A Course in Bayesian Econometrics

University of Queensland

July, 2008

Slides for Lecture on

An Overview of Bayesian Econometrics

Gary Koop, University of Strathclyde

1 Bayesian Theory

- Reading: Chapter 1 of textbook and Appendix B, section B.1.

- We will begin with broad outlines of general concepts in Bayesian theory before getting to practical models such as the regression model. If you know these general concepts you will never get lost.

- What does an econometrician do? i) Estimate parameters in a model (e.g. regression coefficients), ii) Compare different models (e.g. hypothesis testing), iii) Prediction.

- Bayesian econometrics does all these things based on a few simple rules of probability.

- Let A and B be two events, $p(B|A)$ is the conditional probability of $B|A$. "summarizes what is known about B given A"

- Bayesians use this rule with $B =$ something known or assumed (e.g. the Data), A is something unknown (e.g. coefficients in a model).

- Let y be the data, y^* be unobserved data (i.e. to be forecast), M_i for $i = 1, .., m$ be the set of models under consideration each of which depends on some parameters, θ^i.

- Learning about parameters in a given model is based on the posterior density: $p(\theta^i|M_i, y)$

- Model comparison is based on posterior model probability: $p(M_i|y)$

- Prediction is based on the predictive density $p(y^*|y)$.

1.1 Bayes Theorem

- I expect you know basics of probability theory from your previous studies, see Appendix B of my textbook if you do not.

- *Definition: Conditional Probability*

The conditional probability of A given B, denoted by $\Pr(A|B)$, is the probability of event A occurring given event B has occurred.

Theorem: Rules of Conditional Probability including Bayes' Theorem

Let A and B denote two events, then

- $\Pr(A|B) = \frac{\Pr(A,B)}{\Pr(B)}$ and

- $\Pr(B|A) = \frac{\Pr(A,B)}{\Pr(A)}$.

These two rules can be combined to yield *Bayes' Theorem*:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}.$$

Note: Above is expressed in terms of two events, A and B. However, they also can be interpreted as holding for two random variables, A and B with probability or probability density functions replacing the $\Pr()$s in the previous formulae.

1.2 Learning About Parameters in a Given Model (Estimation)

- Assume we are working with a single model (e.g. a regression model with a particular set of explanatory variables) which depends on parameters θ

- So we want to figure out properties of the posterior $p(\theta|y)$

- It is convenient to use Bayes' rule to write the posterior in a different way.

- Bayes' rule lies at the heart of Bayesian econometrics:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}.$$

- Replace B by θ and A by y to obtain:

$$p(\theta|y) = \frac{p(y|\theta)p\left(\theta\right)}{p(y)}.$$

- Bayesians treat $p(\theta|y)$ as being of fundamental interest. That is, it directly addresses the question "Given the data, what do we know about θ?".

- The treatment of θ as a random variable is controversial among some econometricians. The chief competitor to Bayesian econometrics, called *frequentist econometrics*, says that θ is not a random variable.

- For estimation we can ignore the term $p(y)$ since it does not involve θ:

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

- $p(\theta|y)$ is referred to as the *posterior density* (i.e. "after or posterior to" seeing the data)

- $p(y|\theta)$ is the *likelihood function*

- $p(\theta)$ as the *prior density*.

- "posterior is proportional to likelihood times prior".

- $p(\theta)$, does not depend on the data. It contains any non-data information available about θ.

- Prior information is a controversial aspect of Bayesian econometrics since it sounds unscientific.

- Bayesian answers (to be elaborated on later):

- i) Often we do have prior information and, if so, we should include it (more information is good)

- ii) Can work with “noninformative” priors

- iii) Can use “empirical Bayes” methods which estimate prior from the data

- iv) Training sample priors

- v) Bayesian estimators often have better frequentist properties than frequentist estimators (e.g. results due to Stein show MLE is inadmissible – but Bayes estimators are admissible)

- vi) Prior sensitivity analysis

1.3 Prediction in a Single Model

- Prediction based on the *predictive density* $p(y^*|y)$

- Since a marginal density can be obtained from a joint density through integration (see Appendix B) we can write:

$$p(y^*|y) = \int p(y^*, \theta|y) d\theta.$$

- Term inside the integral can be rewritten using another rule of probability as:

$$p(y^*|y) = \int p(y^*|y, \theta) p(\theta|y) d\theta.$$

- Prediction involves the posterior and $p(y^*|y, \theta)$ (more description provided later).

1.4 Model Comparison (Hypothesis testing)

- Models denoted by M_i for $i = 1, .., m$. M_i depends on parameters θ^i.

- *Posterior model probability* is $p(M_i|y)$.

- Using Bayes rule with $B = M_i$ and $A = y$ we obtain:

$$p(M_i|y) = \frac{p(y|M_i)p(M_i)}{p(y)}.$$

- $p(M_i)$ is referred to as the *prior model probability*.

- $p(y|M_i)$ is called the *marginal likelihood.*

- How is marginal likelihood calculated?

- Posterior can be written as:

$$p(\theta^i|y, M_i) = \frac{p(y|\theta^i, M_i)p(\theta^i|M_i)}{p(y|M_i)}$$

- Integrate both sides of previous equation with respect to θ^i, use the fact that $\int p(\theta^i|y, M_i)d\theta^i = 1$ (since probability density functions integrate to one) and rearrange we obtain:

$$p(y|M_i) = \int p(y|\theta^i, M_i)p(\theta^i|M_i)d\theta^i.$$

- Note that the marginal likelihood depends only on the prior and likelihood.

