Econometrics II

Lecture 1 - Multivariate Regression

Lasha Chochua

2026

ISET

Introduction

- **Multivariate regression** is a system of regression equations.
- Used in various contexts:
  - Reduced form models for instrumental variable estimation
  - Demand systems (demand for multiple goods)
- Also known as **systems of regression equations**.
- Related to **Seemingly Unrelated Regressions (SUR)**.
- Many single-equation regression tools generalize to multivariate regression.
- A major difference: **new notation to handle matrix estimators**.

Regression Systems

- A **univariate** linear regression equation:

$$Y = X'\beta + e$$

where Y is scalar and X is a vector.

- **Multivariate regression** is a system of m linear regressions:

$$Y_j = X_j'\beta_j + e_j \quad (1)$$

for $j = 1, \ldots, m$.

- Here, j denotes the j^{th} **dependent variable**, not the i^{th} individual.

Regression Systems (Cont.)

- Y_j could represent **household expenditures** on different goods (e.g., food, housing, clothing).
- The regressors:
  - X_j is a $k_j \times 1$ vector.
  - The coefficient vectors β_j are $k_j \times 1$.
  - The error term e_j is included.
- Total number of coefficients:

$$\bar{k} = \sum_{j=1}^{m} k_j$$

Regression System - Example (Cont.)

- Consider a **household expenditure model**.
- The equations for **housing, food, and clothing expenditures** are:

$$\begin{aligned} \textit{housing} &= \beta_{10} + \beta_{11}\textit{houseprc} + \beta_{12}\textit{foodprc} + \beta_{13}\textit{clothprc} + \beta_{14}\textit{income} \\ &\quad + \beta_{15}\textit{size} + \beta_{16}\textit{age} + e_1. \\ \textit{food} &= \beta_{20} + \beta_{21}\textit{houseprc} + \beta_{22}\textit{foodprc} + \beta_{23}\textit{clothprc} + \beta_{24}\textit{income} \\ &\quad + \beta_{25}\textit{size} + \beta_{26}\textit{age} + e_2. \\ \textit{clothing} &= \beta_{30} + \beta_{31}\textit{houseprc} + \beta_{32}\textit{foodprc} + \beta_{33}\textit{clothprc} + \beta_{34}\textit{income} \\ &\quad + \beta_{35}\textit{size} + \beta_{36}\textit{age} + e_3. \end{aligned}$$

- Each equation represents expenditure on a category (*housing, food, clothing*).
- The independent variables (*houseprc, foodprc, clothprc, income, size, age*) affect all three dependent variables. The error terms e_1, e_2, e_3 capture unexplained variations.

Regression Systems (Cont.)

- **Regressors can be common** across j or vary across j.
  - Example: Household expenditure models often use common regressors like **income, family size, and demographic factors**.
- When $m = 1$, the system **reduces to univariate regression**.

Error Structure and Covariance

- Define the **error vector**:

$$e = (e_1, \dots, e_m)'$$

- The **error covariance matrix**:

$$\Sigma = E[ee']$$

  - **Diagonal elements**: variances of errors e_j.
  - **Off-diagonal elements**: covariances across variables.

Matrix Form of the System

- The m equations from (1) can be grouped into a **single equation**.
- Define the **vector of dependent variables**:

$$Y = (Y_1, \ldots, Y_m)'$$

- Define the $m \times \bar{k}$ **regressor matrix**:

$$\bar{X} = \begin{pmatrix} X_1' & 0 & \cdots & 0 \\ \vdots & X_2' & \vdots & \vdots \\ 0 & 0 & \cdots & X_m' \end{pmatrix}$$

Stacked Coefficient Vector and System Representation

- The $\bar{k} \times 1$ **stacked coefficient vector**:

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}.$$

- The m regression equations can be **jointly written** as:

$$Y = \bar{X}\beta + e. \tag{2}$$

- This represents a **system of m equations**.

Matrix Representation for n Observations

- For n observations, the joint system can be written in **matrix notation** by stacking:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}, \quad \bar{\mathbf{X}} = \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_n \end{pmatrix}$$

Matrix Representation for n Observations

- For n observations, the joint system can be written in **matrix notation** by stacking:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}, \quad \bar{\mathbf{X}} = \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_n \end{pmatrix}$$

- Dimensions:
  - $\mathbf{Y}$ is $mn \times 1$
  - $\mathbf{e}$ is $mn \times 1$
  - $\bar{\mathbf{X}}$ is $mn \times \bar{k}$
- The system can be written as:

$$\mathbf{Y} = \bar{\mathbf{X}}\beta + \mathbf{e}.$$

Special Case: Common Regressors

- In many applications, the regressors X_j are **common across variables** j, meaning $X_j = X$ and $k_j = k$.
- This implies the same variables enter each equation with **no exclusion restrictions**.
- A useful simplification is:

$$Y_i = \mathbf{B}' X_i + e_i \tag{3}$$

where $\mathbf{B} = (\beta_1, \beta_2, \dots, \beta_m)$ is $k \times m$.

