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

Jacob Thompson

University of Oregon

December 01, 2023

Motivation

- Most DSGE modeling assumes Rational Expectations (RE)
- Alternative literature allows agents to use linear models to form expectations
  - Models updated using Adaptive Learning (AL) algorithms
- DSGE Models very commonly estimated with Bayesian econometric methods
  - Authors starting with Milani (2005) applied this to AL models
  - Bayesian methods offers simple ranking of model performance according to *in-sample* data fit based upon priors

Motivation and Research Question

- Estimated DSGE models that relax RE find consistent improvements in model performance over RE baseline
- Beliefs in AL models are updated *recursively*
- Recursive structure raises question: How ought one choose *initial* beliefs?
- Very little work on initial beliefs
  - Papers including Milani (2005), Slobodyan and Wouters (2012) often initialize beliefs around RE solution
  - Some including Milani (2014) initialize using training sample
- Some simulation work using GMM estimation from Berardi and Galimberti (2017)
- No systematic Bayesian evaluation of initial beliefs

Motivation and Research Methods

- To investigate this question, I estimate a 5-equation New Keynesian DSGE model with mechanical lags in the Inflation and the Output Gap processes
- The set of estimated models includes:
  - Two information sets
  - Three initialization schemes for a total of six estimated AL models plus one estimated RE model
- I newer Bayesian algorithm called *Sequential Monte Carlo* (SMC) instead of more common Markov Chain Monte Carlo (MCMC).

Important Results

- Joint Estimation provides the highest marginal data density
- Estimated Inflation indexation and habit persistence are reduced but not eliminated
- Equilibrium-based initial beliefs, a common technique, still improves upon RE baseline

Literature Review: Estimated AL DSGE models

- Seminal work by Milani (2005) who estimates a small-scale New Keynesian DSGE model with lags in inflation and output.
- I estimate this same model
- Important results include:
  - Near-disappearance of lags in inflation and output processes.
  - Improvement in marginal data density over RE baseline.
- Both result from updating of beliefs over time
  - AL allows agents to update beliefs
  - RE only allows agents to update beliefs if the model itself updates

Literature Review: Estimated AL DSGE models

- Similar results obtained by Slobodyan and Wouters (2012)

- Authors estimate a medium-scale DSGE model based on model of Smets and Wouters (2007)
  - Smets Wouters model includes 12 forecasted variables and 36 variables
  - Many frictions including price and wage-stickiness, capital formation, variable capital utilization, etc

Literature Review: My Contribution

- Can DSGE modelers improve model fit with the right choice of initial beliefs?

- Initial beliefs in estimated DSGE models usually chosen based upon pre-sample data or upon the RE solution implied by model parameters

- No comparison of model fit between two choices in the literature

- No Bayesian DSGE model with initial beliefs estimated jointly with model parameters in the literature

Literature Review: My Contribution

- I provide a tractable procedure to jointly estimate initial beliefs that improves further model fit within a Bayesian framework

- Better fitting DSGE models may make better forecasting models

Overview of the Model

- I estimate a 5 equation New Keynesian Model:
  - New Keynesian IS Curve relating Output and Interest Rates with an output shock process (household consumption problem)
  - New Keynesian Phillips Curve relating Inflation and Output with a natural interest rate shock process (firm price-setting problem)
  - Taylor Rule Monetary Policy
  - Autoregressive natural interest rate
  - Autoregressive output shock
- Woodford (2003) provides a detailed derivation, which I will recount briefly

The Model: Optimal Consumption and the IS Curve

- Each i-th member of a continuum of households uniformly distributed along [0,1] maximizes the expected discounted sum of within-period utilities:

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

- Each period household only decides how much to work
- Driven by aggregate preference shocks that increase marginal utility of consumption and reduce marginal disutility of work
- $0 \leq \eta \leq 1$ is the degree of persistence in consumption

The Model: Optimal Consumption and the IS Curve

- After solving for the first-order conditions and log-linearizing around the steady state, we obtain the household's Euler Equation

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

The Model: Optimal Consumption and the IS Curve

where

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

and the circumflex operator ˆ denotes log deviations from the steady-state value.

- $\sigma \equiv \frac{U_c}{\bar{C} U_{cc}} > 0$ is the intertemporal elasticity of substitution of consumption
- All output is consumed so $C_t = X_t$
- First linear equation of the model follows: the New Keynesian IS Curve:

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

where

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

The Model: Optimal Price Setting and the Phillips Curve

- Model assumes a continuum of monopolistically competitive firms who adjust prices as in Calvo (1983).

