A Course in Bayesian Econometrics

University of Queensland

July, 2008

Slides for Lecture on

Bayesian Analysis of the Linear Regression Model

Gary Koop, University of Strathclyde

1 Summary

- Readings: Chapter 2 and 3 of textbook.

- The Normal linear regression model with a single explanatory variable (no matrix algebra required)

- The Normal linear regression model with several explanatory variables (matrix algebra used)

- In these lectures we will use a so-called "natural conjugate prior" which yields analytical results.

- No posterior simulation necessary.

- We will compare analytical results with Monte Carlo integration results

2 Posterior Analysis

Remember: the posterior is proportional to the prior times the likelihood function.

2.1 The Likelihood Function

- Let y_i and x_i denote the observed data on the dependent and explanatory variables, respectively, for individual i for $i = 1, .., N$.

- We will work with the linear regression model (without intercept to simplify mathematics):

$$y_i = \beta x_i + \varepsilon_i,$$

where ε_i is an error term.

- Assumptions about ε_i and x_i determine the form of the likelihood function. The classical assumptions (which we will free up in later lectures) are:

1. ε_i is i.i.d. $N\left(0, \sigma^2\right)$ where i.i.d. stands for "independent and identically distributed".

2. The x_i are fixed (i.e. not random variables)

- The likelihood function is defined as the joint probability density function for all the data conditional on the unknown parameters: $p\left(y|\beta, \sigma^2\right)$

- To start introducing some matrix algebra, stack all observations of the dependent variable into a vector of length N:

$$
y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_N \end{bmatrix},
$$

or, equivalently (and more compactly), $y = (y_1, y_2, .., y_N)'$.

- Similarly, for the explanatory variable, we define $x = (x_1, x_2, .., x_N)'$.

- Likelihood function can be obtained by using basic rules of probability and the regression assumptions.

- $p(y_i|\beta, \sigma^2)$ is Normal (see Appendix B, Theorem B.10).

- $E(y_i|\beta, \sigma^2) = \beta x_i$ (see Appendix B, Theorem B.2).

- $var(y_i|\beta, \sigma^2) = \sigma^2$ (see Appendix B, Theorem B.2).

- Likelihood function is given by:

$$p(y|\beta, \sigma^2) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma^N} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \beta x_i)^2\right].$$

- Likelihood can be written in a slightly different way using:

$$\sum_{i=1}^{N} (y_i - \beta x_i)^2 = \nu s^2 + \left(\beta - \widehat{\beta}\right)^2 \sum_{i=1}^{N} x_i^2,$$

where

$$\nu = N - 1,$$

$$\widehat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

and

$$s^2 = \frac{\sum \left(y_i - \widehat{\beta} x_i\right)^2}{\nu}.$$

- Aside: $\widehat{\beta}$, s^2 and ν are the ordinary least squares (OLS) estimator for β, standard error and degrees of freedom, respectively. They are also sufficient statistics.

- For many Bayesian derivations, it is easier to work with the error precision rather than the variance. The error precision is defined as: $h = \frac{1}{\sigma^2}$.

- Likelihood function is:

$$p(y|\beta, h) = \frac{1}{(2\pi)^{\frac{N}{2}}}$$
$$\left\{ h^{\frac{1}{2}} \exp\left[-\frac{h}{2} \left(\beta - \widehat{\beta}\right)^2 \sum_{i=1}^{N} x_i^2 \right] \right\} \left\{ h^{\frac{\nu}{2}} \exp\left[-\frac{h\nu}{2s^{-2}} \right] \right\}$$

- For future reference, note that the first term in curly brackets looks like the kernel of a Normal density for β and the second term looks almost like a Gamma density for h.

2.2 The Prior

- Priors can be any form. However, it is common to choose particular classes of priors which are easy to interpret and/or make computation easier.

- Conjugate prior: prior and posterior both have same class of distributions.

- Prior can be interpreted as arising from a fictitious data set from the same process which generated the actual data.

- We must elicit a prior for β and h which we denote by $p(\beta, h)$.

- Laws of probability imply $p(\beta, h) = p(\beta|h)p(h)$

- The form of the likelihood function suggests conjugate prior has a Normal distribution for $\beta|h$ and a Gamma distribution for h.

