A STRAIGHTFORWARD APPROACH TO THE PROBLEM OF MULTICOLLINEARITY IN ECONOMETRICS

By Alexis Lazaridis* (PH. D.)

Preface

Given an econometric model with two or more variables in the explanatory list being linearly dependent, then the coefficients of the model cannot be estimated through using the standard econometric methods.

Many techniques have been developed to face this problem of mulicollinearity, but all of them call for a radical reform of the initial data set. Besides most of these techniques are rather based on heuristic consideration.

In this paper we develop a procedure for obtaining efficient estimators of the parameters when we have to work with a data set which is characterized even by extreme multicollinearity.

According to this procedure the parameter vector can be estimated by avoiding the explicit imposition of additional linear resitctions which call for a priori information.

Given the linear model

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{u} \tag{1}$$

the estimator vector $\hat{\mathbf{b}}$ is obtained by minimizing (least-squares method) the scalar function

$$S = \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 \Rightarrow S = \|\mathbf{Y} - \mathbf{X}\hat{\mathbf{b}}\|^2 \tag{2}$$

$$\Rightarrow S = \mathbf{Y}'\mathbf{Y} - 2\hat{\mathbf{b}}'\mathbf{X}'\mathbf{Y} + \hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} \tag{3}$$

(primes denote transposition)

The minimization of (3) with respect to $\hat{\mathbf{b}}$ yields

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{4}$$

---

* Ministry of Co-ordination and Planning
Mailing address:
49, Proxenou Koromila street,
Thessaloniki,
Greece.

For the solution to be unique, (3) must be strict convex. This implies that the Hessian matrix of (3) must be positive definite.[1] It is very easy to show that the Hessian matrix under consideration has the form

$$H = 2X'X \tag{5}$$

For the matrix $(X'X)$ to be positive definite it is necessary that matrix X to have full column rank. Thus the inverse of $X'X$ in (4) exists.

A serious problem arises (known as multicollinearity in econometric literature) if the full column rank condition of X is not satisfied.

It is the purpose of this paper to develop a procedure for obtaining the estimator vector b^* regardless of the column rank of matrix X.

Denoting the rank of the latter matrix by r(X) it will be shown that if X, which is defined on $E^m \times E^n$, has $r(X) = n$, then $b^* = b$. if $1 \leqslant r(X) < n$, then the obtained solution is unique in the sence that b^* has least norm[2] and minimizes the sum of squares

$$\|Y - Xb^*\|^2 \tag{6}$$

Given a non-singular square matrix A it is known that A^{-1} satisfies the equations:

$AA^{-1}A = A$, A^{-1}, $AA^{-1} = A^{-1}$, $(A^{-1})^{-1} = A$,
$(AA^{-1})' = (A^{-1}A)' = A^{-1}A$

If A is singular and not necessarily square it is possible to compute a matrix Z (known as generalized inverse and denoted by $Z = A^+$) which satisfies the above equations, i.e.

$AA^+A = A$, $A^+AA^+ = A^+$, $(A^+)^+ = A$,
$(A^+A)' = A^+A$, $(AA^+)' = AA^+$

If A is square and symmetric then $A^+A = AA^+$.

If A is not square $(m \neq n)$ but has full row or column rank then A^+ is the right or left inverse of A, i.e.

$$\begin{aligned} r(A) = m &\Rightarrow AA^+ = I_m \\ r(A) = n &\Rightarrow A^+A = I_n \end{aligned} \tag{7}$$

If A is square $(n \times n)$ with $r(A) = n$, it is easy to verify from the above identities that $A^{-1} = A^+$.

The main point regarding the generalized inverse, is that it always exists and is unique (Greville, 1960).

To show that if $r(X) = n$ then $b^* = b$, we consider the relations (2), (6) and (7).

---

(1) All pricipal minors to be greater than zero.

(2) Throughout this paper the Euclidean norm is considered.