- $0 < 1 - \alpha < 1$ of firms adjust their price $p_j(t)$ according to

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

- $0 \leq \gamma \leq 1$ measures the degree of indexation

The Model: Optimal Price Setting and the Phillips Curve

- Firms face a common demand curve $y_t(i) = Y_t \left(\frac{p_t(i)}{P_t}\right)^{-\theta}$
- Aggregate output $Y_t = \left[\int_0^1 y_t(i)^{\frac{\theta-1}{\theta}} di\right]^{\frac{\theta}{\theta-1}}$
- P_t is the aggregate price index.
- Price-adjusting firms set a common $p_t(i) = p_t^*$.
- Thus the aggregate price level follows the process:

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

The Model: Optimal Price Setting and the Phillips Curve

- Firms maximize the present-discounted sum of future profits:

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

- $Q_{t,T} = \beta^{T-t} \left(\frac{P_t}{P_T} \right) \left(\frac{\lambda_T}{\lambda_t} \right)$ is a stochastic discount factor.

- $\Pi_T(\cdot)$ denotes period-T nominal profits.

The Model: Optimal Price Setting and the Phillips Curve

- Nominal profits are:

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

The Model: Optimal Price Setting and the Phillips Curve

- Log linearization of FOCs yields the New Keynesian Phillips Curve:

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

wherein

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

- $\alpha, \beta, \theta, \omega$ are the Calvo, discount, disutility of work, and substitution of consumption parameters respectively

The Model: Monetary Policy and Autoregressive shocks

- Monetary Policy is assumed to follow a Taylor Rule:

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

- Shocks follow univariate AR(1) processes:

$$\begin{aligned} r_t^n &= \phi^r r_{t-1}^n + v_t^r \quad \text{(Natural Interest Rate process)} \\ u_t &= \phi^u u_{t-1} + v_t^u \quad \text{(Productivity shock process)} \end{aligned}$$

The Model: Five Equations plus two definitions

$$
\begin{aligned}
&\tilde{x}_t = \hat{\mathbb{E}}_t \tilde{x}_{t+1} - (1-\beta\eta)\sigma[i_t - \hat{\mathbb{E}}_t \pi_{t+1} - r_t^n] && \text{(NK IS Curve)} \\
&\tilde{\pi}_t = \xi_p[\omega \tilde{x}_t + [(1-\eta\beta)\sigma]^{-1}\tilde{x}_t] + \beta \hat{\mathbb{E}}_t \tilde{\pi}_{t+1} + u_t && \text{(NK Phillips Curve)} \\
&i_t = \rho i_{t-1} + (1-\rho)(\psi_\pi \tilde{\pi}_t + \psi_x \tilde{x}_t) + \varepsilon_t && \text{(Taylor Rule)} \\
&r_t^n = \phi^r r_{t-1}^n + v_t^r && \text{(Natural Interest Rate)} \\
&u_t = \phi^u u_{t-1} + v_t^u && \text{(Productivity shock)}
\end{aligned}
$$

The Model: Five Equations plus two definitions

$$\tilde{\pi}_t \equiv \pi_t - \gamma \pi_{t-1} \quad \text{(Inflation Indexation)}$$
$$\tilde{x}_t \equiv (x_t - \eta x_{t-1}) - \beta \eta \hat{\mathbb{E}}_t (x_{t+1} - \eta x_t) \quad \text{(Habit Persistence)}$$

How Learning is incorporated into the model

- Learning is incorporated through the expectations operator $\hat{\mathbb{E}}_t$.
- Instead of using the RE solution, I substitute agents' expectations formed from small, linear forecasting models
- Setup generally called *Euler Equation Learning* in AL literature
- These are VAR(1) models of the form:

Incorporating Learning into the Model

$$\begin{pmatrix} x_t \\ \pi_t \\ i_t \\ r_t^n \\ u_t \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ 0 & 0 & 0 & b_{44} & 0 \\ 0 & 0 & 0 & 0 & b_{55} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ \pi_{t-1} \\ i_{t-1} \\ r_{t-1}^n \\ u_{t-1} \end{pmatrix}$$

$$+ \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \epsilon_t^i \\ \epsilon_t^{r^n} \\ \epsilon_t^u \end{pmatrix}$$

or, more compactly

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

Incorporating Learning into the Model

- Coefficients are updated via recursive least squaures.
- Agents use the simpler linear model to update coefficients.

$$Z_t = \begin{pmatrix} x_t \\ \pi_t \\ i_t \\ r^n_t \\ u_t \end{pmatrix} = \underbrace{\Phi}_{5\times 8} \begin{pmatrix} 1 \\ x_{t-1} \\ \pi_{t-1} \\ i_{t-1} \\ r^n_{t-1} \\ u_{t-1} \\ \epsilon^i_t \\ \epsilon^{r^n}_t \\ \epsilon^u_t \end{pmatrix} = \Phi \times X_t$$

