TA Session for Econometrics I 2025

8.1

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

June 20, 2025

Positive Definite Matrices in Statistics

Definition

A matrix $A \in \mathbb{R}^{n \times n}$ is called **positive definite** if for any non-zero vector $x \in \mathbb{R}^n$,

$$x^\top A x > 0.$$

Key contexts where this matters:

- Covariance matrices
- Quadratic forms in estimation
- Optimisation problems

Variance-Covariance Matrices

Variance-covariance matrices must be positive definite.

- A covariance matrix Σ summarises the variability and relationships among variables.
- For any linear combination of variables $A^\top X$, the variance is

  $$\mathrm{Var}(A^\top X) = A^\top \Sigma A.$$

  where the variance of random bvariable X is Σ.
- To ensure variance is always positive (except for $A = 0$), Σ must be positive definite.

Note: X is given.

$$\begin{aligned}\beta_{ols} &= (X^\top X)^{-1} X^\top y \\ &= \beta + (X^\top X)^{-1} X^\top u\end{aligned}$$

$$\begin{aligned}Var(\beta_{ols}) &= Var(\beta + (X^\top X)^{-1} X^\top u) \\ &= (X^\top X)^{-1} X^\top Var(u) X (X^\top X)^{-1} \\ &= (X^\top X)^{-1} X^\top \Sigma X (X^\top X)^{-1}\end{aligned}$$

Model: $y = \rho W y + X\beta + u$

$$\begin{aligned} y &= \rho W y + X\beta + u \\ (I - \rho W)y &= X\beta + u \end{aligned}$$

Assuming $u \sim (0, \sigma^2 I)$**, the variance of** $\hat{\beta}$**:**

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 (X^\top X)^{-1} X^\top (I - \rho W) I (I - \rho W)^\top X (X^\top X)^{-1}$$

Note:
The matrix $(I - \rho W)$ should be **non-singular**, and the entire variance-covariance matrix must be **positive definite** for valid statistical inference.

This is guaranteed when $I - \rho W$ is positive definite.

Optimisation Perspective

- Many estimation problems reduce to minimising a quadratic form:

$$(y - X\beta)^\top W (y - X\beta)$$

- If W is positive definite, the function is strictly convex.
- This guarantees a unique minimum.

A matrix A is positive definite if:

- For all non-zero vectors x, $x^\top Ax > 0$.

Practical criteria:

- **All eigenvalues are positive.**
  If A is symmetric and all eigenvalues $\lambda_i > 0$, then A is positive definite.
- **All leading principal minors are positive.**
  That is, the determinants of the top-left $k \times k$ submatrices for $k = 1, \ldots, n$.
- **Cholesky decomposition exists.**
  If A can be decomposed as $A = LL^\top$, where L is lower triangular with positive diagonal entries.

- $\forall x \neq 0$,

$$x^\top A x = \sum_{i=1}^{n} \lambda_i y_i^2,$$

where $y = Q^\top x$. If $\lambda_i > 0$ (for all i) holds, we have $x^\top A x > 0$.

Cholesky Decomposition and Positive Definiteness

Cholesky decomposition:
For a real symmetric matrix $A \in \mathbb{R}^{n \times n}$,
if A is positive definite, then

$$A = LL^\top$$

where L is a lower triangular matrix with positive diagonal entries.

- During decomposition, we compute square roots of diagonal elements.
- These are well-defined **only if** those elements are positive.
- So if any diagonal entry becomes zero or negative, the process fails.

Cholesky decomposition exists **if and only if** A is positive definite.

Example: Cholesky Decomposition Intuition

Consider a symmetric matrix:

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix}$$

Cholesky decomposition gives:

$$L = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}, \quad A = LL^\top$$

- The diagonal elements (2 and 1) are both positive.
- $x^\top Ax > 0$ for any $x \neq 0$.

If A were not positive definite, the square roots would not exist or become imaginary.

Variance (Scalar)

Definition

Let X be a real-valued random variable such that its first moment exists, i.e., $E[X] < \infty$. Then the variance of X is defined as follows:

$$\mathrm{Var}(X) = E\left[(X - E[X])^2\right] \geq 0$$

The variance measures the spread of X around its mean. It is always non-negative, and it equals zero only when X is almost surely constant.

Variance-Covariance Matrix

Definition

Let $\mathbf{X} \in \mathbb{R}^n$ be a random vector such that $E[\mathbf{X}] < \infty$. Then the variance-covariance matrix of $\mathbf{X}$ is defined by

$$\Sigma = E\left[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^\top\right]$$

The diagonal elements of Σ represent the variances of individual components of $\mathbf{X}$, and the off-diagonal elements represent the covariances between components.

Positive Semidefiniteness of Σ

Key Property

For any non-zero vector $\mathbf{a} \in \mathbb{R}^n$, the variance of the linear combination $\mathbf{a}^\top \mathbf{X}$ is given by:

$$\mathrm{Var}(\mathbf{a}^\top \mathbf{X}) = \mathbf{a}^\top \Sigma \mathbf{a} \geq 0$$

- This property implies that Σ is a **positive semidefinite matrix**.
- If the inequality is strict for all non-zero $\mathbf{a}$, then Σ is **positive definite**.

Why Positive (Semi)Definiteness Matters

- Variance cannot be negative — this holds for all directions in $\mathbb{R}^n$.
- The matrix Σ must ensure that every linear combination $\mathbf{a}^\top \mathbf{X}$ has non-negative variance.
- This is guaranteed only when Σ is positive semidefinite.
- In statistical modeling, a covariance matrix that is not at least positive semidefinite leads to invalid results.

Intuition: Linear Combination and Variance

Linear combination of random variables

For any vector $\mathbf{a} \in \mathbb{R}^n$, we can construct a new random variable:

$$Y = \mathbf{a}^\top \mathbf{X}$$

This is called a **linear combination** of the components of $\mathbf{X}$.

- Y is a scalar random variable.
- Its variance is given by:

$$\mathrm{Var}(Y) = \mathrm{Var}(\mathbf{a}^\top \mathbf{X}) = \mathbf{a}^\top \Sigma \mathbf{a} \in \mathbb{R}$$

since $Var(X) = \Sigma \in \mathbb{R}^{n \times n}$.

Why $\mathbf{a}^\top \Sigma \mathbf{a} \geq 0$?

- $\mathrm{Var}(Y)$ is the variance of a scalar random variable, and therefore must be **non-negative**.
- Since this must be true for **any** vector $\mathbf{a}$,

$$\mathbf{a}^\top \Sigma \mathbf{a} \geq 0 \quad \text{for all } \mathbf{a} \in \mathbb{R}^n$$

- This is exactly the definition of a **positive semidefinite matrix**.

Conclusion

A variance-covariance matrix must be positive semidefinite so that all possible linear combinations of variables yield valid (non-negative) variances.