Introduction to Econometric Analysis

Osman DOĞAN

Outline: Econometric analysis framework

- Econometric analysis framework:
  1. What is econometrics? Why is there a field called "econometrics"?
  2. What type of questions econometric analysis can help answer?
  3. What do we mean by "causal analysis" or "causal effect"?
  4. What are experimental data and observational data? How do they differ?
  5. What are different kinds of data sets?
  6. Sum and product operators.
  7. Some useful results from Linear Algebra.
- Readings:
  1. Stock and Watson (2020, Chapter 1).
  2. Hanck et al. (2021, Chapter 1).
  3. Optional: Strang (2023) for Linear Algebra.

Econometrics

- **Four definitions**:
  (1) Econometrics is the science of testing economic theories.
  (2) Econometrics is the set of tools used for forecasting future values of economic variables, such as a firm's sales, the overall growth of the economy, or stock prices.
  (3) Econometrics is the process of fitting mathematical economic models to real-world data.
  (4) Econometrics is the science and art of using historical data to make numerical, or quantitative, policy recommendations in government and business.
- Stock and Watson (2020) give the following definition.

Definition 1

Econometrics is the science and art of using economic theory and statistical techniques to analyze economic data.

What type of questions?

- Economics suggests important relationships, often with policy implications, but generally never suggests quantitative magnitudes of causal effects.
- Consider the following questions:
  - What is the quantitative effect of reducing class size on student achievement?
  - How does another year of education change earnings?
  - What is the price elasticity of cigarettes?
  - What is the effect on output growth of a 1 percentage point increase in interest rates by the central bank?
- In order to get quantitative answers for these questions, we need to use data to estimate econometric models.

What type of questions?

- In Chapters 1 through 9, we will try to answer:

  What is the quantitative effect of reducing class size on student achievement?

- With fewer students in the classroom, hypothetically,
  - each student can get more of the teacher's attention,
  - there are fewer class disruptions, learning is enhanced,
  - and grades improve.

- But what, precisely, is the effect on elementary school education of reducing class size?

- This is important because reducing class size costs money, and the policy maker must weigh the costs such as hiring more teachers against the benefits.

- To weigh costs and benefits, however, the decision maker must have a precise quantitative understanding of the likely benefits.

What type of questions?

- There are a lot of factors that can effect student achievement. Let's take a minute here and consider some other factors (confounders) than class size that we think might be relevant and lead students to perform better on standardized tests.
- Some of these factors: Outside learning opportunities, parent's education, peer effects, income, tech resources in classrooms, demographics, and (unobserved) personal traits.
- Some of these can be observed and some cannot, and if we are lucky we might have data on the observed ones.
- Our goal is to measure the effect of class size on student achievement after controlling for the effects of all factors that can affect student achievement.
- Obviously, this will be challenging if not impossible.
- The conceptual framework used in this course is the multiple regression model.

Estimation of Causal Effects

- Econometrics is mostly about **causal** relationships among variables.

Definition 2

In common usage, an action is said to **cause** an outcome if the outcome is the direct result, or consequence, of that action. **Causality** means that a specific action (applying fertilizer) leads to a specific, measurable consequence (more tomatoes).

- For example, what is the causal effect on tomato yield (measured in kilograms) of applying a certain amount of fertilizer, say, 100 grams of fertilizer per plot?
- In general, how should we measure the causal effect of a variable on another variable?

Estimation of Causal Effects

- One way to measure the causal effect is to conduct a randomized controlled experiment.

Definition 3 (Randomized Controlled Experiment)

A randomized controlled experiment is a type of scientific experiment that aims to measure the effectiveness of some intervention (treatment or policy) through randomly allocating subjects to two or more groups, treating them differently, and then comparing them with respect to a measured response.

- Thus, the causal effect is defined to be the effect on an outcome of a given action or treatment, as measured in an ideal randomized controlled experiment (RCE). We will return to this topic in Chapter 13 of the textbook.

Prediction, Forecasting, and Causality

- In econometrics, we use regression models to measure the causal effect:

$$Y = \beta_0 + \beta_1 X + u, \tag{1}$$

where
- Y is the dependent variable and X is the independent variable,
- β_0 is intercept parameter and β_1 is the slope parameter,
- u is the error term.

