Linear Algebra for Estimation

PLSC 30700 — Pre-Midterm Review

Robert Gulotty
University of Chicago

Winter 2026

This sheet collects the linear algebra concepts used in Lectures 1–7 (before the midterm). Calculus skills are covered in a separate document. Each section lists the key definitions and results, and indicates where they appear in the course.

Abbreviations Used in This Document

OLS	Ordinary Least Squares	GLS	Generalized Least Squares
SSE	Sum of Squared Errors ($\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}$)	SSR	Sum of Squares due to Regression ($\hat{\boldsymbol{y}}'\hat{\boldsymbol{y}}$)
BLP	Best Linear Predictor	CEF	Conditional Expectation Function
ANOVA	Analysis of Variance	FWL	Frisch-Waugh-Lovell (Theorem)
PD	Positive Definite	PSD	Positive Semi-Definite
HC	Heteroskedasticity-Consistent		

Contents

1 Vectors and Basic Operations

1.1 Vectors and Notation

A **column vector** $\boldsymbol{a} \in \mathbb{R}^k$ is a $k \times 1$ array of numbers. Its **transpose** $\boldsymbol{a}'$ is $1 \times k$.

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}, \qquad \boldsymbol{a}' = (a_1,\ a_2,\ \dots,\ a_k)$$

Where it appears:

- **Day 1:** The coefficient vector $\boldsymbol{\beta}$ ($k \times 1$), the outcome vector $\boldsymbol{y}$ ($n \times 1$), the observation vector $\boldsymbol{x}_i$ ($k \times 1$).
- **Day 3 onward:** The residual vector $\hat{\boldsymbol{e}}$ ($n \times 1$).

1.2 Inner Product (Dot Product)

For $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^k$:

$$\boldsymbol{a}'\boldsymbol{b} = \sum_{j=1}^{k} a_j b_j$$

- $\boldsymbol{a}'\boldsymbol{b} = 0$ means $\boldsymbol{a}$ and $\boldsymbol{b}$ are **orthogonal**.
- $\|\boldsymbol{a}\| = \sqrt{\boldsymbol{a}'\boldsymbol{a}}$ is the **Euclidean norm** (length).

Where it appears:

- **Day 1:** Covariance is an inner product of demeaned vectors. OLS residuals are orthogonal to $\boldsymbol{X}$: $\boldsymbol{X}'\hat{\boldsymbol{e}} = \mathbf{0}$.
- **Day 3:** SSE $= \hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} = \|\hat{\boldsymbol{e}}\|^2$. The ANOVA decomposition is an orthogonal decomposition of $\boldsymbol{y}$.

2 Matrices

2.1 Matrix Multiplication

If $\boldsymbol{A}$ is $k \times r$ and $\boldsymbol{B}$ is $r \times s$, the product $\boldsymbol{AB}$ is $k \times s$ with entries:

$$(\boldsymbol{AB})_{ij} = \sum_{\ell=1}^{r} a_{i\ell}\, b_{\ell j}$$

Key properties:

- $\boldsymbol{AB} \neq \boldsymbol{BA}$ in general (not commutative).
- $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$ (transpose reverses order).
- $\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$ (associative).
- $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$ (distributive).

Where it appears:

- **Day 1:** The compact regression notation $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}$. The Gram matrix $\boldsymbol{X}'\boldsymbol{X}$.
- **Day 3:** Expanding $(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$ using the distributive and transpose rules.
- **Day 5:** Variance formulas: $\mathrm{Var}(\boldsymbol{Ay}) = \boldsymbol{A}\,\mathrm{Var}(\boldsymbol{y})\boldsymbol{A}'$.

2.2 The Design Matrix $\boldsymbol{X}$

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1,k-1} \\ 1 & x_{21} & \cdots & x_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{n,k-1} \end{pmatrix} \quad (n \times k)$$

- Rows are observations ($\boldsymbol{x}_i'$), columns are variables.
- First column is $\boldsymbol{\iota} = (1, 1, \ldots, 1)'$ (the intercept).

Where it appears: Every lecture from Day 1 onward. The full rank condition ($\mathrm{rank}(\boldsymbol{X}) = k$) is Assumption 3 for OLS.

2.3 Matrix Inverse

A square matrix $\boldsymbol{A}$ ($k \times k$) is **invertible** (nonsingular) if there exists $\boldsymbol{A}^{-1}$ with $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_k$.

