ON THE USE OF SOME EXTRANEOUS INFORMATION IN THE ESTIMATION OF THE COEFFICIENTS OF REGRESSION

YUKIO SUZUKI*

1. Introduction and summary

In this paper, we treat the estimation of regression coefficients, using the extraneous or prior information in addition to the sample information. A similar problem was discussed by Raiffa and Schlaifer [1] and, recently, by Theil [2]. However, their approaches are different. Raiffa and Schlaifer [1] based their discussion on Bayes theorem, whereas Theil [2] applied the generalized least squares method to his problem. Strictly speaking, Theil is concerned mainly with the special regression model which was not treated by Raiffa and Schlaifer. From the Baysian point of view, we wish to discuss the special, but useful regression model which Theil considered of. Special interest will be the case where the prior information is given only for a subset of regression coefficients. In other words, this is the case where only partial prior information is available. Further, when the precision of the regression model which supplies the sample information is unknown, our problem becomes complicated. We will give the exact form of the posterior expected values of regression coefficients, though it is neither manageable nor practical. Then, an approximation and its bias are derived from it.

2. Statement of the problem

The estimation of regression coefficients which is based on some extraneous information as well as the sample information may be considered to be special, but it occurrs in the field of econometric analysis. Also, it has some application in theoretical statistics which is not treated here. For example, when an estimate of the income elasticity of the demand of the goods in question is given extraneously, it seems reasonable to use this information for the estimation of the elasticities (parameters) in the demand equation with which we are concerned. Usually, the extraneous information will arise from the budget investigations, namely,

* On leave from the Institute of Statistical Mathematics, Tokyo.

the cross-section survey.

Let us assume that the data-generating process of the regression model is

(2.1) $$\tilde{y}_t = x_t'\beta + \tilde{\varepsilon}_t{}^{*)}$$

where t is a discrete parameter and

$x_t' = (x_{t1}, x_{t2}, \cdots, x_{tp})$: predetermined variables,

$\beta' = (\beta_1, \beta_2, \cdots, \beta_p)$: regression coefficients.

We assume that $\tilde{\varepsilon}_t$ follows a normal distribution $N(0, \sigma_t^2)$ for all t, and $\tilde{\varepsilon}_t$ and $\tilde{\varepsilon}_{t'}$ are mutually independent if $t \neq t'$. Let a sample of size n be observed and denote it by (y_t, x_t') $(t=1, 2, \cdots, n)$ or simply by (y, X), where $y' = (y_1, \cdots, y_n)$ and X is a matrix of which the ith row is x_i' $(i=1, 2, \cdots, n)$. The observed value (y, X) can be regarded as obtained from the following model:

(2.2) $$\tilde{y} = X\beta + \tilde{\varepsilon}$$

where $\tilde{\varepsilon}' = (\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \cdots, \tilde{\varepsilon}_n)$. Now let us consider the partition of β: $\beta' = (\beta_1', \beta_2')$, where β_1 and β_2 are r and $p-r$ dimensional, respectively. Corresponding to this partition of β, X is partitioned, say (X_1, X_2). Therefore, X_1 is made up of the first r columns of X and X_2 the remaining $p-r$ columns of X. Thus, (2.2) becomes

(2.3) $$\tilde{y} = X_1\beta_1 + X_2\beta_2 + \tilde{\varepsilon}\ .$$

Suppose that β_1 is accurately known extraneously. Then, β_2 will be estimated from the renewed regression model $\tilde{y} - X_1\beta_1 = X_2\beta_2 + \tilde{\varepsilon}$. This is an extreme case. On the other hand, if we have no prior information about β_1 and β_2, the ordinary method of estimation of β is applicable to the model (2.2). This is the other extreme case. In this paper we will consider the intermediate of these two extreme cases. More precisely, we are going to discuss about the situation where we have some prior (or extraneous) information on β_1 for some r $(1 \leqq r \leqq p)$, but no prior information is available on β_2.

When we talk about prior information, we might as well ask about its sources. Of course, different sources of supply of prior information are possible, depending on the relevant situations. However, in the sequel, we restrict ourselves to a special source of supply of the prior information, which will be described in the next section.

In order to combine the prior and the sample information, we take

*) The tilde "$\sim$" on y_t and ε_t means that $\tilde{y}_t$ and $\tilde{\varepsilon}_t$ are random variables.

the Baysian approach. Although we have no prior information on β_2, it is convenient to introduce an artificial prior information on β_2 for the use of Bayes theorem and then eliminate it from the posterior distribution by considering an appropriate limiting distribution.

