J. Natn. Sci. Coun. Sri Lanka 1982 10(1) : 1—11

On the use of Complete and Incomplete Information in Regression Analysis

S. WEERAHANDI
Department of Mathematics, University of Sri Jayewardenepura, Nugegoda, Sri Lanka.

(Date of receipt : 1 December 1980)
(Date of acceptance : 30 December 1981)

Abstract : Theil's Theory of restricted least squares is extended to the case of mixed linear and quadratic constraints on the parameters of the multiple regression model which is, for instance, the case arising in estimating a class of nonlinear models. The 'mixed estimation problem' of Theil and Goldberger (1961) and Theil (1963) is also considered when the estimates of the parameters are required to satisfy any exact linear restriction, and explicit formulae for 'restricted least squares estimates in mixed estimation' are established.

1. Introduction

Suppose the statistician has some prior knowledge on the Parameters of a regression in addition to the sample observations. He may have knowledge of the sum of some coefficients, of the values of some coefficients, of the certain relationships between the coefficients, or of merely the signs of some coefficients. The statistician may have derived this information from economic theory or perhaps from previous statistical work. Such knowledge that the statistician might have in addition to the prevailing sample information is refered to as extraneous information. The advantage of incorporating this extraneous information is clear: it is intuitive that a gain in efficiency will result provided that the available information is utilised properly and efficiently. This extraneous information may be used to improve the estimates of the unknown parameters.

The first three sections of this paper will be devoted to discuss the restricted least squares. The estimation results of the restricted least squares under exact linear restrictions is reproduced in Section 2. The purpose of the third section is to establish an estimation procedure when the coefficients are related by a quadratic equation whereas in the fourth section the same is done under restrictions on the coefficients up to an equation of second order. A numerical example is given in each of these two sections to illustrate the procedure. The last two sections are concerned with the mixed estimation. The results of the classical mixed estimation are reproduced in Section 4. As an extension, in section 5, similar results are obtained for mixed estimation under linear restrictions on the coefficients.

Throughout this paper we shall make the usual assumptions on the regression model $y = X\beta + \varepsilon$, that the disturbances are uncorrelated, homoscedastic, and each has Zero means.

$$E(\varepsilon) = O$$
$$E(\varepsilon\varepsilon') = \sigma^2 I$$

with the N x K matrix of X of rank K fixed in repeated samples where y and E are N x I vectors and β is the KxI parameter vector.

2. Exact Linear Restrictions, Restricted Least Squares

Suppose we know a set of linear relationships between the parameters. Situations of this kind arise often in Economic theory, for example, the sum of the exponents in a Cobb-Douglas production function is known to be one, sum of the money income and price elasticities in a demand function is known to be zero. One way of dealing with this kind of a problem is to incorporate the restriction in the fitting process in such a way that the restriction is exactly satisfied by the estimated coefficients. A general formula for the least squares estimators of the parameters subject to exact linear restrictions was first developed by Theil (1) in 1961. The following is an outline of that proof.

The given set of linear restrictions on the parameters may be expressed compactly as,

$$\mathbf{r} = \mathbf{R}\beta$$

where $\mathbf{r}$ is a known column vector of $g<k$ elements, g being the number of restrictions, and $\mathbf{R}$ is a $g \times k$ known matrix. In order to find $\mathbf{b}$, the desired "restricted least squares estimator" of β, the sum of squared residuals $(\mathbf{y} - \mathbf{Xb})(\mathbf{y} - \mathbf{Xb})$ is minimized subject to $(\mathbf{Rb} - \mathbf{r})$. Using the Lagrangian minimization it is shown that this occurs when,

$$\mathbf{b} = \hat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(\mathbf{r} - \mathbf{R}\hat{\beta})$$

where $\hat{\beta} = (X'X)^{-1} X'y$ is the ordinary (unrestricted) least squares estimator. It can be easily shown comparing the variances of the estimates $\mathbf{b}$ and $\hat{\beta}$ that there has been a gain in efficiency in using the extraneous information which consists of exact linear restrictions on the parameters.

