This joint posterior density for β and h does not take the form of any well-known and understood density and, hence, cannot be directly used in a simple way for posterior inference. For instance, a researcher might be interested in presenting the posterior mean and variance of β. Unfortunately, there is not a simple analytical formula for these posterior features which can be written down and, hence, posterior simulation is required.

If we treat (4.3) as a joint posterior density for β and h, it does not take a convenient form. The conditionals of the posterior are, however, simple. That is, $p(\beta|y, h)$ can be obtained by treating (4.3) as a function of β for a fixed value of h.[1] If we do matrix manipulations similar to those used in derivation of the posterior for the natural conjugate prior, we can write the key term in the first line of (4.3) as:

$$
\begin{aligned}
&h(y - X\beta)'(y - X\beta) + (\beta - \underline{\beta})'\underline{V}^{-1}(\beta - \underline{\beta}) \\
&= (\beta - \overline{\beta})'\overline{V}^{-1}(\beta - \overline{\beta}) + Q
\end{aligned}
$$

where

$$\overline{V} = (\underline{V}^{-1} + hX'X)^{-1} \tag{4.4}$$

$$\overline{\beta} = \overline{V}(\underline{V}^{-1}\underline{\beta} + hX'y) \tag{4.5}$$

and

$$Q = hy'y + \underline{\beta}'\underline{V}^{-1}\underline{\beta} - \overline{\beta}'\overline{V}^{-1}\overline{\beta}$$

Plugging this expression into (4.3) and ignoring the terms that do not involve β (including Q), we can write

$$p(\beta|y, h) \propto \exp\left[-\frac{1}{2}(\beta - \overline{\beta})'\overline{V}^{-1}(\beta - \overline{\beta})\right] \tag{4.6}$$

which is the kernel of a multivariate Normal density. In other words,

$$\beta|y, h \sim N(\overline{\beta}, \overline{V}) \tag{4.7}$$

$p(h|y, \beta)$ is obtained by treating (4.3) as a function of h. It can be seen that

$$p(h|y, \beta) \propto h^{\frac{N+\underline{\nu}-2}{2}} \exp\left[-\frac{h}{2}\left\{(y - X\beta)'(y - X\beta) + \underline{\nu}\underline{s}^2\right\}\right]$$

By comparing this with the definition of the Gamma density (see Appendix B, Definition B.22) it can be verified that

$$h|y, \beta \sim G(\overline{s}^{-2}, \overline{\nu}) \tag{4.8}$$

[1]Formally, the rules of probability imply $p(\beta|y, h) = \frac{p(\beta, h|y)}{p(h|y)}$. However, since $p(h|y)$ does not depend upon β, $p(\beta, h|y)$ gives the kernel of $p(\beta|y, h)$. Since a density is defined by its kernel, examination of the form of $p(\beta, h|y)$, treating h as fixed, will tell us what $p(\beta|y, h)$ is.

function estimation, the restriction that increasing an input will increase output is typically an inequality constraint of the form $\beta_j > 0$. In this chapter, we show how such constraints can be imposed through the prior and a technique called *importance sampling* used to carry out Bayesian inference.

The likelihood function used with both these priors will be the same as that used in the previous chapter. Hence, unlike previous chapters, we will not have a separate section discussing the likelihood function. The reader is referred to (3.3)–(3.7) for a reminder of what it looks like.

4.2 THE NORMAL LINEAR REGRESSION MODEL WITH INDEPENDENT NORMAL-GAMMA PRIOR

4.2.1 The Prior

The Normal linear regression model is defined in Chapter 3 (see (3.2)–(3.7)), and depends upon the parameters β and h. In that chapter, we used a natural conjugate prior where $p(\beta|h)$ was a Normal density and $p(h)$ a Gamma density. In this section, we use a similar prior, but one which assumes prior independence between β and h. In particular, we assume $p(\beta, h) = p(\beta)p(h)$ with $p(\beta)$ being Normal and $p(h)$ being Gamma:

$$p(\beta) = \frac{1}{(2\pi)^{\frac{k}{2}}} |\underline{V}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\beta - \underline{\beta})'\underline{V}^{-1}(\beta - \underline{\beta})\right] \tag{4.1}$$

and

$$p(h) = c_G^{-1} h^{\frac{\underline{\nu}-2}{2}} \exp\left(-\frac{h\underline{\nu}}{2\underline{s}^{-2}}\right) \tag{4.2}$$

where c_G is the integrating constant for the Gamma p.d.f. given in Appendix B, Definition B.22. For simplicity, we are using the same notation as in the previous chapters. That is, $\underline{\beta} = E(\beta|y)$ is still the prior mean of β and the prior mean and degrees of freedom of h are still $\underline{s}^{-2}$ and $\underline{\nu}$, respectively. However, be careful to note that $\underline{V}$ is now simply the prior covariance matrix of β, whereas in the previous chapter we had $var(\beta|h) = h^{-1}\underline{V}$.

