Lecture Notes 2

1. Review of Matrices

The ability to manipulate matrices is critical in economics.

1. Matrix – a rectangular array of numbers, parameters, or variables placed in rows and columns.
   Matrices are associated with linear equations.
   Elements a_{ij} denotes the element in row i and column j.
   column vector – one column of elements
   row vector – one row of elements

   Examples

$$A = \begin{bmatrix} 9 & 8 & 7 \\ 4 & 5 & 6 \end{bmatrix} \quad B = \begin{bmatrix} 3 \\ 2 \\ 6 \\ 8 \end{bmatrix} \quad C = \begin{bmatrix} -1 & 2 & -3 & -4 \end{bmatrix} \quad D = \begin{bmatrix} 3 & 2 & 0 & 5 \\ 5 & 3 & 4 & 7 \\ 6 & 7 & 3 & 8 \\ 1 & 7 & 1 & 5 \end{bmatrix}$$

   (a) What is the dimension of each matrix? Write as $G_{m,n}$.

   (b) Which matrix is a square matrix? Column vector? Row vector?

   (c) What is the value for the following elements $a_{2,3}$, $b_{1,1}$, $c_{2,1}$, and $d_{3,3}$?

2. Multiplication by a scalar – multiplying every element of a matrix by the scalar
   Scalar is a number

   If w = 3, then wA = ?

3. Two matrices can be added or subtracted only if they have the same dimensions.

The commutative law of addition hold for matrices, H + K = K + H

$$A = \begin{bmatrix} 9 & 8 & 7 \\ 4 & 5 & 6 \end{bmatrix} \quad E = \begin{bmatrix} 2 & 21 & 1 \\ 4 & 10 & 4 \end{bmatrix}$$

Find A – 2E?

4. matrix multiplication – requires a conformability condition.

$$H = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 8 & 7 \end{bmatrix} \quad and \quad R = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

We want to multiply S = H R

Check the dimensions for $H_{3,2}$ and $R_{2,1}$.

The number of columns in H equals the number of rows in R

Therefore, the matrices are conformable for multiplication.

The dimension of $S_{3,1}$.

$$S = H \cdot R = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 8 & 7 \end{bmatrix} \times \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 4 \cdot 3 \\ 2 \cdot 5 + 6 \cdot 3 \\ 8 \cdot 5 + 7 \cdot 3 \end{bmatrix} = \begin{bmatrix} 17 \\ 28 \\ 61 \end{bmatrix}$$

Note – The commutative law of multiplication never holds for matrix multiplication, $A \cdot B \neq B \cdot A$

The distributive law holds for matrix multiplication,

A(B+C)=AB+AC or (B+C)A=BA+BC?

5. An **identity** matrix (usually denoted by I) is a square matrix with ones in its principle diagonal (the diagonal running northwest to southeast) and zeros everywhere else.

Write identity matrices of dimensions: 3 x 3, 5 x 5, and 3 x 6.

Any matrix multiplied by the identity matrix gets that same matrix again

$$H \cdot I = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 8 & 7 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1\cdot 1+4\cdot 0 & 1\cdot 0+4\cdot 1 \\ 2\cdot 1+6\cdot 0 & 2\cdot 0+6\cdot 1 \\ 8\cdot 1+7\cdot 0 & 8\cdot 0+7\cdot 1 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 8 & 7 \end{bmatrix}$$

6. Transpose – interchange the rows and columns of a matrix.
   Denoted by a prime symbol, ′, or a superscript T
   Excel =transpose(…)

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

Find A^T

$$A^T = \begin{bmatrix} 9 & 8 & 7 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 9 & 4 \\ 8 & 5 \\ 7 & 6 \end{bmatrix}$$

7. **Inverse** – denoted by A^{-1} and has to be a square matrix
   A **singular** matrix – one or more rows are a linearly combination of another row.
   Or one or more columns are a linearly combination of another column
   Do not need them – textbook has definitions for determinant, cofactor matrix, and adjoint matrix
   We can use Excel to calculate an inverse, =minverse(?,?)
   However, this returns a scalar, you have to add a index command, so
   =index(minverse(?,?), i, j)
   The $ sign command is handy here

Example

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{Thus, } A^{-1} = \begin{bmatrix} -0.3333 & 0.6666 \\ 0.6666 & -0.3333 \end{bmatrix}$$

$$A \cdot A^{-1} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -0.3333 & 0.6666 \\ 0.6666 & -0.3333 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

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

8. Use matrices to write a system of equations

$$\begin{aligned} -2x + 2y + 4z &= 9 \\ 5x + y + 3z &= 12 \\ 7x - 2y - 5z &= 2. \end{aligned} \qquad \begin{bmatrix} -2 & 2 & 4 \\ 5 & 1 & 3 \\ 7 & -2 & -5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 9 \\ 12 \\ 2 \end{bmatrix}$$

