[2025-1] 계량경제학 - Midterm

Jiho KWAK

June 7, 2025

Chapter 1 Linear Algebra

1.1 Matrix

- square matrix: the number of rows and columns is the same
- diagonal matrkx: all elements except the diagonal elements are 0
- triangular matrix: all sub/upper-diagonal elements are 0
- zero matrix: all elements are 0
- identity matrix: a diagonal matrix with all its diagonal elements equal to 1
- transition probability matrix: a square matrix representing the transition probabilities of a stochastic system with the following properties
  - $0 \leq a_{ij} \leq 1$: elements are probabilities
  - $\sum_{j=1}^{m} a_{ij} = 1$: the sum of the elements in each row is 1

1.2 Basic Operations

- transpose
  - reflexive: $(\mathbf{A}')' = \mathbf{A}$
  - transpose of partitioned matrix: transpose each submatrix
- trace: sum of the diagonal elements
  - $tr(\mathbf{A}) = tr(\mathbf{A}')$
  - $tr(\mathbf{AB}) = tr(\mathbf{BA})$
  - $tr(\mathbf{A}'\mathbf{B}) = tr(\mathbf{B}'\mathbf{A}) = tr(\mathbf{AB}') = tr(\mathbf{BA}')$
- addition/subtraction
- equality, null matrix
- multiplication
  - inner product: given two column vectors $\mathbf{a}, \mathbf{b}$, $\mathbf{a} \cdot \mathbf{b} = \mathbf{a}'\mathbf{b} = \sum_{i=1}^{m} a_i b_i$

- given $p \times q$ matrix $\mathbf{A}$,

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pq} \end{bmatrix} = (\mathbf{A}')' = \begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_p' \end{bmatrix} \text{ since } \mathbf{A}' = \begin{bmatrix} a_{11} & \cdots & a_{p1} \\ \vdots & \ddots & \vdots \\ a_{1q} & \cdots & a_{pq} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_p \end{bmatrix}$$

  and $q \times r$ matrix $\mathbf{B}$,

$$\mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1r} \\ \vdots & \ddots & \vdots \\ b_{q1} & \cdots & b_{qr} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_r \end{bmatrix}$$

$$\mathbf{A}\mathbf{b}_i = \begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_p' \end{bmatrix} b_i = \begin{bmatrix} \mathbf{a}_1'\mathbf{b}_i \\ \vdots \\ \mathbf{a}_p'\mathbf{b}_i \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{b}_i \\ \vdots \\ \mathbf{a}_p \cdot \mathbf{b}_i \end{bmatrix} \text{ and } \mathbf{a}_j\mathbf{B} = \mathbf{a}_j' \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_r \end{bmatrix} = \begin{bmatrix} \mathbf{a}_j'\mathbf{b}_1 & \cdots & \mathbf{a}_j'\mathbf{b}_r \end{bmatrix}$$

  - (row vector) $\times$ (column vector) = scalar (inner product)
  - (column vector) $\times$ (row vector) = matrix (outer product)
  - (matrix) $\times$ (column vector) = column vector
  - (row vector) $\times$ (matrix) = row vector
- inequality: $\mathbf{A} \geq \mathbf{B} \Leftrightarrow a_{ij} \geq b_{ij}$
- power of a matrix
- Hadamard product: $\mathbf{A} \cdot \mathbf{B} = a_{ij}b_{ij}$
- Cartesian product: $\mathbf{A} \otimes \mathbf{B} = a_{ij}\mathbf{B}$

1.3 Symmetric Matrix

- basis: a set of vectors(B) in a vector space V is called a basis if every element of V may be written in a unique way as a finite linear combination of element of B
- standard basis: the set of vectors, each of whose components are all zeros except one that equals 1
  - $\mathbf{I}_n = \sum_{i=1}^{n} \mathbf{e}_i\mathbf{e}_i'$
- orthogonal matrix: square matrix whose columns and rows are orthonormal vectors
  - $\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$
  - $\mathbf{A}' = \mathbf{A}^{-1}$
  - the determinant of any orthogonal matrix is either 1 or -1
  - as a linear transformation, an orthogonal matrix preserves the inner product of vectos
    $\Rightarrow$ act as an isometry of Euclidean space (e.g. rotation, reflection)
  - the norm of vector is $||\mathbf{x}|| = \mathbf{x}'\mathbf{x} = \sqrt{\sum_{i=1}^{n} x_i^2}$
  - when $||\mathbf{x}|| = 1$, we can call it as a normal vector
- symmetric matrix
  - symmetric $\Leftrightarrow \mathbf{A} = \mathbf{A}'$

