ECON 452* -- The Skinny on Heteroskedasticity-Robust Inference (Note 11)

Heteroskedasticity-Robust Inference: The Bare Bones

- **General Wald F-Statistic:**

$$F_{WALD} = \frac{1}{q} W = \frac{\left(R\hat{\beta} - r\right)^T \left(R \hat{V}_{\hat{\beta}} R^T\right)^{-1} \left(R\hat{\beta} - r\right)}{q} \qquad (1)$$

where:

$\hat{\beta}$ = a ***consistent unrestricted*** **estimator of β**, such as the OLS estimator;

$\hat{V}_{\hat{\beta}}$ = a ***consistent*** **estimator of** $\mathbf{V}_{\hat{\beta}}$.

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

- **OLS Wald F-Statistic:** set $\hat{V}_{\hat{\beta}} = \hat{V}_{OLS} = \hat{\sigma}^2_{OLS}\left(X^TX\right)^{-1}$

$$F_W = \frac{1}{q}W_{OLS} = \frac{\left(R\hat{\beta}-r\right)^T\left(R\hat{V}_{OLS}R^T\right)^{-1}\left(R\hat{\beta}-r\right)}{q} \sim F[q, N-K] \text{ under } H_0 \qquad (2)$$

where

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

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

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

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

Heteroskedastic Errors: What do they do?

Answer: They change the form of the error covariance matrix, and hence the formula for the covariance matrix of the unrestricted OLS coefficient estimator $\hat{\beta}_{OLS}$.

Assuming Homoskedastic (and Nonautoregressive) Errors – Assumption A3

- The **error covariance matrix V** takes the form:

$$V = V(u|X) = \underset{(N\times N)}{\begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}} = \sigma^2 \underset{(N\times N)}{\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}} = \sigma^2 I_N$$

- The **covariance matrix of the unrestricted OLS coefficient estimator** $\hat{\beta}_{OLS}$ takes the form:

$$V_{\hat{\beta}} = V\left(\hat{\beta}_{OLS} \,|\, X\right) = \sigma^2\left(X^T X\right)^{-1}$$

- The **OLS estimator of the covariance matrix of** $\hat{\beta}_{OLS}$ is an unbiased and consistent estimator of $V_{\hat{\beta}}$:

$$\hat{V}_{OLS} = \hat{\sigma}^2_{OLS}\left(X^T X\right)^{-1}$$ is an ***unbiased*** **and** ***consistent*** estimator of $V_{\hat{\beta}}$

where

$$\hat{\sigma}^2_{OLS} = \frac{RSS_1}{N-K} = \frac{\hat{u}^T\hat{u}}{N-K} = \frac{\sum_{i=1}^{N}\hat{u}_i^2}{N-K}$$

= the unrestricted OLS estimator of σ^2

= an ***unbiased*** **and** ***consistent*** **estimator of** $\boldsymbol{\sigma^2}$ if the equation random errors u_i are **homoskedastic**

Assuming Heteroskedastic (and Nonautoregressive) Errors

- The **error covariance matrix V** takes the form:

$$V = V(u \mid X) = \text{diag}\left(\sigma_1^2 \quad \sigma_2^2 \quad \sigma_3^2 \quad \cdots \quad \sigma_N^2\right) = \underset{(N\times N)}{\begin{bmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix}}$$

- The **covariance matrix of the unrestricted OLS coefficient estimator** $\hat{\beta}_{OLS}$ takes the form:

$$V_{\hat{\beta}} = V\left(\hat{\beta}_{OLS} \mid X\right) = \left(X^T X\right)^{-1} X^T V X \left(X^T X\right)^{-1} \neq \sigma^2 \left(X^T X\right)^{-1}$$

Note: When the error terms u_i are heteroskedastic (have non-constant variances), the **covariance matrix of the unrestricted OLS coefficient estimator** $\hat{\beta}$ ***does not equal*** $\sigma^2\left(X^T X\right)^{-1}$; i.e., $V_{\hat{\beta}} \neq \sigma^2\left(X^T X\right)^{-1}$

Consequences of Heteroskedastic Errors

- **Consequences of heteroskedastic errors for *statistical inference*** based on OLS estimators of β and $V_{\hat{\beta}}$:

1. The OLS estimator of $V_{\hat{\beta}}$, $\hat{V}_{OLS} = \hat{\sigma}^2_{OLS}\left(X^T X\right)^{-1}$, is ***biased* and *inconsistent***.

