Advanced Econometrics
01 Review of Matrix Algebra

Eduard Brüll

Fall 2025

Course Structure

Main Lectures

- Tue & Thu 10:15–11:45
  0129 (Tue), 0048 (Thu)
  Dates: Oct 7 – Dec 4, 2025

Computer Sessions

- Wed 15:30–17:00
  L7, 3–5, Room 358
  Dates: Oct 8 – Dec 3, 2025

Exercises / Tutorials

- Fri 10:15–11:45
  0048 (Schloss Ostflügel)
  Dates: Oct 10 – Dec 5, 2025
- **Collect bonus points:** Hand-in exercises (counts for grade in final exam)
- Effort counts

Assessment, Materials & Instructors

Assessment

- Written exam in late Dec 2025
  If everything works out: 16.12.2025
  Second exam: early Feb 2026

Materials

- **Main reference:** Greene, Econometric Analysis, 7th edition
- Additional readings and datasets available on ILIAS

Instructors

- Eduard Brüll — `Eduard.Bruell@zew.de`
- Mishelle Segui — `mishelle.segui@uni-mannheim.de`
- Jan Kemper — `jan.kemper@zew.de`

Overview of Topics

Review and Foundations

- Review of Matrix Algebra
- Probability and Distribution Theory

Estimation Techniques

- Linear Regression Model and OLS
- Maximum Likelihood Estimation
- General Method of Moments (GMM)

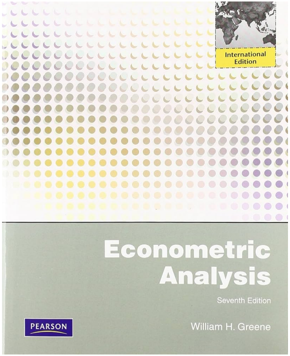
Applications and Extensions

- Time Series Models
- Models for Panel Data
- Difference-in-Differences and Event Studies

Remarks on Readings

Main reference:

Greene, Econometric Analysis
7th edition



- Lecture serves as a guideline. The order of topics may differ from Greene.
- Greene is a modular book.
  - + **Advantage:** Excellent reference.
  - − **Downside:** Dense; more detail than one can digest on a first read.
- Get a "second opinion" from Greene on each topic covered in lecture.
- Take note of material not discussed in class. It helps to know what you don't know.
- Lectures emphasize the essential and sometimes "ugly" parts. Don't skip the "nice" parts in Greene: Introductions, examples, and context matter.

Lecture 1

Advanced Econometrics

What is Econometrics?

"Three viewpoints, that of statistics, economic theory, and mathematics, [are] a necessary, but not by [themselve] a sufficient, condition for a real understanding of the quantitative relations in modern economic life. It is the unification of all three that is powerful. And it is this unification that constitutes econometrics."

First issue of Econometrica, Ragnar Frisch (1933)

... and most recently of machine learning

What is Econometrics?

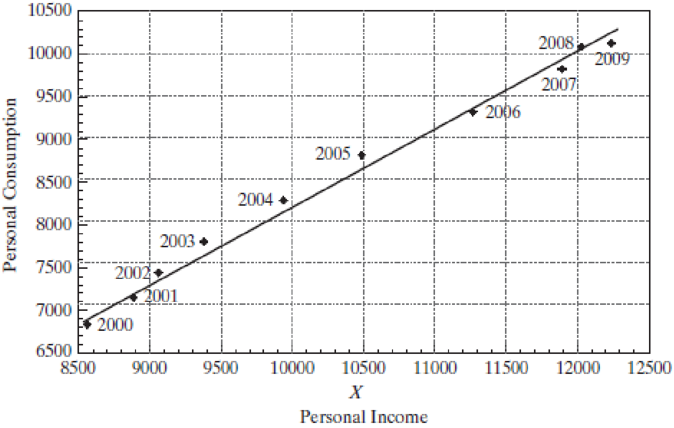
An economic model of consumption: **Keynes's Consumption Function**

- We shall therefore define what we shall call the propensity to consume as the functional relationship f between X, a given level of income, and C, the expenditure on consumption out of the level of income, so that $C = f(X)$.
- The fundamental psychological law upon which we are entitled to depend with great confidence, both a priori from our knowledge of human nature and from the detailed facts of experience, is that men are disposed, as a rule and on the average, to increase their consumption as their income increases, but not by as much as the increase in their income. That is, . . . $dC/dX = \beta$ is positive and less than unity.

An econometric model of consumption

- $C = \alpha + X\beta$
- $0 < \beta < 1, \alpha > 0$

Keynes's Consumption Function



What is Econometrics?

Contributions often awarded

- Ragnar Frisch in 1969, Lawrence Klein in 1980,
- Trygve Haavelmo in 1989,
- James Heckman and Daniel McFadden in 2000, and
- Robert Engle and Clive Granger in 2003
- Lars Peter Hansen (with R. Shiller and E. Fama) in 2013

...and in 2021 Card, Angrist und Imbens:

What is Econometrics?

Two specializations

- Theoretical econometrics
- Applied econometrics

Special focus

- Macroeconometrics
- Forecasting
- Microeconometrics
  (Causal inference, cross sectional, panel/longitudinal analysis)

...tools heavily used in academia and by practitioners.

1.1: Relationships Between Two Variables

The multiple linear regression model assumes a linear (in parameters) relationship between a dependent variable y_i and a set of explanatory variables $x_{i0}, x_{i1}, \ldots, x_{iK}$.
x_{ik} is also called

- an independent variable,
- a covariate or
- a regressor.

The first regressor $x_{i0} = 1$ is a constant unless otherwise specified.

Consider a sample of N observations on individuals $i = 1, \ldots, N$. Every single observation i follows in the multiple regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK} + u_i,$$

where $\beta_0, \beta_1, \ldots, \beta_K$ are $K+1$ parameters and u_i is called the error term.

In more concise matrix notation

$$\underset{N\times 1}{\boldsymbol{y}} = \underset{N\times(K+1)}{\boldsymbol{X}} \underset{(K+1)\times 1}{\boldsymbol{\beta}} + \underset{N\times 1}{\boldsymbol{u}}$$

The bivariate regression model is a special case with only one regressor:

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

1.2: Matrix Fundamentals

Matrix fundamentals

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

- A matrix is a rectangular array of numbers.
- Size: (rows)$\times$(columns). E.g. the size of **A** is $2\times$ 3.
- The size of a matrix is also known as the dimension.
- The element in the *i*th row and *j*th column of **A** is referred to as a_{ij}.
- The matrix **A** can also be written as $\mathbf{A} = (a_{ij})$.

