ECON 452* -- NOTE 11

Heteroskedasticity-Robust Inference in Linear Regression Models Estimated by OLS

This note does three things:

1. identifies the ***nature and consequences of heteroskedasticity*** for estimation and inference based on OLS estimation;
2. reviews ***alternative remedies*** available for estimation and inference in heteroskedastic errors models;
3. outlines how to perform ***heteroskedasticity-robust inference*** in linear regression models estimated by OLS.

1. Nature of Heteroskedasticity

Assumption A3.1 of the classical linear regression model specifies that the error terms $\{u_i : i = 1, \ldots, N\}$ are homoskedastic, meaning that they have the same variance for all observations.

$$\text{Var}\left(u_i \mid x_i^T\right) = E\left(u_i^2 \mid x_i^T\right) = E\left(u_i^2 \mid 1, X_{i1}, X_{i2}, \ldots, X_{ik}\right) = \sigma^2 > 0 \quad \forall\ i$$

where σ^2 is a **finite positive (unknown)** ***constant***.

Heteroskedasticity is a violation of assumption A3.1 of the classical linear regression model (or CLRM). Violation of assumption A3.1 means in general that

$$\text{Var}\left(u_i \mid x_i^T\right) = E\left(u_i^2 \mid x_i^T\right) = E\left(u_i^2 \mid 1, X_{i1}, X_{i2}, \ldots, X_{ik}\right) = \sigma_i^2 > 0 \quad \forall\ i \qquad (1)$$

The "i" subscript on σ_i^2 indicates that the value of the error variance is no longer constant, but instead is now a variable that assumes different (finite positive) values for different observations, i.e., for different sets of regressor values.

The Error Variance-Covariance Matrix

- Under the assumption of **spherical errors (homoskedastic and nonautoregressive errors)**, the error variance-covariance matrix is a constant scalar diagonal matrix with the constant error variance σ^2 along the principal diagonal:

$$V(u) = E\left(uu^T\right) = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 I_N \quad (2)$$

  where I_N is an **N×N identity matrix** with 1s along the principal diagonal and 0s in all the off-diagonal cells.

- Under the assumption of **heteroskedastic and nonautoregressive errors**, the error variance-covariance matrix is an N×N diagonal matrix with σ_i^2 as the i-th diagonal element:

$$V = V(u) = E\left(uu^T\right) = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix} = \text{diag}\left(\sigma_1^2 \quad \sigma_2^2 \quad \sigma_3^2 \quad \cdots \quad \sigma_N^2\right).$$

… (3)

2. Consequences of Heteroskedasticity for OLS

- Under the assumptions of the CLRM, which include the homoskedasticity (constant error variance) assumption, the **OLS coefficient estimators $\hat{\beta}_j$ are BLUE** – meaning they are the Best Linear Unbiased Estimators of the regression coefficients β_j (j = 0, 1, …, k).

  In other words, the OLS coefficient estimators $\hat{\beta}_j$ exhibit the ***small-sample* properties of *unbiasedness* and *efficiency***: they have ***minimum variance*** in the class of all *linear unbiased* estimators of the regression coefficients.

  The OLS coefficient estimators $\hat{\beta}_j$ also exhibit the ***large-sample* property of *consistency***.

- **What are the consequences of heteroskedastic errors** for the desirable statistical properties of the OLS coefficient estimators?

  Heteroskedastic errors have **two sets of consequences for OLS estimation**:

  1. **consequences for the OLS coefficient estimators** $\hat{\beta}_j$, j = 0, 1, …, k;

  2. **consequences for statistical inference based on OLS estimation**.

2.1 Consequences for OLS Coefficient Estimation

In the presence of heteroskedastic errors, the OLS coefficient estimators retain some of their desirable statistical properties.

1. The **OLS coefficient estimators $\hat{\beta}_j$ (j = 0, 1, …, k) are still *unbiased*** (a small sample property):

- In scalar terms: $E(\hat{\beta}_j) = \beta_j$ for all j = 0, 1, …, k.
- In matrix terms: $E(\hat{\beta}) = \beta$.

2. The **OLS coefficient estimators $\hat{\beta}_j$ (j = 0, 1, …, k) are still *consistent*** (a large sample property):

- In scalar terms: $\text{plim}(\hat{\beta}_j) = \beta_j$ for all j = 0, 1, …, k.
- In matrix terms: $\text{plim}(\hat{\beta}) = \beta$.

However, in the presence of heteroskedastic errors, the OLS coefficient estimators lose one of their desirable statistical properties, namely the *efficiency* property.

3. The **OLS coefficient estimators $\hat{\beta}_j$ (j = 0, 1, …, k) are no longer efficient**, meaning they are no longer the minimum variance estimators in the class of all linear unbiased estimators of the regression coefficients, either in small samples or in large samples.