- The *Normal-Gamma* distribution is discussed in Appendix B.

$$\beta|h \sim N(\underline{\beta}, h^{-1}\underline{V})$$

and

$$h \sim G(\underline{s}^{-2}, \underline{\nu})$$

then the natural conjugate prior for β and h is denoted by:

$$\beta, h \sim NG\left(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}\right).$$

- Researcher chooses particular values of *prior hyper-parameters* $\underline{\beta}, \underline{V}, \underline{s}^{-2}$ and $\underline{\nu}$ to reflect prior information.

2.3   The Posterior

- Posterior proportional to likelihood times the prior density.

- Very similar derivations in Exercises 2.4 and 2.5 of Bayesian Econometric Methods

- Posterior density is also of Normal-Gamma form. So the prior is a natural conjugate one.

-
$$\beta, h | y \sim NG\left(\overline{\beta}, \overline{V}, \overline{s}^{-2}, \overline{\nu}\right),$$

where

$$\overline{V} = \frac{1}{\underline{V}^{-1} + \sum x_i^2},$$

$$\overline{\beta} = \overline{V}\left(\underline{V}^{-1}\underline{\beta} + \widehat{\beta}\sum x_i^2\right),$$

$$\overline{\nu} = \underline{\nu} + N$$

and $\overline{s}^{-2}$ is defined implicitly through

$$\overline{\nu s}^2 = \underline{\nu s}^2 + \nu s^2 + \frac{\left(\widehat{\beta} - \underline{\beta}\right)^2}{\underline{V} + \left(\frac{1}{\sum x_i^2}\right)}.$$

- Our notation is such that lower bars (e.g. $\underline{\beta}$) denote prior hyperparameters and upper bars (e.g. $\overline{\beta}$) posterior ones.

2.4 Properties of the Posterior

- What are $E\left(\beta|y\right)$ and $var\left(\beta|y\right)$?

- $p\left(\beta|y,h\right)$ is Normal (properties well-known). But we want $p(\beta|y)$.

- Exercise 2.8 of Bayesian Econometric Methods shows that the marginal posterior distribution for β is a t distribution

$$\beta|y \sim t\left(\overline{\beta}, \overline{s}^2\overline{V}, \overline{\nu}\right),$$

and it follows from the definition of the t distribution that

$$E(\beta|y) = \overline{\beta}$$

and

$$var\left(\beta|y\right) = \frac{\overline{\nu}s^2}{\overline{\nu} - 2}\overline{V}.$$

- Bayesian point estimate, $\overline{\beta}$, is a weighted average of the OLS estimate and the prior mean, $\underline{\beta}$.

- The weights are proportional to $\sum x_i^2$ and $\underline{V}^{-1}$, respectively. The latter of these reflects the confidence in the prior and data information. (Note: in frequentist econometrics $(\sum x_i^2)^{-1}$ is proportional to the variance of $\widehat{\beta}$.)

- Alternative intuition can be obtained by considering the simplest case where $x_i = 1$ for $i = 1, ..., N$. Then $\sum x_i^2 = N$ and the weight attached to $\widehat{\beta}$ will simply be the sample size, a reasonable measure for the amount of information in the data.

- In frequentist econometrics, the variance of the OLS estimator is $s^2(\sum x_i^2)^{-1}$ (e.g. the frequentist t-statistic for testing $\beta = 0$ is $\dfrac{\widehat{\beta}}{\sqrt{s^2(\sum x_i^2)^{-1}}}$).

- The Bayesian analogue is the posterior variance of β has very similar form, but incorporates both prior and data information.

- "posterior precision is an average of prior precision $(\underline{V}^{-1})$ and data precision $(\sum x_i^2)$"

- Natural conjugate prior implies that the prior can be interpreted as arising from a fictitious data set (e.g. $\underline{\nu}$ and N play the same role in equations. See Exercise 2.6 of Bayesian Econometric Methods)

- $\underline{\nu}$ can be interpreted as a prior sample size.

- These facts can be used in prior elicitation (i.e. choosing particular values for $\underline{\beta}, \underline{V}, \underline{s}^{-2}$ and $\underline{\nu}$)

- *Prior sensitivity analysis.* Empirical results can be presented using various priors. If empirical results are basically the same for various sensible priors, then the reader is reassured that researchers with different beliefs can, after looking at the data, come to agreement. If results are sensitive to choice of prior, then the data is not enough to force agreement on researchers with different prior views. The Bayesian approach allows for the scientifically honest finding of such a state of affairs.

