A bayesian approach in multiple regression analysis with inequality constraints

Chowdhury, S.R.

Publication date:
1969

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

S. R. Chowdhurry

A bayesian approach in multiple regression analysis with inequality constraints

R41

T Bayesian statistics
T multiple regression analysis

Research memorandum

ECONIMIC INSTITUTE TILBURG
DEPARTMENT OF ECONOMETRICS

A BAYESIAN APPROACH IN MULTIPLE REGRESSION ANALYSIS WITH INEQUALITY CONSTRAINTS

BY

S. R.CHOWDHURY

1. Introduction

We consider those cases in multiple regression analysis, where our only prior knowledge is, that a subset of the parameters have finite, definite and known bounds. Examples of this type often occur in Econometric Analysis, e.g. the marginal propensity to consume in consumption equations lies between 0 and 1. It may happen, that a least squares method, when applied to the above situations, produce estimates of the parameters, which are inconsistent with our prior knowledge, i.e. some or all of the estimates may fall outside the known bounds. This is clearly unacceptable to the experimentor. The reasons of this inconsistency may be due to multicollinearity, inadequacy of the sample data or otherwise.

The method given here is essentially a Bayesian one, and will take care of the above situations. The estimates will be always consistent with the prior knowledge. Even if the least squares estimates are consistent, the estimation procedure which incorporates the apriori information explicitly is more justified and efficient than the procedure which treats the parameters as unrestricted.

2. Bayesian estimates of the parameters

We take the single equation regression model,

(2.1) $$y = X\boldsymbol{\beta} + u$$

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.
$\boldsymbol{\beta}$ 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 $\boldsymbol{\sigma}^2$.

The likelihood function of the sample is given by,

(2.2) $$\ell(\beta,\sigma\,|\,y) = \frac{1}{\sigma^{T}(2\pi)^{T/2}} \operatorname{Exp}\left\{-\frac{1}{2\sigma^2}\left[(y - X\beta)'(y - X\boldsymbol{\beta})\right]\right\}$$

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

$$Q(\boldsymbol{\beta},\boldsymbol{\alpha},A) \equiv (\beta\ \ \alpha)'A(\beta\ \ \alpha)$$

The likelihood function (2.2) can now be written as:

(2.3) $$\ell(\beta,\sigma\,|\,y) = \frac{1}{\sigma^{T}(2\pi)^{T/2}} \operatorname{Exp}\left\{-\frac{1}{2\sigma^2}\left[Q(\beta,\hat{\beta},V) + (T - p)S^2\right]\right\}$$

where:

$$V = (X'X) \quad ,$$

$$\hat{\beta} = V^{-1}X'y \text{ (L.S. estimate of } \beta)$$

$$(T - p)S^2 = (y - X\hat{\beta})'(y - X\hat{\beta})$$

and

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

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

Bayesian solution: σ is known

As regards the prior distribution, we assume that only the bounds of a subset β_1, of the parameters β are finite and definitely known. The method essentially remains the same if the bounds are either $+\infty$ or $-\infty$ e.g. when the parameters are restricted to be positive or negative. Following Jeffreys [3], Zellner and Tiao [5 & 6], we assume that the elements of β_1 and β_2 are locally independent and uniform in their respective ranges. This type of prior is usually called diffuse or non-informative in the literature.

The following prior distributions on β_1 and β_2 is taken,

(2.4) $\quad p(\beta_1,\beta_2) \propto \text{constant}$

with:

$c \leqslant \beta_1 \leqslant d$

$-\infty < \beta_2 < \infty$

c and d are rx1 vectors with known elements.