- skew-symmetric $\Leftrightarrow -\mathbf{B} = \mathbf{B}'$
- given real matrices $\mathbf{P}, \mathbf{Q}, \mathbf{X}$, $\mathbf{PXX}' = \mathbf{QXX}' \Rightarrow \mathbf{PX} = \mathbf{QX}$

  Proof.

$$\begin{aligned}(\mathbf{PXX}' - \mathbf{QXX}')(\mathbf{P}' - \mathbf{Q}') &= (\mathbf{PX} - \mathbf{QX})\mathbf{X}'(\mathbf{P}' - \mathbf{Q}') \\ &= (\mathbf{PX} - \mathbf{QX})(\mathbf{X}'\mathbf{P}' - \mathbf{X}'\mathbf{Q}') \\ &= (\mathbf{PX} - \mathbf{QX})(\mathbf{PX} - \mathbf{QX})' = 0\end{aligned}$$

■

- given real matrices $\mathbf{P}, \mathbf{Q}, \mathbf{X}$, $\mathbf{X}'\mathbf{XP} = \mathbf{X}'\mathbf{XQ} \Rightarrow \mathbf{XP} = \mathbf{XQ}$

  Proof.

$$\begin{aligned}(\mathbf{P}' - \mathbf{Q}')(\mathbf{X}'\mathbf{XP} - \mathbf{X}'\mathbf{XQ}) &= (\mathbf{P}' - \mathbf{Q}')\mathbf{X}'(\mathbf{XP} - \mathbf{XQ}) \\ &= (\mathbf{P}'\mathbf{X}' - \mathbf{QX}')(\mathbf{XP} - \mathbf{XQ}) \\ &= (\mathbf{XP} - \mathbf{XQ})'(\mathbf{XP} - \mathbf{XQ}) = 0\end{aligned}$$

■

- idempotent matrix: $\mathbf{AA} = \mathbf{A}$ (e.g. identity matrix, zero matrix, projection matrix)
- centering matrix: $\mathbf{C}_n$ is a symmetric and idempotent matrix, which when multiplied with a same effect as subtracting the mean of the components of vectors from every component of that vector
  - mean: $\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\mathbf{1}_n'\mathbf{x}$ where $\mathbf{1}_n$ is column vector of ones (summing vector, $\mathbf{1}_n'\mathbf{1}_n = n$)
  - variance: $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n}\mathbf{x}'\mathbf{x} - \frac{1}{n^2}\mathbf{x}'\mathbf{1}_n\mathbf{1}_n'\mathbf{x} = \frac{1}{n}\mathbf{x}'(\mathbf{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n')\mathbf{x} = \frac{1}{n}\mathbf{x}'\mathbf{C}_n\mathbf{x}$
- quadratic form: $\mathbf{x}'\mathbf{Ax}$
  - in summation form,

$$\begin{aligned}\mathbf{x}'\mathbf{Ax} &= \sum_{i=1}^{n} a_{ii}x^2 + \sum_{j=1}^{n}\sum_{\substack{i=1,\\ i\neq j}}^{n} a_{ij}x_ix_j \\ &= \sum_{i=1}^{n} a_{ii}x^2 + \sum_{j=1}^{n}\sum_{i=j+1}^{n}(a_{ij} + a_{ji})x_ix_j\end{aligned}$$

  - it is possible to have $\mathbf{x}'\mathbf{Ax} = \mathbf{x}'\mathbf{Bx}$ with $\mathbf{A} \neq \mathbf{B}$
  - with symmetric matrix, $\mathbf{x}'\mathbf{Ax} = \mathbf{x}'(\frac{1}{2}\mathbf{A} + \frac{1}{2}\mathbf{A}')\mathbf{x}$

$$\begin{aligned}\mathbf{x}'\mathbf{Ax} &= \sum_{i=1}^{n} a_{ii}x^2 + \sum_{j=1}^{n}\sum_{\substack{i=1,\\ i\neq j}}^{n} a_{ij}x_ix_j \\ &= \sum_{i=1}^{n} a_{ii}x^2 + 2\sum_{j=1}^{n}\sum_{i=j+1}^{n} a_{ij}x_ix_j\end{aligned}$$

- positive definite: $\mathbf{x}'\mathbf{Ax}$ is positive for every nonzero real column vector $\mathbf{x}$
  positive semi-definite: $\mathbf{x}'\mathbf{Ax}$ is non-negative for every nonzero real column vector $\mathbf{x}$

- every symmetric idempotent matrix is positive semi-definite

  Proof.