3. Prior distribution of β

We assume that the prior information on β_1 is given in the form of the observed values on the following data-generating process:

$$\tilde{u}_i = w_i'\beta_1 + \tilde{\eta}_i \qquad (i=1, 2, \cdots, m) \tag{3.1}$$

where $w_i'=(w_{i1}, w_{i2}, \cdots, w_{ir})$ and $\tilde{\eta}_i$ are independently and identically distributed according to the normal distribution $N(0, \sigma_\eta^2)$ $(i=1, 2, \cdots, m)$. Further, the observed values $\{u_i, w_i'\}$ $(i=1, 2, \cdots, m)$ are written as (u, W), where $u'=(u_1, u_2, \cdots, u_m)$, $W'=(w_1, w_2, \cdots, w_m)$ and so (u, W) is regarded as a realized matrix of the random matrix $(\tilde{u}, W)$ which is constrained by the following equation:

$$\tilde{u} = W\beta_1 + \tilde{\eta} \tag{3.2}$$

where $\tilde{\eta}'=(\tilde{\eta}_1, \tilde{\eta}_2, \cdots, \tilde{\eta}_m)$. Clearly, the sufficient statistic with respect to β_1 is (b_1, M), where b_1 is the solution of the normal equation, that is,

$$b_1 = M^{-1}W'u, \quad M = W'W \tag{3.3}$$

under the assumption $|M| \neq 0$.

The distribution of $\tilde{b}_1 = M^{-1}W'\tilde{u}$ is Gaussian with the mean vector β_1 and the variance matrix $M^{-1}\sigma_\eta^2$. Now, let us denote the density function of the p-dimensional normal distribution with the mean vector μ and the variance matrix Σ by $f^{(p)}(x \mid \mu, \Sigma)$. Then, the density function of $\tilde{b}_1$ is written as $f^{(r)}(b_1 \mid \beta_1, M^{-1}\sigma_\eta^2)$.

When the (u, W) is observed and, therefore, the sufficient statistic (b_1, M) is obtained, it seems natural for us to think that the most suitable prior probability measure to be assigned on β_1 is the probability measure with the density function $f^{(r)}(\beta_1 \mid b_1, M^{-1}\sigma_\eta^2)$. Further, we assume the prior probability measure of β_2 to be the normal distribution $f^{(p-r)}(\beta_2 \mid 0, I\sigma^2)$, where, without loss of generality, the mean vector can be regarded as the zero vector 0 of $p-r$ dimensions and I is the unit matrix of order $p-r$. We assume that the two random vectors $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are mutually independent with respect to the prior probability distribution of β. Hence, the joint density function of $\tilde{\beta}_1$ and $\tilde{\beta}_2$ is

$$f^{(r)}(\beta_1 \mid b_1, M^{-1}\sigma_\eta^2) f^{(p-r)}(\beta_2 \mid 0, I\sigma^2) = f^{(p)}\left[\beta \,\middle|\, \begin{pmatrix} b_1 \\ 0 \end{pmatrix}, \begin{pmatrix} M^{-1}\sigma_\eta^2 & 0 \\ 0 & I\sigma^2 \end{pmatrix}\right] \equiv p'_\sigma(\beta) . \tag{3.4}$$

Now, the situation where no prior information about β_2 is available corresponds to the extreme case where σ tends to infinity. Therefore, after the necessary procedures are conducted, we will make σ go to infinity.

4. The posterior distribution of β when σ_ϵ is known

When we have an observed value (y, X) in the model (2.2), the sufficient statistic of β is (b, N) which is obtained by

$$b = N^{-1}X'y , \quad N = X'X \tag{4.1}$$

where we assume that N is nonsingular and σ_ϵ^2 in the model (2.2) is known. Since the distribution of $\tilde{b} = N^{-1}X\tilde{y}$ is normal with the density function:

$$f^{(p)}(b \mid \beta, N^{-1}\sigma_\epsilon^2) \equiv p(b \mid \beta) , \tag{4.2}$$

the posterior density function of β with respect to $p'_\sigma(\beta)$, denoted by $p''_\sigma(\beta \mid b)$, is obtained by Bayes theorem, that is,

$$p''_\sigma(\beta \mid b) = p(b \mid \beta) p'_\sigma(\beta) / \int p(b \mid \beta) p'_\sigma(\beta) d\beta . \tag{4.3}$$

By (3.4), (4.2) and (4.3), we have

$$p''_\sigma(\beta \mid b) = f^{(p)}[\beta \mid a^{(\sigma)}, (N\sigma_\epsilon^{-2} + Q_\sigma)^{-1}] , \tag{4.4}$$

where

$$\begin{aligned} a^{(\sigma)} &= (N\sigma_\epsilon^{-2} + Q_\sigma)^{-1}\left[\sigma_\epsilon^{-2}Nb + \begin{pmatrix} \sigma_\eta^{-2}Mb_1^{(1)} \\ 0 \end{pmatrix}\right] \\ &= b + (N\sigma_\epsilon^{-2} + Q_\sigma)^{-1}\left[\begin{pmatrix} \sigma_\eta^{-2}Mb_1^{(1)} \\ 0 \end{pmatrix} - Q_\sigma b\right] \\ Q_\sigma &= \begin{pmatrix} M\sigma_\eta^{-2} & 0 \\ 0 & I\sigma^{-2} \end{pmatrix} . \end{aligned} \tag{4.5}$$

