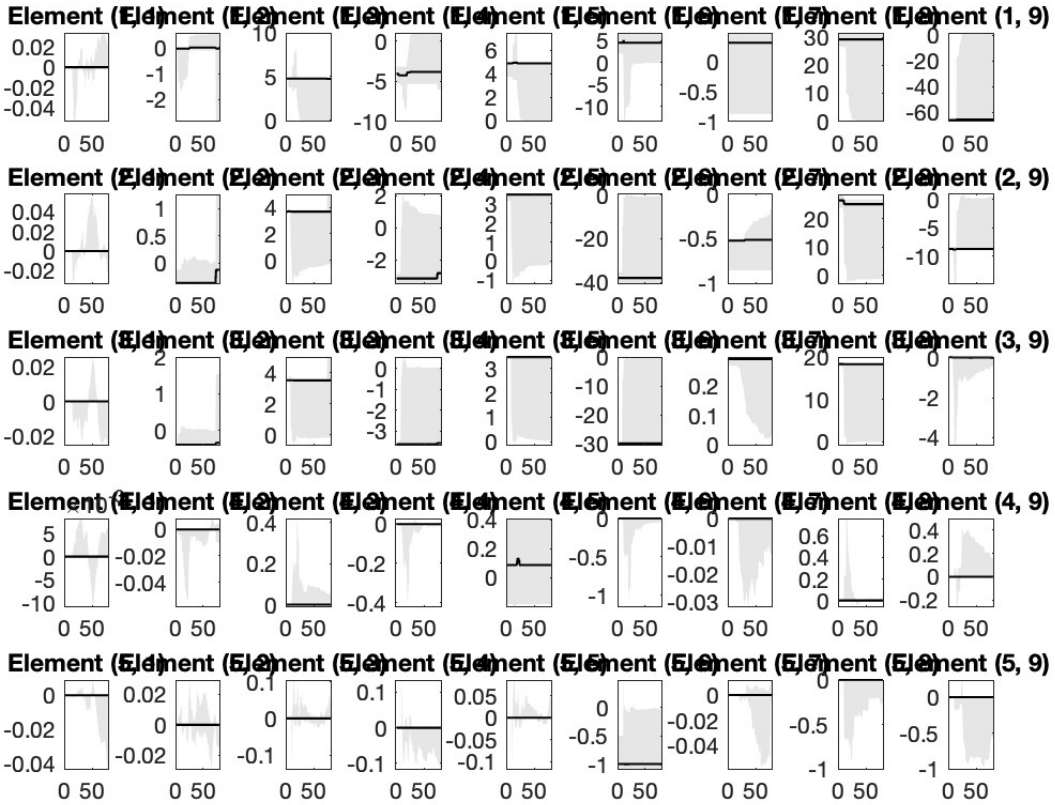

Figure 10: Beliefs Evolution, Training Sample Inits, Full Info

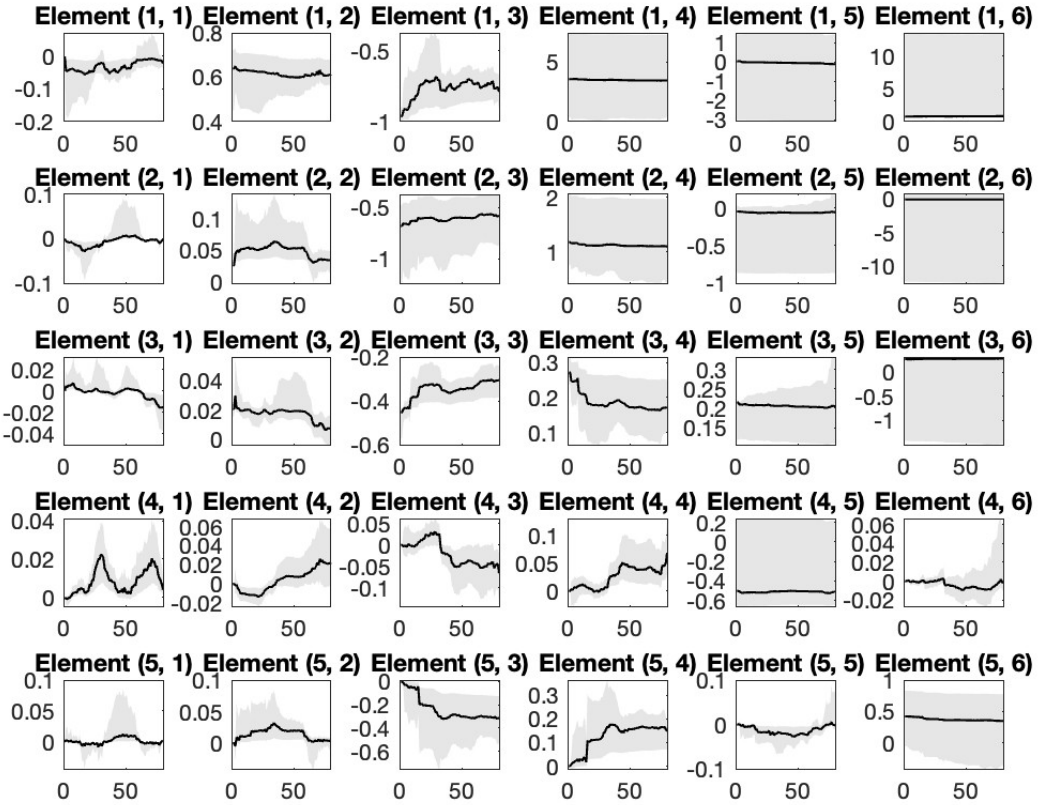

Figure 11: Beliefs Evolution, Training Sample, Limited Info