- $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$
- $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ (reverses order)
- $\boldsymbol{A}$ is invertible iff $\text{rank}(\boldsymbol{A}) = k$ iff $\det(\boldsymbol{A}) \neq 0$.

Where it appears:

- **Day 1:** OLS requires $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ to exist, which needs full column rank.
- **Day 2:** BLP requires $\boldsymbol{Q}_{XX} = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}']$ to be invertible.
- **Day 4:** Partitioned inverse of $\boldsymbol{X}'\boldsymbol{X}$ (Schur complement).

2.4 The Identity Matrix

$\boldsymbol{I}_k$ is the $k \times k$ matrix with 1s on the diagonal and 0s elsewhere.

$$\boldsymbol{I}_k\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_k = \boldsymbol{A}$$

Where it appears: Homoskedasticity is $\text{Var}(\boldsymbol{e}|\boldsymbol{X}) = \sigma^2\boldsymbol{I}_n$ (Days 3, 5, 6). The annihilator is $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$ (Day 3).

3 Symmetric Matrices and Quadratic Forms

3.1 Symmetric Matrices

$\boldsymbol{A}$ is **symmetric** if $\boldsymbol{A}' = \boldsymbol{A}$. Equivalently, $a_{ij} = a_{ji}$ for all i, j.

Where it appears: $\boldsymbol{X}'\boldsymbol{X}$ is always symmetric. The variance-covariance matrix $\boldsymbol{\Sigma}$ is symmetric. The projection matrix $\boldsymbol{P}$ is symmetric (Day 1, 3).

3.2 Quadratic Forms

A **quadratic form** is a scalar of the form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is symmetric:

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_i a_{ii}x_i^2 + 2\sum_{i<j} a_{ij}x_ix_j$$

Where it appears:

- **Day 1:** SSR is a quadratic form: $\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$. The SSE bowl is the surface plot of $\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{y}'\boldsymbol{y}$.
- **Day 3:** $\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} = \boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}$ is a quadratic form in $\boldsymbol{e}$ (used in proving unbiasedness of s^2).
- **Day 7:** $\boldsymbol{z}'\boldsymbol{A}\boldsymbol{z} \sim \chi^2(\nu)$ when $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{A}$ is idempotent with $\text{rank}(\boldsymbol{A}) = \nu$.

3.3 Positive Definiteness

A symmetric matrix $\boldsymbol{A}$ is:
- **Positive definite** (PD) if $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} > 0$ for all $\boldsymbol{c} \neq \boldsymbol{0}$.
- **Positive semi-definite** (PSD) if $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} \geq 0$ for all $\boldsymbol{c} \neq \boldsymbol{0}$.

Key facts:
- PD $\iff$ all eigenvalues are positive $\iff$ invertible.
- If $\boldsymbol{A}$ is PD, then $\boldsymbol{A}^{-1}$ is also PD.
- If $\boldsymbol{X}$ has full column rank, then $\boldsymbol{X}'\boldsymbol{X}$ is PD.
- $\boldsymbol{C}'\boldsymbol{C}$ is always PSD (PD if $\boldsymbol{C}$ has full column rank).

Where it appears:

- **Day 1:** PD of $\boldsymbol{X}'\boldsymbol{X}$ guarantees a unique OLS solution and that the SSE is strictly convex (the "bowl" has a unique bottom).
- **Day 2:** $\boldsymbol{Q}_{XX} = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}']$ must be PD for the BLP to be identified.
- **Day 3:** Second-order condition: $2\boldsymbol{X}'\boldsymbol{X}$ is PD, confirming $\hat{\boldsymbol{\beta}}$ is a minimum.
- **Day 5:** Gauss-Markov proof: $\boldsymbol{A}'\boldsymbol{A} - (\boldsymbol{X}'\boldsymbol{X})^{-1} = \boldsymbol{C}'\boldsymbol{C} \geq 0$ shows OLS has smallest variance.
- **Day 5:** Comparing estimators via PD: "$\boldsymbol{A} - \boldsymbol{B}$ is PD" means $\boldsymbol{A}$ has "larger" variance.