The direct minimization of (6) using the generalized inverse, yields the optimally estimated vector $\hat{\mathbf{b}}^*$ defined by

$$\hat{\mathbf{b}}^* = X^+ \mathbf{Y} \tag{8}$$

Given that $X^+ X = I$, since $r(X) = n$, it is implied that

$$X^+ = (X' X)^{-1} X' \tag{9}$$

so that

$$\hat{\mathbf{b}}^* = (X' X)^{-1} X' \mathbf{Y} \tag{10}$$

Eq. (9) can also be written as

$$X^+ = (X' X)^+ X' \Rightarrow \hat{\mathbf{b}}^* = (X' X)^+ X' \mathbf{Y} = \hat{\mathbf{b}}$$
$$\text{since } (X' X)^{-1} = (X' X)^+.$$

This is important from the computational point of view, since $(X' X)^+$ can be computed much easier than X^+, given the dimension and the properties of the former matrix.

The case of interest arises when $1 \leqslant r(X) < n$, which implies that matrix $(X' X)$ is singular.

In the appendix we develop the procedure for computing the generalized inverse of a singular matrix.

In such singular cases $(X^+ X) \neq I_n$ but the products $(X^+ X) = (X' X)^+ (X' X)$ have the properties of an idempotent matrix with $r(X^+ X) = r(X)$. In addition it is easy to verify that

$$(X' X)^+ (X' X) = (X' X)\ (X' X)^+ \tag{11}$$

$$X(X' X)^+ (X' X) = XX^+ X = X \tag{12}$$

In what follows we establish the properties of the optimally estimated vector

$$\hat{\mathbf{b}}^* = X^+ \mathbf{Y} = (X' X)^+ X' \mathbf{Y} \tag{13}$$

Substitution of eq. (1) into (13) for $\mathbf{Y}$ yields

$$\hat{\mathbf{b}}^* = X^+ (X\mathbf{b} + \mathbf{u}) = X^+ X\mathbf{b} + X^+ \mathbf{u}$$

and

$$E(\mathbf{b}^*) = X^+ X\mathbf{b} \qquad (\text{since } E(\mathbf{u}) = 0) \tag{14}$$

Considering the norm of eq. (14) we have

$$\| E(\mathbf{b}^*) \| \leqslant \|X^+ X\| \ \|\mathbf{b}\| \Rightarrow \|E(\hat{\mathbf{b}}^*)\|^2 \leqslant \|X^+ X\|^2 \ \|\mathbf{b}\|^2 \tag{15}$$

Since $(X^+ X)$ is an idempotent matrix it is implied that

$$\|X^+ X\|^2 = tr(X^+ X) = r(X^+ X) = r(X)$$

Even if $r(X) = 1$, $E(\mathbf{\hat{b}}^*)$ has least norm as it is verified from (15).

In general, for any $\mathbf{\tilde{b}}$, either $\|\mathbf{Y} - X\mathbf{\tilde{b}}\| > \|\mathbf{Y} - X\mathbf{\hat{b}}^*\|$ or $\|\mathbf{Y} - X\mathbf{\tilde{b}}\| = \|\mathbf{Y} - X\mathbf{\hat{b}}^*\|$ and $\|\mathbf{\tilde{b}}\| \geqslant \|\mathbf{\hat{b}}^*\|$, which implies that eq. (8) gives the minimal norm solution (see appendix).

To prove that $\mathbf{\hat{b}}^*$ is the best unbiased estimator of $\mathbf{b}$ it is assumed that there exists another linear estimator denoted by

$$\mathbf{\tilde{b}} = A\mathbf{Y} = A(X\mathbf{b} + \mathbf{u}) \tag{16}$$

Hence

$$E(\mathbf{\tilde{b}}) = AE(\mathbf{Y}) = AX\mathbf{b} \tag{17}$$

Matrix A is defined on $E^n \times E^m$.

For the vector $\mathbf{\tilde{b}}$ to be an unbiased estimator of $\mathbf{b}$ it is required that

$$AX = I \tag{18}$$

so that

$$E(\mathbf{\tilde{b}}) = \mathbf{b} \tag{19}$$

It is implied that if eq. (19) holds, then it must also be true that

$$XE(\mathbf{\tilde{b}}) = X\mathbf{b} \tag{20}$$

In other words for the case at hand the condition $AX = I$ changes to

$$XAX = X \tag{21}$$

Although it will be shown later that this is the case, we assume at this point that the product XA is a symmetric matrix so that $XA = A'X'$. Hence eq. (21) can be written as

$$A'X'X = X \tag{22}$$

The variance - converiance matrix of $\mathbf{\tilde{b}}$, denoted by $\Sigma_{\tilde{b}}$ is