Now, let us consider the limiting case: $\lim_{\sigma \to \infty} p''_\sigma(\beta \mid b) = p''_\infty(\beta \mid b)$, then

$$p''_\sigma(\beta \mid b) = f^{(p)}[\beta \mid a^{(\infty)}, (N\sigma_\epsilon^{-2} + Q_\infty)^{-1}] \tag{4.6}$$

where

$$a^{(\infty)} = (N\sigma_\epsilon^{-2} + Q_\infty)^{-1}\left[\sigma_\epsilon^{-2}Nb + \begin{pmatrix} \sigma_\eta^{-2}Mb_1^{(1)} \\ 0 \end{pmatrix}\right] \tag{4.7}$$

$$=b+(N\sigma_\epsilon^{-2}+Q_\infty)^{-1}\begin{pmatrix}\sigma_\eta^{-2}M(b_1^{(1)}-b_1)\\ 0\end{pmatrix}$$

$$Q_\infty=\begin{pmatrix}M\sigma_\eta^{-2} & 0\\ 0 & 0\cdot I\end{pmatrix}$$

where $b'=(b_1', b_2')$ and b_1 consists of the first r components of b. Further, let us partition N as follows: $N=\begin{pmatrix}N_{11} & N_{12}\\ N_{21} & N_{22}\end{pmatrix}$, where N_{11}, N_{12}, N_{21} and N_{22} are $r\times r$, $r\times(p-r)$, $(p-r)\times r$ and $(p-r)\times(p-r)$, respectively. Then, as easily seen, we have

$$(N\sigma_\epsilon^{-2}+Q_\infty)^{-1}=\begin{pmatrix}R^{-1} & -R^{-1}N_{12}N_{22}^{-1}\\ -N_{22}^{-1}N_{21}R^{-1} & (N_{22}\sigma_\epsilon^{-2}-N_{21}R^{-1}N_{12}\sigma_\epsilon^{-4})^{-1}\end{pmatrix} \tag{4.8}$$

where $R=N_{11}^*\sigma_\epsilon^{-2}+M\sigma_\eta^{-2}$ and $N_{11}^*=N_{11}-N_{12}N_{22}^{-1}N_{21}$. Therefore, (4.7) is written as follows:

$$a^{(\infty)}=b+\begin{pmatrix}R^{-1}\\ -N_{22}^{-1}N_{21}R^{-1}\end{pmatrix}\sigma_\eta^{-2}M(b_1^{(1)}-b_1)\,. \tag{4.9}$$

Thus we have

(4.10)

$$a_1^{(\infty)}=b_1+\sigma_\eta^{-2}R^{-1}M(b_1^{(1)}-b_1)=b_1+\left(N_{11}^*\frac{\sigma_\eta^2}{\sigma_\epsilon^2}+M\right)^{-1}M(b_1^{(1)}-b_1)$$

$$a_2^{(\infty)}=b_2-\sigma_\eta^{-2}N_{22}^{-1}N_{21}R^{-1}M(b_1^{(1)}-b_1)=b_2-N_{22}^{-1}N_{21}\left(N_{11}^*\frac{\sigma_\eta^2}{\sigma_\epsilon^2}+M\right)^{-1}M(b_1^{(1)}-b_1)\,.$$

Clearly, we can see that $a^{(\infty)}\to b$ when $\sigma_\eta\to\infty$, and $a^{(\infty)}\to b$ when $b_1^{(1)}\to b_1$. Further, if $\frac{\sigma_\epsilon}{\sigma_\eta}\gg 1$, $a_1^{(\infty)}\doteqdot b_1^{(1)}$ and $a_2^{(\infty)}\doteqdot b_2-N_{22}^{-1}N_{21}(b_1^{(1)}-b_1)$. Having obtained the posterior distribution of β, we can take $a^{(\infty)}$ as an estimate of β which is unbiased, of minimum variance and of maximum likelihood in the sense of the posterior distribution.

5. The case where σ_ϵ is unknown

Theil [2] treated the estimation of β when σ_ϵ is unknown, the test of the compativility of prior and sample information, and also the shares of prior and sample information in the posterior precision. Concerning the estimation of β, he introduced f-estimate $\hat\beta_f$, which corresponds to $a^{(\sigma)}$ in (4.5) or $a^{(\infty)}$ in (4.7) with f in the place of σ_ϵ^{-2}. Let $\hat\beta_\varphi$ be the f-estimate of which the suffix φ equals σ_ϵ^{-2}. In other words, $\hat\beta_\varphi$ is the estimate of β obtained in section 4 when σ_ϵ^2 is known. Theil shows that, if $\varphi-f=O(n^{-1/2})$, $\hat\beta_f=\hat\beta_\varphi+O(n^{-1})$, where n is the sample size. Since $\hat\beta_\varphi$

is unbiased, $\hat{\beta}_f$ is asymptotically unbiased, if $\varphi-f=O(n^{-1/2})$. An obvious choise of f is $\frac{1}{s^2}=\frac{n-p}{y'y-y'X(X'X)^{-1}X'y}$. Now, we will derive analogous results by employing Baysian approach. First of all, the prior information is supplied by the observed value (u, W) of the regression model (3.2). However, it must be noted that this prior information gives us no information about the unknown σ_ε^2. As far as our problem is concerned, only the sample information about σ_ε^2 is available to us.

