

2.8 RESTRICTED AND UNRESTRICTED REGRESSION

In previous sections we made use of the LS and ML principles to derive estimators of the unknown parameters of the MLRM. In using these principles, we assumed that our information level was only the sample information, so it was considered there was no a priori information on the parameters of the model. However, in some situations it is possible to have some non-sample information (a priori information on the parameters), which can be of several kinds. Now we focus only on exact a priori information about β coefficients (useful references for this topic are [Fomby, Carter, and Johnson \(1984\)](#) and [Judge, Griffiths, Carter, Lutkepohl and Lee \(1985\)](#)).

In general, this previous information on the coefficients can be expressed as follows:

$$R\beta = r \quad (2.153)$$

where R and r are the matrix and the vector which was defined to establish the test given in (2.134). Now (2.153) can be thought of as the way of expressing the a priori information about the elements of the β vector.

In this section, our objective consists of estimating the parameters of the MLRM by considering the a priori information. Basically, there are two equivalent ways of carrying out such an estimation. One of them consists of incorporating the a priori information into the specified model, in such a way that a transformed model is obtained whose unknown parameters are estimated by OLS or ML. The other way of operating consists of applying either what we call the restricted least squares (RLS) method, or what we call the restricted maximum likelihood (RML) method.

2.8.1 Restricted Least Squares and Restricted Maximum Likelihood Estimators

Given the MLRM

$$y = X\beta + u$$

and the a priori information about β expressed as $R\beta = r$, we try to find the vector $\hat{\beta}_R$ which minimizes the squared sum of residuals (if we use the LS method) or maximizes the likelihood function (in the case of the ML method), subject to $R\hat{\beta}_R = r$. Then, the estimator which we obtain by combining all the information is called *Restricted Least Squares* or *Restricted Maximum Likelihood*, respectively.

The conditioned optimization problem can be solved through the classical Lagrangian procedures. If

we first consider the LS method, the corresponding Lagrange function is:

$$\begin{aligned}\mathfrak{S} &= (y - X\hat{\beta}_R)^\top (y - X\hat{\beta}_R) - 2\lambda^\top (R\hat{\beta}_R - r) = \\ &y^\top y - 2\hat{\beta}_R^\top X^\top y + \hat{\beta}_R^\top X^\top X \hat{\beta}_R - 2\lambda^\top (R\hat{\beta}_R - r)\end{aligned}\quad (2.154)$$

where λ is the $q \times 1$ vector of Lagrange multipliers. The 2 in the last term appears to make the derivation easier and does not affect the outcome.

To determine the optimum values, we set the partial derivatives of \mathfrak{S} with respect to β and λ equal to zero:

$$\frac{\partial \mathfrak{S}}{\partial \hat{\beta}_R} = -2X^\top y + 2X^\top X \hat{\beta}_R - 2R^\top \lambda = 0 \quad (2.155)$$

$$\frac{\partial \mathfrak{S}}{\partial \lambda} = -2(R\hat{\beta}_R - r) = 0 \quad (2.156)$$

We substitute $\hat{\beta}_R$ by $\hat{\beta}_R$ in order to obtain the value of $\hat{\beta}_R$ which satisfies the first-order conditions.

Then, from (2.155) we have:

$$X^\top X \hat{\beta}_R = X^\top y + R^\top \hat{\lambda} \quad (2.157)$$

In premultiplying the last expression by $(X^\top X)^{-1}$ we get:

$$\begin{aligned}(X^\top X)^{-1} X^\top X \hat{\beta}_R &= (X^\top X)^{-1} (X^\top y + R^\top \hat{\lambda}) \\ \Rightarrow \hat{\beta}_R &= \hat{\beta} + (X^\top X)^{-1} R^\top \hat{\lambda}\end{aligned}\quad (2.158)$$

where $\hat{\beta}$ is the unrestricted least squares estimator which was obtained in (2.25).