Joint System in $n \times m$ Notation

- The joint system for n **observations** is written as:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$$

where:

$$\mathbf{Y} = \begin{pmatrix} Y_1' \\ \vdots \\ Y_n' \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} e_1' \\ \vdots \\ e_n' \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} X_1' \\ \vdots \\ X_n' \end{pmatrix}.$$

Kronecker Product Representation

- A convenient simplification when **regressors are common**:

$$\bar{X} = \begin{pmatrix} X' & 0 & \cdots & 0 \\ 0 & X' & 0 & \cdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & X' \end{pmatrix} = \mathbf{I}_m \otimes X'$$

where $\otimes$ is the **Kronecker product**.

The Kronecker Product - Example

- The **Kronecker product** of two matrices $A \otimes B$ expands A by multiplying each element of A with the entire matrix B.
- **Example:**
- Given matrices:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix}$$

- The **Kronecker product** $A \otimes B$ is:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} 1 \cdot \mathbf{B} & 2 \cdot \mathbf{B} \\ 3 \cdot \mathbf{B} & 4 \cdot \mathbf{B} \end{pmatrix} = \begin{pmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{pmatrix}$$

Least Squares Estimator

- The equations (1) can be estimated using **least squares**.
- The estimator for β_j is given by:

$$\hat{\beta}_j = \left(\sum_{i=1}^{n} X_{ji} X_{ji}' \right)^{-1} \left(\sum_{i=1}^{n} X_{ji} Y_{ji} \right).$$

- An estimator of the **stacked coefficient vector** β is:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_m \end{pmatrix}.$$

Least Squares Estimator in Systems Notation

- The least squares estimator can also be written in **systems notation**:

$$\hat{\beta} = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}(\bar{\mathbf{X}}'\mathbf{Y}) = \left(\sum_{i=1}^{n} \bar{X}_i'\bar{X}_i\right)^{-1}\left(\sum_{i=1}^{n} \bar{X}_i' Y_i\right). \quad (4)$$

Derivation of $\bar{\mathbf{X}}'\bar{\mathbf{X}}$

- To see this, observe that:

$$\bar{\mathbf{X}}'\bar{\mathbf{X}} = \begin{pmatrix} \bar{X}_1' & \dots & \bar{X}_n' \end{pmatrix} \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_n \end{pmatrix}$$

- Expanding:

$$\bar{\mathbf{X}}'\bar{\mathbf{X}} = \sum_{i=1}^{n} \bar{X}_i' \bar{X}_i.$$

Note

- For a partitioned column matrix, the transpose rule is:

$$\overline{\mathbf{X}} = \begin{pmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \vdots \\ \overline{X}_n \end{pmatrix} \quad \Rightarrow \quad \overline{\mathbf{X}}' = \begin{pmatrix} \overline{X}_1' & \overline{X}_2' & \cdots & \overline{X}_n' \end{pmatrix}$$

- You **both** change the arrangement from column to row **and** transpose each individual block.

Derivation of $\bar{\mathbf{X}}'\bar{\mathbf{X}}$ (Cont.)

- Since $\bar{X}_i$ has a **block-diagonal structure**, we write:

$$\sum_{i=1}^{n} \begin{pmatrix} X_{1i} & 0 & \cdots & 0 \\ 0 & X_{2i} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{mi} \end{pmatrix} \begin{pmatrix} X'_{1i} & 0 & \cdots & 0 \\ 0 & X'_{2i} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X'_{mi} \end{pmatrix}$$

- The result is:

$$\bar{\mathbf{X}}'\bar{\mathbf{X}} = \begin{pmatrix} \sum_{i=1}^{n} X_{1i}X'_{1i} & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^{n} X_{2i}X'_{2i} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{i=1}^{n} X_{mi}X'_{mi} \end{pmatrix}.$$

Computing $\bar{\mathbf{X}}'\mathbf{Y}$

- We showed that:

$$\bar{\mathbf{X}}'\mathbf{Y} = \begin{pmatrix} \bar{X}_1' & \dots & \bar{X}_n' \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

- Expanding:

$$\bar{\mathbf{X}}'\mathbf{Y} = \sum_{i=1}^{n} \bar{X}_i' Y_i.$$

Computing $\bar{\mathbf{X}}'\mathbf{Y}$

- Given the block structure of $\bar{X}_i$:

$$\sum_{i=1}^{n} \begin{pmatrix} X_{1i} & 0 & \cdots & 0 \\ 0 & X_{2i} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{mi} \end{pmatrix} \begin{pmatrix} Y_{1i} \\ \vdots \\ Y_{mi} \end{pmatrix}$$

- This results in:

$$\bar{\mathbf{X}}'\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^{n} X_{1i}Y_{1i} \\ \vdots \\ \sum_{i=1}^{n} X_{mi}Y_{mi} \end{pmatrix}.$$

Final Expression for $\hat{\beta}$

- We now compute:

$$(\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}(\bar{\mathbf{X}}'\mathbf{Y})$$

- Expanding:

$$\left(\sum_{i=1}^{n} X_i X_i'\right)^{-1}\left(\sum_{i=1}^{n} X_i Y_i\right)$$

Inverse of a Block Diagonal Matrix

- Let

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m \end{pmatrix}$$

  be a **block diagonal matrix**, where each block A_j is square.