2. **t-tests** and **F-tests** based on $\hat{V}_{OLS} = \hat{\sigma}^2_{OLS}\left(X^T X\right)^{-1}$ are ***invalid***.

- **Consequences of heteroskedastic errors** for the **statistical properties of the *unrestricted* OLS estimator** $\hat{\beta}_{OLS}$ of the regression coefficient vector β:

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

$E(\hat{\beta}_{OLS}) = \beta$.

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

$\text{plim}(\hat{\beta}_{OLS}) = \beta$.

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.

What We Need for Valid Statistical Inference Based on the OLS Coefficient Estimator $\hat{\beta}_{OLS}$

- For **statistical inference based on** $\hat{\beta}_{OLS}$, we need a ***consistent*** **estimator** of the **covariance matrix of** $\hat{\beta}_{OLS}$, which in the presence of heteroskedastic errors takes the form:

$$V_{\hat{\beta}} = V\left(\hat{\beta}_{OLS} \mid X\right) = \left(X^T X\right)^{-1} X^T V X \left(X^T X\right)^{-1}$$

- A **heteroskedasticity-consistent (or heteroskedasticity-robust) estimator of the covariance matrix of** $\hat{\beta}_{OLS}$ is White's estimator of $V_{\hat{\beta}}$:

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

where

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

$\hat{u}_i^2 = (Y_i - x_i^T \hat{\beta})^2 =$ the **squared *unrestricted* OLS residuals** for i = 1, …, N

Problem: In small samples, $\hat{V}_{HC}$ is a ***downward-biased*** **estimator** of the covariance matrix $V_{\hat{\beta}}$ of $\hat{\beta}_{OLS}$.

- **An Adjusted Heteroskedasticity-Consistent Estimator of $\mathbf{V(\hat{\beta}_{OLS})}$**

  To mitigate the small-sample downward bias of the HC covariance matrix estimator $\hat{V}_{HC}$, it is common practice to apply a ***degrees-of-freedom correction*** to the matrix formula for $\hat{V}_{HC}$.

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

  The ***degrees-of-freedom adjusted*** **heteroskedasticity-consistent estimator of $\mathbf{V(\hat{\beta}_{OLS})}$** 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 $\mathbf{\hat{V}_{HC}}$ and $\mathbf{\hat{V}_{HC1}}$**

  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 $\hat{V}_{HC1}$, simply enter the following ***regress*** command with the ***robust*** option:

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

- The ***regress*** command computes all coefficient standard errors, t-ratios and confidence intervals using the ***adjusted* HC covariance estimator** $\hat{V}_{HC1}$.
- The ***matrix*** command saves $\hat{V}_{HC1}$ in the matrix VHC1, which in this case is a 5×5 symmetric positive definite matrix.
- The ***matrix list*** command displays the adjusted heteroskedasticity-consistent covariance matrix estimator $\hat{V}_{HC1}$ in the matrix VHC1.

- **Interpreting the Elements of $\hat{V}_{HC1}$**

- $\hat{\mathbf{V}}_{\mathbf{HC1}}$ for the OLS sample regression equation

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{\beta}_3 X_{i3} + \hat{\beta}_4 X_{i4} + \hat{u}_i$$

is a **square, symmetric 5×5 positive definite matrix**, the elements of which are the estimated variances and covariances of the OLS coefficient estimates $\hat{\beta}_j$ j = 0, 1, …, 4. The symmetry of $\hat{V}_{HC1}$ follows from the fact that

$$\hat{Cov}(\hat{\beta}_f, \hat{\beta}_g) = \hat{Cov}(\hat{\beta}_g, \hat{\beta}_f) \quad \text{for all } f \neq g.$$