The following is the presentation of sample likelihood and the posterior distribution of β which is obtained by the combination of the prior and sample informations. When the precision $h=\sigma_\varepsilon^{-2}$ of the regression model (2.2) is unknown, the sufficient statistic of β and h is (b, v, N, ν) for the observed value (y, X), where $n=\nu+p$, $v=\frac{1}{\nu}(y-Xb)'(y-Xb)$, $p=\text{rank}(X)$, $b=N^{-1}X'y$ and $N=X'X$. As a matter of fact, the sample likelihood function, say $p(b, v\mid\beta, h)$, is

$$p(b, v\mid\beta, h)=(2\pi)^{-n/2}|N|^{1/2}e^{-(1/2)h(b-\beta)'N(b-\beta)}h^{p/2}e^{-(1/2)h\nu v}h^{\nu/2}. \tag{5.1}$$

Since the prior distribution of β is normal with the mean vector $\begin{pmatrix} b_1 \\ 0 \end{pmatrix}$ and the variance matrix Q_σ for an arbitrary positive σ as section 4, we have

$$p_\sigma'(\beta)=(2\pi)^{-p/2}|M|^{1/2}\sigma_\eta^{-r}\sigma^{r-p}\exp\left\{-\frac{1}{2}\left[\beta-\begin{pmatrix} b_1^{(1)} \\ 0 \end{pmatrix}\right]'Q_\sigma\left[\beta-\begin{pmatrix} b_1^{(1)} \\ 0 \end{pmatrix}\right]\right\}. \tag{5.2}$$

Therefore, the posterior density function $p_\sigma''(\beta, h\mid b, v)$ of β and h is obtained by Bayes theorem, that is

$$p_\sigma''(\beta, h\mid b, v)=\frac{p_\sigma'(\beta)p(b, v\mid\beta, h)}{\int_0^\infty dh\int_{E^{(p)}}p_\sigma'(\beta)p(b, v\mid\beta, h)d\beta} \tag{5.3}$$

where $E^{(p)}$ is the p-dimensional Euclidean space. From (5.1) and (5.2), we have

$$\begin{aligned} p_\sigma'(\beta)p(b, v\mid\beta, h)&=(2\pi)^{-(p+n)/2}|N|^{1/2}|M|^{1/2}\sigma_\eta^{-r}\sigma^{r-p} \\ &\cdot\exp\left\{-\frac{1}{2}(b-b^*)'Q_\sigma(b-b^*)\right\} \\ &\cdot\exp\left\{-\frac{1}{2}(\beta-a^{(\sigma)})'(Q_\sigma+hN)(\beta-a^{(\sigma)})\right\} \\ &\cdot h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-K_\sigma)\right\} \end{aligned} \tag{5.4}$$

where $K_\sigma=(b-b^*)'Q_\sigma(Q_\sigma+hN)^{-1}Q_\sigma(b-b^*)$, $b^{*\prime}=(b_1^{(1)\prime}, 0)$: p-dimensional. For the simplicity, let us denote

$$(5.5)\qquad \begin{aligned} p'_\sigma(\beta)p(b, v\,|\,\beta, h)&=C(2\pi)^{-p/2}\exp\left\{-\frac{1}{2}(\beta-a^{(\sigma)})'(Q_\sigma+hN)(\beta-a^{(\sigma)})\right\} \\ &\quad\cdot h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-K_\sigma)\right\}, \end{aligned}$$

where $C=(2\pi)^{-n/2}|N|^{1/2}|M|^{1/2}\sigma_\eta^{-r}\sigma^{r-p}\exp\left\{-\frac{1}{2}(b-b^*)'Q_\sigma(b-b^*)\right\}$.