By Bayes theorem, the joint posterior distribution is given by,

(2.5) $\quad p(\beta_1,\beta_2 | y) \propto \ell(\beta_1,\beta_2 | y)p(\beta_1,\beta_2)$

or, combining (2.3) and (2.4) we get,

$$(2.6) \qquad p(\beta_1, \beta_2 \mid y) \propto \sigma^{-T} \exp \left\{ -\frac{1}{2\sigma^2} \left[Q(\beta, \hat{\beta}, V) + (T-p)S^2 \right] \right\}$$

Without loss of generality, let β_1 be the first r elements of β, and β_2 consists of the remaining p-r elements. Thus $\beta = \binom{\beta_1}{\beta_2}$. The matrix V is accordingly partitioned as,

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \equiv \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}$$

where

X_1 is a T x r matrix

X_2 is a T x (p-r) matrix

$X = (X_1 \; X_2)$.

The quadratic form $Q(\beta, \hat{\beta}, V)$ in (2.6) can be further written as,

$$(2.7) \qquad Q(\beta, \hat{\beta}, V) = (\beta - \hat{\beta})'V(\beta - \hat{\beta}) = Q(\beta_2, \hat{\beta}_2 - V_{22}^{-1}V_{21}(\beta_1 - \hat{\beta}_1), V_{22}) + Q(\beta_1, \hat{\beta}_1, V_{11} - V_{12}V_{22}^{-1}V_{21})$$

Here the quadratic form $Q(\beta, \hat{\beta}, V)$ is split into two quadratic forms, one containing β_1 only and the other containing β_1 and β_2.

Taking account of (2.7), the joint posterior distribution of β_1 and β_2 in (2.6) is expressed as,

$$(2.8) \qquad p(\beta_1, \beta_2 \mid y) \propto \sigma^{-T} \exp \left\{ -\frac{1}{2\sigma^2} \left[Q(\beta_2, \hat{\beta}_2 - V_{22}^{-1}V_{21}(\beta_1 - \hat{\beta}_1), V_{22}) + Q(\beta_1, \hat{\beta}_1, V_{11} - V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2 \right] \right\}$$

Using the properties of multivariate normal distribution, β_2 is integrated out from (2.8), when we get the marginal posterior distribution of β_1 as,

$$(2.9) \qquad p(\beta_1|y) \propto \sigma^{-(T-p+r)} \exp\{-\frac{1}{2\sigma^2}\left[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2\right]\}$$

Since σ is known and $(T-p)S^2$ is constant, we can write,

$$(2.10) \qquad p(\beta_1|y) \propto \exp\{-\frac{1}{2\sigma^2}\left[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})\right]\}, c \leq \beta_1 \leq d$$

From (2.10) it is seen that the marginal posterior distribution of β_1 is in the form of a multivariate r dimensional normal distribution, but truncated.

It is well known that the Bayesian estimates of the parameters are the means of the marginal posterior distributions, when the loss function is a quadratic one.

With the assumption of a quadratic loss function, the Bayesian estimate of β_1 can be evaluated from,

$$(2.11) \qquad \tilde{\beta}_1 = \frac{\int_c^d \beta_1 \exp\{-\frac{1}{2\sigma^2}\left[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})\right]\} d\beta_1}{\int_c^d \exp\{-\frac{1}{2\sigma^2}\left[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})\right]\} d\beta_1}$$

where

$\tilde{\beta}_1$ is the posterior mean of (2.10). The denominator in (2.11) is the normalising constant for (2.10).

(2.11) can be further written as:

$$\tilde{\beta}_1 = \frac{\int_c^d (\beta_1-\hat{\beta}_1)\exp\{-\frac{1}{2\sigma^2}[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})]\,d\beta_1 + \hat{\beta}_1\int_c^d\exp\{-\frac{1}{2\sigma^2}[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})]\}\,d\beta_1}{\int_c^d\exp\{-\frac{1}{2\sigma^2}[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})]\}\,d\beta_1}$$

$$= \hat{\beta}_1 - \frac{\sigma^2}{V_{11}-V_{12}V_{22}^{-1}V_{21}}\left\{\frac{\int_c^d\exp\left[-\frac{1}{2\sigma^2}[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})]\right]d\left(-\frac{1}{2\sigma^2}[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})]\right)}{\int_c^d\exp\left[-\frac{1}{2\sigma^2}[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})]\right]d\beta_1}\right\}$$

$$= \hat{\beta}_1 - \frac{\sigma^2}{V_{11}-V_{12}V_{22}^{-1}V_{21}}\left\{\frac{\exp\{-\frac{1}{2\sigma^2}[Q(d,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})]\} - \exp\{Q(c,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})\}}{\int_c^d\exp\{-\frac{1}{2\sigma^2}[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})]\}\,d\beta_1}\right\}$$

One has to apply numerical integration procedure to evaluate $\tilde{\beta}_1$.