The solution is: $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -2 & 2 & 4 \\ 5 & 1 & 3 \\ 7 & -2 & -5 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 12 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.68 \\ 11.36 \\ -2.59 \end{bmatrix}$

What about this system of equations?

$$\begin{aligned} -2x + 2y &= 9 \\ -x + y &= 2. \end{aligned}$$

A unique inverse does not exist for this system, because the second row is a linear combination of the first row

2. Multiple Linear Regression

A regression with two variables is not useful
Many economic relationships have several variables

We have N-paired observations

Y is the dependent variable and it is paired with multiple, k, independent variables

Example

(1)

y_i	x_1	x_3	$\cdots$	x_k
99	10	-1	$\cdots$	50
100	15	2	$\cdots$	25
110	25	-2	$\cdots$	75

Equation is

(2) $$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \cdots + \beta_k x_{tk} + u_t$$

Linear Equation

OLS is used to estimate linear equations.

Equations must be linear in the parameters.

We can write them as three equations are:

(3)
$$\begin{aligned} 99 &= \beta_1 + \beta_2 10 + \beta_3(-1) + \cdots + \beta_k 50 + u_1 \\ 100 &= \beta_1 + \beta_2 15 + \beta_3 2 + \cdots + \beta_k 25 + u_2 \\ 110 &= \beta_1 + \beta_2 25 + \beta_3(-2) + \cdots + \beta_k 75 + u_3 \end{aligned}$$

where the β's are unknown parameters

u's are the error or residual terms.

Each equation has error term

x variables begins with two, because of the intercept

(4)
$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_k x_{1k} + u_1 \\ y_2 &= \beta_1 + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_k x_{2k} + u_2 \\ y_3 &= \beta_1 + \beta_2 x_{32} + \beta_3 x_{33} + \cdots + \beta_k x_{3k} + u_3 \ . \\ &\vdots \qquad\qquad \vdots \\ y_n &= \beta_1 + \beta_2 x_{n2} + \beta_3 x_{n3} + \cdots + \beta_k x_{nk} + u_n \end{aligned}$$

Note – in regression, n > k

If n = k, then we have n equations and k variables

If equations are independent, then we can solve for a unique solution

Note – we use matrices to write these equations

Equations written in matrix form

(5)
$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & x_{13} & x_{14} & \cdots & x_{1k} \\ 1 & x_{22} & x_{23} & x_{24} & \cdots & x_{2k} \\ 1 & x_{32} & x_{33} & x_{34} & \cdots & x_{3k} \\ & & \vdots & & & \\ 1 & x_{n2} & x_{n3} & x_{n4} & \cdots & x_{nk} \end{bmatrix} . \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix}$$

Y and U are column vectors of dimension n x 1

β is column vector of dimension k x 1

X is a matrix of dimension n x k.

Equation is re-written using

(6)
$$\underset{(n\,x\,1)}{Y} = \underset{(n\,x\,k)}{X}\ \underset{(k\,x\,1)}{\beta} + \underset{(n\,x\,1)}{U}.$$

1) each row corresponds to an individual observation,

2) the column of ones in the X matrix represent the intercept term

Ordinary Least Squares (OLS)

We solve for the residual,

(7)
$$\hat{u}_i = y_i - (\hat{\beta}_1 + \beta_2 x_{i2} + \hat{\beta}_3 x_{i3} + \cdots + \hat{\beta}_k x_{ik})$$

where the hat denotes an estimated value for the parameter.

However, we use matrices

(8) $$\hat{U} = Y - \hat{Y} = Y - X\hat{\beta}$$

Residuals are a column vector

The objective of the OLS estimator is to minimize the sum of the squared errors.

(9) $$\underset{(1x1)}{SSR} = \sum_{i=1}^{n} \hat{u}_i^2 = \underset{(1xn)}{\hat{U}^T} \underset{(nx1)}{\hat{U}}$$

SSR is a scalar, or a number

OLS is written as

(10) $$\min_{w.r.t\ \hat{\beta}} \hat{U}^T\hat{U} = (Y - X\hat{\beta})^T (Y - X\hat{\beta}).$$

Use matrix algebra to reduce equation
Apply the transpose to all terms within parenthesis

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

Multiply matrices

$$\begin{aligned}\min \hat{U}^T\hat{U} &= Y^TY - Y^TX\hat{\beta} - \hat{\beta}^TX^TY + \hat{\beta}^TX^TX\hat{\beta} \\ &= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}\end{aligned}$$

$$\min \hat{U}^T\hat{U} = Y^TY - Y^TX\hat{\beta} - \hat{\beta}^TX^TY + \hat{\beta}^TX^TX\hat{\beta}$$

(11) $$\min \hat{U}^T\hat{U} = Y^TY - 2\hat{\beta}^TX^TY + \hat{\beta}^TX^TX\hat{\beta}$$

The last step relies on the $\hat{\beta}^T X^T Y = \left(Y^T X\hat{\beta}\right)^T = Y^T X\hat{\beta}$.
Note - this matrix multiplication results in a scalar.