2.5   Posterior Analysis Using a Noninfor-

   mative Prior

- A noninformative prior can be obtained by setting $\underline{\nu} = 0$ and $\underline{V}^{-1} = 0$ (i.e. $\underline{V} \to \infty$). Prior no longer enters posterior – just contains likelihood information.

- In this case $\beta, h|y \sim NG\left(\overline{\beta}, \overline{V}, \overline{s}^{-2}, \overline{\nu}\right)$, where

$$\overline{V} = \frac{1}{\sum x_i^2},$$

$$\overline{\beta} = \widehat{\beta},$$

$$\overline{\nu} = N$$

and

$$\overline{\nu s}^2 = \nu s^2.$$

- These are OLS results.

- But different interpretation: Bayesian says β is random variable. Frequentist says $\widehat{\beta}$ random variable.

- Aside: this prior "density" does not integrate to one. Such priors are referred to as *improper*.

- There are some issues relating to the use of improper priors which we will discuss later.

3 Model Comparison

- Two models, M_1 and M_2.

- M_j for $j = 1, 2$ have different explanatory variables:

$$y_i = \beta_j x_{ji} + \varepsilon_{ji},$$

- Normal-Gamma natural conjugate priors as:

$$\beta_j, h_j | M_j \sim NG\left(\underline{\beta}_j, \underline{V}_j, \underline{s}_j^{-2}, \underline{\nu}_j\right),$$

which implies posteriors of the form:

$$\beta_j, h_j | y, M_j \sim NG\left(\overline{\beta}_j, \overline{V}_j, \overline{s}_j^{-2}, \overline{\nu}_j\right),$$

- Remember the posterior odds ratio:

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

- The prior model probabilities, $p(M_i)$, selected before seeing the data. The noninformative choice, $p(M_1) = p(M_2) = \frac{1}{2}$, is commonly made.

- The marginal likelihood, $p(y|M_j)$, is calculated as:

$$p\left(y|M_j\right) = \int\int p(y|\beta_j, h_j)p\left(\beta_j, h_j\right)d\beta_j dh_j.$$

- This can be calculated analytically

$$p(y|M_j) = c_j \left(\frac{\overline{V}_j}{\underline{V}_j}\right)^{\frac{1}{2}} \left(\overline{\nu}_j \overline{s}_j^2\right)^{-\frac{\overline{\nu}_j}{2}},$$

for $j = 1, 2$, where c_j is a constant which does not depend on the data (see textbook, page 41 for details).

- The posterior odds ratio comparing M_1 to M_2 becomes:

$$PO_{12} = \frac{c_1 \left(\frac{\overline{V}_1}{\underline{V}_1}\right)^{\frac{1}{2}} \left(\overline{\nu}_1 \overline{s}_1^2\right)^{-\frac{\overline{\nu}_1}{2}} p(M_1)}{c_2 \left(\frac{\overline{V}_2}{\underline{V}_2}\right)^{\frac{1}{2}} \left(\overline{\nu}_2 \overline{s}_2^2\right)^{-\frac{\overline{\nu}_2}{2}} p(M_2)}.$$

- What factors enter a Bayesian comparison of models?

- First, the greater is the prior odds ratio, $\frac{p(M_1)}{p(M_2)}$, the higher the support for M_1.

- $\overline{\nu}_j \overline{s}_j^2$ contains the term $\nu_j s_j^2$ which is the sum of squared errors (a measure of model fit).

- Other things being equal, the posterior odds ratio will indicate support for the model where there is the greatest coherency between prior and data information (i.e. $\left(\widehat{\beta}_j - \underline{\beta}_j\right)^2$ enters $\overline{\nu}_j \overline{s}_j^2$).

- $\left(\dfrac{\overline{V}_1}{\underline{V}_1}\right)$ is the ratio of posterior to prior variances. This term can be interpreted as saying, all else being equal, the model with more prior information (i.e. smaller prior variance) relative to posterior information receives most support.

- With noninformative prior the marginal likelihood is not defined and, hence, the posterior odds ratio is undefined (we will return to this issue – use of noninformative priors for model comparison — later).

