The Econometrics of DSGE Models

Giuseppe Ragusa
Luiss University

EIEF
Lecture 3: A modicum of Bayesian theory

April 24, 2015

Estimating (Linearized) DSGE Models

1. Linearize the model

$$\Gamma_0(\theta)E[x_{t+1}] = \Gamma_1(\theta)x_t + C + \Psi(\theta)z_t$$

2. Solve the model

$$x_{t+1} = G_0(\theta) + G(\theta)x_t + M(\theta)\varepsilon_{t+1}$$

3. Attach data to the problem

$$\underbrace{\underbrace{y_t}_{\text{observables}} = H(\theta)x_t \underbrace{(+m(\theta)\eta_t)}_{\text{meas. error}}}_{\text{observation equation}}$$

4. Obtain the likelihood function (Kalman Filter)

$$f(y^T;\theta) = \prod_{t=1}^{T} f(y_t|y_{t-1};\theta)$$

Think in terms of computer code

```
likelihood(theta) {

linearize(theta)
solve(theta)
kalman_filter(theta)

}


theta_hat = maximize(likelihood, theta)
```

Outline

- Bayes' theorem
- Simple example to motivate the use of Bayesian methods
- The basics of the Bayesian logic
- Readings

Bayes' theorem

Suppose X and Y are a pair of random variables. The Bayes's theorem says that

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx}$$

If one of the random variable is a "parameter" θ

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

- $p(y|\theta)$ likelihood function
- $p(\theta)$ prior distribution
- $\int p(y|\theta)p(\theta)d\theta$ marginal density

Notation and terminology

We often write posterior as

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

where "$\propto$" is the proportionality symbol.

- "$\propto$" the proportionality symbol is to be taken for functions of θ (not of y)
- the proportionality comes from the fact that the denominator in the Bayes' theorem does not depend on θ
- This notation is extensively used in the literature

Notation and terminology

$$p(y) = \int p(y|\theta)p(\theta)d\theta$$

- It is sometimes called the prior predictive distribution of y, because it is the distribution that should be used in predicting y, taking into account both the uncertainty about the value of and the residual uncertainty about y when θ is known
- The predictive distribution can useful in model criticism and revision
    - If $p(y)$ is "small" for the observed y, we may reconsider either $p(y|\theta)$ and/or $p(\theta)$

Example: inference μ

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

and assume σ^2 is known.

Frequentist inference based on:

- Point estimate

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$$

- Sampling distribution

$$\sqrt{T}(\bar{y} - \mu) \sim N(0, \sigma^2)$$

- (95%) Confidence interval:

$$\left[\bar{y} - 1.96\sqrt{\frac{\sigma^2}{T}}, \bar{y} + 1.96\sqrt{\frac{\sigma^2}{T}}\right]$$

Example: inference μ

Bayesian inference based on:

- *eliciting* a prior distribution for μ, e.g.,

$$\mu \sim N(\mu_0, \sigma_0^2)$$

- Calculate the posterior distribution

$$p(\mu|y) = \frac{p(y|\mu)p(\mu)}{p(y)}$$

with

$$p(y|\mu) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma}} e^{-(y_t-\mu)^2/2\sigma^2}$$

and

$$\mu \sim N(\mu_0, \sigma_0^2)$$

.

Example: inference on μ

- Point estimate

$$\min_{a \in A} \int \mathscr{L}(\mu - a) p(\mu|y) d\mu$$

- Credible region

$$\int_{C(y)} p(\mu|y) d\mu = .95$$

Example: inference on μ

Let $\lambda = \sigma^{-2}$ and $\lambda_0 = \sigma_0^{-2}$. Then,

$$\begin{aligned}
p(\mu|x) &\propto p(y|\mu)p(\mu) \\
&\propto \exp\left[-\frac{T}{2}\lambda(\mu^2 - 2\bar{y}\mu)\right]\exp\left[-\frac{1}{2}\lambda_0(\mu^2 - 2\mu\mu_0)\right] \\
&= \exp\left[-\frac{T}{2}\lambda(\mu^2 - 2\bar{y}\mu) - \frac{1}{2}\lambda_0(\mu^2 - 2\mu\mu_0)\right] \\
&= \exp\left\{-\frac{T}{2}(\lambda + \lambda_0)\left[\mu^2 - 2\mu\left(\frac{T\lambda\bar{y} + \lambda_0\mu_0}{\lambda + \lambda_0}\right)\right]\right\} \\
&\propto \exp\left\{-\frac{T}{2}(\lambda + \lambda_0)\left[\mu^2 - 2\mu\left(\frac{T\lambda\bar{y} + \lambda_0\mu_0}{\lambda + \lambda_0}\right) + \left(\frac{T\lambda\bar{y} + \lambda_0\mu_0}{\lambda + \lambda_0}\right)^2\right]\right\} \\
&\propto \underbrace{\exp\left[-\frac{T}{2}(\lambda + \lambda_0)\left(\mu - \frac{T\lambda\bar{y} + \lambda_0\mu_0}{\lambda + \lambda_0}\right)^2\right]}_{\text{kernel of normal pdf}}
\end{aligned}$$

