Bayesian analysis in linear regression with different priors

Chowdhury, S.R.

Publication date:
1970

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Chowdhury, S. R. (1970). *Bayesian analysis in linear regression with different priors*. (EIT Research Memeorandum). Stichting Economisch Instituut Tilburg.

S. R. Chowdhury

Bayesian analysis in linear regression with different priors

Research memorandum

TILBURG INSTITUTE OF ECONOMICS
DEPARTMENT OF ECONOMETRICS

Bayesian analysis in linear regression with different priors

by

S.R. CHOWDHURY

Bayesian analysis in linear regression with different priors

by

S.R. CHOWDHURY

1. Introduction

In this paper two types of priors are used to estimate the parameters. One type of prior is just the Jeffreys' prior i.e. a prior of ignorance or which is sometimes called a diffuse prior. The other prior is in the form of a multivariate normal distribution. In both the cases joint posterior distributions, marginal posterior distributions etc. of the parameters are derived analytically. It may be mentioned that these are the priors which are mostly used in econometric analysis.

2. The Model and the estimation of parameters

We take the single equation regression model,

(1) $$y = X\beta + u$$

> y is a Tx1 vector of observations on dependent variable.
> X is a Txp matrix of observations on the explanatory variables, with fixed elements and rank p.
> β is a px1 vector of unknown parameters.
> u is a Tx1 vector of random disturbances.
> Each element of u is independently and normally distributed with mean zero and variance σ^2.

The likelihood function of the sample is given by,

(2) $$l(\beta,\sigma|y) = \frac{1}{\sigma^T(2\pi)^{T/2}} \exp\{-\frac{1}{2\sigma^2}[(y - X\beta)'(y - X\beta)]\}$$

Throughout this paper we shall use the symbol $Q(\beta, \alpha, A)$ to denote a quadratic form in variables β centred at α and with matrix A, namely

$$Q(\beta, \alpha, A) \equiv (\beta-a)'A(\beta-\alpha)$$

The likelihood function (2) can now be written as:

(3) $$l(\beta,\sigma|y) = \frac{1}{\sigma^T(2\pi)^{T/2}} \exp\{-\frac{1}{2\sigma^2}[Q(\beta, \hat{\beta}, V) + (T - p)s^2]\}$$

where:

$V = (X'X)$,
$\hat{\beta} = V^{-1}X'y$ (L.S. estimator of β)
$(T - p)s^2 = (y - X\hat{\beta})'(y - X\hat{\beta})$ (Residual S.S)

and

$$(y - X\beta)'(y - X\beta) = (\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta}) + (y - X\hat{\beta})'(y - X\beta)$$

$$= Q(\beta, \hat{\beta}, V) + (T - p)s^2$$

Prior distributions

2.1 Jeffreys' prior [3]

Log σ, and the elements of β are assumed to be locally, uniformly and independently distributed:

(4) $$p(\beta, \sigma) \propto \frac{1}{\sigma} \qquad 0 < \sigma < \infty$$

$$- \infty < \beta < + \infty$$

Combining prior and likelihood by Bayes theorem, we get the joint posterior distribution of β and σ as:

(5) $$p(\beta, \sigma | y) \propto \sigma^{-(T+1)} \exp \{- \frac{1}{2\sigma^2} [Q(\beta, \hat{\beta}, V) + (T-p)s^2]\}$$

$$0 < \sigma < \infty$$

$$- \infty < \beta < + \infty$$

Integrating out σ, we get the marginal posterior distribution of β as:

(6) $$p(\beta | y) \propto [Q(\beta, \hat{\beta}, V) + (T-p)s^2]^{-\frac{T}{2}} \qquad - \infty < \beta < + \infty$$

This is in the form of a p-dimensional multivariate "t" distribution. Under the assumption of quadratic loss, the posterior mean of β is the Bayesian estimator. So the Bayesian estimator of β is $\hat{\beta}$ (posterior mean), i.e. same as the L.S. estimator. The marginal distribution of any one element of β is univariate "t" and can be easily obtained from (6) by integration.

2.2 Multinormale prior

We assume that the prior of β is in the form of a multivariate normal distribution. The prior of σ is just like before i.e. log σ is uniform, and locally independent of the prior of β:

$$(7) \qquad p(\beta,\sigma) \propto \frac{1}{\sigma} \exp\{-\frac{1}{2}(\beta - \tilde{\beta})'S(\beta - \tilde{\beta})\} \qquad 0 < \sigma < \infty$$
$$- \infty < \beta < + \infty$$

β is assumed to follow a multivariate normal with mean $\tilde{\beta}$ and covariance matrix S^{-1}.