4 Prediction

- We want to predict y^* at a point x^*:

$$y^* = \beta x^* + \varepsilon^*,$$

where y^* is not observed.

$$p\left(y^*|y\right) = \int \int p\left(y^*|y, \beta, h\right) p(\beta, h|y) d\beta dh.$$

- Exercise 7.7 of Bayesian Econometric Methods shows that:

$$y^*|y \sim t\left(\overline{\beta} x^*, \overline{s}^2 \left\{1 + \overline{V} x^{*2}\right\}, \overline{\nu}\right).$$

These results can be used to provide point predictions and measures of uncertainty associated with the point prediction (e.g. the predictive standard deviation).

- Aside: this is a logical place to introduce an important Bayesian concept: model averaging.

- Suppose you have $p(M_j|y)$, for $j = 1, 2$.

- *Bayesian model averaging* involves averaging over all models.

- Rules of probability imply:

$$p\left(y^*|y\right) = p\left(y^*|y, M_1\right)p\left(M_1|y\right) + p\left(y^*|y, M_2\right)p\left(M_2|y\right).$$

- You should not simply choose one model and work with, e.g., $p\left(y^*|y, M_1\right)$, but rather average results over the two models with weights given by the posterior model probabilities.

5   Empirical Illustration

- Artificial data with $N = 50$, $\beta = 2$ and $h = 1$.

- Two priors, the noninformative one and the informative one with $\underline{\beta} = 1.5$, $\underline{V} = .25$, $\underline{\nu} = 10$ and $\underline{s}^{-2} = 1$. (choice of prior hyperparameter values are purely illustrative).

Table 2.1: Prior and Posterior Properties of β			
	Prior	Posterior	
	Informative	Using Noninf Prior	Using Inf Prior
Mean	1.50	2.06	1.96
St. Dev	0.56	0.24	0.22

Table 2.2: Prior and Posterior Properties of h			
	Prior	Posterior	
	Informative	Using Noninf Prior	Using Inf Prior
Mean	1.00	1.07	1.04
St. Dev	0.45	0.21	0.19

Figure 2.1: Marginal Prior and Posteriors for β
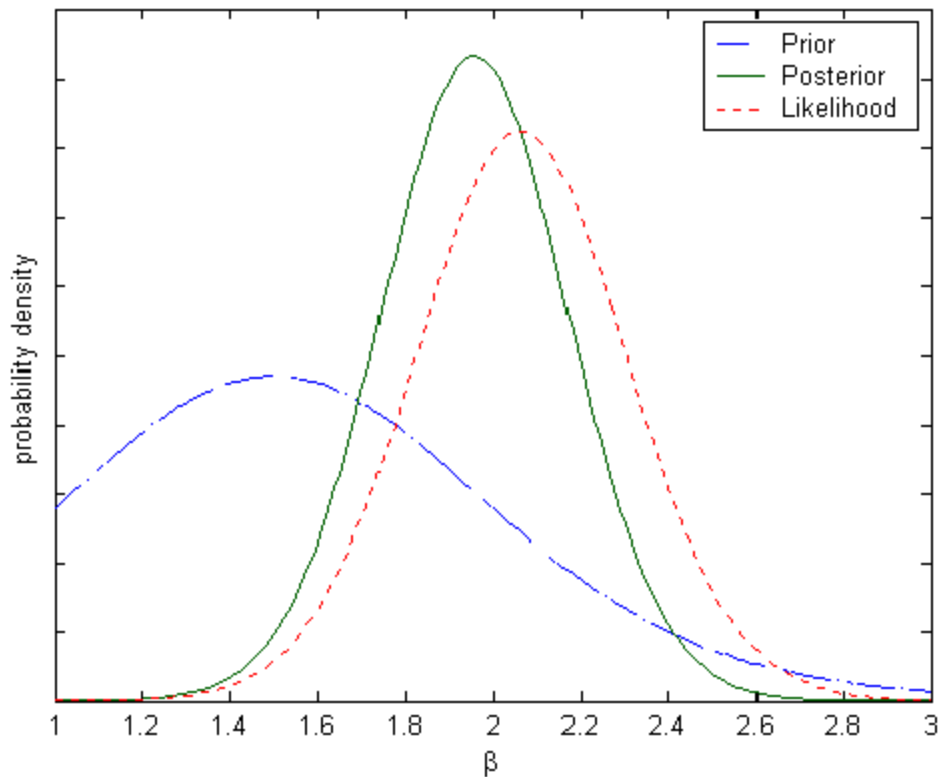
- Bayesian inference involves combining prior and data information to form a posterior.

