# Econometrics 2

## Heteroskedasticity

Lasha Chochua

2026

# Heteroskedasticity and Homoskedasticity

# What Are Heteroskedasticity and Homoskedasticity?

- The **error term** $u_i$ is called:
  - **Homoskedastic** if $\text{Var}(u_i \mid X_i)$ is **constant** for all $i = 1, \ldots, n$
  - That is, the variance **does not depend** on the value of $X_i$
  - **Heteroskedastic** if the variance of $u_i$ **changes with** $X_i$
- Homoskedasticity is one of the **Gauss–Markov assumptions**
- Violation of this assumption leads to **inefficient OLS estimates**

# Addressing Heteroskedasticity – GLS and FGLS

- **Generalized Least Squares (GLS)**:
  - Provides **efficient estimates** in the presence of known heteroskedasticity
  - Transforms the model to satisfy classical assumptions:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \quad \mathrm{Var}(\mathbf{u} \mid \mathbf{X}) = \sigma^2$$

- GLS transformation uses $\Omega^{-1/2}$ to "whiten" the errors
- **Feasible GLS (FGLS)**:
  - Used when $\Omega$ is **unknown**
  - Step 1: Estimate $\hat{\Omega}$ from OLS residuals
  - Step 2: Apply GLS using $\hat{\Omega}$ instead
  - Asymptotically efficient under correct specification

# Generalized Least Squares

# Generalized Least Squares (GLS)

- Consider the linear regression model in matrix form:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e} \tag{1}$$

- Assume a generalized setting where the errors may be correlated or heteroskedastic:

$$E[\mathbf{e} \mid \mathbf{X}] = 0, \quad \text{var}[\mathbf{e} \mid \mathbf{X}] = \Sigma\sigma^2 \tag{2}$$

- The matrix $\Sigma$ is $n \times n$ (possibly a function of $\mathbf{X}$), symmetric, and positive semi-definite

# Properties of the OLS Estimator

**Assumption**

The error term $\mathbf{e}$ satisfies $E[\mathbf{e} \mid \mathbf{X}] = 0$ and $\text{var}[\mathbf{e} \mid \mathbf{X}] = \Sigma\sigma^2$.

- Expectation of $\hat{\beta}$:

$$E[\hat{\beta} \mid \mathbf{X}] = \beta \tag{3}$$

- Variance of $\hat{\beta}$:

$$\text{var}[\hat{\beta} \mid \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\Sigma\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \tag{4}$$

# Derivation of the GLS Estimator

- Multiply the regression equation (1) by $\Sigma^{-1/2}$:

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\beta + \tilde{\mathbf{e}} \tag{5}$$

where $\tilde{\mathbf{Y}} = \Sigma^{-1/2}\mathbf{Y}$, $\tilde{\mathbf{X}} = \Sigma^{-1/2}\mathbf{X}$, and $\tilde{\mathbf{e}} = \Sigma^{-1/2}\mathbf{e}$

- The GLS estimator is:

$$\tilde{\beta}_{gls} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}$$

- Let's derive the result!

# Derivation of the GLS Estimator (Cont.)

- Expanding the transformation:

$$(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}} = \left((\Sigma^{-1/2}\mathbf{X})'(\Sigma^{-1/2}\mathbf{X})\right)^{-1}(\Sigma^{-1/2}\mathbf{X})'(\Sigma^{-1/2}\mathbf{Y})$$

- Simplifying:

$$\tilde{\beta}_{gls} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y} \tag{6}$$

# Properties of the GLS Estimator

- Expectation:

$$E[\tilde{\beta}_{gls} \mid \mathbf{X}] = \beta \tag{7}$$

- Variance:

$$\text{var}[\tilde{\beta}_{gls} \mid \mathbf{X}] = \sigma^2(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \tag{8}$$

Important

**Aitken's Theorem (1935)**
The GLS estimator achieves the lowest possible efficiency bound among all linear unbiased estimators.

