Problem Set 2: Bayesian Theory in the Linear Regression Model

This problem sheet solves a series of Bayesian problems relating to the linear regression model (in matrix notation). Let me offer a quick review of this model. The linear regression model is the workhorse of econometrics. In addition to being important in its own right, this model is an important component of other, more complicated, models. The linear regression model posits a linear relationship between the dependent variable y_i and a $k-$vector of explanatory variables, x_i, where $i = 1, .., N$ indexes the relevant observational unit (e.g. individual, firm, time period, etc.). In matrix notation, the linear regression model can be written as:

$$y = X\beta + \varepsilon, \tag{1}$$

where $y = (y_1, .., y_N)'$ is an $N-$vector

$$X = \begin{bmatrix} x_1' \\ \cdot \\ \cdot \\ x_N' \end{bmatrix},$$

is an $N \times k$ matrix and $\varepsilon = (\varepsilon_1, .., \varepsilon_N)'$ is an $N-$vector of errors. Assumptions about ε and X define the likelihood function. The questions in this problem sheet we will assume that ε and X satisfy what we will refer to as the classical assumptions. With regards to the explanatory variables, we assume that they are not random.

With regards to the errors, we assume they are independently Normally distrubuted with mean 0 and common variance σ^2. That is,

$$\varepsilon \sim N\left(0_N, \sigma^2 I_N\right) \tag{2}$$

where 0_N is an $N-$vector of zeroes, I_N is the $N \times N$ identity matrix and $N()$ denotes the multivariate Normal distribution (see Appendix B to the textbook). It proves more convenient to work with the error precision, $h \equiv \sigma^{-2}$, and, thus, the Normal linear regression model depends on the parameter vector $\left(\beta', h\right)'$. Using the properties of the multivariate Normal distribution, it follows that $p\left(y|\beta, h\right) = f_N\left(y|X\beta, h^{-1}I_N\right)$ and, thus, the likelihood function is given by:

$$p\left(y|\beta, h\right) \equiv \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \exp\left[-\frac{h}{2}\left(y - X\beta\right)'\left(y - X\beta\right)\right]. \tag{3}$$

It is often convenient to write the likelihood function in terms of ordinary least squares (OLS) quantities:

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

and

$$SSE = \left(y - X\widehat{\beta}\right)' \left(y - X\widehat{\beta}\right).$$

This can be done by using the fact that

$$(y - X\beta)' (y - X\beta) = SSE + \left(\beta - \widehat{\beta}\right)' X'X \left(\beta - \widehat{\beta}\right)$$

and, thus,

$$p\left(y|\beta, h\right) \equiv \frac{h^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}}} \exp\left[-\frac{h}{2}\left\{SSE + \left(\beta - \widehat{\beta}\right)' X'X \left(\beta - \widehat{\beta}\right)\right\}\right]. \qquad (4)$$

Exercise 1

For the Normal linear regression model under the classical assumptions, use a Normal-Gamma prior (i.e.the prior for β and h is $NG\left(\underline{\beta}, \underline{Q}, \underline{s}^{-2}, \underline{\nu}\right)$, see the textbook Appendix B). Derive the posterior for β and h and, thus, show that the Normal-Gamma prior is a conjugate prior for this model.

Solution to Exercise 1

Using Bayes' rule and the properties of the Normal-Gamma density:

$$\begin{aligned} p\left(\beta, h|y\right) &\propto p\left(\beta, h\right) l\left(\beta, h\right) \\ &\propto f_N\left(\beta|\underline{\beta}, h^{-1}\underline{Q}\right) f_G\left(h|\underline{s}^{-2}, \underline{\nu}\right) f_N\left(y|X\beta, h^{-1}I_N\right), \end{aligned}$$

where $f_G\left(\right)$ denotes the Gamma density (see the textbook's Appendix B). This question can be solved by plugging in the forms for each of the densities in the above expressions, rearranging them (using some theorems in matrix algebra) and recognizing that the result is the kernel of Normal-Gamma density (details are provided below). Since the posterior and prior are both Normal-Gamma, conjugacy is established. The steps are elaborated in the remainder of this solution.