The joint posterior distribution of β and σ is given by:

$$(8) \qquad p(\beta,\sigma|y) \propto \sigma^{-(T+1)} \exp\{-\frac{1}{2\sigma^2}[Q(\beta, \hat{\beta}, V) + (T - p)s^2]\}.$$
$$\exp\{-\frac{1}{2}(\beta - \tilde{\beta})'S(\beta - \tilde{\beta})\}$$

Marginal posterior distribution for the multinormal prior.

2.2.1 Case 1: σ is known.

Since σ is known and $(T - p)s^2$ is constant, we can write for the marginal posterior distribution of β from (8):

$$(9) \qquad p(\beta|y) \propto \exp\{-\frac{1}{2}[(\beta - \hat{\beta})'\frac{V}{\sigma^2}(\beta - \hat{\beta}) + (\beta - \tilde{\beta})'S(\beta - \tilde{\beta})]\}.$$

Denote $\frac{V}{\sigma^2} = V_1$ $\qquad H = V_1 + S$

$$\bar{\beta} = H^{-1}(V_1\hat{\beta} + S\tilde{\beta})$$

So

$$(10) \qquad p(\beta|y) \propto \exp\{-\frac{1}{2}Q(\beta, \bar{\beta}, H)\}$$

Now (10) is in the form of a multivariate normal distribution. The posterior mean of β is $\overline{\beta}$, which is the Bayesian estimator, again with the assumption of quadratic loss. The marginal posterior distributions of each element of β is normal and can be easily derived.

2.2.2 Case 2: σ is unknown.

From the joint posterior distribution of β and σ in (8), by integrating out σ we get the marginal posterior distribution of β as:

$$(11) \qquad p(\beta|y) \; \alpha \; [Q(\beta, \hat{\beta}, V) + (T - p)s^2]^{-T/2} . \; [\exp\{-\tfrac{1}{2}(\beta-\tilde{\beta})'S(\beta-\tilde{\beta})\}]$$

(11) is the product of a multivariate "t" distribution with a multivariate normal distribution.

As $\nu = T - p$ tends to infinity, the multivariate "t" distribution becomes a multivariate normal and $\overline{\beta}$ is the limiting mean of the distribution in (11).

Now from (11) we get:

$$(12) \qquad p(\beta|y) \; \alpha \left[1 + \frac{Q(\beta, \hat{\beta}, V)}{\nu s^2}\right]^{-\frac{(\nu+p)}{2}} \left[\exp\{-\tfrac{1}{2}(\beta - \tilde{\beta})'S(\beta - \tilde{\beta})\}\right]$$

Write $Q_1 = \dfrac{Q(\beta, \hat{\beta}, V)}{s^2}$ and $Q_2 = (\beta - \tilde{\beta})'S(\beta - \tilde{\beta})$

$V_1 = \dfrac{V}{s^2}$. Then $Q_1 = Q(\beta, \hat{\beta}, V_1)$.

The expression $\left\{1 + \dfrac{Q_1}{\nu}\right\}^{-\frac{(\nu+p)}{2}}$ can be written as

$$(13) \qquad \left\{1 + \frac{Q_1}{\nu}\right\}^{-\frac{1}{2}(\nu+p)} = \exp\left\{-\frac{1}{2}Q_1\right\} \exp\left\{\frac{1}{2}Q_1 - \frac{\nu + p}{2} \ln\left[1 + \frac{Q_1}{\nu}\right]\right\}$$

On expanding the second factor on the right in powers of ν^{-1}, we obtain from (13):

(14) $$\left\{1 + \frac{Q_1}{\nu}\right\}^{-\frac{1}{2}(\nu+p)} = \exp\left\{-\frac{1}{2} Q_1\right\} \sum_{i=0}^{\infty} p_i \nu^{-i}$$

where:

$$p_0 = 1, \; p_1 = \frac{1}{4} [Q_1^2 - 2pQ_1]$$

$$p_2 = \frac{1}{96} [3Q_1^4 - 4(3p + 4)Q_1^3 + 12p(p + 2)Q_1^2]$$

The expression (12) can now be wrtiten as:

(15) $$p(\beta|y) \;\alpha\; \exp\{-\frac{1}{2}(Q_1 + Q_2)\} \sum_{i=0}^{\infty} p_i \nu^{-i}$$

or

(16) $$p(\beta|y) \;\alpha\; \frac{|H|^{1/2}}{(2\pi)^{p/2}} \exp\{-\frac{1}{2} Q(\beta, \bar{\beta}, H)\} \sum_{i=0}^{\infty} p_i \nu^{-i}$$

where $H = V_1 + S$; $\bar{\beta} = H^{-1}(V_1 \hat{\beta} + S\tilde{\beta})$
or

(17) $$p(\beta|y) = W^{-1} h(\beta),$$

where $h(\beta) = \frac{|H|^{1/2}}{(2\pi)^{p/2}} \exp\{-\frac{1}{2} Q(\beta, \bar{\beta}, H)\} \sum_{i=0}^{\infty} p_i \nu^{-i}$

and

(18) $$W = \int h(\beta) d\beta$$

The integral W in (18) can be integrated term by term. Each term is, in fact, a polynomial in the moments of the quadratic form $Q_1 = Q(\beta,\hat{\beta},V_1)$, where the variables β have a multivariate normal distribution with mean $\bar{\beta}$ and covariance matrix V^{-1}. The moments are found out indirectly from the cumulants. (COOK, M.B. [2], KENDALL, M.G. and A. STUART [4])

The cumulant generating function of Q_1 is

(19) $$K(t) = \log \int_R \frac{|H|^{1/2}}{(2\pi)^{p/2}} \exp\{tQ_1 - \frac{1}{2} Q(\beta, \bar{\beta}, H)\}d\beta$$

$$= -\frac{1}{2}\log|I - 2H^{-1}(tV_1)| + t\eta^1 V_1\eta + 2(tV_1\eta)'(H-2tV_1\eta)^{-1}(tV_1\eta)$$

where $\eta = \bar{\beta} - \hat{\beta}$

On differentiating (19) and after some algebratic reductions we get

$$K_1 = \text{tr}\, H^{-1} V_1 + \eta' V_1 \eta$$

$$K_r = 2^{r-1}(\gamma-1)!\ \{\text{tr}(H^{-1}V_1)^{\gamma} + \gamma\eta' H(H^{-1}V_1)^{\gamma}\eta\}$$

where K's are the cumulants.

Now $W = \Sigma\, b_i v^{-i}$

where $b_o = 1$; $b_1 = \frac{1}{4}[K_2 + K_1^2 - 2pK_1]$

(20) $$b_2 = \frac{1}{96}[3(K_4 + 4K_3K_1 + 3K_2^2 + 6K_2K_1^2 + K_1^4) - 16K_3 - 48K_2K_1$$
$$- 16K_1^3 - 3p(4K_3 + 12K_2K_1 + 4K_1^3 - 8K_2 - 8K_1^2)$$
$$+ 12p^2(K_2 + K_1^2)]$$

Substituting the results of (20) in (17) we get the following asymptotic expression for the posterior distribution of β.

(21) $$p(\beta|y) = \frac{|H|^{1/2}}{(2\pi)^{p/2}} \exp\{-\frac{1}{2} Q(\beta, \bar{\beta}, H)\} \sum_{i=o}^{\infty} c_i\, v^{-i}$$

where $c_o = 1$

$$c_1 = p_1 - b_1$$

$$c_2 = p_2 - b_2 - p_1 b_1 + b_1^2$$

The terms c_i are found out from the expression:

$$\left[\sum_{o}^{\infty} b_i \; \nu^{-i}\right]\left[\sum_{o}^{\infty} p_i \; \nu^{-i}\right]$$

As ν tends to infinity, the mean of the posterior distribution in (21) tends to $\bar{\beta}$. For other values of ν, the mean can be found out from (21).

2.2.2.1 Marginal posterior distribution of a single parameter

$$\text{Let } \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_{(2)} \end{pmatrix} \qquad \text{where } \beta_{(2)} = \begin{pmatrix} \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

From the joint posterior distribution of β_1 and $\beta_{(2)}$, we get the marginal posterior distribution for β_1 as

$$(22) \qquad p(\beta_1|y) = \frac{|H|^{1/2}}{(2\pi)^{p/2}} \int_{R'} \exp\{-\tfrac{1}{2} Q(\beta, \bar{\beta}, H)\} \sum_{i=o}^{\infty} c_i \; \nu^{-i} \, d\beta_{(2)}$$

The following partition is made:

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \qquad H^{-1} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

