Contents

Fourth International Derive TI-89/92 Conference

Liverpool John Moores University, July 12 – 15, 2000

An Application of the Moore-Penrose Inverse of a Matrix to Linear Regression

Karsten Schmidt
Faculty of Management and Economics
University of Applied Science, Schmalkalden, Germany
e-mail: kschmidt@wi.fh-schmalkalden.de

Moore-Penrose inverse of a matrix

It is well-known that the unique inverse $\mathbf{A}^{-1}$ of a matrix $\boldsymbol{A}$, which satisfies the condition

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

only exists if $\boldsymbol{A}$ is square and nonsingular, i.e. $\operatorname{rank}\left(\underset{n\times n}{\mathbf{A}}\right) = n$.

However, the also unique Moore-Penrose inverse $\underset{n\times m}{\mathbf{A}^{+}}$, which satisfies the four conditions

$$\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A} \tag{1}$$
$$\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+} \tag{2}$$
$$\left(\mathbf{A}^{+}\mathbf{A}\right)' = \mathbf{A}^{+}\mathbf{A} \tag{3}$$
$$\left(\mathbf{A}\mathbf{A}^{+}\right)' = \mathbf{A}\mathbf{A}^{+} \tag{4}$$

exists for every matrix $\underset{m\times n}{\mathbf{A}}$, regardless of its dimension and rank. A *DERIVE 5* function for the computation of the Moore-Penrose inverse is presented in the next section.

A wealth of properties holds for the Moore-Penrose inverse of which the following three will be useful later in this paper:

$$\left(\mathbf{A}'\mathbf{A}\right)^{+}\mathbf{A}' = \mathbf{A}^{+} \tag{5}$$
$$\mathbf{A}'\mathbf{A}\mathbf{A}^{+} = \mathbf{A}' \tag{6}$$
$$\operatorname{rank}\left(\underset{m\times n}{\mathbf{A}}\right) = n \quad \Leftrightarrow \quad \mathbf{A}^{+} = \left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}' \quad \Leftrightarrow \quad \mathbf{A}^{+}\mathbf{A} = \underset{n\times n}{\mathbf{I}} \tag{7}$$

The Moore-Penrose inverse (or, more generally, any generalized inverse, i.e. any matrix satisfying condition (1)) of a matrix can be used to determine the solution(s) of a system of linear equations

$$\underset{m\times n}{\mathbf{A}}\;\underset{n\times 1}{\mathbf{x}} = \underset{m\times 1}{\mathbf{b}}$$

Such a system is consistent if and only if

$$\mathbf{A}\mathbf{A}^{+}\mathbf{b} = \mathbf{b}$$

If $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent, its general solution is given by

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{b} + \left(\mathbf{I} - \mathbf{A}^{+}\mathbf{A}\right)\mathbf{z}$$

where $\underset{n\times1}{\mathbf{z}}$ is an arbitrary vector.

In a later section below we will see how the Moore-Penrose inverse can be used to show that the so-called system of normal equations, which is derived by applying the principle of least squares in linear regression, is consistent regardless of the rank of the regressor matrix. We then provide the general solution of the system of normal equations, and highlight the fact that the usual rank assumption on the regressor matrix assures a unique solution.

Computation of the Moore-Penrose inverse

In a previous paper (Schmidt 1998) the Greville algorithm for the computation of the Moore-Penrose inverse was described and implemented in *DERIVE 4* for matrices with $\min(m,n)\le 2$, i.e. vectors, and matrices which have either only two rows or only two columns. For the sake of convenience we repeat here the description of the Greville algorithm, which leads to the unique Moore-Penrose inverse in a finite number of iterations.

We start with a simple formula to calculate the Moore-Penrose inverse if $\mathbf{A}=\underset{n\times1}{\mathbf{a}}$ is a vector:

$$\mathbf{a}^{+}=\begin{cases}\frac{1}{\mathbf{a}'\mathbf{a}}\mathbf{a}' & \text{if } \mathbf{a}\neq\mathbf{0}\\ \mathbf{0}' & \text{if } \mathbf{a}=\mathbf{0}\end{cases} \tag{8}$$

We now consider the column notation of $\boldsymbol{A}$:

$$\underset{m\times n}{\mathbf{A}}=\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix}$$

and denote the submatrix, which comprises the first k columns of $\boldsymbol{A}$, by

$$\underset{m\times k}{\mathbf{A}_k}=\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_k\end{bmatrix}$$

Hence

$$\mathbf{A}_k=\begin{bmatrix}\mathbf{A}_{k-1} & \mathbf{a}_k\end{bmatrix}$$

Moreover, we define the following vectors for $j\ge 2$:

$$\mathbf{d}_j{}'=\mathbf{a}_j{}'\mathbf{A}_{j-1}^{+}{}'\mathbf{A}_{j-1}^{+}$$

$$\mathbf{c}_j=\left(\mathbf{I}-\mathbf{A}_{j-1}\mathbf{A}_{j-1}^{+}\right)\mathbf{a}_j$$

$$\mathbf{b}_j{}'=\mathbf{c}_j^{+}+\frac{1-\mathbf{c}_j^{+}\mathbf{c}_j}{1+\mathbf{d}_j{}'\mathbf{a}_j}\mathbf{d}_j{}'$$

