Restricted Least Squares Estimator in Under-Identified Models.

Giri Parameswaran

Princeton University

April 5, 2011

Abstract

The need often arises to estimate linear models whose regressors do not satisfy the standard rank conditions - for example in models with dummy variables. The problem of singularity is not that the model cannot be estimated, but rather that it cannot be done so uniquely. A common practice is to omit sufficiently many regressors as is required to alleviate the rank deficiency. However, under such an approach, the interpretation of the estimates of the remaining parameters is often found wanting. An alternative approach is to impose appropriate restrictions upon the parameter space that help identify a unique solution (from the many). This note provides conditions on the set of such parametric restrictions that are sufficient for a unique solution to exist, and provides an explicit characterisation of the restricted least square estimators.
1 Introduction

The need often arises to estimate linear models whose regressors do not satisfy the standard rank conditions. Common examples include linear models with dummy variables (the ‘dummy variable trap’), models with interactive terms and under-identified systems in GMM estimation. However, there has been little attention paid to the problem of estimating such models. Greene and Seaks (1991) elaborate on this point rather forcefully. The usual procedure when confronted with singularity is to omit some of the correlated variables. Whilst this allows for estimation using standard techniques, the interpretation of the parameters in such a model are often found to be wanting. Suits (1984), for example, provides an elaborate discussion of the problems of interpreting dummy coefficients, when the base-category dummy variable is omitted.

Several authors (Sweeney & Ulveling 1972), (Suits 1984), (Kennedy 1986)) have proposed alternative methods for estimating models with dummy variables, that generate more plausible or useful estimators. For example, Kennedy (1986) suggests including all dummy variables in the model (including the base category) and imposing the restriction that the weighted sum of the dummy coefficients equal zero. This gives the dummy coefficients the more natural interpretation of the average effect of being in a particular group (relative to the mean) - rather than relative to some arbitrarily chosen base category.

The problem of singularity is not that the model cannot be estimated, but rather that it cannot be estimated uniquely. The purpose of parametric restrictions in this context is to select a unique estimator. Indeed, the ‘usual’ procedure of ‘omitting’ variables from the model merely disguises a covert set of restrictions in which the parameters corresponding to the omitted variables are constrained to equal zero. In this sense, omitting ‘redundant’ variables is merely a special case of restricted least square estimation. This observation, coupled with the interpretive challenges that the ‘omission’-procedure creates, motivates the need for a generalised theory of restricted least square estimation in models with singularity.

Greene and Seaks (1991) extend the analysis of the above authors to a more general class of models (than merely those with dummy variables). They posit that the restricted least square estimators are the solution to a particular linear system - however they neither find an expression for the solution, nor show that the system can be solved uniquely. This note extends the analysis in Greene and Seaks (1991) by addressing
each of these problems, and then outlines the properties of the restricted least square estimators in models with singularity.

2 The Model

Consider a linear model, \( y = X\beta + u \) where the \( n \times k \) matrix of regressors \( X \) is singular. For concreteness, let \( r(X) = k - h \), so that \( h \) regressors are redundant - i.e. they are linearly dependent upon the remaining \( k - h \) regressors. The ordinary least square (OLS) estimators are the solutions to the normal equations
\[
(X^TX)^T\hat{\beta} = X^Ty.
\]
Given the rank properties, this system will not have a unique solution. A unique estimator can be identified by imposing appropriate restrictions on the parameter space. In this note, I restrict attention to the class of linear parametric constraints. Let \( R\beta = r \) be a system of \( h \) linear constraints, where \( R \) is an \( h \times k \) matrix of the form
\[
R = \begin{bmatrix}
s_1 & \cdots & s_h
\end{bmatrix}^T,
\]
with full row rank.