Definition 4 (Estimation)

Estimation is the process of using data on Y and X to learn about the unknown coefficients β_0 and β_1.

- Thus, our goal is to use data on Y and X to estimate β_0 and β_1.
- β_1 measures the average effect of one unit increase in X on Y.
- Under which conditions, this effect can have a causal interpretation in the sense of Definition 2?
- We will explore these conditions throughout this course.

Prediction and forecasting

- Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the estimated coefficients obtained from (1).
- We can use the estimated regression model to get predicted values on the dependent variable:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

  where $\hat{Y}$ denotes the predicted values.
- Thus, **prediction** is the process of using information on some variables to make statement about the value of another variable.
- A **forecast** is a prediction about the value of a variable in the future. We will see more on this topic in the next semester.
- Prediction/forecasting does not necessitate figuring out a causal relationship.
- For example, you can predict whether it is raining by whether pedestrians are using umbrellas, but the act of using an umbrella does not cause rain.

Data

- In econometrics, data come from one of two sources: experiments or nonexperimental observations of the world.

Definition 5 (Experimental data)

Experimental data come from experiments designed to evaluate a treatment or policy or to investigate a causal effect.

Definition 6 (Observational data)

Data obtained by observing actual behavior outside an experimental setting are called observational data.

- Whether the data are experimental or observational, data sets come in three main types:
  1. Cross-sectional data
  2. Time series data
  3. Panel data

Data

- Data on different entities-workers, consumers, firms, governmental units, and so forth-for a single time period are called **cross-sectional data**.

TABLE 1.1 Selected Observations on Test Scores and Other Variables for California School Districts in 1999

Observation (District) Number	District Average Test Score (fifth grade)	Student-Teacher Ratio	Expenditure per Pupil ($)	Percentage of Students Learning English
1	690.8	17.89	$6385	0.0%
2	661.2	21.52	5099	4.6
3	643.6	18.70	5502	30.0
4	647.7	17.36	7102	0.0
5	640.8	18.67	5236	13.9
⋮	⋮	⋮	⋮	⋮
418	645.0	21.89	4403	24.3
419	672.2	20.20	4776	3.0
420	655.8	19.04	5993	5.0

Note: The California test score data set is described in Appendix 4.1.

Data

- **Time series data** are data for a single entity (person, firm, country) collected at multiple time periods.

TABLE 1.2 Selected Observations on the Growth Rate of GDP and the Term Spread in the United States: Quarterly Data, 1960:Q1–2017:Q4

Observation Number	Date (year: quarter)	GDP Growth Rate (% at an annual rate)	Term Spread (percentage points)
1	1960:Q1	8.8%	0.6
2	1960:Q2	−1.5	1.3
3	1960:Q3	1.0	1.5
4	1960:Q4	−4.9	1.6
5	1961:Q1	2.7	1.4
⋮	⋮	⋮	⋮
230	2017:Q2	3.0	1.4
231	2017:Q3	3.1	1.2
232	2017:Q4	2.5	1.2

Note: The United States GDP and term spread data set is described in Appendix 15.1.

Data

- **Panel data**, also called longitudinal data, are data for multiple entities in which each entity is observed at two or more time periods.

TABLE 1.3 Selected Observations on Cigarette Sales, Prices, and Taxes, by State and Year for U.S. States, 1985–1995

Observation Number	State	Year	Cigarette Sales (packs per capita)	Average Price per Pack (including taxes)	Total Taxes (cigarette excise tax + sales tax)
1	Alabama	1985	116.5	$1.022	$0.333
2	Arkansas	1985	128.5	1.015	0.370
3	Arizona	1985	104.5	1.086	0.362
⋮	⋮	⋮	⋮	⋮	⋮
47	West Virginia	1985	112.8	1.089	0.382
48	Wyoming	1985	129.4	0.935	0.240
49	Alabama	1986	117.2	1.080	0.334
⋮	⋮	⋮	⋮	⋮	⋮
96	Wyoming	1986	127.8	1.007	0.240
97	Alabama	1987	115.8	1.135	0.335
⋮	⋮	⋮	⋮	⋮	⋮
528	Wyoming	1995	112.2	1.585	0.360

Note: The cigarette consumption data set is described in Appendix 12.1.