3. Least Squares Under Quadratic Constraints

It is of mathematical interest now to find the restricted least squares estimators when the extraneous information consists of exact quadratic restrictions on the coefficients, because of its simplicity. Such a treatment is useful, in particular, in estimating the parameters of nonlinear models such as $Y = \theta + (\alpha X_1 + \beta X_2)(\gamma + X_3)$ which is equivalent to the linear model $\lambda = \theta + \alpha X_1 X_3 + \beta X_2 X_3 + \lambda_1 X_1 + \lambda_2 X_2$ with the restriction $\alpha\lambda_1 - \beta\lambda_2 = O$. First we consider the case where only a single constraint on the coefficients of the classical regression model is present.

Let the single quadratic relationship between the coefficients of the linear regression model $y = X\beta + \varepsilon$, be

$$C = \beta' R \beta$$

where C is a known scaler and R is a known symmetric matrix of order k. For example, if we wish to incorporate the restriction $\beta^2 + 4\,\beta\beta_2 + 2\,\beta\beta_3 + 3\beta_3^2 = 3$ we would set,

$$C = 3 \quad \text{and} \quad R = \begin{pmatrix} 1 & 2 & 1 & 0 & 0 & . & . & 0 \\ 2 & 0 & 0 & 0 & 0 & . & . & 0 \\ 1 & 0 & 3 & 0 & 0 & . & . & 0 \\ 0 & 0 & 0 & 0 & 0 & . & . & 0 \end{pmatrix}$$

It should be pointed out that it is not always possible to find the restricted least squares subject to a quadratic constraint on the parameters by direct substitution, unlike in linear constraints. However we can make use of the Lagrange minimization to do this. The estimates of the parameters obtained by restricted least squares in this case, will not be unbiased. Nevertheless, these estimates possess smaller mean square error over the unrestricted least square estimates.

We seek for the estimated coefficient vector, b, to satisfy the restrictions, so we must find b that minimizes $(y - Xb)'(\,(y - Xb)$ subject to the restriction $C = b'R\,b$. therefore we minimize,

$$F(b,\lambda) = (y - Xb)'(y - Xb) - \lambda\,(c - b\,R\,b)$$

where λ is a Lagrange multiplier which is a scalar. The points at which a local minimum , a local maximum or a saddle point may occur, can be evaluated by setting the respective derivatives of F (b, λ) with respect to b and λ equal to zero;

$$\frac{\partial F}{\partial b} = -2X'y + 2(X'X)b + 2\lambda\,Rb = 0$$

$$\frac{\partial F}{\partial b} = -C + b'Rb = 0$$

Hence the required values of b and λ are given by the quadratic equations,

$$\left.\begin{aligned} (X'X + \lambda R)b &= X'y \\ b'Rb &= C \end{aligned}\right\} \qquad (3.1)$$

or equivalently by

$$\text{with} \quad \left.\begin{aligned} (I_k + \lambda\,(X'X)^{-1}R)b &= \hat{\beta} \\ \lambda &= 1/C\,(b'X'Xb - y'Xb) \end{aligned}\right\} \qquad (3.2)$$

where $\hat{\beta}$ is the unrestricted least squares estimator of β .

These equations cannot be solved explicitly for b, however, they can be solved numerically for a given problem, under certain conditions on R. To illustrate this fact, a numerical example will be discussed at the end of this section.

In order to choose the value of b which indeed minimizes the Lagrangian function $F(b, \lambda)$ the following second order condition is used.

In order to minimize the sum of squares of residuals subject to the quadratic constraints we must choose the values of λ for which the matrix

$$\frac{\partial^2 F}{\partial b\, \partial b'} = 2\,(X'X + \lambda R) \text{ is positive definite.}$$

In case where there are more than one local minima of $F(b, \lambda)$ occur the one which gives the least value of F is to be chosen. The restricted least squares estimate of β is the value of b that satisfies the equation 3.1 with this particular value of λ.