- Bayesians often present *posterior odds ratio* to compare two models:

$$PO_{ij} = \frac{p(M_i|y)}{p(M_j|y)} = \frac{p(y|M_i)p(M_i)}{p(y|M_j)p(M_j)}.$$

- Note that, since $p(y)$ is common to both models, do not need to work it out. Can use fact that $p(M_1|y) + p(M_2|y) + ... + p(M_m|y) = 1$ and posterior odds ratios to calculate the posterior model probabilities.

- For instance, if we have $m = 2$ models then we can use the two equations:

$$p(M_1|y) + p(M_2|y) = 1$$

and

$$PO_{12} = \frac{p(M_1|y)}{p(M_2|y)}$$

to work out

$$p(M_1|y) = \frac{PO_{12}}{1 + PO_{12}}$$

and

$$p(M_2|y) = 1 - p(M_1|y).$$

- The *Bayes Factor* is defined as:

$$BF_{ij} = \frac{p(y|M_i)}{p(y|M_j)}.$$

1.5 Summary

On one level, this course could end right here. These few pages have outlined all the basic theoretical concepts required for the Bayesian to learn about parameters, compare models and predict. We stress what an enormous advantage this is. Once you accept that unknown things (i.e. θ, M_i and y^*) are random variables, the rest of Bayesian approach is non-controversial.

What are going to do in rest of this course?

See how these concepts work in commonly-used models (e.g. the regression model).

Bayesian computation.

2 Bayesian Computation

- How do you present results from a Bayesian empirical analysis?

- $p(\theta|y)$ is a p.d.f. Especially if θ is a vector of many parameters cannot present a graph of it.

- Want features analogous to frequentist point estimates and confidence intervals.

- A common point estimate is the mean of the posterior density (or *posterior mean*).

- Let θ be a vector with k elements, $\theta = (\theta_1, .., \theta_k)'$. The posterior mean of any element of θ is:

$$E(\theta_i|y) = \int \theta_i p(\theta|y) d\theta.$$

- Aside *Definition B.8: Expected Value*

Let $g()$ be a function, then the *expected value* of $g(X)$, denoted $E[g(X)]$, is defined by:

$$E[g(X)] = \sum_{i=1}^{N} g(x_i)\, p(x_i)$$

if X is a discrete random variable with sample space $\{x_1, x_2, x_3, .., x_N\}$ and

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, p(x)\, dx$$

if X is a continuous random variable (provided $E[g(X)] < \infty$).

- Most common measure of dispersion is the *posterior standard deviation* which is the square root of the *posterior variance*. The latter is calculated as:

$$var(\theta_i|y) = E(\theta_i^2|y) - \{E(\theta_i|y)\}^2,$$

which requires evaluation posterior mean as well as:

$$E(\theta_i^2|y) = \int \theta_i^2 p(\theta|y) d\theta.$$

- Many other possible features of interest. E.g. what is probability that a coefficient is positive?

$$p(\theta_i \geq 0|y) = \int_0^\infty p(\theta|y) d\theta$$

- All of these posterior features which the Bayesian may wish to calculate have the form:

$$E\left[g\left(\theta\right)|y\right] = \int g(\theta)p(\theta|y)d\theta,$$

where $g(\theta)$ is a *function of interest*.

- All of these features have integrals in them. Marginal likelihood and predictive density also involved integrals.

- Apart from a few simple cases, it is not possible to evaluate these integrals analytically, and we must turn to the computer.

2.1 Posterior Simulation

- The integrals involved in Bayesian analysis are usually evaluated using simulation methods. We will develop many of these later. Here we just describe the basic idea to give you some intuition.

- From your study of frequentist econometrics you should know some asymptotic theory and, in particular, the idea of a Law of Large Numbers (LLN) and a Central Limit Theorem (CLT).

- A typical LLN would say "consider a random sample, $Y_1, ..Y_N$, as N goes to infinity, the average converges to its expectation" (e.g. $\overline{Y} \rightarrow \mu$)

- Bayesians use a LLN as "consider a random sample from the posterior, $\theta^{(1)}, ..\theta^{(S)}$, as S goes to infinity, the average of these converges to $E[\theta|y]$"

- Note: Bayesians use asymptotic theory, but asymptotic in S (under control of researcher) not N

- Example: Monte Carlo integration.

Let $\theta^{(s)}$ for $s = 1, .., S$ be a random sample from $p(\theta|y)$ and define

$$\widehat{g}_S = \frac{1}{S}\sum_{s=1}^{S} g\left(\theta^{(s)}\right),$$

then $\widehat{g}_S$ converges to $E\left[g\left(\theta\right)|y\right]$ as S goes to infinity.

- Monte Carlo integration can be used to approximate $E\left[g\left(\theta\right)|y\right]$, but only if S were infinite would the approximation error go to zero.

- We can choose any value for S (although larger values of S will increase the computational burden).

- To gauge size of approximation error, we can use a CLT to obtain a numerical standard error.

- In practice, most Bayesians write their own programs (e.g. using Gauss, Matlab or R) to do posterior simulation

- BUGS (Bayesian Analysis Using Gibbs Sampling) and BACC (Bayesian Analysis, Computation and Communication) are popular Bayesian packages, but only have limited set of models (or require substantial programming to adapt to other models)

- Bayesian work cannot (easily) be done in standard econometric packages like Microfit, Eviews or Stata.