The restricted least square (RLS) estimators minimise the residual sum of squares, subject to the parameter constraints. Taking first order conditions gives the restricted normal equations:
\[
\begin{bmatrix}
X^TX & R^T \\
R & 0
\end{bmatrix}
\begin{bmatrix}
\hat{\beta} \\
\hat{\lambda}
\end{bmatrix}
=
\begin{bmatrix}
X^Ty \\
r
\end{bmatrix},
\]
where \( \lambda \) is the \( h \times 1 \) vector of multipliers. If \( X \) has full column rank and \( R \) has full row rank, then this system has a well known unique solution. (See (Greene 2003)). However, if \( X \) is singular, then the restricted normal matrix (which I denote by \( N \)) need not be invertible. In the following sections, I find conditions under which this matrix is invertible and so a unique solution exists. I then characterise this solution.

3 Analysis

For any matrix \( A \in \mathbb{R}^{m \times n} \), let \( \mathcal{M}(A) \) denote the range-space of \( A \), and let \( \mathcal{N}(A) \) denote the null-space of \( A \).

\[
\mathcal{M}(A) = \{ y \in \mathbb{R}^m \mid y = Ax, \quad x \in \mathbb{R}^n \}
\]
\[
\mathcal{N}(A) = \{ x \in \mathbb{R}^n \mid Ax = 0 \}
\]
The following proposition provides sufficient conditions on the set of parametric restrictions to guarantee that the restricted normal matrix is invertible:

**Proposition 1** Let $V$ be an $n \times h$ matrix whose columns form a basis for $N(X)$. If $RV$ is non-singular, then the restricted normal matrix $N$ is invertible. Furthermore, $W = (X^TX + R^TR)^{-1}$ exists and is a generalised inverse of $X^TX$.\(^1\)

**Proof.** See Appendix. ■

The requirement that $RV$ is non-singular implies the set of estimators that satisfy the normal equations are - in a sense - orthogonal to the set of estimators that satisfy the parametric restrictions. Hence, even thought there are infinitely many candidate estimates $\hat{\beta}$ that satisfy the normal equations, only one will also satisfy the parameter restriction. Having shown that the restricted normal matrix $N$ is invertible, it remains to characterise the inverse of this matrix.

**Proposition 2** The inverse of the restricted normal matrix is given by:

\[
\begin{bmatrix}
X^TX & R^T \\
R & 0
\end{bmatrix}^{-1} = \begin{bmatrix}
[I - V [RV]^{-1} R] [X^TX + R^TR]^{-1} V [RV]^{-1} \\
[VTR^T]^{-1} V^T & 0
\end{bmatrix} = \begin{bmatrix}
[X^TX + R^TR]^{-1} [I - R^T [VTR^T]^{-1} V] & V [RV]^{-1} \\
[VTR^T]^{-1} V^T & 0
\end{bmatrix}
\]

where the top-left blocks are equivalent and symmetric.

**Proof.** By direct verification, using Lemma 5 (see Appendix) to simplify expressions. Lemma 5 also implies that the two expressions in the top-left blocks are identical, and moreover, that these are both equivalent to $W - WR^TRW$, which is clearly symmetric. ■

\(^1\)This proposition draws upon insights from Rao (1973)
4 Properties of RLS Estimators

In this section, I compute the RLS estimator, and compare the properties of the RLS and OLS estimators.

The restricted least square estimators are given by:

$$
\begin{pmatrix}
\hat{\beta}_1 \\
\hat{\lambda}
\end{pmatrix}
= 
\begin{bmatrix}
X^T & R^T \\
R & 0
\end{bmatrix}^{-1} 
\begin{pmatrix}
X^Ty \\
r
\end{pmatrix}
= 
\begin{pmatrix}
[X^TX + R^TR]^{-1} X^Ty + V[RV]^{-1}r \\
0
\end{pmatrix}
$$

The fitted values are given by \( \hat{y} = X\hat{\beta} = P^*y \), where:

$$
P^* = X \left[ X^TX + R^TR \right]^{-1} X^T
$$

is the restricted projection matrix and is analogous to the projection matrix in typical OLS analysis. (Recall that \( [X^TX + R^TR]^{-1} \) is a generalised inverse of \( X^TX \).) Using (1), Lemma (5b) and the fact that \( XV = 0 \), it can be shown that \( P^* \) is idempotent with rank \( k - h \).

