namely, if the main objective is forecasting, the appropriate model is that one with the smallest variance of the one-step-ahead predictive distribution, but if the main goal is to estimate the regression parameters, the appropriate model is the one with the "largest" precision matrix of the marginal posterior distribution of the regression parameters θ. The "largest" precision matrix will be explained later, so for now let us consider the predictive method.

The precision of the predictive distribution of one future observation w is calculated for each of the four models, then the most appropriate model is the one with the largest precision. Suppose one wants to predict y when $x = (1, x^*_1, x^*_2)$ and the Jeffreys' prior density is used, then one would substitute the appropriate values of α, β, p, X, Y, and Z into the precision

$$P(w|s) = (n + 2\alpha)A\,(C - B'A^{-1}B)^{-1},$$

where A, B, and C are given by (1.39). The following substitutions are:

Model 1:

$$\alpha = -\tfrac{1}{2},\ \beta = P = 0,\ x = (1, 1, \ldots, 1)',$$
$$y = (y_1, y_2, \ldots, y_n),\ z = 1.$$

Model 2:

$$\alpha = -\tfrac{2}{2},\ \beta = 0,\ P = 0(2 \times 2),\ x = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{12} \\ & \vdots \\ 1 & x_{1n} \end{pmatrix},$$

$$z = (1, x^*_1).$$

Model 3:

$$\alpha = -\tfrac{2}{2},\ \beta = 0,\ P = 0(2 \times 2),\ x = \begin{pmatrix} 1 & x_{21} \\ 1 & x_{22} \\ & \vdots \\ 1 & x_{2n} \end{pmatrix},$$

$$z = (1, x^*_2).$$

Model 4:

$$\alpha = -\tfrac{3}{2},\ \beta = 0,\ P = 0(3 \times 3),\ x = \begin{pmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ & \vdots & \\ 1 & x_{1n} & x_{2n} \end{pmatrix},$$

$$z = (1, x^*_1, x^*_2).$$

With this way to select the model, model 4 is usually chosen as the most appropriate model, thus if the precision of model 4 is only slightly larger than the precisions of models 2 or 3, then perhaps the final choice of model should be model 2, if the precision of model 3 is larger than that of 2. If model 3 has a larger precision than that of 2, then perhaps model 3 should be the choice of a final model. Obviously subjectivity is heavily involved in the choice of the model. To summarize the selection procedure, let $P_k(w|s)$ be the precision of the predictive distribution of one future observation w, predicted by model k, where k = 1, 2, 3, 4, then the appropriate model is model L, L = 1, 2, 3, 4, where

$$P_L(w|s) > \max_{k \neq L}\{P_k(w|s)\},$$

and $P_k(w|s)$ is calculated by the above substitution scheme.

If estimation is one's main goal in the regression analysis, one might choose the model which has the "largest" precision matrix in the marginal posterior distribution of the regression coefficients θ; however, "largest" precision matrix is vague and one must clarify the meaning of "large." In general the precision matrix of the marginal posterior distribution of θ is, from (1.22),

$$P(\theta|s) = (n - p)(x'x)(y'y - y'x\,(x'x)^{-1}\,x'y)^{-1}, \tag{1.22}$$

where p is the number of parameters in the model which is p = m + 1, where m is the number of independent variables in the model, and assuming a Jeffreys' prior density for the parameters. Thus one way to choose a model is that one with the largest $|P(\theta|s)|$.

The two criteria for choosing an appropriate model will not always produce the same model. Also, the two selection procedures were developed assuming a Jeffreys' improper prior density, but the general idea is easily extended to regression models with any number of independent variables and when the prior information of the parameters is expressed by a conjugate prior density.

when the prior information of the parameters is expressed by a conjugate prior density.

