Econometrics I TA Session

Wang Xin*

April 14, 2022

Contents

1 Terminology

Suppose that there is a simple regression model

$$y_i = x_i\beta + u_i,$$

where i = 1,2,3,. . . , n, y_i and x_i are observations.

When we arrange those observations like

$$\begin{gathered} y_1 = x_1\beta + u_1 \\ y_2 = x_2\beta + u_2 \\ \dots \\ y_n = x_n\beta + u_n, \end{gathered}$$

*If you have any questions, please notice us on the class or contact us by e-mail: wx1184097947@yahoo.com (Wang Xin)/ zhengxuzhu_u@yahoo.co.jp (Zheng Xuzhu)

we can use vectors to rewrite the above equations into one:

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{u} \tag{1}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$, $\mathbf{u} = (u_1, u_2, \ldots, u_n)'$ are vectors.

A vector is an ordered set of numbers arranged either in a row or a column. In view of the preceding, a row vector is also a matrix with one row, whereas a column vector is a matrix with one column.

Similarly, for a multiple regression model

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \ldots + x_{ik}\beta_k + u_i,$$

where y_i and x_{ij} are observations and $i = 1, 2, 3, \ldots, n$, $j = 1, 2, 3, \ldots, k$, we arrange them in a series

$$\begin{gathered} y_1 = x_{11}\beta_1 + x_{12}\beta_2 + x_{13}\beta_3 + \ldots + x_{1k}\beta_k + u_1 \\ y_2 = x_{21}\beta_1 + x_{22}\beta_2 + x_{23}\beta_3 + \ldots + x_{2k}\beta_k + u_2 \\ \ldots \\ y_n = x_{n1}\beta_1 + x_{n2}\beta_2 + x_{n3}\beta_3 + \ldots + x_{nk}\beta_k + u_n. \end{gathered}$$

Then we extract x and β from right side

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \ldots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \ldots & x_{1k} \\ x_{21} & x_{22} & x_{23} & \ldots & x_{2k} \\ x_{31} & x_{32} & x_{33} & \ldots & x_{3k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{n1} & x_{n2} & x_{n3} & \ldots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \ldots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \ldots \\ u_n \end{pmatrix}$$

so x can be compacted into a matrix

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{2}$$

A matrix can also be viewed as a set of column vectors or as a set of row vectors. The dimensions of a matrix are the numbers of rows and columns it contains. "**X** is an n × K matrix" will always mean that **X** has n rows and K columns. If n equals K, then **X** is a **square matrix**.

- A **symmetric matrix** is one in which $x_{ij} = x_{ji}$ for all i and j.
- A **diagonal matrix** is a square matrix whose only nonzero elements appear on the main diagonal, that is, moving from upper left to lower right.
- An **identity matrix** is a scalar matrix with ones on the diagonal. This matrix is always denoted as **I**. A subscript is sometimes included to indicate its size, or order. For example, $\mathbf{I}_4$ indicates a 4 × 4 identity matrix.

2 Algebraic Manipulation of Matrices

2.1 Equality of Matrices

Matrices (or vectors) $\mathbf{A}$ and $\mathbf{B}$ are equal if and only if they have the same dimensions and each element of $\mathbf{A}$ equals the corresponding element of $\mathbf{B}$. That is, $\mathbf{A} = \mathbf{B}$ if and only if $a_{ik} = b_{ik}$ for all i and k.

2.2 Transposition

The transpose of a matrix $\mathbf{A}$, denoted $\mathbf{A}'$, is obtained by creating the matrix whose k_{th} row is the k_{th} column of the original matrix. Thus, if $\mathbf{B} = \mathbf{A}'$, then each column of $\mathbf{A}$ will appear as the corresponding row of $\mathbf{B}$. If $\mathbf{A}$ is n $\times$ K, then $\mathbf{A}'$ is K $\times$ n.

An equivalent definition of the transpose of a matrix is $\mathbf{B} = \mathbf{A}' \Longleftrightarrow b_{ik} = a_{ki}$ for all i and k. The definition of a symmetric matrix implies that

$$\text{if (and only if) } \mathbf{A} \text{ is symmetric, then } \mathbf{A} = \mathbf{A}'.$$

It also follows from the definition that for any $\mathbf{A}$,

$$(\mathbf{A}')' = \mathbf{A}.$$

Finally, the transpose of a column vector, $\mathbf{a}$, is a row vector:

$$\mathbf{a}' = (a_1 \; a_2 \; \ldots\ldots \; a_n)$$

2.3 Addition

The operations of addition and subtraction are extended to matrices by defining

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = [a_{ik} + b_{ik}] \tag{3}$$

$$\mathbf{A} - \mathbf{B} = [a_{ik} - b_{ik}]$$

Matrices cannot be added unless they have the same dimensions, in which case they are said to be conformable for addition. A zero matrix or null matrix is one whose elements are all zero. It follows from (3) that matrix addition is commutative,