Note that $\mathbf{d}_j{}'$ is a row vector, $\mathbf{c}_j$ a column vector (and hence $\mathbf{c}_j^{+}$ a row vector) and $\mathbf{b}_j{}'$ a row vector. Then we have

$$\mathbf{A}_j^{+}=\begin{bmatrix}\mathbf{A}_{j-1} & \mathbf{a}_j\end{bmatrix}^{+}=\begin{bmatrix}\mathbf{A}_{j-1}^{+}-\mathbf{A}_{j-1}^{+}\mathbf{a}_j\mathbf{b}_j{}'\\ \mathbf{b}_j{}'\end{bmatrix} \tag{9}$$

Since $\mathbf{A}_1 = \mathbf{a}_1$ is a matrix having only one column, its Moore-Penrose inverse is easily calculated by (8). Using (9) we can then iteratively calculate $\mathbf{A}_2^+$, $\mathbf{A}_3^+$, ..., $\mathbf{A}_n^+ = \mathbf{A}^+$.
The following *DERIVE 5* functions can be used to compute the Moore-Penrose inverse of any matrix:

```
mpi(A) :=
  Prog
    APLUS := mpiv(A COL [1])
    J := 2
    Loop
      If J > DIM(A`)
         RETURN APLUS
      aj := A COL [J]
      dt := aj`·APLUS`·APLUS
      c := (IDENTITY_MATRIX(DIM(A)) - A COL [1, ..., J - 1]·APLUS)·aj
      bt := mpiv(c) + (1 - mpiv(c)·c)/(1 + dt·aj)·dt
      APLUS := APPEND(APLUS - APLUS·aj·bt, bt)
      J :+ 1

mpiv(a) :=
  Prog
    If a`·a = 0
       RETURN 0·a`
    RETURN a`/ELEMENT(a`·a, 1, 1)
```

The function `mpiv(a)`, which is used repeatedly inside the function `mpi(A)`, returns the Moore-Penrose inverse of a (column) vector ***a*** passed as parameter. The function `mpi(A)` calculates the Moore-Penrose inverse of a matrix ***A*** passed as parameter.

Linear Regression and the Moore-Penrose inverse

We consider the (multiple) linear regression model

$$\underset{N\times 1}{\mathbf{y}} = \underset{N\times K}{\mathbf{X}} \underset{K\times 1}{\mathrm{b}} + \underset{N\times 1}{\mathbf{u}}$$

where ***y*** is the vector of observations on the dependent variable, ***X*** the regressor matrix, $\boldsymbol{\beta}$ a vector of unknown parameters, and ***u*** a vector of disturbances with

$$\mathrm{E}[\mathbf{u}] = 0; \quad \mathrm{D}[\mathbf{u}] = \sigma^2 \mathbf{I}$$

Denoting an estimator of the unknown parameter vector $\boldsymbol{\beta}$ by $\tilde{\mathrm{b}}$, we have

$$\tilde{\mathbf{y}} = \mathbf{X}\tilde{\mathrm{b}}$$

$$\tilde{\mathbf{u}} = \mathbf{y} - \tilde{\mathbf{y}}$$

The most popular estimator for $\boldsymbol{\beta}$ is the least squares estimator which minimises the sum of squared residuals

$$\varphi\left(\tilde{\mathrm{b}}\right) = \sum_{i=1}^{N} \tilde{u}_i^{\,2} = \tilde{\mathbf{u}}'\tilde{\mathbf{u}} = \left(\mathbf{y} - \mathbf{X}\tilde{\mathrm{b}}\right)' \left(\mathbf{y} - \mathbf{X}\tilde{\mathrm{b}}\right) \to \min_{\tilde{\mathrm{b}}}$$

Note that

$$
\begin{aligned}
\varphi(\tilde{b}) &= (\mathbf{y}-\mathbf{X}\tilde{b})'(\mathbf{y}-\mathbf{X}\tilde{b}) \\
&= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\tilde{b} - \tilde{b}'\mathbf{X}'\mathbf{y} + \tilde{b}'\mathbf{X}'\mathbf{X}\tilde{b} \\
&= \tilde{b}'\mathbf{X}'\mathbf{X}\tilde{b} - 2\mathbf{y}'\mathbf{X}\tilde{b} + \mathbf{y}'\mathbf{y}
\end{aligned}
$$

is a convex function since $\mathbf{X}'\mathbf{X}$ is a nonnegative definite matrix. Therefore, finding its first derivative

$$
\begin{aligned}
\frac{\partial\varphi(\tilde{b})}{\partial\tilde{b}} &= -2\mathbf{y}'\mathbf{X} + \tilde{b}'\left(\mathbf{X}'\mathbf{X} + (\mathbf{X}'\mathbf{X})'\right) \\
&= -2\mathbf{y}'\mathbf{X} + 2\tilde{b}'\mathbf{X}'\mathbf{X}
\end{aligned}
$$

and setting it equal to $\mathbf{0}$ is necessary and sufficient for determining the minimum of $\varphi(\tilde{b})$:

$$
\begin{aligned}
-2\mathbf{y}'\mathbf{X} + 2\tilde{b}'\mathbf{X}'\mathbf{X} &= \underset{1\times K}{\mathbf{0}} \quad \Leftrightarrow \\
-\mathbf{X}'\mathbf{y} + \mathbf{X}'\mathbf{X}\tilde{b} &= \underset{K\times 1}{\mathbf{0}} \quad \Leftrightarrow \\
\mathbf{X}'\mathbf{X}\tilde{b} &= \mathbf{X}'\mathbf{y}
\end{aligned}
$$

