LINEAR ALGEBRA AND PROPERTIES OF ESTIMATORS

Advanced Econometrics

Outline and Objectives

Outline:

- review of linear algebra
- linear algebra and the least squares estimator
- properties of estimators
- a small quiz

Objectives:

- develop a common language for communicating
- apply linear algebra to a familiar setting
- understand what constitutes a “good” estimator

Matrices and Vectors

- A *matrix* is a rectangular array of elements:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NK} \end{bmatrix}$$

- $\mathbf{X}$ has N rows and K columns – $N \times K$.
- Element in row i and column j of $\mathbf{X}$: x_{ij}
- Matrices for which $N = K$ are called *square*.
- Two special sets of matrices: *column vectors* and *row vectors*:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad \text{and} \qquad \mathbf{z} = \begin{bmatrix} z_1 & z_2 & \cdots & z_K \end{bmatrix}$$

Equality and Transposition

- **Matrix Equality:** $\mathbf{X} = \mathbf{Z}$ if they contain the exact same elements at the exact same positions: $x_{ij} = z_{ij}, \forall i, j$
- **Transpose of a Matrix:** For any matrix (or vector) $\mathbf{X}$, its transpose is the matrix obtained by interchanging the rows and columns of $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NK} \end{bmatrix} \Leftrightarrow \mathbf{X}' = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{N1} \\ x_{12} & x_{22} & \cdots & x_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1K} & x_{2K} & \cdots & x_{NK} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \Leftrightarrow \mathbf{y}' = \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix}$$

Symmetry and Addition

- **Symmetric Matrices:** A square matrix is *symmetric* if it is equal to its transpose:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 7 \end{bmatrix} \Longrightarrow \mathbf{A}' = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 7 \end{bmatrix} = \mathbf{A}$$

- **Matrix Addition:** $\mathbf{W} = \mathbf{X} + \mathbf{Z}$ is defined as:

$$\mathbf{W} = \mathbf{X} + \mathbf{Z} = \begin{bmatrix} x_{11} + z_{11} & x_{12} + z_{12} & \cdots & x_{1K} + z_{1K} \\ x_{21} + z_{21} & x_{22} + z_{22} & \cdots & x_{2K} + z_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} + z_{N1} & x_{N2} + z_{N2} & \cdots & x_{NK} + z_{NK} \end{bmatrix}$$

  - $\mathbf{X}$ and $\mathbf{Z}$ should have the same dimensions – conformable
  - Matrix addition follows the same rules as scalar addition
  - There exists a zero matrix, $\mathbf{0}$, such that $\mathbf{X} + \mathbf{0} = \mathbf{X}$. All the elements of $\mathbf{0}$ are equal to 0.

Multiplication

- **Scalar-Matrix Multiplication:** Let a be a real number. $\mathbf{W} = a \cdot \mathbf{X}$ is defined as:

$$\mathbf{W} = a\mathbf{X} = \begin{bmatrix} ax_{11} & ax_{12} & \cdots & ax_{1K} \\ ax_{21} & ax_{22} & \cdots & ax_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ ax_{N1} & ax_{N2} & \cdots & ax_{NK} \end{bmatrix}$$

- **Matrix Multiplication:** Let $\mathbf{X}$ an $N \times K$ matrix and $\mathbf{Z}$ a $K \times L$ matrix. Then $\mathbf{W} = \mathbf{X} \cdot \mathbf{Z}$ is defined as an $N \times L$ matrix with elements:

$$w_{i\ell} = \sum_{k=1}^{K} x_{ik} \cdot z_{k\ell} \quad \forall i = 1, \ldots N, \quad \ell = 1, \ldots, L$$

Multiplication

- For the product to be defined, the two matrices need to be conformable for multiplication:

$$[N \times K] \cdot [K \times L]$$

- Visually:

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & \cdots & x_{i\ell} & \cdots & x_{iK} \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \cdots & z_{1\ell} & \cdots \\ \vdots & \vdots & \vdots \\ \vdots & z_{i\ell} & \vdots \\ \vdots & \vdots & \vdots \\ \cdots & z_{K\ell} & \cdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \vdots \\ \vdots & \sum_{k=1}^{K} x_{ik} z_{k\ell} & \vdots \\ \vdots & \vdots & \vdots \end{bmatrix}$$

- The ith row of $\mathbf{X}$ is a row vector. The ℓth column of $\mathbf{Z}$ is a column vector: $w_{i\ell}$ is the *inner product* of the two.