$$\begin{aligned} \Sigma_{\tilde{b}} &= E\{(\mathbf{\tilde{b}} - E(\mathbf{\tilde{b}}))(\mathbf{\tilde{b}} - E(\mathbf{\tilde{b}}))'\} \\ &= E\{(A\mathbf{Y} - AE(\mathbf{Y}))(A\mathbf{Y} - AE(\mathbf{Y}))'\} \\ &= E\{A(\mathbf{Y} - E(\mathbf{Y}))(A(\mathbf{Y} - E(\mathbf{Y}))'\} \\ &= E\{A\mathbf{uu}'A'\} = AE(\mathbf{uu}')A' \\ &= A\sigma^2 IA' = \sigma^2 AA' \end{aligned} \tag{23}$$

To solve the problem (see the appendix for details)

min $\sigma^2 AA'$

s.t.

$A'X'X = X$

we form the vector - valued fanction (Lagrangean)

$$L = \text{diag}\,[\sigma^2 A' A] + \text{diag}\,[(X - A' X' X)\Lambda'] \tag{24}$$

Where Λ, the matrix of Lagrange multipliers, is defined on $E^m \times E^n$. The resulting normal equations from (24) are of the form:

$$2\sigma^2 A' = \Lambda X' X \tag{25}$$

$$A' X' X = X \tag{26}$$

From eq. (25) we get:

$$A' = \frac{1}{2\sigma^2} \Lambda X' X \tag{27}$$

and

$$A = \frac{1}{2\sigma^2} X' X \Lambda' \tag{27a}$$

Eq. (27) is inserted into eq. (26) to yield

$$\frac{1}{2\sigma^2} \Lambda X' X X' X = X \Rightarrow \Lambda X' X X' X = 2\sigma^2 X \tag{28}$$

Eq. (28) holds iff $\Lambda = 2\sigma^2 X(X' X)^+ (X' X)^+$. Hence eq. (28) can be written as

$$X(X' X)^+ (X' X)^+ (X' X)\ (X' X) = X \tag{29}$$

Since the matrices in brackets are symmetric and due to the fact that $(X' X)^+ (X' X)$ is an idempotent matrix, the left-side of eq. (29) reduces to

$$X(X' X)^+ (X' X)\ (X' X)^+ (X' X) = X(X' X)^+ (X' X) = XX^+ X = X$$

Since $\Lambda = 2\sigma^2 X(X' X)^+ (X' X)^+$ it is implied that $\Lambda' = 2\sigma^2 (X' X)^+ \cdot (X' X)^+ X'$ which is inserted into eq. (27a) to yield

$$A = \frac{2\sigma^2}{2\sigma^2} (X' X)\ (X' X)^+ (X' X)^+ X' = (X' X)^+ X' = X^+ \tag{30}$$

Hence, the assumption made earlier is entirely justified, since $XX^+ = X(X' X)^+ X'$ so that $XX^+ = (X^+)' X'$.

By inserting the value of A, in eq. (30), into eq. (16) we get

$$\tilde{b} = X^+ Y = b^*$$

which indicates that the minimum variance property is also satisfied.

The direct substitution of the value of A and A′ into eq. (23) yields

$$\Sigma_{\hat{b}} = \sigma^2 AA' = \sigma^2 X^+ (X^+)'$$
$$= \sigma^2 (X'X)^+ X' (X'X)^+ X')'$$
$$= \sigma^2 (X'X)^+ (X'X) (X'X)^+$$
$$= \sigma^2 (X'X)^+$$

The estimator of the scalar variance σ^2 is obtained according to the known way, i.e.

$$\begin{aligned}\hat{\mathbf{u}}^* &= \mathbf{Y} - X\hat{\mathbf{b}}^* \\ &= X\mathbf{b} + \mathbf{u} - XX^+\mathbf{Y} \\ &= X\mathbf{b} + \mathbf{u} - X(X'X)^+ X'\mathbf{Y} \\ &= X\mathbf{b} + \mathbf{u} - X(X'X)^+ X'(X\mathbf{b} + \mathbf{u}) \\ &= X\mathbf{b} + \mathbf{u} - X(X'X)^+ (X'X)\mathbf{b} - X(X'X)^+ X'\mathbf{u} \\ &= X\mathbf{b} + \mathbf{u} - X\mathbf{b} - X(X'X)^+ X'\mathbf{u}, \text{ since } X(X'X)^+ X'X = X\end{aligned}$$

and

$$\begin{aligned}\hat{\mathbf{u}}^* &= \mathbf{u} - X(X'X)^+ X\mathbf{u} \\ &= \{I_m - X(X'X)^+ X'\}\mathbf{u} = \{I - XX^+\}\mathbf{u}\end{aligned}$$