- To illustrate posterior odds, compare to another linear regression model which contains only an intercept (i.e. in this second model $x_i = 1$ for $i = 1, .., 50$).

- For both models, we use the same informative prior described above

- Posterior odds ratio of $3,749.7$. (overwhelming support for the correct first model).

- The posterior odds ratio implies that $p(M_1|y) = 0.9997$ and $p(M_2|y) = 0.0003$.

- Bayesian model averaging would attach **99.97%** weight to results from the first model and only **0.03%** weight to results from the second.

- Predictive inference at the $x^* = 0.5$. Using the informative prior

$$y^*|y \sim t\,(0.98, 0.97, 60)\,.$$

Using the noninformative prior:

$$y^*|y \sim t\left(1.03, 0.95, 50\right).$$

6 The Normal Linear Regression Model with Natural Conjugate Prior and Many Explanatory Variables

6.1 The Linear Regression Model in Matrix Notation

- Now assume k explanatory variables, $x_{i1},..,x_{ik}$ for $i = 1, .., N$ and regression model:

$$y_i = \beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i.$$

- Note x_{i1} is implicitly set to 1 to allow for an intercept.

- In matrix notation we have $N \times 1$ vectors

$$
y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_N \end{bmatrix}
$$

and

$$
\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_N \end{bmatrix},
$$

the $k \times 1$ vector

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_k \end{bmatrix},$$

and the $N \times k$ matrix

$$X = \begin{bmatrix} 1 & x_{12} & . & . & x_{1k} \\ 1 & x_{22} & . & . & x_{2k} \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & x_{N2} & . & . & x_{Nk} \end{bmatrix},$$

and the regression model can be written as:

$$y = X\beta + \varepsilon.$$

6.2   The Likelihood Function

- The likelihood can be derived under the classical assumptions:

1. ε has a multivariate Normal distribution with mean 0_N and covariance matrix $\sigma^2 I_N$. Notation for this is: ε is $N(0_N, h^{-1}I_N)$ where $h = \sigma^{-2}$.

2. All elements of X are either fixed (i.e. not random variables).

- Exercise 10.1 for Bayesian Econometric Methods shows that the likelihood function can be written in terms of OLS quantities:

$$\nu = N - k,$$

$$\widehat{\beta} = \left(X'X\right)^{-1} X'y$$

and

$$s^2 = \frac{\left(y - X\widehat{\beta}\right)' \left(y - X\widehat{\beta}\right)}{\nu}.$$

- Likelihood function:

$$p(y|\beta, h) = \frac{1}{(2\pi)^{\frac{N}{2}}}$$

$$\left\{ h^{\frac{1}{2}} \exp\left[-\frac{h}{2}\left(\beta - \widehat{\beta}\right)' X'X \left(\beta - \widehat{\beta}\right)\right] \right\} \left\{ h^{\frac{\nu}{2}} \exp\left[-\frac{h\nu}{2s^{-2}}\right] \right\} .$$

6.3   The Prior

- Normal-Gamma prior is the same as before, except β conditional on h is now multivariate Normal:

$$\beta|h \sim N(\underline{\beta}, h^{-1}\underline{V})$$

and a prior for h of the form

$$h \sim G(\underline{s}^{-2}, \underline{\nu}),$$

- Same as before except that $\underline{\beta}$ is now a $k-$vector containing the prior means for the k regression coefficients, $\beta_1, .., \beta_k$, and $\underline{V}$ is now a $k \times k$ positive definite prior covariance matrix.

6.4   The Posterior

- The posterior is derived by multiplying the likelihood by the prior and collecting terms (see Bayesian Econometrics Methods Exercise 10.1). All similar to before, except matrices used.