Incorporating Learning into the Model

- Coefficients updated according to *Constant Gain* recursive least squares formulae.

$$\hat{\phi}_t = \hat{\phi}_{t-1} + \bar{g} R_t^{-1} X_t (Z_t - \hat{\phi}'_{t-1} X_t)'$$
$$R_t = R_{t-1} + \bar{g}(X_t X_t' - R_{t-1})$$

- R_t is $E(X_t X_t')$.
- $\bar{g}$ is the "learning gain" parameter and is estimated with parameters.
- Constant gain weighs more recent observations more heavily and allows for faster updating of beliefs compared to decreasing gain least-squares

Incorporating Learning into the Model

- The recursive structure of agents' updating procedure raises the question of initial beliefs
- Initial beliefs are choices of $\hat{\phi}_0, R_0$
- I investigate three methods of choosing $\hat{\phi}_0, R_0$:
  - Equilibrium based initials
  - Training-Sample based initials
  - Jointly-Estimated initials
- I also estimate each choice of initial belief under a complete and incomplete information or observation set
  - Under complete information, agents observe contemporaenous i.i.d. shocks $(\epsilon_t^i, \epsilon_t^{r^n}, \epsilon_t^u)'$ and lagged endogenous variables $(x_{t-1}, \pi_{t-1}, i_{t-1}, r_{t-1}^n, u_{t-1})'$

Initial Beliefs: Equilibrium Based Initial Beliefs

- Under RE and AL, estimation involves drawing randomly thousands of different vectors of possible values for the true parameter vector.

- For each parameter draw θ_i for which there exists a unique RE solution, I compute the stationary VAR(1) representation of the model and use the elements from that equation as the elements of my $\hat{\phi}_0$ matrix

- One can also compute the stationary variance-covariance matrix since θ_i also includes the variance of the shock processes.

- I substitute the elements of this variance-covariance matrix into my R_0 matrix.

Initial Beliefs: Training-Sample based Initial Beliefs

- Using 40 quarters of pre-sample data, I estimate via Maximum Likelihood a state-space model with the following measurement and transition equations:

$$\text{Observation:} \quad \begin{pmatrix} x_t \\ \pi_t \\ i_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_t \\ \pi_t \\ i_t \\ r_t^n \\ u_t \end{pmatrix}$$

Initial Beliefs: Training-Sample based Initial Beliefs

$$\text{State:} \quad \begin{pmatrix} x_{t+1} \\ \pi_{t+1} \\ i_{t+1} \\ r^n_{t+1} \\ u_{t+1} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ 0 & 0 & 0 & a_{44} & 0 \\ 0 & 0 & 0 & 0 & a_{55} \end{pmatrix} \begin{pmatrix} x_t \\ \pi_t \\ i_t \\ r^n_t \\ u_t \end{pmatrix}$$

$$+ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \end{pmatrix}.$$

Initial Beliefs: Training-Sample based Initial Beliefs

- I also estimate the variances of the shock processes $\eta_{1,t}, \eta_{2,t}, \eta_{3,t}$
- After obtaining MLE estimates I substitute the elements of matrix $\widehat{\mathbf{a}}_{\mathsf{MLE}}$ into ϕ_0 for each parameter draw θ_i.
- Compute variance-covariance matrix of the state equation and substitue the elements of that matrix into R_0 matrix
- VERY important for interpreting final results: ϕ_0, R_0 do not vary across parameter draws θ_i.

Initial Beliefs: Jointly-Estimated Initial Beliefs

- Joint estimation treats each element of the matrix **a** and the variances of the shock processes $\eta_{1,t}, \eta_{2,t}, \eta_{3,t}$ as estimated parameters

- When jointly estimating initial beliefs, draw augmented parameter vector $(\theta_i', vec(\hat{\phi}_0))'$

- Avoid estimating elements of R_0 to reduce the number of estimated parameters.

- Procedure equires explicit prior distribution for $vec(\hat{\phi}_0))$.

