S.F. Ellison

Review of Matrix Notation & Operations

Matrix: A is $K \times L$ (K rows, L columns) matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1L} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{K1} & \cdots & & a_{KL} \end{bmatrix} = [a_{ij}] \quad \begin{matrix} i=1,\dots,K \\ j=1,\dots,L \end{matrix}$$

(Column) Vector: Y is a K- (column) vector if it's a $K\times 1$ matrix

Matrix equality:

$A=B$ means A & B are both $K\times L$ matrices, and

$$a_{ij} = b_{ij} \quad \forall \begin{matrix} i=1,\dots,K \\ j=1,\dots,L \end{matrix} \quad \begin{matrix} A=[a_{ij}], \\ B=[b_{ij}] \end{matrix}$$

Matrix addition:

If both $A=[a_{ij}]$ and $B=[b_{ij}]$ are $K\times L$ then

$$A+B = [a_{ij} + b_{ij}]$$

Scalar multiplication:

If $\alpha \in \mathbb{R}$ then

$$\alpha A = [\alpha a_{ij}]$$

Matrix subtraction:

If both $A = [a_{ij}]$ and $B = [b_{ij}]$ are $K \times L$ then

$$A - B = A + (-1)B = [a_{ij} - b_{ij}]$$

Matrix multiplication:

If $A = [a_{ij}]$ is $K \times L$ and $B = [b_{ij}]$ is $L \times M$ then

$$AB \equiv C = [c_{ik}]$$

where

$$c_{ik} = \sum_{j=1}^{L} a_{ij} b_{jk}$$

e.g. $(X_{0i}, \dots, X_{Ki}) \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_K \end{pmatrix} = \sum_{j=0}^{K} \beta_j X_{ji}$

Transposition:

If $A = [a_{ij}]$ is $K \times L$ then

A' (A^T) = $[a_{ji}]$ is an $L \times K$ matrix

Symmetry:

A is symmetric if $A = A^T$ ($\Rightarrow K = L$)

e.g. $X = \begin{bmatrix} X_{11} & \cdots & X_{1k} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{nk} \end{bmatrix}$

X is $n \times k$ $\Rightarrow$ X^T is $k \times n$

$\Rightarrow$ $X^T X$ is $k \times k$

$$X^T X = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & & & \vdots \\ \vdots & & & \\ X_{1k} & \cdots & & X_{nk} \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & & & \vdots \\ \vdots & & & \\ X_{n1} & \cdots & & X_{nk} \end{bmatrix}$$

$$= \begin{bmatrix} \sum^n X_{i1}^2 & \sum X_{i1} X_{i2} & \cdots & \sum X_{i1} X_{ik} \\ \sum X_{i2} X_{i1} & \sum X_{i2}^2 & & \vdots \\ \vdots & & & \\ \sum X_{ik} X_{i1} & \cdots & & \sum X_{ik}^2 \end{bmatrix}$$

Matrix multiplication & transposes:

If $C = AB$ exists then

$$C^T = (AB)^T = B^T A^T$$

Identity matrix:

$$I_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \equiv [I_{[i=j]}]$$

If A is $K \times K$ then

$$I_K A = A I_K = A$$

If A is $K \times L$ then

$$I_K A = A I_L = A$$

Linear dependence / independence :

The vectors $v_1, \dots, v_k$ are linearly dependent if $\exists$ a choice of scalars $c_1, \dots, c_k$ other than $c_1 = c_2 = \dots = c_k = 0$ $\ni$

$$\sum_{i=1}^{k} c_i v_i = 0$$

If there exists no such choice, $v_1, \dots, v_k$ are linearly independent

Rank :

An $L \times K$ matrix, $L > K$, composed of K linearly independent vectors has rank K

Matrix Inversion :

If A is a square matrix $(K \times K)$ then its inverse exists if $\exists\ A^{-1} \ni$

$$A^{-1} A = A A^{-1} = I_K$$

Inverse exists $\Leftrightarrow$ rank$(A) = K$ $\Leftrightarrow$ rows (columns) of A linearly indep.