On the other hand, we have

$$(5.6)\qquad \begin{aligned} &\int_0^\infty dh\int_{E^{(p)}}p'_\sigma(\beta)p(b, v\,|\,\beta, h)d\beta \\ &=C\int_0^\infty\Big[h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-K_\sigma)\right\}|Q_\sigma+hN|^{-1/2}(2\pi)^{-p/2}|Q_\sigma+hN|^{1/2} \\ &\qquad\cdot\int_{E^{(p)}}\exp\left\{-\frac{1}{2}(\beta-a^{(\sigma)})'(Q_\sigma+hN)(\beta-a^{(\sigma)})\right\}d\beta\Big]dh \\ &=C\int_0^\infty|Q_\sigma+hN|^{-1/2}h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-K_\sigma)\right\}dh\,. \end{aligned}$$

Hence,

(5.7) $p''_\sigma(\beta, h\,|\,b, v)$

$$=\frac{(2\pi)^{-p/2}\exp\left\{-\frac{1}{2}(\beta-a^{(\sigma)})'(Q_\sigma+hN)(\beta-a^{(\sigma)})\right\}h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-K_\sigma)\right\}}{\int_0^\infty|Q_\sigma+hN|^{-1/2}h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-K_\sigma)\right\}dh}.$$

When $\sigma\to\infty$, we obtain $p''_\infty(\beta, h\,|\,b, v)$ by replacing Q_σ and $a^{(\sigma)}$ in (5.7) respectively by Q_∞ and $a^{(\infty)}$, where $Q_\infty=\lim_{\sigma\to\infty}Q_\sigma$ and $a^{(\infty)}=\lim_{\sigma\to\infty}a^{(\sigma)}$.

When the loss function of our estimation problem is quadratic, the optimal estimate of β will be the mean of the posterior distribution of β. Let us denote it by $E''\tilde{\beta}^{(\sigma)}$, that is,

$$E''\tilde{\beta}^{(\sigma)}=\iint\beta p''_\sigma(\beta, h\,|\,b, v)d\beta dh\,.$$

Using (5.7), we have

$$(5.8)\qquad E''\tilde{\beta}^{(\sigma)}=\frac{\int_0^\infty a^{(\sigma)}|Q_\sigma+hN|^{-1/2}h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-K_\sigma)\right\}dh}{\int_0^\infty|Q_\sigma+hN|^{-1/2}h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-K_\sigma)\right\}dh}.$$

Since $a^{(\sigma)}=b-(Q_\sigma+hN)^{-1}Q_\sigma(b-b^*)$, we obtain

$$E''\tilde{\beta}^{(\sigma)}=b-\frac{A_\sigma}{d_\sigma}Q_\sigma(b-b^*)\,, \tag{5.9}$$

where

$$\begin{aligned} A_\sigma&=\int_0^\infty (Q_\sigma+hN)^{-1}|Q_\sigma+hN|^{-1/2}h^{-n/2}\exp\left\{-\frac{1}{2}(h\nu v-K_\sigma)\right\}dh \\ d_\sigma&=\int_0^\infty |Q_\sigma+hN|^{-1/2}h^{-n/2}\exp\left\{-\frac{1}{2}(h\nu v-K_\sigma)\right\}dh\,. \end{aligned} \tag{5.10}$$

For the limiting case, that is, when $\sigma\to\infty$, we have

$$E''\tilde{\beta}=b-\frac{A_\infty}{d_\infty}Q_\infty(b-b^*)\,, \tag{5.11}$$

where $A_\infty=\lim_{\sigma\to\infty}A_\sigma$ and $d_\infty=\lim_{\sigma\to\infty}d_\sigma$.

Since Q_∞ is non-negative definite and N is positive definite, there exists a nonsingular matrix F such that

$$F'NF=I \quad\text{and}\quad F'Q_\infty F=D \tag{5.12}$$

where I is the identity matrix of order p and D is a diagonal matrix with diagonal elements λ_i $(i=1, 2, \cdots, p)$. Therefore,

$$F'(Q_\infty+hN)F=D+hI,\qquad F^{-1}(Q_\infty+hN)^{-1}(F^{-1})'=(D+hI)^{-1}$$

and

$$|Q_\infty+hN|^{-1}=|F||D+hI|^{-1}|F'|=|F|^2\prod_{i=1}^{p}(\lambda_i+h)^{-1}.$$

Furthermore, we have

$$\begin{aligned} K_\infty=\lim_{\sigma\to\infty}K_\sigma&=(b-b^*)'Q_\infty(Q_\infty+hN)^{-1}Q_\infty(b-b^*) \\ &=(b-b^*)'Q_\infty F(D+hI)^{-1}F'Q_\infty(b-b^*) \\ &=\sum_{i=1}^{p}\rho_i^2(\lambda_i+h)^{-1} \end{aligned}$$

where $\rho'=(\rho_1, \rho_2, \cdots, \rho_p)$ is defined by $\rho=F'Q_\infty(b-b^*)$. Hence, we have

$$\begin{aligned} A_\infty=\int_0^\infty &F(D+hI)^{-1}F'|F|\left\{\prod_{i=1}^{p}(\lambda_i+h)^{-1/2}\right\}h^{-n/2} \\ &\cdot\exp\left\{-\frac{1}{2}\left(h\nu v-\sum_{i=1}^{p}\rho_i^2(\lambda_i+h)^{-1}\right)\right\}dh \end{aligned} \tag{5.13}$$