- If every block A_j is **invertible**, then

$$A^{-1} = \begin{pmatrix} A_1^{-1} & 0 & \cdots & 0 \\ 0 & A_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_m^{-1} \end{pmatrix}.$$

- The inverse of a block diagonal matrix is the block diagonal matrix of the inverses.

Final Expression for $\hat{\beta}$

- Since $\bar{\mathbf{X}}'\bar{\mathbf{X}}$ is block diagonal:

$$\begin{pmatrix} \left(\sum_{i=1}^{n} X_{1i}X_{1i}'\right)^{-1}\left(\sum_{i=1}^{n} X_{1i}Y_{1i}\right) \\ \vdots \\ \left(\sum_{i=1}^{n} X_{mi}X_{mi}'\right)^{-1}\left(\sum_{i=1}^{n} X_{mi}Y_{mi}\right) \end{pmatrix}$$

- Thus, we obtain:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_m \end{pmatrix}.$$

Residual Vector and Error Covariance Matrix

- The $m \times 1$ **residual vector** for the i^{th} observation is:

$$\hat{e}_i = Y_i - \bar{\mathbf{X}}_i' \hat{\beta}.$$

- The **least squares estimator of the** $m \times m$ **error covariance matrix** is:

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i \hat{e}_i'. \tag{5}$$

Least Squares Coefficients with Common Regressors

- When regressors are **common** across equations, the least squares estimator for β_j is:

$$\hat{\beta}_j = \left(\sum_{i=1}^{n} X_i X_i' \right)^{-1} \left(\sum_{i=1}^{n} X_i Y_{ji} \right).$$

- The stacked coefficient matrix $\hat{B}$ is:

$$\hat{\mathbf{B}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}). \tag{6}$$

Expectation and Variance of $\hat{\beta}$

- The **finite-sample expectation** of $\hat{\beta}$ under the conditional expectation assumption:

$$E[e \mid X] = 0. \tag{7}$$

- Here, X is the union of the regressors X_j.
- Equation (7) implies that:

$$E[Y_j \mid X] = X_j' \beta_j,$$

which means the **regression model is correctly specified**.

Centering the Estimator

- We express the centered estimator as:

$$\hat{\beta} - \beta = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}(\bar{\mathbf{X}}'\mathbf{e}) = \left(\sum_{i=1}^{n} \bar{X}_i'\bar{X}_i\right)^{-1}\left(\sum_{i=1}^{n} \bar{X}_i' e_i\right).$$

- Taking expectations:

$$E[\hat{\beta} \mid \mathbf{X}] = \beta.$$

Variance of the Estimator

- Define the **conditional covariance matrix of the errors** for the i^{th} observation:

$$E[e_i e_i' \mid X_i] = \Sigma_i.$$

- If observations are **mutually independent**, then:

$$E[\mathbf{e}\mathbf{e}' \mid \mathbf{X}] = E\left[\begin{bmatrix} e_1 e_1' & e_1 e_2' & \cdots & e_1 e_n' \\ \vdots & \ddots & \vdots & \vdots \\ e_n e_1' & e_n e_2' & \cdots & e_n e_n' \end{bmatrix} \middle| \mathbf{X} \right] = \begin{bmatrix} \Sigma_1 & 0 & \cdots & 0 \\ 0 & \Sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_n \end{bmatrix}.$$

Variance of the Estimator

- By **independence across observations**:

$$\text{var}\left[\sum_{i=1}^{n} \bar{X}_i' e_i \middle| \mathbf{X}\right] = \sum_{i=1}^{n} \text{var}\left[\bar{X}_i' e_i \mid X_i\right] = \sum_{i=1}^{n} \bar{X}_i' \Sigma_i \bar{X}_i.$$

- It follows that:

$$\text{var}[\widehat{\beta} \mid \mathbf{X}] = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1} \left(\sum_{i=1}^{n} \bar{X}_i' \Sigma_i \bar{X}_i\right) (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}.$$

Special Case: Common Regressors

- When regressors are **common**, so that $\bar{X}_i = \mathbf{I}_m \otimes X_i'$, then the **covariance matrix** simplifies to:

$$\text{var}[\hat{\beta} \mid \mathbf{X}] = \left(\mathbf{I}_m \otimes (\mathbf{X}'\mathbf{X})^{-1}\right) \left(\sum_{i=1}^{n} (\Sigma_i \otimes X_i' X_i) \right) \left(\mathbf{I}_m \otimes (\mathbf{X}'\mathbf{X})^{-1}\right).$$

Homoskedastic Case

- If **errors are conditionally homoskedastic**:

$$E[ee' \mid \mathbf{X}] = \Sigma. \tag{8}$$

- The **covariance matrix** then simplifies to:

$$\text{var}[\hat{\beta} \mid \mathbf{X}] = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1} \left(\sum_{i=1}^{n} \bar{X}_i' \Sigma \bar{X}_i \right) (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}.$$

Final Simplification

- If **both conditions hold** (common regressors and homoskedasticity), we get:

$$\text{var}[\hat{\beta} \mid \mathbf{X}] = \Sigma \otimes (\mathbf{X}'\mathbf{X})^{-1}.$$

Asymptotic Distribution

- For asymptotics, we consider the **equation-by-equation projection model**, where:

$$E\left[X_j e_j\right] = 0. \tag{9}$$

Consistency of $\hat{\beta}_j$

- Since $\hat{\beta}_j$ are standard least squares estimators, they are **consistent** for the projection coefficients β_j.

