An Overview of Bayesian Econometrics

Bayesian Theory

- Begin with general concepts in Bayesian theory before getting to specific models.
- If you know these general concepts you will never get lost.
- What does 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 these 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 $A =$ something known or assumed (e.g. the Data), B is something unknown (e.g. coefficients in a model).
- Let y be data, y^* be unobserved data (i.e. to be forecast), M_i for $i = 1, .., m$ be set of models each of which depends on some parameters, θ^i.
- Learning about parameters in a model is based on the posterior density: $p(\theta^i|M_i, y)$
- Model comparison based on posterior model probability: $p(M_i|y)$
- Prediction based on the predictive density $p(y^*|y)$.

Bayes Theorem

- I expect you know basics of probability theory from 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*:

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

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

Learning About Parameters in a Given Model (Estimation)

- Assume a single model which depends on parameters θ
- 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:

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

- Bayesians treat $p(\theta|y)$ as being of fundamental interest: "Given the data, what do we know about θ?".
- Treatment of θ as a random variable is controversial among some econometricians.
- Competitor to Bayesian econometrics, called *frequentist econometrics*, says that θ is not a random variable.
- For estimation 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*
- $p(y|\theta)$ 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 controversial aspect 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 hierarchical priors which treat prior hyperparameters as parameters and estimates them
- 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

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:

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

- Term inside integral can be rewritten 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)

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 with respect to θ^i, use fact that $\int p(\theta^i|y, M_i)d\theta^i = 1$ and rearrange:

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

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

- *Posterior odds ratio* compares 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: $p(y)$ is common to both models, no need to calculate.

- Can use fact that $p(M_1|y) + p(M_2|y) + ... + p(M_m|y) = 1$ and PO_{ij} to calculate the posterior model probabilities.
- E.g. suppose $m = 2$ models and you know:

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

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

- imply

$$p(M_1|y) = \frac{PO_{12}}{1 + PO_{12}}$$
$$p(M_2|y) = 1 - p(M_1|y).$$

- The *Bayes Factor* is:

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

Example: Deriving Posterior When Data Has Bernoulli Distribution

- Background:
- Experiment repeated T times
- Each time the outcome can be "success" or "failure"
- y_t for $t = 1, .., T$ are random variables for each repetition of experiment
- Realization of y_t can be 1 or 0
- Probability of success is θ (hence probability of failure is $1 - \theta$)
- The goal is to estimate θ

Example (cont.): The Bernoulli Likelihood function

- Notation for things above is: $y_t \in \{0,1\}$, $0 \leq \theta \leq 1$ and

$$p(y_t|\theta) = \begin{cases} \theta & \text{if } y_t = 1 \\ 1-\theta & \text{if } y_t = 0. \end{cases}$$

- Let m be the number of successes in T repetitions of experiment
- Likelihood function is:

$$\begin{aligned} p(y|\theta) &= \prod_{t=1}^{T} p(y_t|\theta) \\ &= \theta^m (1-\theta)^{T-m} \end{aligned}$$

Example (cont.): The Beta Prior

- View this likelihood in terms of θ: proportional to p.d.f. of a Beta distribution
- See definition in textbook Appendix B or Wikipedia
- Most common distribution for random variables bounded to lie in the interval $[0, 1]$
- Commonly used for parameters which are probabilities (like θ)
- Bayesians need prior
- Let us also Beta distribution for prior
- Prior beliefs concerning θ are represented by

$$p(\theta) \propto \theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\delta}-1}$$

Example (cont.): Eliciting a Prior

- The researcher chooses prior hyperparameters $\underline{\alpha} > 0$ and $\underline{\delta} > 0$ to reflect beliefs
- Called prior elicitation
- Properties of Beta distribution imply prior mean is

$$E(\theta) = \frac{\underline{\alpha}}{\underline{\alpha} + \underline{\delta}}$$

- Suppose you believe, a priori, that success and failure are equally likely
- $E(\theta) = \frac{1}{2}$ obtained by setting $\underline{\alpha} = \underline{\delta}$
- If I look on Wikipedia I see $\underline{\alpha} = \underline{\delta} = 2$ has mean at $E(\theta) = \frac{1}{2}$ but spreads probability widely over interval $[0, 1]$
- So I might be "relatively noninformative" and choose this for my prior

Example (cont.): A Noninformative Prior

- Or I might set $\underline{\alpha} = \underline{\delta} = 1$ and be completely noninformative
- Note: $\underline{\alpha} = \underline{\delta} = 1$ implies $p(\theta) \propto 1$
- Uniform distribution over interval $[0, 1]$
- Every value for θ receives same probability (equally likely) = noninformative prior

