will converge to $E[g(\theta)|y]$ as S increases.[4] This line of reasoning suggests that, if we can find $y^{*(s)}$ for $s = 1, \ldots, S$ which are draws from $p(y^*|y)$, then

$$\widehat{g}_Y = \frac{1}{S}\sum_{s=1}^{S} g(y^{*(s)}) \tag{4.32}$$

will converge to $E[g(y^*)|y]$. This is indeed the case.

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 $p(y^*|y, \beta^{(s)}, h^{(s)})$. Since the latter density is Normal (see (4.29)), such a strategy is quite simple. We now have $\beta^{(s)}$, $h^{(s)}$ and $y^{*(s)}$ for $s = 1, \ldots, S$. The rules of probability say that $p(\beta, h, y^*|y) = p(y^*|y, \beta, h)p(\beta, h|y)$ and, hence, the strategy of first drawing from the posterior, then drawing from $p(y^*|y, \beta, h)$ will yield draws from $p(\beta, h, y^*|y)$. Hence, our set of draws $\beta^{(s)}$, $h^{(s)}$ and $y^{*(s)}$ thus created can be used to evaluate any posterior feature of interest using (4.11) and any predictive feature of interest using (4.32).[5]

The strategy outlined in this section can be used for any model where a posterior simulator is used to provide draws from $p(\theta|y)$ and $p(y^*|y, \theta)$ has a form that is easy to work with. Almost all of the models discussed in the remainder of this book fall in this category. Hence, in future chapters you will often see a very brief discussion of prediction, including a sentence of the form 'Predictive inference in this model can be carried out using the strategy outlined in Chapter 4'.

4.2.7 Empirical Illustration

We use the house price data set introduced in Chapter 3 to illustrate the use of Gibbs sampling in the Normal linear regression model with independent Normal-Gamma prior. The reader is referred to Chapter 3 (Section 3.9) for a precise description of the dependent and explanatory variables for this data set. Section 3.9 discusses prior elicitation using a natural conjugate prior. This discussion implies that sensible values for the hyperparameters of the independent Normal-Gamma prior would be $\underline{\nu} = 5$ and $\underline{s}^{-2} = 4.0 \times 10^{-8}$, and

$$\underline{\beta} = \begin{bmatrix} 0.0 \\ 10 \\ 5000 \\ 10\,000 \\ 10\,000 \end{bmatrix}$$

[4] As discussed above, with Gibbs sampling you may wish to omit some initial burn-in draws and, hence, the summation would go from $S_0 + 1$ through S.

[5] This result uses the general rule that, if we have draws from the joint density $p(\theta, y^*|y)$, then the draws of θ considered alone are draws from the marginal $p(\theta|y)$ and the draws of y^* considered alone are draws from $p(y^*|y)$.

These values are identical to those used in the previous chapter, and have the same interpretation. However, we stress that $\underline{V}$ has a different interpretation in this chapter than the previous one. With the independent Normal-Gamma prior we have

$$var(\beta) = \underline{V}$$

while with the natural conjugate prior we had

$$var(\beta) = \frac{\underline{\nu}\underline{s}^2}{\underline{\nu} - 2}\underline{V}$$

Accordingly, to have a prior comparable to that used in the previous chapter, we set

$$\underline{V} = \begin{bmatrix} 10\,000^2 & 0 & 0 & 0 & 0 \\ 0 & 5^2 & 0 & 0 & 0 \\ 0 & 0 & 2500^2 & 0 & 0 \\ 0 & 0 & 0 & 5000^2 & 0 \\ 0 & 0 & 0 & 0 & 5000^2 \end{bmatrix}$$

Note that, with the independent Normal-Gamma prior, it is usually easy to elicit $\underline{V}$, since it is simply the prior variance of β. With the natural conjugate prior, the prior dependence between β and h means that the prior variance of β depends upon the prior you have elicited for h as well as $\underline{V}$.

Bayesian inference in this model can be done using Gibbs sampling. Most common Bayesian computer software (e.g. Jim LeSage's Econometrics Toolbox or BACC; see Section 1.3 of Chapter 1) allows for a thorough analysis of this model. The interested reader, at this stage, may wish to download and use this software. Alternatively, for the reader with some knowledge of computer programming, writing your own programs is a simple option. The website associated with this book contains MATLAB code for such a program (although for the MCMC convergence diagnostics a function from Jim LeSage's Econometrics Toolbox has been used). The structure of this program is very similar to the Monte Carlo integration program of the previous chapter, although it sequentially draws from $p(\beta|y, h)$ and $p(h|y, \beta)$, instead of simply drawing from $p(\beta|y)$.