Bayesian estimate of β_2

To find the Bayesian estimate of β_2, we need to find first the marginal posterior distribution of β_2. From (2.8), the marginal posterior distribution of β_2 is obtained by integrating out β_1. Thus

$$(2.12) \quad p(\beta_2|y) \propto \sigma^{-T} \int_c^d \exp\{-\frac{1}{2\sigma^2}[Q(\beta_2,\hat{\beta}_2-V_{22}^{-1}V_{21}(\beta_1-\hat{\beta}_1),V_{22}) + Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2]\}\,d\beta_1$$

The Bayesian estimate of β_2, which is the posterior mean of β_2, is,

$$(2.13) \quad \tilde{\beta}_2 = \frac{\int_{-\infty}^{\infty} \beta_2\{\int_c^d \exp\{-\frac{1}{2\sigma^2}[Q(\beta_2,\hat{\beta}_2-V_{22}^{-1}V_{21}(\beta_1-\hat{\beta}_1),V_{22}) + Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2]\}\,d\beta_1\}\,d\beta_2}{\int_{\infty}^{\infty}\left[\int_c^d[\exp\{-\frac{1}{2\sigma^2}[Q(\beta_2,\hat{\beta}_2-V_{22}^{-1}V_{21}(\beta_1-\hat{\beta}_1),V_{22}) + Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21})+(T-p)S^2]\}]\,d\beta_1\right]d\beta_2}$$

Changing the order of integrals, and considering the properties of the multivariate normal distribution, we obtain after simplification the following simple relation,

$$(2.14) \quad \tilde{\beta}_2 = \hat{\beta}_2 - V_{22}^{-1}V_{21}(\tilde{\beta}_1-\hat{\beta}_1)$$

From (2.14), $\tilde{\beta}_2$ can be easily calculated, once $\tilde{\beta}_1$ is calculated by numerical integrations procedure. It is to be noted that when the prior informations about β_1 are also non informative like β_2 i.e. $p(\beta_1,\beta_2) \propto$ Constant with $-\infty<\beta_1<\infty$, $-\infty<\beta_2<\infty$, then $\hat{\beta}_1$ and $\hat{\beta}_2$ are respectively equal to $\tilde{\beta}_1$ and $\tilde{\beta}_2$, and this fact is also corroborated by the relation (2.14).

Bayesian solution: σ is unknown

In this case, in addition to the prior distributions on β_1 and β_2, we have to assume the prior distribution on σ.

Again following Jeffreys [3] , Zellner and Tiao [5 & 6] , we take the most logical prior distributions on β_1, β_2 and σ as

$$(2.15) \qquad p(\beta_1,\beta_2,\sigma) \propto \frac{1}{\sigma} \qquad \begin{array}{c} c \leqslant \beta_1 \leqslant d \\ -\infty < \beta_2 < \infty \\ 0 < \sigma < \infty \end{array}$$

The elements of β_1, β_2 and log σ are assumed to be uniformly, and locally independently distributed. This type of prior follows from Invariance theory given by Jeffreys.