The final equation constitutes the so-called system of normal equations.
Under the assumption that $\operatorname{rank}(\mathbf{X}) = K$ we can easily derive the Least Squares estimator from the normal equations

$$
\begin{aligned}
\mathbf{X}'\mathbf{X}\tilde{b} &= \mathbf{X}'\mathbf{y} \quad \Leftrightarrow \\
\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}}_{\mathbf{I}}\tilde{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \Leftrightarrow \\
\tilde{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}
\end{aligned}
$$

One could get the impression that the system of normal equations is inconsistent if $\operatorname{rank}(\mathbf{X}) < K$. However, this is not true.
Observe that the system of normal equations is essentially a system of linear equations:

$$
\underbrace{\mathbf{X}'\mathbf{X}}_{\mathbf{A}}\underbrace{\tilde{b}}_{\mathbf{x}} = \underbrace{\mathbf{X}'\mathbf{y}}_{\mathbf{b}}
$$

Using properties (5) and (6) of the Moore-Penrose inverse, it is easily shown that the system of normal equations is consistent:

$$
\mathbf{A}\mathbf{A}^{+}\mathbf{b} = \mathbf{b} \quad \Rightarrow \quad \mathbf{X}'\mathbf{X}\underbrace{(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'}_{\mathbf{X}^{+}}\mathbf{y} = \underbrace{\mathbf{X}'\mathbf{X}\mathbf{X}^{+}}_{\mathbf{X}'}\mathbf{y} = \mathbf{X}'\mathbf{y}
$$

Hence, its general solution is given by

$$
\mathbf{x} = \mathbf{A}^{+}\mathbf{b} + (\mathbf{I} - \mathbf{A}^{+}\mathbf{A})\mathbf{z} \quad \Rightarrow \quad
\begin{aligned}
\tilde{b} &= \underbrace{(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'}_{\mathbf{X}^{+}}\mathbf{y} + \left(\mathbf{I} - \underbrace{(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'}_{\mathbf{X}^{+}}\mathbf{X}\right)\mathbf{z} \\
&= \mathbf{X}^{+}\mathbf{y} + (\mathbf{I} - \mathbf{X}^{+}\mathbf{X})\mathbf{z}
\end{aligned}
$$

where $\underset{K\times 1}{\mathbf{z}}$ is an arbitrary vector.

The number of solutions, however, depends on the rank of the regressor matrix. If $\operatorname{rank}(\mathbf{X}) = K$, it follows from (7) that $\mathbf{X}^{+}\mathbf{X} = \mathbf{I}$, and the general solution simplifies to the unique solution

$$\tilde{b} = \mathbf{X}^{+}\mathbf{y} + \left(\mathbf{I} - \underbrace{\mathbf{X}^{+}\mathbf{X}}_{\mathbf{I}}\right)\mathbf{z}$$
$$= \mathbf{X}^{+}\mathbf{y}$$

i.e. the Least Squares estimator is simply the product of the Moore-Penrose inverse of the regressor matrix and the vector of the observations on the dependent variable.
If, however, $\operatorname{rank}(\mathbf{X}) < K$, we have an infinite number of solutions.

Therefore, it is not the consistency of the system of normal equations that is guaranteed by assuming ***X*** to be of full column rank, but the uniqueness of its solution.

References

Ben-Israel, A. & T.N.E. Greville (1974), *Generalized Inverses: Theory and Applications*, New York (Wiley).

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

Moore, E.H. (1920), On the Reciprocal of the General Algebraic Matrix (Abstract), *Bulletin of the American Mathematical Society* ***26***, 394-395.

Penrose, R. (1955), A Generalized Inverse for Matrices, *Proceedings of the Cambridge Philosophical Society* ***51***, 406-413.

Rao, C.R. (1962), A Note on a Generalized Inverse of a Matrix with Applications to Problems in Mathematical Statistics, *Journal of the Royal Statistical Society B* ***24***, 152-158.

Rao, C.R. & S.K. Mitra (1971), *Generalized Inverse of Matrices and Its Applications*, New York (Wiley).

Schmidt, K. (1998), The Use of a Computer Algebra System in Modern Matrix Algebra, *Proceedings of the 3rd International DERIVE and TI-92 Conference*.

Schmidt, K. & G. Trenkler (1998), *Moderne Matrix-Algebra*, Berlin (Springer).