Initial Beliefs: Jointly-Estimated Initial Beliefs

- I seek a prior distribution that is informed by the other estimated models
- Use draws from the simulated posterior under RE and saved $vec(\hat{\phi}_0))$ for each θ_i and fitted a multivariate normal distribution to this vector
- This multivariate normal serves as the prior distribution for elements of $vec(\hat{\phi}_0))$
- For elements which have sample variances less than 1, I set their prior variances equal to 1
  - I set variance equal to one so as not to increase prior density
- I also estimate agents' perceived variances $\eta_{1,t}, \eta_{2,t}, \eta_{3,t}$.
  - use IG prior with mean 1, variance .5

Bayesian Estimation

- To match common conventions in DSGE literature I use the following prior distributions for the structural 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]
gain	Learning Gain	BETA[.031, .022]

Bayesian Estimation

- Bayesian Estimation of DSGE models entails simulating through numerical methods a posterior distribution of parameters
- Simulation of the posterior distribution only requires computation of the prior density and the likelihood function
- The model is linear and the i.i.d. shocks are normal
- I compute the likelihood function through a Kalman Filter prediction-error variance decomposition.

Bayesian Estimation

- Metropolis Hastings Random Walk (MHRW) algorithm most common algorithm
  - Used by Milani (2005), Slobodyan and Wouters (2012), Smets and Wouters (2007), Giannoni and Woodford (2004), and many, many others.
  - Included in Dynare package.
- I depart from this convention owing to the irregular posterior exhibited by DSGE models with AL agents.

Bayesian Estimation (Cont.)

- I use Sequential Monte Carlo (SMC) algorithm for estimation
  - Described in detail by Herbst and Schorfheide (2013)
  - Code based on supplement to Herbst and Schorfheide (2016)
- Two advantages of SMC for the present study:
  - Efficient simulation of irregular posterior distributions.
  - Highly parallelizable.

Bayesian Estimation: Generic SMC Algorithm

Algorithm 1 Generic SMC with Likelihood Tempering - Part 1

1: **Initialization.** ($\phi_0 = 0$). Draw $\theta_1^i \sim p(\theta)$ i.i.d. and set $W_1^i = 1$, $i = 1, \ldots, N$.

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

1. **Correction.** Incremental weights:

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

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

Bayesian Estimation: Generic SMC Algorithm

Algorithm 2 Generic SMC with Likelihood Tempering - Part 2

1: **Recursion (Contd.).**

2 **Selection.**

1 If $\rho_n = 1$: Resample $\{\hat{\theta}\}_i^N$ via multinomial resampling.

2 If $\rho_n = 0$: Set $\hat{\theta}_n^i = \theta_{n-1}^i$ and $W_n^i = \tilde{W}_n^i$.

3 **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 $\pi_n(\theta)$.

Advantages of the SMC method over MHRW

- Single-chained MHRW cannot sample multimodal posterior
  - Can be solved by random blocking
- SMC particles initially distributed across prior distribution, all modes likely to have some mass of initial particles
- Single chained MHRW chain not parallelizable; no benefit to greater computing power
- SMC highly parallelizable
  - This allows one to take advantage of high performance computing.
  - Results in present paper obtained using EC2 instance with 96 physical CPU cores
- posterior simulated by normalized weights

Data Description

- I use FRED data for the output gap, inflation, and the federal funds rate from 1982-2002 for estimating deep parameters
- I use de-meaned values for the inflation rate and the federal funds rate
  - computation of model moments imply that observed inflation and federal funds rates are zero
  - Federal reserve has positive inflation target
  - ZLB means federal funds rate also cannot have zero mean
- Output Gap plausibly zero mean process with different observed mean.
- Could follow An and Schorfheide (2007) and add mean inflation and federal funds rates as estimated parameters.

Results: Model Comparison

- I compare the model fit according to the estimated marginal data density
- The marginal data density average likelihood function weighted by the prior distribution
- Marginal data density gives the "unconditional" likelihood of data appearing, given the model
- Bayes Factor the ratio of the marginal data densities between two models; Bayesian analogue to likelihood ratio teset

Results: model comparison, short data series

Table 1: 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)

Results: model comparison

- Under full information, Equilibrium Initial beliefs have Bayes Factor of 10 over RE baseline.

- Under Full and Limited information, Jointly Estimated Initial Beliefs have Bayes Factor of 18 and 315 over RE baseline respectively.

Results: model comparison

- Training Sample initial beliefs perform very poorly because beliefs are not allowed to vary across parameter draws.
- Beliefs are also not able to update because of projection facility
- One does not want interesting results to be driven by projection facility “hits”, so researchers often penalize such hits in the likelihood function.
    - I follow this convention by reducing the log posterior value by 10 for each projection facility hit.

