A Course in Bayesian Econometrics

University of Queensland

July, 2008

Slides for Lecture on

The Linear Regression Model with General Error Covariance Matrix

Gary Koop, University of Strathclyde

1 Summary

- Readings: Chapter 6 of textbook. I will cover the general theory and three special cases: the regression model with autocorrelated errors, Student-t errors and the seemingly unrelated regressions (SUR) model.

- The textbook discusses heteroskedasticity (but I will not cover this here)

- All fall into the class where, conditional on Ω (to be defined shortly), the model becomes a Normal linear regression model.

- Can draw on results from previous lecture for $p(\beta, h|y, \Omega)$.

- So, if we knew Ω, we could do Bayesian inference.

- But, in practice, Ω will be unknown. How to proceed? Use Gibbs sampling.

1.1 Bayesian Computation: The Gibbs Sampler

- The Gibbs sampler is a powerful tool for posterior simulation which is used in many econometric models.

- Bayesian Econometric Methods, Exercises 11.6 through 11.16 all relate to Gibbs sampling.

- We will motivate the basic ideas in a very general context before returning to the regression model.

- General notation: θ is a $p-$vector of parameters and $p\left(y|\theta\right), p\left(\theta\right)$ and $p\left(\theta|y\right)$ are the likelihood, prior and posterior, respectively.

- Let θ be partitioned into various *blocks* as $\theta = \left(\theta'_{(1)}, \theta'_{(2)}, .., \theta'_{(B)}\right)'$ where $\theta_{(j)}$ is a scalar or vector, $j = 1, 2, .., B$.

- E.g. in regression model, $B = 2$ with $\theta_{(1)} = \beta$ and $\theta_{(2)} = h$.

- Intuition: i) Monte Carlo integration takes draws from $p(\theta|y)$ and averages them to produce estimates of $E[g(\theta)|y]$ for any function of interest $g(\theta)$.

- ii) In many models, it is not easy to directly draw from $p(\theta|y)$. However, it often is easy to randomly draw from

$$p\left(\theta_{(1)}|y, \theta_{(2)}, .., \theta_{(B)}\right), p\left(\theta_{(2)}|y, \theta_{(1)}, \theta_{(3)}.., \theta_{(B)}\right), ..., \\ p\left(\theta_{(B)}|y, \theta_{(1)}, .., \theta_{(B-1)}\right).$$

- Note: Preceding distributions are referred to as *full conditional posterior distributions* since they define a posterior for each block conditional on all the other blocks.

- iii) Drawing from the full conditionals will yield a sequence $\theta^{(1)}, \theta^{(2)}, .., \theta^{(s)}$ which can be averaged to produce estimates of $E\left[g\left(\theta\right)|y\right]$ in the same manner as Monte Carlo integration did.

1.1.1 More motivation for the Gibbs sampler

- Let $B = 2$ and suppose you have one random draw from $p\left(\theta_{(2)}|y\right)$. Call this draw $\theta_{(2)}^{(0)}$.

- Since $p\left(\theta|y\right) = p\left(\theta_{(1)}|y, \theta_{(2)}\right) p\left(\theta_{(2)}|y\right)$, it follows that a random draw from $p\left(\theta_{(1)}|y, \theta_{(2)}^{(0)}\right)$ is a valid draw of $\theta_{(1)}$ from $p\left(\theta|y\right)$. Call this draw $\theta_{(1)}^{(1)}$.

- Since $p\left(\theta|y\right) = p\left(\theta_{(2)}|y, \theta_{(1)}\right) p\left(\theta_{(1)}|y\right)$, it follows that a random draw from $p\left(\theta_{(2)}|y, \theta_{(1)}^{(1)}\right)$ is a valid draw of $\theta_{(2)}$ from $p\left(\theta|y\right)$.

- Hence, $\theta^{(1)} = \left(\theta_{(1)}^{(1)\prime}, \theta_{(2)}^{(1)\prime}\right)^{\prime}$ is a valid draw from $p\left(\theta|y\right)$.

- You can continue this reasoning indefinitely.

- Hence, if you can successfully find $\theta_{(2)}^{(0)}$, then sequentially drawing from the posterior of $\theta_{(1)}$ conditional on the previous draw for $\theta_{(2)}$, then $\theta_{(2)}$ given the previous draw for $\theta_{(1)}$, will yield a sequence of draws from the posterior.