$\text{Var}\left(\hat{\beta}_j\right) \geq \text{Var}\left(\tilde{\beta}_j\right)$ (j = 0, 1, …, k), where $\hat{\beta}_j$ denotes the OLS estimator of β_j and $\tilde{\beta}_j$ denotes an alternative estimator of β_j that properly takes account of heteroskedasticity.

The **OLS coefficient estimators $\hat{\beta}_j$ (j = 0, 1, …, k) are *inefficient*** in finite samples of any given size.

The **OLS coefficient estimators $\hat{\beta}_j$ (j = 0, 1, …, k) are also *asymptotically inefficient*** in large samples.

2.2 Consequences for OLS Inference Procedures - Scalar Analysis

Consequence 1: The **OLS formulas for Var($\hat{\beta}_j$), j = 0, 1, …, k**, the variances of the OLS coefficient estimators $\hat{\beta}_j$, **are *incorrect*** in the presence of heteroskedastic errors.

Assumming Homoskedastic (and Nonautoregressive) Errors

- **The OLS formula for Var($\hat{\beta}_j$),** the variance of the OLS coefficient estimator $\hat{\beta}_j$, can be written as:

$$\text{Var}\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\text{RSS}_j} = \frac{\sigma^2}{\text{TSS}_j(1-R_j^2)} \qquad j = 0, 1, \ldots, k. \tag{4}$$

  where σ^2 is the error variance (incorrectly assumed to be constant), and the quantities RSS_j, TSS_j and R_j^2 in formula (4) are sample statistics from the auxiliary OLS regression of X_j on all the other regressors in the regression model, including the constant term.

- This **auxiliary OLS sample regression equation for regressor X_j** can be written for the i-th sample observation as:

$$X_{ij} = \hat{b}_{j0} + \hat{b}_{j1}X_{i1} + \hat{b}_{j2}X_{i2} + \cdots + \hat{b}_{j,j-1}X_{i,j-1} + \hat{b}_{j,j+1}X_{i,j+1} + \cdots + \hat{b}_{jk}X_{ik} + \hat{r}_{ij} \tag{5}$$

  where:

  $\hat{r}_{ij}$ = the OLS residual for observation i from auxiliary regression (5);

  $\text{RSS}_j = \sum_{i=1}^{N} \hat{r}_{ij}^2$ = the residual sum-of-squares from auxiliary regression (5);

  $\text{TSS}_j = \sum_{i=1}^{N} \left(X_{ij} - \bar{X}_j\right)^2$ = the total sum-of-squares of the regressand X_{ij} in auxiliary regression (5);

  $R_j^2 = 1 - \left(\text{RSS}_j / \text{TSS}_j\right)$ = the R-squared from auxiliary regression (5).

Assuming Heteroskedastic (and Nonautoregressive) Errors

- **In the presence of heteroskedastic errors, the *correct* formula for Var($\hat{\beta}_j$),** the variance of the OLS coefficient estimator $\hat{\beta}_j$, is:

$$\mathrm{Var}\left(\hat{\beta}_j\right) = \frac{\sum_{i=1}^{N} \hat{r}_{ij}^2 \sigma_i^2}{(\mathrm{RSS}_j)^2} = \frac{\sum_{i=1}^{N} \hat{r}_{ij}^2 \sigma_i^2}{\left(\sum_{i=1}^{N} \hat{r}_{ij}^2\right)^2} \qquad j = 0, 1, \ldots, k. \tag{6}$$

where

$\sigma_i^2 = \mathrm{Var}\left(u_i \mid \underline{x}_i^T\right)$ = the unknown conditional error variance for the i-th random error in regression model (1);

$\hat{r}_{ij}$ = the i-th OLS residual from auxiliary regression (5);

$\mathrm{RSS}_j = \sum_{i=1}^{N} \hat{r}_{ij}^2$ = the residual sum-of-squares from auxiliary regression (5).

Consequence 2: The **OLS estimator of the error variance**, denoted as $\hat{\sigma}^2$, **is *biased* and *inconsistent*** in the presence of heteroskedastic errors.

- The formula for the OLS error variance estimator $\hat{\sigma}^2$ is:

$$\hat{\sigma}^2 = \frac{\mathrm{RSS}}{N - K} = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N - K}. \tag{7}$$

- The bias of $\hat{\sigma}^2$ means that $E(\hat{\sigma}^2) \neq \sigma^2$.
- The inconsistency of $\hat{\sigma}^2$ means that $\mathrm{plim}(\hat{\sigma}^2) \neq \sigma^2$.

<u>Consequence 3</u>: Consequences 1 and 2 imply that **the OLS estimator of Var($\hat{\beta}_j$)**, the variance of the OLS coefficient estimator $\hat{\beta}_j$, **is *biased* and *inconsistent*** in the presence of heteroskedastic errors.