4 Trace, Rank, and Determinant

4.1 Trace

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}$$

Key properties:
- $\mathrm{tr}(\boldsymbol{A} + \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$
- $\mathrm{tr}(c\boldsymbol{A}) = c\,\mathrm{tr}(\boldsymbol{A})$
- $\mathrm{tr}(\boldsymbol{A}') = \mathrm{tr}(\boldsymbol{A})$
- $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$ (cyclic property—the most used)
- $\mathrm{tr}(\boldsymbol{A}) = \sum_i \lambda_i$ (sum of eigenvalues)

Where it appears:

- **Day 1:** $\mathrm{tr}(\boldsymbol{P}) = k$, $\mathrm{tr}(\boldsymbol{M}) = n - k$. This is the algebraic origin of degrees of freedom.
- **Day 3:** Proving $\mathbb{E}[\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}|\boldsymbol{X}] = \sigma^2(n-k)$ via the "trace trick":
$$\mathbb{E}[\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}|\boldsymbol{X}] = \mathrm{tr}(\boldsymbol{M}\,\mathbb{E}[\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}]) = \sigma^2\,\mathrm{tr}(\boldsymbol{M}) = \sigma^2(n-k)$$
- **Day 4:** $\mathrm{tr}(\boldsymbol{P}) = k$ proved via the cyclic property: $\mathrm{tr}(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') = \mathrm{tr}(\boldsymbol{I}_k) = k$.
- **Day 5:** Unbiasedness of $\hat{\sigma}^2$: $\mathbb{E}[\hat{\sigma}^2|\boldsymbol{X}] = \frac{1}{n}\,\mathrm{tr}(\boldsymbol{M}\boldsymbol{D})$.

4.2 Rank

rank($\boldsymbol{A}$) = number of linearly independent columns (or rows).

- rank($\boldsymbol{X}$) $= k$ means full column rank (no perfect multicollinearity).
- $\text{rank}(\boldsymbol{AB}) \leq \min(\text{rank}(\boldsymbol{A}), \text{rank}(\boldsymbol{B}))$.
- For idempotent $\boldsymbol{A}$: $\text{rank}(\boldsymbol{A}) = \text{tr}(\boldsymbol{A})$.

Where it appears:

- **Day 1:** Full rank of $\boldsymbol{X}$ is needed for OLS; rank deficiency = perfect multicollinearity.
- **Day 6:** Near multicollinearity inflates $\text{Var}(\hat{\beta}_j)$ via $(1-\rho^2)^{-1}$.

4.3 Determinant

$\det(\boldsymbol{A})$ measures signed volume scaling. Key facts:

- $\det(\boldsymbol{A}) = 0 \iff \boldsymbol{A}$ is singular (not invertible).
- $\det(\boldsymbol{A}) = \prod_i \lambda_i$ (product of eigenvalues).

Where it appears:

- **Day 1:** Mentioned as a diagnostic for invertibility.
- **Day 7:** The multivariate normal density involves $|\boldsymbol{\Sigma}| = \det(\boldsymbol{\Sigma})$.

5 Eigenvalues and Eigenvectors

If $\boldsymbol{Au} = \lambda \boldsymbol{u}$ for nonzero $\boldsymbol{u}$, then λ is an **eigenvalue** and $\boldsymbol{u}$ is the corresponding **eigenvector**.

For symmetric $\boldsymbol{A}$:

- All eigenvalues are real.
- Eigenvectors corresponding to distinct eigenvalues are orthogonal.
- $\boldsymbol{A}$ is PD $\iff$ all $\lambda_i > 0$.
- $\boldsymbol{A}$ is singular $\iff$ some $\lambda_i = 0$.
- $\text{tr}(\boldsymbol{A}) = \sum \lambda_i, \quad \det(\boldsymbol{A}) = \prod \lambda_i$.

Where it appears:

- **Day 1:** Eigenvalues diagnose PD of $\boldsymbol{X}'\boldsymbol{X}$; idempotent matrices have eigenvalues in $\{0, 1\}$.
- **Day 5:** Spectral decomposition $\boldsymbol{\Sigma} = \boldsymbol{C\Lambda C}'$ used to construct $\boldsymbol{\Sigma}^{-1/2}$ for GLS.

