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

Jacob Thompson

University of Oregon

October, 27, 2023

Introduction

- Overview of Presentation
  - Explanation of Rational Expectations and how it contrasts with Adaptive Learning Framework
    - Explanation of my own Adaptive Learning algorithm
- Review Estimation of DSGE models
  - Review previous empirical work in Adaptive Learning literature
  - Introduce Sequential Monte Carlo method
- Results: Marginal likelihoods and important parameter estimates
  - Equilibrium-based initials
  - Training-sample based initials
  - Jointly estimated initials

The Theory of Adaptive Learning

- Detailed in (Evans and Honkapohja 2001)
- Suppose there is some function that maps from agents beliefs, ϕ, to the behavior of the economy, $T(\phi)$
- The fixed point ϕ^* where $T(\phi^*) = \phi^*$, if it exists, is the rational expectations solution
  - Recall the method of undetermined coefficients

The Theory of Adaptive Learning

- Let Z_t be some vector of forecasted variables, X_t be some vector of regressors, in our model $(1, Z_{t-1})$ or $(1, Z_{t-1}, \varepsilon_t)$ so that agents have the model $Z_t = \phi' X_t$.
- Agents update their beliefs ϕ_t according to the recursive least-squares formulae.

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

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

- where Σ_t is $E(X_t X_t')$

Adaptive Learning

- Empirical work, including that presented today, replaces $\frac{1}{t}$ with a constant $\bar{g}$
- This weighs more recent observations more heavily and allows beliefs to change more rapidly
- Especially useful in models of regime-switching

Asymptotic Properties of AL algorithms

- Letting τ be notional time
- *E-stability principle*
  - A fixed-point ϕ^* of the T-map $T(\phi)$ is expectationally stable if the differential equation $\frac{d\phi}{d\tau} = T(\phi) - \phi$ is asymptotically stable
- (Marcet and Sargent 1989) show that for expectationally stable rational expectations equilibria with a suitable projection facility, agents' beliefs will converge with probability of one to the rational expectations solution
- Empirical work leans on this result for simulation and estimation of DSGE models
  - Without a projection facility, beliefs become explosive leading to undefined likelihood values, confounding estimation

Bayesian Estimation of DSGE Models

- Fundamental object of interest is Posterior distribution of parameters, or $p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}$
  - $p(Y|\theta)$ is the likelihood function
  - $p(\theta)$ is the prior density
  - $p(Y)$ is the marginal likelihood or model evidence, and is used to compare the evidence in favor of one model or another
  - Since $p(Y)$ does not depend on θ, one only needs to compute $p(Y|\theta)p(\theta)$ to simulate the posterior distribution since $p(Y|\theta)p(\theta) \propto \frac{p(Y|\theta)p(\theta)}{p(Y)}$

Bayesian Estimation of DSGE Models

- Two primary concerns in present review
  - Computation of the likelihood function
  - Simulating the Posterior distribution
- Since the model is conditionally linear, I compute the likelihood function using a kalman filter
- As the posterior function is fairly irregular, I use Sequential Monte Carlo rather than Markov Chain Monte Carlo to simulate the posterior

Bayesian Estimation of DSGE models

- The Likelihood Function is computed using a Kalman Filter
- For any parameter draw θ_i for which there is a unique solution to the DSGE model, there is a state-space representation of the following form

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, H_t)$$

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t, \eta_t \sim \mathcal{N}(0, Q_t)$$

- y_t is the vector of observables
- α_t is the vector of latent variables
- $Z_t, d_t, T_t, R_t, H_t, Q_t$ are the system matrices and are all functions of θ_i, y_t

Bayesian Estimation of DSGE models

- let $a_t = E(\alpha_t | y^t)$ be the optimal forecast of the state given information through time t
- $P_t = E((a_t - \alpha_t)(a_t - \alpha_t)' | y^t)$ be the variance of that optimal forecast. The Kalman filter consists of prediction and updating equations.

$$a_{t|t-1} = T_t a_{t-1} + c_t$$

$$P_{t|t-1} = Z_t P_{t-1} Z_t' + R_t Q_t R_t'$$

Bayesian Estimation of DSGE models

- We get the optimal predictor of y_t based on the information set $\{y_t\}_0^{t-1}$ and the prediction error variance

$$y_{t|t-1} = Z_t a_{t|t-1} + d_t$$

$$v_t = y_t - y_{t|t-1}$$

$$E(v_t v_t') = F_t = Z_t P_{t|t-1} Z_t' + H_t$$

Bayesian Estimation of DSGE models

- From these we get the updating equations for a_t, P_t

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

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

- and, because the prediction errors follow a multivariate normal distribution, the likelihood function

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

Bayesian Estimation of DSGE models

- Once one has a likelihood function and a prior density one must choose a sampler algorithm in order to simulate the posterior density
- The most commonly used sampler is the Metropolis Hastings Random Walk Algorithm
- Because of time constraints and irregularity of the Posterior density of models with Adaptive Learning, I opt instead for a Sequential Monte Carlo sampler.