To see the biasedness of b rewrite 3.1 as

$$b = \hat{\beta} - \lambda\,(X'X)^{-1} R\, b$$

Therefore $E(b) = \beta - E\,(\lambda (X'X)^{-1} R\, b)$

Therefore the direction of the bias is determined by the term $E\,(\lambda\,(X'X)^{-1} Rb)$.

These results can be generalized when there are moie than one quadratic relationship between the parameters. Suppose that the extraneous information consists of the quadratic constraints:

$$C_i = \beta' R_1 \beta, \quad i = 1, 2, \ldots\ldots\ldots l$$

with l is less than k

where C_i's are known scalars and R_i's are k x k symmetrical matrices such that

$$\begin{pmatrix} R_i & b \\ \vdots & \vdots \\ R_l & b \end{pmatrix} \quad \text{is of full rank.}$$

Again we wish to minimize $(y - Xb)'(y - Xb)$ subject to $C_i = \beta R_1 \beta$ $i = 1, 2, \ldots\ldots\ldots l$. Define

$$F\,(b, \lambda_i;\, i=1, 2, \ldots l) = \,|(y - Xb)\;\;(y - Xb) - \sum_{i=1}^{l} \lambda_i\,(C_i - b' R_1\, b)$$

Setting the derivatives of F with respect to b and λ_i's equal to zero gives for the points of inflection,

$$\frac{\partial F}{\partial b} = -2X'y + 2X'Xb' + 2\sum_{i=1}^{l} \lambda_i R_i b$$

$$\frac{\partial F}{\partial \lambda_i} = -C_i + b'R_i b = 0$$

Hence the restricted least squares estimator of β in this case is given by

$$(X'X + \sum_{i=1}^{l} \lambda_i^{\circ} R_i)\, b = X'y \qquad (3.3)$$

$$\text{and} \qquad b'R_i b = C_i, \quad i = 1, 2, \ldots\ldots l \qquad (3.4)$$

where λ_1°'s are chosen so that the matrix $(X'X + \Sigma \lambda_i^{\circ} R_i)$ is positive definite.

A numerical illustration;

As an illustration, consider a two variable relationship

$$y = \alpha + \beta X$$

and suppose we wish to incorporate the quadratic constraint[1] $2\alpha\beta = 5$, which can be written as, $C = \beta R. \beta$

$$\text{where } R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad C = 5.$$

Let us suppose that 20 sample observations yield

$$X'X = \begin{pmatrix} 20 & 10 \\ 10 & 3.2 \end{pmatrix} \text{ and } X'y = \begin{pmatrix} 20 \\ 8 \end{pmatrix}$$

Therefore we require to compute $\hat{\alpha}$, $\hat{\beta}$ and λ such that

$$(X'X+\lambda R)\, b = \begin{pmatrix} 20 & 10+\lambda \\ 10+\lambda & 3.2 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} 20 \\ 8 \end{pmatrix}$$

$$2\hat{\alpha}\hat{\beta} = 5$$

$$20 + \hat{\alpha}(10+\lambda)\hat{\beta} = 20$$

$$(10+\lambda)\hat{\alpha} + 3.2\hat{\beta} = 8$$

1. This constraint may have arisen in the estimation of the model $2\alpha(Y - \alpha) = 5X$

Then we may obtain $(\hat{\alpha}^2 + \hat{\alpha} + 1)(\hat{\alpha} - 1)^2 = 0$,

whose only real solution is $\hat{\alpha} = 1$.

Thus the corresponding values of $\hat{\beta}$ and λ which satisfy the Constraint equations are,

$$\hat{\beta} = \frac{5}{2}, \lambda = -10$$

Now observing that, this particular value of λ makes the matrix,

$$(X'X = \lambda\ R) = \begin{pmatrix} 20 & 0 \\ 0 & 3.2 \end{pmatrix} \text{ positive definite,}$$

we conclude that

$$\left.\begin{aligned} \hat{\alpha} &= 1 \\ \hat{\beta} &= 5/2 \end{aligned}\right\}$$ minimize the sum of the squared residuals subject to the constraint $(\hat{\alpha}\ \hat{\beta})\ R\ (\hat{\alpha}\ \hat{\beta})' = C$.