Summation operator

- For ease of notation, we will use $\sum$ to denote the sum:

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + X_3 + \cdots + X_n.$$

- Suppose $n = 4$, and $X_1 = 9$, $X_2 = 7.5$, $X_3 = 13$, $X_4 = 10$. Then,

$$\sum_{i=1}^{4} X_i = X_1 + X_2 + X_3 + X_4 = 9 + 7.5 + 13 + 10 = 39.5.$$

- The simple average of these number is $\bar{X} = \frac{1}{4}\sum_{i=1}^{4} X_i = 39.5/4 = 9.875.$

Summation operator

- Some properties of the summation operator:
  - For any constant c, $\sum_{i=1}^{n} c = n \times c$.
  - For any constant c, $\sum_{i=1}^{n}(c \times X_i) = c \times (\sum_{i=1}^{n} X_i)$.
  - Let $\{X_i\}_1^n$ and $\{Y_i\}_1^n$ be two sequences of numbers and let a and b be any two constants. Then,

$$\sum_{i=1}^{n}(a \times X_i + b \times Y_i) = a \times \left(\sum_{i=1}^{n} X_i\right) + b \times \left(\sum_{i=1}^{n} Y_i\right).$$

  - The sum of ratios is not equal to the ratio of sums:

$$\sum_{i=1}^{n}(X_i/Y_i) \neq \left(\sum_{i=1}^{n} X_i\right) / \left(\sum_{i=1}^{n} Y_i\right).$$

  - In general, the sum of squares is not equal to the square of the sum:

$$\sum_{i=1}^{n} X_i^2 \neq \left(\sum_{i=1}^{n} X_i\right)^2.$$

Summation operator

- Some properties of the summation operator:
  - Let $\{X_i\}_1^n$ and $\{Y_j\}_1^m$ be two sequences of numbers and let a and b be any two constants. Then,

$$(X_1 + X_2 + \cdots + X_n)(Y_1 + Y_2 + \cdots + Y_m) = \sum_{i=1}^{n}\sum_{j=1}^{m}(X_i \times Y_j)$$

  - For example, $X_1 = 2$, $X_2 = -2$, $X_3 = 3$, $X_4 = 5$, $X_5 = -10$ and $Y_1 = 4$, $Y_2 = 2$, $Y_3 = 1$, $Y_4 = -1$, $Y_5 = 6$, $Y_6 = -3$. Then,

$$(X_1 + X_2 + \cdots + X_5)(Y_1 + Y_2 + \cdots + Y_6) = -2 \times 9 = -18$$

$$\sum_{i=1}^{5}\sum_{j=1}^{6}(X_i \times Y_j) = \sum_{i=1}^{5} X_i (Y_1 + Y_2 + \cdots + Y_6) = \sum_{i=1}^{5} X_i \times 9 = -2 \times 9 = -18.$$

Summation operator

- Some properties of the summation operator:
  - $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \implies \sum_{i=1}^{n} X_i = n\bar{X}$.
  - The order of summation does not matter for finite sums:
  $$\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i \times Y_j) = \sum_{j=1}^{m}\sum_{i=1}^{n}(X_i \times Y_j)\,.$$
  - Also:
  $$\left(\sum_{i=1}^{n} X_i\right)^2 = (X_1 + X_2 + \cdots + X_n)(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n}\sum_{j=1}^{n}(X_i \times X_j)\,.$$
  - The sum of natural numbers from 1 to n is $\sum_{j=1}^{n} j = n \times (n+1)/2$.
  - The sum of first n terms of a geometric series:
  $1 + X + X^2 + X^3 + \cdots + X^n = (1 - X^{n+1})/(1 - X)$ for $X \neq 1$.

Product operator

- We use $\prod$ for the product of the elements of sequence $\{X_i\}_{i=1}^{n}$:

$$\prod_{i=1}^{n} X_i = X_1 \times X_2 \times \cdots \times X_n.$$

- Some properties of product operator are:
  1. $\prod_{i=1}^{n} c = c^n$, where c is a constant,
  2. $\prod_{i=1}^{n} X_i^p = \left(\prod_{i=1}^{n} X_i\right)^p$, and
  3. $\prod_{i=1}^{n} X_i Y_i = \left(\prod_{i=1}^{n} X_i\right) \times \left(\prod_{i=1}^{n} Y_i\right)$.