We can now write the marginal posterior distribution as

$$p(\beta_1|y) = \frac{|S_{11}|^{-1/2}}{(2\pi)^{1/2}} \quad \exp\{-\tfrac{1}{2} Q(\beta_1, \bar{\beta}, S_{11}^{-1})\} \; f(\beta_1|y)$$

where,

$$f(\beta_1|y) = \frac{|H_{22}|^{1/2}}{(2\pi)^{1/2(p-1)}} \int_{R'} \exp\{-\tfrac{1}{2} Q(\beta_{(2)}, \; \theta, H_{22})\} \sum_{i=o}^{\infty} c_i \nu^{-i} d\beta_{(2)}$$

with $\theta = \bar{\beta}_{(2)} - H_{22}^{-1} H_{21}(\beta_1 - \bar{\beta}_1)$

The evaluation of $f(\beta_1|y)$ will be done in the same way as before i.e. by finding out the cumulants first and then making inversion.
For this, β_1 is considered fixed and $\beta_{(2)}$ is considered to have a multivariate normal distribution with mean θ and covariance matrix H_{22}^{-1}.

The following partition is made:

$$V_1 = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \quad (\text{columns } 1,\ p-1) \quad ; \qquad V_1^{-1} = \begin{bmatrix} M^{11} & M^{12} \\ M^{21} & M^{22} \end{bmatrix}$$

The cumulants of $Q(\beta, \hat{\beta}, V_1)$ are as follows:

$$(23) \qquad W_1 = \text{tr } H_{22}^{-1} M_{22} + \gamma' M_{22} \gamma + Q(\beta_1, \hat{\beta}_1, (M^{11})^{-1})$$

$$W_r = 2^{r-1}(r-1)! \; \{ \text{tr}(H_{22}^{-1} M_{22})^r + r\gamma' H_{22}(H_{22}^{-1} M_{22})^r \gamma \}$$

where $\gamma = \theta - \hat{\beta}_{(2)} + M_{22}^{-1} M_{21}(\beta_1 - \hat{\beta}_1)$.

Using the results of 23, we can express the marginal posterior distribution of β_1 as

$$p(\beta_1|y) = \frac{|S_{11}|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \exp\{ -\tfrac{1}{2} Q(\beta_1, \bar{\beta}_1, S_{11}^{-1}) \} \sum_{o}^{\infty} \delta_i \, v^{-i}$$

where $\delta_o = 1$ $\qquad \delta_2 = g_2 - b_2 g_1 b_1 + b_1^2$

$\delta_1 = g_1 - b_1$ $\qquad$ ------------------------------

where g_i's have the same relation with W_i's as b_i's has with K_i's.

3. Construction of the Multinormal prior

In the previous section we have given all the theory that is required for the estimation of parameters. The procedure assumed the existence of a multinormal prior.

Multinormal priors can be constructed from the past samples. Least squares estimates of β and its covariance matrix for the past sample can be utilized to build up the prior mean and prior covariance matrix. Another way may be to assume Jeffreys' prior for the previous sample and take the posterior distribution of β as the prior for the current one.

4. Acknowledgement

Many thanks are due to my colleague W.H. Vandaele for preparing this note.

5. References

[1] ANDERSON, T.W. An Introduction to Multivariate Statistical Analysis. New York, John Wiley & Sons, 1958, 374 pp.

[2] COOK, M.B. "Bivariate K-statistics and cumulants of their joint sampling distribution", Biometrika, Vol. 38, 1951, pp. 179-195.

[3] JEFFREYS, H. Theory of Probability. Oxford, Clarendon Press, 1961, 3rd edition, 459 pp.

[4] KENDALL, M.G. and A. STUART. The Advanced Theory of Statistics: Vol. 1, Distribution Theory. London, Charles Griffin and Company Ltd, 1963, 433 pp.

[5] ROTHENBERG, T. A Bayesian analysis of simultaneous systems. Rotterdam, Econometric Institute Report 6315, 1963, 20 pp.

[6] TIAO, George C. and Arnold ZELLNER "Bayes's theorem and the use of prior knowledge in Regression Analysis", Biometrika, Vol. 51, 1964, nrs 1/2, pp. 219-230.

[7] ZELLNER, Arnold and George C. TIAO "Bayesian Analysis of the Regression Model with Autocorrelated Errors", Journal of the American Statistical Association, Vol. 59, September 1964, nr. 307, pp. 763-778.

E.I.T. 1970