Matrix addition and subtraction

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}; \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

Matrix Addition and Subtraction

- Dimensions must match:
  $(r \times c) \pm (r \times c) \implies (r \times c)$
- **A** and **B** are both 2× 3 matrices, so

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & a_{13} + b_{13} \\ a_{21} + b_{21} & a_{22} + b_{22} & a_{23} + b_{23} \end{bmatrix}$$

- More generally we can write:

$$\mathbf{A} \pm \mathbf{B} = (a_{ij}) \pm (b_{ij})$$

Matrix multiplication

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}; \mathbf{D} = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \\ d_{31} & d_{32} \end{bmatrix}$$

Matrix Multiplication

- Inner dimensions need to match:

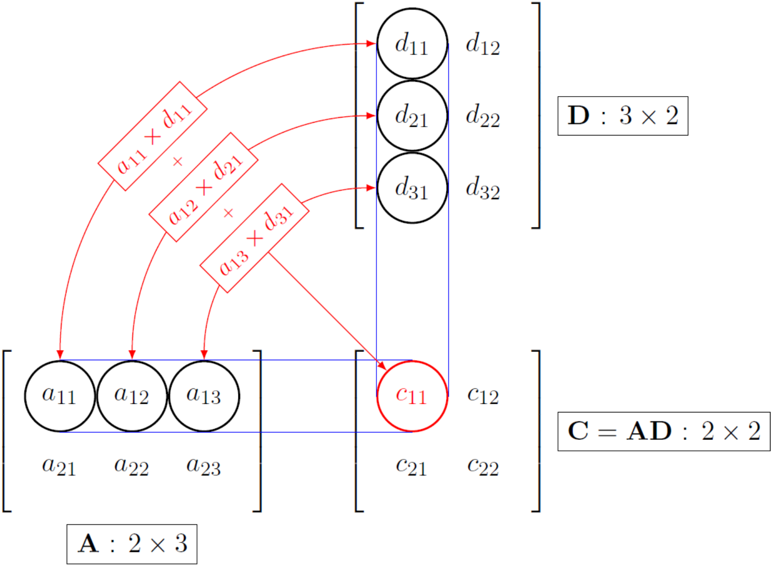
  $$(r \times c) \times (c \times p) \implies (r \times p)$$

- **A** is a 2×3 and **D** is a 3×2 matrix, so the inner dimensions match and we have: $\mathbf{C} = \mathbf{A} \times \mathbf{D} =$

  $$\begin{bmatrix} a_{11}d_{11} + a_{12}d_{21} + a_{13}d_{31} & a_{11}d_{12} + a_{12}d_{22} + a_{13}d_{32} \\ a_{21}d_{11} + a_{22}d_{21} + a_{23}d_{31} & a_{21}d_{12} + a_{22}d_{22} + a_{23}d_{32} \end{bmatrix}$$

- Look at the pattern in the terms above.

Matrix multiplication



Identity matrix

- An identity matrix is the matrix analogue of the number 1.
- If you multiply any matrix (or vector) with a conformable identity matrix the result will be the same matrix (or vector).

Example for a 2×2 matrix

.

$$\mathbf{AI} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} \times 1 + a_{12} \times 0 & a_{11} \times 0 + a_{12} \times 1 \\ a_{21} \times 1 + a_{22} \times 0 & a_{21} \times 0 + a_{22} \times 1 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \mathbf{A}.$$

Vectors

Vectors are matrices with only one row or column. For example, the column vector:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}; \qquad \mathbf{i} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Transpose Operator

. Turns columns into rows (and vice versa):

$$\mathbf{x}' = \mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$$

Sidenote: Sum of Values and Squares

.

$$\mathbf{i}'\mathbf{x} = \sum_{i=1}^{n} x_i; \qquad \frac{1}{n}\mathbf{i}'\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}; \qquad \mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2$$

Transpose

Say we have some $m \times n$ matrix:

$$\mathbf{A} = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

Transpose Operator

- Flips the rows and columns of a matrix:

$$\mathbf{A}' = (a_{ji})$$

- The subscripts gets swapped.
- $\mathbf{A}'$ is a $n \times m$ matrix: the columns in $\mathbf{A}$ are the rows in $\mathbf{A}'$.

Symmetry

Square Matrix

A matrix, **P** is square if it has the same number of rows as columns. I.e.

$$dim(\mathbf{P}) = n \times n$$

for some $n \geq 1$.

Symmetric Matrix

. A square matrix, **P** is symmetric if it is equal to its transpose:

$$\mathbf{P} = \mathbf{P}'$$

Determinant

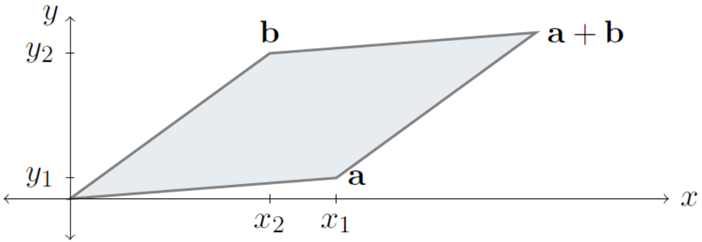
The **determinant** $\det(A)$ of a square matrix A is a scalar computed from its entries that measures how the associated linear map scales volume and whether it preserves or flips orientation.

- It is zero exactly when the matrix is not invertible.
- $\det(A) \neq 0 \iff A$ is invertible $\iff$ rows/columns are linearly independent.
- Row/column swap $\Rightarrow$ sign flips; scaling a row/column by $k \Rightarrow$ determinant scales by k.
- Adding a multiple of one row/column to another does not change the determinant.

Geometric intuition: Determinant as an area

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}.$$

- For a 2×2 matrix, $\det(A)$ is the oriented area of the parallelogram with vertices at $\mathbf{0} = (0, 0), \mathbf{a} = (x_1, y_1), \mathbf{a} + \mathbf{b} = (x_1 + x_2, y_1 + y_2)$, and $\mathbf{b} = (x_2, y_2)$.



Basic computational definition of a determinant

Cofactor expansion (basic idea)

- Let $\mathbf{C} = (c_{ij})$ be an $n \times n$ square matrix.
- Define a cofactor matrix, C_{ij}, be the determinant of the square matrix of order $(n-1)$ obtained from $\mathbf{C}$ by removing row i and column j multiplied by $(-1)^{i+j}$.
- For fixed i, i.e. focusing on one row: $\det(\mathbf{C}) = \sum_{j=1}^{N} c_{ij} C_{ij}$.
- For fixed j, i.e. focusing on one column: $\det(\mathbf{C}) = \sum_{i=1}^{N} c_{ij} C_{ij}$.
- Note that this is a recursive formula.
- The trick is to pick a row (or column) with a lot of zeros (or better yet, use a computer)!

2×2 Determinant

Apply the general formula to a 2×2 matrix: $\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$.