6 Linear Independence, Span, and Column Space

- Vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are **linearly independent** if $c_1\boldsymbol{x}_1 + \cdots + c_k\boldsymbol{x}_k = \mathbf{0}$ implies $c_1 = \cdots = c_k = 0$.
- The **span** (or column space) of $\boldsymbol{X}$ is $\text{col}(\boldsymbol{X}) = \{c_1\boldsymbol{x}_1 + \cdots + c_k\boldsymbol{x}_k : c_j \in \mathbb{R}\}$.
- The **dimension** of a subspace is the number of vectors in any basis.
- **Orthogonal complement**: $\mathcal{W}^\perp = \{\boldsymbol{v} : \boldsymbol{v}'\boldsymbol{w} = 0 \text{ for all } \boldsymbol{w} \in \mathcal{W}\}$.
- $\dim(\mathcal{W}) + \dim(\mathcal{W}^\perp) = n$.

Where it appears:

- **Day 1:** Fitted values $\hat{\boldsymbol{y}} \in \text{col}(\boldsymbol{X})$ (dimension k); residuals $\hat{\boldsymbol{e}} \in \text{col}(\boldsymbol{X})^\perp$ (dimension $n-k$). This is the geometric origin of degrees of freedom: $n-k$.
- **Day 1:** Linear dependence among columns of $\boldsymbol{X}$ $\Leftrightarrow$ perfect multicollinearity $\Leftrightarrow$ no unique OLS.

7 Projection Matrices

7.1 The Hat Matrix $\boldsymbol{P}$

$$\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

Properties:

- **Symmetric**: $\boldsymbol{P}' = \boldsymbol{P}$.
- **Idempotent**: $\boldsymbol{P}^2 = \boldsymbol{P}$ (projecting twice does nothing extra).
- $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$ (anything in the column space is unchanged).
- $\boldsymbol{P}\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{y}}$ (fitted values).
- $\text{tr}(\boldsymbol{P}) = \text{rank}(\boldsymbol{P}) = k$.
- Diagonal entries: $h_{ii} = \boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i \in [0,1]$ (leverage).

Where it appears:

- **Day 1:** Introduced as the geometric heart of OLS—projection of $\boldsymbol{y}$ onto $\text{col}(\boldsymbol{X})$.
- **Day 3:** Computing fitted values, residuals, and showing $\hat{\boldsymbol{y}}'\hat{\boldsymbol{e}} = \mathbf{0}$.
- **Day 4:** Diagonal entries h_{ii} measure how much each observation influences fitted values.

7.2 The Annihilator Matrix $\boldsymbol{M}$

$$\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{P} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

Properties:

- Symmetric and idempotent.
- $\boldsymbol{M}\boldsymbol{X} = \mathbf{0}$ (annihilates anything in $\text{col}(\boldsymbol{X})$).
- $\boldsymbol{M}\boldsymbol{y} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \hat{\boldsymbol{e}}$ (residuals).
- $\text{tr}(\boldsymbol{M}) = n - k$.
- $\hat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{e}$ (residuals are a linear function of the true errors).

Where it appears:

- **Day 3:** Residuals as $\boldsymbol{My}$; demeaning as $\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{\iota}\bar{Y}$; variance decomposition.
- **Day 4:** Frisch-Waugh-Lovell: $\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{y}$. The residual maker $\boldsymbol{M}_1$ "partials out" $\boldsymbol{X}_1$.
- **Day 5:** $\text{Var}(\hat{\boldsymbol{e}}|\boldsymbol{X}) = \boldsymbol{MDM}$ (heteroskedastic case) or $\sigma^2\boldsymbol{M}$ (homoskedastic).

7.3 Idempotent Matrices

$\boldsymbol{A}$ is **idempotent** if $\boldsymbol{A}^2 = \boldsymbol{A}$.

- Eigenvalues of an idempotent matrix are either 0 or 1.
- $\text{rank}(\boldsymbol{A}) = \text{tr}(\boldsymbol{A}) =$ (number of eigenvalues equal to 1).
- Both $\boldsymbol{P}$ and $\boldsymbol{M}$ are symmetric and idempotent.

Where it appears:

- **Day 1:** Proof that eigenvalues of an idempotent matrix are 0 or 1.
- **Day 3:** Simplification: $\boldsymbol{M}'\boldsymbol{M} = \boldsymbol{M}^2 = \boldsymbol{M}$, so $\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} = \boldsymbol{e}'\boldsymbol{Me}$.
- **Day 7:** If $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{A}$ is idempotent, then $\boldsymbol{z}'\boldsymbol{Az} \sim \chi^2(\text{tr}(\boldsymbol{A}))$.