The residual values are given by: \( \hat{u} = y - \hat{y} = M^*y \), where \( M^* = I - P^* \) is the restricted residual matrix and is analogous to the residual matrix in typical least square analysis. Clearly, \( M^* \) is symmetric idempotent with rank \( n - (k - h) \). By (1) and Lemma (5b), the restricted projection and residual matrices satisfy the usual properties that: \( P^*X = X \), \( M^*X = 0 \) and \( P^*M^* = 0 \).

The next proposition shows that the fitted values and the residuals (and measures of goodness-of-fit, such as \( R^2 \)) from RLS regression do not depend upon the choice of parametric restrictions.

**Proposition 3** \( P^* \) and \( M^* \) are invariant to the matrix of restrictions \( R \).

**Proof.** Let \( R_1 \neq R_2 \) be arbitrary restriction matrices satisfying \( |R_iV| \neq 0 \), let \( P_1^* \) and \( P_2^* \) be the corresponding projection matrices and let \( W_i = (X^TX + R_i^TR_i)^{-1} \) for each \( i \). Since \( W_i \) is a generalised inverse, \( X^TXW_iX^TX = X^TX = X^TXW_2X^TX \). Then \( XW_1X^T = XW_2X^T \) (since by Lemma 2, \( ABB^T = CBB^T \Rightarrow AB = CB \)). Hence \( P_1^* = P_2^* \). The proof for \( M^* \) is straightforward. \( \blacksquare \)
Proposition 3 confirms the intuition that the parameter restrictions do not affect the performance of the model - they serve simply to identify the coefficient estimates appropriately. This is also consistent with the result that the Lagrange multipliers $\lambda^*$ - which provide a measure of the loss of fit associated with each parameter restriction - are zero for any choice of parametric restrictions.

So far, no assumptions have been made regarding the nature of the stochastic error term. Suppose errors satisfy $E[u] = 0$ and $V[u|X] = \sigma^2 I$. Then the mean and variance-covariance of the RLS estimators are given by:

\[
E[\hat{\beta}] = \beta + V[RV]^{-1}(r - R\beta)
\]
\[
V[\hat{\beta}] = \sigma^2 \left[ X^T X + R^T R \right]^{-1} X^T X \left[ X^T X + R^T R \right]
\]

This implies that if the parameter restrictions are appropriately chosen (in some sense), then the restricted least square estimators are unbiased. Furthermore, the expression for the variance-covariance matrix of the estimators is analogous to the variance of OLS estimators (since $\left[ X^T X + R^T R \right]^{-1}$ is a generalised inverse of $X^T X$). I note that the variance/covariance matrix for the restricted least square estimators is simply the upper-left block of the inverse of the restricted normal matrix, which confirms the result in Green and Seaks (1991).

Finally, I show that the RLS estimators for different sets of parametric restrictions are simply linear combinations of one another. It follows that, having derived the estimators under one set of parametric restrictions, the estimates for any other set of restrictions can be found without needing to repeat the above procedure.

**Proposition 4** Let $R_1 \beta = r_1$ and $R_2 \beta = r_2$ be two linear restrictions on the parameter-space that satisfy the conditions in Proposition 1, and let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the corresponding parameter estimates. Then:

\[
\hat{\beta}_2 = \left[ I - V[RV]^{-1} R_2 \right] \hat{\beta}_1 + V[RV]^{-1} r_2
\]

**Proof.** By direct verification, noting that $\left[ X^T X + R^T R \right]^{-1} X^T X = \left[ I - V[RV]^{-1} R \right]$ and $XV = 0$. ■
5 Conclusion

This note considered the problem of finding the restricted least square estimators in models with singularity. I found sufficient conditions for a set of parametric restrictions to uniquely identify a set least square estimators and provided a closed-form expression for these estimators. Unsurprisingly, the restricted least square estimators are analogous in many ways to the ordinary least square estimators in the classical model. Importantly, I showed that the performance of the model was independent of the parametric restrictions, and so the imposition of parametric restrictions comes at no cost, but facilitates the estimation of coefficients that can be more naturally interpreted.