- Posterior is

$$\beta, h|y \sim NG\left(\overline{\beta}, \overline{V}, \overline{s}^{-2}, \overline{\nu}\right),$$

where

$$\overline{V} = \left(\underline{V}^{-1} + X'X\right)^{-1},$$

$$\overline{\beta} = \overline{V}\left(\underline{V}^{-1}\underline{\beta} + X'X\widehat{\beta}\right),$$

$$\overline{\nu} = \underline{\nu} + N$$

and $\overline{s}^{-2}$ is defined implicitly through

$$\overline{\nu s}^2 = \underline{\nu s}^2 + \nu s^2 + \left(\widehat{\beta} - \underline{\beta}\right)' \left[\underline{V} + \left(X'X\right)^{-1}\right]^{-1} \left(\widehat{\beta} - \underline{\beta}\right).$$

- Marginal posterior for β is a multivariate t distribution:

$$\beta | y \sim t \left(\overline{\beta}, \overline{s}^2 \overline{V}, \overline{\nu}\right),$$

- Useful results for estimation:

$$E(\beta | y) = \overline{\beta}$$

and

$$var(\beta|y) = \frac{\overline{\nu}\overline{s}^2}{\overline{\nu} - 2}\overline{V}.$$

- Similar intuition as before, except now matrices. For instance, $\widehat{\beta}$ is now a vector instead of a scalar, the matrix $(X'X)^{-1}$ plays the role that the scalar $\frac{1}{\sum x_i^2}$ did in simple regression model, $\overline{V}$ is now a $k \times k$ matrix, etc.. $\widehat{\beta}$ is a matrix weighted average of information in the prior $(\underline{V}^{-1})$ and the data.

6.5   A Noninformative Prior

- As before we can create a noninformative prior by setting $\underline{\nu} = 0$ and setting $\underline{V}^{-1}$ to a small value.

- But there is not a unique way of doing the latter (see Exercise 10.4 in Bayesian Econometric Methods).

- A common way is to set $\underline{V}^{-1} = cI_k$ where c is a scalar and then let c go to zero.

- This noninformative prior is improper and can be written as:

$$p(\beta, h) \propto \frac{1}{h}.$$

- With this choice we get OLS results.

$$\beta, h|y \sim NG\left(\overline{\beta}, \overline{V}, \overline{s}^{-2}, \overline{\nu}\right)$$

where

$$\overline{V} = \left(X'X\right)^{-1},$$

$$\overline{\beta} = \widehat{\beta},$$

$$\overline{\nu} = N$$

and

$$\overline{\nu s}^2 = \nu s^2.$$

7 Model Comparison

7.1 Model Comparison Involving Inequality Restrictions

- Consider inequality restrictions of the form:

$$R\beta \geq r,$$

where R is a known $J \times k$ matrix and r is a known J-vector.

- Two models of the form:

$$M_1 : R\beta \geq r$$

and

$$M_2 : R\beta \ngeq r,$$

where the notation in the equation defining M_2 means that one or more of the J inequality restrictions in M_1 are violated.

- Posterior odds ratios is easy, and the use of noninformative priors is not a problem. That is,

$$PO_{12} = \frac{p(M_1|y)}{p(M_2|y)} = \frac{p(R\beta \geq r|y)}{p(R\beta \ngeq r|y)}.$$

Since the posterior for β has a multivariate t distribution, it follows that $p(R\beta|y)$ also has a t distribution.

7.2   Equality Restrictions

- Case 1: M_1 which imposes $R\beta = r$ to M_2 which does not have this restriction (nested).

- Case 2: $M_1 : y = X_1\beta_{(1)} + \varepsilon_1$ to $M_2 : y = X_2\beta_{(2)} + \varepsilon_2$, where X_1 and X_2 are matrices containing possibly different explanatory variables (non-nested).

- Both cases can be handled by defining models as (for $j = 1, 2$):

$$M_j : y_j = X_j\beta_{(j)} + \varepsilon_j$$

- Non-nested model comparison involves $y_1 = y_2$.

- Nested model comparison defines M_2 as the unrestricted regression. M_1 imposes $R\beta = r$ which involves a redefinition of explanatory and dependent variable (see textbook page 40 for details).

- Posterior odds can be calculated by calculating marginal likelihoods for the two regression models (where j subscripts denote models).