Begin by writing out each density (ignoring integrating constants not involving the parameters) and using the expression for the likelihood function in (4):

$$\begin{aligned} p\left(\beta, h|y\right) \propto\; & \left\{h^{\frac{k}{2}} \exp\left[-\frac{h}{2}\left(\beta - \underline{\beta}\right)' \underline{Q}^{-1}\left(\beta - \underline{\beta}\right)\right]\right\} \\ & \left\{h^{\frac{\underline{\nu}-2}{2}} \exp\left[-\frac{h\underline{\nu}\underline{s}^2}{2}\right]\right\} \left\{h^{\frac{N}{2}} \exp\left[-\frac{h}{2}\left\{SSE + \left(\beta - \widehat{\beta}\right)' X'X \left(\beta - \widehat{\beta}\right)\right\}\right]\right\} \\ =\; & h^{\frac{\overline{\nu}+k-2}{2}} \exp\left[-\frac{h}{2}\left\{\underline{\nu}\underline{s}^2 + SSE + \left(\beta - \underline{\beta}\right)' \underline{Q}^{-1}\left(\beta - \underline{\beta}\right) + \left(\beta - \widehat{\beta}\right)' X'X \left(\beta - \widehat{\beta}\right)\right\}\right], \end{aligned}$$

where $\overline{\nu} = \underline{\nu} + N$. In this expression, β enters only in the terms in the exponent, $\left(\beta - \underline{\beta}\right)' \underline{Q}^{-1}\left(\beta - \underline{\beta}\right) + \left(\beta - \widehat{\beta}\right)' X'X \left(\beta - \widehat{\beta}\right)$. Tedious, but straightfoward matrix manipulations allus to write this term as:

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

where

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

and

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

Thus, β enters only through the term $\left(\beta-\overline{\beta}\right)'\overline{Q}^{-1}\left(\beta-\overline{\beta}\right)$ and we can establish that the kernel of $\beta|y,h$ is given by:

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

Since this is the kernel of a Normal density we have established that $\beta|y,h\sim N\left(\overline{\beta},h^{-1}\overline{Q}^{-1}\right)$.

We can derive $p\left(h|y\right)$ by using the fact that

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

Since p.d.f.s integrate to one we can integrate out the component involving Normal density for $p\left(\beta|y,h\right)$ and we are left with:

$$\begin{aligned} p\left(h|y\right) &\propto h^{\frac{\overline{\nu}-2}{2}}\exp\left[-\frac{h}{2}\left\{\underline{\nu s}^2+SSE+\left(\widehat{\beta}-\underline{\beta}\right)'X'X\overline{Q}\underline{Q}^{-1}\left(\widehat{\beta}-\underline{\beta}\right)\right\}\right] \\ &= h^{\frac{\overline{\nu}-2}{2}}\exp\left[-\frac{h}{2}\overline{\nu s}^2\right]. \end{aligned}$$

But this is the kernel of a Gamma density and we have established that $h|y\sim G\left(\overline{s}^{-2},\overline{\nu}\right)$.

Since $\beta|y,h$ is Normal and $h|y$ is Gamma, it follows immediately that the posterior for β and h is $NG\left(\overline{\beta},\overline{Q},\overline{s}^{-2},\overline{\nu}\right)$. Since prior and posterior are both Normal-Gamma, conjugacy has been established.

Exercise 2

Suppose you have a Normal linear regression model with *partially informative* natural conjugate prior where prior information is available only on $J\leq k$ linear combinations of the regression coefficients and the prior for h is the standard noninformative one: $p\left(h\right)\propto\frac{1}{h}$. Thus, $R\beta|h\sim N(r,h^{-1}\underline{V}_r)$, where R is a known $J\times k$ matrix with rank$(R)=J$, r is a known J-vector and $\underline{V}_r$ is a $J\times J$ positive definite matrix. Show that the posterior is given by

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

where

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

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

$$\widetilde{\nu} = N$$

and

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

Solution to Exercise 2

Partition $X = [X_1, X_2]$ and $R = [R_1, R_2]$ where X_1 is $N \times (k - J)$, X_2 is $N \times J$, R_1 is $J \times (k - J)$ and R_2 is $J \times J$. Partion $\beta = \left(\beta_1', \beta_2'\right)'$ conformably. The linear regression model in (1) can be written as:

$$y = Z\gamma + \varepsilon,$$

where $Z = XA^{-1}$ and $\gamma = A\beta$ for any $k \times k$ nonsingular matrix A. If we take

$$A = \begin{bmatrix} I_{k-J} & 0_{(k-J)\times J} \\ R_1 & R_2 \end{bmatrix}$$

then

$$\begin{aligned} Z &= XA^{-1} = [Z_1, Z_2] \\ &= \left[X_1 - X_2R_2^{-1}R_1, X_2R_2^{-1}\right] \end{aligned}$$

and $\gamma = [\gamma_1', \gamma_2']' = \left[\beta_1', (R\beta)'\right]'$. In words, we have transformed the model so that the only prior information $\gamma_2 = R\beta$. The remainder of the proof is essentially the same as for Question 1. Assuming a Normal-Gamma natural conjugate prior for γ and h leads to a Normal-Gamma posterior. Taking noninformative limiting cases for the priors for γ_1 and h and then transforming back to the original parameterization (e.g. using $\beta = A^{-1}\gamma$ and $X = ZA$) yields the expressions given in the question.