Also note the middle term in (11) is linear and the last term is quadratic

Derivation of solution

1. $\min \hat{U}^T\hat{U} = Y^TY - 2\hat{\beta}^T X^T Y + \hat{\beta}^T X^T X\hat{\beta}$ Original problem to min. SSR

2. $\partial\hat{U}^T\hat{U}/\partial\beta = \partial(Y^TY - 2\hat{\beta}^T X^T Y + \hat{\beta}^T X^T X\hat{\beta})/\partial\beta$ Take partial derivative

3. $\partial\hat{U}^T\hat{U}/\partial\beta = 0 - 2X^TY + 2X^TX\hat{\beta} = 0$ Set partial to zero

4. $X^TX\hat{\beta} = X^TY$ Divide both sides by 2

5. $(X^TX)^{-1}(X^TX)\hat{\beta} = (X^TX)^{-1}X^TY$ Multiply by inverse

6. $\hat{\beta} = (X'X)^{-1}X'Y$ Multiply by inverse

Second order condition

Take second partial with respect to b

Involves a matrix

Don't worry, we are indeed minimizing the squared residuals

3. Algebraic Properties of the OLS Estimator

Algebraic Property 1. The sum of the estimated residuals (error terms) is equal to zero:

$$\sum_{i=1}^{n}\hat{u}_i = 0.$$

Thus, the mean equals zero

Algebraic Property 2. The point $(\overline{Y}, \overline{X})$ is always on the estimated regression.

Algebraic Property 3. The sample covariance between each individual x_i and the OLS residual $\hat{u}_i$ is equal to zero.

Algebraic Property 4. The mean of the variable, y, will equal the mean of the $\hat{y}$.

4. Five Assumptions of the OLS Estimator

OLS estimator has five assumptions

Assumption A - Linear in Parameters

Linear in parameters

Examples – Linear in parameters

$$y = \beta_1 + \beta_2 \frac{1}{x}$$
$$y = \beta_1 + \beta_2 \log(x)$$
$$y = \beta_1 + \beta_2 x^2$$

Examples – Not linear in parameters.

$$y = \beta_1 + \beta_2^2 x$$
$$y = \beta_1 + (\beta_2 / \beta_3) \log(x)$$
$$y = \beta_1 \beta_2 x^2$$

Note – equation can be nonlinear in the x's.
Note – there is nonlinear least squares

Assumption B - Random Sample of n Observations

1. *The sample consists of n-paired observations that are drawn randomly from the population.*

Y is a dependent variable, and x's are independent variables, x's or $\{y_i : x_{2i}, x_{3i}, \cdots x_{ki}\}$

2. *The number of observations is greater than the number of parameters to be estimated, usually written n > k.*

If n = k, the number of observations (equations) will equal the number of unknowns.

3. *The independent variables (x's) are nonstochastic, whose values are fixed.*

Assumption C – Zero Conditional Mean

There is no relationship between the error terms and the independent variables or $E(U \mid X) = 0$.

Assumption D – No Perfect Multicollinearity

Perfect Multiollinearity – *there is an exact linear relationship among the independent variables*.

Cannot take the inverse of the matrix, $X^T X$

Correlations – could be used to determine presence of multicollinearity

If two variables are highly correlated

Multicollinearity could still be present is more than two variables are involved

Assumption E - Homoskedasticity

The error terms all have the same variance and are not correlated with each other.

Use the term independent and identically distributed (iid)

$$\text{var}(u_i \mid X) = \sigma^2 \quad \textit{and}$$
$$\text{cov}(u_i u_j \mid X) = 0 \quad \textit{for } i \neq j$$

where var represents the variance, cov the covariance,

$$\text{var}(u_i) = \sigma^2 \quad \textit{and}$$
$$\text{cov}(u_i u_j) = 0 \quad \textit{for } i \neq j$$

We can say $u_i \sim iid\left(o, \sigma^2\right)$

5. Properties of OLS

These properties relate to the assumptions of OLS

1. OLS is an Unbiased Estimator, β

$$E(\hat{\beta}) = \beta \text{ .}$$

By adding the two assumptions B-3 and C

Proof

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad \text{and} \quad Y = X\beta + U$$

Substitute Y into the estimator:

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

Using properties of matrices,

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T (X\beta + U) = (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T U \\ &= I\beta + (X^T X)^{-1} X^T U \\ &= \beta + (X^T X)^{-1} X^T U\end{aligned}.$$