Matrix multiplication & inversion:

If A^{-1} & B^{-1} exist then

$$(BA)^{-1} = A^{-1}B^{-1}$$

Matrix inversion & symmetry:

A symmetric $(A = A^T)$. If A^{-1} exists then $(A^{-1})^T = A^{-1}$

Notes on Notation:

i) Conflict between statistics & linear algebra conventions

statistics: upper case is random variable
lower case is realization

linear algebra: upper case is matrix
lower case is vector or element

ii) Subscripts

linear algebra: 1st subscript is row
2nd subscript is column

econometrics: 1st subscript is variable
2nd subscript is observation

but we array them in a matrix with observations down a column & variables across a row

Multiple Regression in Matrix Form

$$Y_i = \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i \qquad i = 1, \dots, n$$

Matrix notation: $\underset{1\times k}{X_i} = (X_{1i}, \dots, X_{ki})$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

so $Y_i = X_i \beta + \varepsilon_i \qquad i = 1, \dots, n$

Go further: $\underset{n\times 1}{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

$$\underset{n\times k}{X} = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ \vdots & & & \vdots \\ X_{1n} & \cdots & & X_{kn} \end{bmatrix}$$

so $\underset{n\times 1}{Y} = \underset{(n\times k)(k\times 1)}{X\beta} + \underset{n\times 1}{\varepsilon}$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ \vdots & & & \vdots \\ X_{1n} & & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Assumptions in Matrix Form

a) true model $Y = X\beta + \varepsilon$

b) X $N \times k$ matrix of regressors

i) $N > k$

ii) X has full column rank k

$\Leftrightarrow$ regressors are linearly indep

$\Leftrightarrow$ X^TX is invertible

c) error assumptions

i) $E(\varepsilon) = 0$, $E(\varepsilon\varepsilon^T)$ $(= Cov(\varepsilon)) = \sigma^2 I_n$

ii) $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$

$$\underbrace{E(\varepsilon\varepsilon^T)}_{N \times N} = \begin{bmatrix} E(\varepsilon_1\varepsilon_1) & \cdots & E(\varepsilon_1\varepsilon_N) \\ \vdots & \ddots & \vdots \\ E(\varepsilon_N\varepsilon_1) & \cdots & E(\varepsilon_N\varepsilon_N) \end{bmatrix}$$

$$= \begin{bmatrix} Var(\varepsilon_1) & \cdots & Cov(\varepsilon_1, \varepsilon_N) \\ \vdots & \ddots & \vdots \\ Cov(\varepsilon_1, \varepsilon_N) & \cdots & Var(\varepsilon_N) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{bmatrix} = \sigma^2 I_n$$

Least Squares in Matrix Form

$\hat{\beta}$ minimizes over $\tilde{\beta}$ $\quad S(\tilde{\beta}) = \underbrace{\tilde{\varepsilon}^T \tilde{\varepsilon}}_{1\times 1} \quad \left(= \sum_{i=1}^{N} \tilde{\varepsilon}_i^2 \right)$

where $\tilde{\varepsilon} = Y - X\tilde{\beta}$

normal equations (get these by differentiating $S(\tilde{\beta}) = \tilde{\varepsilon}^T\tilde{\varepsilon} = (Y - X\tilde{\beta})^T (Y - X\tilde{\beta})$ with respect to $\tilde{\beta}$ and setting equal to 0)

$$-2X^T(Y - X\hat{\beta}) = 0$$

$$-2 \underbrace{\begin{bmatrix} X_{11} & \cdots & X_{1N} \\ \vdots & & \vdots \\ X_{K1} & \cdots & X_{KN} \end{bmatrix}}_{K\times N} \underbrace{\begin{bmatrix} Y_1 - X_{\cdot 1}\hat{\beta} \\ Y_2 - X_{\cdot 2}\hat{\beta} \\ \vdots \\ Y_N - X_{\cdot N}\hat{\beta} \end{bmatrix}}_{N\times 1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\left.\begin{aligned} -2 \sum_i X_{1i}(Y_i - X_{\cdot i}\hat{\beta}) &= 0 \\ &\vdots \\ -2 \sum_i X_{Ki}(Y_i - X_{\cdot i}\hat{\beta}) &= 0 \end{aligned}\right\} \text{K of them}$$