An Example with Two Regressor Variables

An example of a simple linear regression analysis was given in the previous section and it was shown how to calculate the posterior estimates of the regression coefficients and how to predict a future observation. Consider the following regression model with two regressors,

$$y_i = \theta_1 + \theta_2 x_{i1} + \theta_3 x_{i2} + e_i, \tag{3.55}$$

i = 1, 2, …, 30, x_{i1} =i, and x_{i2} = 1, 3, 1, 5, 0, 6, 5, 10, 7, 13, 4, 8, 17, 2, 10, 20, 15, 17, 22, 23, 15, 18, 30, 25, 10, 19, 22, 20, 25, and 36. The errors e_i and n.i.d. (0, 1) and θ_1 =1, θ_2 = 2, and θ_3 = 3.

Thirty Y values were generated from this model and are given in Table 3.2 and the first six observations (x_{i1}, x_{i2}, y_i), i = 1, 2, 3, 4, 5, 6, provided a way to fit the parameters μ, p (3 × 3), α and β to a normal-gamma prior density, where

$$\mu = (x^{*\prime}x^*)^{-1}x^{*\prime}z,$$
$$P = x^{*\prime}x^*/s^2,$$
$$s^2 = \frac{z'z - z'x^*(x^{*\prime}x^*)^{-1}x^{*\prime}z}{n-3},\ n = 6,$$
$$\alpha = 1$$

and

$$\beta = (s^2)^{-1}.$$

Also, $Z = (y_1, y_2, y_3, y_4, y_5, y_6,)'$ and

$$x^* = \begin{pmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & & \vdots \\ 1 & x_{16} & x_{26} \end{pmatrix}.$$

and using the formulas the values of the hyperparameters were calculated to be:

Table 3.2
Y Values for Multiple Linear Regression

y_1 = 5.70178	y_{16} = 80.3337
y_2 = 13.0713	y_{17} = 87.2295
y_3 = 10.6488	y_{18} = 106.561
y_4 = 22.5712	y_{19} = 109.683
y_5 = 12.31	y_{20} = 87.5413
y_6 = 12.31	y_{21} = 100.251
y_7 = 29.7727	y_{22} = 136.989
y_8 = 46.6751	y_{23} = 122.513
y_9 = 40.1657	y_{24} = 80.6753
y_{10} = 59.4161	y_{25} = 112.06
y_{11} = 36.6735	y_{26} = 120.948
y_{12} = 50.1047	y_{27} = 114.828
y_{13} = 77.4773	y_{28} = 135.531
y_{14} = 34.2187	y_{29} = 138.048
y_{15} = 94.3475	y_{30} = 138.903

$$p = \begin{pmatrix} 12.1431 & 42.5008 & 32.3815 \\ & 184.17 & 133.574 \\ \prime\prime & & 145.717 \end{pmatrix}, \tag{3.56}$$
$$\alpha = 1,$$
$$\mu = \begin{pmatrix} .26964 \\ 2.47146 \\ 2.6549 \end{pmatrix}$$

and

$\beta = .722657.$

Combining the prior density with parameters (3.56) and the likelihood function for μ, p, α and β given $S = \{(x_{i1}, x_{i2}, y_i): i = 7, \ldots, 29\}$ yields the joint posterior density of the parameters and the marginal posterior distribution of $\theta = (\theta_1, \theta_2, \theta_3)^1$ is a t with $n + 2\alpha$ (=25) degrees of freedom, location vector

$$E(\theta|s) = \begin{pmatrix} .951843 \\ 2.02176 \\ 2.98288 \end{pmatrix} \tag{3.57}$$

and precision matrix

$$P(\theta|s) = \begin{pmatrix} 27.3105 & 354.758 & 302.598 \\ & 6720.7 & 5722.41 \\ \text{''} & & 5387.58 \end{pmatrix}. \tag{3.58}$$

It follows that the marginal posterior distribution of is also a t with 25 degrees of freedom, location

$$E(\theta_1|s) = .951843 \tag{3.59}$$

and precision

$$P(\theta_1|s) = 8.5838 \tag{3.60}$$

As for θ_2, it has mean

$$E(\theta_2|s) = 2.02176 \tag{3.61}$$

precision

$$P(\theta_2|s) = 534.799 \tag{3.62}$$

and 25 degrees of freedom.

