Bayesian Analysis of DSGE Models

Frank Schorfheide
Department of Economics, University of Pennsylvania

Challenges for DSGE Model Estimation

- Misspecification: potentially invalid cross-coefficient restrictions on moving-average representations may lead to
  - inferior fit compared to (good!) VARs, e.g., Del Negro, Schorfheide, Smets, and Wouters (2004);
  - "dilemma of absurd parameter estimates."
- Identification: *"(...) A lot of your posteriors look exactly like the priors (...)"* Richard Blundell, when awarding Frank Smets and Raf Wouters with the Hicks-Tinbergen Medal at the 2004 EEA Meetings.
- Size: GEM (IMF) and SIGMA (Federal Reserve Board) provide computational challenges for estimation procedures.

How Can Bayesian Analysis Help?

Misspecification (I)

- Objects of interest: (i) parameter values; (ii) MA representations of time series in terms of structural shocks; (iii) effects of parameter changes on law of motion for data Y?
- Under misspecification there is no single value of θ that delivers the best answers to all three questions. Loss functions matter, see Schorfheide (JAE 2000).
- Likelihood-based estimates tend to minimize the Kullback-Leibler distance between the "truth" and the model, White (Ecta, 1982). It's an important metric as it relates to time series fit of the model, but not always the best to answer (ii) and (iii).

Misspecification (II)

- Suppose we have additional information X, e.g., micro-level data on price-setting behavior, that is informative about θ but no joint likelihood function for Y and X.

- Information in X might be at odds with information in Y-likelihood function: "dilemma of absurd parameter estimates."

Misspecification (III)

- Information in X and Y can be combined using Bayes theorem. Use X to specify a prior $p(\theta)$ then update $p(\theta|Y) \propto \mathcal{L}(\theta|Y)p(\theta)$.

- Bayesian analysis formalizes combination of information and allows us to account for less than conclusive evidence in X / illustrate sensitivity to weight placed on X.

- "Tensions" between X (through prior) and Y (through likelihood) are reflected in low marginal data densities used to weight models: $p(Y) = \int \mathcal{L}(\theta|Y)p(\theta)d\theta$.

Identification (I)

- Some identification issues are fairly obvious...
- If we linearize our DSGE model, the Phillips curve is given by

$$\hat{\pi}_t = \beta I\!E_t[\hat{\pi}_{t+1}] + \kappa(\hat{y}_t - \hat{g}_t) \tag{1}$$

where

$$\kappa = \tau \frac{1-\nu}{\nu \pi^2 \varphi}. \tag{2}$$

- Others are more subtle...

Identification (II)

- $\mathcal{M}_1$ has serially correlated u_t's:

$$y_t = \frac{1}{\alpha}I\!\!E_t[y_{t+1}] + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t, \quad \epsilon_t \sim iid\left(0, (1-\rho\alpha)^2\right). \tag{3}$$

- $\mathcal{M}_2$ has backward-looking term:

$$y_t = \frac{1}{\alpha}I\!\!E_t[y_{t+1}] + \phi y_{t-1} + u_t, \quad u_t = \epsilon_t, \quad \epsilon_t \sim iid\left(0, \left[\frac{\alpha + \sqrt{\alpha^2 - 4\phi\alpha}}{2\alpha}\right]^2\right). \tag{4}$$

- For both specifications, the law of motion of y_t is

$$y_t = \psi y_{t-1} + \eta_t, \quad \eta_t \sim iid(0,1). \tag{5}$$

- Imposing determinacy we obtain:

$$\mathcal{M}_1 : \ \psi = \rho, \quad \mathcal{M}_2 : \ \psi = \frac{1}{2}(\alpha - \sqrt{\alpha^2 - 4\phi\alpha}).$$

Identification (III)

- Strong auxiliary assumptions on the distribution of error terms are needed to distinguish between classes of models. Some references: Sims (Ecta, 1980), Lubik and Schorfheide (AER, 2004), Beyer and Farmer (2004), Canova and Sala (2005).

- Limited information approaches that try to avoid these assumptions are often unable to identify the structural parameters they claim to identify. They only recover reduced form parameters.

- Likelihood-based approaches make auxiliary assumptions transparent. Still likelihood might be flat in some dimensions. Difficult to summarize information in likelihood.

Identification (III)

- Bayesian approach adds prior information that might be helpful to discriminate hypotheses. However, prior is not updated if likelihood is flat.

- Suppose

$$\mathcal{L}(\theta_1, \theta_2|Y^T) = \mathcal{L}(\theta_1, \theta_2'|Y^T) = \tilde{\mathcal{L}}(\theta_1|Y^T) \tag{6}$$

  Straightforward manipulations with Bayes theorem lead to

$$\begin{aligned} p(\theta_1, \theta_2|Y^T) &= \frac{\mathcal{L}'(\theta_1|Y^T)p(\theta_1)p(\theta_2|\theta_1)}{\int_{\theta_1}\left[\mathcal{L}'(\theta_1|Y^T)p(\theta_1)\int_{\theta_2} p(\theta_2|\theta_1)d\theta_2\right]d\theta_1} \\ &= \frac{\mathcal{L}'(\theta_1|Y^T)p(\theta_1)}{\int \mathcal{L}'(\theta_1|Y^T)p(\theta_1)d\theta_1} p(\theta_2|\theta_1) \\ &= p(\theta_1|Y^T)p(\theta_2|\theta_1). \end{aligned}$$

Identification (III)

- Thus, there is updating of the marginal distribution of θ_1, but not of the conditional distribution of $\theta_2|\theta_1$. Nevertheless, if the prior is proper, the posterior is proper. Posterior provides coherent measure of uncertainty.

- From a numerical perspective: the prior can be used to introduce "curvature" into objective function. Maximization of posterior is "easier" than maximization of likelihood function.