RESTRICTED LEAST-SQUARES REGRESSION

Sometimes, we find that there is a set of a priori restrictions on the elements of the vector $\beta$ of the regression coefficients which can be taken into account in the process of estimation. A set of $j$ linear restrictions on the vector $\beta$ can be written as $R\beta = r$, where $r$ is a $j \times k$ matrix of linearly independent rows, such that $\text{Rank}(R) = j$, and $r$ is a vector of $j$ elements.

To combine this a priori information with the sample information, we adopt the criterion of minimising the sum of squares $(y - X\beta)'(y - X\beta)$ subject to the condition that $R\beta = r$. This leads to the Lagrangean function

$$L = (y - X\beta)'(y - X\beta) + 2\lambda'(R\beta - r)$$

On differentiating $L$ with respect to $\beta$ and setting the result to zero, we get the following first-order condition $\partial L/\partial \beta = 0$:

$$2\beta'X'X - 2y'X + 2\lambda'R = 0,$$

whence, after transposing the expression, eliminating the factor 2 and rearranging, we have

$$X'X\beta + R'\lambda = X'y.$$

When these equations are compounded with the equations of the restrictions, which are supplied by the condition $\partial L/\partial \lambda = 0$, we get the following system:

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \lambda \end{bmatrix} = \begin{bmatrix} X'y \\ r \end{bmatrix}.$$

For the system to have a unique solution, that is to say, for the existence of an estimate of $\beta$, it is not necessary that the matrix $X'X$ should be invertible—it is enough that the condition

$$\text{Rank} \begin{bmatrix} X \\ R \end{bmatrix} = k$$

should hold, which means that the matrix should have full column rank. The nature of this condition can be understood by considering the possibility of estimating $\beta$ by applying ordinary least-squares regression to the equation

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} \varepsilon \\ 0 \end{bmatrix},$$

which puts the equations of the observations and the equations of the restrictions on an equal footing. It is clear that an estimator exits on the condition that $(X'X + R'R)^{-1}$ exists, for which the satisfaction of the rank condition is necessary and sufficient.
RESTRICTED LEAST-SQUARES REGRESSION

Let us simplify matters by assuming that \((X'X)^{-1}\) does exist. Then equation (3) gives an expression for \(\beta\) in the form of

\[
\beta^* = (X'X)^{-1}X'y - (X'X)^{-1}R'\lambda
\]

(7)

where \(\hat{\beta}\) is the unrestricted ordinary least-squares estimator. Since \(R\beta^* = r\), premultiplying the equation by \(R\) gives

\[
r = R\hat{\beta} - R(X'X)^{-1}R'\lambda,
\]

(8)

from which

\[
\lambda = \{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r).
\]

(9)

On substituting this expression back into equation (7), we get

\[
\beta^* = \hat{\beta} - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - r).
\]

(10)

This formula is more intelligible than it might appear to be at first, for it is simply an instance of the prediction-error algorithm whereby the estimate of \(\beta\) is updated in the light of the information provided by the restrictions. The error, in this instance, is the divergence between \(R\hat{\beta}\) and \(E(R\hat{\beta}) = r\).

The sampling properties of the restricted least-squares estimator are easily established. Given that \(E(\hat{\beta} - \beta) = 0\), which is to say that \(\hat{\beta}\) is an unbiased estimator, it follows that \(E(\beta^* - \beta) = 0\), so that \(\beta^*\) is also unbiased.

Next consider the expression

\[
\beta^* - \beta = [I - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R](\hat{\beta} - \beta)
\]

(11)

\[
= (I - P_R)(\hat{\beta} - \beta),
\]

where

\[
P_R = (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R.
\]

(12)

The expression comes from taking \(\beta\) from both sides of (10) and from recognising that \(R\hat{\beta} - r = R(\hat{\beta} - \beta)\). We may observe that \(P_R\) is an idempotent matrix which is subject to the conditions that

\[
P_R = P_R^2, \quad P_R(I - P_R) = 0 \quad \text{and} \quad P'_RX'X(I - P_R) = 0.
\]

(13)

From equation (11), we deduce that

\[
D(\beta^*) = (I - P_R)E((\hat{\beta} - \beta)(\hat{\beta} - \beta)'(I - P_R)
\]

(14)

\[
= \sigma^2(I - P_R)(X'X)^{-1}(I - P_R)
\]

\[
= \sigma^2[(X'X)^{-1} - (X'X)^{-1}R'\{R(X'X)^{-1}R'\}^{-1}R(X'X)^{-1}].
\]