Hence

$$\hat{\mathbf{u}}^{*\prime}\hat{\mathbf{u}}^* = \mathbf{u}'M\mathbf{u} = \mathrm{tr}(M\mathbf{u}\,\mathbf{u}')$$

where $M = \{I - XX^+\}$ is an idempotent matrix and tr(A) denotes the trace of matrix A.

Finally it is:

$$E(\hat{\mathbf{u}}^{*\prime}\hat{\mathbf{u}}^*) = \mathrm{tr}\{ME(\mathbf{u}\,\mathbf{u}')\} = \sigma^2\,\mathrm{tr}(M), \text{ since } E(\mathbf{u}\,\mathbf{u}') = \sigma^2 I$$

and

$$\begin{aligned}E(\hat{\mathbf{u}}^{*\prime}\hat{\mathbf{u}}^*) &= \sigma^2\mathrm{tr}(I_m) - \sigma^2\mathrm{tr}(X(X'X)^+ X') \\ &= \sigma^2 m - \sigma^2\mathrm{tr}(X(X'X)^+ X'X) \\ &= \sigma^2 m - \sigma^2 r((X'X)^+ X'X) \\ &= \sigma^2(m - r(X))\end{aligned}$$

since $\mathrm{tr}((X'X)^+ X'X) = r((X'X)^+ X'X) = r(X)$

If the estimator of σ^2 is denoted by S^2 then it is determined from:

$$S^2 = \frac{\hat{\mathbf{u}}^{*\prime}\hat{\mathbf{u}}^*}{m - r(X)} \qquad \text{so that } E(S^2) = \sigma^2$$

We have applied the developed procedure using a set of data characterized by extreme multicollinearity. The data used and the results obtained are ilustrated in the table which follows.

The model: $Y_i = b_1 + b_2 X_{2i} + b_3 X_{3i} + u_i$

i	Y_i	X_{2i}	$X_{3i} = 2X_{2i}$	$\hat{Y}_i^*$	$\hat{u}_i^*$
1	18	4	8	21.5883	−3.5883
2	19	3	6	18.8240	0.1760
3	25	5	10	24.2942	07058
4	28	6	12	27.0001	0.9999
5	33	10	20	37.8237	−4.8237
6	39	8	16	32.4119	6.5881

$\Sigma \hat{u}_i^* = 0.0578$, $\Sigma \hat{u}_i^{*2} = 81.07597$

SST $= 330$, SSR $= 248.924$

$R^2 = 0.754$, $S^2 = 20.26899$

The estimated model is:

$$\hat{Y}_i^* = \underset{\substack{(4.98392)\\(2.15989)}}{10.7647} + \underset{\substack{(0.154389)\\(3.50507)}}{0.541144}\ X_{2i} + \underset{\substack{(0.308846)\\(3.50459)}}{1.08238}\ X_{3i}$$

The figures in brackets are standrad errors of coefficients and the values of the computed t, respectively.

In the econometric literature it is suggested that the above model, i.e.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 (2X_{2i}) + u_i$$

to be written as

$$Y_i = \beta_1 + (\beta_2 + 2\beta_3) X_{2i} + u_i \Rightarrow Y_i = \beta_1 + \gamma X_{2i} + u_i \quad (31)$$

It is also mentioned that although we can obtain the estimators for β_1 and $\gamma = \beta_2 + \beta_3$, "there is no way to estimate the coefficients β_2 and β_3 separately" (Christou, 1978, p. 282).

From the given set of data, $\hat{\gamma}$ is found to be $\hat{\gamma} = 2.7059$. Recalling that

$$\hat{\gamma} = \hat{\beta}_2 + 2\hat{\beta}_3 \quad (32)$$

it can be easily seen that the estimated coefficients β_2^* and β_3^*, according to the procedure developed in this paper, satisfy equation (32), i.e. $\hat{\gamma} = \beta_2^* + 2\beta_3^*$.