8 Partitioned Matrices

8.1 Block Matrix Multiplication

If $\boldsymbol{X} = [\boldsymbol{X}_1 \;\; \boldsymbol{X}_2]$, then:

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{pmatrix}, \qquad \boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} \boldsymbol{X}_1'\boldsymbol{y} \\ \boldsymbol{X}_2'\boldsymbol{y} \end{pmatrix}$$

8.2 Schur Complement and Partitioned Inverse

For a block matrix $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}$, the **Schur complement** of $\boldsymbol{A}_{22}$ is:

$$\boldsymbol{A}_{11\cdot 2} = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$$

The $(1,1)$ block of $\boldsymbol{A}^{-1}$ is $\boldsymbol{A}_{11\cdot 2}^{-1}$.

Where it appears:

- **Day 2:** Partitioned regression: $\beta_1 = \boldsymbol{Q}_{11\cdot 2}^{-1}\boldsymbol{Q}_{1Y\cdot 2}$ — the coefficient on X_1 after removing the linear influence of X_2.
- **Day 4:** Full derivation of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ in block form. Foundation for FWL theorem and partitioned regression.

8.3 Frisch-Waugh-Lovell (FWL) Theorem

In the regression $\boldsymbol{y} = \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2 + \hat{\boldsymbol{e}}$:

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{y}$$

where $\boldsymbol{M}_1 = \boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'$.

Interpretation: $\hat{\boldsymbol{\beta}}_2$ equals the coefficient from regressing the residuals of $\boldsymbol{y}$ on $\boldsymbol{X}_1$ against the residuals of $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$.

Where it appears:

- **Day 4:** Proved via the normal equations and the partitioned inverse. Shows how projection matrices decompose multivariate regression into sequential bivariate regressions.

9 Variance-Covariance Matrices

9.1 Definition and the Sandwich Formula

For a random vector $\boldsymbol{X}$:

$$\text{Var}(\boldsymbol{X}) = \mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])'] = \boldsymbol{\Sigma}$$

$\boldsymbol{\Sigma}$ is symmetric and PSD. For linear transformations:

$$\text{Var}(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$$

Applied to OLS:

$$\text{Var}(\hat{\boldsymbol{\beta}}|\boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \quad \text{(sandwich)}$$

where $\boldsymbol{D} = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$.
Under homoskedasticity ($\boldsymbol{D} = \sigma^2\boldsymbol{I}$): $\text{Var}(\hat{\boldsymbol{\beta}}|\boldsymbol{X}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

Where it appears:

- **Day 1:** $\text{Var}(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$ introduced as a recurring formula.
- **Day 3:** Derivation of $\text{Var}(\hat{\boldsymbol{\beta}}|\boldsymbol{X})$ using $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}$.
- **Day 5:** The sandwich formula with general $\boldsymbol{D}$; homoskedastic simplification; Gauss-Markov proof that $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is the smallest possible variance.
- **Day 5 (GLS):** $\text{Var}(\hat{\boldsymbol{\beta}}_{GLS}) = \sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$, the efficiency lower bound.
- **Day 6:** Different choices of diagonal matrix $\boldsymbol{D}$ in the sandwich formula (e.g., $\text{diag}(\hat{e}_i^2)$, $\text{diag}(\hat{e}_i^2/(1 - h_{ii}))$) yield different variance estimators.

9.2 Diagonal Matrices

$\boldsymbol{D} = \text{diag}(d_1, \dots, d_n)$ has d_i on the diagonal and 0s elsewhere.

- $\boldsymbol{Dx}$ scales each element: $(\boldsymbol{Dx})_i = d_i x_i$.
- $\boldsymbol{D}^{-1} = \text{diag}(1/d_1, \dots, 1/d_n)$ (if all $d_i \neq 0$).
- $\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X} = \sum_{i=1}^n d_i \boldsymbol{x}_i \boldsymbol{x}_i'$.

Where it appears:

- **Day 5:** $\boldsymbol{D} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ in the heteroskedastic variance formula.
- **Day 4:** $\text{diag}((1 - h_{11})^{-1}, \dots, (1 - h_{nn})^{-1})$ arises from inverting diagonal entries of $\boldsymbol{I} - \boldsymbol{P}$.
- **Day 6:** Block-diagonal structure $\boldsymbol{\Sigma} = \text{blockdiag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_G)$ arises when observations are independent across groups but correlated within.