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})'\mathbf{A}\mathbf{x} \geq 0$$

■

1.4 Inverse Matrix

- determinant: a scalar-valued function of the entries of a square matrix
  - Laplace expansion: the minor $\mathbf{M}_{ij}$ is $(n-1)\times(n-1)$ matrix that results from $\mathbf{A}$ by removing the ith row and jth column

$$|\mathbf{A}| = \sum_{j=1}^{n}(-1)^{i+j}a_{ij}|\mathbf{M}_{ij}| = \sum_{i=1}^{n}(-1)^{i+j}a_{ij}|\mathbf{M}_{ij}| \text{ for any } i, j$$

  - $|\mathbf{A}| = |\mathbf{A}'|$
  - $|\mathbf{A}| = 0$ if two rows or columns of $\mathbf{A}$ are equal
  - $|\mathbf{A}| = \prod_{i=1}^{n} a_{ii}$ if $\mathbf{A}$ is the lower(upper) triangular matrix
  - $|\mathbf{A}\mathbf{B}| = |\mathbf{A}||\mathbf{B}| = |\mathbf{B}||\mathbf{A}| = |\mathbf{B}\mathbf{A}|$
  - $|\mathbf{A}^2| = |\mathbf{A}|^2$
  - $|\mathbf{A}| = 0$ or 1 if $\mathbf{A}$ is an idempotent matrix
- inverse matrix
  - square matrix $\mathbf{A}$ is called invertible if there exists a square matrix $\mathbf{B}$ such that $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} = \mathbf{I}$
  - the square matrix $\mathbf{B}(= \mathbf{A}^{-1})$, the inverse matrix of $\mathbf{A}$ is unique.
  - $|\mathbf{A}^{-1}| = \dfrac{1}{|\mathbf{A}|}$
  - $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
  - $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$
  - $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
  - if $\mathbf{A}$ is symmetry, $\mathbf{A}^{-1}$ is also symmetry
  - if $\mathbf{D} = \text{diag}(x_i)$, $\mathbf{D}^{-1} = \text{diag}(\dfrac{1}{x_i})$
- linear combinations of vectors and scalars

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \mathbf{X}\mathbf{a}$$

  where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ and $\mathbf{a} = (a_1, a_2, \ldots, a_n)$
  - $\mathbf{X}$ is the transformation matrix
- a sequence of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is said to be linearly dependent if there exist scalars $a_1, a_2, \ldots, a_n$, not all zero, such that $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \mathbf{0}$ where $\mathbf{0}$ is zero vector
  - with at least one non-zero scalar ($a_1 \neq 0$), the above equation is able to be written as

$$\mathbf{x}_1 = -\frac{a_2}{a_1}\mathbf{x}_2 - \cdots - \frac{a_n}{a_1}\mathbf{x}_n$$

  - therefore, a set of vectors is linearly dependent
    $\iff$ one of them is $\mathbf{0}$ or a linear comination of the others

- a sequence of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is said to be linearly independent, if it is not linearly dependent. that is, $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \mathbf{0}$ can only be satisfied by $a_i = 0$ for $i = 1, \ldots, n$
  - a set of vectors is linearly independent
    $\iff$ $\mathbf{0}$ can be represented as a linear combination of its vectors in a unique way
  - the statements below are identical
    1. a sequence of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is linearly independent
    2. if $\mathbf{Xa} = 0,\ \mathbf{a} = 0$
    3. $\mathbf{X}$ is invertible (the inverse matrix of $\mathbf{X}, \mathbf{X}^{-1}$ exists)

- rank: the number of linearly independent row or column vectors
  - the column rank and the row rank are always equal
  - $\mathbf{A}$ has a full rank if its rank is equal to the lesser of the number of row and columns
  - if and only if $\mathbf{A}$ has a full rank, $\mathbf{A}$ is invertible

- vector space: given $\mathbf{x}, \mathbf{y} \in S$ and for all scalar, S is a vector space when $a\mathbf{x} + b\mathbf{y} \in S$ holds
  - $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ span S: all the elements are represented by the linear combinations of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$
  - a set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is called as spanning set
  - when vectors in spanning set are linearly independent, the spanning set is called as basis