Exercise 3

Problems with Bayes Factors using Noninformative Priors.

Suppose you have two Normal linear regression models:

$$M_j : y = X_j\beta + \varepsilon_j,$$

where $j = 1, 2$, X_j is an $N \times k_j$ matrix of explanatory variables, β_j is a k_j-vector of regression coefficients and ε_j is an N-vector of errors distributed as $N\left(0_N, h_j^{-1}I_N\right)$. If natural conjuage priors are used for both models (i.e. $\beta_j, h_j|M_j \sim NG\left(\underline{\beta}_j, \underline{Q}_j, \underline{s}_j^{-2}, \underline{\nu}_j\right)$), then the posterior is $\beta_j, h_j|y, M_j \sim NG\left(\overline{\beta}_j, \overline{Q}_j, \overline{s}_j^{-2}, \overline{\nu}_j\right)$ (where $\overline{\beta}_j, \overline{Q}_j, \overline{s}_j^{-2}$ and $\overline{\nu}_j$ are as given in the solution to Question1) and the Bayes factor comparing M_2 to M_1 is given by:

$$BF_{21} = \frac{c_2\left(\frac{|\overline{Q}_2|}{|\underline{Q}_2|}\right)^{\frac{1}{2}}\left(\overline{\nu}_2\overline{s}_2^2\right)^{-\frac{\overline{\nu}_2}{2}}}{c_1\left(\frac{|\overline{Q}_1|}{|\underline{Q}_1|}\right)^{\frac{1}{2}}\left(\overline{\nu}_1\overline{s}_1^2\right)^{-\frac{\overline{\nu}_1}{2}}}.$$

where

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

a) Consider a noninformative prior created by letting $\underline{\nu}_j \to 0$, $\underline{Q}_j^{-1} = cI_{k_j}$ and letting $c \to 0$ for $j = 1, 2$. Show that the Bayes factor reduces to:

$$\begin{cases} 0 \text{ if } \quad k_2 > k_1 \\ \left[\frac{|X_2'X_2|}{|X_1'X_1|}\right]^{-\frac{1}{2}}\left(\frac{SSE_2}{SSE_1}\right)^{-\frac{N}{2}} \text{ if } \quad k_2 = k_1 \\ \infty \text{ if } \quad k_2 < k_1 \end{cases}.$$

b) Consider a noninformative prior created by setting $\underline{\nu}_j \to 0$, $\underline{Q}_j^{-1} = \left(c^{\frac{1}{k_j}}\right)I_{k_j}$ and letting $c \to 0$ for $j = 1, 2$. Show that the Bayes factor reduces to:

$$\left[\frac{|X_2'X_2|}{|X_1'X_1|}\right]^{-\frac{1}{2}}\left(\frac{SSE_2}{SSE_1}\right)^{-\frac{N}{2}}.$$

c) Consider a noninformative prior created by setting $\underline{\nu}_j \to 0$, $\underline{Q}_j^{-1} = \left(c^{\frac{1}{k_j}}\right)X_j'X_j$ and letting $c \to 0$ for $j = 1, 2$. Show that the Bayes factor reduces to:

$$\left(\frac{SSE_2}{SSE_1}\right)^{-\frac{N}{2}}.$$

Solution to Exercise 3

In all cases, if $\underline{\nu}_1 = \underline{\nu}_2 \to 0$ at the same rate then $c_1 = c_2$ and these integrating constants cancel out in the Bayes factor. Furthermore, under the various assumptions about $\underline{Q}_j^{-1}$ if can be seen that:

$$\overline{\nu}_j\overline{s}_j^2 = SSE_j,$$

$\overline{\nu}_j = N$ and $\overline{Q}_j = \left(X_j'X_j\right)^{-1}$ for $j = 1, 2$. Thus, in all cases, the Bayes factor reduces to:

$$BF_{21} = \left(\frac{|\underline{Q}_2^{-1}|}{|\underline{Q}_1^{-1}|}\right)^{\frac{1}{2}} \left[\frac{|X_2'X_2|}{|X_1'X_1|}\right]^{-\frac{1}{2}} \left(\frac{SSE_2}{SSE_1}\right)^{-\frac{N}{2}}.$$

In part a) $|\underline{Q}_j^{-1}| = c^{k_j}$ and, hence, $\left(\frac{|\underline{Q}_2^{-1}|}{|\underline{Q}_1^{-1}|}\right) = c^{k_2-k_1}$. If $k_1 = k_2$ then $\left(\frac{|\underline{Q}_2^{-1}|}{|\underline{Q}_1^{-1}|}\right) = 1$ for all c. If $k_1 > k_2$ then $\left(\frac{|\underline{Q}_2^{-1}|}{|\underline{Q}_1^{-1}|}\right) \to \infty$ as $c \to 0$. If $k_1 < k_2$ then $\left(\frac{|\underline{Q}_2^{-1}|}{|\underline{Q}_1^{-1}|}\right) \to 0$ as $c \to 0$. Thus, the result in part a) is established.