Asymptotic Normality

- Single equation theory implies $\hat{\beta}_j$ are **asymptotically normal**.
- However, this does not provide a **joint distribution** for $\hat{\beta}_j$ across j.

Derivation of the Joint Distribution

- Consider the vector:

$$\bar{X}_i' e_i = \begin{pmatrix} X_{1i} e_{1i} \\ \vdots \\ X_{mi} e_{mi} \end{pmatrix}.$$

- Since this vector is **i.i.d. across** i and has **mean zero** under (9), the **central limit theorem** implies:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{X}_i' e_i \xrightarrow{d} N(0, \Omega).$$

- Where:

$$\Omega = E\left[\bar{X}_i' e_i e_i' \bar{X}_i\right] = E\left[\bar{X}_i' \Sigma_i \bar{X}_i\right].$$

Covariance Matrix Ω

- The matrix Ω is the **covariance matrix** of the variables $X_{ji}e_{ji}$ across equations.
- Under **conditional homoskedasticity** (8), the matrix Ω simplifies to:

$$\Omega = E\left[\bar{X}_i' \Sigma \bar{X}_i\right]. \tag{10}$$

Special Cases

- When the **regressors are common**, Ω simplifies further:

$$\Omega = E\left[ee' \otimes XX'\right]. \tag{11}$$

- Under **both conditions** (homoskedasticity and common regressors), Ω simplifies to:

$$\Omega = \Sigma \otimes E\left[XX'\right]. \tag{12}$$

Asymptotic Distribution of $\hat{\beta}$

Assumption 7.2 (Hansen, *Econometrics*)

Let (Y_i, X_i), $i = 1, \ldots, n$, denote the sample observations. The following conditions hold:

1. The variables (Y_i, X_i) are **independent and identically distributed (i.i.d.)**.
2. The fourth moment of the dependent variable exists: $E[Y^4] < \infty$
3. The regressors have finite fourth moments: $E\|X\|^4 < \infty$
4. The second moment matrix of the regressors $Q_{XX} = E[XX']$ is **positive definite**.

Asymptotic Distribution of $\hat{\beta}$ (Cont.)

Theorem 1 (Asymptotic Distribution of $\hat{\beta}$)

Under Assumption 7.2,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \mathbf{V}_\beta),$$

where:

$$\mathbf{V}_\beta = \mathbf{Q}^{-1}\Omega\mathbf{Q}^{-1},$$

and

$$Q = E\left[\bar{\mathbf{X}}'\bar{\mathbf{X}}\right] = \begin{pmatrix} E[X_1X_1'] & 0 & \cdots & 0 \\ 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & E[X_mX_m'] \end{pmatrix}.$$

Special Cases

- When regressors are **common**, the matrix Q simplifies as:

$$\mathbf{Q} = \mathbf{I}_m \otimes E[XX'] \tag{13}$$

- If regressors are common **and** errors are **conditionally homoskedastic** (8), then:

$$\mathbf{V}_\beta = \Sigma \otimes \left(E[XX']\right)^{-1}. \tag{14}$$

Asymptotic Distribution of $\hat{\theta}$

Assumption 7.3 (Hansen, Econometrics)

Let

$$r(\beta) : \mathbb{R}^k \to \mathbb{R}^q$$

denote the vector of restrictions.

Assume:

1. The function $r(\beta)$ is **continuously differentiable** at the true parameter value β.
2. The Jacobian matrix

$$R = \frac{\partial}{\partial \beta} r(\beta)'$$

has **rank** q.

Asymptotic Distribution of $\hat{\theta}$ (Cont.)

- Sometimes we are interested in parameters $\theta = r(\beta_1, \ldots, \beta_m) = r(\beta)$, which are **functions of the coefficients** from multiple equations.
- In this case, the least squares estimator of θ is $\hat{\theta} = r(\hat{\beta})$.
- The **asymptotic distribution** of $\hat{\theta}$ can be obtained from Theorem 1 by the **delta method**.

Theorem 2 (Asymptotic Distribution of Functions of Parameters)
Under Assumptions 7.2 and 7.3 in the book, as $n \to \infty$,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathbf{V}_\theta),$$

where: $\mathbf{V}_\theta = \mathbf{R}'\mathbf{V}_\beta\mathbf{R}$, and $\mathbf{R} = \frac{\partial}{\partial \beta} r(\beta)'$.