10 Spectral Decomposition and Matrix Square Roots

Any symmetric PD matrix $\boldsymbol{\Sigma}$ can be decomposed as:

$$\boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}' \quad \text{where} \quad \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n), \quad \boldsymbol{C}'\boldsymbol{C} = \boldsymbol{I}$$

$\boldsymbol{C}$ is the matrix of orthonormal eigenvectors; $\boldsymbol{\Lambda}$ contains the eigenvalues.

The matrix square root and its inverse:

$$\boldsymbol{\Sigma}^{1/2} = \boldsymbol{C}\boldsymbol{\Lambda}^{1/2}\boldsymbol{C}', \qquad \boldsymbol{\Sigma}^{-1/2} = \boldsymbol{C}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{C}'$$

Key property: $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\,\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{I}$.

Where it appears:

- **Day 5 (GLS):** The transformation $\tilde{\boldsymbol{y}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}$, $\tilde{\boldsymbol{X}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}$ converts a heteroskedastic model into a homoskedastic one. OLS on the transformed data gives GLS:

$$\hat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$$

- **Day 5:** The efficiency lower bound proof transforms any linear estimator through $\boldsymbol{\Sigma}^{-1/2}$ and then applies Gauss-Markov in the spherical world.

11 The Normal Distribution and Linear Transformations

If $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{z} = \boldsymbol{g} + \boldsymbol{H}\boldsymbol{y}$ where $\boldsymbol{H}$ has full row rank, then:

$$\boldsymbol{z} \sim N(\boldsymbol{g} + \boldsymbol{H}\boldsymbol{\mu},\ \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}')$$

Special case: if $\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta},\ \sigma^2\boldsymbol{I})$, then $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ gives:

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta},\ \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$$

Under joint normality, uncorrelated $\Rightarrow$ independent.

Where it appears:

- **Day 7:** The distribution of $\hat{\boldsymbol{\beta}}$ is derived as a linear transformation of the multivariate normal $\boldsymbol{y}$.
- **Day 7:** Independence of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{e}}$ (they are jointly normal and uncorrelated because $\text{Cov}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \hat{\boldsymbol{e}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M} = \boldsymbol{0}$). This independence is what makes the t-statistic work.
- **Day 7:** Conditional distributions of multivariate normals are obtained via block partitioning of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$; the conditional mean is linear in the conditioning variable.

12 Summary: Concepts by Lecture Day

Lecture	Linear Algebra Concepts Used
Day 1 (Review)	Vectors, inner product, orthogonality; matrix multiplication, transpose; design matrix $\boldsymbol{X}$; matrix inverse; positive definiteness; quadratic forms; column space, span, basis, dimension; projection $\boldsymbol{P}$ and annihilator $\boldsymbol{M}$; idempotent matrices; eigenvalues; trace and rank; degrees of freedom as $\dim(\text{col}(\boldsymbol{X})^{\perp})$
Day 2 (CEF/BLP)	$\boldsymbol{Q}_{XX} = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}']$ and its inverse; partitioned $\boldsymbol{Q}$ matrices; Schur complement for multivariate regression
Day 3 (OLS)	Expanding $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$; projection and annihilator matrices; idempotent simplifications ($\boldsymbol{M}'\boldsymbol{M} = \boldsymbol{M}$); trace trick ($\mathbb{E}[\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}] = \text{tr}(\boldsymbol{A}\boldsymbol{\Sigma})$); orthogonal decomposition ($\boldsymbol{y} = \hat{\boldsymbol{y}} + \hat{\boldsymbol{e}}$)
Day 4 (Sensitivity)	Partitioned inverse (Schur complement); block matrix multiplication; diagonal entries of $\boldsymbol{P}$ (h_{ii})
Day 5 (GLS)	Sandwich formula ($\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$); diagonal matrices; PSD matrix ordering ($\boldsymbol{C}'\boldsymbol{C} \geq 0$); spectral decomposition; matrix square root $\boldsymbol{\Sigma}^{-1/2}$
Day 6 (Heteroskedasticity)	Weighted quadratic forms ($\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}$); block-diagonal matrices; diagonal scaling
Day 7 (MLE/Normal)	Multivariate normal distribution; linear transformations of normals; independence via uncorrelatedness under joint normality; quadratic forms in normal vectors (χ^2 from idempotent matrices)