- This strategy of sequentially drawing from full conditional posterior distributions is called Gibbs sampling.

- Problem with steps above is that it is not possible to find such an initial draw $\theta_{(2)}^{(0)}$. (if we knew how to easily take random draws from $p\left(\theta_{(2)}|y\right)$, we could use this and $p\left(\theta_{(1)}|\theta_{(2)},y\right)$ to do Monte Carlo integration and have no need for Gibbs sampling.

- However, subject to weak conditions, the initial draw $\theta_{(2)}^{(0)}$ does not matter in the sense that the Gibbs sampler will converge to a sequence of draws from $p\left(\theta|y\right)$.

- In practice, choose $\theta_{(2)}^{(0)}$ in some manner and then run the Gibbs sampler for S replications. However, the first S_0 of these are discarded as so-called *burn-in replications* and the remaining S_1 retained for the estimate of $E\left[g\left(\theta\right)|y\right]$, where $S_0 + S_1 = S$.

- After dropping the first S_0 of these to eliminate the effect of $\theta^{(0)}$, remaining S_1 draws can be averaged to create estimates of posterior features of interest. That is, if

$$\widehat{g}_{S_1} = \frac{1}{S_1} \sum_{s=S_0+1}^{S} g\left(\theta^{(s)}\right),$$

then $\widehat{g}_{S_1}$ converges to $E\left[g(\theta)|y\right]$ as S_1 goes to infinity.

- There are various "MCMC Diagnostics" which you can use to make sure you have taken enough draws (and discarded enough burn-in draws). See textbook pages 64-68.

- Gibbs sampler popular since many models logically break into blocks. Many posteriors can be written as $p\left(\beta,h,z|y\right)$ where z is something else (often a vector of latent data). Gibbs sampling involving $p\left(\beta,h|y,z\right)$ and $p\left(z|y,\beta,h\right)$ can be used (where $p\left(\beta,h|y,z\right)$ uses results for linear regression model).

- Examples: tobit, probit, stochastic frontier model, random effects panel data model, SUR, error correction models, state space models, threshold autoregressive models, Markov switching models, some semiparametric regression models, etc. etc. etc.

2 The Model with General Ω

- Now return to regression model:

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

- Before we assumed ε was $N(0_N, h^{-1}I_N)$.

- Now we will assume:

$$\varepsilon \sim N(0_N, h^{-1}\Omega).$$

where Ω is an $N \times N$ positive definite matrix.

- Many models can be put in this form (including random effects panel data models, SUR models, ARMA models and the ones we will discuss below).

- Appendix A, Theorem A.10 says that an $N \times N$ matrix P exists with the property that $P\Omega P' = I_N$.

- Multiply both sides of the regression model by P:

$$y^{\dagger} = X^{\dagger}\beta + \varepsilon^{\dagger},$$

where $y^{\dagger} = Py$, $X^{\dagger} = PX$ and $\varepsilon^{\dagger} = P\varepsilon$.

- It can be verified that $\varepsilon^{\dagger}$ is $N(0_N, h^{-1}I_N)$.

- Hence, the transformed model is identical to the Normal linear regression model.

- If Ω is known, Bayesian analysis of the Normal linear regression model with non-scalar error covariance matrix is straightforward (simply work with transformed model).

- If Ω is unknown, often can use Gibbs sampling

- For instance, if the prior for β and h is $NG\left(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}\right)$, then all the results of previous lecture are applicable *conditional upon* Ω.

- E.g. $p(\beta|y, \Omega)$ is a multivariate t distribution and this, combined with a posterior simulator for $p(\Omega|y, \beta)$ can be used to set up a Gibbs sampler.

- Note: what if $p(\Omega|y, \beta, h)$ does not have a convenient form to draw from? Metropolis-Hastings algorithms are popular (see pages 92-99 of textbooks). "Metropolis-within-Gibbs" algorithms popular.

2.1 Posterior Inference in General Case

- In last lecture, we used a *natural conjugate* Normal-Gamma prior.

- To illustrate another prior we will use an *independent* Normal-Gamma prior for β and h

- At this stage use general notation, $p(\Omega)$, to indicate the prior for Ω.