$$d_\infty=\int_0^\infty |F| \left\{\prod_{i=1}^{p} (\lambda_i+h)^{-1/2}\right\} h^{n/2} \exp\left\{-\frac{1}{2}(h\nu v-\sum_{i=1}^{p}\rho_i^2(\lambda_i+h)^{-1}\right\} dh$$

and

$$E''\tilde{\beta}=b-FGF'Q_\infty(b-b^*)\,, \tag{5.14}$$

where G is a diagonal matrix: $G=(g_{ij})$ $(i, j=1, 2, \cdots, p)$,

(5.15)

$$g_{ii}=\frac{\int_0^\infty(\lambda_i+h)^{-1}\left\{\prod_{i=1}^{p}(\lambda_i+h)^{-1/2}\right\}h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-\sum_{i=1}^{p}\rho_i^2(\lambda_i+h)^{-1})\right\}dh}{\int_0^\infty\left\{\prod_{i=1}^{p}(\lambda_i+h)^{-1/2}\right\}h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-\sum_{i=1}^{p}\rho_i^2(\lambda_i+h)^{-1})\right\}dh}$$

$g_{ij}=0$ if $i\neq j$.

(N.B.) In (5.12), we can assume without loss of generality that $\lambda_i>0$ for $i=1, 2, \cdots, r$ and $\lambda_i=0$ for $i=r+1, \cdots, p$, where $r=$ rank of M. Then, it follows that $\rho_i=0$ for $i=r+1, \cdots, p$. For, $\rho=F'Q_\infty(b-b^*)=DF^{-1}(b-b^*)$.

Let us denote $g_{ii}=\left(\lambda_i+\frac{1}{v}\right)^{-1}+t_i$ and

(5.16) $t_i=$

$$\frac{\int_0^\infty\left[(\lambda_i+h)^{-1}-(\lambda_i+v^{-1})^{-1}\right]\left\{\sum_{i=1}^{p}(\lambda_i+h)^{-1/2}\right\}h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-\sum\rho_i^2(\lambda_i+h)^{-1})\right\}dh}{\int_0^\infty\left\{\sum_{i=1}^{p}(\lambda_i+h)^{-1/2}\right\}h^{n/2}\exp\left\{-\frac{1}{2}(h\nu v-\sum\rho_i^2(\lambda_i+h)^{-1})\right\}dh}$$

$(i=1, 2, \cdots, p)$.

Then, we have

$$G=(D+v^{-1}I)^{-1}+T \tag{5.17}$$

where T is a diagonal matrix with the ith diagonal element t_i $(i=1, 2, \cdots, p)$. Noticing that $F(D+v^{-1}I)^{-1}F'=(Q_\infty+v^{-1}N)^{-1}$, we obtain

$$\begin{aligned} E''\tilde{\beta}&=b-(Q_\infty+v^{-1}N)^{-1}Q_\infty(b-b^*)-FTF'Q_\infty(b-b^*) \\ &=b-(Q_\infty+v^{-1}N)^{-1}Q_\infty(b-b^*)-FT\rho\,. \end{aligned} \tag{5.18}$$

As easily seen, $b-(Q_\infty+v^{-1}N)^{-1}Q_\infty(b-b^*)$ is the same as $a^{(\infty)}$ in (4.7) or (4.9) if σ_t^{-2} is replaced by v^{-1}, so let us denote it by $a_v^{(\infty)}$. The formula (5.18) tells us that, when we use $a_v^{(\infty)}$ as an estimate of β, the bias is

$-FT\rho$. However, it is difficult to obtain the exact values of t_i, diagonal elements of T. Therefore, in the next section, we will give inequalities to evaluate the size of the bias of $a_v^{(\infty)}$ from $E''\tilde{\beta}$.

6. Evaluation of the bias of $a_v^{(\infty)}$ from $E''\tilde{\beta}$

Since $\rho_i=0$ if $i>r$, we have $(T\rho)'=(t_1\rho_1, t_2\rho_2, \cdots, t_r\rho_r, 0, \cdots, 0)$. Let us denote the ith column vector of F' by f_i $(i=1, 2, \cdots, p)$. Further, let f_i^0 be the column vector which consists of the first r elements of f_i, and let $(t_1\rho_1, t_2\rho_2, \cdots, t_r\rho_r)=\gamma'$, and let $(\rho_1, \rho_2, \cdots, \rho_r)=\rho^0$. Then,

$$[FT\rho]_i=(f_i^0, \gamma), \quad (i=1, 2, \cdots, p).^{*)} \tag{6.1}$$

Therefore, we have the following inequalities:

$$\begin{aligned} |[FT\rho]_i| &= |(f_i^0, \gamma)| \\ &\leqq \|f_i^0\| \|\gamma\| \quad \text{(Schwartz' inequality)} \\ &\leqq \{\max_{1\leqq k\leqq r}|t_k|\} \|f_i^0\| \|\rho^0\| & (6.2) \\ &\leqq \{\max_{1\leqq k\leqq r}|t_k|\} \|f_i\| \|\rho\| \qquad (i=1, 2, \cdots, p). & (6.3) \end{aligned}$$