- Recall that the OLS estimator of Var($\hat{\beta}_j$) is obtained by substituting the OLS error variance estimator $\hat{\sigma}^2$ for σ^2 in the OLS formula for Var($\hat{\beta}_j$) given by equation (4):

$$\hat{\mathrm{Var}}\left(\hat{\beta}_j\right) = \frac{\hat{\sigma}^2}{\mathrm{RSS}_j} = \frac{\hat{\sigma}^2}{\mathrm{TSS}_j(1 - R_j^2)} \qquad j = 0, 1, \ldots, k. \qquad (8)$$

  - The bias of $\hat{\mathrm{Var}}(\hat{\beta}_j)$ means that $E\left(\hat{\mathrm{Var}}(\hat{\beta}_j)\right) \neq \mathrm{Var}(\hat{\beta}_j)$.
  - The inconsistency of $\hat{\mathrm{Var}}(\hat{\beta}_j)$ means that $\mathrm{plim}\left(\hat{\mathrm{Var}}(\hat{\beta}_j)\right) \neq \mathrm{Var}(\hat{\beta}_j)$.

- **<u>Intuition</u>:** There are two distinct reasons why $\hat{\mathrm{Var}}(\hat{\beta}_j)$, the OLS estimator of Var($\hat{\beta}_j$), is biased and inconsistent in the presence of heteroskedastic errors.

  <u>Reason 1</u>: **The OLS variance estimator $\hat{\mathrm{Var}}(\hat{\beta}_j)$ uses the wrong formula for Var($\hat{\beta}_j$).** In the presence of heteroskedastic errors,

$$\mathrm{Var}\left(\hat{\beta}_j\right) = \frac{\sum_{i=1}^{N} \hat{r}_{ij}^2 \sigma_i^2}{(\mathrm{RSS}_j)^2} \neq \frac{\sigma^2}{\mathrm{RSS}_j} \qquad j = 0, 1, \ldots, k.$$

  <u>Reason 2</u>: **The OLS variance estimator $\hat{\mathrm{Var}}(\hat{\beta}_j)$ uses a *biased* and *inconsistent* estimator of the error variance σ^2.**

$$E(\hat{\sigma}^2) \neq \sigma^2 \quad \Rightarrow \quad \hat{\sigma}^2 \text{ is a } \textit{biased} \text{ estimator of } \sigma^2$$
$$\mathrm{plim}\,(\hat{\sigma}^2) \neq \sigma^2 \quad \Rightarrow \quad \hat{\sigma}^2 \text{ is an } \textit{inconsistent} \text{ estimator of } \sigma^2$$

<u>RESULT</u>: *OLS statistical inference procedures* for hypothesis testing and interval estimation are *invalid* in the presence of heteroskedastic errors.

2.3 Consequences for OLS Inference Procedures - Matrix Analysis

Consequence 1: The **OLS formula for the variance-covariance matrix of the OLS coefficient estimator is *incorrect*** in the presence of heteroskedastic errors.

Assuming Homoskedastic (and Nonautoregressive) Errors

- The OLS formula for the **variance-covariance matrix of the OLS coefficient estimator** $\hat{\beta}$ is:

$$V(\hat{\beta}) = \sigma^2(X^TX)^{-1} = \text{a } K \times K \text{ symmetric positive definite matrix.} \qquad (9)$$

Remember that formula (9) for $V(\hat{\beta})$ is derived using the assumption that the error covariance matrix takes the form

$$V(u) = E(uu^T) = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 I_N$$

where I_N is an **N×N identity matrix** with 1s along the principal diagonal and 0s in all the off-diagonal cells.

Assuming Heteroskedastic (and Nonautoregressive) Errors

- The correct formula of the **variance-covariance matrix of the OLS coefficient estimator** $\hat{\beta}$ in heteroskedastic and nonautoregressive errors models is:

$$V(\hat{\beta}) = V_{\hat{\beta}} = (X^TX)^{-1}X^TVX(X^TX)^{-1} \qquad (10)$$

where V is a general heteroskedastic error covariance matrix with diagonal elements σ_i^2 i = 1, …, N and zeros in all off-diagonal cells.

The general error covariance matrix for heteroskedastic and nonautoregressive errors models is:

$$V = \text{diag}\left(\sigma_1^2 \quad \sigma_2^2 \quad \sigma_3^2 \quad \cdots \quad \sigma_N^2\right) = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix}.$$

Since V is an N×N diagonal matrix, its inverse V^{-1} takes the form:

$$V^{-1} = \text{diag}\left(\frac{1}{\sigma_1^2} \quad \frac{1}{\sigma_2^2} \quad \frac{1}{\sigma_3^2} \quad \cdots \quad \frac{1}{\sigma_N^2}\right) = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma_3^2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma_N^2} \end{bmatrix}.$$

<u>**Consequence 2**</u>**:** The **OLS estimator of the error variance**, denoted as $\hat{\sigma}^2$, **is *biased* and *inconsistent*** in the presence of heteroskedastic errors.