- Thus prior used is

$$p(\beta, h, \Omega) = p(\beta)\, p(h)\, p(\Omega)$$

where

$$p(\beta) = f_N\left(\beta|\underline{\beta}, \underline{V}\right)$$

and

$$p(h) = f_G\left(h|\underline{\nu}, \underline{s}^{-2}\right).$$

- Exercise 13.1 of Bayesian Econometric Methods show that posterior conditionals are (in terms of transformed model):

$$\beta|y, h, \Omega \sim N\left(\overline{\beta}, \overline{V}\right),$$

where

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

and

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

$$h|y, \beta, \Omega \sim G(\overline{s}^{-2}, \overline{\nu}),$$

where $\widehat{\beta}\left(\Omega\right)$ is the GLS estimator

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

and

$$\overline{s}^2 = \frac{\left(y - X\beta\right)' \Omega^{-1} \left(y - X\beta\right) + \underline{\nu s}^2}{\overline{\nu}}.$$

The posterior for Ω conditional on β and h has a kernel of the form:

$$p\left(\Omega|y,\beta,h\right)\propto$$
$$p\left(\Omega\right)\left|\Omega\right|^{-\frac{1}{2}}\left\{\exp\left[-\frac{h}{2}\left(y-X\beta\right)'\Omega^{-1}\left(y-X\beta\right)\right]\right\}. \tag{*}$$

- In general, this conditional posterior does not take any easily recognized form. Note that, if we could take posterior draws from $p\left(\Omega|y,\beta,h\right)$, then a Gibbs sampler for this model could be set up in a straightforward manner since $p\left(\beta|y,h,\Omega\right)$ is Normal and $p\left(h|y,\beta,\Omega\right)$ is Gamma.

3 Heteroskedasticity of an Unknown Form: Student-t Errors

- It turns out that we have heteroskedasticity of an unknown form in the Normal linear regression model it is equivalent to a regression model with Student-t errors.

- This is a simple example of a *mixture model*.

- Mixture models are very popular right now in many fields as a way of making models more flexible (e.g. non-Normal errors, “nonparametric” treatment of regression line, etc.).

- Heteroskedasticity occurs if:

$$\Omega = \begin{bmatrix} \omega_1 & 0 & . & . & 0 \\ 0 & \omega_2 & 0 & . & . \\ . & 0 & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & \omega_N \end{bmatrix}$$

- In other words, $var(\varepsilon_i) = h^{-1}\omega_i$ for $i = 1, .., N$.

- With N observations and $N+k+1$ parameters to estimate (i.e. β, h and $\omega = (\omega_1, .., \omega_N)'$), treatment of heteroskedasticity of unknown form may sound like a difficult task.

- Solution: use a *hierarchical prior* (ω_is drawn from some common distribution – parameters of that distribution estimated from the data).

- Hierarchical priors are commonly used as a way of making flexible, parameter-rich models more amenable to statistical analysis.

- Allows us to free up the assumption of Normal errors that we have used so far.

3.1 A Hierarchical Prior for the Error Variances

- We begin by eliciting $p(\omega)$.

- Work with error precisions rather than variances and, hence, we define $\lambda \equiv (\lambda_1, \lambda_2, .., \lambda_N)'$

$\equiv \left(\omega_1^{-1}, \omega_2^{-1}, .., \omega_N^{-1}\right)'$.

- Consider the following prior for λ:

$$p(\lambda) = \prod_{i=1}^{N} f_G(\lambda_i | 1, \nu_\lambda). \qquad (**)$$

Note f_G is the Gamma p.d.f.

- The prior for λ depends on a hyperparameter, ν_λ, and assumes each λ_i comes from the same distribution.

- In other words, λ_is are i.i.d. draws from the Gamma distribution.

- This assumption (or something similar) is necessary to deal with the problems caused by the high-dimensionality of λ.

- Why should the λ_is be i.i.d. draws from the Gamma distribution with mean 1.0? This model is *exactly the same* as the linear regression model with i.i.d. Student-t errors with ν_λ degrees of freedom (Bayesian Econometric Methods Exercise 15.1)..

- In other words, if we had begun by assuming:

$$p(\varepsilon_i) = f_t\left(\varepsilon_i|0, h^{-1}, \nu_\lambda\right)$$

for $i = 1, .., N$, we would have ended up with exactly the same posterior.