Thus the restricted least square estimators of α and β are $\hat{\alpha} = 1$ and $\hat{\beta} = 5/2$ respectively.

4. Least Squares Under Mixed Linear and Quadratic Constraints

In this section we consider a more general situation where the parameters of the classical regression model are known to be related by second order equations such as the two equations $2\beta_1 + \beta_1^2 + \beta_1\beta_3 + \beta_3 = 0$ and $2\beta_1 = \beta_4^2$. We will establish a general formula to determine the restricted least squares estimates when the constraint on the parameters consists of only a single equation of this type. The generalization to the case where there are more than one equation of that type is straightforward but will not be undertaken here.

The mixed linear and quadratic constraint under consideration may be expressed in the form,

$$C = \beta'\ R\ \beta + r'\beta$$

where C is a known scalar, R is a known symmetric matrix of order k and r is a column vector of k coefficients.

To incorporate this information the method of restricted least squares is proposed. Therefore we require the estimated coefficient vector b, that minimizes $(y-Xb)'$ $(y-Xb)$ subject to $C = b'Rb + r'b$. Thus the appropiate Lagrangian function becomes,

$$F(b, \lambda) = (y - Xb)'(y - Xb)\lambda - (C - b'Rb - r'b)$$

where λ is a Lagrange multiplier. The stationary points of this function are found by setting the derivatives of F with respect to b and λ equal to zero.

$$\frac{\partial F}{\partial b} = -2X'y + (2X'X)b + \lambda(2Rb + r) = 0$$

$$\frac{\partial F}{\partial \lambda} = -(C - b'Rb - r'b) = 0$$

The second order condition for a minimum is satisfied if

$$\frac{\partial^2 F}{\partial b \partial b^1} = 2\,(X'X - \lambda R) \text{ is positive definite.}$$

Hence the restricted least squares estimator of the parameter vector β is given by the equations,

$$(X'X + \lambda^* R)\, b = X'y - \frac{\lambda^*}{2} r$$

$$C - b'R\,b + \lambda^* r'b$$

where λ^* is chosen so that the matrix $(X'X + \lambda^* R)$ is positive definite. In case where there are more than one local minima the one which gives the least value of F (b,λ) should be chosen.

A numerical Illustration

Suppose we want to find the restricted least squares estimators of α and β appearing in the two variable linear regression model,

$y = \alpha + \beta x + u$, given the extraneous information that $\alpha + \beta^2 = 3$. Let us also suppose that 20 sample observations yield,

$$X'X = \begin{pmatrix} 20 & 40\frac{2}{3} \\ 40\frac{2}{3} & 163 \end{pmatrix}, \quad X'y = \begin{pmatrix} 68\frac{1}{2} \\ 284 \end{pmatrix}$$

using the usual notation we have,

$$R = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \text{ and } r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

therefore the above two implies that,

$$\begin{pmatrix} 20 & 40\frac{2}{3} \\ 40\frac{2}{3} & 63 + \lambda \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} 68\frac{1}{2} & - \frac{\lambda}{2} \\ 284 & \end{pmatrix}$$

and $\hat{\alpha} + \hat{\beta}^2 = 3$

Eliminating λ and $\hat{\alpha}$ from these equations we have

$$(5\hat{\beta} - 9) \quad (4\hat{\beta}^2 - 5\hat{\beta} + 9) = 0.$$

The only real solution of this equation is $\hat{\beta} = 1.8$

The corresponding values of $\hat{\alpha}$ and $\hat{\lambda}$ are $\hat{\alpha} = -0.24$ and $\lambda = 0.2$

Now

$$(X'X + \lambda R) = \begin{pmatrix} 20 & 40\frac{2}{3} \\ 40\frac{2}{3} & 163 + \lambda \end{pmatrix}$$

is positive definite when $\lambda = 0.2$

Therefore we conclude that $\hat{\beta} = 1.8$, $\hat{\alpha} = -0.24$, $\lambda = 0.1$ minimizes the Lagrangian function F (b, λ). Hence the restricted least squares estimators of α and β are $\hat{\alpha} = -0.24$, $\hat{\beta} = 1.8$ respectively.