Covariance Matrix Estimation

- From **finite sample and asymptotic theory**, we can construct **appropriate estimators** for the variance of $\hat{\beta}$.
- In the general case, we have:

$$\hat{\mathbf{V}}_{\hat{\beta}} = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1} \left(\sum_{i=1}^{n} \bar{X}_i' \hat{e}_i \hat{e}_i' \bar{\mathbf{X}}_i \right) (\bar{\mathbf{X}}'\bar{X})^{-1}.$$

- Under **conditional homoskedasticity** (8), an appropriate estimator is:

$$\hat{\mathbf{V}}^0_{\hat{\beta}} = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1} \left(\sum_{i=1}^{n} \bar{X}_i' \hat{\Sigma} \bar{X}_i \right) (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}.$$

Special Case: Common Regressors

- When regressors are **common**, the estimators simplify to:

$$\hat{\mathbf{V}}_{\hat{\beta}} = \left(I_m \otimes (X'X)^{-1}\right) \left(\sum_{i=1}^{n} (\hat{e}_i \hat{e}_i' \otimes X_i' X_i)\right) \left(I_m \otimes (X'X)^{-1}\right).$$

- And for the homoskedastic case:

$$\hat{\mathbf{V}}^0_{\hat{\beta}} = \hat{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}.$$

Covariance Estimators for $\hat{\theta}$

- The **covariance matrix estimators for** $\hat{\theta}$ are:

$$\hat{\mathbf{V}}_{\hat{\theta}} = \hat{\mathbf{R}}'\hat{\mathbf{V}}_{\hat{\beta}}\hat{\mathbf{R}}$$

$$\hat{\mathbf{V}}^0_{\hat{\theta}} = \hat{\mathbf{R}}'\hat{\mathbf{V}}^0_{\hat{\beta}}\hat{\mathbf{R}}$$

$$\hat{\mathbf{R}} = \frac{\partial}{\partial\beta} r(\hat{\beta})'.$$

Asymptotic Behavior of Variance Estimators

Theorem 3 (Asymptotic Behavior of Variance Estimators)
Under Assumption 7.2 in the book,

$$n\hat{\mathbf{V}}_{\hat{\beta}} \xrightarrow{p} \mathbf{V}_{\beta} \quad \text{and} \quad n\hat{\mathbf{V}}^0_{\hat{\beta}} \xrightarrow{p} \mathbf{V}^0_{\beta}.$$

Seemingly Unrelated Regression (SUR)

- Consider the **systems regression model** under the **conditional expectation and homoskedasticity** assumptions:

$$Y = \bar{X}\beta + e$$

$$E[e \mid X] = 0$$

$$E[ee' \mid X] = \Sigma. \tag{15}$$

- Since **errors are correlated across equations**, we use **Generalized Least Squares (GLS)**.

Generalized Least Squares (GLS) Estimator

- To derive the estimator:
  - **Premultiply** equation (15) by $\Sigma^{-1/2}$.
  - The transformed error vector becomes **i.i.d.** with covariance matrix $\mathbf{I}_m$.
  - Apply **least squares** and rearrange to find:

$$\hat{\beta}_{\text{gls}} = \left(\sum_{i=1}^{n} X_i' \Sigma^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{n} X_i' \Sigma^{-1} Y_i \right). \tag{16}$$

Vector Representation of GLS

- Alternatively, using the **vector representation**:

$$\mathbf{Y} = \bar{\mathbf{X}}\beta + \mathbf{e}$$

- The **error variance** is:

$$E[\mathbf{e}\mathbf{e}'] = \mathbf{I}_n \otimes \Sigma.$$

- **Premultiply** the equation by $\mathbf{I}_n \otimes \Sigma^{-1/2}$ so that the transformed error has covariance matrix $\mathbf{I}_{nm}$.
- Then apply **least squares** to find:

$$\hat{\beta}_{\text{gls}} = \left(\bar{\mathbf{X}}'(\mathbf{I}_n \otimes \Sigma^{-1})\bar{\mathbf{X}}\right)^{-1}\left(\bar{\mathbf{X}}'(\mathbf{I}_n \otimes \Sigma^{-1})\mathbf{Y}\right). \quad (17)$$

Equivalence of GLS Representations

- Expressions **(16) and (17) are algebraically equivalent**.
- To see the equivalence, observe that:

$$\bar{\mathbf{X}}'(\mathbf{I}_n \otimes \Sigma^{-1})\bar{\mathbf{X}} = \begin{pmatrix} \bar{X}_1' & \cdots & \bar{X}_n' \end{pmatrix} \begin{pmatrix} \Sigma^{-1} & 0 & \cdots & 0 \\ 0 & \Sigma^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma^{-1} \end{pmatrix} \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_n \end{pmatrix}$$

- Which simplifies to:

$$\sum_{i=1}^{n} \bar{X}_i' \Sigma^{-1} \bar{X}_i.$$

Equivalence of GLS Representations (Cont.)