- Note: we now have model with more flexible error distribution, but we are still our familiar Normal linear regression model framework.

- Chapter 10 of textbook discusses several ways of making models more flexible: *mixture of Normals* distributions. Our treatment of heteroskedasticity is *scale mixture of Normals*.

- If ν_λ is unknown, need a prior $p(\nu_\lambda)$.

- Note that now the prior for λ is specified in two steps, the first being (**), the other being $p(\nu_\lambda)$. Alternatively, the prior for λ can be written as $p(\lambda|\nu_\lambda)\,p(\nu_\lambda)$. Priors written in two (or more) steps in this way are referred to as hierarchical priors.

- See discussion of $p(\nu_\lambda)$ in textbook pages 126-127.

3.2 Bayesian Computation with Student-t Model

- Geweke (1993, JAE) develops a Gibbs sampler for taking draws of the parameters in the model: β, h, λ and ν_λ.

- $p(\beta|y, h, \lambda)$ and $p(h|y, \beta, \lambda)$ are as discussed in last week.

- Focus on $p(\lambda|y, \beta, h, \nu_\lambda)$ and $p(\nu_\lambda|y, \beta, h, \lambda)$.

- Bayesian Econometric Methods, Exercise 15.1 derives posterior conditionals for λ_is as

$$p(\lambda_i|y, \beta, h, \nu_\lambda) = f_G\left(\lambda_i|\frac{\nu_\lambda + 1}{h\varepsilon_i^2 + \nu_\lambda}, \nu_\lambda + 1\right).$$

- $p(\nu_\lambda|y,\beta,h,\lambda)$ depends on $p(\nu_\lambda)$. Geweke uses the exponential density which is simply the Gamma with two degrees of freedom:

$$p(\nu_\lambda) = f_G(\nu_\lambda|\underline{\nu}_\lambda, 2).$$

$$p(\nu_\lambda|y,\beta,h,\lambda) \propto \left(\frac{\nu_\lambda}{2}\right)^{\frac{N\nu_\lambda}{2}} \Gamma\left(\frac{\nu_\lambda}{2}\right)^{-N} \exp(-\eta\nu_\lambda),$$

where

$$\eta = \frac{1}{\underline{\nu}_\lambda} + \frac{1}{2}\sum_{i=1}^{N}\left[\ln\left(\lambda_i^{-1}\right) + \lambda_i\right]$$

- Geweke derives a method of drawing from this density (thus completing the Gibbs sampler). My textbook treatment slightly different.

4 Autocorrelated Errors

- Assume errors in a regression model follow an *autoregressive process of order 1* or *AR(1)* process:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t,$$

where u_t is i.i.d. $N\left(0, h^{-1}\right)$ and $-1 < \rho < 1$.

- Using standard results from time series we can write covariance matrix of ε as $h^{-1}\Omega$, where

$$\Omega = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & . & \rho^{T-1} \\ \rho & 1 & \rho & . & . \\ \rho^2 & \rho & . & . & \rho^2 \\ . & . & . & . & \rho \\ \rho^{T-1} & . & \rho^2 & \rho & 1 \end{bmatrix}.$$

- Thus, the regression model with AR(1) errors falls into the class of regression models with General Error Covariance Matrix.

- Extension to AR(p) errors is straightforward. Extension to ARMA(p,q) errors also (relatively) straightforward.

- Assuming independent Normal-Gamma prior for regression part, then Gibbs sampler can be set up involving $p\left(\Omega|y,\beta,h\right)$, $p\left(\beta|y,h,\Omega\right)$ and $p\left(h|y,\beta,\Omega\right)$.

4.1 Bayesian Computation in Regression Model with AR Errors

- Same idea as for all models in this chapter: $p\left(\beta|y,h,\Omega\right)$ and $p\left(h|y,\beta,\Omega\right)$ have familiar forms (Normal and Gamma) and we need only focus on $p\left(\Omega|y,\beta,h\right) = p\left(\rho|y,\beta,h\right)$.

- To motivate results, write the regression model as:

$$y_t = x_t\beta + \varepsilon_t$$

where x_t is a scalar.