In part b) $|\underline{Q}_j^{-1}| = c$ and, hence, $\left(\frac{|\underline{Q}_2^{-1}|}{|\underline{Q}_1^{-1}|}\right) = 1$ for all c regardless of what k_j is.Thus, the result in part b) is established.

In part c) $|\underline{Q}_j^{-1}| = c\,|X_j'X_j|$and, hence, $\left(\frac{|\underline{Q}_2^{-1}|}{|\underline{Q}_1^{-1}|}\right) = \frac{|X_2'X_2|}{|X_1'X_1|}$ for all c regardless of what k_j is. Using this results, the Bayes factor simplifies to the expression given in part c).

Exercise 4

Multicollinearity

Consider the Normal linear regression model with natural conjugate prior: $NG\left(\underline{\beta}, \underline{Q}, \underline{s}^{-2}, \underline{\nu}\right)$.Assume in addition that $Xc = 0$ for some non-zero vector of constants c. Note that this is referred to as a case of *perfect multicollinearity.* It implies the matrix X is not of full rank and $\left(X'X\right)^{-1}$ does not exist.

a) Show that, despite this pathology, the posterior exists if $\underline{Q}$ is positive definite.

b) Define

$$\alpha = c'\underline{Q}^{-1}\beta.$$

Show that, given h, the prior and posterior distributions of α are both identical and equal to:

$$N\left(c'\underline{Q}^{-1}\underline{\beta}, h^{-1}c'\underline{Q}^{-1}c\right).$$

Hence, although prior information can be used to surmount the problems caused by perfect multicollinearity, there are some combinations of the regression coefficients about which learning does not occur.

Solution to Exercise 4

part a). The solution to this question is essentially the same as to Exercise 1. The key thing to note is that, since $\underline{Q}$ is positive definite, $\underline{Q}^{-1}$ exists and, hence, $\overline{Q}^{-1}$ exists despite the fact that $X'X$ is rank deficient. In the manner, it can be shown that the posterior for β and h is $NG\left(\overline{\beta}, \overline{Q}, \overline{s}^{-2}, \overline{\nu}\right)$ with:

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

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

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

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

part b). The properties of the Normal-Gamma distribution imply:

$$\beta|h \sim N\left(\underline{\beta}, h^{-1}\underline{Q}\right).$$

The properties of the Normal distribution imply

$$\alpha|h \equiv c'\underline{Q}^{-1}\beta \sim N\left(c'\underline{Q}^{-1}\underline{\beta}, h^{-1}c'\underline{Q}^{-1}c\right),$$

which establishes that the prior has the required form.

Using the result from part a), the relevant posterior has the form:

$$\beta|y, h \sim N\left(\overline{\beta}, h^{-1}\overline{Q}\right)$$

which implies

$$\alpha|y, h \sim N\left(c'\underline{Q}^{-1}\overline{\beta}, h^{-1}c'\underline{Q}^{-1}\overline{Q}\underline{Q}^{-1}c\right).$$

The mean can be written as:

$$\begin{aligned}
c'\underline{Q}^{-1}\overline{\beta} &= c'\left(\overline{Q}^{-1} - X'X\right)\overline{Q}\left(\underline{Q}^{-1}\underline{\beta} + X'y\right) \\
&= c'\underline{Q}^{-1}\underline{\beta} + c'X'y - c'X'X\overline{Q}\left(\underline{Q}^{-1}\underline{\beta} + X'y\right) \\
&= c'\underline{Q}^{-1}\underline{\beta},
\end{aligned}$$

since $Xc = 0$.

The variance can be written as:

$$\begin{aligned}
h^{-1}c'\underline{Q}^{-1}\overline{Q}\underline{Q}^{-1}c &= h^{-1}c'\underline{Q}^{-1}\left[\left(\underline{Q}^{-1} + X'X\right)^{-1}\underline{Q}^{-1}\right]c \\
&= h^{-1}c'\underline{Q}^{-1}\left[I - \left(\underline{Q}^{-1} + X'X\right)^{-1}X'X\right]c \\
&= h^{-1}c'\underline{Q}^{-1}c,
\end{aligned}$$

since $Xc = 0$. Note that this derivation uses a standard theorem in matrix algebra that says, if G and H are $k \times k$ matrices and $(G + H)^{-1}$ exists, then

$$(G+H)^{-1}H = I_k - (G+H)^{-1}G.$$

Combining these derivations, we see that the posterior is also $\alpha|y,h \sim N\left(c'\underline{Q}^{-1}\underline{\beta}, h^{-1}c'\underline{Q}^{-1}c\right)$ as required.