- Recall that the formula for $\hat{\sigma}^2$ is:

$$\hat{\sigma}^2 = \frac{RSS}{N-K} = \frac{\hat{u}^T\hat{u}}{N-K} = \frac{\sum_{i=1}^{N}\hat{u}_i^2}{N-K}. \tag{7}$$

  - The bias of $\hat{\sigma}^2$ means that $E(\hat{\sigma}^2) \neq \sigma^2$.
  - The inconsistency of $\hat{\sigma}^2$ means that $\text{plim}(\hat{\sigma}^2) \neq \sigma^2$.

<u>Consequence 3</u>: Consequences 1 and 2 imply that **the OLS *estimator* of $\mathbf{V(\hat{\beta})}$**, the variance-covariance matrix of the OLS coefficient estimator $\hat{\beta}$, **is *biased* and *inconsistent*** in the presence of heteroskedastic errors.

- The **OLS estimator of $\mathbf{V(\hat{\beta})}$**, the variance-covariance matrix of $\hat{\beta}$, is obtained by replacing the unknown scalar constant σ^2 in formula (9) with the estimator $\hat{\sigma}^2 = \hat{u}^T\hat{u} \big/ N - K = \sum_{i=1}^{N} \hat{u}_i^2 \Big/ N - K$:

$$\hat{V}_{OLS}\left(\hat{\beta}\right) = \hat{V}_{OLS} = \hat{\sigma}^2\left(X^TX\right)^{-1}. \tag{11}$$

- **<u>Intuition</u>:** There are two distinct reasons why $\hat{V}_{OLS}(\hat{\beta})$, the OLS estimator of $V(\hat{\beta})$, is biased and inconsistent in the presence of heteroskedastic errors.

***<u>Reason 1</u>:* The OLS variance estimator $\mathbf{\hat{V}_{OLS}(\hat{\beta})}$ uses the wrong formula for $\mathbf{V(\hat{\beta})}$.** In the presence of heteroskedastic errors,

$$V_{OLS}\left(\hat{\beta}\right) = V_{OLS} = \left(X^TX\right)^{-1}X^TVX\left(X^TX\right)^{-1} \neq \sigma^2\left(X^TX\right)^{-1}.$$

***<u>Reason 2</u>:* The OLS variance estimator $\mathbf{\hat{V}_{OLS}(\hat{\beta})}$ uses a *biased* and *inconsistent* estimator of the error variance $\boldsymbol{\sigma^2}$.**

$E(\hat{\sigma}^2) \neq \sigma^2 \quad \Rightarrow \quad \hat{\sigma}^2$ is a *biased* estimator of σ^2

$\text{plim}\,(\hat{\sigma}^2) \neq \sigma^2 \quad \Rightarrow \quad \hat{\sigma}^2$ is an *inconsistent* estimator of σ^2

Implication: $\hat{\sigma}^2\left(X^TX\right)^{-1}$ is a ***biased*** **and *inconsistent* estimator** of $\sigma^2\left(X^TX\right)^{-1}$.

<u>RESULT</u>: *OLS statistical inference procedures* for hypothesis testing and interval estimation are *invalid* in the presence of heteroskedastic errors.

3. Strategies for Dealing with Heteroskedasticity

There are **two basic strategies** for dealing with linear regression models in which the errors are suspected of being heteroskedastic.

<u>Strategy 1</u>: G*LS estimation that fully accounts for heteroskedasticity*

- Use an alternative estimator that is efficient relative to OLS and that properly accounts for the heteroskedasticity in both estimation and inference.

  Generalized Least Squares estimators, or GLS estimators, constitute such a class of alternative estimators.

- ***Advantages:*** GLS estimation has two advantages over OLS.

  1. GLS estimation procedures are efficient relative to OLS, at least in sufficiently large samples. In other words, GLS is **asymptotically efficient** relative to OLS.

  2. GLS estimation procedures account for heteroskedastic errors in estimating the coefficient variances and covariances, and therefore **yield valid inferences**.

- ***Disadvantage:*** The major drawback of GLS estimation is that it **requires prior knowledge of the form of the heteroskedasticity** on the part of the investigator.

  This requirement is often difficult or impossible to satisfy in practice because the range of forms that heteroskedastic errors can take is very large.

<u>Strategy 2</u>: ***OLS estimation with heteroskedasticity-robust inference procedures***

- Use the OLS estimator even though it is inefficient, but modify the conventional OLS coefficient variance estimators to make them consistent in the presence of heteroskedasticity of unknown form.

- ***Advantages:***

  1. Using heteroskedasticity-consistent estimators of the variances and covariances of the OLS coefficient estimates means that **inferences based on the OLS estimates will be valid**, at least in sufficiently large samples.

  2. The **investigator does not need to know the specific form the heteroskedasticity** in order to make valid statistical inferences – i.e., in order to perform valid hypothesis tests.

- ***Disadvantages:*** The major disadvantage of this strategy is that the **OLS coefficient estimator is *inefficient*** in the presence of heteroskedastic errors.

  However, the inefficiency of OLS relative to GLS can be at least partially mitigated by using sufficiently large samples of data.