Multiplication

- If $\mathbf{x}$ and $\mathbf{z}$ are $K \times 1$ vectors, the *inner* and *outer* products are:

$$w = \mathbf{x}'\mathbf{z} = \mathbf{z}'\mathbf{x} \qquad \text{and} \qquad \mathbf{W} = \mathbf{x}\mathbf{z}' \neq \mathbf{z}\mathbf{x}' = \mathbf{W}'$$

- In general:

$$\mathbf{XZ} \neq \mathbf{ZX} \qquad \text{but} \qquad (\mathbf{XZ})' = \mathbf{Z}'\mathbf{X}'$$

  which can be used to deduce that $\mathbf{X}'\mathbf{X}$ is symmetric
- There exists a square matrix $\mathbf{I}$ such that $\mathbf{X} \cdot \mathbf{I} = \mathbf{X}$ and $\mathbf{I} \cdot \mathbf{X} = \mathbf{X}$.
- $\mathbf{I}$ is called the *identity matrix*:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Rank and Inversion

- **Rank of a Matrix:** The *rank* of a matrix is the number of linearly independent columns or rows.
  - Let $\mathbf{X}$ be $N \times K$, with $N \geq K$. $\mathbf{X}$ has *full rank* if $\operatorname{rank}(\mathbf{X}) = K$
  - It holds:

$$\operatorname{rank}(\mathbf{X}) = \operatorname{rank}(\mathbf{X}\mathbf{X}') = \operatorname{rank}(\mathbf{X}'\mathbf{X})$$

- **Matrix Inversion** The inverse of a square matrix $\mathbf{A}$, if it exists, is a matrix $\mathbf{A}^{-1}$ with the property:

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

  - Not all matrices can be inverted!
  - To be invertible, a matrix has to be square <u>and</u> to have full rank – *non-singular*
  - If $\mathbf{A}$ and $\mathbf{B}$ are invertible, it holds:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \qquad \text{and} \qquad (\mathbf{A}')^{-1} = \left(\mathbf{A}^{-1}\right)'$$

Positive and Negative Definiteness

- **Positive and Negative Definite Matrices:** Let $\mathbf{A}$ be a $K \times K$ symmetric matrix and $\mathbf{v}$ be a $K \times 1$ vector. $\mathbf{A}$ is positive definite if:

$$\mathbf{v}'\mathbf{A}\mathbf{v} > 0 \qquad \forall \mathbf{v} \neq \mathbf{0}$$

$\mathbf{A}$ is positive semi-definite if:

$$\mathbf{v}'\mathbf{A}\mathbf{v} \geq 0 \qquad \forall \mathbf{v} \neq \mathbf{0}$$

Negative definiteness and negative semi-definiteness are defined similarly – different direction of the inequality

Why Matrix Algebra in this Course?

- We will deal with econometric models – theoretical model plus data.
- We need a compact notation to represent the model – store the data in matrices.
- The linear model with K explanatory variables:

$$y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + x_{i,3}\beta_3 + \ldots + x_{i,K}\beta_K + \varepsilon_i$$

could be written as:

$$y_i = \sum_{k=1}^{K} x_{i,k}\beta_k + \varepsilon_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

where:

$$\mathbf{x}_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,K} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

Why Matrix Algebra in this Course?

- We could go further. we can write $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ for N potential observations as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_N' \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NK} \end{bmatrix}$$

- The ordinary least squares estimator can be written as:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y}) = \left(\sum_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{x}_i y_i\right)$$

Introductory Example

- We want to estimate the average share of household income spent on food products, in the Netherlands
- This share is a random variable, Y
- We use a sample of 100 households
- We consider two estimators for the mean of the distribution of Y:
  1. the sample mean
  2. the sample median

- Which one would you use?
- Why?