Example: inference on μ

The posterior distribution is

$$\mu | y \sim N(\underline{\mu}, \underline{\sigma}^2)$$

where

$$\underline{\mu} = \frac{T\lambda\bar{y} + \lambda_0\mu_0}{T\lambda + \lambda_0}$$

$$\underline{\sigma}^2 = \left[\frac{T}{\sigma^2} + \frac{1}{\sigma_0^2}\right]^{-1}.$$

Notice that, as $T \to \infty$

$$\underline{\mu} - \bar{y} \to 0$$

$$\underline{\sigma}^2 \approx \frac{\sigma^2}{T} \to 0$$

- Special case of a general results: as $T \to \infty$, the prior does not influence the posterior: "as the sample grows, the posterior density becomes increasingly concentrated around the maximum likelihood estimate of μ".

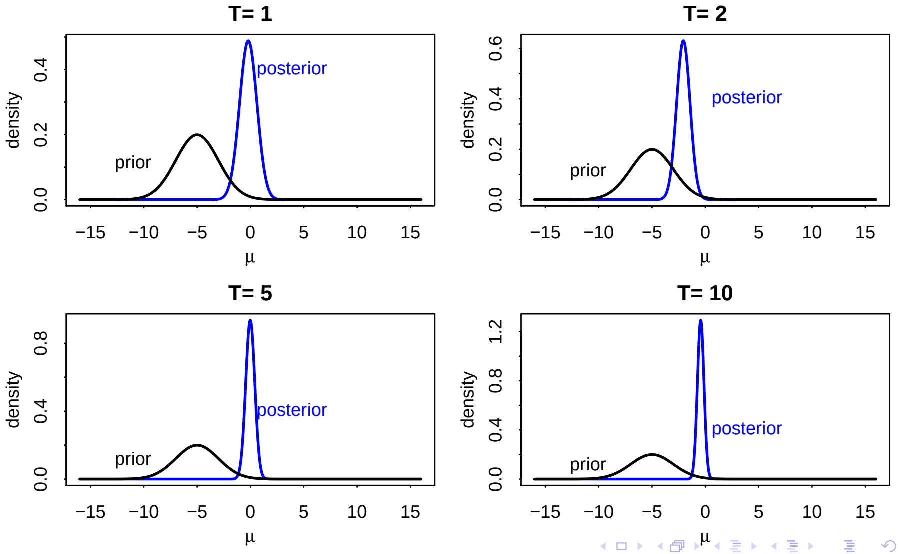
Example: inference on μ

The parameters of the posterior distribution are easy to remember and understand:

- the posterior mean is a weighted average of the prior mean and *datum* value, the weights being proportional to the prior and datum precisions
- the posterior precision is the sum of the prior precision and the datum precision

In a picture

$$y_t \sim N(\mu, 1), \quad \mu \sim N(-5, 2), \quad \mu = 0$$



Conjugate Priors

Definition

A prior is said to be conjugate to the likelihood if the posterior derived from this prior and likelihood is in the same class of distribution as the prior

In the previous example, we elicited a normal prior and we obtained a normal posterior. We say that the normal prior is a conjugate prior for estimating the location parameter μ of normal distribution.

- set of problems for which there exists usable conjugate prior is very limited
- seemingly minor change to model destroys conjugacy
- important model do not have conjugate priors (logistic)

Distributions

Bayesian theory makes heavy use of parametric distributions.