1.5 Eigenvalues

- $\mathbf{A}\nu = \mathbf{w} = \lambda\nu \iff (\mathbf{A} - \lambda\mathbf{I})\nu = 0$
  - ν is an eigenvector of the linear transformation $\mathbf{A}$
  - the scale factor λ is the eigenvalue corresponding to the eigenvector
  - the equation has a non-zero solution ν if and only if $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ ($\mathbf{A} - \lambda\mathbf{I}$ is not a full rank)
  - $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ can be factored into the product of n linear terms

    $$\det(\mathbf{A} - \lambda\mathbf{I}) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)$$

    where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{A}$
  - $tr(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i$
  - $\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$
  - the matrix $\mathbf{A}$ is invertible if and only if every eigenvalue is nonzero
  - if $\mathbf{A}$ is invertible, then the eigenvalues of $\mathbf{A}^{-1}$ are $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n}$
  - if $\mathbf{A}$ is unitary($\mathbf{A}' = \mathbf{A}^{-1}$), every eigenvalue has absolute value $|\lambda_i| = 1$

- eigendecomposition
  - define a square matrix $\mathbf{Q}$ whose columns are the n linearly independent eigenvectors of $\mathbf{A}$

    $$\mathbf{Q} = \begin{bmatrix} \nu_1 & \nu_2 & \cdots & \nu_n \end{bmatrix}$$

then we have

$$\mathbf{AQ} = \begin{bmatrix} \lambda_1\nu_1 & \lambda_2\nu_2 & \cdots & \lambda_n\nu_n \end{bmatrix}$$

- define a diagonal matrix $\mathbf{\Lambda}$ where each diagonal element is the eigenvalue associated with the ith column of $\mathbf{Q}$

$$\mathbf{AQ} = \mathbf{Q\Lambda}$$

- since the columns of $\mathbf{Q}$ are linearly independent, $\mathbf{Q}$ is invertible

$$\mathbf{A} = \mathbf{Q\Lambda Q}^{-1}$$

- $\mathbf{A}$: decomposed into (1) a matrix composed of its eigenvectors (2) a diagnoal matrix with its eigenvalues along the diagonal, and (3) the inverse of the matrix of eigenvectors
- such a matrix A is said to be diagonalizable
- the eigenvectors of $\mathbf{A}$ form a basis $\iff$ $\mathbf{A}$ is diagonalizable

Chapter 2 Regression

2.1 Projection and Regressions

- the conditional mean function is the regression function

$$Y = E[Y \mid X] + (Y - E[Y \mid X]) = E[Y \mid X] + \epsilon$$

where $E[\epsilon \mid X] = 0$

- suppose that there exists a linear function such that $Y = \alpha + \beta X + \epsilon$ where $E[\epsilon \mid X] = 0$
  - $Cov(X, Y) = Cov(X, \alpha) + \beta Cov(X, X) + Cov(X, \epsilon) = 0 + \beta Cov(X, X) + 0$
  - $\beta = \dfrac{Cov(X, Y)}{Cov(X, X)} = \dfrac{Cov(X, Y)}{Var(X)}$
  - $\alpha = E[Y] - \beta E[X] - E[\epsilon] = E[Y] - \beta E[X] = E[\epsilon \mid X] = E[Y] - \beta E[X]$ ($\because E[E[\epsilon \mid X]] = E[\epsilon] = 0$)
  - note that this does not mean $E[Y \mid X] = \alpha + \beta X$, but this is the linear projection of Y on X
  - simple regression is closely related with covariance

2.2 Least Squares Estimation

- estimate α and β to minimize

$$S(\alpha, \beta) = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 = \sum_{i=1}^{n} \epsilon_i^2$$

- least-squares normal equations

$$\frac{\partial S}{\partial \alpha}\Big|_{\hat{\alpha},\hat{\beta}} = -2\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \implies n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\frac{\partial S}{\partial \beta}\Big|_{\hat{\alpha},\hat{\beta}} = -2\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)x_i = 0 \implies \hat{\alpha}\sum_{i=1}^{n} x_i + \hat{\beta}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$

- LS estimators: $\begin{cases} \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \\ \hat{\beta} = \dfrac{S_{xy}}{S_{xx}} \end{cases}$ where $S_{xx} = \sum(x_i - \bar{x})^2$ and $S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$
- note that coefficient β is not unit free (cf. correlation coefficient is unit free)
- by multiplying some terms,