Vectors and Matrices

- In this section, we will review some properties of vectors and matrices that are useful for econometric analysis.

Definition 7

A vector is a collection of n numbers or elements, collected either in a column (a column vector) or in a row (a row vector).

- The n-dimensional column vector $\boldsymbol{b}$ and the n-dimensional row vector $\boldsymbol{c}$ are

$$\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \text{ and } \boldsymbol{c} = \begin{pmatrix} c_1 & c_2 & \cdots & c_n \end{pmatrix}. \tag{2}$$

Definition 8

A matrix is a collection, or an array, of numbers or elements, in which the elements are laid out in columns and rows. The dimension of a matrix is $n \times m$, where n is the number of rows and m is the number of columns. The $n \times m$ matrix $\boldsymbol{A}$ is

$$\boldsymbol{A}_{m\times n} = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

where a_{ij} is the (i, j)th element and a_{ii} is the ith diagonal element.

Types of Matrices

- A matrix is said to be square if the number of rows equals the number of columns.
- A square matrix is said to be symmetric if its (i, j)th element equals its (j, i)th element.
- A diagonal matrix is a square matrix in which all the off-diagonal elements equal 0; that is, if the square matrix $\boldsymbol{A}$ is diagonal, then $a_{ij} = 0$ for $i \neq j$.
- A diagonal square matrix is called the identity matrix, if it has 1s on the diagonal: the $n \times n$ identity matrix is written as

$$\boldsymbol{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

- The null matrix, $\mathbf{0}_{n \times m}$, is the $n \times m$ matrix with all elements equal to 0.
- A square matrix having zeros as elements below the diagonal is called an upper triangular matrix. A square matrix having zeros as elements above the diagonal is called a lower triangular matrix.

Types of Matrices

- The transpose of a matrix $\boldsymbol{A}$ is obtained by interchanging its rows to columns (or columns to rows).
- Suppose $\boldsymbol{A} = \begin{pmatrix} 1 & 3 & 3 \\ 2 & 4 & 1 \end{pmatrix}$. Then, its transpose is

$$\boldsymbol{A}^{'} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 3 & 1 \end{pmatrix}$$

.

- In some textbooks, the transpose matrix of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^{T}$.
- If $\boldsymbol{a}$ is an $n \times 1$ column vector, then its transpose $\boldsymbol{a}^{'}$ is a $1 \times n$ row-vector.

Addition and Multiplication

- The sum of two matrices of the same dimensions, $\boldsymbol{C}_{m\times n} = \boldsymbol{A}_{m\times n} + \boldsymbol{B}_{m\times n}$, is the matrix with elements $c_{ij} = a_{ij} + b_{ij}$.
- The product of a scalar κ with a matrix $\boldsymbol{B}_{m\times n}$, written $\kappa \times \boldsymbol{B}_{m\times n}$, is defined as the matrix with elements $\kappa \times b_{ij}$.
- Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two $n \times 1$ column vectors. Then, $\boldsymbol{a}^{'}\boldsymbol{b} = \sum_{i=1}^{n} a_i b_i$ and $\boldsymbol{a}^{'}\boldsymbol{a} = \sum_{i=1}^{n} a_i^2$.
- The product $\boldsymbol{AB} = \boldsymbol{C}$ is defined when the row order of $\boldsymbol{B}$ matches the column order of $\boldsymbol{A}$. Then, $\boldsymbol{A}$ and $\boldsymbol{B}$ are said to be conformable for multiplication.
- If $\boldsymbol{A}$ is $m \times n$ and $\boldsymbol{B}$ is $n \times p$, then $\boldsymbol{C}$ is $m \times p$ matrix with the elements