- Defining $y_t^\dagger = y_t - \rho y_{t-1}$ and $x_t^\dagger = x_t - \rho x_{t-1}$ we obtain:

$$y_t^\dagger = x_t^\dagger\beta + u_t.$$

- We have assumed that u_t is i.id. $N\left(0, h^{-1}\right)$. This transformed model is simply a Normal linear regression model with i.i.d. errors.

- Aside: treatment of initial condition.

- Prior for ρ can be anything, here assume Normal, truncated to the stationary region. That is,

$$p(\rho) \propto f_N\left(\rho|\underline{\rho}, \underline{V}_\rho\right) 1\left(\rho \in \Phi\right),$$

where $1\left(\rho \in \Phi\right)$ is the indicator function which equals 1 for the stationary region and zero otherwise.

- Intuition for $p(\rho|y, \beta, h)$. Conditional on β, can use

$$\varepsilon_t = y_t - x_t\beta,$$

to get ε_t. But then the AR(1) equation:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t,$$

is just like a regression model.

- Using standard regression derivations we have:

$$p\left(\rho|y,\beta,h\right) \propto f_N\left(\rho|\overline{\rho},\overline{V}_\rho\right) 1\left(\rho \in \Phi\right),$$

where

$$\overline{V}_\rho = \left(\underline{V}_\rho^{-1} + hE'E\right)^{-1},$$

$$\overline{\rho} = \overline{V}\rho \left(\underline{V}_{\rho}^{-1} \underline{\rho} + hE'\varepsilon \right)$$

and E is a $(T-p) \times k$ matrix with t^{th} row given by $\left(\varepsilon_{t-1}, .., \varepsilon_{t-p} \right)$.

- Exercise 13.4 of Bayesian Econometric Methods gives exact derivations (and an empirical application).

- Key thing: Gibbs sampler involves drawing from full conditional posteriors: $p(\beta|y,h,\rho)$ and $p(h|y,\beta,\rho)$ and $p(\rho|y,\beta,h)$. All of these have forms the computer can easily draw from.

- Remember, once you have S_1 Gibbs sampling draws (discarding S_0 burn-in draws), you can simply average them to produce any feature of interest you want.

- For instance if β_j is a regression coefficient

$$\frac{1}{S_1} \sum_{s=S_0+1}^{S} \beta_j{}^{(s)},$$

converges to $E\left(\beta_j|y\right)$, a popular point estimate.

$$\frac{1}{S_1} \sum_{s=S_0+1}^{S} \left(\beta_j{}^{(s)}\right)^2,$$

converges to $E\left(\beta_j^2|y\right)$, which can be used to calculate $var\left(\beta_j|y\right)$ (i.e. $var\left(\beta_j|y\right) = E\left(\beta_j^2|y\right) - \left[E\left(\beta_j|y\right)\right]^2$).

etc. etc. etc.

4.2 Prediction Using the Gibbs Sampler

- In last lecture we worked out that the predictive density for the Normal regression model with natural conjugate prior had t distribution. But in other cases predictive density may not have convenient form.

- Gibbs sampling can be used. The strategy below works with any Gibbs sampler, but let me illustrate with regression model with the independent Normal-Gamma prior (for simplicity set $\Omega = I$).

- Want to predict T unobserved values of the dependent variable $y^* = \left(y_1^*, .., y_T^*\right)'$, which are generated according to:

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

- The predictive density is $p\left(y^*|y\right)$ but cannot be derived analytically.

- But we do know:

$$p\left(y^{*}|\beta,h\right)=\frac{h^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}}\exp\left[-\frac{h}{2}\left(y^{*}-X^{*}\beta\right)'\left(y^{*}-X^{*}\beta\right)\right].$$

- Predictive features of interest can be written as $E\left[g\left(y^{*}\right)|y\right]$ for some function $g\left(.\right)$.

- E.g. Predictive mean of y_{i}^{*} implies $g\left(y^{*}\right)=y_{i}^{*}$,

- But, using same reasoning as for Monte Carlo integration, if we can find $y^{*(s)}$ for $s=1,..,S$ which are draws from $p\left(y^{*}|y\right)$, then

$$\widehat{g}_{Y}=\frac{1}{S}\sum_{s=1}^{S}g\left(y^{*(s)}\right),$$

will converge to $E\left[g\left(y^{*}\right)|y\right]$.