4. OLS with Heteroskedasticity-Robust Inference

In this section we show how to obtain consistent estimators of the variances of OLS coefficient estimates in the presence of heteroskedasticity of unknown form. The analysis is presented first in scalar terms, then in more general matrix terms.

4.1 OLS with Heteroskedasticity-Robust Inference - Scalar Analysis

- In the presence of heteroskedastic and nonautoregressive errors, the correct formula for the variance of the OLS coefficient estimator $\hat{\beta}_j$ is:

$$\mathrm{Var}\left(\hat{\beta}_j\right) = \frac{\sum_{i=1}^{N} \hat{r}_{ij}^2 \sigma_i^2}{(\mathrm{RSS}_j)^2} = \frac{\sum_{i=1}^{N} \hat{r}_{ij}^2 \sigma_i^2}{\left(\sum_{i=1}^{N} \hat{r}_{ij}^2\right)^2} \qquad j = 0, 1, \ldots, k. \qquad (12)$$

where

$\sigma_i^2 = \mathrm{Var}\left(u_i \middle| \underline{x}_i^T\right)$ = the unknown conditional error variance for the i-th random error in regression model (1).

$\hat{r}_{ij}$ = the i-th OLS residual from auxiliary regression (5);

$\mathrm{RSS}_j = \sum_{i=1}^{N} \hat{r}_{ij}^2$ = the residual sum-of-squares from auxiliary regression (5);

and auxiliary regression (5) is the OLS regression of the regressor X_j on all the other k−1 regressors in the regression model, including the constant term:

$$X_{ij} = \hat{b}_{j0} + \hat{b}_{j1}X_{i1} + \hat{b}_{j2}X_{i2} + \cdots + \hat{b}_{j,j-1}X_{i,j-1} + \hat{b}_{j,j+1}X_{i,j+1} + \cdots + \hat{b}_{jk}X_{ik} + \hat{r}_{ij} \qquad (5)$$

Note: Formula (12) cannot be computed from sample data because the conditional error variances $\{\sigma_i^2 : i = 1, \ldots, N\}$ are unknown.

- **A Heteroskedasticity-Consistent Estimator of Var($\hat{\beta}_j$)**

White (1980) showed that a consistent estimator of Var($\hat{\beta}_j$) in the presence of heteroskedasticity of unknown form is

$$\hat{\text{Var}}_{HC}\left(\hat{\beta}_j\right) = \frac{\sum_{i=1}^{N} \hat{r}_{ij}^2 \hat{u}_i^2}{(RSS_j)^2} = \frac{\sum_{i=1}^{N} \hat{r}_{ij}^2 \hat{u}_i^2}{\left(\sum_{i=1}^{N} \hat{r}_{ij}^2\right)^2} \qquad j = 0, 1, \ldots, k. \qquad (13)$$

where:

$\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}$
= the i-th OLS residual from OLS estimation of regression equation (1);

$\hat{r}_{ij}$ = the i-th OLS residual from auxiliary regression (5);

$RSS_j = \sum_{i=1}^{N} \hat{r}_{ij}^2$ = the residual sum-of-squares from auxiliary regression (5).

Comments

- The key difference between Var($\hat{\beta}_j$) in (12) and $\hat{\text{Var}}_{HC}(\hat{\beta}_j)$ in (13) is that the ***unknown* error variances {σ_i^2: i = 1, …, N} in (12) have been replaced in (13) by the *squared* OLS residuals {$\hat{u}_i^2$: i = 1, …, N}** from OLS estimation of the regression equation $Y_i = x_i^T\beta + u_i$.

- A technical problem with formula (13) for the heteroskedasticity-consistent estimator of Var($\hat{\beta}_j$) is that it can yield estimates of Var($\hat{\beta}_j$) that, though consistent, are sometimes downward biased, or "too small", in small samples.

  The fix for this problem is to apply a ***degrees-of-freedom correction*** to $\hat{\text{Var}}_{HC}(\hat{\beta}_j)$ in (13).

- **Finite Sample Adjustment to $\hat{Var}_{HC}(\hat{\beta}_j)$**

It is common practice to apply a **degrees-of-freedom adjustment** to $\hat{Var}_{HC}(\hat{\beta}_j)$ in (13) to correct for its small-sample bias. The most common adjustment consists of **multiplying the formula in (13) by the ratio $N/N - K$**.

The **degrees-of-freedom adjusted heteroskedasticity-consistent estimator of $Var(\hat{\beta}_j)$** is therefore:

$$\hat{Var}_{HC1}\left(\hat{\beta}_j\right) = \frac{N}{N-K}\hat{Var}_{HC}\left(\hat{\beta}_j\right) = \frac{N}{N-K}\frac{\sum_{i=1}^{N}\hat{r}_{ij}^2\hat{u}_i^2}{(RSS_j)^2} \quad j = 0, 1, \ldots, k. \qquad (14)$$

Note: $\hat{Var}_{HC1}(\hat{\beta}_j) > \hat{Var}_{HC}(\hat{\beta}_j)$ because $\dfrac{N}{N-K} > 1$.