Estimators and Properties of Estimators

- In econometrics, we have a model for the population
- We assume we know the data generating process, up to a vector of parameters $\boldsymbol{\theta}$
- We take a sample from the population $\rightarrow$ random sample implies random variables
- We approximate $\boldsymbol{\theta}$ using the data
- An estimator is a formula which is used to approximate the parameters of the population using information from the sample – $\hat{\boldsymbol{\theta}}$.
- $\boldsymbol{\theta}$ is fixed!
- data are random $\Longrightarrow$ $\hat{\boldsymbol{\theta}}$ is a random variable
- $\hat{\boldsymbol{\theta}}$ has all the characteristics of a random variable – mean, variance, pdf, etc

Finite Sample Properties – Unbiasedness & Efficiency

- **Unbiasedness:** An estimator is *unbiased* if its expected value is equal to the parameter it is estimating:

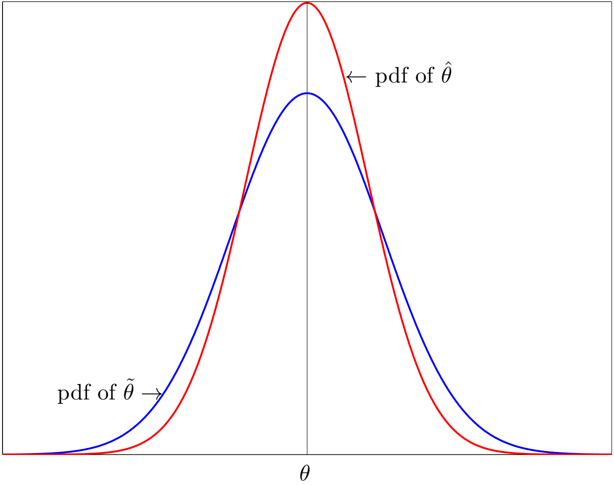
$$\mathrm{E}\left(\hat{\boldsymbol{\theta}}\right) = \boldsymbol{\theta}$$

In repeated samples, the estimator is on average correct.

- **Efficiency:** An unbiased estimator is *efficient* if it has the smallest possible variance within a class of unbiased estimators.
  - Efficiency is a desirable property: the smaller the variance of an estimator the more trust we can place on an estimate.
  - Efficiency is a relative concept – there also exist absolute measures

Verbeek – §2.3

Finite Sample Properties – Unbiasedness & Efficiency



- Both $\tilde{\theta}$ and $\hat{\theta}$ are unbiased
- $\hat{\theta}$ has smaller variance than $\tilde{\theta}$

Finite Sample Properties – Normality

- **Normality:** For an estimator $\hat{\boldsymbol{\theta}}$ that has a multivariate normal distribution with mean $\boldsymbol{\theta}$ and variance $\boldsymbol{\Sigma}$ we will write:

$$\hat{\boldsymbol{\theta}} \sim \mathrm{N}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right)$$

  - Normality facilitates post-estimation inference – t tests require $\hat{\theta}_k$ to have a normal distribution

Theorem

Let $\hat{\boldsymbol{\theta}}$ be a K-dimensional normally distributed random vector. We can partition $\hat{\boldsymbol{\theta}}$ into two parts such that:

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_1 \\ \hat{\boldsymbol{\theta}}_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

Then: $\qquad \hat{\boldsymbol{\theta}}_1 \sim \mathrm{N}\left(\boldsymbol{\theta}_1, \boldsymbol{\Sigma}_{11}\right) \qquad \text{and} \qquad \hat{\boldsymbol{\theta}}_2 \sim \mathrm{N}\left(\boldsymbol{\theta}_2, \boldsymbol{\Sigma}_{22}\right)$

Asymptotic Properties – Consistency

What happens to the estimator as the sample size increases, $N \to \infty$?