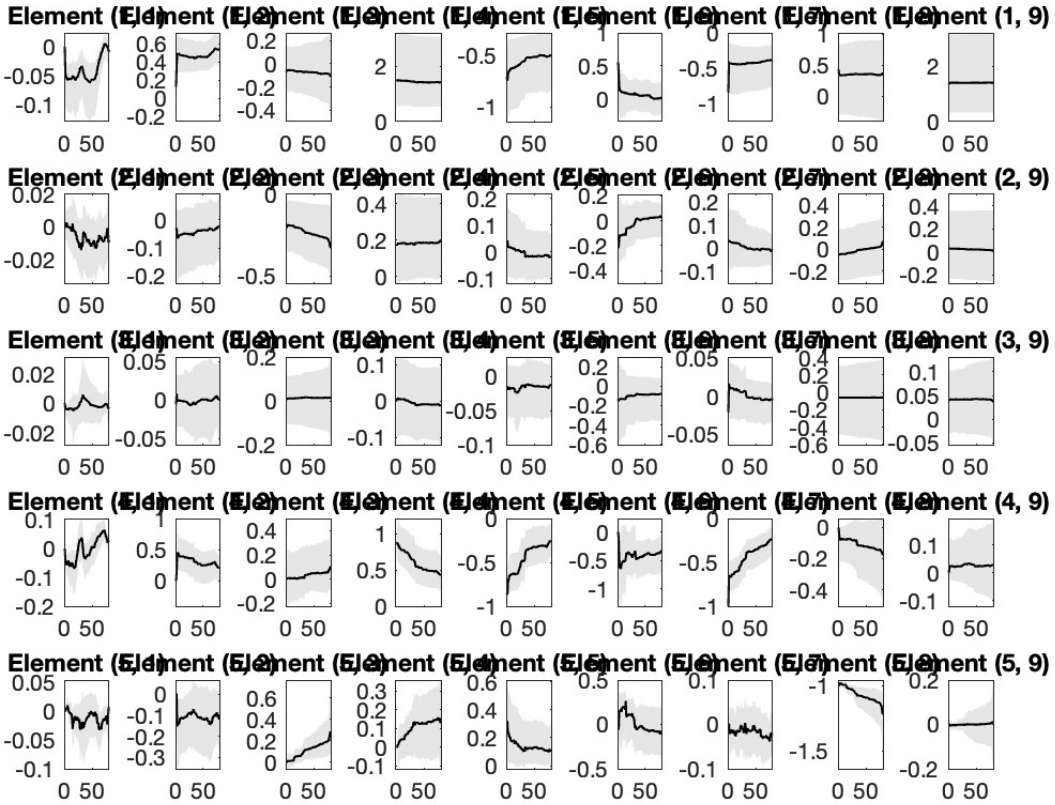

Figure 12: Beliefs Evolution, Jointly Estimated Inits, Full Info

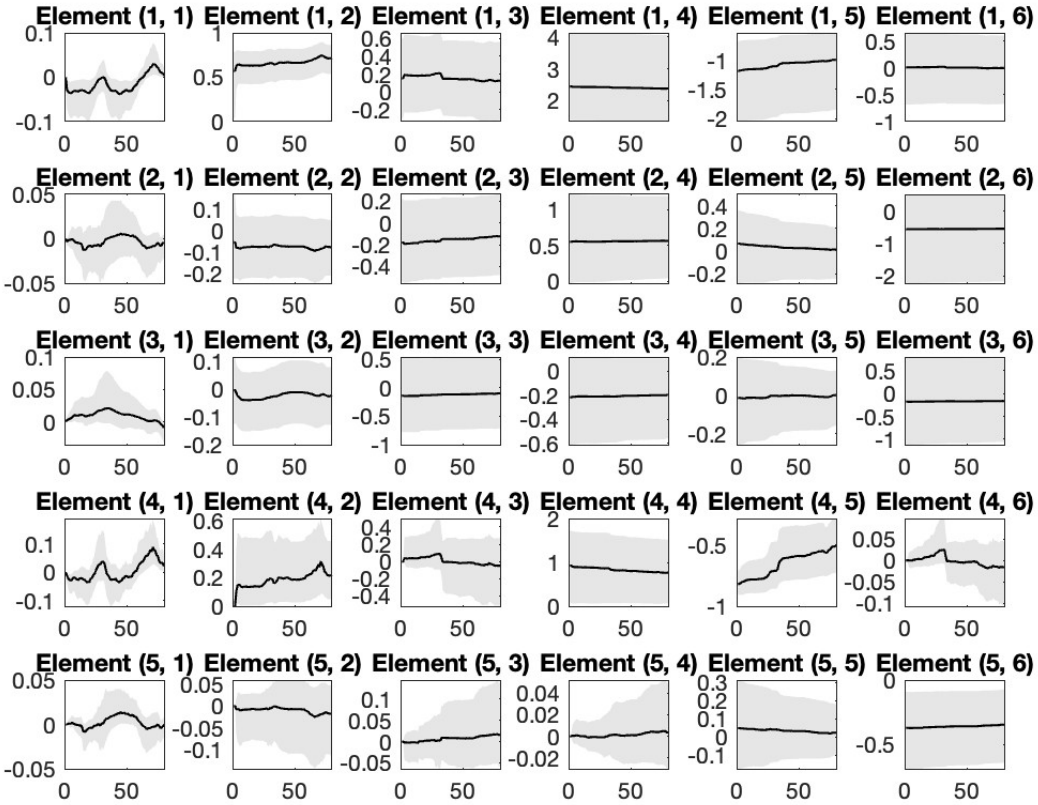

Figure 13: Beliefs Evolution, Jointly Estimated Inits, Limited Info