# GLS with Known Conditional Variances

- If $\Sigma = D = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$, then:

$$\tilde{\beta}_{gls} = (\mathbf{X}'D^{-1}\mathbf{X})^{-1}\mathbf{X}'D^{-1}\mathbf{Y} \tag{9}$$

- Expanding:

$$\tilde{\beta}_{gls} = \left(\sum_{i=1}^{n} \sigma_i^{-2} X_i X_i'\right)^{-1} \left(\sum_{i=1}^{n} \sigma_i^{-2} X_i Y_i\right) \tag{10}$$

- If $\sigma_i^2 = \sigma^2$ for all $i$, this reduces to OLS

# Feasibility of the GLS Estimator

- The assumption $\Sigma > 0$ reduces to $\sigma_i^2 > 0$ for $i = 1, \dots, n$
- In most settings, the matrix $\Sigma$ is unknown, making the GLS estimator (6) infeasible
- The structure of the GLS estimator motivates **feasible GLS**, where $\Sigma$ is replaced with a suitable estimator

# Estimation of Error Variance

# Estimation of Error Variance

- The error variance $\sigma^2 = E[e^2]$ represents the variation in the **unexplained** part of the regression
- Define the **bias-corrected estimator**:

$$s^2 = \frac{1}{n-k} \sum_{i=1}^{n} \hat{e}_i^2 \tag{11}$$

# Covariance Matrix Estimation Under Homoskedasticity

# Covariance Matrix Estimation Under Homoskedasticity

- For inference, we need an estimator of the covariance matrix $V_{\hat{\beta}}$ of the least squares estimator
- Under homoskedasticity, the covariance matrix has the form:

$$V_{\hat{\beta}}^{0} = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

# Estimating $V_{\hat{\beta}}$

- The covariance matrix is known **up to the scale** $\sigma^2$
- We have discussed one possible estimator of $\sigma^2$
- The most commonly used choice is $s^2$ from (11), leading to the classic covariance matrix estimator:

$$\hat{V}^0_{\hat{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} s^2 \tag{12}$$

# Unbiasedness of $\widehat{V}^0_{\hat{\beta}}$

- Since $s^2$ is **conditionally unbiased** for $\sigma^2$, we compute:

$$E[\widehat{V}^0_{\hat{\beta}} \mid \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} E[s^2 \mid \mathbf{X}]$$

- Substituting $E[s^2 \mid \mathbf{X}] = \sigma^2$:

$$(\mathbf{X}'\mathbf{X})^{-1}\sigma^2 = V_{\hat{\beta}}$$

# Practical Importance

Important

**Key Point**

The covariance estimator $\hat{V}^0_{\hat{\beta}}$ from (12) is widely used in applied econometrics. It is the **default** in Stata, R, and other statistical packages for OLS inference.

# Covariance Matrix Estimation Under Heteroskedasticity

# Covariance Matrix Estimation Under Heteroskedasticity

- The **classic covariance matrix estimator** (12) can be highly biased if **homoskedasticity fails**
- We now construct **heteroskedasticity-robust** covariance matrix estimators
- The general form for the covariance matrix is:

$$V_{\hat{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'D\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \tag{13}$$

# Definition of $D$

- The matrix $D$ depends on the **unknown** variances of the error terms:

$$D = \text{diag}(\sigma_1^2, \dots, \sigma_n^2) = E[\mathbf{ee}' \mid \mathbf{X}] = E[\tilde{D} \mid \mathbf{X}] \tag{14}$$

- Since $\tilde{D} = \text{diag}(e_1^2, \dots, e_n^2)$ is a **conditionally unbiased estimator** for $D$, we use it for estimation

# Estimating $V_{\hat{\beta}}$ – The Ideal Estimator

- If squared errors $e_i^2$ were observable, we could estimate $V_{\hat{\beta}}$ as:

$$\widehat{V}_{\hat{\beta}}^{\text{ideal}} = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{n} X_i X_i' e_i^2\right)(\mathbf{X}'\mathbf{X})^{-1} \tag{15}$$

# Unbiasedness of $\widehat{V}_{\hat{\beta}}^{\text{ideal}}$

- Taking expectation of (15):

$$E[\widehat{V}_{\hat{\beta}}^{\text{ideal}} \mid \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} E\left[\sum_{i=1}^{n} X_i X_i' e_i^2 \mid \mathbf{X}\right] (\mathbf{X}'\mathbf{X})^{-1}$$