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})(y_i - \bar{y})} \frac{(y_i - \bar{y})}{(y_i - \bar{y})}$$

  - the last term is the slope of the line that connects the t-th point to the average of all points
  - $\hat{\beta}$ is the weighted average of $\dfrac{(y_i - \bar{y})}{(y_i - \bar{y})}$ where the weight is $\dfrac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})(y_i - \bar{y})}$

2.3 Prediction Errors

- prediction: $\hat{\alpha} + \hat{\beta}X_t = \hat{Y}_t$
- prediction error: $Y_t - \hat{Y}_t = Y_t - \hat{\alpha} + \hat{\beta}X_t = e_t$
- R^2 measures how much variations of y_i can be explained by the regression

$$\underbrace{\sum_{t=1}^{T}(Y_t - \bar{T})^2}_{\text{Total Sum of Squares (TSS)}} = \underbrace{\sum_{t=1}^{T}(\hat{Y}_t - \bar{Y})^2}_{\text{Explained Sum of Squares (ESS)}} + \underbrace{\sum_{t=1}^{T} e_t^2}_{\text{Unexplained Sum of Squares (USS)}}$$

  - $R^2 = \dfrac{\text{ESS}}{\text{TSS}} = 1 - \dfrac{\text{USS}}{\text{TSS}},\ 0 \le R^2 \le 1$

2.4 Multiple Regression

- remark. $\dfrac{\partial(\mathbf{a}'\mathbf{x})}{\partial\mathbf{x}} = \mathbf{a}$, $\dfrac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial\mathbf{x}} = 2\mathbf{A}\mathbf{x}$
- adjusted $R^2 = 1 - \dfrac{\text{USS}/(n-k-1)}{\text{TSS}/(n-1)}$
- wish to find $\hat{\beta}$ that minimize

$$\begin{aligned} S(\beta) &= \sum_{i=1}^{n} \epsilon_i^2 \\ &= \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} \\ &= (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) \end{aligned}$$

- $S(\beta)$ can be expressed as

$$\begin{aligned} S(\beta) &= (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) \\ &= \mathbf{y}^{\mathbf{y}} - \beta^T\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\beta + \beta^T\mathbf{X}^T\mathbf{X}\beta \\ &= \mathbf{y}^{\mathbf{y}} - 2\mathbf{y}^T\mathbf{X}\beta + \beta^T\mathbf{X}^T\mathbf{X}\beta \quad (\because \mathbf{y}^T\mathbf{X}\beta \text{ is a scalar}) \end{aligned}$$

- LS estimators must satisfy

$$\left.\frac{\partial S}{\partial \beta}\right|_{\hat{\beta}} = -2\mathbf{X}^T + 2\mathbf{X}^T\mathbf{X}\hat{\beta} = 0 \iff \mathbf{X}^T\mathbf{X}\hat{\beta} = \mathbf{X}^T\mathbf{y}$$

  thus, $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

Chapter 3 Gauss-Markov Theorem

3.1 Projection

- $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{P}_X\mathbf{y}$
  - $\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is projection matrix on to the regression space
  - $\mathbf{P}_X$ is symmetric and idempotent
- $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{P}_X\mathbf{y} = (\mathbf{I} - \mathbf{P}_X)\mathbf{y} = \mathbf{M}_X(\mathbf{X}\beta^* + \epsilon)$
  - $\mathbf{I} - \mathbf{P}_X$ is idempotent
  - since $\mathbf{M}_X\mathbf{X}\beta^* = \mathbf{X}\beta^* - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta^* = 0$, $\mathbf{e} = \mathbf{M}_X\epsilon$

3.2 Estimator

- properties
  - unbiased: $E(\hat{\beta}) = \beta^*$
  - consistent: $\operatorname{plim}_{n\to\infty} \hat{\beta} = \beta^*$
  - efficient: the estimator with the smallest variance among unbiased estimators
  - unbiased $\Rightarrow$ consistent
- MSE(Mean Squared Error)

$$MSE(\hat{\beta}) = E[(\hat{\beta} - \beta)^2] = Var(\hat{\beta}) + (Bias(\hat{\beta}, \beta))^2$$

  - e.g. 1) $\frac{1}{T}\sum_{t=1}^{T}(e_t - \bar{e})^2$ is consistent while 2) $\frac{1}{T-1}\sum_{t=1}^{T}(e_t - \bar{e})^2$ is unbiased
  - MSE: 1) $\frac{(2n-1)\sigma^4}{n^2} < 2)\ \frac{2\sigma^4}{n-1}$
- classical assumptions in DGP

  A0. X_t is not random (not stochastic) cf. ϵ_t is random

  A1. $E[\epsilon_t] = 0$

  A2. $Var(\epsilon_t) = \sigma^2$ (homoskedasticity)

  A3. $Cov(\epsilon_t, \epsilon_s) = 0$ for $t \neq s$. no serial correlation

  A4. $\epsilon_t \sim N(0, \sigma^2)$
- statistical properties in OLS

  1. Unbiasedness: $E(\hat{\beta}) = \beta^*$

Proof.