Since $N^{-1}=FF'$, we have

$$\|f_i\| = \sqrt{d_i(N^{-1})} \quad (i=1, 2, \cdots, p) \tag{6.4}$$

where $d_i(N^{-1})$ stands for the ith element on the main diagonal of N^{-1}. Further, as $\rho=F'Q_\infty(b-b^*)$, we have

$$\begin{aligned} \|\rho\|^2 &= (b-b^*)'Q_\infty FF'Q_\infty(b-b^*) \\ &= (b-b^*)'Q_\infty N^{-1}Q_\infty(b-b^*) \\ &= (b-b_1^{(1)})'MN_{11}^*M(b-b_1^{(1)})\sigma_\eta^{-4} \end{aligned} \tag{6.5}$$

where $N_{11}^*=N_{11}-N_{12}N_{22}^{-1}N_{21}$. Hence, we have

$$|[E''\tilde{\beta}-a_v^{(\infty)}]_i| \leqq \{\max_{1\leqq k\leqq r}|t_k|\}[d_i(N^{-1})]^{1/2}[(b_1-b_1^{(1)})'MN_{11}^*M(b_1-b_1^{(1)})]^{1/2}\sigma_\eta^{-2}. \tag{6.6}$$

Now, let us go to the evaluation of $|t_k|$ $(1\leqq k\leqq r)$. Let us consider the probability density function $f(h)$ which is defined as follows:

$$f(h)=\frac{h^{(r+\nu)/2}\prod_{i=1}^{r}(\lambda_i+h)^{-1/2}\exp\left\{-\frac{1}{2}(h\nu v-\sum_{i=1}^{r}\rho_i^2(\lambda_i+h)^{-1})\right\}}{\int_0^\infty h^{(r+\nu)/2}\prod_{i=1}^{r}(\lambda_i+h)^{-1/2}\exp\left\{-\frac{1}{2}(h\nu v-\sum_{i=1}^{r}\rho_i^2(\lambda_i+h)^{-1})\right\}dh}. \tag{6.7}$$

---

*) For a vector $x'=(x_1, x_2, \cdots, x_n)$, we define $[x]_i=x_i$ and $\|x\|=\left(\sum_{i=1}^{n}x_i^2\right)^{1/2}$.

From (5.16), if $\tilde{h}$ is a random variable with the density function $f(h)$, t_k is expressed as follows:

$$t_k = E[(\lambda_k+\tilde{h})^{-1}-(\lambda_k+v^{-1})^{-1}]$$
$$=(\lambda_k+v^{-1})^{-1}E[(v^{-1}-\tilde{h})(\lambda_k+\tilde{h})^{-1}].$$

By Schwartz' inequality, we have

$$t_k^2 \leqq (\lambda_k+v^{-1})^{-2}E(v^{-1}-\tilde{h})^2E(\lambda_k+\tilde{h})^{-2}. \tag{6.8}$$

Since $\lambda_k>0$ if $1\leqq k\leqq r$, the following inequalities hold; $E(\lambda_k+\tilde{h})^{-2}<\lambda_k^{-2}$ $(k=1, 2, \cdots, r)$. Thus, we have

$$t_k^2<(\lambda_k^2+\lambda_k v^{-1})^{-2}E(\tilde{h}-v^{-1})^2 \quad (k=1, 2, \cdots, r). \tag{6.9}$$

Now, let us evaluate $E(\tilde{h}-v^{-1})^2$. For the sake of simplicity let us define I, I_1, and I_2 as follows:

$$I=\int_0^\infty h^{(\nu+r)/2}K(h)dh$$
$$I_1=\int_0^\infty h^{(\nu+r)/2+1}K(h)dh\,,\quad I_2=\int_0^\infty h^{(\nu+r)/2+2}K(h)dh\,, \tag{6.10}$$

where $\quad K(h)=\exp\left[-\frac{1}{2}(h\nu v+\sum_1^r \log(\lambda_i+h)-\sum_1^r \rho_i^2(\lambda_i+h)^{-1})\right].$

Then we have

$$E\tilde{h}=I_1/I, \quad E\tilde{h}^2=I_2/I. \tag{6.11}$$

Now, employing integration by parts, we have

$$I=-\frac{2}{\nu+r+2}\int_0^\infty h^{(\nu+r)/2+1}K'(h)dh\,. \tag{6.12}$$

By performing integration by parts once again, we get

$$I=\frac{4}{(\nu+r+2)(\nu+r+4)}\int_0^\infty h^{(\nu+r)/2+2}K''(h)dh\,. \tag{6.13}$$