- Keep the first row fixed, i.e. set $i = 1$.
- General formula when $i = 1$ and $N = 2$: $\det(\mathbf{C}) = \sum_{j=1}^{2} c_{1j} C_{1j}$
- When $j = 1$, C_{11} is one cofactor matrix of **C**, i.e. the determinant after removing the first row and first column of **C** multiplied by $(-1)^{i+j} = (-1)^2$. So

$$C_{11} = (-1)^2 det(c_{22}) = c_{22}$$

  as c_{22} is a scalar and the determinant of a scalar is itself.
- $C_{12} = (-1)^3 det(c_{21}) = -c_{21}$ as c_{21} is a scalar and the determinant of a scalar is itself.
- Put it all together and you get the familiar result:

$$det(C) = c_{11}C_{11} + c_{12}C_{12} = c_{11}c_{22} - c_{12}c_{21}$$

3×3 Determinant

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}.$$

- Keep the first row fixed ($i = 1$):

$$\det(\mathbf{B}) = b_{11}B_{11} + b_{12}B_{12} + b_{13}B_{13}.$$

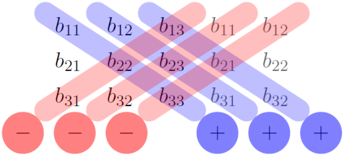
- Example: $B_{12} = (-1)^{1+2} \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix} = - \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix}.$
- $\det(\mathbf{B}) = b_{11} \begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix} - b_{12} \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix} + b_{13} \begin{vmatrix} b_{21} & b_{22} \\ b_{31} & b_{32} \end{vmatrix}.$

Sarrus' scheme for the determinant of a 3×3

- French mathematician: Pierre Frédéric Sarrus (1798–1861)

$$\det(\mathbf{B}) = \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} = b_{11} \begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix} - b_{12} \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix} + b_{13} \begin{vmatrix} b_{21} & b_{22} \\ b_{31} & b_{32} \end{vmatrix}$$

$$= (b_{11}b_{22}b_{33} + b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32}) - (b_{13}b_{22}b_{31} + b_{11}b_{23}b_{32} + b_{12}b_{21}b_{33})$$



Write the first two columns of the matrix again to the right of the original matrix. Multiply the diagonals together and then add or subtract.

Inverting Matrices

Inverse

- Requires a square matrix i.e. dimensions: $r \times r$
- For a 2×2 matrix, $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$,

$$\mathbf{A}^{-1} = \frac{1}{det(\mathbf{A})} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

- More generally, a square matrix $\mathbf{A}$ is invertible or nonsingular if there exists another matrix $\mathbf{B}$ such that

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}$$

- If this occurs then $\mathbf{B}$ is uniquely determined by $\mathbf{A}$ and is denoted $\mathbf{A}^{-1}$, i.e. $\mathbf{AA}^{-1} = \mathbf{I}$.

Rank

- The rank of a matrix **A** is the maximal number of linearly independent rows or columns of **A**.
- A family of vectors is linearly independent if none of them can be written as a linear combination of finitely many other vectors in the collection.

Example: A tyical dummy variable trap

.

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ are independent but $\mathbf{v}_4 = \mathbf{v}_1 - \mathbf{v}_2 - \mathbf{v}_3$.

Rank

- The maximum rank of an $m \times n$ matrix is $\min(m, n)$.
- A full rank matrix is one that has the largest possible rank, i.e. the rank is equal to either the number of rows or columns (whichever is smaller).
- In the case of an $n \times n$ square matrix $\mathbf{A}$, then $\mathbf{A}$ is invertible if and only if $\mathbf{A}$ has rank n (that is, $\mathbf{A}$ has full rank).
- For some $n \times k$ matrix with $k \leq n$, $\mathbf{X}$, $\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}'\mathbf{X})$
- If you use dummy variables, you need to drop one of the dummy categories otherwise $\mathbf{X}$ is not of full rank and therefore you cannot find the inverse of $\mathbf{X}'\mathbf{X}$. This is why the dummy variable trap exists.

Trace

Trace of a Matrix

The trace of an $n \times n$ matrix **A** is the sum of the elements on the main diagonal: $\text{tr}(A) = a_{11} + a_{22} + \ldots + a_{nn} = \sum_{i=1}^{n} a_{ii}$.

- $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$
- $\text{tr}(c\mathbf{A}) = c\,\text{tr}(\mathbf{A})$
- If **A** is an $m \times n$ matrix and **B** is an $n \times m$ matrix then

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$$

- More generally, for conformable matrices:

$$\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CBA}) = \text{tr}(\mathbf{BCA})$$

  BUT: $\text{tr}(\mathbf{ABC}) \neq \text{tr}(\mathbf{ACB})$. You can only move from the front to the back (or back to the front)!

Idempotent

Idempotent

A square matrix, $\mathbf{P}$ is idempotent if when multiplied by itself, yields itself. I.e.

$$\mathbf{PP} = \mathbf{P}.$$

1. When an idempotent matrix is subtracted from the identity matrix, the result is also idempotent, i.e. $\mathbf{M} = \mathbf{I} - \mathbf{P}$ is idempotent.
2. The trace of an idempotent matrix is equal to the rank.
3. $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is an idempotent matrix.

Order of operations

- Matrix multiplication is non-commutative, i.e. the order of multiplication is important: $\mathbf{AB} \neq \mathbf{BA}$.
- Matrix multiplication is associative, i.e. as long as the order stays the same, $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.
- $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
- $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$

Example: Order

. Let $\mathbf{A}$ be a $k \times k$ matrix and $\mathbf{x}$ and $\mathbf{c}$ be $k \times 1$ vectors:

$$\begin{aligned} \mathbf{Ax} &= \mathbf{c} \\ \mathbf{A}^{-1}\mathbf{Ax} &= \mathbf{A}^{-1}\mathbf{c} \qquad \text{(PRE-multiply both sides by } \mathbf{A}^{-1}\text{)} \\ \mathbf{Ix} &= \mathbf{A}^{-1}\mathbf{c} \\ \mathbf{x} &= \mathbf{A}^{-1}\mathbf{c} \end{aligned}$$

Note: $\mathbf{A}^{-1}\mathbf{c} \neq \mathbf{cA}^{-1}$

Matrix Differentiation

If $\boldsymbol{\beta}$ and $\mathbf{a}$ are both $k \times 1$ vectors, then

$$\frac{\partial \boldsymbol{\beta}'\mathbf{a}}{\partial \boldsymbol{\beta}} = \mathbf{a}.$$

$$\boldsymbol{\beta}' = \begin{pmatrix} \beta_1 & \beta_2 & \ldots & \beta_k \end{pmatrix}; \mathbf{a}' = \begin{pmatrix} a_1 & a_2 & \ldots & a_k \end{pmatrix}$$

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}'\mathbf{a}) &= \frac{\partial}{\partial \boldsymbol{\beta}}(\beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_k a_k) \\
&= \begin{bmatrix} \frac{\partial}{\partial \beta_1}(\beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_k a_k) \\ \frac{\partial}{\partial \beta_2}(\beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_k a_k) \\ \vdots \\ \frac{\partial}{\partial \beta_k}(\beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_k a_k) \end{bmatrix} \\
&= \mathbf{a}.
\end{aligned}$$