5 Mixed Estimation

When unbiased estimates of some of the parameters of a linear regression is available from outside of the sample, a technique to use this information so as to obtain more efficient estimates was suggested by Durbin[1] in 1953. In fact this approach can be extended further to tackle various problems concerning extraneous information. This procedure of "mixed estimation" was developed by Theil and Goldberger[4] in 1961 and it was extended by Theil[3] in 1963. One may make prior expectations on the results of a regression on the basis of extraneous information. The expectations that we derived from theoretical considerations are called prior information, the positivity of the marginal productivities in a production function being an example. The kind of expectations which were derived from previous statistical work, for example some unbiased estimates of certain parameters, are called statistical prior information. The estimation process that combines the prior information together with the sample data is called mixed estimation.

The prior knowledge is formulated in terms of prior estimates of the parameters which are assumed to be unbiased, namely $r = R\beta + v$, where v is the error of the prior information with $E(v) = 0$ and the g-element vector r estimates $R\beta$, R being a known gxk matrix. Particularly if r is an unbiased estimate of β, a vector formed by g coefficients of β, we would set $R = (I_g\, 0)$ because $r = \beta_1 + v$, where 0 is $g \times (k\text{-}g)$ matrix containing only zeros. Let us assume that $E(vv^1) = \psi$ is known and that $E(v\epsilon^1) = 0$. We combine sample observations and the prior information to write

$$\begin{pmatrix} y \\ r \end{pmatrix} = \begin{pmatrix} X \\ R \end{pmatrix} \beta + \begin{pmatrix} \epsilon \\ v \end{pmatrix}$$

where

$$E\begin{pmatrix} \epsilon \\ v \end{pmatrix} = 0 \text{ and } E\begin{pmatrix} \epsilon \\ v \end{pmatrix} (\epsilon' \ v') = \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \psi \end{pmatrix}$$

It can be shown that the application of Generalized Least Squares to the above linear regression model yields

$$b = \left(\frac{1}{\sigma^2} X'X + R^1 \psi^{-1} R\right)^{-1} \left(\frac{1}{\sigma^2} X'y + R'\psi^{-1} r\right) \text{ and that}$$

$$\text{Var}(b) = \left(\frac{1}{\sigma} X^1X + R^1\psi^{-1}R\right)^{-1}$$

This estimator b is a best linear unbiased estimator of β, where "best" refers to the sample and prior information taken together.

6. Mixed Estimation Under Linear Constraints

Now suppose that two kinds of extraneous information are available to the statistician, some exact linear relationships between the coefficients are known and on the other hand a prior information, for instance some unbiased estimates of some of the coefficients, is available. The statistician wishes to incorporate his knowledge on the parameters in order to improve the estimate of the coefficient vector. Let the known linear restrictions on the coefficients be $S = L\beta$, where S is a known vector of h elements and L is a $h \times k$ known matrix of rank $h < k$. Assume that the prior knowledge can be formulated as,

$$r = R\beta + v$$

with $E(v) = O$, where v is the error of the prior information, r is a known gxl vector and R is a known g x k matrix. Let us assume that $E(v\epsilon') = O$ and that $E(vv') = \psi$ is known. In order to apply generalized least squares we write the prior and sample information together as,

$$\begin{pmatrix} y \\ r \end{pmatrix} = \begin{pmatrix} \lambda \\ R \end{pmatrix} \beta + \begin{pmatrix} \epsilon \\ v \end{pmatrix}$$