But the difference between $\hat{Var}_{HC1}(\hat{\beta}_j)$ and $\hat{Var}_{HC}(\hat{\beta}_j)$ becomes negligible in large samples because

$$\lim_{N\to\infty}\frac{N}{N-K} = 1.$$

4.2 OLS with Heteroskedasticity-Robust Inference - Matrix Analysis

- The correct formula of the **variance-covariance matrix of the OLS coefficient estimator** $\hat{\beta}$ in heteroskedastic and nonautoregressive errors models is:

$$V(\hat{\beta}) = V_{\hat{\beta}} = \left(X^T X\right)^{-1} X^T V X \left(X^T X\right)^{-1} \qquad (15)$$

where $V = V(u) = E(uu^T)$ is the diagonal error covariance matrix

$$V = \text{diag}\left(\sigma_1^2 \quad \sigma_2^2 \quad \sigma_3^2 \quad \cdots \quad \sigma_N^2\right) = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix}. \qquad (16)$$

- ***Note:*** The matrix $V_{\hat{\beta}} = V(\hat{\beta})$ in (15) cannot be computed from sample data because the conditional error variances $\{\sigma_i^2 : i = 1, \ldots, N\}$ are unknown.

- ***Special Case:*** The form of the matrix $V_{\hat{\beta}} = V(\hat{\beta})$ for homoskedastic and nonautoregressive errors models – the classical linear regression model – is a special case of the above matrix.

  For homoskedastic and nonautoregressive errors models, the error covariance matrix V takes the form

$$V = V(u) = E\left(uu^T\right) = \sigma^2 I_N = \text{ a constant scalar diagonal matrix.}$$

Substituting $V = \sigma^2 I_N$ into formula (15) for $V_{\hat\beta}$ yields:

$$
\begin{aligned}
V_{\hat\beta} &= (X^TX)^{-1}X^TVX(X^TX)^{-1} && \text{from formula (15)}\\
&= (X^TX)^{-1}X^T(\sigma^2 I_N)X(X^TX)^{-1} && \text{setting } V = \sigma^2 I_N\\
&= \sigma^2(X^TX)^{-1}X^TI_NX(X^TX)^{-1} && \text{since } \sigma^2 \text{ is a scalar constant}\\
&= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} && \text{since } X^TI_NX = X^TX\\
&= \sigma^2(X^TX)^{-1}I_N && \text{since } X^TX(X^TX)^{-1} = I_N\\
&= \sigma^2(X^TX)^{-1} && \text{since } (X^TX)^{-1}I_N = (X^TX)^{-1}.
\end{aligned}
$$

- **A Heteroskedasticity-Consistent Estimator of $V(\hat\beta)$**

- **White's (1980) heteroskedasticity-consistent estimator of the matrix** $V_{\hat\beta}$ = **$V(\hat\beta)$** for heteroskedastic and nonautoregressive errors models is obtained by simply replacing the unknown error variances { σ_i^2 : i = 1, …, N} in error covariance matrix (16) with the corresponding ***squared*** **OLS residuals** { $\hat u_i^2$: i = 1, …, N} from OLS estimation of the regression equation $Y_i = x_i^T\beta + u_i$.

That is, replace matrix $V = \text{diag}(\sigma_1^2 \quad \sigma_2^2 \quad \sigma_3^2 \quad \cdots \quad \sigma_N^2)$ with

$$
\hat V = \text{diag}(\hat u_1^2 \quad \hat u_2^2 \quad \hat u_3^2 \quad \cdots \quad \hat u_N^2) = \begin{bmatrix} \hat u_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \hat u_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \hat u_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \hat u_N^2 \end{bmatrix}. \qquad (17)
$$

Then substitute $\hat V$ in (17) for V in matrix formula (15):

$$
V(\hat\beta) = V_{\hat\beta} = (X^TX)^{-1}X^TVX(X^TX)^{-1} \qquad (15)
$$

The result of this substitution is:

$$\hat{V}_{HC} = \left(X^T X\right)^{-1} X^T \hat{V} X \left(X^T X\right)^{-1}. \tag{18}$$

Matrix equation (18) is the ***heteroskedasticity-consistent*** **estimator of the coefficient variance-covariance matrix $V_{\hat{\beta}} = V(\hat{\beta})$** for heteroskedastic and nonautoregressive errors models.

- **The Adjusted Heteroskedasticity-Consistent Estimator of $V(\hat{\beta})$**

It is common practice to apply a ***degrees-of-freedom correction*** to the matrix in formula (18) to mitigate the small-sample downward bias of the HC covariance matrix estimator $\hat{V}_{HC}$.

The most widely used adjustment consists of multiplying the matrix estimator $\hat{V}_{HC}$ in (18) by the ratio $N/N-K$.