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

and associative,

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}),$$

and that

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$$

2.4 Multiplication

Matrices are multiplied by using the inner product. The inner product, or dot product, of two vectors, **a** and **b**, is a scalar and is written

$$\mathbf{a}'\mathbf{b} = a_1b_1 + a_2b_2 + \quad + a_nb_n. \tag{4}$$

Note that the inner product is written as the transpose of vector **a** times vector **b**, a row vector times a column vector. In (A-12), each term a_jb_j equals b_ja_j ; hence

$$\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a}.$$

For an n × K matrix **A** and a K × M matrix **B**, the product matrix, $\mathbf{C} = \mathbf{AB}$, is an n × M matrix whose ik_{th} element is the inner product of row i of **A** and column k of **B**. Thus, the product matrix **C** is

$$\mathbf{C} = \mathbf{AB} \Rightarrow c_{ik} = \mathbf{a_i'b_k}.$$

To multiply two matrices, the number of columns in the first must be the same as the number of rows in the second, in which case they are **conformable for multiplication**.

Scalar multiplication of a matrix is the operation of multiplying every element of the matrix by a given scalar. For scalar c and matrix **A**,

$$c\mathbf{A} = [ca_{ik}].$$

The product of a matrix and a vector is written

$$\mathbf{c} = \mathbf{Ab}$$

the number of elements in **b** must equal the number of columns in **A**; the result is a vector with number of elements equal to the number of rows in **A**.

The product of a matrix and a vector is a linear combination of the columns of the matrix where the coefficients are the elements of the vector. For the general case,

$$\mathbf{c} = \mathbf{Ab} = b_1\mathbf{a_1} + b_2\mathbf{a_2} + ... + b_k\mathbf{a_k} \tag{5}$$

In the calculation of a matrix product $\mathbf{C} = \mathbf{AB}$, each column of **C** is a linear combination of the columns of **A**, where the coefficients are the elements in the corresponding column of **B**. That is,

$$\mathbf{C} = \mathbf{AB} \Longleftrightarrow \mathbf{c_k} = \mathbf{Ab_k}.$$

- Associative law: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
- Distributive law: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
- Transpose of a product: $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$
- Transpose of an extended product: $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$

3 Some Definitions about Matrices

3.1 Inverse Matrix

Let matrix $\mathbf{X}$ and $\mathbf{Y}$ be n×n matrices. If

$$\mathbf{XY} = \mathbf{YX} = \mathbf{I_n}, \tag{6}$$

then matrix $\mathbf{Y}$ is a inverse matrix of $\mathbf{X}$, which can be written as

$$\mathbf{Y} = \mathbf{X^{-1}},$$

Matrix $\mathbf{X}$ has inverse matrix if and only if the determinant of $\mathbf{X}$ is not 0.

Moreover if $\mathbf{X}$ has inverse matrix, then it is called to be regular.

If $\mathbf{A}$ and $\mathbf{B}$ are regular matrices, $\mathbf{AB}$ and $(\mathbf{AB})^{-1}$ are regular matrices and $\mathbf{B^{-1}A^{-1}}$ is equal to $(\mathbf{AB})^{-1}$

3.2 Rank

The rank of $\mathbf{X}$ is the maximum number of linearly independent column (or row) vectors of $\mathbf{X}$, which is denoted by $rank(\mathbf{X})$.

For example, $rank(A) = ?$

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & -3 & 1 \\ 1 & -1 & 2 & -1 \\ 4 & -4 & 3 & -2 \\ 2 & -2 & -11 & 4 \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 1 \\ 3 \\ 10 \\ 0 \end{pmatrix}$$

$$\mathbf{y} = \mathbf{Ax} \Rightarrow \begin{cases} 1 = x_1 - x_2 - 3x_3 + x_4 \\ 3 = x_1 - x_2 + 2x_3 - x_4 \\ 10 = 4x_1 - 4x_2 + 3x_3 - 2x_4 \\ 0 = 2x_1 - 2x_2 - 11x_3 + 4x_4 \end{cases}$$

$$\begin{cases} 1 = x_1 - x_2 - 3x_3 + x_4 \\ 2 = 5x_3 - 2x_4 \\ 6 = 15x_3 - 6x_4 \\ -2 = -5x_3 + 2x_4 \end{cases}$$