- Using $E[e_i^2 \mid \mathbf{X}] = \sigma_i^2$:

$$= (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{i=1}^{n} X_i X_i' \sigma_i^2\right) (\mathbf{X}'\mathbf{X})^{-1}$$

- This simplifies to:

$$(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'D\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} = V_{\hat{\beta}}$$

# Unbiasedness of $\widehat{V}_{\hat{\beta}}^{\text{ideal}}$ (Cont.)

Important

**Key Point**

The estimator $\widehat{V}_{\hat{\beta}}^{\text{ideal}}$ from (15) is **unbiased** for $V_{\hat{\beta}}$.

- This estimator **removes bias** present in homoskedasticity-based methods
- In practice, we use **heteroskedasticity-robust estimators**, since $e_i^2$ is not directly observable

# Heteroskedasticity-Consistent (HC) Estimators

# Heteroskedasticity-Consistent Covariance Matrix Estimators

- The **ideal estimator** (15) is not feasible because the errors $e_i^2$ are unobserved
- Instead, we replace $e_i^2$ with the squared residuals $\hat{e}_i^2$
- This substitution leads to the **HC0 estimator**:

$$\widehat{V}_{\hat{\beta}}^{\mathsf{HC0}} = (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_{i=1}^{n} X_i X_i' \hat{e}_i^2 \right) (\mathbf{X}'\mathbf{X})^{-1} \tag{16}$$

# Interpretation of HC0

**Definition: HC0**

The label "HC" refers to "heteroskedasticity-consistent." HC0 is the baseline heteroskedasticity-consistent covariance matrix estimator.

- The HC0 estimator accounts for **heteroskedasticity** but does not correct for small-sample bias

# Bias Correction – HC1 Estimator

- $\hat{e}_i^2$ is **biased downward**
- To correct for this, we scale by $\frac{n}{n-k}$
- Making the same adjustment to (16), we obtain the **HC1 estimator**:

$$\hat{V}_{\hat{\beta}}^{\text{HC1}} = \left(\frac{n}{n-k}\right)(\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{i=1}^{n} X_i X_i' \hat{e}_i^2\right)(\mathbf{X}'\mathbf{X})^{-1} \tag{17}$$

- $\hat{V}_{\hat{\beta}}^{\text{HC0}}$ is the **baseline** heteroskedasticity-consistent covariance estimator
- $\hat{V}_{\hat{\beta}}^{\text{HC1}}$ adjusts for **small-sample bias** by scaling residuals
- These estimators are widely used in **econometrics** and **applied regression analysis**

# Standard Errors

# Standard Errors

**Definition: Standard Error**

A **standard error** $s(\hat{\beta})$ for a real-valued estimator $\hat{\beta}$ is an estimator of the standard deviation of the distribution of $\hat{\beta}$.

- A variance estimator such as $\hat{V}_{\hat{\beta}}$ estimates the **variance** of the distribution of $\hat{\beta}$
- A more interpretable measure of spread is the **square root** of the variance – the **standard deviation**
- Estimates of standard deviation for parameter estimators are called **standard errors**

# Computing Standard Errors

- When $\beta$ is a **vector** with estimator $\hat{\beta}$ and covariance matrix estimator $\hat{V}_{\hat{\beta}}$:
  - **Standard errors** for individual elements of $\hat{\beta}$ are the **square roots** of the diagonal elements of $\hat{V}_{\hat{\beta}}$:

$$s(\hat{\beta}_j) = \sqrt{\hat{V}_{\hat{\beta}_j}} = \sqrt{[\hat{V}_{\hat{\beta}}]_{jj}}$$

# Standard Errors Under Homoskedasticity

- When the **classical covariance matrix estimator** (12) is used, standard errors simplify to:

$$s(\hat{\beta}_j) = s\sqrt{[(\mathbf{X}'\mathbf{X})^{-1}]_{jj}} \tag{18}$$

- This is the **default** approach in standard OLS inference

# Required Reading

- Wooldridge (2022), *Introductory Econometrics*, Chapter 8, Esp. Section 8.3 – Testing for Heteroskedasticity.