- The following strategy will provide the required draws of y^*.

- For every $\beta^{(s)}$ and $h^{(s)}$ provided by the Gibbs sampler, take a draw, $y^{*(s)}$ from $p\left(y^*|\beta^{(s)}, h^{(s)}\right)$ (a Normal density)

- We now have draws $\beta^{(s)}$, $h^{(s)}$ and $y^{*(s)}$ for $s = 1, .., S$ which we can use for posterior or predictive inference.

- Why are these the correct draws? Simply use rules of conditional probability (see pages 72-73 of textbook for details).

5 The Seemingly Unrelated Regressions Model

- Seemingly unrelated regressions (SUR) are multiple equation models:

$$y_{mi} = x'_{mi}\beta_m + \varepsilon_{mi},$$

with $i = 1, .., N$ observations for $m = 1, .., M$ equations.

- y_{mi} is the i^{th} observation on the dependent variable in equation m, x_{mi} is a k_m-vector containing the i^{th} observation of the vector of explanatory variables in the m^{th} equation and β_m is a k_m-vector of regression coefficients for the m^{th} equation.

- SUR model can be written using matrices in a familiar form.

- Stack all equations into vectors/matrices as $y_i = (y_{1i}, .., y_{Mi})'$, $\varepsilon_i = (\varepsilon_{1i}, .., \varepsilon_{Mi})'$,

$$\beta = \begin{pmatrix} \beta_1 \\ . \\ . \\ \beta_M \end{pmatrix},$$

$$X_i = \begin{pmatrix} x'_{1i} & 0 & . & . & 0 \\ 0 & x'_{2i} & 0 & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & x'_{Mi} \end{pmatrix}.$$

and define $k = \sum_{m=1}^{M} k_m$.

- SUR model can be written as:

$$y_i = X_i \beta + \varepsilon_i.$$

- Stack all the observations together as:

$$y = \begin{pmatrix} y_1 \\ . \\ . \\ y_N \end{pmatrix},$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ . \\ . \\ \varepsilon_N \end{pmatrix},$$

$$X = \begin{pmatrix} X_1 \\ . \\ . \\ X_N \end{pmatrix}$$

and write

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

- Thus, the SUR model can be written as our familiar linear regression model.

- If we were to assume ε_{mi} to be i.i.d. $N\left(0, h^{-1}\right)$ for all i and m, then we would simply have the Normal linear regression model of Chapters 2, 3 and 4.

- However, it is common for the errors to be correlated across equations and, thus, we assume ε_i to be i.i.d. $N\left(0, H^{-1}\right)$ for $i = 1, .., N$ where H is an $M \times M$ error precision matrix.

- Thus, ε is $N(0, \Omega)$ where Ω is an $NM \times NM$ block-diagonal matrix given by:

$$\Omega = \begin{pmatrix} H^{-1} & 0 & . & . & 0 \\ 0 & H^{-1} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & H^{-1} \end{pmatrix}.$$

- Hence, the SUR model lies in the class of models being studied in this lecture.

5.1 Bayesian Inference in the SUR Model

- Any prior can be used, here we use a popular one which is an extended version of our familiar independent Normal-Gamma prior.

- The independent Normal-Wishart prior:

$$p(\beta, H) = p(\beta)\, p(H)$$

where

$$p(\beta) = f_N\left(\beta|\underline{\beta}, \underline{V}\right)$$

and

$$p(H) = f_W\left(H|\underline{\nu}, \underline{H}\right).$$

- The Wishart distribution, which is a matrix generalization of the Gamma distribution, is defined/discussed in Appendix B, Definition B.27 of textbook.

- Bayesian computation involves a Gibbs sampler using following posterior conditionals:

$$\beta|y, H \sim N\left(\overline{\beta}, \overline{V}\right),$$

where formula for $\overline{\beta}, \overline{V}$ are on page 140 of textbook.

- And the posterior for H conditional on β is Wishart:

$$H|y, \beta \sim W\left(\overline{\nu}, \overline{H}\right)$$

where

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

and

$$\overline{H} = \left[\underline{H}^{-1} + \sum_{i=1}^{N} \left(y_i - X_i \beta \right) \left(y_i - X_i \beta \right)' \right]^{-1}.$$

- Empirical illustration provided in textbook.