- Similarly:

$$\bar{\mathbf{X}}'(\mathbf{I}_n \otimes \Sigma^{-1})\mathbf{Y} = \begin{pmatrix} \bar{X}_1' & \dots & \bar{X}_n' \end{pmatrix} \begin{pmatrix} \Sigma^{-1} & 0 & \cdots & 0 \\ 0 & \Sigma^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma^{-1} \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

- Which simplifies to:

$$\sum_{i=1}^{n} \bar{X}_i' \Sigma^{-1} Y_i.$$

Feasible Generalized Least Squares (FGLS) Estimator

- Since Σ **is unknown**, it must be replaced by an estimator $\hat{\Sigma}$ from (5).
- Using this, we obtain the **feasible GLS estimator**:

$$\hat{\beta}_{\text{sur}} = \left(\sum_{i=1}^{n} \bar{X}_i' \hat{\Sigma}^{-1} \bar{X}_i \right)^{-1} \left(\sum_{i=1}^{n} \bar{X}_i' \hat{\Sigma}^{-1} Y_i \right)$$

$$= \left(\bar{\mathbf{X}}' (\mathbf{I}_n \otimes \hat{\Sigma}^{-1}) \bar{\mathbf{X}} \right)^{-1} \left(\bar{\mathbf{X}}' (\mathbf{I}_n \otimes \hat{\Sigma}^{-1}) \mathbf{Y} \right). \quad (18)$$

- This is the **Seemingly Unrelated Regression (SUR) estimator**, as introduced by **Zellner (1962)**.

Iterated SUR Estimator

- The estimator $\hat{\Sigma}$ can be updated by calculating the **SUR residuals**:

$$\hat{e}_i = Y_i - \bar{X}_i \hat{\beta}_{\textsf{sur}}.$$

- The covariance matrix estimator is:

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i \hat{e}_i'.$$

- Substituting this into (18), we obtain an **iterated SUR estimator**, which can be repeated **until convergence**.

Asymptotic Distribution of $\hat{\beta}_{\textsf{sur}}$

Theorem 11.4 (Asymptotic Distribution of $\hat{\beta}_{\textsf{sur}}$)
Under Assumption 7.2 and (8),

$$\sqrt{n}(\hat{\beta}_{\textsf{sur}} - \beta) \xrightarrow{d} N(0, \mathbf{V}_\beta^*),$$

where:

$$\mathbf{V}_\beta^* = \left(E\left[\bar{X}'\Sigma^{-1}\bar{X}\right]\right)^{-1}.$$

SUR vs. OLS Efficiency

- Under these assumptions, **SUR is more efficient than least squares**.

Theorem 5 (Efficiency of the SUR Estimator)
Under Assumption 7.2 in the book and (8),

$$\mathbf{V}_\beta^* = \left(E\left[\bar{X}'\Sigma^{-1}\bar{X}\right]\right)^{-1} \leq \left(E\left[\bar{X}'\bar{X}\right]\right)^{-1} E\left[\bar{X}'\Sigma\bar{X}\right]\left(E\left[\bar{X}'\bar{X}\right]\right)^{-1} = \mathbf{V}_\beta,$$

and thus $\widehat{\beta}_{\text{sur}}$ **is asymptotically more efficient than** $\widehat{\beta}_{\text{ols}}$.

Variance Estimation for $\hat{\beta}_{\text{sur}}$

- An appropriate estimator of the variance of $\hat{\beta}_{\text{sur}}$ is:

$$\hat{\mathbf{V}}_{\hat{\beta}} = \left(\sum_{i=1}^{n} \bar{X}_i' \hat{\Sigma}^{-1} \bar{X}_i \right)^{-1}.$$

Theorem 6 (Consistency of the Variance Estimator)
Under Assumption 7.2 in the book and (8),

$$n\hat{V}_{\hat{\beta}} \xrightarrow{p} V_{\beta}.$$

When SUR = OLS (Case 1: Common Regressors)

- Assume regressors are **common across equations**, so $X_{ji} = X_i$ and $k_j = k$ for all j.
- For each observation i:

$$\overline{X}_i = I_m \otimes X_i' \quad (m \times mk)$$

- Its transpose:

$$\overline{X}_i' = I_m \otimes X_i \quad (mk \times m)$$

1: Start from the SUR Estimator

- The SUR estimator is:

$$\hat{\beta}_{sur} = \left(\sum_{i=1}^{n} \overline{X}_i' \Sigma^{-1} \overline{X}_i \right)^{-1} \left(\sum_{i=1}^{n} \overline{X}_i' \Sigma^{-1} Y_i \right)$$

- This is the GLS formula applied observation by observation.
- We keep it in terms of $\overline{X}_i'$ and $\overline{X}_i$ for now — no substitution yet.

2: Establish the Key Identity

- We want to simplify $\overline{X}_i' \Sigma^{-1}$.
- Dimensions: $\overline{X}_i'$ is $mk \times m$, Σ^{-1} is $m \times m$, so the product is $mk \times m$.
- **Claim:**

$$\overline{X}_i' \Sigma^{-1} = \left(\Sigma^{-1} \otimes I_k\right) \overline{X}_i'$$

- We verify that **both sides equal** $\Sigma^{-1} \otimes X_i$.

2: Establish the Key Identity (Cont.)

- **Right-hand side** – substitute $\overline{X}_i' = I_m \otimes X_i$
- Apply the mixed-product rule $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$:

$$(\Sigma^{-1} \otimes I_k)(I_m \otimes X_i) = (\Sigma^{-1} I_m) \otimes (I_k X_i) = \Sigma^{-1} \otimes X_i.$$

2: Establish the Key Identity (Cont.)