4.2.2 The Posterior

The posterior is proportional to the prior times the likelihood. Hence, if we multiply (3.2), (4.1) and (4.2) and ignore terms that do not depend upon β and h, we obtain:

$$p(\beta, h|y) \propto \left\{\exp\left[-\frac{1}{2}\left\{h(y - X\beta)'(y - X\beta) + (\beta - \underline{\beta})'\underline{V}^{-1}(\beta - \underline{\beta})\right\}\right]\right\}$$
$$h^{\frac{N+\underline{\nu}-2}{2}} \exp\left[-\frac{h\underline{\nu}}{2\underline{s}^{-2}}\right] \tag{4.3}$$

where

$$\overline{\nu} = N + \underline{\nu} \tag{4.9}$$

and

$$\overline{s}^2 = \frac{(y - X\beta)'(y - X\beta) + \underline{\nu}\underline{s}^2}{\overline{\nu}} \tag{4.10}$$

These formulae look quite similar to those for the Normal linear regression model with natural conjugate prior (compare with (3.9)–(3.13)). Indeed, at an informal level, the intuition for how the posterior combines data and prior information is quite similar. However, it must be stressed that (4.4)–(4.10) do not relate directly to the posterior of interest, $p(\beta, h|y)$, but rather to the conditional posteriors, $p(\beta|y, h)$ and $p(h|y, \beta)$. Since $p(\beta, h|y) \neq p(\beta|y, h)p(h|y, \beta)$, the conditional posteriors in (4.7) and (4.8) do not directly tell us everything about $p(\beta, h|y)$. Nevertheless, there is a posterior simulator, called the *Gibbs sampler*, which uses conditional posteriors like (4.7) and (4.8) to produce random draws, $\beta^{(s)}$ and $h^{(s)}$ for $s = 1, \ldots, S$, which can be averaged to produce estimates of posterior properties just as with Monte Carlo integration.

4.2.3 Bayesian Computation: The Gibbs Sampler

The Gibbs sampler is a powerful tool for posterior simulation which is used in many econometric models. We will motivate the basic ideas in a very general context before returning to the Normal linear regression model with independent Normal-Gamma prior. Accordingly, let us temporarily adopt the general notation of Chapter 1, where θ is a p-vector of parameters and $p(y|\theta)$, $p(\theta)$ and $p(\theta|y)$ are the likelihood, prior and posterior, respectively. In the linear regression model, $p = k + 1$ and $\theta = (\beta', h)'$. Furthermore, let θ be partitioned into various *blocks* as $\theta = (\theta'_{(1)}, \theta'_{(2)}, \ldots, \theta'_{(B)})'$, where $\theta_{(j)}$ is a scalar or vector, $j = 1, 2, \ldots, B$. In the linear regression model, it is convenient to set $B = 2$ with $\theta_{(1)} = \beta$ and $\theta_{(2)} = h$.

Remember that Monte Carlo integration involves taking random draws from $p(\theta|y)$ and then averaging them to produce estimates of $E[g(\theta)|y]$ for any function of interest $g(\theta)$ (see Chapter 3, Theorem 3.1). In many models, including the one discussed in the present chapter, it is not easy to directly draw from $p(\theta|y)$. However, it often is easy to randomly draw from $p(\theta_{(1)}|y, \theta_{(2)}, \ldots, \theta_{(B)})$, $p(\theta_{(2)}|y, \theta_{(1)}, \theta_{(3)}, \ldots, \theta_{(B)}), \ldots, p(\theta_{(B)}|y, \theta_{(1)}, \ldots, \theta_{(B-1)})$. The preceding distributions are referred to as *full conditional posterior distributions*, since they define a posterior for each block conditional on all the other blocks. In the Normal linear regression model with independent Normal-Gamma prior $p(\beta|y, h)$ is Normal and $p(h|y, \beta)$ is Gamma, both of which are simple to draw from. It turns out that drawing from the full conditionals will yield a sequence $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(s)}$ which can be averaged to produce estimates of $E[g(\theta)|y]$ in the same manner as did Monte Carlo integration.