As before, the joint posterior distribution of β_1, β_2 and σ is,

$$(2.16) \qquad p(\beta_1,\beta_2,\sigma|y) \propto \sigma^{-(T+1)}\exp\{-\frac{1}{2\sigma^2}\left[Q(\beta_2,\hat{\beta}_2-V_{22}^{-1}V_{21}(\beta_1-\hat{\beta}_1),V_{22}) + Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2\right]\}$$

Integrating out β_2 from (2.16), will give the joint posterior distribution of β_1 and σ,

$$(2.17) \qquad p(\beta_1,\sigma|y) \propto \sigma^{-(T-p+r+1)}\exp\{-\frac{1}{2\sigma^2}\left[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2\right]\}$$

Finally integrating (2.17) with respect to σ, we get the marginal posterior distribution of β_1 as,

$$(2.18) \qquad p(\beta_1|y) \propto \left[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2\right]^{-\frac{(T-p+r)}{2}} \qquad c \leqslant \beta_1 \leqslant d$$

The expression (2.18) is in the form of a multivariate 't' distribution, but truncated.

The Bayesian estimate $\tilde{\beta}_1$ which is the mean of (2.18), is given by the following expression,

$$(2.19)\quad \tilde{\beta}_1 = \frac{\int_c^d \beta_1 \left[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2\right]^{-\frac{(T-p+r)}{2}} d\beta_1}{\int_c^d \left[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2\right]^{-\frac{(T-p+r)}{2}} d\beta_1}$$

$$= \hat{\beta}_1 - \frac{1}{(T-p+r-2)(V_{11}-V_{12}V_{22}^{-1}V_{21})}\left\{\frac{\left[Q(d,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2\right]^{-\frac{T-p+r}{2}+1} - \left[Q(c,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2\right]^{-\frac{T-p+r}{2}+1}}{\int_c^d \left[Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2\right]^{-\frac{T-p+r}{2}} d\beta_1}\right\}$$

The evaluation of $\tilde{\beta}_1$ is to be done by numerical integration.

Bayesian estimate of β_2

The joint posterior distribution of β_2 and σ is given by,

$$(2.20)\quad p(\beta_2,\sigma|y) \propto \int_c^d \sigma^{-(T+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[Q(\beta_2,\hat{\beta}_2 - V_{22}^{-1}V_{21}(\beta_1-\hat{\beta}_1),V_{22}) + Q(\beta_1,\hat{\beta}_1,V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2\right]\right\} d\beta_1$$

The marginal posterior distribution of β_2 is obtained by integrating out σ from (2.20) :

$$(2.21)\quad p(\beta_2|y) \propto \int_0^\infty \left[\int_c^d \sigma^{-(T+1)} \exp\{ - \frac{1}{2\sigma^2} [Q(\beta_2, \hat{\beta}_2 - V_{22}^{-1}V_{21}(\beta_1-\hat{\beta}_1), V_{22}) + Q(\beta_1, \hat{\beta}_1, V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2]\} d\beta_1 \right] d\sigma$$

Finally, the Bayesian estimate of β_2 is given by,

$$(2.22)\quad \tilde{\beta}_2 = \frac{\int_\infty^\infty \beta_2 \left[\int_0^\infty \left[\int_c^d \sigma^{-(T+1)} \exp\{ - \frac{1}{2\sigma^2} [Q(\beta_2, \hat{\beta}_2 - V_{22}^{-1}V_{21}(\beta_1-\hat{\beta}_1), V_{22}) + Q(\beta_1, \hat{\beta}_1, V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2]\} d\beta_1 \right] d\sigma \right] d\beta_2}{\int_\infty^\infty \left[\int_0^\infty \left[\int_c^d \sigma^{-(T+1)} \exp\{ - \frac{1}{2\sigma^2} [Q(\beta_2, \hat{\beta}_2 - V_{22}^{-1}V_{21}(\beta_1-\hat{\beta}_1), V_{22}) + Q(\beta_1, \hat{\beta}_1, V_{11}-V_{12}V_{22}^{-1}V_{21}) + (T-p)S^2]\} d\beta_1 \right] d\sigma \right] d\beta_2}$$

As before, simplyfying we get,

$$(2.23)\quad \tilde{\beta}_2 = \hat{\beta}_2 - V_{22}^{-1} V_{21}(\tilde{\beta}_1-\hat{\beta}_1)$$

The relation (2.23) is same as (2.14). Both $\tilde{\beta}_1$ and $\tilde{\beta}_2$ when σ is known will differ from $\tilde{\beta}_1$ and $\tilde{\beta}_2$ when σ is unknown. This is evident from the expressions of $\tilde{\beta}_1$ in two cases (vide (2.11) & (2.19)). The forms of the distributions in two cases are different, the former involves multivariate normal, whereas the latter involves multivariate 't'.