Expression (2.158) is premultiplied by R and we get:

$$R\hat{\beta}_R = R\hat{\beta} + R(X^\top X)^{-1}R^\top \hat{\lambda} \quad (2.159)$$

Since $(X^\top X)^{-1}$ is a positive definite matrix, $R(X^\top X)^{-1}R^\top$ is also positive definite, and moreover, its rank is $q \leq k$ and it is nonsingular. Then, from (2.159) we may obtain:

$$\hat{\lambda} = [R(X^\top X)^{-1}R^\top]^{-1}(R\hat{\beta}_R - R\hat{\beta})$$

or

$$\hat{\lambda} = [R(X^\top X)^{-1}R^\top]^{-1}(r - R\hat{\beta}) \quad (2.160)$$

because from (2.156), the restricted minimization problem must satisfy the side condition $R\hat{\beta}_R = r$.

Using the value (2.160) for the vector λ , we get from (2.158) the estimator:

$$\hat{\beta}_R = \hat{\beta} + (X^\top X)^{-1}R^\top[R(X^\top X)^{-1}R^\top]^{-1}(r - R\hat{\beta}) \quad (2.161)$$

which is denoted as the restricted least squares (RLS) estimator.

Given that $(X^\top X)^{-1}R^\top[R(X^\top X)^{-1}R^\top]^{-1}$ is a matrix of constant elements, from (2.161) we can see that the difference between $\hat{\beta}_R$ and $\hat{\beta}$ is a linear function of the $(r - R\hat{\beta})$ vector.

Moreover, we deduce that this difference increases the further $\hat{\beta}$ (unrestricted LS) is from satisfying the restriction.

According to the RLS estimator, the residuals vector can be defined as:

$$\hat{u}_R = y - X\hat{\beta}_R \quad (2.162)$$

and, analogously to the procedure followed to obtain $\hat{\sigma}^2$, the RLS estimator of σ^2 is given by:

$$\hat{\sigma}_R^2 = \frac{\hat{u}_R^\top \hat{u}_R}{n - k + q} \quad (2.163)$$

which is an unbiased estimator of σ^2 , given that $E(\hat{u}_R^\top \hat{u}_R) = \sigma^2(n - (k - q))$.

Having obtained the expressions of the RLS estimators of the parameters in the MLRM, we now have the required information in order to prove the equivalence between (2.139) and (2.140), established in the previous section. In order to show such equivalence, we begin by adding and subtracting $X\hat{\beta}$ to and from (2.162):

$$\hat{u}_R = y - X\hat{\beta}_R + X\hat{\beta} - X\hat{\beta} = (y - X\hat{\beta}) - X(\hat{\beta}_R - \hat{\beta}) = \hat{u} - X(\hat{\beta}_R - \hat{\beta}) \quad (2.164)$$

and then, given that $X^\top \hat{u} = \hat{u}^\top X = 0$ (an algebraic property of the LS method, described in the estimation section), we have:

$$\hat{u}_R^\top \hat{u}_R = [\hat{u} - X(\hat{\beta}_R - \hat{\beta})]^\top [\hat{u} - X(\hat{\beta}_R - \hat{\beta})] = \hat{u}^\top \hat{u} + (\hat{\beta}_R - \hat{\beta})^\top X^\top X (\hat{\beta}_R - \hat{\beta}) \quad (2.165)$$

From (2.165), we can write:

$$\hat{u}_R^\top \hat{u}_R - \hat{u}^\top \hat{u} = (\hat{\beta}_R - \hat{\beta})^\top X^\top X (\hat{\beta}_R - \hat{\beta}) \quad (2.166)$$

and if we substitute $(\hat{\beta}_R - \hat{\beta})$ according to (2.161), we have

$$\hat{u}_R^\top \hat{u}_R - \hat{u}^\top \hat{u} = (r - R\hat{\beta})^\top [R(X^\top X)^{-1}R^\top]^{-1} (r - R\hat{\beta}) \quad (2.167)$$

This last expression allows us to conclude that (2.139) and (2.140) are equivalent. Additionally, from (2.167) it is satisfied that $\hat{u}_R^\top \hat{u}_R > \hat{u}^\top \hat{u}$, given that $[R(X^\top X)^{-1}R^\top]^{-1}$ is a positive definite matrix.