Matrix Differentiation

Let β be a $k \times 1$ vector and $\mathbf{A}$ be a $k \times k$ symmetric matrix, then

$$\frac{\partial \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = 2\mathbf{A}\boldsymbol{\beta}.$$

Say $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, then

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta}) &= \frac{\partial}{\partial \boldsymbol{\beta}}(\beta_1^2 a_{11} + 2a_{12}\beta_1\beta_2 + \beta_2^2 a_{22}) \\
&= \begin{bmatrix} \frac{\partial}{\partial \beta_1}(\beta_1^2 a_{11} + 2a_{12}\beta_1\beta_2 + \beta_2^2 a_{22}) \\ \frac{\partial}{\partial \beta_2}(\beta_1^2 a_{11} + 2a_{12}\beta_1\beta_2 + \beta_2^2 a_{22}) \end{bmatrix} \\
&= \begin{bmatrix} 2\beta_1 a_{11} + 2a_{12}\beta_2 \\ 2\beta_1 a_{12} + 2a_{22}\beta_2 \end{bmatrix} \\
&= 2\mathbf{A}\boldsymbol{\beta}.
\end{aligned}$$

Matrix Differentiation

Let $\boldsymbol{\beta}$ be a $k \times 1$ vector and $\mathbf{A}$ be a $n \times k$ matrix, then

$$\frac{\partial \mathbf{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}'} = \mathbf{A}.$$

Say $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, then

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}'}(\mathbf{A}\boldsymbol{\beta}) &= \frac{\partial}{\partial \boldsymbol{\beta}'} \begin{bmatrix} a_{11}\beta_1 + a_{12}\beta_2 \\ a_{21}\beta_1 + a_{22}\beta_2 \end{bmatrix} \\
&= \begin{bmatrix} \begin{bmatrix} \frac{\partial}{\partial \beta_1} & \frac{\partial}{\partial \beta_2} \end{bmatrix} (a_{11}\beta_1 + a_{12}\beta_2) \\ \begin{bmatrix} \frac{\partial}{\partial \beta_1} & \frac{\partial}{\partial \beta_2} \end{bmatrix} (a_{21}\beta_1 + a_{22}\beta_2) \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial}{\partial \beta_1}(a_{11}\beta_1 + a_{12}\beta_2) & \frac{\partial}{\partial \beta_2}(a_{11}\beta_1 + a_{12}\beta_2) \\ \frac{\partial}{\partial \beta_1}(a_{21}\beta_1 + a_{22}\beta_2) & \frac{\partial}{\partial \beta_2}(a_{21}\beta_1 + a_{22}\beta_2) \end{bmatrix} \\
&= \mathbf{A}.
\end{aligned}$$

Eigenvalues

- An eigenvalue λ and an eigenvector $\mathbf{x} \neq 0$ of a square matrix $\mathbf{A}$ is defined as

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}.$$

- Since the eigenvector $\mathbf{x}$ is different from the zero vector (i.e. $\mathbf{x} \neq 0$) the following is valid:

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0} \rightarrow \det(\mathbf{A} - \lambda \mathbf{I}) = 0.$$

- We know $\det(\mathbf{A} - \lambda \mathbf{I}) = \mathbf{0}$ because:
- if $(\mathbf{A} - \lambda \mathbf{I})^{-1}$ existed, we could just pre-multiply both sides by $(\mathbf{A} - \lambda \mathbf{I})^{-1}$ and get the solution $\mathbf{x} = \mathbf{0}$.
- but we have assumed $\mathbf{x} \neq 0$ so we require that $(\mathbf{A} - \lambda \mathbf{I})$ is NOT invertible which implies[1] that $\det(\mathbf{A} - \lambda \mathbf{I}) = \mathbf{0}$.
- To find the eigenvalues, we can solve $\det(\mathbf{A} - \lambda \mathbf{I}) = \mathbf{0}$.

[1] A matrix is invertible if and only if the determinant is non-zero.

Eigenvalues

Example: Finding eignevalues

. Say $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. We can find the eigenvalues of $\mathbf{A}$ by solving

$$\begin{aligned} \det(\mathbf{A} - \lambda \mathbf{I}) &= 0 \\ \det\left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) &= 0 \\ \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} &= 0 \\ (2-\lambda)(2-\lambda) - 1 \times 1 &= 0 \\ \lambda^2 - 4\lambda + 3 &= 0 \\ (\lambda - 1)(\lambda - 3) &= 0 \end{aligned}$$

The eigenvalues are the roots of this quadratic: $\lambda_1 = 1$ and $\lambda_2 = 3$.

Why do we care about eigenvalues?

- An $n \times n$ matrix $\mathbf{A}$ is positive definite if all eigenvalues of $\mathbf{A}$, $\lambda_1, \lambda_2, \ldots, \lambda_n$ are positive.
- A matrix is negative-definite, negative-semidefinite, or positive-semidefinite if and only if all of its eigenvalues are negative, non-positive, or non-negative, respectively.
- The eigenvectors corresponding to different eigenvalues are linearly independent. So if a $n \times n$ matrix has n nonzero eigenvalues, it is of full rank.
- The trace of a matrix is the sum of the eigenvectors: $\text{tr}(\mathbf{A}) = \lambda_1 + \lambda_2 + \ldots + \lambda_n$.
- The determinant of a matrix is the product of the eigenvectors: $\det(\mathbf{A}) = \lambda_1 \lambda_2 \ldots \lambda_n$.
- The eigenvectors and eigenvalues of the covariance matrix of a data set data are also used in principal component analysis (similar to factor analysis).

1.3: Quadratic Forms & Definiteness

Quadratic forms

- A quadratic form on $\mathbb{R}^n$ is a real-valued function of the form

$$Q(x_1, \dots, x_n) = \sum_{i \leq j} a_{ij} x_i x_j.$$

- E.g. in $\mathbb{R}^2$ we have $Q(x_1, x_2) = a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2$.
- Quadratic forms can be represented by a symmetric matrix $\boldsymbol{A}$ such that:

$$Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$$

- E.g. if $\boldsymbol{x} = (x_1, x_2)'$, then

$$\begin{aligned} Q(\boldsymbol{x}) &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} \\ \frac{1}{2}a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= a_{11}x_1^2 + \frac{1}{2}(a_{12} + a_{21})x_1x_2 + a_{22}x_2^2 \end{aligned} \tag{1}$$

but $\boldsymbol{A}$ is symmetric, i.e. $a_{12} = a_{21}$, so we can write,

Quadratic forms

If $\boldsymbol{x} \in \mathbb{R}^3$, i.e. $\boldsymbol{x} = (x_1, x_2, x_3)'$, then the general three dimensional quadratic form is:

$$\begin{aligned} Q(\boldsymbol{x}) &= \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \\ &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} & \frac{1}{2}a_{13} \\ \frac{1}{2}a_{12} & a_{22} & \frac{1}{2}a_{23} \\ \frac{1}{2}a_{13} & \frac{1}{2}a_{23} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 \end{aligned} \tag{2}$$

Quadratic Forms and Sum of Squares

Recall sums of squares can be written as $\boldsymbol{x}'\boldsymbol{x}$ and quadratic forms are $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$. Quadratic forms are like generalised and weighted sum of squares. Note that if $\boldsymbol{A} = \boldsymbol{I}$ then we recover the sums of squares exactly.