Sequential Monte Carlo Methods

- Sequential Monte Carlo methods are a newer class of Bayesian estimation methods of DSGE models
- Superficially similar to running thousands of MHRW chains with starting points distributed throughout the prior distribution
- Far more efficient for sampling multi-modal posterior distributions: sampler does not get “stuck” in local maxima.
- SMC Highly Parallelizable and can take advantage of HPC
- Estimates I present today were obtained using an AWS EC2 instance with 96 physical CPU cores

Sequential Monte Carlo Algorithm

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

Sequential Monte Carlo Algorithm (cont'd)

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:** If $\rho_n = 1$, resample the particles via multinomial resampling. If $\rho_n = 0$, let $\hat{\theta}_n^i = \theta_{n-1}^i$ and $W_n^i = \tilde{W}_n^i$, $i = 1, \ldots, N$.

Sequential Monte Carlo Algorithm (cont'd)

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

Sequential Monte Carlo Algorithm (cont'd)

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

$$\bar{h}_{N_\phi,N} = \sum_{i=1}^{N} h(\theta^i_{N_\phi}) W^i_{N_\phi}.$$

Sequential Monte Carlo

- Researcher chooses the cooling schedule, number of stages, and number of particles
- Accuracy of estimates scales linearly with number of particles
- Parallelizability of algorithm allows the use of HPC to estimate high-dimensional models.

Sequential Monte Carlo

- Sequential Monte Carlo also allows for sequential computation of marginal data density

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

Which implies

Sequential Monte Carlo

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

The Model

- I estimate a small scale New Keynesian Model with two sources of mechanical persistence, developed in detail by (Woodford 2003) and first estimated under AL by (Milani 2005)
  - Habit persistence in consumption by households
  - Inflation indexation by price-setting firms
- In the end we will have five equations:
  - A New Keynesian Phillips Curve
  - A New Keynesian IS curve
  - Taylor Rule for Monetary Policy
  - Productivity Shock
  - Aggregate Demand shock

The Model

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

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

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

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

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

Integrating Learning within the Model

- Solving the model under bounded rationality is mechanical and straightforward, no solution by undetermined coefficients or eigenvalue or schur decompositions needed!
- Agents have a "Perceived Law of Motion", a belief about how the economy operates
  - I model this as a VAR(1) process $\alpha_{t+1} = c + \Phi\alpha_t + \Psi\varepsilon_t$
- One can solve easily for the period t expectations of period $t+1$ variables: $E_t\alpha_{t+1} = c + \Phi\alpha_t$
- If agents only have information up to time $t-1$ then $E_{t-1}\alpha_{t+1} = c + \Phi c + \Phi^2\alpha_{t-1}$
- If simulating the model, simply substitute these $E_{t-n}\alpha_{t+1}$ terms into the agents' Euler equations and one has a solution.

Integrating Learning within the model

- For estimation, recall the state-space representation of a model

$$y_t = Z_t\alpha_t + d_t + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, H_t)$$

$$\alpha_t = T_t\alpha_{t-1} + c_t + R_t\eta_t, \eta_t \sim \mathcal{N}(0, Q_t)$$

- One can cast any DSGE model into the following form: $s_t = P + Qs^e_{t+1} + Rs_{t-1} + S\varepsilon_t$
- To make the model amenable to estimation, then, I take the coefficients from $E_{t-1}\alpha_{t+1} = c + \Phi c + \Phi^2\alpha_{t-1}$ and simply substitute Φ^2 for Q and this gives me the VAR(1) representation of the model
- This gives time-varying coefficients as well in the system matrices.

Choice of Initial beliefs: Equilibrium-Based initials

- For every θ_i that has a unique Rational Expectations solution, there is a state-space representation

$$y_t = Z\alpha_t + d + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, H_t)$$

$$\alpha_t = T\alpha_{t-1} + c + R\eta_t, \eta_t \sim \mathcal{N}(0, Q_t)$$

- When choosing Equilibrium-based initials, then, for every θ_i drawn from the posterior distribution, I substitute the corresponding element of $T(\theta_i)$ into b_0 in agents PLM $\alpha_t = a_0 + b_0\alpha_{t-1} + \epsilon_t$