which can be rewritten as,

$y_1 = X_1 \beta + u$, where y_1 is a (n + g) XI vector X_1 is a (n + g) x k matrix and $u = \begin{pmatrix} e \\ v \end{pmatrix}$. We have for the covariance matrix of the extended "disturbance",

$$E(uu') = \begin{pmatrix} \sigma^2 I_n & 0 \\ 0 & \psi \end{pmatrix} = \Omega$$

The symmetric positive definite matrix Ω can be expressed in the form,

$$\Omega = PP', \text{ where P is non-singular.}$$

Now premultiply the model,

$$y_1' = X_1' \beta + u \text{ by } P^{-1} \text{ to give}$$
$$y^* = X^* \beta + u^* \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (6.1)$$

where $y^* = P^{-1}y$ $X^* = P^{-1} X_1$ and $u^* = P^{-1}u$. It is easily seen that,

$$E(u^*) = 0, \text{ and that}$$
$$E(u^* u^{*\prime}) = \sigma^2 I \text{ because}$$

$P^{-1} \Omega P^{-1\prime} = I_{(nx=g)}$ so that (6.1) satisfies all the assumptions of a classical least squares model. Now it follows from section (2) that the restricted least squares estimator of β under $S = L\beta$ is,

$$b = \hat{\beta} + (X^{*\prime} X^*)^{-1} L [L (X^{*\prime} X^*)^{-1} L']^{-1} (S - L\hat{\beta}) \text{ and that}$$

$$V(b) = V - VL' (LVL')^{-1} LV, \text{ where } V = \sigma^2 (X^{*\prime} X^*)^{-1}$$

and $\hat{\beta} = (X^{*\prime} X^*)^{-1} X^* y$. Moreover we know that b is the BLUE of β in the sense that its elements have the minimum variance within the class of all unbiased estimators which are linear functions of y and S, where "best" refers to the sample and prior information taken together. The b and var (b) can be expressed in terms of X as follows,

Since $X^* = p^{-1} X_1$ and $y^* = P^{-1} y_1$,
we have, $X^{*\prime} X^* = X_1' \Omega^{-1} X_1$ and $X^{*1} y^* = X_1' \Omega^{-1} y_1$

and in turn we have,

$$X^{*\prime} X^* = \left(\frac{1}{\sigma^2} X' X + R' \psi^{-1} R\right)$$

$$\text{and } X^{*\prime} y = \left(\frac{1}{\sigma^2} X' y + R' \psi^{-1} r\right).$$

Hence $\hat{\beta} = \left(\frac{1}{\sigma^2} X' X + R' \psi R\right)^{-1} \left(\frac{1}{\sigma^2} X' y + R' \psi^{-1} r\right)$

and

$$b = \hat{\beta} + \left(\frac{1}{\sigma^2} X' X + R' \psi^{-1} R\right)^{-1} L'\left[L\left(\frac{1}{\sigma^2} X' X + R' \psi^{-1} R\right)^{-1} L'\right]^{-1}$$

$(S - L \hat{B})$.

The variance of b also reduces to,

$$V(b) = V(\hat{\beta}) - V(\hat{\beta}) L' (L V(\hat{\beta}) L')^{-1} L V(\hat{\beta})$$

where

$$V(\hat{\beta}) = \sigma^2\left(\frac{1}{\sigma^2} X' X + R' \psi^{-1} R\right)^{-1}.$$

References

1. DURBIN, J. (1953). *A note on Regression when there is Extraneous Information about one of the Coefficients.* Journal of the American Statistical Association **48**, pp 799-808.

2. THEIL, H. (1961), *"Economic Forecasts and Policy" second edition Amsterdam* : North Holland Publishing Co.

3. THEIL, H. (1963). *On the use of Incomplete Prior Information in Regression Analysis.* Journal of the American Statistical Association **58**, pp 401 - 414.

4. THEIL, H. & GOLDBERGER. A. S. (1961). *On pure and mixed statistical Estimation in Economics.* Int. Econ. Rev. **2**, pp 65 - 78.