- **Left-hand side** – inspect the (j,ℓ) block of $\overline{X}_i'\Sigma^{-1}$:
  - Since $\overline{X}_i' = I_m \otimes X_i$ is block-diagonal, its (j,ℓ) block is $\delta_{j\ell}X_i$, where $\delta_{j\ell}$ is the **Kronecker delta** – equal to 1 if $j=\ell$ and 0 otherwise.
  - Post-multiplying by Σ^{-1} weights each block by the corresponding entry of Σ^{-1}, which is just a scalar, and which we denote by $[\Sigma^{-1}]_{j\ell}$.
  - So the (j,ℓ) block of $\overline{X}_i'\Sigma^{-1}$ is:

$$\sum_{s=1}^{m} \delta_{js}\, X_i \cdot [\Sigma^{-1}]_{s\ell} = X_i \cdot [\Sigma^{-1}]_{j\ell}$$

    where the sum collapses because $\delta_{js} = 0$ unless $s = j$.
  - The (j,ℓ) block of $\Sigma^{-1} \otimes X_i$ is also $[\Sigma^{-1}]_{j\ell} \cdot X_i$ by definition of the Kronecker product.
  - Since $[\Sigma^{-1}]_{j\ell}$ is a scalar, both expressions are equal.
- Both sides equal $\Sigma^{-1} \otimes X_i$. $\blacksquare$

3: Simplify the Quadratic Term

- Right-multiply both sides of the identity by $\overline{X}_i$:

$$\overline{X}_i'\Sigma^{-1}\overline{X}_i = [(\Sigma^{-1} \otimes I_k)\overline{X}_i']\overline{X}_i = (\Sigma^{-1} \otimes I_k)[\overline{X}_i'\overline{X}_i]$$

  where the last step uses matrix associativity.

- Since $(\Sigma^{-1} \otimes I_k)$ does not depend on i, it factors out of the sum:

$$\sum_{i=1}^{n} \overline{X}_i'\Sigma^{-1}\overline{X}_i = (\Sigma^{-1} \otimes I_k)\sum_{i=1}^{n} \overline{X}_i'\overline{X}_i$$

4: Invert the Quadratic Term

- Both $(\Sigma^{-1} \otimes I_k)$ and $\sum_i \overline{X}_i' \overline{X}_i$ are square $mk \times mk$ matrices:
  - $(\Sigma^{-1} \otimes I_k)$ is invertible by positive definiteness of Σ.
  - $\sum_i \overline{X}_i' \overline{X}_i$ is invertible by the full-rank assumption on regressors.
- So $(AB)^{-1} = B^{-1} A^{-1}$ applies:

$$\left(\sum_{i=1}^{n} \overline{X}_i' \Sigma^{-1} \overline{X}_i \right)^{-1} = \left(\sum_{i=1}^{n} \overline{X}_i' \overline{X}_i \right)^{-1} (\Sigma^{-1} \otimes I_k)^{-1}$$

4: Invert the Quadratic Term (Cont.)

- Using $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$:

$$(\Sigma^{-1} \otimes I_k)^{-1} = \Sigma \otimes I_k$$

- Therefore:

$$\left(\sum_{i=1}^{n} \overline{X}_i' \Sigma^{-1} \overline{X}_i \right)^{-1} = \left(\sum_{i=1}^{n} \overline{X}_i' \overline{X}_i \right)^{-1} (\Sigma \otimes I_k)$$

5: Simplify the Linear Term

- Apply the same identity from Step 2, now right-multiplied by Y_i $(m \times 1)$:

$$\overline{X}_i' \Sigma^{-1} Y_i = (\Sigma^{-1} \otimes I_k) \overline{X}_i' Y_i$$

- Since $(\Sigma^{-1} \otimes I_k)$ does not depend on i, it factors out of the sum:

$$\sum_{i=1}^{n} \overline{X}_i' \Sigma^{-1} Y_i = (\Sigma^{-1} \otimes I_k) \sum_{i=1}^{n} \overline{X}_i' Y_i$$

6: Substitute Back into SUR

- Combining Steps 4 and 5:

$$\hat{\beta}_{sur} = \left[\left(\sum_{i=1}^{n} \overline{X}_i' \overline{X}_i\right)^{-1} (\Sigma \otimes I_k)\right] \left[(\Sigma^{-1} \otimes I_k) \sum_{i=1}^{n} \overline{X}_i' Y_i\right]$$

- The two middle Kronecker terms are adjacent — apply the mixed-product rule:

$$(\Sigma \otimes I_k)(\Sigma^{-1} \otimes I_k) = (\Sigma\Sigma^{-1}) \otimes (I_k I_k) = I_m \otimes I_k = I_{mk}$$

- They cancel completely.