3.3 Trace

The trace of a square K × K matrix $\mathbf{A}$ is the sum of its diagonal elements:

$$tr(\mathbf{A}) = \sum_{i=1}^{k} a_{ii} \tag{7}$$

Some easily proven results are

$$tr(c\mathbf{A}) = c(tr(\mathbf{A}))$$

$$tr(\mathbf{A}') = tr(\mathbf{A})$$
$$tr(\mathbf{A}+\mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$$
$$tr(\mathbf{I}_k) = \mathbf{K}$$
$$tr(\mathbf{AB}) = tr(\mathbf{BA})$$
$$\mathbf{a}'\mathbf{a} = tr(\mathbf{a}'\mathbf{a}) = tr(\mathbf{a}\mathbf{a}')$$
$$tr(\mathbf{A}'\mathbf{A}) = \sum_{k=1}^{K} \mathbf{a}'_k \mathbf{a}_k = \sum_{i=1}^{K}\sum_{k=1}^{K} a_{ik}^2$$

The permutation rule can be extended to any cyclic permutation in a product:

$$tr(\mathbf{ABCD}) = tr(\mathbf{BCDA}) = tr(\mathbf{CDAB}) = tr(\mathbf{DABC})$$

4 Differentiation of Matrices

A linear function can be written

$$\mathbf{y} = \mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{a} = \sum_{i=1}^{n} \mathbf{a}_i \mathbf{x}_i,$$

so

$$\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}'\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} \tag{8}$$

Note, in particular, that $\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$, not $\mathbf{a}'$.

$$\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial x_1 a_1 + x_2 a_2 + \ldots + x_n a_n}{\partial x_1} \\ \frac{\partial x_1 a_1 + x_2 a_2 + \ldots + x_n a_n}{\partial x_2} \\ \ldots \\ \frac{\partial x_1 a_1 + x_2 a_2 + \ldots + x_n a_n}{\partial x_n} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \ldots \\ a_n \end{pmatrix} = \mathbf{a}$$

Similarly, for a matrix $\mathbf{A}$, $\mathbf{A} = (\mathbf{a}'_1 \mathbf{a}'_2 \ldots \mathbf{a}'_n)'$,

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}'} = \begin{pmatrix} \frac{\partial \mathbf{a}'_1\mathbf{x}}{\partial \mathbf{x}'} \\ \frac{\partial \mathbf{a}'_2\mathbf{x}}{\partial \mathbf{x}'} \\ \ldots \\ \frac{\partial \mathbf{a}'_n\mathbf{x}}{\partial \mathbf{x}'} \end{pmatrix} = \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \ldots \\ \mathbf{a}'_n \end{pmatrix} = \mathbf{A}$$

Further, a quadratic form is

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{ij}$$

For example, $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$.

$$\begin{aligned}
\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= \begin{pmatrix} \dfrac{\partial\left(ax_1^2 + bx_1x_2 + cx_1x_2 + dx_2^2\right)}{\partial x_1} \\ \dfrac{\partial\left(ax_1^2 + bx_1x_2 + cx_1x_2 + dx_2^2\right)}{\partial x_2} \end{pmatrix} \\
&= \begin{pmatrix} 2ax_1 + bx_2 + cx_2 \\ bx_1 + cx_1 + 2dx_2 \end{pmatrix} \\
&= \begin{pmatrix} 2a & b+c \\ b+c & 2d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= \left(\mathbf{A} + \mathbf{A}'\right)\mathbf{x}
\end{aligned}$$

when A is a symmetric matrix, $\dfrac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$

5 Assumptions of the Linear Regression Model

1. **Linearity**: $\mathbf{y} = \mathbf{x}\beta + \mathbf{u}$. The model specifies a linear relationship between $\mathbf{y}$ and $\mathbf{x}$.

2. **Nonstochasticity and Full Rank**: The independent variables are observable and not stochastic. Moreover, there is no exact linear relationship among any of the independent variables in the model.

3. **Exogeneity of the independent variables**: $E[u_i|x_{i1}, x_{i2}, ..., x_{ik}] = 0$. This states that the expected value of the disturbance at observation i in the sample is not a function of the independent variables observed at any observation, including this one. This means that the independent variables will not carry useful information for prediction of u_i.

4. **Homoscedasticity and nonautocorrelation**: $Var(u_i^2) = \sigma^2 < \infty$ for all i and $Cov(u_i, u_j) = 0, \forall i \neq j$. Each disturbance, u_i has the same finite variance, σ^2, and is uncorrelated with every other disturbance, u_j . This assumption limits the generality of the model, and we will want to examine how to relax it in the chapters to follow.

5. **Normal distribution**: The disturbances are normally distributed. Once again, this is a convenience that we will dispense with after some analysis of its implications.