Definiteness of quadratic forms

- A quadratic form always takes on the value zero at the point $\boldsymbol{x} = \mathbf{0}$. This is not an interesting result!
- For example, if $\boldsymbol{x} \in \mathbb{R}$, i.e. $\boldsymbol{x} = x_1$ then the general quadratic form is ax_1^2 which equals zero when $x_1 = 0$.
- Its distinguishing characteristic is the set of values it takes when $\boldsymbol{x} \neq \mathbf{0}$.
- We want to know if $\boldsymbol{x} = \mathbf{0}$ is a max, min or neither.
- Example: when $\boldsymbol{x} \in \mathbb{R}$, i.e. the quadratic form is ax_1^2,

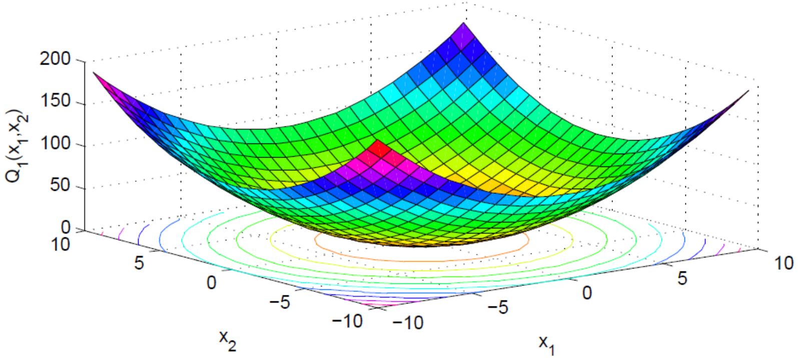
  $a > 0$ means $ax^2 \geq 0$ and equals 0 only when $x = 0$. Such a form is called positive definite; $x = 0$ is a global minimiser.

  $a < 0$ means $ax^2 \leq 0$ and equals 0 only when $x = 0$. Such a form is called negative definite; $x = 0$ is a global maximiser.

Positive definite

If $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ then $Q_1(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 + x_2^2$.

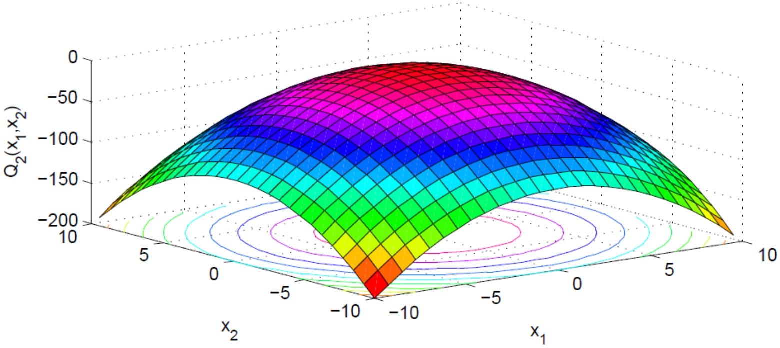
- Q_1 is greater than zero at $\boldsymbol{x} \neq \mathbf{0}$ i.e. $(x_1, x_2) \neq (0, 0)$.
- The point $\boldsymbol{x} = \mathbf{0}$ is a global minimum.
- Q_1 is called positive definite.



Negative definite

If $\mathbf{A} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ then $Q_2(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = -x_1^2 - x_2^2$.

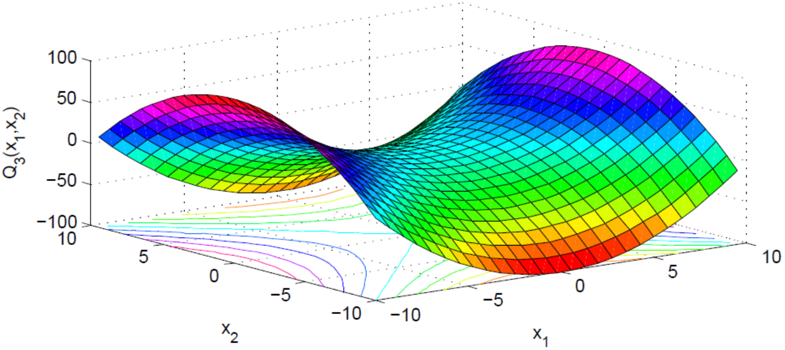
- Q_2 is less than zero at $\boldsymbol{x} \neq \mathbf{0}$ i.e. $(x_1, x_2) \neq (0, 0)$.
- The point $\boldsymbol{x} = \mathbf{0}$ is a global maximum.
- Q_2 is called negative definite.



Indefinite

If $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ then $Q_3(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 - x_2^2$.

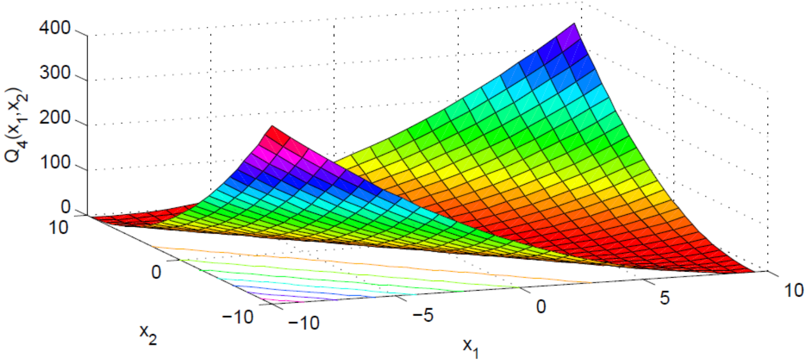
- Q_3 can be take both positive and negative values.
- E.g. $Q_3(1,0) = +1$ and $Q_3(0,1) = -1$.
- Q_3 is called indefinite.



Positive semidefinite

If $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ then $Q_4(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 + 2x_1x_2 + x_2^2$.