7: Final Result

- Thus:

$$\hat{\beta}_{sur} = \left(\sum_{i=1}^{n} \overline{X}_i' \overline{X}_i \right)^{-1} \sum_{i=1}^{n} \overline{X}_i' Y_i$$

- But this is exactly the **stacked OLS estimator**:

$$\hat{\beta}_{ols} = \left(\sum_{i=1}^{n} \overline{X}_i' \overline{X}_i \right)^{-1} \sum_{i=1}^{n} \overline{X}_i' Y_i$$

- Therefore $\hat{\beta}_{sur} = \hat{\beta}_{ols}$. ■

Key Result

Note

If **all equations share the same regressors**, then

$$\hat{\beta}_{sur} = \hat{\beta}_{ols}.$$

SUR does **not change coefficient estimates**. Moreover, in this case, GLS weighting does not provide any additional efficiency gain over OLS. The cross-equation covariance matrix Σ cancels from the estimator, so SUR and OLS coincide.

When SUR = OLS (Case 2: No Cross-Equation Correlation)

- Suppose the covariance matrix of the errors is **diagonal**:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m^2 \end{pmatrix}$$

- Then $\Omega = I_n \otimes \Sigma$ and $\Omega^{-1} = I_n \otimes \Sigma^{-1}$.

1: Start from the SUR Estimator

- The SUR estimator is:

$$\hat{\beta}_{sur} = (\overline{\mathbf{X}}'\Omega^{-1}\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\Omega^{-1}\mathbf{Y}.$$

- Substitute $\Omega^{-1} = I_n \otimes \Sigma^{-1}$:

$$\hat{\beta}_{sur} = (\overline{\mathbf{X}}'(I_n \otimes \Sigma^{-1})\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'(I_n \otimes \Sigma^{-1})\mathbf{Y}.$$

2: Diagonal Structure of Σ^{-1}

- Since Σ is diagonal, its inverse is also diagonal:

$$\Sigma^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_m^2 \end{pmatrix}$$

- Since $\overline{X}_i$ is block-diagonal and Σ^{-1} is diagonal, the quadratic term is block-diagonal with j-th block:

$$\sum_{i=1}^{n} \overline{X}_i' \Sigma^{-1} \overline{X}_i \quad \Rightarrow \quad j\text{-th block: } \frac{1}{\sigma_j^2} \sum_{i=1}^{n} X_{ji} X_{ji}'$$

3: Effect on the Estimator

- Similarly, the j-th block of the linear term is:

$$\frac{1}{\sigma_j^2}\sum_{i=1}^{n} X_{ji}Y_{ji}$$

- So the j-th block of $\widehat{\beta}_{sur}$ is:

$$\left(\frac{1}{\sigma_j^2}\sum_{i=1}^{n} X_{ji}X_{ji}'\right)^{-1}\left(\frac{1}{\sigma_j^2}\sum_{i=1}^{n} X_{ji}Y_{ji}\right) = \left(\sum_{i=1}^{n} X_{ji}X_{ji}'\right)^{-1}\left(\sum_{i=1}^{n} X_{ji}Y_{ji}\right)$$

- The scalar $1/\sigma_j^2$ **cancels** from numerator and denominator.
- This is exactly $\widehat{\beta}_j^{ols}$ – the OLS estimator for equation j.

Result

- Therefore:

$$\hat{\beta}_{sur} = \hat{\beta}_{ols}$$

- when the covariance matrix Σ is **diagonal**.

Intuition

- SUR improves efficiency only when:
  - errors across equations are **correlated**, and
  - regressors **differ across equations**.
- If Σ is diagonal, errors across equations are **uncorrelated**.
- This means knowing the residual in one equation tells you **nothing** about another — there is no information to borrow across equations.
- OLS already exploits all available information, so SUR adds nothing.

Summary without Formulas

OLS vs SUR: Two Ways to Estimate a System

- We often observe **multiple related regression equations** (a system).
- Example: demand for housing, food, and clothing.
- The key issue: **shocks affecting one equation may also affect others**.
- Two estimation approaches:

OLS (Ordinary Least Squares)

- Estimate **each equation separately**
- Treat equations as if they are independent
- Ignore possible **correlation of shocks across equations**

SUR (Seemingly Unrelated Regression)

- Estimate the **entire system jointly**
- Allow **correlation of shocks across equations**
- Use this information during estimation

Properties of OLS and SUR

- Under standard assumptions:
  - **OLS is unbiased and consistent**
  - **SUR is also unbiased and consistent**
- This means that with large samples **both estimators converge to the true parameters**.
- The main difference is **efficiency**.
- Compared with OLS, SUR can produce:
  - **smaller variance**
  - **smaller standard errors**
  - **more precise estimates**
- Important distinction:
  - **Robust standard errors fix inference**
  - **They do not change coefficient estimates**

When SUR Matters

- SUR can produce different coefficient estimates from OLS when it exploits the covariance of errors across equations.
- This can improve **predictive performance**, since prediction depends on the estimated coefficients.
- Robust standard errors **cannot improve prediction** because they do not change the estimator.
- SUR differs from OLS **only when two conditions hold simultaneously**:
  - **Errors across equations are correlated**
  - **Regressors differ across equations**
- If either condition fails:
  - Same regressors in all equations → **SUR = OLS**
  - No cross-equation correlation → **SUR = OLS**
- **Key intuition**
  - SUR helps only when **correlated shocks provide additional information across equations**.

Required Reading

- Hansen, Econometrics - Chapter 11.1 - 11.8