Taking the expectation of both sides of the equation, and expected value of the error is zero, then

$$\begin{aligned}E(\hat{\beta}) &= E[\beta + (X^T X)^{-1} X^T U] \\ &= E(\beta) + E[(X^T X)^{-1} X^T U] \\ &= \beta + (X^T X)^{-1} X^T E(U) \\ &= \beta\end{aligned}.$$

2. Gauss-Markov Theorem – OLS is one of the strongest and most used estimators for unknown parameters. The Gauss-Markov Theorem is

Given the assumptions A – E, the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

Uses all five assumptions, A – E

Best – means the estimator gives the lowest variance for the parameter estimates

Linear – the estimator is linear

Unbiased – the estimator is unbised

If the error terms are distributed normally, U ~ N(0, Φ^2) or U ~ N(0, $\sigma^2\mathbf{I}$), then the OLS estimator is the Best Unbiased Estimator (BUE).

Note - Gauss-Markov Theorem do not imply that OLS has the minimum variance among all potential estimators.

Biased estimators may have smaller variance

In forecasting, it is how well the estimator predicts

3. Unbiased Estimator of Φ - estimate the variances of the betas

The simple formula for calculating the variance of a random number is:

$$\text{var}(d) = E(d - E(d))^2$$

where E is the expectation operator.
Statistics reduces equation to:

$$\text{var}(d) = \frac{1}{(n-1)} \sum_i (d_i - \bar{d})^2$$

where n-1 is the degrees and freedom
Lose one piece of information
$\bar{d}$ is the mean of the observations

Applying this concept to the residuals

$$\text{var}(U) = E(U - E(U))^2 = E(U)^2 .$$

We do not know the true variance, thus we estimate it:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-k} = \frac{(\hat{U}^T \hat{U})}{n-k} = \frac{SSE}{n-k}$$

OLS had k parameters, thus we lose k pieces of information to estimate the k variances

Variance / Covariance Matrix for $\hat{\beta}$

$$\text{cov}(w,z) = E[(w - E[w])(z - E[z])] = \frac{\sum [(w - E[w])(z - E[z])]}{n-1}$$

\

The variance / covariance matrix is:

$$V(\hat{\beta}) = \begin{bmatrix} \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1\hat{\beta}_2) & \text{cov}(\hat{\beta}_1\hat{\beta}_2) & \cdots & \text{cov}(\hat{\beta}_1\hat{\beta}_k) \\ \text{cov}(\hat{\beta}_2\hat{\beta}_1) & \text{var}(\hat{\beta}_2) & \text{cov}(\hat{\beta}_2\hat{\beta}_3) & \cdots & \text{cov}(\hat{\beta}_2\hat{\beta}_k) \\ \text{cov}(\hat{\beta}_3\hat{\beta}_1) & \text{cov}(\hat{\beta}_3\hat{\beta}_2) & \text{var}(\hat{\beta}_3) & \cdots & \text{cov}(\hat{\beta}_3\hat{\beta}_k) \\ & & \vdots & & \\ \text{cov}(\hat{\beta}_k\hat{\beta}_1) & \text{cov}(\hat{\beta}_k\hat{\beta}_2) & \text{cov}(\hat{\beta}_k\hat{\beta}_3) & \cdots & \text{var}(\hat{\beta}_k) \end{bmatrix}$$

where var denotes variance and cov denotes covariance
Square, symmetric matrix with dimensions k x k
The diagonal is the variances

The variance tells how the estimated parameters vary and are given by:

$$\underset{k\,x\,k}{V(\hat{\beta})} = \underset{1\,x\,1}{\hat{\sigma}^2} \quad \underset{k\,x\,n \quad n\,x\,k}{(X^T X)^{-1}}$$

Derivation of the Variance / Covariance Matrix.

Start with what we know:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad \text{and} \quad Y = X\beta + U$$

Substitute one equation into another

$$\hat{\beta} = (X^T X)^{-1} X^T \left(X\beta + U \right) = \beta + (X^T X)^{-1} X^T U$$

$$\hat{\beta} - \beta = (X^T X)^{-1} X^T U \,.$$

Use the matrix transpose:

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

Substituting these results into the covariance equation we obtain:

$$
\begin{aligned}
\mathrm{cov}(\hat{\beta}) &= E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T] \\
&= E[((X^T X)^{-1} X^T U)\ (U^T X (X^T X)^{-1})] \\
&= (X^T X)^{-1} X'^T E\left[UU'^T\right] X (X^T X)^{-1} \\
&= (X^T X)^{-1} X^T\ \sigma^2 I\ X (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1} I \\
&= \sigma^2 (X^T X)^{-1}
\end{aligned}
$$

Note – Assuming each error is homoskedasticity. Thus, each error term has the same variance

We estimate the variance σ^2 by using estimator, $\hat{\sigma}^2$