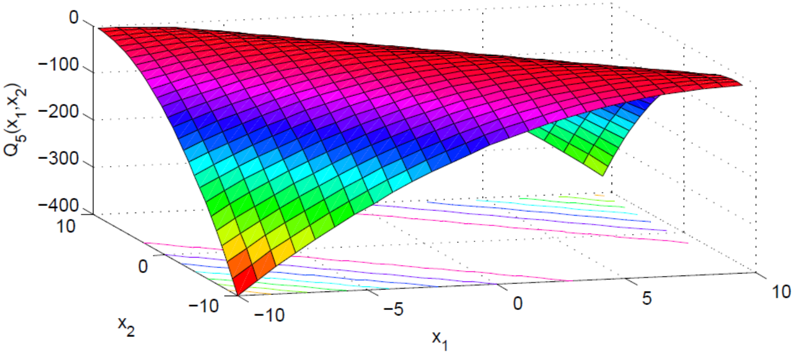
- Q_4 is always ≥ 0 but does equal zero at some $\boldsymbol{x} \neq \mathbf{0}$.
- E.g. $Q_4(10, -10) = 0$.
- Q_4 is called positive semidefinite.



Negative semidefinite

If $\mathbf{A} = \begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix}$ then $Q_5(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = -(x_1 + x_2)^2$.

- Q_5 is always ≤ 0 but does equal zero at some $\boldsymbol{x} \neq \mathbf{0}$.
- E.g. $Q_5(10, -10) = 0$.
- Q_5 is called negative semidefinite.



Definite symmetric matrices

A symmetric matrix, $\boldsymbol{A}$, is called positive definite, positive semidefinite, negative definite, etc. according to the definiteness of the corresponding quadratic form $Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$.
Let $\boldsymbol{A}$ be a $n \times n$ symmetric matrix, then $\boldsymbol{A}$ is

1. positive definite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$ in $\mathbb{R}^n$
2. positive semidefinite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq \mathbf{0}$ in $\mathbb{R}^n$
3. negative definite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} < 0$ for all $\boldsymbol{x} \neq \mathbf{0}$ in $\mathbb{R}^n$
4. negative semidefinite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leq 0$ for all $\boldsymbol{x} \neq \mathbf{0}$ in $\mathbb{R}^n$
5. indefinite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for some $\boldsymbol{x} \neq \mathbf{0}$ in $\mathbb{R}^n$ and < 0 for some other $\boldsymbol{x}$ in $\mathbb{R}^n$

- We can check the definiteness of a matrix by showing that one of these definitions holds as in the example
- You can find the eigenvalues to check definiteness

How else to check for definiteness?

You can check the sign of the sequence of determinants of the leading principal minors:

Positive Definite

. An $n \times n$ matrix $\boldsymbol{M}$ is positive definite if all the following matrices have a positive determinant:

- the top left 1×1 corner of $\boldsymbol{M}$ (1st order principal minor)
- the top left 2×2 corner of $\boldsymbol{M}$ (2nd order principal minor)

  $\vdots$
- $\boldsymbol{M}$ itself.

In other words, all of the leading principal minors are positive.

Negative Definite

. A matrix is negative definite if all kth order leading principal minors are negative when k is odd and positive when k is even.

Why do we care about definiteness?

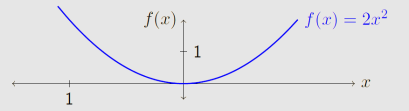
Useful for establishing if a (multivariate) function has a maximum, minimum or neither at a critical point.

- If we have a function, $f(x)$, we can show that a minimum exists at a critical point, i.e. when $f'(x) = 0$, if $f''(x) > 0$.

Positive Definite

$$f(x) = 2x^2.$$

- $f'(x) = 4x$
- $f'(x) \stackrel{!}{=} 0 \quad \rightarrow \quad x = 0$
- $f''(x) = 4 > 0 \quad \rightarrow \quad$ minimum at $x = 0$.



Why do we care about definiteness?

- In the special case of a univariate function $f''(x)$ is a 1×1 Hessian matrix and showing that $f''(x) > 0$ is equivalent to showing that the Hessian is positive definite.
- If we have a bivariate function $f(x, y)$ we find critical points when the first order partial derivatives are equal to zero:
  1. Find the first order derivatives and set them equal to zero
  2. Solve simultaneously to find critical points
- We can check if max or min or neither using the Hessian matrix, $\boldsymbol{H}$, the matrix of second order partial derivatives:

$$\boldsymbol{H} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

  1. (If necessary) evaluate the Hessian at a critical point
  2. Check if $\boldsymbol{H}$ is positive or negative definite:
     - $|\boldsymbol{H}| > 0$ and $f_{xx} > 0$ $\rightarrow$ positive definite $\rightarrow$ minimum
     - $|\boldsymbol{H}| > 0$ and $f_{xx} < 0$ $\rightarrow$ negative definite $\rightarrow$ maximum
  3. Repeat for all critical points

Why do we care about definiteness?

- If we find the second order conditions and show that it is a positive definite matrix then we have shown that we have a minimum.
- Positive definite matrices are non-singular, i.e. we can invert them. So if we can show $\boldsymbol{X}'\boldsymbol{X}$ is positive definite, we can find $[\boldsymbol{X}'\boldsymbol{X}]^{-1}$.
- Application: showing that the Ordinary Least Squares (OLS) minimises the sum of squared residuals.

Matrices as systems of equations

- A system of equations:

$$\begin{aligned} y_1 &= x_{11}b_1 + x_{12}b_2 + \ldots + x_{1k}b_k \\ y_2 &= x_{21}b_1 + x_{22}b_2 + \ldots + x_{2k}b_k \\ &\vdots \\ y_n &= x_{n1}b_1 + x_{n2}b_2 + \ldots + x_{nk}b_k \end{aligned} \tag{3}$$

- The matrix form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}. \tag{4}$$

Matrices as systems of equations

- More succinctly: $\mathbf{y} = \mathbf{X}\mathbf{b}$ where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}; \quad \mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix}$$

for $i = 1, 2, \ldots, n$ and

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix}$$

- $\mathbf{x}_i$ is the "covariate vector" for the ith observation.