6 Appendix

The following Lemmata will be used in the proof of Proposition 1:

Lemma 1 Let $A$ be an $m \times n$ matrix and let $x, z \in \mathbb{R}^n$. Then:

1. $z \in M(A^T)$ if and only if $x^Tz = 0 \ \forall x \in N(A)$

2. $x \in N(A)$ if and only if $x^Tz = 0 \ \forall z \in M(A^T)$

Proof. To prove (1), suppose $z \in M(A^T)$. Then there exists $y \in \mathbb{R}^m$ such that $z = A^Ty$. Then for every $x \in N(A)$, $x^Tz = x^TA^Ty = 0$ (since $Ax = 0$). Conversely, suppose $x^Tz = 0 \ \forall x \in N(A)$. Let the columns of $V$ form a basis for $N(A)$. Then $\alpha V^Tz = 0 \ \forall \alpha \in \mathbb{R}^{n-r(A)}$ and so $V^Tz = 0$. This implies $z \in N(V^T)$. Note further that $V^TA^Ty = 0 \ \forall y \in \mathbb{R}^m$ and so the columns of $A^T$ span $N(V^T)$. But then $\exists y \in \mathbb{R}^m$ s.t. $z = A^Ty$. Hence $z \in M(A^T)$. (2) can be shown in a similar way.

Lemma 2 For any matrix $A$, $N(A^TA) = N(A)$ and $M(AA^T) = M(A)$

Proof. Clearly $N(A) \subseteq N(A^TA)$. Suppose $x \in N(A^TA)$. Then $x^TA^TAx = 0$, which implies $(Ax)^T(Ax) = 0$. Hence $Ax = 0$ and so $x \in N(A)$. Hence $N(A^TA) \subseteq N(A)$, which proves the first claim. It is similarly clear that $M(AA^T) \subseteq M(A)$ (since $A^Tz \in \mathbb{R}^n$ for $\forall z \in \mathbb{R}^m$). Suppose $z \in M(A)$. Then, by Lemma 1, $x^Tz = 0 \ \forall x \in N(A^T) = N(AA^T)$. Then again by Lemma 1, $z \in M(AA^T)$ and so $M(A) \subseteq M(AA^T)$. □
Lemma 3 Let $A \in \mathbb{R}^{m \times n}$ with $r(A) = n - s$, let $B \in \mathbb{R}^{s \times n}$ with $r(B) = s$ and let the columns of $V \in \mathbb{R}^{n \times s}$ form a basis for $\mathcal{N}(A)$. Then $\mathcal{M}(A^T) \cap \mathcal{M}(B^T) = \{0\}$ if $|BV| \neq 0$.

Proof. Clearly $0 \in \mathcal{M}(A^T) \cap \mathcal{M}(B^T)$. Suppose $BV$ is non-singular. Take any $z \neq 0$ and suppose $z \in \mathcal{M}(A^T) \cap \mathcal{M}(B^T)$. Since $z \in \mathcal{M}(B^T)$, there is some $y \neq 0 \in \mathbb{R}^s$ such that $z = B^Ty$. Further, $V^TB^Ty \neq 0$, since $BV$ is non-singular, and so there is some $a \in \mathbb{R}^s$ such that $a^TV^TB^Ty \neq 0$. Now, by Lemma 1, $x^Tz = 0 \forall x \in \mathcal{N}(A)$. For every $a \in \mathbb{R}^s$, there is some $x \in \mathcal{N}(A)$, such that $x = Va$. But this implies $a^TV^TB^ty = 0 \forall a \in \mathbb{R}^s$, which is a contradiction. □

Lemma 4 Let $A \in \mathbb{R}^{m \times n}$ with $r(A) = n - s$ and $B \in \mathbb{R}^{s \times n}$ with $r(B) = s$. If $\mathcal{M}(A^T) \cap \mathcal{M}(B^T) = \{0\}$, then $(A^TA + B^TB)$ is invertible.