The parameters of the posterior distribution of θ_3 are

$$E(\theta_3|s) = 2.98288 \tag{3.63}$$

and

$$P(\theta_3|s) = 515.142 \tag{3.64}$$

In order to forecast Y_{30} when $x_{1,30} = 30$ and $x_{2,30} = 36$, the predictive distribution of Y_{30} is a univariate t with 25 degrees of freedom, location (1.39)

$$E(w|s) = 168.988 \tag{3.65}$$

and precision

$$P(w|s) = .597336 \tag{3.66}$$

The prior estimate μ, (3.56), of θ is close to θ_2 and θ_3 but not as close to θ_1 (=1); however, as expected, the posterior mean (3.57) of θ is quite close to the "true" value of θ; however, the posterior mean of θ_1 is not as close to $\theta_1 = 1$ as are the posterior means of θ_2 and θ_3 to $\theta_2 = 2$ and $\theta_3 = 3$ respectively.

HPD intervals for these parameters are easily found. For example, consider a 90% interval estimate of θ_1. The general formula is

$$E(\theta_1|s) \pm t_{n+2\alpha,.05}\sqrt{P^{-1}(\theta_1|s)} \tag{3.67}$$

or

$$.951843 \pm .58297$$

and a 90% HPD interval for θ_2 is 2.02176 ± .07385. What is the 90% HPD interval for θ_3 and how would one test for significance of regression ($\theta_1 = \theta_2 = \theta_3 = 0$)?

Suppose one desires to forecast a future value of Y_{30} (=138.903) given $x_{1,30} = 30$ and $x_{2,30} = 36$. We have seen a point estimate of Y_{30} is given by $E(w|s) = 168.988$ with a predictive precision of .597336 thus a 90% prediction interval for Y_{30} is

$$168.988 \pm 2.20993$$

and the actual observation of 138.048 lies quite far outside the prediction interval, because the predictive precision is very small.

and the actual observation of 138.048 lies quite far outside the prediction interval, because the predictive precision is very small.

This example was constructed so that and have the same "true" values as θ_1 and θ_2, the slope and intercept, respectively, of the simple linear regression model (3.41) and the first regressor in both assume the same values, namely $x_{i1} = i$, $i = 1, 2, \ldots, 30$. The addition of the second regressor affects the marginal posterior distributions of and θ_1. For instance $P(\theta_1 | s) = 5.35085$ with the simple linear regression model and a data-based prior, while $P(\theta_1 | s) = 8.5838$ with the regression containing two regressions and a data-based prior.

NONLINEAR REGRESSION ANALYSIS

Suppose

$$y_i = f(x_i, \theta) + e_i, \qquad i = 1, 2, \ldots, n \tag{3.68}$$

where x_i is a $m \times 1$ vector of independent variables, θ is a $p \times 1$ real unknown parameter vector, $x_i \in R^m$, $\theta \in R^p$, $f: R^m \times R^p \to R$ is a known function with domain $R^m \times R^p$ and range contained in R, and the e_i are n.i.d. $(0, \tau^{-1})$, where $\tau > 0$ and unknown. if

$$f(x, \theta) = x\theta, \qquad x \in R^m, \qquad \theta \in R^p \tag{3.69}$$

where x is a $1 \times m$ vector, then the model is the general linear model, but if $f(x, \theta)$ is not linear in θ, the model (3.68) is called a nonlinear regression model, and the mean of the dependent variable y is a nonlinear function of θ.