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Addition and Multiplication

- Some properties of matrix addition and multiplication:
  1. $\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{A}$, $(\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}=\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C})$,
  2. $(\boldsymbol{A}+\boldsymbol{B})^{'}=\boldsymbol{A}^{'}+\boldsymbol{B}^{'}$,
  3. $\boldsymbol{A}_{n\times n}\boldsymbol{I}_n=\boldsymbol{A}_n$ and $\boldsymbol{A}_{n\times m}\boldsymbol{0}_{m\times p}=\boldsymbol{0}_{n\times p}$,
  4. $\boldsymbol{A}(\boldsymbol{B}\boldsymbol{C})=(\boldsymbol{A}\boldsymbol{B})\boldsymbol{C}$ and $(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C}=\boldsymbol{A}\boldsymbol{C}+\boldsymbol{B}\boldsymbol{C}$,
  5. $(\boldsymbol{A}\boldsymbol{B})^{'}=\boldsymbol{B}^{'}\boldsymbol{A}^{'}$.
- Generally, $\boldsymbol{AB}\neq\boldsymbol{BA}$. Suppose $\boldsymbol{A}=\begin{pmatrix}1&3&3\\2&4&1\end{pmatrix}$ and $\boldsymbol{B}=\begin{pmatrix}2&4\\1&5\\6&2\end{pmatrix}$.

  Let's compute $\boldsymbol{AB}$ and $\boldsymbol{BA}$.

$$\boldsymbol{AB}=\begin{pmatrix}(1\cdot 2+3\cdot 1+3\cdot 6) & (1\cdot 4+3\cdot 5+3\cdot 2)\\(2\cdot 2+4\cdot 1+1\cdot 6) & (2\cdot 4+4\cdot 5+1\cdot 2)\end{pmatrix}=\begin{pmatrix}23 & 25\\14 & 30\end{pmatrix},$$

$$\boldsymbol{BA}=\begin{pmatrix}(2\cdot 1+4\cdot 2) & (2\cdot 3+4\cdot 4) & (2\cdot 3+4\cdot 1)\\(1\cdot 1+5\cdot 2) & (1\cdot 3+5\cdot 4) & (1\cdot 3+5\cdot 1)\\(6\cdot 1+2\cdot 2) & (6\cdot 3+2\cdot 4) & (6\cdot 3+2\cdot 1)\end{pmatrix}=\begin{pmatrix}10 & 22 & 10\\11 & 23 & 8\\10 & 26 & 20\end{pmatrix}.$$

Determinant

- The determinant of a square matrix $\boldsymbol{A}_{n\times n}$ is the scalar quantity calculated from the elements of $\boldsymbol{A}$. It is usually denoted as $|\boldsymbol{A}|$ and also sometimes $\det(\boldsymbol{A})$.

- It can be calculated as

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

where $|\boldsymbol{M}_{ij}|$ (the minor of a_{ij}) is the determinant of the remaining $(n-1)\times(n-1)$ matrix when the ith row and the jth column of $\boldsymbol{A}$ are deleted.

- $C_{ij} = (-1)^{i+j}|\boldsymbol{M}_{ij}|$ is called the cofactor of a_{ij}.

Determinant

- Suppose $\boldsymbol{A}_{2\times 2}$. Then, $|\boldsymbol{A}| = a_{11}a_{22} - a_{12}a_{21}$.
- If $\boldsymbol{A}_{3\times 3}$ and holding column $j = 1$ fixed

$$C_{11} = (-1)^{1+1} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix},$$

$$C_{21} = (-1)^{2+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix},$$

$$C_{31} = (-1)^{3+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix},$$

  and $|\boldsymbol{A}| = a_{11}C_{11} + a_{21}C_{21} + a_{31}C_{31}$.
- A square matrix $\boldsymbol{A}$ is called singular if $|\boldsymbol{A}| = 0$. Otherwise, it is said to be nonsingular.

Properties of determinant

- Let $\boldsymbol{A}_{n\times n}$ and $\boldsymbol{B}_{n\times n}$ be square matrices, and c be a scalar. Some properties of the determinant are

  1. $|\boldsymbol{A}^{'}| = |\boldsymbol{A}|$,
  2. $|c\boldsymbol{A}| = c^n|\boldsymbol{A}|$,
  3. $|\boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{A}||\boldsymbol{B}|$,
  4. $|\boldsymbol{A}^p| = |\boldsymbol{A}|^p$,
  5. If $\boldsymbol{A}$ is diagonal or triangular, then $|\boldsymbol{A}| = \prod_{i=1}^{n} a_{ii}$.