- Marginal likelihood is:

$$p(y_j|M_j) = c_j \left(\frac{|\overline{V}_j|}{|\underline{V}_j|} \right)^{\frac{1}{2}} \left(\overline{\nu}_j \overline{s}_j^2 \right)^{-\frac{\overline{\nu}_j}{2}},$$

for $j = 1, 2$, where

$$c_j = \frac{\Gamma \left(\frac{\overline{\nu}_j}{2} \right) \left(\underline{\nu}_j \underline{s}_j^2 \right)^{\frac{\underline{\nu}_j}{2}}}{\Gamma \left(\frac{\underline{\nu}_j}{2} \right) \pi^{\frac{N}{2}}}.$$

- The posterior odds ratio comparing M_1 to M_2 is:

$$PO_{12} = \frac{c_1 \left(\frac{|\overline{V}_1|}{|\underline{V}_1|}\right)^{\frac{1}{2}} \left(\overline{\nu}_1 \overline{s}_1^2\right)^{-\frac{\overline{\nu}_1}{2}} p(M_1)}{c_2 \left(\frac{|\overline{V}_2|}{|\underline{V}_2|}\right)^{\frac{1}{2}} \left(\overline{\nu}_2 \overline{s}_2^2\right)^{-\frac{\overline{\nu}_2}{2}} p(M_2)}.$$

- The posterior odds ratio depends on the prior odds ratio and contains rewards for model fit, coherency between prior and data information and parsimony.

7.3 Model Comparison with Noninformative Priors

- Important rule of thumb: *When comparing models using posterior odds ratios, it is acceptable to use noninformative priors over parameters which are common to all models. However, informative, proper priors should be used over all other parameters.*

- Why?

- If we set $\underline{\nu}_1 = \underline{\nu}_2 = 0$. Posterior odds ratio still has a sensible interpretation involving model fit, the coherency between prior and data information, etc.. In short, using a noninformative prior for the error precisions in the two models is perfectly reasonable.

- But, using noninformative priors for the $\beta_{(j)}$'s causes major problems which occur largely when $k_1 \neq k_2$. (Exercise 10.4 of Bayesian Econometric Methods)

- Briefly, noninformative prior for $\beta_{(j)}$ based on $\underline{V}_j^{-1} = cI_{k_j}$ and letting $c \to 0$. Since $|\underline{V}_j| = \frac{1}{c^{k_j}}$ terms involving k_j do not cancel out. If $k_1 < k_2$, PO_{12} becomes infinite, while if $k_1 > k_2$, PO_{12} goes to zero (regardless of data).

7.4 Highest Posterior Density Intervals

- Loosely analogous to confidence intervals and, thus, sometimes used for testing point null hypotheses.

- General definition in textbook, here define for β_j a single regression coefficient.

- A 95% credible interval for β_j is any interval, $[a, b]$ such that:

$$p\left(a \leq \beta_j \leq b|y\right) = 0.95.$$

- There are numerous possible credible intervals. E.g. if $\beta_j|y$ is $N(0, 1)$. Then, $[-1.96, 1.96]$, $[-1.75, 2.33]$ and $[-1.64, \infty)$, etc. are all 95% credible intervals.

- A **95%** highest posterior density interval is a **95%** credible interval with the property that it has a smaller area than any other **95%** credible interval.

- In example, $[-1.96, 1.96]$ is the shortest credible interval.

- HPDIs exist any time the posterior exists. Thus, they can be used with the noninformative prior discussed.

7.5   Prediction

- Want to predict:

$$y^* = X^*\beta + \varepsilon^*$$

- As before, prediction is based on:

$$p\left(y^*|y\right) = \int \int p\left(y^*|y, \beta, h\right) p(\beta, h|y) d\beta dh.$$

- The resulting predictive:

$$y^*|y \sim t\left(X^*\overline{\beta}, \overline{s}^2\left\{I_T + X^*\overline{V}X^{*\prime}\right\}, \overline{\nu}\right).$$

8 Computational Methods: Monte Carlo Integration

- Model comparison, prediction and posterior inference about β can all be done analytically. So no need for posterior simulation in this model. However, let us illustrate Monte Carlo integration in this model.

- *Theorem 3.1: Monte Carlo Integration*

Let $\beta^{(s)}$ for $s = 1, .., S$ be a random sample from $p(\beta|y)$ and $g\left(.\right)$ be any function and define

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

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

- Outline of computer code:

Step 1: Take a random draw, $\beta^{(s)}$ from the posterior for β using a random number generator for the multivariate t distribution.