Table 4.1 contains empirical results relating to β, including MCMC convergence diagnostics, for the Normal linear regression model with the independent Normal-Gamma prior specified above. We set the initial draw for the error precision to be equal to the inverse of the OLS estimate of σ^2 (i.e. $h^{(0)} = \frac{1}{s^2}$). We discard an initial $S_0 = 1000$ burn-in replications and include $S_1 = 10\,000$ replications. For the sake of brevity, we do not present results for h.

The posterior means and standard deviations are similar to those in Table 3.1, reflecting the fact that we have used similarly informative priors in the two chapters. The column labelled 'NSE' contains numerical standard errors for the

Table 4.1 Prior and Posterior Results for β (standard deviations in parentheses)

	Prior	Posterior	NSE	Geweke's CD	Post. Odds for $\beta_j = 0$
β_1	0 (10 000)	−4063.08 (3259.00)	28.50	−0.68	1.39
β_2	10 (5)	5.44 (0.37)	0.0029	0.11	6.69×10^{-42}
β_3	5000 (2500)	3214.09 (1057.67)	12.45	−0.57	0.18
β_4	10 000 (5000)	16 132.78 (1617.34)	15.56	0.55	2.06×10^{-19}
β_5	10 000 (5000)	7680.50 (979.09)	8.44	1.22	3.43×10^{-12}

approximation of $E(\beta_j|y)$ for $j = 1, \ldots, 5$, calculated using (4.13).[6] They can be interpreted as in the previous chapter, and indicate that our estimates are quite accurate. Of course, if a higher degree of accuracy is desired, the researcher can increase S_1. The column labeled 'Geweke's CD' is described in (4.14), and compares the estimate of $E(\beta_j|y)$ based on the first 1000 replications (after the burn-in replications) to that based on the last 4000 replications. If the effect of the initial condition has vanished and an adequate number of draws have been taken, then these two estimates should be quite similar. Noting that CD is asymptotically standard Normal, a common rule is to conclude that convergence of the MCMC algorithm has occurred if CD is less than 1.96 in absolute value for all parameters. Using this rule, Table 4.1 indicates that convergence of the MCMC algorithm has been achieved.

Table 4.1 also contains posterior odds ratios comparing two regression models: $M_1 : \beta_j = 0$ to $M_2 : \beta_j \neq 0$. As in Chapter 3, the restricted model uses an informative prior, which is identical to the unrestricted prior, except that $\underline{\beta}$ and $\underline{V}$ become 4×1 and 4×4 matrices, respectively, with prior information relating to β_j omitted. A prior odds ratio of one is used. The model comparison information in Table 4.1 is qualitatively similar to that in Table 3.1. That is, there is overwhelming evidence that β_2, β_4 and β_5 are non-zero, but some uncertainty as to whether β_1 and β_3 are zero. Note, however, that, if we compare empirical

[6]For the reader who knows spectral methods, $S(0)$ is calculated using a 4% autocovariance tapered estimate.

results in Table 4.1 with those in Chapter 3, the model comparison results are more different than the posterior means. This is a common finding. That is, prior information tends to have a bigger effect on posterior odds ratios than it does on posterior means and standard deviations. Hence, the slight difference in prior between Chapters 3 and 4 reveals itself more strongly in posterior odds ratios than in posterior means.

Working out the predictive density of the price of a house with given characteristics can be done using the methods outlined in Section 4.2.6. As in the previous chapter, we consider the case where the researcher is interested in predicting the sales price of a house with a lot size of 5000 square feet, two bedrooms, two bathrooms and one storey. Unlike with the natural conjugate prior, with the independent Normal-Gamma prior analytical results for the predictive distribution are unavailable. Nevertheless, the properties of the predictive can be calculated by making minor modifications to our posterior simulation program. That is, if we add one line of code which takes a random draw of $y^{*(s)}$ conditional on $\beta^{(s)}$ and $h^{(s)}$ using (4.29) and save the resulting draws, $y^{*(s)}$ for $s = S_0 + 1, \ldots, S$, we can then calculate any predictive property we wish using (4.32). Using these methods, we find that the predictive mean of a house with the specified characteristics is \$69 750 and the predictive standard deviation is 18 402. As expected, the figures are quite close to those obtained in the previous chapter.