- **Consistency:** An estimator $\hat{\boldsymbol{\theta}}$ is *consistent* if the probability of the estimator being even very little off the true value in the population becomes zero, when the sample size increases:

  $$\operatorname{plim} \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$$

  For plim and a continuous function g it holds:

  $$\operatorname{plim} g\left(\hat{\boldsymbol{\theta}}\right) = g\left(\operatorname{plim} \hat{\boldsymbol{\theta}}\right)$$

- As the sample size approaches infinity the variance of a consistent estimator goes to zero and the distribution of the estimator collapses to the true parameter value

Asymptotic Properties – Asymptotic Normality

- **Asymptotic Normality:** An estimator is asymptotically normal if it *converges in distribution* to a multivariate normal:

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \mathbf{W}\right)$$

  - As the sample size increases, the distribution of $\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)$ becomes indistinguishable from a normal.
  - In such a case we will write:

$$\hat{\boldsymbol{\theta}} \overset{A}{\sim} \mathrm{N}\left(\boldsymbol{\theta}, \mathbf{W}/N\right)$$

  - We call $\mathbf{V} = \mathbf{W}/N$ the *asymptotic variance* of $\hat{\boldsymbol{\theta}}$
  - Asymptotic normality is a desirable property for the same reasons as normality.

Asymptotic Properties – Asymptotic Efficiency

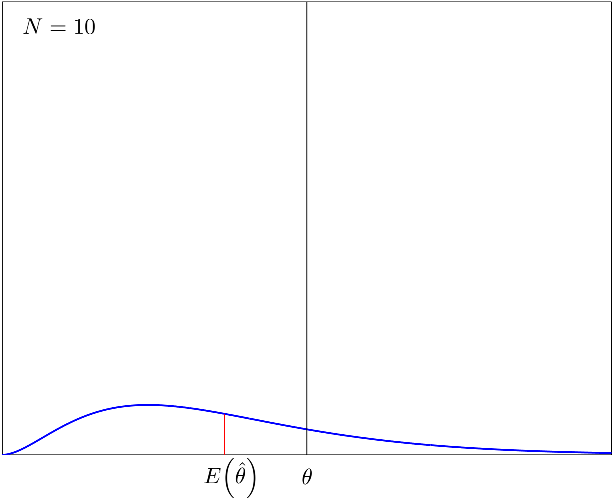
- **Asymptotic Efficiency:** A consistent and asymptotically normally distributed estimator is *asymptotically efficient* withing a class of consistent and asymptotically normally distributed estimators if it has the smallest possible asymptotic variance within this class.

Note: For an estimator to be asymptotically efficient it needs to:

- be consistent
- be asymptotically normal
- have the smallest possible variance among consistent and asymptotically normal estimators (within the reference class)

A Small Quiz

- We want to estimate a parameter θ from a population.
- We have a sample of N observations from the population.
- We consider two alternative estimators:
  - $\hat{\theta}$ which is unbiased and has positive variance
  - $\tilde{\theta}$ which is biased and has zero variance
- Is $\tilde{\theta}$ more efficient than $\hat{\theta}$?


$N = 10$
$E\left(\hat{\theta}\right)$
θ

Summary

- reviewed linear algebra and saw the OLS estimator in this notation
- defined what is an estimator and its desirable properties:

Unbiasedness $\mathrm{E}\left(\hat{\boldsymbol{\theta}}\right) = \boldsymbol{\theta}$	**Consistency** $\operatorname{plim} \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$
Normality $\hat{\boldsymbol{\theta}} \sim \mathrm{N}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}\right)$	**Asymptotic Normality** $\hat{\boldsymbol{\theta}} \overset{A}{\sim} \mathrm{N}\left(\boldsymbol{\theta}, \mathbf{W}/N\right)$
Efficiency Smallest possible variance	**Asymptotic Efficiency** Smallest possible asymptotic variance

- In what comes next we will:
  - use linear algebra extensively to communicate
  - use the properties of estimators to decide whether an estimator is "good" or not in specific situations