Another point of interest is that $V(\hat{\gamma}) = S^2_{\hat{\gamma}}$ has found to be (form eq. (

$$V(\hat{\gamma}) = 0.596$$

Since $\hat{\gamma} = \alpha_2\hat{\beta}^*_2 + \alpha_3\hat{\beta}^*_3$ (where $\alpha_2 = 1$ and $\alpha_3 = 2$)
it is recalled that

$$V(\hat{\gamma}) = \sum_{i=2}^{3} \alpha^2_i V(\hat{\beta}^*_i) + 2\sum_{i<j} \alpha_i \alpha_j C(\hat{\beta}^*_i, \hat{\beta}^*_j) = 0.596$$

since $C(\hat{\beta}^*_2, \hat{\beta}^*_3) = 0.047693$.

Hence the estimated vector $\mathbf{b}^*$ has the following properties:

– It has least norm

– Minimizes $\hat{\mathbf{u}}^{*\prime} \hat{\mathbf{u}}^* = \hat{\mathbf{u}}' \hat{\mathbf{u}}$ which refer to the models:

$$\hat{Y}^*_i = \beta^*_1 + \beta^*_2 X_{2i} + \beta^*_3 X_{3i} + \hat{u}^*_i$$
$$\hat{Y}_i = \beta_1 + (\beta_2 + 2\beta_3) X_{2i} + \hat{u}_i \text{ since } X_{3i} = 2X_{2i}$$

(it is also noted that $\hat{Y}^*_i = \hat{Y}_i$).

– Its elements have minimum variances which satisfy eq. (33).

Furthermore and contrary to what is mentioned in the econometric literature
Theil, 1970, p. 148), the mean vector of the dependent variables is unique,

$$E(\mathbf{Y}) = XE(\mathbf{b}^*) = X\mathbf{b}$$

since we have shown that $XE(\mathbf{b}^*) = X\mathbf{b}$

The above features, together with some additional ones mentioned in the text
the main outcomes of this paper.

APPENDIX

On the computation of the generalized inverse of a singular symmetric ma

It it assumed that the generalized inverse of matrix

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \text{ is required.}$$

Step 1

Compute the egenvalues and the asociated eigenvectors of $A'A = AA'$.

Denoting by R the diagonal matrix of eigenvalues and by V the matrix of
associated eigenvectors we get for the above example

$$R = \begin{bmatrix} 25 & \\ & 0 \end{bmatrix} \quad V = [V_1\ V_2] = \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}$$

It is required that matrix V to be orthonormalized. The conditions to be satisfied are: $<V_1, V_1> = 1$, $<V_2, V_2> = 1$ and $<V_1, V_2> = 0$.

The orthonormalized V matrix is

$$V = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{bmatrix}$$

Step 2

Define a diagonal matrix F with elements $f_{ii} = f_i$, which are the non-negative square roots of the eigenvalues of A′ A. If the rank of A is $r \leqslant n$, then

$$f_1 \geqslant f_2 \geqslant ....f_r > 0$$

and

$$f_{r+1} = f_{r+2} = ... f_n = 0$$

For the above example matrix F has the form

$$F = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}$$

Now matrix A can be written as

$$A = VFV'$$

which is called the singular value decomposition of A.

Step 3

Define the diagonal matrix F* with elements

$$f^*_{ii} = f^*_i = 1/f_i \text{ for } i \leqslant r$$

and

$$f^*_i = 0 \text{ for } i > r$$

F* for the case at hand is

$$F^* = \begin{bmatrix} 1/5 & 0 \\ 0 & 0 \end{bmatrix}$$

Step 4

Determine the generalized inverse A^+ of the singular matrix A from

$$A^+ = VF^* V'$$

For the above example A^+ is

$$A^+ = \begin{bmatrix} 0.04 & 0.08 \\ 0.08 & 0.16 \end{bmatrix}$$

and satisfies all the equations mentioned in the text.

we have developed a routine which is available at the University of Thessaloniki Computer Centre, for computing the generalized inverse of any real matrix.