The Bayesian estimators are optimal with respect to the prior distributions and loss functions assumed, for they minimise the average risk. They are also BAN and efficient in comparison to the OLS.

3. Numerical Example [†]

To illustrate the working of the formulas, a consumption-equation relating to the figures 1948-1966 of the Belgian economy is taken:

$$C_t = \beta_1 + \beta_2 W_t + \beta_3 Z_{t-\frac{1}{4}} + \beta_4 L_{t-1} + \beta_5 i_{t-1} + \beta_6 \Delta c_{t-}$$

Explanation of the symbols

All the variables are expressed as relative changes:

$x_t = \dfrac{\tilde{x}_t - \tilde{x}_{t-1}}{\tilde{x}_{t-1}}$, where absolute quantities are indicated by $\sim$.

C	=	private consumption: current value ;
W	=	disposable labour income ;
Z	=	disposable non-labour income ;
L	=	primary and secondary liquidities;
i	=	interest on long dated government securities ;
Δc_t	=	$c_t - c_{t-1}$

From past experience, we can accept the bounds as $.4 \leqslant \beta_2 \leqslant .6$ and $0 \leqslant \beta_4 \leqslant .3$. The other parameters are taken to be unrestricted.

First ordinary least squares (O.L.S.) is applied, and then with the relevant data, numerical integrations and other calculations are performed to obtain the Bayesian estimates.

---

† I am indebted to J. Pompen of the Computer Centre and E. Borghers of Economic faculty of Katholieke Hogeschool, Tilburg, for making necessary programmes and calculations on a I.B.M. 1620^{II} computer.

Parameters	O.L.S.	Bayes Estimators	
		Bounds: $.4<\beta_2<6$; $.0<\beta_4<.3$	
		a)	b)
β_1	- .38877	.78212	.71993
β_2	.55211	.43887	.44129
β_3	.29055	.36131	.35731
β_4	.17748	.05212	.06348
β_5	- .13678	- .13549	- .13529
β_6	- .32183	- .30261	- .30237
$\bar{R}$ †	.90612	.86746	.87179
S ††	1.02610	1.20678	1.18830

† $\bar{R}$ = Multiple correlation coefficient, adjusted for degress of freedom

†† S = least squares estimates of the standard deviation of the error terms

a)b) The numerical integrations are done with trapezoidal rule. The columns a) and b) differ only in that the figures of b) are made more accurate by taking smaller intervals for integrations.

Though in this example O.L.S. estimates are reasonable i.e. they lie already within the bounds according to our apriori belief, nevertheless Bayesian method is applied to show how the estimates can differ in two cases when the apriori informations are explicitly taken into account.

4. Conclusions

The method of estimation given in the preceding sections is quite general and is applicable to the class of problems in regression analysis where a subset of parameters is known to lie within certain ranges apriori. The cases of positive and negative restrictions of the parameters are also incorporated into the method. The only trouble is computational, but with powerful computers this is not impossible.

5. Acknowledgement

I am very much grateful to Prof. T. Rothenberg of University of California, Berkeley for reading this paper and giving valuable suggestions. I also give my thanks to a referee for showing me some simplifications. I am much indebted to my colleague W.H. Vandaele for active encouragement at every stage of the preparation of this paper.

References

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

[2] CHETTY, V.K. "Pooling of Time Series and Cross-Section Data", Econometrica. Vol. 36, April 1968, nr 2, pp. 279-290.

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

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

[5] 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.

[6] 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. 1969