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta+\epsilon) \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \\ E[\hat{\beta}] &= E[\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon] \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\epsilon] \quad (\because \text{A0}) \\ &= \beta \quad (\because \text{A1})\end{aligned}$$

■

2. Variance: $Var(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

Proof.

$$\begin{aligned}Var(\hat{\beta}) &= Var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta+\epsilon)) \\ &= Var(\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon) \\ &= Var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon\epsilon')\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1} \quad (\because \text{A2, A3}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

■

3.3 Gauss-Markov Theorem

- if A1, A2, A3 are satisfied, then the OLS estimator has the smallest variance among all linear and unbiased estimators for β^*

  Proof. Let $\tilde{\beta}$ be the unbiased and linear. We need to show $Var(\tilde{\beta}) - Var(\hat{\beta})$ is positive semi-definite.

  (1) $\tilde{\beta} = \mathbf{C}\mathbf{y}$: given linear assumption

  (2) $E(\tilde{\beta}) = \beta^*$: given unbiased assumption

  Since $Var(\mathbf{y}) = Var(\mathbf{X}\beta+\epsilon) = Var(\epsilon)$,

$$Var(\tilde{\beta}) = \mathbf{C}Var(\mathbf{y})\mathbf{C}' = \sigma^2\mathbf{C}\mathbf{C}'$$

  Let $\mathbf{A} = \mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Then

$$\mathbf{C} = \mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{A} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

  Thus,

$$\begin{aligned}\sigma^2\mathbf{C}\mathbf{C}' &= \sigma^2(\mathbf{A} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{A} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' \\ &= \sigma^2(\mathbf{A}\mathbf{A}' + \mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}' + (\mathbf{X}'\mathbf{X})^{-1})\end{aligned}$$

  Since $E(\tilde{\beta}) = E(\mathbf{C}\mathbf{y}) = E(\mathbf{C}\mathbf{X}\beta + \mathbf{C}\epsilon) = \mathbf{C}\mathbf{X}\beta + 0 = \beta$, $\mathbf{C}\mathbf{X} = \mathbf{I}$ and thus

$$\mathbf{A}\mathbf{X} = \mathbf{C}\mathbf{X} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{C}\mathbf{X} - \mathbf{I} = 0$$

Thus,

$$\begin{aligned}\sigma^2\mathbf{CC}' &= \sigma^2(\mathbf{AA}' + \mathbf{AX}(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}' + (\mathbf{X}'\mathbf{X})^{-1}) \\ &= \sigma^2\mathbf{AA}' + \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = Var(\tilde{\beta})\end{aligned}$$

Then, $Var(\tilde{\beta}) - Var(\hat{\beta}) = \sigma^2\mathbf{AA}'$.

To show that $\mathbf{AA}'$ is positive semi-definite (i.e. $\mathbf{x}'\mathbf{AA}'\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$), define $\mathbf{Z} = \mathbf{A}'\mathbf{x}$.

$$\mathbf{x}'\mathbf{AA}'\mathbf{x} = \mathbf{Z}'\mathbf{Z} = (z_1^2 + z_2^2 + \cdots + z_T^2) \geq 0$$

Therefore, OLS estimator is BLUE.