At this point it is constractive to recall that for a real symmetric ($n \times n$) matrix:

- all eigenvalues and the corresponding eigenvectors are real
- two eigenvectors corresponding to distinct eigenvalues are orthogonal
- the eigenvectors form a basis for E^n
- there exists at least one orghonormal set of eigenvectors which provides a basis for E^n

---

On the minimization of $(\sigma^2 AA' \mid A' X' X = X)$.

Assuming that for i=1, 2,...,n, the vectors $\mathbf{a}_{r-i} \in E^m$ are the n rows of matrix A and the vectors $\mathbf{a}_{c-j} \in E^n$ are the m columns (j=1, 2,..,m) of A, then in order to minimize $(\sigma AA' \mid A' X' X = X)$ we require

$$\text{either } \sigma^2 \| \mathbf{a}_{r-i} \|^2 \text{ to be minimum, for } i=1, 2,....., n$$

$$\text{or } \quad \sigma \| \mathbf{a}_{c-j} \|^2 \text{ to be minimum, for } j=1, 2,...., m$$

subject to the given set of constraints.

For the first case the lagrangean (vector-valued function) will be

$$L = (\text{diag}\{\sigma^2 AA'\})' + (\text{diag}\{\Lambda' (X - A' X' X)\})'$$

The notation diag (Z) denotes that the diagonal elements of the square matrix Z are only considered.

Λ, the matrix of lagrange multipliers, is defined on $E^m \times E^n$. In order to derive the analytical partial derivatives of each l_i with respect to the components of each $\mathbf{a}_{r-i}$, all the n columns of the row-vector $(\text{diag}\{\Lambda' (X - A' X' X)\})'$ must be considered.

For the second case, the Lagrangean will be

$$L = \text{diag}\{\sigma^2 AA'\} + \text{diag}\{(X - A' X' X)\Lambda'\}$$

The column-vectors in the above specification are m-dimensional. In this case, the gradient of each l_j with respect to each $\underset{c-j}{\mathbf{a}}$ (i.e. $\nabla_{\mathbf{a}_{c-j}} l_j$) is determined from the corresponding j^{th} row of the right-side vectors.

In both cases the resulting normal equations will be of the form

$$2\sigma^2 A' = \Lambda X' X$$
$$A' X' X = X$$

The property of minimal norm

We want to minimize

$$J_{\tilde{\mathbf{b}}} = \|\mathbf{Y} - X\tilde{\mathbf{b}}\|^2 \qquad (34)$$

given that the real matrix X is defined on $E^m \times E^n$ ($m > n$). If $r(X) < n$ the minimizing $\tilde{\mathbf{b}}$ is not usually unique unless one imposes additional restrictions such as the condition that $\|\tilde{\mathbf{b}}\|$ is also minimal.

The vector $\tilde{\mathbf{b}}^*$ given by

$$\tilde{\mathbf{b}}^* = X^+\mathbf{Y} \text{ or } \tilde{\mathbf{b}}^* = (X'X)^+ X'\mathbf{Y} \qquad (35)$$

is the one with minimum norm, as it will be shown in what follows.

Denoting X′ X by D we can write

$$\tilde{\mathbf{b}} = \tilde{\mathbf{b}}_1 + \tilde{\mathbf{b}}_2$$

where

$$\tilde{\mathbf{b}}_1 \in R(D) \quad \text{(the range space of D)}$$
$$\tilde{\mathbf{b}}_2 \in N(D) \quad \text{(the null space of D)}$$

Since D is symetric, R(D) and N(D) are orthogonal, so that

$$\|\tilde{\mathbf{b}}\|^2 = \|\tilde{\mathbf{b}}_2\|^2 + \|\tilde{\mathbf{b}}_2\|^2 \qquad (36)$$

To derive (35) and to prove the minimal norm property we expand (34), assuming that $r(X) = n$.