Choice of Initial Beliefs: Training Sample

- This involves estimating a reduced-form state-space model based on pre-sample data
- The state-space model in my case has three observable variables and five latent variables $y_t = (\pi_t, x_t, i_t)$ or inflation, the output gap, and the federal funds rate and the latent variables are the inflation process, output process, interest rate process, natural interest rate process, and productivity shock process

Choice of Initial Beliefs: Training Sample

- I also exploit the restrictions on the effect of the autoregressive shocks

$$
\begin{bmatrix} \pi_t \\ x_t \\ i_t \\ r_t^n \\ g_t \end{bmatrix} = \mathbf{a} + \begin{bmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{15} \\ b_{21} & b_{22} & b_{23} & \dots & b_{25} \\ b_{31} & b_{32} & b_{33} & \dots & b_{35} \\ 0 & 0 & 0 & b_{44} & 0 \\ 0 & 0 & 0 & 0 & b_{55} \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ x_{t-1} \\ i_{t-1} \\ r_{t-1}^n \\ g_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \varepsilon_t
$$

- After finding the likelihood maximum, I substitute the coefficients into agents' PLM

Choice of Initial Beliefs: Joint Estimation

- One could estimate initial beliefs along with the rest of the model parameters
- This requires researcher to choose an explicit prior distribution
- It also increases substantially the number of estimated parameters, which can confound Bayesian samplers.

Joint Estimation, How to compute the prior density

- I take the simulated Posterior distribution of θ under Rational Expectations
- For each θ_i I then back out the VAR-1 representation of the model

$$\begin{bmatrix} \pi_t \\ x_t \\ i_t \\ r_t^n \\ g_t \end{bmatrix} = \mathbf{a} + \begin{bmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{15} \\ b_{21} & b_{22} & b_{23} & \dots & b_{25} \\ b_{31} & b_{32} & b_{33} & \dots & b_{35} \\ 0 & 0 & 0 & b_{44} & 0 \\ 0 & 0 & 0 & 0 & b_{55} \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ x_{t-1} \\ i_{t-1} \\ r_{t-1}^n \\ g_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \varepsilon_t$$

Joint Estimation

- I then gather elements $(b_{11}, b_{12}, ...)'$ into a single vector for each θ_i
- I then fit a multivariate normal distribution over the vectors from the simulated posterior distribution from the model under rational expectations
- As the posterior density can be made arbitrarily large by shrinking the prior sample space, I set each diagonal element of the covariance matrix equal to the maximum of the sample covariance and one

Priors on Parameters

- I seek to understand just the effects of initial beliefs, and so I seek priors that match the usual conventions in the DSGE literature

The Data

- The estimates I report are based on 1982-2002 data, but some robustness checks are performed with 1960-2008 data
- Data are the time series including inflation, output gap, and the federal funds rate
- Because the latent variables are zero-mean processes but the observables are not zero-mean processes due to the Fed's inflation rate, I use de-meaned values for the federal funds rate and the inflation rate
- Another solution used by (An and Schorfheide 2007) treats those means as parameters to be estimated along with structural parameters

Results: Rational Expectations Baseline

- Parameter estimates under Rational Expectations match somewhat closely those found in (Milani 2005)

Table 1: 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

Equilibrium-based Initial Beliefs

Table 2: 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
gain	0.0218	0.0049	0.0484

Equilibrium-based Initial Beliefs

Table 3: 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
gain	0.0176	0.0088	0.0313

Training Sample Initial Beliefs

- Extremely low estimated gain value due to projection facility

Table 4: 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
gain	0.0061	0.0012	0.0153

Training Sample

Table 5: 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
gain	0.0106	0.0025	0.0212

Joint Estimation Results

Table 6: 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
gain	0.0145	0.0079	0.0256

Joint Estimation Results

Table 7: 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
gain	0.0062	0.0011	0.0138

Summary of Important Parameter Estimates

- Somewhat consistent reductions in Inflation Indexation compared to RE
- Less so with Habit persistence
- Unable to match the near-elimination of mechanical persistence in (Milani 2005)
- Fairly close to those estimated in (Cole and Milani 2019)
- Unusually low gain value compared to (Milani 2005)
- Likely driven by data, as post-1982 estimated model reports extremely low constant gain of .0058

Comparison of Model Fit

- Marginal Data Density allows one to rank models according to in-sample forecasting performance
- They tell researchers, given the prior density, how likely are the data to appear
- Driven by priors, so want to use “reasonable” priors
- Use different priors, or even use different data sets

Comparison of Model Fit

Table 8: 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)

Comparison of Model Fit

Table 9: Mean and Standard Deviation of Natural Logarithms of the Marginal Likelihoods, 1961-2006 data