Proof. By Lemma 2, $r(A^TA) = n - s$ and $r(B^TB) = s$. Let the columns of $P \in \mathbb{R}^{n \times (n-s)}$ form a basis for $\mathcal{M}(A^T)$ and let the columns of $Q \in \mathbb{R}^{n \times s}$ form a basis for $\mathcal{M}(B^T)$. Since $\mathcal{M}(A^T) \cap \mathcal{M}(B^T) = \{0\}$, the columns of $P$ and $Q$ must be jointly linearly independent. Let $S = [P,Q]$ and note that $r(S) = n$. For any $y \in \mathbb{R}^n$, there is some $x = (x_p, x_q)$ with $x_p \in \mathbb{R}^{n-s}$ and $x_q \in \mathbb{R}^s$, such that $(A^TA + B^TB)y = A^TAy + B^TBy = Px_p + Qx_q = Sx$. Hence $S$ forms a basis for $(A^TA + B^TB)$, which implies $(A^TA + B^TB)$ has full rank. □

The above lemmata establish the rank properties of matrices that are constructed from an underlying matrix which has less than full rank. In the RLS framework, the idea is to extend the normal matrix using appropriate restrictions in such a way that the resulting expression has full rank, even though the underlying expression $(X^TX)$ is singular.

Proof of Proposition 1. First, I show that the restricted normal matrix $(N)$ is invertible. Since $RV$ is non-singular, Lemma 3 implies that $\mathcal{M}(X^T) \cap \mathcal{M}(R^T) = \{0\}$. For any $y = (y_k, y_h) \in \mathbb{R}^{k+h}$, $Ny = (X^TXy_k + R^T y_h, R y_k)^T$. It can be shown using a similar argument to that in the proof of Lemma 4, that: $[X^TX, R^T]$ spans $\mathbb{R}^k$ and $[R,0]$ spans $\mathbb{R}^h$ (since $r(R) = s$). Hence $N$ spans $\mathbb{R}^{k+h}$ and so the restricted normal matrix is invertible.
Next, I show that $W$ is a generalised inverse of $X^T X$. Note that Lemmas 3 and 4 imply that $W$ exists. Then, since

$$WR^T R = I - WX^T X$$

(1)

it suffices to show that $XTXWR^T R = 0$. Let $e_j$ be the $j^{th}$ elementary vector and let:

$$y_j = WR^T Re_j$$

Then, $R^T Re_j = (X^T X + R^T R)y_j$. Now, since $M(X^T) \cap M(R^T) = \{0\}$, then $X^T Xy_j = 0$. Hence $e_i X^T XWR^T Re_j = 0$. Since this is true for $\forall i, j \in \{1, ..., n\}$,

$$X^T XWR^T R = 0$$

(2)

The following Lemma will be used in the proof of Proposition 2:

**Lemma 5** Let $R, V, W$ and $X$ be defined as above. Then:

a. $RWR^T = I$

b. $WR^T = V[RV]^{-1}$

**Proof.** (1) and (2) imply that $R^T RWR^T R = R^T R$. Since $RR^T$ is invertible, pre and post-multiplying this expression by $[RR^T]^{-1} R$ and $R^T [RR^T]^{-1}$ (respectively) implies (a). Now, by (2) and Lemma 2, $WR^T R \in N(X)$ and so $WR^T R = VD$ for some matrix $D$. Premultiplying by $R$ and using (a) implies $D = [RV]^{-1} R$, and so $WR^T R = V[RV]^{-1} R$. Then, post-multiplying by $R^T [RR^T]^{-1}$ gives (b). ■

**References**