The ***degrees-of-freedom adjusted*** **heteroskedasticity-consistent estimator of $Var(\hat{\beta})$** is therefore:

$$\hat{V}_{HC1} = \frac{N}{N-K}\hat{V}_{HC} = \frac{N}{N-K}\left(X^T X\right)^{-1} X^T \hat{V} X \left(X^T X\right)^{-1}. \tag{19}$$

- **Computation of the HC Covariance Matrix Estimators $\hat{V}_{HC}$ and $\hat{V}_{HC1}$**

Tedious matrix manipulations would be required to calculate from scratch the value of $\hat{V}_{HC}$ in (18) or $\hat{V}_{HC1}$ in (19) for any OLS sample regression equation.

Fortunately, modern econometric software makes such laborious computations unnecessary. Options on OLS estimation commands usually make it very simple to compute heteroskedasticity-consistent estimates of the variances and covariances of OLS coefficient estimates.

- **Computing the Adjusted HC Covariance Matrix Estimator $\hat{V}_{HC1}$ in *Stata***

Stata incorporates a ***robust*** option on the ***regress*** command to compute the adjusted HC covariance matrix estimator $\hat{V}_{HC1}$ in (19).

For example, to estimate by OLS the regression equation

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + u_i$$

and compute the adjusted HC coefficient covariance estimator $\hat{V}_{HC1}$ in (19), simply enter the following ***regress*** command with the ***robust*** option:

```
regress y x1 x2 x3 x4, robust
matrix VHC1 = e(V)
```

- The ***regress*** command computes all coefficient standard errors, t-ratios and confidence intervals using the covariance estimator $\hat{V}_{HC1}$ in (19).

- The ***matrix*** command saves $\hat{V}_{HC1}$ in the matrix VHC1, which in this case is a 5×5 symmetric positive definite matrix.

5. Heteroskedasticity-Robust Hypothesis Tests with OLS

❑ General Setup for Tests of Linear Coefficient Restrictions

1. All t-tests and F-tests of linear coefficient restrictions on the regression coefficient vector β in linear regression equations of the form $Y_i = x_i^T\beta + u_i$ can be formulated in general terms as tests of the following null and alternative hypotheses:

Null hypothesis H_0: $R\beta = r$ (20)

Alternative hypothesis H_1: $R\beta \neq r$

where:

R = a q×K matrix of specified constants;

β = the K×1 regression coefficient vector;

r = a q×1 vector of specified constants.

2. **Wald F-statistic:** In *Note 10*, we introduced the following **general Wald F-statistic** for testing the q linear coefficient restrictions $R\beta = r$:

$$F_{WALD} = \frac{1}{q}W = \frac{\left(R\hat{\beta} - r\right)^T \left(R\hat{V}_{\hat{\beta}} R^T\right)^{-1}\left(R\hat{\beta} - r\right)}{q} \sim F[q, N-K] \text{ under } H_0 \qquad (21)$$

where:

$W = \left(R\hat{\beta} - r\right)^T \left(R\hat{V}_{\hat{\beta}} R^T\right)^{-1}\left(R\hat{\beta} - r\right) \sim \chi^2[q]$ = the general Wald statistic;

$\hat{V}_{\hat{\beta}}$ = a *consistent* estimator of OLS coefficient covariance matrix $V_{\hat{\beta}}$;

$F[q, N-K]$ = the F-distribution with q numerator and N–K denominator degrees of freedom.

3. The **OLS Wald F-statistic** is obtained by using the OLS coefficient covariance matrix estimator in place of $\hat{V}_{\hat{\beta}}$ in (21):

$$F_W = \frac{1}{q}W_{OLS} = \frac{\left(R\hat{\beta} - r\right)^T \left(R\hat{V}_{OLS} R^T\right)^{-1}\left(R\hat{\beta} - r\right)}{q} \sim F[q, N-K] \text{ under } H_0 \qquad (22)$$

where

$W_{OLS} = \left(R\hat{\beta} - r\right)^T \left(R\hat{V}_{OLS} R^T\right)^{-1}\left(R\hat{\beta} - r\right) \overset{a}{\sim} \chi^2[q]$ = the OLS Wald statistic

$\hat{\beta} = \left(X^T X\right)^{-1} X^T y$ = the unrestricted OLS estimator of β;

$\hat{V}_{OLS} = \hat{\sigma}^2\left(X^T X\right)^{-1}$ = the OLS estimator of $V_{\hat{\beta}}$, the covariance matrix of the unrestricted OLS estimator $\hat{\beta}$ of β,

$\hat{\sigma}^2 = \dfrac{RSS_1}{N-K} = \dfrac{\hat{u}^T\hat{u}}{N-K} = \dfrac{\sum_{i=1}^{N}\hat{u}_i^2}{N-K}$ = the unrestricted OLS estimator of σ^2.