Properties of inverse matrix

- Nonsingularity of $\boldsymbol{A}$ implies that there exists a unique $n \times n$ matrix $\boldsymbol{A}^{-1}$ such that
$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n.$$
- $\boldsymbol{A}^{-1}$ is called the inverse of $\boldsymbol{A}$. Suppose $\boldsymbol{A}_{2\times 2}$ is nonsingular. Then,
$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$
- Let $\boldsymbol{A}_{n\times n}$ and $\boldsymbol{B}_{n\times n}$ be nonsingular square matrices, and c be a scalar. Some properties of the inverse operator are as follow.
  1. $(c\boldsymbol{A})^{-1} = c^{-1}\boldsymbol{A}^{-1}$.
  2. $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.
  3. $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$.
  4. $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$.
  5. $|\boldsymbol{A}^{-1}| = |\boldsymbol{A}|^{-1}$.

Linear independence and rank

- A set of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are linearly independent if and only if the only solution to $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_n\boldsymbol{x}_n = 0$ is $c_1 = c_2 = \cdots = c_n = 0$.
- Consider $\boldsymbol{x}_1 = (3,1)^{'}$ and $\boldsymbol{x}_2 = (1,2)^{'}$. These vectors are linearly independent because

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 = \mathbf{0} \implies 3c_1 + c_2 = 0, \quad c_1 + 2c_2 = 0 \implies c_1 = c_2 = 0.$$

- The rank of $\mathbf{A}_{m\times n}$ is the number of linearly independent rows or columns, and is denoted by $rank(\mathbf{A})$. If $rank(\mathbf{A}) = n$, we say that $\mathbf{A}$ is full column rank.
- Some properties of the rank operator are
  1. $0 \leq rank(\boldsymbol{A}) \leq \min(m,n)$,
  2. $rank(\boldsymbol{A}) = rank(\boldsymbol{A}^{'})$,
  3. $rank(\boldsymbol{A}) = rank(\boldsymbol{A}^{'}\boldsymbol{A}) = rank(\boldsymbol{A}\boldsymbol{A}^{'})$.

Norms

- The typical norm used for Euclidean space $\mathbb{R}^n$ is the Euclidean norm:

$$\|\boldsymbol{a}\| = (\boldsymbol{a}^{'}\boldsymbol{a})^{1/2} = \left(\sum_{i=1}^{n} a_i^2\right)^{1/2}.$$

- An alternative norm is the p-norm (for $p \geq 1$):

$$\|\boldsymbol{a}\|_p = \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p}.$$

Matrix Calculus

Definition 9

Let $f:\mathbb{R}^n\to\mathbb{R}$. Then, we define

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}=\begin{pmatrix}\frac{\partial f(\boldsymbol{x})}{\partial x_1}\\ \vdots\\ \frac{\partial f(\boldsymbol{x})}{\partial x_n}\end{pmatrix}_{n\times 1}\quad\text{and}\quad \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}'}=\begin{pmatrix}\frac{\partial f(\boldsymbol{x})}{\partial x_1} & \dots & \frac{\partial f(\boldsymbol{x})}{\partial x_n}\end{pmatrix}_{1\times n}. \tag{3}$$

- The vector of partial derivatives $\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}$ is called the gradient of f. In Definition 9, note that

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}'}=\left(\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)'.$$

Matrix Calculus

- In the special case where $f(\boldsymbol{x}) = \boldsymbol{a}^{'}\boldsymbol{x}$, we have

$$\frac{\partial \boldsymbol{a}^{'}\boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial \boldsymbol{a}^{'}\boldsymbol{x}}{\partial x_1} \\ \vdots \\ \frac{\partial \boldsymbol{a}^{'}\boldsymbol{x}}{\partial x_n} \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \boldsymbol{a}, \tag{4}$$

$$\frac{\partial \boldsymbol{a}^{'}\boldsymbol{x}}{\partial \boldsymbol{x}^{'}} = \begin{pmatrix} \frac{\partial \boldsymbol{a}^{'}\boldsymbol{x}}{\partial x_1} & \ldots & \frac{\partial \boldsymbol{a}^{'}\boldsymbol{x}}{\partial x_n} \end{pmatrix} = \begin{pmatrix} a_1 & \ldots & a_n \end{pmatrix} = \boldsymbol{a}^{'}. \tag{5}$$

- Since $\boldsymbol{a}^{'}\boldsymbol{x} = \boldsymbol{x}^{'}\boldsymbol{a}$, we also have $\frac{\partial \boldsymbol{a}^{'}\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}^{'}\boldsymbol{a}}{\partial \boldsymbol{x}} = \boldsymbol{a}$.