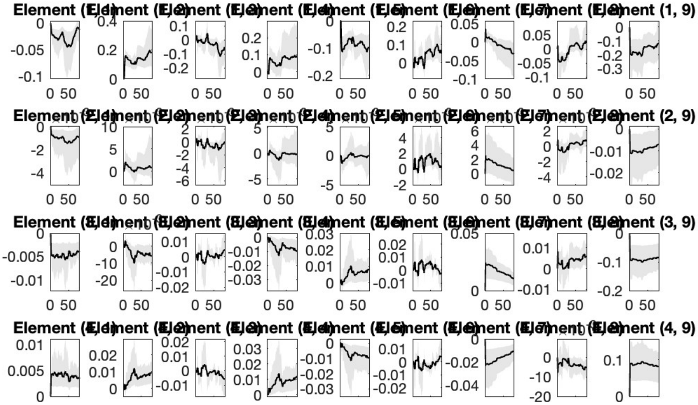
	Full Information	Limited Information
Rational Expectations	-839.2973 (3.4430)	N/A
Equilibrium Initials	-833.9175 (0.8594)	-838.5927 (0.7877)
Training Sample Initials	-2764.7828 (0.6802)	-859.9934 (0.3554)
Jointly Estimated Initials	-885.0740 (12.0577)	-833.8648 (1.7775)

Evolution of Beliefs

- Want to see if results are driven by the choice of projection facility
- If beliefs are driven by projection facility then results are trivial

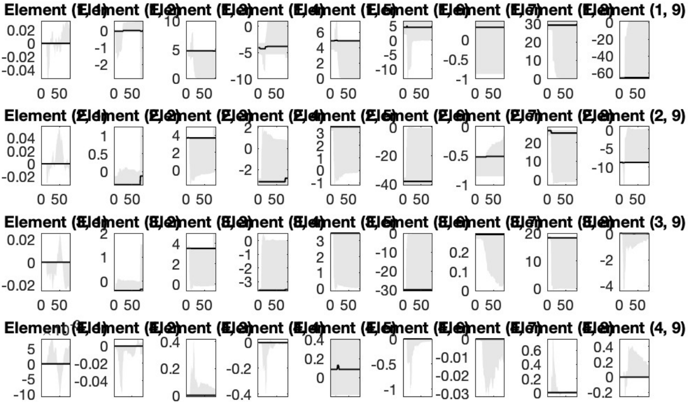
Evolution of Beliefs

- For most models beliefs are moving, and thus not hitting projection facility



Evolution of Beliefs

- For training sample inits with full information, beliefs (nearly) always hitting projection facility and thus not moving at all



Evolution of Beliefs

- Impact of projection facility unsurprising for training sample beliefs
- Projection facility forces beliefs to not update when entering explosivity
- Training sample forces initial beliefs for all θ_i to be a single ϕ_i, Σ_i

Summary

- SMC results unable to eliminate mechanical persistence
- Equilibrium Initials generally at least as good or better than Rational Expectations Baseline in terms of marginal likelihood
- Joint Estimation performs the best
- Training sample performs poorly
- I propose this rule-of-thumb for choosing initial beliefs
  - For models wherein agents use very small forecasting models, jointly estimate initial beliefs
  - For models wherein agents use richer forecasting models a la (Smets and Wouters 2007), consider using equilibrium-based initials

References I

An, Sungbae, and Frank Schorfheide. 2007. "Bayesian Analysis of DSGE Models." *Econometric Reviews* 26 (2-4): 113–72. https://doi.org/10.1080/07474930701220071.

Cole, Stephen J., and Fabio Milani. 2019. "THE MISSPECIFICATION OF EXPECTATIONS IN NEW KEYNESIAN MODELS: A DSGE-VAR APPROACH." *Macroeconomic Dynamics* 23 (3): 974–1007. https://doi.org/10.1017/S1365100517000104.

Evans, and Honkapohja. 2001. *Learning and Expectations in Macroeconomics*. Princeton University Press. http://www.jstor.org/stable/j.ctt7s6t9.

References II

Marcet, Albert, and Sargent. 1989. "Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models." *Journal of Economic Theory* 48 (2): 337–68. https://doi.org/https://doi.org/10.1016/0022-0531(89)90032-X.

Milani, Fabio. 2005. "Expectations, Learning and Macroeconomic Persistence." Macroeconomics 0510022. University Library of Munich, Germany. https://ideas.repec.org/p/wpa/wuwpma/0510022.html.

Smets, Frank, and Rafael Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97 (3): 586–606. https://doi.org/10.1257/aer.97.3.586.

References III

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ [u.a.]: Princeton Univ. Press. http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+368740471&sourceid=fbw_bibsonomy.