Matrices as systems of equations

- We can write $\mathbf{y} = \mathbf{X}\mathbf{b}$ as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix} \mathbf{b}.$$

- Returning to the original system, we can write each individual equation using vectors:

$$\begin{aligned} y_1 &= \mathbf{x}_1'\mathbf{b} \\ y_2 &= \mathbf{x}_2'\mathbf{b} \\ &\vdots \\ y_n &= \mathbf{x}_n'\mathbf{b}. \end{aligned}$$

1.4: Mixing Matrices, Vectors & Summation

Often we want to find $\mathbf{X}'\mathbf{X}$ or $\mathbf{X}'\mathbf{u}$. A convenient way to write this is as a sum of vectors. Say we have a 3×2 matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \mathbf{x}_3' \end{bmatrix}; \quad \mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix}; \text{and} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

We can write,

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{33} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \mathbf{x}_3' \end{bmatrix} \\ &= \mathbf{x}_1\mathbf{x}_1' + \mathbf{x}_2\mathbf{x}_2' + \mathbf{x}_3\mathbf{x}_3' \\ &= \sum_{i=1}^{3} \mathbf{x}_i\mathbf{x}_i'. \end{aligned}$$

In a similar fashion, we can also show that $\mathbf{X}'\mathbf{u} = \sum_{i=1}^{3} \mathbf{x}_i u_i$.

$$
\begin{aligned}
\mathbf{X}'\mathbf{u} &= \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{33} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \\
&= \begin{bmatrix} x_{11}u_1 + x_{21}u_2 + x_{31}u_3 \\ x_{12}u_1 + x_{22}u_2 + x_{32}u_3 \end{bmatrix} \\
&= \mathbf{x}_1 u_1 + \mathbf{x}_2 u_2 + \mathbf{x}_3 u_3 \\
&= \sum_{i=1}^{3} \mathbf{x}_i u_i .
\end{aligned}
$$

1.5: Applications in Econometrics

Application: variance-covariance matrix

- For the univariate case, $var(y) = \mathbb{E}([y - \mu]^2)$.
- In the multivariate case $\mathbf{y}$ is a vector of n random variables.
- Without loss of generality, assume $\mathbf{y}$ has mean zero, i.e. $\mathbb{E}(\mathbf{y}) = \boldsymbol{\mu} = \mathbf{0}$. Then,

$$
\begin{aligned}
cov(\mathbf{y}, \mathbf{y}) &= var(\mathbf{y}) = \mathbb{E}([\mathbf{y} - \boldsymbol{\mu}][\mathbf{y} - \boldsymbol{\mu}]') \\
&= \mathbb{E}\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} \right) \\
&= \mathbb{E} \begin{bmatrix} y_1^2 & y_1 y_2 & \dots & y_1 y_n \\ y_2 y_1 & y_2^2 & \dots & y_2 y_n \\ \vdots & \vdots & & \vdots \\ y_n y_1 & y_n y_2 & \dots & y_n^2 \end{bmatrix}
\end{aligned}
$$

Application: variance-covariance matrix

- Hence, we have a variance-covariance matrix:

$$var(\mathbf{y}) = \begin{bmatrix} var(y_1) & cov(y_1, y_2) & \dots & cov(y_1, y_n) \\ cov(y_2, y_1) & var(y_2) & \dots & cov(y_2, y_n) \\ \vdots & \vdots & & \vdots \\ cov(y_n, y_1) & cov(y_n, y_2) & \dots & var(y_n) \end{bmatrix}$$

- What if we weight the random variables with a vector of constants, **a**?

$$\begin{aligned} var(\mathbf{a}'\mathbf{y}) &= \mathbb{E}([\mathbf{a}'\mathbf{y} - \mathbf{a}'\mu][\mathbf{a}'\mathbf{y} - \mathbf{a}'\mu]') \\ &= \mathbb{E}(\mathbf{a}'[\mathbf{y} - \mu](\mathbf{a}'[\mathbf{y} - \mu])') \\ &= \mathbb{E}(\mathbf{a}'[\mathbf{y} - \mu][\mathbf{y} - \mu]'\mathbf{a}) \\ &= \mathbf{a}'\mathbb{E}([\mathbf{y} - \mu][\mathbf{y} - \mu]')\mathbf{a} \\ &= \mathbf{a}'var(\mathbf{y})\mathbf{a}. \end{aligned}$$

Let $\mathbf{y} = (y_1, y_2)'$ be a vector of random variables and $\mathbf{a} = (a_1, a_2)'$ be some constants,

$$\mathbf{a}'\mathbf{y} = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = a_1 y_1 + a_2 y_2.$$

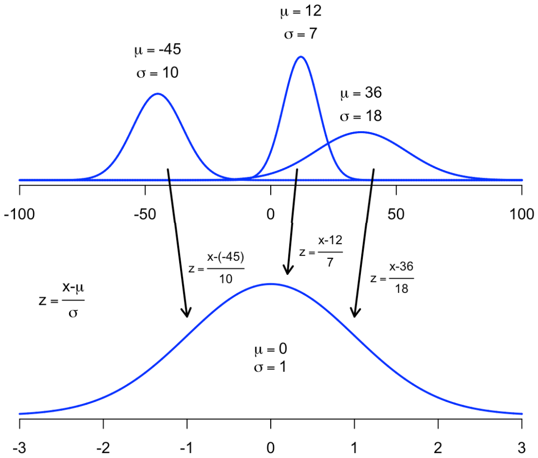
Now, $var(a_1 y_1 + a_2 y_2) = var(\mathbf{a}'\mathbf{y}) = \mathbf{a}' var(\mathbf{y}) \mathbf{a}$ where

$$var(\mathbf{y}) = \begin{bmatrix} var(y_1) & cov(y_1, y_2) \\ cov(y_1, y_2) & var(y_2) \end{bmatrix}$$

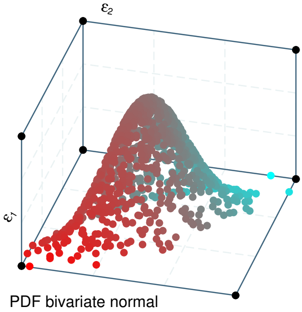
is the (symmetric) variance-covariance matrix.

$$\begin{aligned} var(\mathbf{a}'\mathbf{y}) &= \mathbf{a}' var(\mathbf{y}) \mathbf{a} \\ &= \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} var(y_1) & cov(y_1, y_2) \\ cov(y_1, y_2) & var(y_2) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \\ &= a_1^2 var(y_1) + a_2^2 var(y_2) + 2 a_1 a_2 cov(y_1, y_2) \end{aligned}$$

Application: Standardizing a Univariate Normally Distributed Vector



Is this also true for multivariate (e.g. bivariate) normal vectors?



The example shows a bivariate normal distribution of ε and ν with density $f(\varepsilon, \nu) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(\varepsilon^2 - 2\rho\,\varepsilon\nu + \nu^2)\right)$ with $N = 1000$, $\sigma_\varepsilon = \sigma_\nu = 1$, and $\rho = 0.8$. The graph was produced using the ado file graph3d.