It has been shown that if f is linear in θ, then the Bayesian approach is well developed, however when f is nonlinear in θ, Bayesian inferences are more difficult to develop. One problem is the nonlinearity of f makes a conjugate family difficult to identify (except in some special cases). This is obvious from the likelihood function

$$L(\theta, \tau | s) \propto \tau^{n/2} \exp\left[-\frac{\tau}{2} \sum_{i=1}^{n} [y_i - f(x_i, \theta)]^2\right], \qquad \theta \in R^p, \quad \tau > 0, \tag{3.70}$$

where the sample is denoted by

$$s = \{x_{1i}, x_{2i}, \cdots x_{mi}, y_i), i = 1, 2, \ldots, m$$

and

$$x_i = \{x_{1i}, x_{2i}, \cdots x_{mi}).$$

If $p = 1$, there is an exact small-sample Bayesian analysis as follows. Suppose the prior density of θ and τ is

$$\xi(\theta, \tau) \propto \tau^{\alpha - 1} e^{-\tau\beta} \xi(\theta), \theta \in R, \tau > 0$$

then the posterior density of θ and τ is

$$\xi(\theta, \tau | s) \propto \xi(\theta) \tau^{(n+2\alpha)/2 - 1} \exp\left[-\frac{\tau}{2} \{\sum_{i=1}^{n} [y_i - f(x_i, \theta)]^2 + 2\beta\}\right], \qquad \theta \in R, \quad \tau > 0, \tag{3.71}$$

where $\xi(\theta)$ is a proper marginal prior density of θ and the marginal prior density of τ is gamma with parameters $\alpha > 0$ and $\beta > 0$.

By integrating (3.71) with respect to τ,

$$\xi(\theta, \tau | s) \propto \frac{\xi(\theta)}{\{2\beta + \sum_{i=1}^{n} [y_i - f(x_i, \theta)]^2\}^{(n+2\alpha)/2}}, \qquad \theta \in R \tag{3.72}$$

is the proper marginal posterior density of θ. Since θ is scalar, $\xi(\theta | s)$ is easily normalized, plotted, and the posterior moments, provided they exist, computed. It is assumed f is such a function that $\xi(\theta | s)$ is integrable with respect to θ over R. The marginal posterior density of θ is difficult to determine (except in special cases) in terms of simple functions, however the conditional posterior distribution of τ is gamma with parameters

$$\alpha' = \frac{n + 2\alpha}{2} \tag{3.73}$$

and $\beta'(\theta)$, where

$$2\beta'(\theta) = 2\beta + \sum_{i=1}^{n} [y_i - f(x_i, \theta)]^2, \tag{3.74}$$

and the marginal posterior mean of τ is

$$E(\tau | s) = E[\tau' | \beta'(\theta)],$$

$$E(\tau|s) = \underset{\theta|s}{E}[\tau'|\beta'(\theta)],$$

where the expectation is with respect to the posterior density of θ and one must assume $\xi(\theta|s)/\beta'(\theta)$ is integrable with respect to θ over R. Provided the other moments of τ exist, they may be computed in a similar fashion, in terms of the conditional posterior moments of τ. Of course, $\xi(\tau|s)$ may be determined by a univariate numerical integration of the joint density $\xi(\theta, \tau|s)$, (3.71), and we see if θ is scalar a complete Bayesian analysis is possible.

What is the Bayesian predictive density of a future observation w of the dependent variable y when x = x*? Assuming w is independent of s, the conditional density of w given θ and τ is

$$g(w|\theta,\tau) \propto \tau^{1/2} e^{-\tau[w-f(x*,\theta)]^2/2}, \qquad w \in R \tag{3.75}$$

and upon forming the product of $g(w|\theta, \tau)$ and $\xi(\theta, \tau|s)$, the joint density of w, θ, and τ given s is

$$h(w,\theta,\tau|s) \propto \xi(\theta)\tau^{(n+2\alpha+1)/2-1}\exp -\tfrac{\tau}{2}\{2\beta + \sum_{i=1}^{n}[y_i - f(x_i,\theta)]^2 + [w - f(x*,\theta)]^2\}, \tag{3.76}$$

where $w \in R$, $\theta \in R$, and $\tau > 0$.

Eliminating τ from (3.76) gives

$$h(w,\theta,\tau|s) \propto \frac{\xi(\theta)}{\{2\beta+\sum_{1}^{n}[y_i-f(x_i,\theta)]^2+[w-f(x*,\theta)]^2\}^{(n}}$$