- Similarly, $\frac{\partial \boldsymbol{a}^{'}\boldsymbol{x}}{\partial \boldsymbol{x}^{'}} = \frac{\partial \boldsymbol{x}^{'}\boldsymbol{a}}{\partial \boldsymbol{x}^{'}} = \boldsymbol{a}^{'}$.

Matrix Calculus

- Let $\boldsymbol{A}$ be a $m \times n$ matrix, $\boldsymbol{x}$ be an $n \times 1$ vector, and $\boldsymbol{z}$ be an $n \times 1$ vector. Then,

$$\frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}^{'}} = \boldsymbol{A} \text{ and } \frac{\partial \boldsymbol{x}^{'}\boldsymbol{A}^{'}}{\partial \boldsymbol{x}} = \boldsymbol{A}^{'}. \tag{6}$$

- Consider $\frac{\partial(\boldsymbol{z}^{'}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}}$. Let $\boldsymbol{c} = \boldsymbol{A}^{'}\boldsymbol{z}$. Then,

$$\frac{\partial(\boldsymbol{z}^{'}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial(\boldsymbol{c}^{'}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{c} = \boldsymbol{A}^{'}\boldsymbol{z}, \tag{7}$$

where the second equality follows from (4).

- Consider $\frac{\partial(\boldsymbol{z}^{'}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{z}}$. Let $\boldsymbol{d} = \boldsymbol{A}\boldsymbol{x}$, then

$$\frac{\partial(\boldsymbol{z}^{'}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{z}} = \frac{\partial(\boldsymbol{z}^{'}\boldsymbol{d})}{\partial \boldsymbol{z}} = \boldsymbol{d} = \boldsymbol{A}\boldsymbol{x}, \tag{8}$$

where the second equality follows from (4).

Matrix Calculus

- Let $\boldsymbol{z}(\boldsymbol{\alpha}) \in \mathbb{R}^m$ and $\boldsymbol{x}(\boldsymbol{\alpha}) \in \mathbb{R}^n$ be functions of $\boldsymbol{\alpha} \in \mathbb{R}^r$. Let $\boldsymbol{A}$ be an $m \times n$ matrix of constant. Then

$$\text{Product rule:} \quad \frac{\boldsymbol{z}'(\boldsymbol{\alpha})\boldsymbol{A}\boldsymbol{x}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \frac{\partial \boldsymbol{x}'(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\boldsymbol{A}'\boldsymbol{z}(\boldsymbol{\alpha}) + \frac{\partial \boldsymbol{z}'(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\boldsymbol{A}\boldsymbol{x}(\boldsymbol{\alpha}). \tag{9}$$

- Consider $\frac{\partial(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}}$. Applying the product rule, we obtain

$$\frac{\partial(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}'}{\partial \boldsymbol{x}}\boldsymbol{A}'\boldsymbol{x} + \frac{\partial \boldsymbol{x}'}{\partial \boldsymbol{x}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}'\boldsymbol{x} + \boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}' + \boldsymbol{A})\boldsymbol{x}. \tag{10}$$

- If $\boldsymbol{A}$ is symmetric in the quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, then

$$\frac{\partial(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}. \tag{11}$$

Bibliography I

Hanck, Christoph et al. (2021). *Introduction to Econometrics with R*. URL: https://www.econometrics-with-r.org/index.html.

Stock, James H. and Mark W. Watson (2020). *Introduction to Econometrics*. Fourth. Pearson.

Strang, Gilbert (2023). *Introduction to Linear Algebra*. Sixth edition. Wellesley-Cambridge Press.