■

Chapter 4 Test

4.1 Hypothesis Testing

H_0: $\beta_i^* = 0$ H_1: $\beta_i^* \neq 0$		**Decision**	
		Accept H_0	Reject H_0
True Statement	H_0: True	OK	Type I Error
	H_0: False	Type II Error	OK

- classical approach to testing
  - set probability (type 1) $= \alpha$
  - try to minimize prob (type 2)
- $\hat{\beta} \sim N(\beta^*, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, $\hat{\beta}_i \sim N(\beta_i^*, \sigma^2 s_{ii}^2)$ where s_{ii}^2 is the element of $(\mathbf{X}'\mathbf{X})^{-1}$
- how to minimize $P(\text{Type II}) = \beta$?
  - with greater T
  - low variance
- usual procedure for testing
  1. find a good estimator for the parameter in H_0 and H_a
  2. using the estimator, develop a sensible test statistics
  3. to control P(type I error), try to find the null distribution of the test statistics
  4. try to find a rejection region that maximizes the power of the test $(= 1 - P(\text{type II error}))$
- If σ^2 is known,

$$Z = \frac{\hat{\beta}_i - \beta^* \text{under } H_0}{\sigma s_{ii}} = \frac{\hat{\beta}_i}{\sigma s_{ii}} \sim N(0,1)$$

given $\alpha = 0.05$, the decision rule is $\begin{cases} |\frac{\hat{\beta}_i}{\sigma s_{ii}}| > 1.96, & \text{reject } H_0 \\ |\frac{\hat{\beta}_i}{\sigma s_{ii}}| \leq 1.96, & \text{fail to reject } H_0 \end{cases}$

- in reality, we have to estimate σ^2

- $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2$ is a biased estimator

  Proof. Since $\mathbf{e} = \mathbf{M}_X\epsilon = (\mathbf{I} - \mathbf{P}_X)\epsilon$,

  $$E(\hat{\sigma}^2) = \frac{1}{T}E(\mathbf{e}'\mathbf{e}) = \frac{1}{T}E(\epsilon'\mathbf{M}_X\mathbf{M}_X\epsilon) = \frac{1}{T}E(\epsilon'\mathbf{M}_X\epsilon)$$

  Since a real number can be thought as a 1×1 matrix and its trace is itself,

  $$\begin{aligned} E(\hat{\sigma}^2) = \frac{1}{T}E(\epsilon'\mathbf{M}_X\epsilon) &= \frac{1}{T}E(tr(\epsilon'\mathbf{M}_X\epsilon)) \\ &= \frac{1}{T}E(tr(\epsilon\epsilon'\mathbf{M}_X)) \\ &= \frac{1}{T}tr(E(\epsilon\epsilon'\mathbf{M}_X)) \\ &= \frac{1}{T}tr(\sigma^2\mathbf{M}_X) \\ &= \frac{1}{T}(\sigma^2 tr(\mathbf{I}_T) - \sigma^2 tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')) \\ &= \frac{1}{T}(\sigma^2 T - \sigma^2 tr(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})) \\ &= \frac{1}{T}(\sigma^2 T - \sigma^2 tr(\mathbf{I}_k)) \\ &= \frac{T-k}{T}\sigma^2 \neq \sigma^2 \end{aligned}$$

  ■

  - the unbiased estimator is $\hat{\sigma}^2 = \frac{1}{T-k}\sum_{t=1}^{T} e_t^2$

- theorem: if $Z_1 \sim N(0,1)$, $Z_2 \sim \chi^2(\nu)$ and $Z_1 \perp\!\!\!\perp Z_2$, $\frac{Z_1}{\sqrt{Z_2/\nu}} \sim t(\nu)$

  - since $\frac{\hat{\beta}_i}{\sigma s_{ii}} \sim N(0,1)$, $\frac{\mathbf{e}'\mathbf{e}}{\sigma^2} \sim \chi^2(T-k)$ and $\frac{\hat{\beta}_i}{\sigma s_{ii}} \perp\!\!\!\perp \frac{\mathbf{e}'\mathbf{e}}{\sigma^2}$,

    $$\frac{\frac{\hat{\beta}_i}{\sigma s_{ii}}}{\sqrt{\frac{\mathbf{e}'\mathbf{e}}{\sigma^2(T-k)}}} = \frac{\frac{\hat{\beta}_i}{\sigma s_{ii}}}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}} \sim t(T-k)$$

Chapter 5 Model Misspecification

5.1 Model Misspecification Error

- implicit assumptions
  1. $rank(\mathbf{X}) = \tilde{k}$
  2. relationship between y_t and X_t: linear
  3. correct specification: we know the true model
- omission of relevant variables
  - DGP: $\mathbf{y} = \beta_0^* + \mathbf{X}_1\beta_1^* + \mathbf{X}_2\beta_2^* + \epsilon_t$ vs. regression: $\mathbf{y} = \mathbf{X}_1\beta_1 + \nu_t$

- β_1 becomes biased

  Proof.