Thus,

$$\begin{aligned}
J_{\tilde{b}} &= \mathbf{Y}'\mathbf{Y} - 2\tilde{\mathbf{b}}' X' Y + \tilde{\mathbf{b}}' D\tilde{\mathbf{b}} \\
&= \mathbf{v}'\mathbf{Y} - 2\mathbf{Y}' XD^{-1} X'\mathbf{Y} + \mathbf{Y}'XD^{-1} DD^{-1}X'\mathbf{Y} \text{ (since } \tilde{b} = \tilde{\mathbf{b}}) \\
&= \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}' XD^{-1} X'\mathbf{Y} + \mathbf{Y}' XD^{-1} X'\mathbf{Y} \\
&= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}' XD^{-1} X'\mathbf{Y} \\
&= (\mathbf{Y}, (I - XD^{-1} X')\mathbf{Y}) \\
&= \|\mathbf{Y}\|^2_{(I - XD^{-1} X')}
\end{aligned}$$

where (**a**, **c**) denotes the inner product, i.e. **a**′ **c**

Considering the scalar function

$$J = \| \tilde{b} - D^+ X' Y \|^2_D + \| Y \|^2_{(I-XD^+X')} - 2(X'Y, (I-D^+D)\tilde{b}) \tag{37}$$

and expanding it, one obtaines [recalling that $D^+DD^+ = D^+$ and $(D^+)' = D^+$ since D is symmetric]

$$J = \tilde{b}' D\tilde{b} - 2Y'XD^+D\tilde{b} + Y'XD^+X'Y + Y'Y - Y'XD^+X'Y - 2Y'X\tilde{b} + 2Y'XD^+D\tilde{b}$$

and

$$J = \tilde{b}' D\tilde{b} - 2Y'X\tilde{b} + Y'Y = J_{\tilde{b}}$$

It is noted that the term $\|\tilde{b} - D^+X'Y\|^2_D$ in (37) never becomes negative and it vanishes when $\tilde{b} = \tilde{b}^*$.

Since $J = J\tilde{b}$ we consider (37), noting that

$$\begin{aligned}(I-D^+D)\tilde{b} &= \tilde{b}_1 + \tilde{b}_2 - D^+D(\tilde{b}_1 + \tilde{b}_2) \\ &= \tilde{b}_1 + \tilde{b}_2 - \tilde{b}_1 \\ &= \tilde{b}_2 \text{ since } D^+D \text{ is restricted to. } R(D).\end{aligned}$$

For the same reason, and due to the orthogonality condition, the first term of the right-side of (37) becomes

$$\| \tilde{b} - D^+X'Y \|^2_D = \| \tilde{b}_1 - D^+X'Y \|^2_D$$

Hence, (34) can be written

$$J_{\tilde{b}} = \|\tilde{b}_1 - D^+X'Y\|^2_D + \|Y\|^2_{(I-XD^+X')} - 2(X'Y, \tilde{b}_2) \tag{38}$$

The first term in (38) is minimized, as it has been mentioned above, by choosing

$$\tilde{b}_1 = D^+X'Y$$

In view of eq. (36) and from the requirement of the minimal norm of $\tilde{b}$, we have

$$\tilde{b}_2 = 0 \tag{39}$$

Hence, $\tilde{b} = D^+X'Y = \tilde{b}^*$

References

Bellman, R., (1961), Introduction to Matrix Analysis, Princeton University Press, New Jersey.

Canon, M.D., C.D. Cullum, Jr., and E. Polak, (1966), "Constrained Minimization Problems in finite Dimensional Space", SIAM Journal of Control pp. 528.547.

Christou, G., (1978), An introduction to Econometrics, 1, School of Industrial Studies, Thessaloniki, Greece (in Greek).

Dhrymes, P.J., (1970), Econometrics: Statistical Foundations and Applications, Happer and Row Publishers, New York.

Greville, T.N.E., (1960), "Some Applications of the Pseudoinverse of a Matrix", J. SIAM Review, 2, pp. 15-22.

Johnston, J., (1972), Econometric Methods, (2nd Edition), McGraw-Hill Book Co., New York.

Kmenta, J., (1971), Elements of Econometrics, the McMillan Company, New York.

McMillan, G., (1975), Mathemtical Programming, (2nd Edition), John Wiley, New York.

Pernrose, R., (1955), "A Generalized Inverse for Matrices", Proccedings of Cambridge Philos. Society, 51, p.p. 406-413.

Penrose, R., (1956), "On Best Approximate Solutions of Linear Matrix Equations", Proc. Cambridge Philos. Soc., 52, pp. 17-19.

Theil, H., (1971), Principles of Econometrics, John Wiley, New York.

Wilkinson, J.H, and C. Reinsch (1971), Handbook for Automatic Computation, Vol II Linear Algebra, Springer-Verlag, Berlin.