For one (or at most two) dimensional features of interest, graphical methods can be a quite effective way of presenting empirical results. Figure 4.1 presents a plot of the predictive density. This figure is simply a histogram of all the draws, $y^{*(s)}$ for $s = S_0 + 1, \ldots, S$. The approximation arises since a histogram

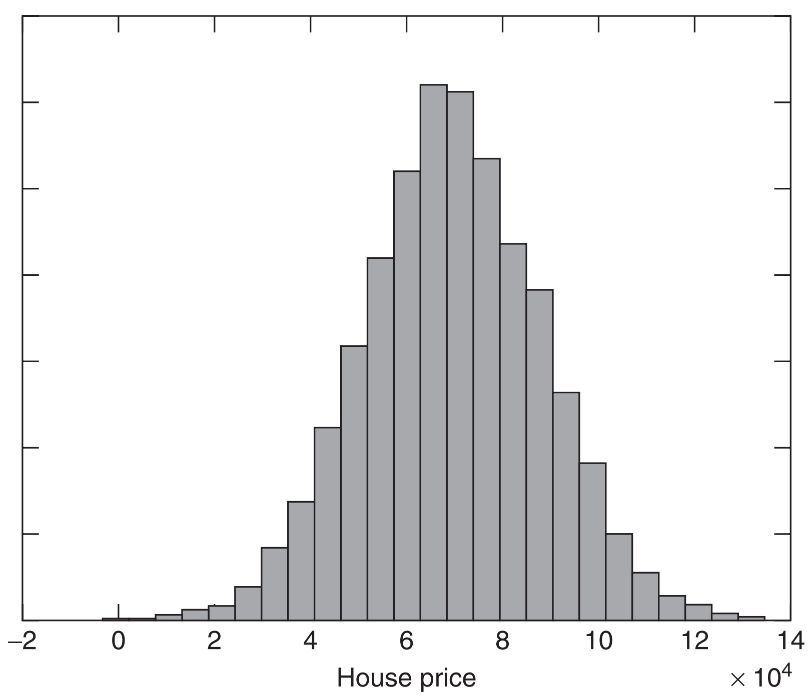

Figure 4.1 Predictive Density

is a discrete approximation to the continuous predictive density. This graph not only allows the reader to make a rough guess at the predictive mean, but also show the fatness of the tails of the predictive distribution. The graph shows that this data set does not allow for very precise predictive inference. Although the researcher's best prediction of the price of a house with a lot size of 5000 square feet, two bedrooms, two bathrooms and one storey is roughly \$70 000, the predictive allocates non-negligible probability to the house price being less than \$30 000 or more than \$110 000.

4.3 THE NORMAL LINEAR REGRESSION MODEL SUBJECT TO INEQUALITY CONSTRAINTS

In this section, we discuss imposing inequality constraints on the coefficients in the linear regression model. This is something that the researcher may often wish to do. For instance, it may be desirable to impose concavity or monotonicity on a production function. In a model with autocorrelated errors (see Chapter 6, Section 6.5) the researcher may wish to impose stationarity. All such cases can be written in the form $\beta \in A$, where A is the relevant region. Bayesian analysis of the regression model subject to such restrictions is quite simple since we can simply impose them through the prior. To carry out posterior inference, we use something called 'importance sampling'. It is worth noting that, for some types of inequality constraints (e.g. linear inequality constraints such as $\beta_j > 0$), slightly simpler methods of posterior analysis are available. However, importance sampling is reasonably simple and works for any type of inequality constraint. Furthermore, importance sampling is a powerful tool which can be used with a wide variety of models, not only those with inequality constraints. Hence, we introduce the concept of importance sampling here, in the context of the familiar regression model. However, we stress it is a useful tool that works with many models. We remind the reader that the likelihood function for this model is the familiar one given in (3.3) or (3.7).

4.3.1 The Prior

It is convenient to introduce inequality restrictions through the prior. That is, saying $\beta \in A$ is equivalent to saying that a region of the parameter space which is not within A is *a priori* impossible and, hence, should receive a prior weight of 0. Such prior information can be combined with any other prior information. For instance, we can combine it with an independent Normal-Gamma or a natural conjugate prior. Here we combine it with the natural conjugate prior given in (3.8). Remember that a special case of this is the noninformative prior given in (3.24). Such a noninformative prior is useful in the common case where the researcher wishes to impose an inequality constraint on β, but has no other prior information.