Solution of normal equations

$$-2X^T(Y - X\hat{\beta}) = 0$$

$$X^T(Y - X\hat{\beta}) = 0$$

$$X^TY = X^TX\hat{\beta}$$

if X^TX invertible,

$$\hat{\beta} = (X^TX)^{-1}X^TY$$

e.g. $k=1$ $\quad X = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ $\quad Y = \underbrace{\beta}_{=\mu_Y} + \varepsilon$

$$X^TX = [1, \dots, 1]\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = N$$

so $(X^TX)^{-1} = \frac{1}{N}$

$$X^TY = [1, \dots, 1]\begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} = \sum_i Y_i$$

so $\hat{\beta} = (X^TX)^{-1}X^TY = \frac{1}{N}\sum_i Y_i = \overline{Y}$

expectation of $\hat{\beta}$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$= (X^T X)^{-1} X^T (X\beta + \varepsilon)$$

$$= \underbrace{(X^T X)^{-1} X^T X \beta}_{=\beta} + (X^T X)^{-1} X^T \varepsilon$$

$$E(\hat{\beta}) = \beta + E((X^T X)^{-1} X^T \varepsilon)$$

$$= \beta + (X^T X)^{-1} X^T \underbrace{E(\varepsilon)}_{=0}$$

$$= \beta$$

variance-covariance matrix of $\hat{\beta}$

$$\underbrace{\mathrm{Cov}(\underbrace{\hat{\beta}}_{K\times 1})}_{K\times K} = \begin{bmatrix} \mathrm{Var}(\hat{\beta}_1) & \mathrm{Cov}(\hat{\beta}_1,\hat{\beta}_2) & \cdots & \mathrm{Cov}(\hat{\beta}_1,\hat{\beta}_K) \\ \vdots & \ddots & & \vdots \\ \mathrm{Cov}(\hat{\beta}_K,\hat{\beta}_1) & \cdots & & \mathrm{Var}(\hat{\beta}_K) \end{bmatrix}$$

$$= E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T\right]$$

recall $\hat{\beta}-\beta = (X^TX)^{-1}X^T\varepsilon$

so

$$= E\left[(X^TX)^{-1}X^T\varepsilon\varepsilon^T X(X^TX)^{-1}\right]$$

$$\left\{ \begin{array}{c} \left[(X^TX)^{-1}X^T\varepsilon\right]^T \\ = \\ \varepsilon^T X\left[(X^TX)^{-1}\right]^T \\ = \\ \varepsilon^T X(X^TX)^{-1} \end{array} \right\}$$

$$= (X^TX)^{-1}X^T E(\varepsilon\varepsilon^T)X(X^TX)^{-1}$$

$$= (X^TX)^{-1}X^T \sigma^2 I X(X^TX)^{-1}$$

$$= \sigma^2 (X^TX)^{-1}X^TX(X^TX)^{-1}$$

$$= \sigma^2 (X^TX)^{-1}$$

$$\left(= \frac{\sigma^2}{N}\left[\frac{1}{N}X^TX\right]^{-1}\right)$$

estimation of σ^2

$SSE = \hat{\varepsilon}^T\hat{\varepsilon}$

biased: $\hat{\sigma}_b^2 = \dfrac{\hat{\varepsilon}^T\hat{\varepsilon}}{N}$

unbiased: $\hat{\sigma}^2 = s^2 = \dfrac{\hat{\varepsilon}^T\hat{\varepsilon}}{N-k}$

(if $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_N)$ then

$\hat{\varepsilon}^T\hat{\varepsilon} \sim \sigma^2 \chi^2_{N-K}$ and

$\hat{\varepsilon}^T\hat{\varepsilon}$ indep of

$\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X^T X)^{-1})$)