Distribution	Notation	Mean	Variance
Normal	$N(\mu,\sigma^2)$	μ	σ^2
Gamma	$\Gamma(\alpha,\beta)$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
t-student	$\mathscr{T}(\mu,\sigma^2)$	μ	σ^2
Inverse Gamma	$\Gamma^{-1}(\alpha,\beta)$	$\frac{\alpha}{\beta-1}$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$
	if $X \sim \Gamma(\alpha,\beta)$, $X^{-1} \sim \Gamma^{-1}(\alpha,\beta)$		
Chi squared	χ^2_ν	ν	2ν
Inverse chi squared	χ_ν^{-2}	$\frac{1}{\nu-2}$	$\frac{2}{(\nu-2)^2(\nu-4)}$
Scale inv. chi squared	$\chi^{-2}(\nu,\tau^2)$	$\frac{\nu\tau^2}{(\nu-2)}$	$\frac{2\nu^2\tau^4}{(\nu-2)^2(\nu-4)}$
	$X \sim \chi^{-2}(\nu,\tau^2)$, then $X \sim \Gamma^{-1}(\nu/2,\nu\tau^2/2)$		

* α and β of the Γ distribution are shape and rate, respectively.

Example: Estimating μ and σ^2

Improper priors

When analyzing a sample from a normal distribution, we are usually uncertain about both the mean and the variance.

- Likelihood

$$p(y|\mu,\sigma^2) = (2\pi\sigma^2)^{-T/2} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(y_i-\mu)^2/\sigma^2\right]$$

- Prior

$$p(\mu,\sigma^2) = p(\mu)p(\sigma^2)$$

  where

$$p(\mu) \propto 1, \quad p(\sigma^2) \propto \sigma^{-2}$$

- These are improper priors, since $\int p(\mu)d\mu = \infty$ and $\int p(\sigma^2)d\sigma^2 = \infty$.
- Improper priors may or may not be usable, depending on whether the posterior density is proper.
- In this case, the posterior is proper even if priors are improper

Example: inference on μ and σ^2

Improper priors

The posterior distribution is:

$$p(\mu, \sigma^2 | y) \propto \sigma^{-T} (2\pi\sigma^2)^{-T/2} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(y_i - \mu)^2/\sigma^2\right] / \sigma^2$$

We have, for $s^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_i - \bar{y})^2$, the following results:

Conditionals

$$\mu | \sigma^2, y \sim N\left(\bar{y}, \frac{\sigma^2}{T}\right)$$

Marginals

$$\mu | y \sim \mathscr{T}_{T-1}(\bar{y}, s^2)$$

$$\sigma^2 | y \sim \Gamma^{-1}\left(\frac{T-1}{2}, \frac{(T-1)s^2}{2}\right)$$

Example: inference on μ and σ^2

Conjugate priors

- Likelihood

$$p(y|\mu,\sigma^2) = (2\pi\sigma^2)^{-T/2} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(y_i-\mu)^2/\sigma^2\right]$$

- Prior: $p(\mu|\sigma^2) \propto N(\mu_0, \sigma^2/\kappa_0)$, $\quad p(\sigma^2) \propto \Gamma^{-1}(\alpha_0, \beta_0)$. Let

$$\mu_T = \frac{\kappa_0\mu_0 + T\bar{y}}{k_0 + T}, \quad \nu_T = \nu_0 + T, \quad \kappa_T = k_0 + T$$

$$\nu_T\sigma_T^2 = \nu_0\sigma_0^2 + (T-1)\sigma^2 + \frac{\kappa_0 T}{\kappa_0 + T}(\mu_0 - \bar{y})^2$$

Conditionals

$$\mu|y,\sigma^2 \sim N\left(\mu_T, \frac{\sigma^2}{\kappa_0 + T}\right)$$

Marginals

$$\mu|y \sim \mathscr{T}_{\nu_T}(\bar{y}, \sigma_T^2/\kappa_T)$$
$$\sigma^2|y \sim \Gamma^{-1}(\nu_T, \sigma_T^2)$$

Example: inference on μ and σ^2

Conjugate priors

- The joint distribution is also available in closed-form.
- It is called the Normal-Inverted-Gamma distribution:

$$p(\mu, \sigma^2 | \mu_0, \kappa_0, \alpha_0, \beta_0, y) = \frac{\sqrt{\kappa_0}}{\sigma\sqrt{2\pi}} \frac{\beta^{\alpha_0}}{\Gamma(\alpha_0)} \left(\frac{1}{\sigma^2}\right)^{\alpha_0+1} \times \exp\left\{-\frac{2\kappa_0\beta_0 + (\mu - \mu_0)}{2\kappa_0\sigma^2}\right\}.$$

- Notice that although we can evaluate the likelihood function, we cannot directly *simulate* from it.

