3.10 Model in Matrix Notation

For many purposes, including computation, it is convenient to write the model and statistics in matrix notation. The n linear equations $Y_i = X_i'\beta + e_i$ make a system of n equations. We can stack these n equations together as

$$\begin{aligned} Y_1 &= X_1'\beta + e_1 \\ Y_2 &= X_2'\beta + e_2 \\ &\vdots \\ Y_n &= X_n'\beta + e_n. \end{aligned}$$

Define

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{pmatrix}, \qquad \boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}.$$

Observe that $\boldsymbol{Y}$ and $\boldsymbol{e}$ are $n \times 1$ vectors and $\boldsymbol{X}$ is an $n \times k$ matrix. The system of n equations can be compactly written in the single equation

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}. \tag{3.19}$$

Sample sums can be written in matrix notation. For example

$$\begin{aligned} \sum_{i=1}^{n} X_i X_i' &= \boldsymbol{X}'\boldsymbol{X} \\ \sum_{i=1}^{n} X_i Y_i &= \boldsymbol{X}'\boldsymbol{Y}. \end{aligned}$$

Therefore the least squares estimator can be written as

$$\widehat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Y}\right).$$

The matrix version of (3.15) and estimated version of (3.19) is

$$\boldsymbol{Y} = \boldsymbol{X}\widehat{\beta} + \widehat{\boldsymbol{e}}.$$

Equivalently the residual vector is

$$\widehat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\beta}.$$

Using the residual vector we can write (3.16) as

$$\boldsymbol{X}'\widehat{\boldsymbol{e}} = 0.$$

It can also be useful to write the sum of squared error criterion as

$$SSE(\beta) = \left(\boldsymbol{Y} - \boldsymbol{X}\beta\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\beta\right).$$

Using matrix notation we have simple expressions for most estimators. This is particularly convenient for computer programming as most languages allow matrix notation and manipulation.

Theorem 3.2 Important Matrix Expressions

$$\widehat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Y}\right)$$
$$\widehat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\beta}$$
$$\boldsymbol{X}'\widehat{\boldsymbol{e}} = 0.$$

Early Use of Matrices

The earliest known treatment of the use of matrix methods to solve simultaneous systems is found in Chapter 8 of the Chinese text *The Nine Chapters on the Mathematical Art*, written by several generations of scholars from the 10^{th} to 2^{nd} century BCE.

3.11 Projection Matrix

Define the matrix

$$\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'.$$

Observe that

$$\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}.$$

This is a property of a **projection matrix**. More generally, for any matrix $\boldsymbol{Z}$ which can be written as $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{\Gamma}$ for some matrix $\boldsymbol{\Gamma}$ (we say that $\boldsymbol{Z}$ lies in the **range space** of $\boldsymbol{X}$), then

$$\boldsymbol{P}\boldsymbol{Z} = \boldsymbol{P}\boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{Z}.$$

As an important example, if we partition the matrix $\boldsymbol{X}$ into two matrices $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ so that $\boldsymbol{X} = [\boldsymbol{X}_1 \quad \boldsymbol{X}_2]$ then $\boldsymbol{P}\boldsymbol{X}_1 = \boldsymbol{X}_1$. (See Exercise 3.7.)

The projection matrix $\boldsymbol{P}$ has the algebraic property that it is **idempotent**: $\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}$. See Theorem 3.3.2 below. For the general properties of projection matrices see Section A.11.

The matrix $\boldsymbol{P}$ creates the fitted values in a least squares regression:

$$\boldsymbol{P}\boldsymbol{Y} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}\widehat{\beta} = \widehat{\boldsymbol{Y}}.$$

Because of this property $\boldsymbol{P}$ is also known as the **hat matrix**.

A special example of a projection matrix occurs when $\boldsymbol{X} = \mathbf{1}_n$ is an n-vector of ones. Then

$$\boldsymbol{P} = \mathbf{1}_n\left(\mathbf{1}_n'\mathbf{1}_n\right)^{-1}\mathbf{1}_n' = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'.$$

Note that in this case

$$\boldsymbol{P}\boldsymbol{Y} = \mathbf{1}_n\left(\mathbf{1}_n'\mathbf{1}_n\right)^{-1}\mathbf{1}_n'\boldsymbol{Y} = \mathbf{1}_n\overline{Y}$$

creates an n-vector whose elements are the sample mean $\overline{Y}$.

The projection matrix $\boldsymbol{P}$ appears frequently in algebraic manipulations in least squares regression. The matrix has the following important properties.

Theorem 3.3 The projection matrix $\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$ for any $n \times k$ $\boldsymbol{X}$ with $n \geq k$ has the following algebraic properties.

1. $\boldsymbol{P}$ is symmetric ($\boldsymbol{P}' = \boldsymbol{P}$).
2. $\boldsymbol{P}$ is idempotent ($\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}$).
3. $\operatorname{tr}\boldsymbol{P} = k$.
4. The eigenvalues of $\boldsymbol{P}$ are 1 and 0. There are k eigenvalues equalling 1 and $n-k$ equalling 0.
5. $\operatorname{rank}(\boldsymbol{P}) = k$.

We close this section by proving the claims in Theorem 3.3. Part 1 holds since

$$\begin{aligned}
\boldsymbol{P}' &= \left(\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right)' \\
&= \left(\boldsymbol{X}'\right)'\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)'(\boldsymbol{X})' \\
&= \boldsymbol{X}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)'\right)^{-1}\boldsymbol{X}' \\
&= \boldsymbol{X}\left((\boldsymbol{X})'\left(\boldsymbol{X}'\right)'\right)^{-1}\boldsymbol{X}' = \boldsymbol{P}.
\end{aligned}$$

To establish part 2, the fact that $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$ implies that

$$\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' = \boldsymbol{P}$$

as claimed. For part 3,

$$\operatorname{tr}\boldsymbol{P} = \operatorname{tr}\left(\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right) = \operatorname{tr}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\right) = \operatorname{tr}\left(\boldsymbol{I}_k\right) = k.$$

See Appendix A.5 for definition and properties of the trace operator.

For part 4, it is shown in Appendix A.11 that the eigenvalues λ_i of an idempotent matrix are all 1 and 0. Since $\operatorname{tr}\boldsymbol{P}$ equals the sum of the n eigenvalues and $\operatorname{tr}\boldsymbol{P} = k$ by part 3, it follows that there are k eigenvalues equalling 1 and the remainder $n-k$ equalling 0.

For part 5, observe that $\boldsymbol{P}$ is positive semi-definite since its eigenvalues are all non-negative. By Theorem A.4.5 its rank equals the number of positive eigenvalues, which is k as claimed.

3.12 Annihilator Matrix

Define

$$\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. Note that

$$\boldsymbol{M}\boldsymbol{X} = \left(\boldsymbol{I}_n - \boldsymbol{P}\right)\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{P}\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = 0. \tag{3.21}$$

Thus $\boldsymbol{M}$ and $\boldsymbol{X}$ are orthogonal. We call $\boldsymbol{M}$ the **annihilator matrix** due to the property that for any matrix $\boldsymbol{Z}$ in the range space of $\boldsymbol{X}$ then

$$\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{Z} - \boldsymbol{P}\boldsymbol{Z} = 0.$$