In order to derive now the RML estimators, the Lagrange function according to the ML principle is written as:

$$\mathfrak{S} = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma_R^2 - \frac{(y - X\beta_R)^\top (y - X\beta_R)}{2\sigma_R^2} + 2\lambda^\top (R\beta_R - r) \quad (2.168)$$

and the first-order conditions are:

$$\frac{\partial \mathfrak{S}}{\partial \beta_R} = -\frac{1}{2\sigma_R^2} (-2X^\top y + 2X^\top X\beta_R) + 2R^\top \lambda = 0 \quad (2.169)$$

$$\frac{\partial \mathfrak{S}}{\partial \sigma_R^2} = -\frac{n}{2\sigma_R^2} + \frac{(y - X\beta_R)^\top (y - X\beta_R)}{2\sigma_R^4} = 0 \quad (2.170)$$

$$\frac{\partial \mathfrak{S}}{\partial \lambda} = -2(R\beta_R - r) = 0 \quad (2.171)$$

From (2.169)-(2.171), and putting \sim to the parameters, in a similar way to the RLS procedure, we deduce:

$$\tilde{\beta}_R = \tilde{\beta} + (X^\top X)^{-1} R^\top [R(X^\top X)^{-1} R^\top]^{-1} (r - R\tilde{\beta}) \quad (2.172)$$

$$\tilde{\sigma}_R^2 = \frac{(y - X\tilde{\beta}_R)^\top (y - X\tilde{\beta}_R)}{n} = \frac{\hat{u}_R^\top \hat{u}_R}{n} \quad (2.173)$$

$$\tilde{\lambda} = \frac{[R(X^\top X)^{-1} R^\top]^{-1} (r - R\tilde{\beta})}{\tilde{\sigma}_R^2} \quad (2.174)$$

so we conclude that, in a MLRM which satisfies the classical assumptions, the RLS estimators of the coefficients are the same as the RML estimators. This allows us to write the equality given in (2.173).

2.8.2 Finite Sample Properties of the Restricted Estimator Vector

Given the equality between $\hat{\beta}_R$ and $\tilde{\beta}_R$, the following proofs are valid for both procedures.

Before deriving some properties, it is convenient to obtain the expectation vector and the variance-covariance matrix of the restricted estimator vector. Using (2.161), the expected value of $\hat{\beta}_R$ is :

$$\begin{aligned} E(\hat{\beta}_R) &= E(\hat{\beta}) + (X^\top X)^{-1} R^\top [R(X^\top X)^{-1} R^\top]^{-1} (r - RE(\hat{\beta})) = \\ &\beta + (X^\top X)^{-1} R^\top [R(X^\top X)^{-1} R^\top]^{-1} (r - R\beta) \end{aligned} \quad (2.175)$$

and the variance-covariance matrix:

$$V(\tilde{\beta}_R) = E[(\hat{\beta}_R - E\hat{\beta}_R)(\hat{\beta}_R - E\hat{\beta}_R)^\top]$$

If we consider expression (2.54) and it is substituted into (2.161), we can write:

$$\begin{aligned}\hat{\beta}_R - E\hat{\beta}_R &= (X^\top X)^{-1}X^\top u - (X^\top X)^{-1}R^\top[R(X^\top X)^{-1}R^\top]^{-1}R(X^\top X)^{-1}X^\top u \\ &= [I_k - (X^\top X)^{-1}R^\top[R(X^\top X)^{-1}R^\top]^{-1}R](X^\top X)^{-1}X^\top u \\ &= \Phi(X^\top X)^{-1}X^\top u\end{aligned}\tag{2.176}$$

The Φ matrix (which premultiplies to $(X^\top X)^{-1}X^\top u$ in (2.176), is a $k \times k$ idempotent matrix of constant elements. From this last expression we obtain:

$$\begin{aligned}V(\hat{\beta}_R) &= E[\Phi(X^\top X)^{-1}X^\top uu^\top X(X^\top X)^{-1}\Phi^\top] \\ &= \sigma^2\Phi(X^\top X)^{-1}\Phi^\top = \sigma^2\Phi(X^\top X)^{-1}\end{aligned}\tag{2.177}$$