Consider the vector of random variables $\boldsymbol{y}$ that is normally distributed with expectation $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{y} \sim N[\boldsymbol{\mu}, \boldsymbol{\Sigma}].$$

Show that the standardized vector

$$\Sigma^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}) = \boldsymbol{z} \sim N[\mathbf{0}, \boldsymbol{I}].$$

Let's weight the random variable in $\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{\mu}$ with a symmetric matrix $\boldsymbol{\Sigma}^{-1/2}$.

$$E(\boldsymbol{\Sigma}^{-1/2} z) = \boldsymbol{\Sigma}^{-1/2} E(\boldsymbol{z}) = \mathbf{0}.$$

$$Var(\boldsymbol{\Sigma}^{-1/2} z) = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1/2} = \boldsymbol{I}.$$

Example

Bivariate vector $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ with $\rho = 1$,$\mu_1 = 1$,$\mu_2 = 8$,$\sigma_1^2 = 13$, $\sigma_2^2 = 5$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right]$$

Show that $E(\boldsymbol{\Sigma}^{-1/2}z) = \boldsymbol{\Sigma}^{-1/2}E(\boldsymbol{z}) = \mathbf{0}$:

$$E(\boldsymbol{z}) = E\begin{pmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Drawing two values of $y_1^1 = -1$, $y_1^2 = 3$ and $y_2^1 = 6$, $y_2^2 = 10$

$$E(\boldsymbol{z}) = 1/2\begin{pmatrix} -1 - 1 \\ 6 - 8 \end{pmatrix} + 1/2\begin{pmatrix} 3 - 1 \\ 10 - 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Application: OLS

Given a linear model $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$ derive the OLS estimator β. Show that β achieves a minimum.

- The OLS estimator β minimises the sum of squared residuals,

$$\boldsymbol{u}'\boldsymbol{u} = \sum_{i=1}^{n} u_i^2 \text{ where } \boldsymbol{u} = \boldsymbol{y} - \boldsymbol{X}\beta \text{ or } u_i = y_i - \boldsymbol{x}_i'\beta.$$

$$\begin{aligned} S(\beta) = \sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\beta)^2 &= (\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta) \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\beta + \beta'\boldsymbol{X}'\boldsymbol{X}\beta \end{aligned}$$

Multiplying $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$ out

$$\ldots - (\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta}\ldots$$

$$\ldots - (\underset{1\times n}{\boldsymbol{y}}{}' \underset{n\times k}{\boldsymbol{X}} \underset{k\times 1}{\boldsymbol{\beta}})' - \underset{1\times n}{\boldsymbol{y}'} \underset{n\times k}{\boldsymbol{X}} \underset{k\times 1}{\boldsymbol{\beta}}\ldots$$

Recall that for a symmetric matrix $(\boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta})' = (\boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta})$

$$\ldots - 2\boldsymbol{y}'\boldsymbol{X}'\boldsymbol{\beta}\ldots$$

Example

$$
\begin{aligned}
\mathbf{y} &= \begin{bmatrix}1\\2\end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix}0.5\\2\end{bmatrix}; \mathbf{X} = \begin{bmatrix}2 & 3\\1 & 0\end{bmatrix} \\
& \ldots - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta}\ldots \\
- & \begin{bmatrix}0.5 & 2\end{bmatrix}\begin{bmatrix}2 & 1\\3 & 0\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}2 & 3\\1 & 0\end{bmatrix}\begin{bmatrix}0.5\\2\end{bmatrix} = -16 \\
&= \ldots - 2\boldsymbol{y}'\boldsymbol{X}'\boldsymbol{\beta}\ldots \\
&= -2\left(\begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}2 & 3\\1 & 0\end{bmatrix}\begin{bmatrix}0.5\\2\end{bmatrix}\right) \\
&= -2 \times \Big((0.5 \times 2 + 2 \times 3) \times 1 + (0.5 \times 1 + 2 \times 0) \times 2\Big) = -16
\end{aligned}
$$

Application: OLS

- Recall

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2 = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$$

- Take the first derivative of $S(\boldsymbol{\beta})$ and set it equal to zero:

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = 0 \rightarrow \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{y}.$$

- Assuming $\boldsymbol{X}$ (and therefore $\boldsymbol{X}'\boldsymbol{X}$) is of full rank (so is $\boldsymbol{X}'\boldsymbol{X}$ invertible) we get,

$$\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

Application: OLS

- For a minimum we need to use the second order conditions:
$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = 2\boldsymbol{X}'\boldsymbol{X}.$$
- The solution will be a minimum if $\boldsymbol{X}'\boldsymbol{X}$ is a positive definite matrix. Let $q = \boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{c}$ for some $\boldsymbol{c} \neq \mathbf{0}$. Then
$$q = \boldsymbol{v}'\boldsymbol{v} = \sum_{i=1}^{n} \nu_i^2, \text{ where } \boldsymbol{v} = \boldsymbol{X}\boldsymbol{c}.$$
- Unless $\boldsymbol{v} = \mathbf{0}$, q is positive. But, if $\boldsymbol{v} = \mathbf{0}$ then $\boldsymbol{v}$ or $\boldsymbol{c}$ would be a linear combination of the columns of $\boldsymbol{X}$ that equals $\mathbf{0}$ which contradicts the assumption that $\boldsymbol{X}$ has full rank.

Application: OLS

- Since $\boldsymbol{c}$ is arbitrary, q is positive for every $\boldsymbol{c} \neq 0$ which establishes that $\boldsymbol{X}'\boldsymbol{X}$ is positive definite.
- Therefore, if $\boldsymbol{X}$ has full rank, then the least squares solution β is unique and minimises the sum of squared residuals.

Matrix Operations

Operation	R	Stata	Mata	Matlab	Python
$\mathbf{A} = \begin{bmatrix} 5 & 7 \\ 10 & 5 \end{bmatrix}$	A=matrix(c(5,7,10,2), ncol=2,byrow=T)	mat A = (5,7\10,5)	A = (5,7\10,5)	A = [5,7;10,2]	A = np.array([[5, 7], [10, 5]])
$r \times r$ identity matrix $\mathbf{I}_r$	diag(1,r)	mat I = I(r)	I(r)	eye(r)	np.eye(r)
$\mathbf{A}^{-1}$	solve(A)	mat C = inv(A)	invsym(A)	inv(A)	np.linalg.inv(A)
$\mathbf{A} + \mathbf{B}$	A + B	mat C = A + B	A + B	A + B	np.add(A,B)
$\mathbf{AB}$	A %*% B	mat D = A*B	A * B	A * B	np.multiply(A,B)
$\mathbf{A}'$	t(A)	mat a = A'	A'	A'	A.T
Kronecker product of **A** and **B**	kronecker.prod(A, B)	mat b = A#B	A#B	A = [5,7;10,2]	np.kron(A, B)
...					

Matrix Functions

Function	R	Stata	Mata	Matlab	Python
eigenvalues of **A** & eigenvectors	eigen(A)	matrix eigenvalues r c = A	eigenvalues(A)	[V,E] = eig(A)	np.linalg.eig(A)
covariance matrix	var(A) or cov(A)	mat r= corr(A)	corr(A)	cov(A)	np.cov(A)
rank(**A**)	qr(A)$rank	—	rank(A)	rank(A)	np.linalg.matrix_rank(A
$\det(\mathbf{A})$	det(A)	scalar det = det(A)	det(A)	det(A)	np.linalg.det(A)
vectorize **A**	b=c(A)	mat b = vec(A)	vec(A)	A = [5,7;10,2]	np.vectorize(A)
...					