Linear model

Consider the linear model:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{1t} + \ldots + \beta_{k-1} x_{(k-1)t} + u_t \\ &= x_t \beta + u_t \end{aligned}$$

or, in matrix form,

$$y = X\beta + u, \quad u \sim N(0, \sigma^2 I_T).$$

We assume that u, X are independent. The formal posterior distribution is thus

$$\begin{aligned} p(\beta | y, X) &\propto p(y, X | \beta) p(\beta) \\ &= p_u(y - X\beta | \beta) p(X | \beta) p(\beta) \end{aligned}$$

If $p(X|\beta) = p(X)$, that is, X is *strictly exogenous*, the posterior is

$$p(\beta | y, X) \propto \exp\left[-\frac{1}{2} \sum_{t=1}^{T} (y_t - x_t \beta)^2 / \sigma^2 \right] p(\beta)$$

Linear model

Conjugate Priors

The conjugate priors for the parameters of the linear model are

$$
\begin{aligned}
p(\beta|\sigma^2) &\sim N(\beta_0, \sigma^2 A_0^{-1}) \\
p(\sigma^2) &\sim \Gamma^{-1}(\nu_0/2, \sigma_0^2/2).
\end{aligned}
$$

Let

$$
\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'Y, \\
s^2 &= (y - X'\hat{\beta})'(y - X'\hat{\beta})/(T-k) \\
A_T &= (A_0 + X'X)^{-1} \\
\tilde{\beta} &= A_T(A_0\beta_0 + X'Y) \\
\nu_T &= \nu_0 + T \\
\nu_T\sigma_T^2 &= \nu_0\sigma_0^2 + (T-k)s^2 + (\beta_0 - \tilde{\beta})A_0(\beta_0 - \tilde{\beta}) + (\hat{\beta} - \tilde{\beta})X'X(\hat{\beta} - \tilde{\beta})
\end{aligned}
$$

Linear model

Conjugate Priors, ctd

Conditionals

$$\beta|y,\sigma^2 \sim N\left(\tilde{\beta},\sigma^2 A_T^{-1}\right)$$

- The conditional posterior of β is a multivariate normal distribution with location $\tilde{\beta}$ and scale $\sigma^2 A_T^{-1}$.

Marginals

$$\beta|y \sim \mathscr{T}_{\nu_T}(\tilde{\beta},\sigma_T^2 A_T^{-1})$$
$$\sigma^2|y \sim \Gamma^{-1}\left(\nu_T/2,\nu_T\sigma_T^2/2\right)$$

- The marginal posterior of β is a multivariate t distribution with location $\hat{\beta}$, scale $\sigma_T^2 A_T^{-1}$, and $\nu_T = \nu_0 + T$ degrees of freedom.

Inference about β and σ

Improper priors

We first consider the case with improper priors

$$p(\beta,\sigma^2) \propto 1/\sigma^2, \quad -\infty < \beta < +\infty, \tau > 0$$

Conditionals

$$\beta|y,\sigma^2 \sim N\left(\hat{\beta},\sigma^2(X'X)^{-1}\right)$$

Marginals

$$\beta|y \sim \mathscr{T}_{T-k}\left(\hat{\beta},s^2(X'X)^{-1}\right)$$

$$\sigma^2|y \sim \Gamma^{-1}((T-k)/2, \\ (T-k)s^2(X'X)^{-1}/2)$$

- The posterior of β is a multivariate t distribution with location $\hat{\beta}$, scale s^2(T-k), and degrees of freedom $T-K$.
- The posterior of σ^2 is a inverted gamma distribution with shape $(T-k)/2$ and scale $s^2(T-k)(X''X)^{-1}/2$

Note on Bayes regression

- The posterior mean of β is shrinkage estimator, in the sense that the least squares estimator is shrank toward the prior mean
- The posterior mean of σ^2 is centered over s^2 which is a weighted average of the prior parameter and a sample quantity (although Ts^2 includes $(\tilde{\beta} - \bar{\beta})A(\tilde{\beta} - \bar{\beta})$, which represents the degree to which the prior mean differs from the OLS.
- As $T \to \infty$, the posterior mean converges to the OLS

Loose ends (Priors)

Three kind of priors:

- "improper"
- proper, but conjugate
- "vague", proper priors with very large variance, e.g., $\beta|\sigma^2, y, X \sim N(0, 1000 I_k)$
  - Vague priors are often improperly referred to as non-informative, but they are sensitive to scaling, e.g. if $\gamma = h(\beta)$,

$$p(\gamma) = p(h^{-1}(\gamma)) \left| \frac{\partial h(\beta)}{\partial \beta} \right|^{-1}$$

  - Jeffrey's invariant priors,

$$p(\theta) \propto \sqrt{\det \left\{ -E \left[\frac{\partial^2 \log p(y|\theta)}{\partial \theta \partial \theta'} \right] \right\}}$$

- Hierarchical priors, e.g.,

$$\mu \sim N(\mu_0, \sigma_0^2), \quad \sigma_0^2 \sim G(\alpha_0, \beta_0)$$

"Estimating" the prior

Consider the prior predictive distribution

$$p(y) = \int p(y|\theta)p(\theta)d\theta$$

- We can choose the prior such to maximize the predictive density

$$\max_{\lambda \in \Lambda} \int p(y|\theta)\gamma(\theta|\lambda)d\theta$$

- Minimum distance, given an estimate of, $p(\theta)$, $\hat{p}(y)$, find $p(\theta)$ that minimizes

$$\min_{\gamma(\theta)} \int \log\left(\frac{\hat{p}(y)}{\int p(y|\theta)\gamma(\theta)d\theta}\right) d\hat{p}(y)$$

Keyword: Empirical Bayes

Point Estimator

Recall that the Bayes' point estimates are defined as

$$\min_{a \in A} \int \mathscr{L}(\theta - a) p(\theta|y) d\theta$$

for given loss functions $\mathscr{L}(\cdot)$.

- if $\mathscr{L}(\theta - a) = k(\theta - a)^2$, the Bayes's point estimate is the posterior mean
- if $\mathscr{L}(\theta - a) = k|\theta - a|^2$, the Bayes's point estimate is the posterior median

Complete class theorem

Any admissible estimator is a Bayes estimator w.r.t. some prior $p(\theta)$. If you are frequentist, you can only consider Bayes estimator

Asymptotic behavior of posteriors

The Bernstein-von Mises theorem

As $T \to \infty$,

$$\sup_B \left| \int_B p(\theta|y)d\theta \to \int_B N(\hat{\theta}^{MLE}, \frac{H^{-1}}{T})d\theta \right| \to 0,$$

The likelihood term dominates the prior and the prior becomes more and more uniform in appearance in the region in which the likelihood is concentrating.

- It is sometimes said, in defense of the Bayesian concept, that the choice of prior distribution is unimportant in practice, because it hardly influences the posterior distribution at all when there are moderate amounts of data. The less said about this defense the better.

Asymptotic behavior of posteriors

The Bernstein-von Mises theorem

As $T \to \infty$,

$$\sup_B \left| \int_B p(\theta|y)d\theta \to \int_B N(\hat{\theta}^{MLE}, \frac{H^{-1}}{T})d\theta \right| \to 0,$$

The likelihood term dominates the prior and the prior becomes more and more uniform in appearance in the region in which the likelihood is concentrating.

- It is sometimes said, in defense of the Bayesian concept, that the choice of prior distribution is unimportant in practice, because it hardly influences the posterior distribution at all when there are moderate amounts of data. The less said about this defense the better.

Identification

- Identification is a great issue from a frequentist point of view
- Asymptotic theory for non-identified models and weakly-identified models need to be modified, often without success
- From a Bayesian point of view, non-identification means that the likelihood is flat in some region of the parameter space, thus the prior will drive inference
- With dependent priors, e.g. $p(\theta_1, \theta_2) \neq p(\theta_1)p(\theta_2)$ non-identification θ_1 might spill-over to other parameters
- Poirier, Dale J. "Revising beliefs in nonidentified models." Econometric Theory 14.4 (1998): 483-509.

Literature

- Textbooks:
  - Robert, Christian. The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Verlag, 2007.
  - Berger, James O. Statistical decision theory and Bayesian analysis. Springer Verlag, 1985.
  - Geweke, John. Contemporary Bayesian econometrics and statistics. Vol. 537. Wiley-Interscience, 2005.
  - Lancaster, Tony. An introduction to modern Bayesian econometrics. Oxford: Blackwell, 2004.
  - Gelman, Andrew, et al. Bayesian data analysis. Chapman & Hall/CRC, 2004.
  - Bernardo, José M., and Adrian FM Smith. Bayesian theory. Vol. 405. Wiley, 2009.