Example (cont.): Deriving the Posterior

- To get posterior multiply prior times likelihood

$$
\begin{aligned}
p(\theta|y) &\propto \theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\delta}-1}\theta^{m}(1-\theta)^{T-m} \\
&= \theta^{\overline{\alpha}-1}(1-\theta)^{\overline{\delta}-1}
\end{aligned}
$$

- where

$$
\begin{aligned}
\overline{\alpha} &= \underline{\alpha} + m \\
\overline{\delta} &= \underline{\delta} + T - m
\end{aligned}
$$

Example (cont.): Interpretation and Terminology

- Posterior same Beta form as prior (terminology = conjugate)
- Posterior has arguments $\overline{\alpha}$ and $\overline{\delta}$ instead of $\underline{\alpha}$ and $\underline{\delta}$
- Arguments have been updated:
- Begin with prior belief ($\underline{\alpha}$ or $\underline{\delta}$) update with data information (m and $T - m$)
- Posterior combines prior and data information
- "Bayesian learning" = learn about θ by combining prior and data information

Example (cont.): Predictive Density

- Derivations of marginal likelihood and predictive density are a bit messier
- Exercise 7.1 in Bayesian Econometric Methods shows predictive density has Beta-Binomial distribution
- Shows
$$E\left(y^{*}|y\right)=\frac{\overline{\alpha}}{\overline{\alpha}+\overline{\delta}}$$
- How do I interpret this?
- Question: If I run the experiment again what is the probability of getting a success?
- Answer: $\frac{\overline{\alpha}}{\overline{\alpha}+\overline{\delta}}$

Summary

- These few pages have outlined all the basic theoretical concepts required for the Bayesian to learn about parameters, compare models and predict.
- This is an enormous advantage: 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 some models of interest.
- First the regression model
- Then time series models of interest for macroeconomics
- Bayesian computation.

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 discrete random variable with sample space $\{x_1, x_2, x_3, .., x_N\}$
- $$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, p(x)\, dx$$
- if X is a continuous random variable (provided $E[g(X)] < \infty$).

- Common measure of dispersion is the *posterior standard deviation* (square root of *posterior variance*)
- Posterior variance:

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

- This requires calculating another expected value:

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

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

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

- All of these posterior features 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 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.

Posterior Simulation

- The integrals involved in Bayesian analysis are usually evaluated using simulation methods.
- Will use several methods later on. Here we provide some intuition.
- Frequentist asymptotic theory uses Laws of Large Numbers (LLN) and a Central Limit Theorems (CLT).
- A typical LLN: "consider a random sample, $Y_1, ..Y_N$, as N goes to infinity, the average converges to its expectation" (e.g. $\overline{Y} \to \mu$)
- Bayesians use LLN: "consider a random sample from the posterior, $\theta^{(1)}, ..\theta^{(S)}$, as S goes to infinity, the average of these converges to $E\left[\theta|y\right]$"
- 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 approximates $E\left[g\left(\theta\right)|y\right]$, but only if S were infinite would the approximation error be zero.
- We can choose any value for S (but larger values of S will increase computational burden).
- To gauge size of approximation error, use a CLT to obtain numerical standard error.

- Most Bayesians write own programs (e.g. using Matlab, Julia, Python, Gauss, R or C++) to do posterior simulation
- Bayesian work cannot (easily) be done in standard econometric packages like Microfit, Eviews or Stata.
- New Stata has some Bayes, but limited (and little for macroeconomics)
- I have a Matlab website for VARs, TVP-VARs and TVP-FAVARs (see my website)
- Dimitris Korobilis: https://sites.google.com/site/dimitriskorobilis/matlab
- Joshua Chan: http://joshuachan.org/
- Haroon Mumtaz: https://sites.google.com/site/hmumtaz77/
- Many more using R see http://cran.r-project.org/web/views/Bayesian.html

Learning Outside of Lectures

- Go through the textbook and readings provided.
- In addition to this:
- Computational methods are the most important thing for the aspiring Bayesian econometrician to learn
- Thus, we devote all of the tutorial hours in this course to computer sessions
- Four computer sessions based on four question sheets
- Computer code will be provided which will "answer" the questions
- Work through/adapt/extend the code
- Idea is to develop skills so as to produce your own code or adapt someone else's for your purposes

Learning Outside of Lectures

- What about proofs/derivations of theoretical results?
- In lectures (with a few exceptions) will not do proofs
- E.g. just state a particular posterior in Normal with formula given for mean and variance
- To use Bayesian methods in practice, this is usually all that is needed
- But if you want to derive posterior for new model or obtain deeper understanding need to learn necessary tools
- These tools best learned by practicing on your own
- I will provide Problem Sheets which give practice problems and ask for derivations of some key results
- Answers are provided, so I will not formally take them up in lectures or tutorials
- Bayesian Econometrics Methods by Chan, Koop, Poirier and Tobias has many more practice problems (and answers)