Results: model comparison, long data series

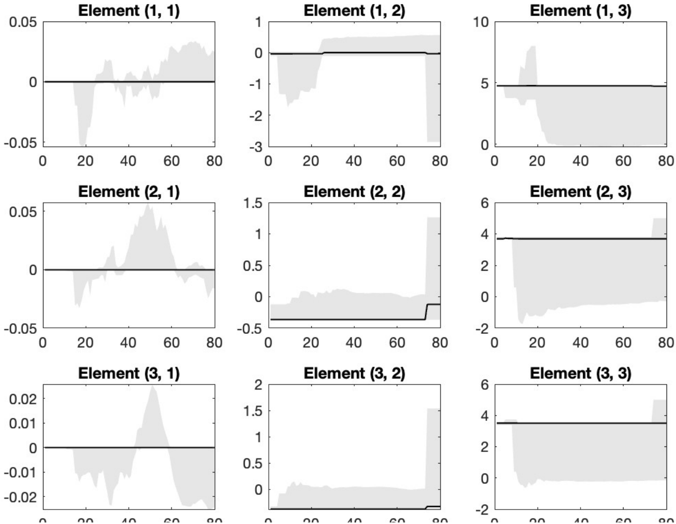
Table 2: Mean and Standard Deviation of Natural Logarithms of the Marginal Likelihoods, 1961-2006 data

	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)

Results: model comparison, long data series

- Jointly estimated initial beliefs still performed the best but not by much.
  - Bayes Factor of Equilibrium based and Jointly Estimated initial beliefs of 217 and 229 over RE baseline respectively.
- Training sample beliefs still did poorly, results uninformative

Evolution of Beliefs: Training Sample Initial Beliefs



Results: Parameter Estimates

- I report estimates based on 1982-2002 data.
- Able to replicate reduction but not elimination of mechanical persistence
  - mechanical persistence parameters statistically greater than zero under all estimated models.
- Would seem to contradict findings in Milani (2005).
- DSGE-VAR model by Cole and Milani (2019) estimates same 5-equation NK-DSGE model with habit persistence and inflation indexation with AL but fails to eliminate mechanical persistence.

Results: Parameter estimates (Part 1)

Table 3: Summary of Parameter Estimates for Different Models (Part 1)

Model Description	**Parameter Estimates $\bar{\mu}$, 5%, 95%)**
Rational Expectations	η: 0.67 (0.52, 0.78) γ: 0.97 (0.91, 1.00) gain: N/A
Equilibrium Initials, Full Info	η: 0.54 (0.33, 0.72) γ: 0.96 (0.88, 1.00) gain: 0.0218 (0.0049, 0.0484)
Equilibrium Initials, Limited Info	η: 0.26 (0.10, 0.45) γ: 0.59 (0.10, 0.97) gain: 0.0176 (0.0088, 0.0313)

Results: Parameter estimates (Part 2)

Table 4: Summary of Parameter Estimates for Different Models (Part 2)

Model Description	**Parameter Estimates $\bar{\mu}$, (5%, 95%)**
Training Sample Initials, Full Info	η: 0.40 (0.14, 0.61) γ: 0.23 (0.02, 0.53) gain: 0.0061 (0.0012, 0.0153)
Training Sample Initials, Limited Info	η: 0.24 (0.09, 0.42) γ: 0.15 (0.01, 0.38) gain: 0.0106 (0.0025, 0.0212)
Jointly Estimated Initials, Full Info	η: 0.28 (0.13, 0.42) γ: 0.93 (0.77, 1.00) gain: 0.0145 (0.0079, 0.0256)
Jointly Estimated Initials, Limited Info	η: 0.39 (0.12, 0.66) γ: 0.37 (0.02, 0.94) gain: 0.0062 (0.0011, 0.0138)

Summary of parameter estimates

- Mechanical persistence parameters usually but not always reduced compared to RE baseline

- Somewhat low estimated learning gain value
  - Possibly due to post-1980 data. Milani (2005) also finds very low estimated learning gain when estimating model based on post-1980 data. Will have to check parameter estimates under longer/alternative data sets

Implications of model comparison and Conclusions

- Joint Estimation can improve model fit substantially even above equilibrium initials but at the cost of impoverishing agents' forecasting model
  - Improved model fit required use of incomplete forecasting model: full forecasting model suffers
  - Result comports with Slobodyan and Wouters (2012) who find that models wherein agents use AR(1) forecasting models perform better than models wherein agents use VAR models.
- Equilibrium initials can still improve over RE baseline if agents use rich forecasting models

Implications of model comparison and Conclusions

- I Propose the following rule of for choosing initial beliefs when estimating DSGE models with AL: For models wherein agents use rich forecasting models, use equilibrium-based initials. For models wherein agents use small or very incomplete forecasting models, use joint estimation.
  - Rich DSGE models needn't assume rich forecasters, nor impoverished DSGE models impoverished forecasters.