$$\begin{aligned}\hat{\beta}_1 &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} \\ &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{X}_1\beta_1^* + \mathbf{X}_2\beta_2^* + \epsilon_t) \\ &= \beta_1^* + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2^* + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\epsilon_t\end{aligned}$$

  Thus,

$$E[\hat{\beta}_1] = \beta_1^* + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2^* \neq \beta_1^*$$

■

- inclusion of irrelevant variable
  - DGP: $\mathbf{y} = \mathbf{X}_1\beta_1^* + \epsilon_t$ vs. regression: $\mathbf{y} = \mathbf{X}_1\hat{\beta}_1 + \mathbf{X}_2\hat{\beta}_2 + \nu_t$
  - By definition, we have

$$\mathbf{y} = \mathbf{X}_1\hat{\beta}_1 + \mathbf{X}_2\hat{\beta}_2 + (\mathbf{I} - \mathbf{P}_X)\mathbf{y}$$

    Multiplying both sides by $\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)$,

$$\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{y} = \mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1\hat{\beta}_1 + \mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_2\hat{\beta}_2 + \mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)(\mathbf{I} - \mathbf{P}_X)\mathbf{y}$$

    Since $(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_2 = 0$ and $(\mathbf{I} - \mathbf{P}_2)(\mathbf{I} - \mathbf{P}_X) = \mathbf{I} - \mathbf{P}_X$,

$$\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{y} = \mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1\hat{\beta}_1$$

  - we have the following estimator: $\begin{cases}\hat{\beta}_1 = (\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{y} \\ \hat{\beta}_2 = (\mathbf{X}_2'(\mathbf{I} - \mathbf{P}_1)\mathbf{X}_2)^{-1}\mathbf{X}_2'(\mathbf{I} - \mathbf{P}_1)\mathbf{y}\end{cases}$
  - variance of $\hat{\beta}_1$

$$\begin{aligned}Var(\hat{\beta}_1) &= Var((\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{y}) \\ &= Var(\beta_1^* + (\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\epsilon) \\ &= (\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)Var(\epsilon)(\mathbf{I} - \mathbf{P}_2)(\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1)^{-1} \\ &= \sigma^2(\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1)^{-1}\end{aligned}$$

  - given positive semi-definite matrix $\mathbf{P}_2$, $\mathbf{X}_1'\mathbf{X} - \mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1 = \mathbf{X}_1'\mathbf{P}_2\mathbf{X}_1$ is positive semi-definite and $(\mathbf{X}_1'(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1)^{-1} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}$ is also positive semi-definite
  - if the columns of $\mathbf{X}_1$ is orthogonal to the columns of $\mathbf{X}_2$, we immediately have $(\mathbf{I} - \mathbf{P}_2)\mathbf{X}_1 = \mathbf{X}_1$, so that the estimator is same as the estimator of OLS ⇒ estimating a more complex specification does not result in efficiency loss

5.2 Model Selection

- sequential search: general to simple / simple to general
- model selection information criterion
  - adjusted $R^2 = 1 - (1 - R^2)\dfrac{n-1}{n-p-1}$
  - AIC $= 2k - 2\ln(\hat{L})$
  - BIC $= k\ln(n) - 2\ln(\hat{L})$: appropriate for selecting the "true model"

Chapter 6 Multicollinearity

6.1 Multicollinearity

- we say there is multicollinearity in data where some explanatory variable have strong inner relationship
  - correlation coefficient $\simeq \pm 1$ ($|c| > 0.7$)
  - violation of A0 assumption ($\mathbf{X}$ is not random)
  - high R^2 with low t-statistic values
- VIF(variance inflation factor)
  1. run the OLS regression for each X variable:
  $$x_{1i} = \alpha_1 + \alpha_2 x_{2i} + \cdots + \alpha_k x_{ki} + \nu_i$$
  2. calculate the VIF for $\hat{\beta}_i$
  $$\text{VIF}(\hat{\beta}_i) = \frac{1}{1 - R_i^2}$$
  3. evaluate each $\text{VIF}(\hat{\beta}_i)$ (rule of thumb: multicollinearity may be present if $\text{VIF}(\hat{\beta}_i) > 5$)
- with perfect multicollinearity, the rank of matrix $\mathbf{X}$ is not full and the parameter estimates are not well-defined
- with near collinearity, the Gauss-Markov theorem tells us that the OLS estimators are BLUE, but the variances of the regression coefficient estimates will increase (lower t-statistics)
- however, forecasting and prediction will be largely unaffected