Step 2: Calculate $g\left(\beta^{(s)}\right)$ and keep this result.

Step 3: Repeat Steps 1 and 2 S times.

Step 4: Take the average of the S draws $g\left(\beta^{(1)}\right), ..., g\left(\beta^{(S)}\right)$.

These steps will yield an estimate of $E\left[g(\beta)|y\right]$ for any function of interest.

- Remember: Monte Carlo integration yields only an approximation for $E\left[g(\beta)|y\right]$ (since you cannot set $S = \infty$). However, by selecting S, the researcher can control the degree of approximation error.

- Using a CLT we can obtain an approximate 95% confidence interval for $E[g(\beta)|y]$ of the form

$$\left[\widehat{g}_S - 1.96\frac{\widehat{\sigma}_g}{\sqrt{S}}, \widehat{g}_S + 1.96\frac{\widehat{\sigma}_g}{\sqrt{S}}\right]. \text{ where } \sigma_g^2 = var(g(\beta)|y).$$

- Alternatively, the numerical standard error, $\frac{\widehat{\sigma}_g}{\sqrt{S}}$, can be reported as implicitly containing the same information.

9 Empirical Illustration

- Data set on $N = 546$ houses sold in Windsor, Canada in 1987.

- y_i = sales price of the ith house measured in Canadian dollars,

- x_{i2} = the lot size of the ith house measured in square feet,

- x_{i3} = the number of bedrooms in the ith house,

- x_{i4} = the number of bathrooms in the ith house,

- x_{i5} = the number of storeys in the ith house.

- Example uses informative and noninformative priors.

- Textbook goes through detailed discussion of how you might elicit a prior for this example.

- Our prior implies statements of the form "if we compare two houses which are identical except the first house has one bedroom more than the second, then we expect the first house to be worth $\$5,000$ more than the second". This yields prior mean, then choose large prior variance to indicate prior uncertainty.

- The following tables present some empirical results (textbook has lots of discussion of how you would interpret them).

	Prior	Posterior	
	Informative	Using Noninf Prior	Using Inf Prior
β_1	0 (10,000)	$-4,009.55$ (3,593.16)	$-4,035.05$ (3,530.16)
β_2	10 (5)	5.43 (0.37)	5.43 (0.37)
β_3	5,000 (2,500)	2,824.61 (1,211.45)	2,886.81 (1,184.93)
β_4	10,000 (5,000)	17,105.17 (1,729.65)	16,965.24 (1,708.02)
β_5	10,000 (5,000)	7,634.90 (1,005.19)	7,641.23 (997.02)

Table 3.1: Prior and Posterior Means for β (standard deviations in parentheses)

Table 3.3: Model Comparison involving β			
Informative Prior			
	$p\left(\beta_j > 0|y\right)$	95% HPDI	Posterior Odds for $\beta_j = 0$
β_1	0.13	$[-10,957, 2,887]$	4.14
β_2	1.00	$[4.71, 6.15]$	2.25×10^{-39}
β_3	0.99	$[563.5, 5,210.1]$	0.39
β_4	1.00	$[13,616, 20,314]$	1.72×10^{-19}
β_5	1.00	$[5,686, 9,596]$	1.22×10^{-11}
Noninformative Prior			
	$p\left(\beta_j > 0|y\right)$	95% HPDI	Posterior Odds for $\beta_j = 0$
β_1	0.13	$[-11,055, 3,036]$	—
β_2	1.00	$[4.71, 6.15]$	—
β_3	0.99	$[449.3, 5,200]$	—
β_4	1.00	$[13,714, 20,497]$	—
β_5	1.00	$[5,664, 9,606]$	—

Table 3.4: Posterior Results for β_2 Calc. Various Ways			
	Mean	Standard Deviation	Numerical St. Error
Analytical	5.4316	0.3662	—
Number of Reps			
$S = 10$	5.3234	0.2889	0.0913
$S = 100$	5.4877	0.4011	0.0401
$S = 1,000$	5.4209	0.3727	0.0118
$S = 10,000$	5.4330	0.3677	0.0037
$S = 100,000$	5.4323	0.3664	0.0012