From the definition of $K(h)$,

$$K'(h)=K(h)\left[-\frac{1}{2}\left(\nu v+\sum_1^r(\lambda_i+h)^{-1}+\sum_1^r \rho_i^2(\lambda_i+h)^{-2}\right)\right] \tag{6.14}$$

$$(6.15)\qquad K''(h)=K(h)\Big[-\frac{1}{2}\Big(\nu v+\sum_1^r(\lambda_i+h)^{-1}+\sum_1^r\rho_i^2(\lambda_i+h)^{-2}\Big)\Big]^2 +K(h)\Big(\frac{1}{2}\sum_1^r(\lambda_i+h)^{-2}+\sum_1^r\rho_i^2(\lambda_i+h)^{-3}\Big).$$

From (6.14) we obtain the following inequality:

$$(6.16)\qquad -K'(h)<\frac{1}{2}K(h)(\nu v+\sum_1^r\lambda_i^{-1}+\sum_1^r\rho_i^2\lambda_i^{-2}).$$

From (6.15) we have

$$(6.17)\qquad K''(h)>\frac{1}{4}K(h)v^2\nu^2.$$

Therefore, from (6.10), (6.12) and (6.13) we have

$$(6.18)\qquad \frac{v^2\nu^2}{(\nu+r+2)(\nu+r+4)}I_2<I<\frac{v\nu+\sum_1^r\lambda_i^{-1}+\sum_1^r\rho_i^2\lambda_i^{-2}}{\nu+r+2}I_1.$$

That is, we have

$$(6.19)\qquad \begin{aligned} E(\tilde{h})&=I_1/I>\frac{\nu+r+2}{v\nu+\sum_1^r(\lambda_i^{-1}+\rho_i^2\lambda_i^{-2})}\\ E(\tilde{h}^2)&=I_2/I<\frac{(\nu+r+2)(\nu+r+4)}{v^2\nu^2}.\end{aligned}$$

Consequently, we have the following inequality:

$$(6.20)\qquad E(\tilde{h}-v^{-1})^2<\frac{1}{v^2}\Big\{\frac{2\{1+v^{-1}\sum_1^r(\lambda_i^{-1}+\rho_i\lambda_i^{-2})\}}{\nu+v^{-1}\sum_1^r(\lambda_i^{-1}+\rho_i\lambda_i^{-2})}+\frac{2(r+3)v^{-1}\sum_1^r(\lambda_i^{-1}+\rho_i^2\lambda_i^{-2})}{\nu\{\nu+v^{-1}\sum_1^r(\lambda_i^{-1}+\rho_i^2\lambda_i^{-2})\}} +\frac{(r+2)(r+4)}{\nu^2}\Big\}.$$

For the sake of simplicity, let us denote the right-hand side of the above inequality by $q^2(\nu, r, v, \lambda, \rho)$; then, from (6.6), (6.9) and (6.20), we obtain

$$(6.21)\qquad |[E''\tilde{\beta}-a_v^{(\infty)}]_i|<\{\max_{1\le k\le r}(\lambda_k^2+\lambda_k v^{-1})^{-1}\}q(\nu, r, v, \lambda, \rho)(d_i(N^{-1}))^{1/2} \cdot[(b_1-b_1^{(1)})'MN_{11}^*M(b_1-b_1^{(1)})]^{1/2}\sigma_\eta^{-2}\quad (i=1, 2, \cdots, p).$$

That is, (6.21) gives an upper bound for the bias of the estimator $a_v^{(\infty)}$.

Now, let us consider the order of the bias when the sample size increases. Obviously, $q(\nu, r, v, \lambda, \rho)=O(\nu^{-1/2})$ from (6.20). As $\lim_{\nu\to\infty} \nu^{-1}X'X$ is a moment matrix, we assume here that it is positive definite. Under this assumption, we have the following results:

(i) Each component of $N=X'X$ is $O(\nu)$

(ii) Each component of N^{-1} is $O(\nu^{-1})$.

Therefore, we have

$$(d_i(N^{-1}))^{1/2}=O(\nu^{-1/2})$$

Each component of $N^*_{11}=O(\nu^{-1})$.

From (6.21), we finally have

$$[E''\tilde{\beta}-a_v^{(\infty)}]_i=O(\nu^{-3/2}) \quad (i=1, 2, \cdots, p).$$

This is more accurate than Theil's evaluation. It is naturally expected because we restricted the model by the assumption of normality of distribution.

7. Acknowledgment

The author is heartily thankful to Dr. Z. Govindarajulu for reading the draft carefully and giving him valuable suggestions.

THE INSTITUTE OF STATISTICAL MATHEMATICS
CASE INSTITUTE OF TECHNOLOGY, OHIO, U.S.A.

REFERENCES

[1] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*, Division of Research, Graduate School of Business Administration, Harvard University, Boston, 1961.

[2] H. Theil, "On the use of incomplete prior information in regression analysis", *J. Amer. Statist. Ass.*, 58 (1963), 401–414.