4. Computation of F_W:

- The Wald F-statistic (22) can be computed using only the ***unrestricted* OLS coefficient estimates** $\hat{\beta}$ and a ***consistent* estimate** $\hat{\mathbf{V}}_{\hat{\beta}} = \hat{\mathbf{V}}(\hat{\beta})$ of the variance-covariance matrix of $\hat{\beta}$.

- Both the coefficient estimator $\hat{\beta}$ and the variance-covariance matrix estimator $\hat{V}_{\hat{\beta}} = \hat{V}(\hat{\beta})$ must at least be *consistent*.

- In heteroskedastic and nonautoregressive errors models that conform to all the other assumptions of the classical linear regression model:
  - the **OLS coefficient estimator** $\hat{\beta}$ is both ***unbiased* and *consistent***;
  - but the **OLS estimator** $\hat{\mathbf{V}}_{\mathbf{OLS}}$ of the variance-covariance matrix for $\hat{\beta}$ given in equation (11) is ***biased* and *inconsistent***.

❑ **Heteroskedasticity-Robust Tests of Linear Coefficient Restrictions**

- **General Idea:** Use either of the heteroskedasticity-consistent, or heteroskedasticity-robust, estimators of $V_{\hat{\beta}}$ in computing the Wald and Wald F-statistics.

- **Heteroskedasticity-robust Wald tests** can be performed using either of the following Wald statistics:

$$W_{HC} = \left(R\hat{\beta} - r\right)^T \left(R\,\hat{V}_{HC}\,R^T\right)^{-1}\left(R\hat{\beta} - r\right) \overset{a}{\sim} \chi^2[q] \text{ under } H_0 \qquad (23)$$

$$W_{HC1} = \left(R\hat{\beta} - r\right)^T \left(R\,\hat{V}_{HC1}R^T\right)^{-1}\left(R\hat{\beta} - r\right) \overset{a}{\sim} \chi^2[q] \text{ under } H_0 \qquad (24)$$

But heteroskedasticity-robust Wald tests are *large sample tests* only. In linear regression models, we normally use heteroskedasticity-robust F tests rather than heteroskedasticity-robust Wald tests.

- **Heteroskedasticity-robust F tests** can be performed using either of the following **heteroskedasticity-robust Wald F-statistics**:

$$F_{HC} = \frac{1}{q} W_{HC} = \frac{\left(R\hat{\beta} - r\right)^T \left(R \hat{V}_{HC} R^T\right)^{-1} \left(R\hat{\beta} - r\right)}{q} \overset{a}{\sim} F[q, N - K] \text{ under } H_0 \quad (25)$$

$$F_{HC1} = \frac{1}{q} W_{HC1} = \frac{\left(R\hat{\beta} - r\right)^T \left(R \hat{V}_{HC1} R^T\right)^{-1} \left(R\hat{\beta} - r\right)}{q} \overset{a}{\sim} F[q, N - K] \text{ under } H_0 \quad (26)$$

❑ ***Summary.*** We have **three alternative estimators** of **the coefficient covariance matrix V($\hat{\beta}$)**:

1. The **OLS estimator** is ***biased*** **and** ***inconsistent*** in the presence of heteroskedastic errors.

$$\hat{V}_{OLS} = \hat{\sigma}^2 \left(X^T X\right)^{-1} \quad (11)$$

where

$$\hat{\sigma}^2 = \frac{\hat{u}^T \hat{u}}{N - K} = \frac{RSS}{N - K} = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N - K} = \text{ the OLS estimator of } \sigma^2.$$

$$\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik} = \text{i-th OLS residual, } i = 1, \ldots, N.$$

2. The ***unadjusted*** **HC estimator** is ***biased*** **but** ***consistent*** in the presence of heteroskedastic errors.

$$\hat{V}_{HC} = \left(X^T X\right)^{-1} X^T \hat{V} X \left(X^T X\right)^{-1} \tag{18}$$

where:

$$\hat{V} = \text{diag}\left(\hat{u}_1^2 \quad \hat{u}_2^2 \quad \hat{u}_3^2 \quad \cdots \quad \hat{u}_N^2\right) = \begin{bmatrix} \hat{u}_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \hat{u}_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \hat{u}_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \hat{u}_N^2 \end{bmatrix}. \tag{17}$$

$\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}$ = i-th OLS residual, i = 1, …, N.

3. The ***adjusted*** **HC estimator** is ***less biased*** **and** ***consistent*** in the presence of heteroskedastic errors.

$$\hat{V}_{HC1} = \frac{N}{N-K}\hat{V}_{HC} = \frac{N}{N-K}\left(X^T X\right)^{-1} X^T \hat{V} X \left(X^T X\right)^{-1} \tag{19}$$

- ***Implication:*** For testing linear coefficient restrictions in linear regression models with heteroskedastic and nonautoregressive errors, use **Wald F-statistics** that employ either the ***unadjusted*** **HC estimator** $\hat{V}_{HC}$ or the ***adjusted*** **HC estimator** $\hat{V}_{HC1}$ of the coefficient covariance matrix $V(\hat{\beta})$.