The last equality of (2.177) is written according to the proof presented in [Judge, Carter, Griffiths, Lutkepohl and Lee \(1988\)](#)

From (2.177), it is possible to deduce the relationship between $V(\hat{\beta}_R)$ and $V(\hat{\beta})$, by replacing Φ by its expression, and thus:

$$\begin{aligned}V(\hat{\beta}_R) &= \sigma^2(X^\top X)^{-1} - \sigma^2(X^\top X)^{-1}R^\top[R(X^\top X)^{-1}R^\top]^{-1}R(X^\top X)^{-1} \Rightarrow \\ V(\hat{\beta}_R) - V(\hat{\beta}) &= \sigma^2(X^\top X)^{-1}R^\top[R(X^\top X)^{-1}R^\top]^{-1}R(X^\top X)^{-1} = \sigma^2 C\end{aligned}$$

with C being a positive semidefinite matrix, as [Fomby, Carter, and Johnson \(1984\)](#) show.

Consequently, the diagonal elements of $V(\hat{\beta}_R)$ (variances of each $\hat{\beta}_{Rj}$) are equal to or less than the corresponding elements of $V(\hat{\beta})$ (variances of each $\hat{\beta}_j$). This means that, if the a priori information is correct (as we will show later, in this case $E(\hat{\beta}_R) = \beta$), the estimator vector $\hat{\beta}_R$ is more efficient than the OLS estimator.

On the basis of these previous results, we can establish the finite properties of the RLS estimator.

- Linearity. If we substitute expression (2.54) of $\hat{\beta}$ into (2.161), and then we group terms, we

obtain:

$$\hat{\beta}_R = (X^\top X)^{-1} R^\top [R(X^\top X)^{-1} R^\top]^{-1} r + \left[(X^\top X)^{-1} X^\top - (X^\top X)^{-1} R^\top [R(X^\top X)^{-1} R^\top]^{-1} R(X^\top X)^{-1} X^\top \right] y \quad (2.178)$$

Given that the first term of the right-hand side of (2.178) is a vector of constant elements, and the second term is the product of a matrix of constant elements multiplied by the vector y , it follows that $\hat{\beta}_R$ satisfies the linearity property.

- Unbiasedness. According to (2.175), $\hat{\beta}_R$ is unbiased only if $r - R\beta = 0$, that is to say, if the a priori information is true.
- BLUE. With correct a priori information, the estimator vector $\hat{\beta}_R$ is the best linear unbiased vector within the class of unbiased estimators that are linear functions of the endogenous variable and that also satisfy the a priori information (2.153). This property, which we will not prove here, is based on the Gauss-Markov Theorem.
- Efficiency. As [Rothemberg \(1973\)](#) shows, when the a priori information is correct, the estimator vector $\hat{\beta}_R$ satisfies the Cramer-Rao inequality, and consequently, it is efficient.

From (2.177) it can be deduced that the expression of $V(\hat{\beta}_R)$ does not change if the a priori information is correct or non correct, which means that $V(\hat{\beta}_R) \leq V(\hat{\beta})$ is maintained whatever the situation.

Finally, given the linearity of $\hat{\beta}_R$ with respect to y , and this vector being normally distributed, if the a priori information is correct, we have:

$$\hat{\beta}_R \sim N(\beta, \sigma^2 \Phi(X^\top X)^{-1}) \quad (2.179)$$

2.8.3 Example

We now consider that we have the following a priori information on the coefficients: $\beta_2 = 1$, in such a way that we calculate the restricted estimators of the coefficients. Jointly with these estimators the quantlet `xEGm1rm06.xpl` computes the F statistic as a function of the restricted and unrestricted squared sum of the residuals which allows us to test $H_0 : \beta_2 = 1$.



Note that the RML estimator satisfies the formulated restriction, and the value of the F statistic is the same as the one obtained in Section [2.7.4](#)