- The $\hat{V}_{HC1}$ matrix for the above OLS sample regression equation is written in general as:

$$\hat{V}_{HC1} = \begin{bmatrix} \hat{Var}(\hat{\beta}_0) & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_2) & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_3) & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_4) \\ \hat{Cov}(\hat{\beta}_1,\hat{\beta}_0) & \hat{Var}(\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_2) & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_3) & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_4) \\ \hat{Cov}(\hat{\beta}_2,\hat{\beta}_0) & \hat{Cov}(\hat{\beta}_2,\hat{\beta}_1) & \hat{Var}(\hat{\beta}_2) & \hat{Cov}(\hat{\beta}_2,\hat{\beta}_3) & \hat{Cov}(\hat{\beta}_2,\hat{\beta}_4) \\ \hat{Cov}(\hat{\beta}_3,\hat{\beta}_0) & \hat{Cov}(\hat{\beta}_3,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_3,\hat{\beta}_2) & \hat{Var}(\hat{\beta}_3) & \hat{Cov}(\hat{\beta}_3,\hat{\beta}_4) \\ \hat{Cov}(\hat{\beta}_4,\hat{\beta}_0) & \hat{Cov}(\hat{\beta}_4,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_4,\hat{\beta}_2) & \hat{Cov}(\hat{\beta}_4,\hat{\beta}_3) & \hat{Var}(\hat{\beta}_4) \end{bmatrix}$$

- The software program *Stata* stores and displays the elements of $\hat{V}_{HC1}$ in a slightly different arrangement than that given above. *Stata* places the estimated variances and covariances involving the intercept coefficient estimate $\hat{\beta}_0$ in the last row and last column of the $\hat{V}_{HC1}$ matrix, rather than in the first row and column.

$$\textit{Stata}\ \hat{V}_{HC1} = \begin{bmatrix} \hat{Var}(\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_2) & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_3) & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_4) & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_0) \\ \hat{Cov}(\hat{\beta}_2,\hat{\beta}_1) & \hat{Var}(\hat{\beta}_2) & \hat{Cov}(\hat{\beta}_2,\hat{\beta}_3) & \hat{Cov}(\hat{\beta}_2,\hat{\beta}_4) & \hat{Cov}(\hat{\beta}_2,\hat{\beta}_0) \\ \hat{Cov}(\hat{\beta}_3,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_3,\hat{\beta}_2) & \hat{Var}(\hat{\beta}_3) & \hat{Cov}(\hat{\beta}_3,\hat{\beta}_4) & \hat{Cov}(\hat{\beta}_3,\hat{\beta}_0) \\ \hat{Cov}(\hat{\beta}_4,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_4,\hat{\beta}_2) & \hat{Cov}(\hat{\beta}_4,\hat{\beta}_3) & \hat{Var}(\hat{\beta}_4) & \hat{Cov}(\hat{\beta}_4,\hat{\beta}_0) \\ \hat{Cov}(\hat{\beta}_0,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_2) & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_3) & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_4) & \hat{Var}(\hat{\beta}_0) \end{bmatrix}$$

Heteroskedastic-Robust Hypothesis Tests with OLS

- All F-tests of linear coefficient restrictions on the regression coefficient vector β can be formulated in general terms as tests of the following null and alternative hypotheses:

Null hypothesis H_0: $R\beta = r \quad \Leftrightarrow \quad R\beta - r = \underline{0}$

Alternative hypothesis H_1: $R\beta \neq r \quad \Leftrightarrow \quad R\beta - r \neq \underline{0}$

- The **general Wald F-statistic** takes the form:

$$F_{WALD} = \frac{1}{q} W = \frac{\left(R\hat{\beta} - r\right)^T \left(R \hat{V}_{\hat{\beta}} R^T\right)^{-1} \left(R\hat{\beta} - r\right)}{q} \tag{1}$$

where:

$\hat{\beta}$ = a ***consistent unrestricted* estimator of β**, such as the OLS estimator $\hat{\beta}_{OLS}$;

$\hat{V}_{\hat{\beta}}$ = a ***consistent* estimator of $\mathbf{V}_{\hat{\beta}}$**.

Question: What consistent covariance matrix estimator should be used in place of $\hat{V}_{\hat{\beta}}$ in formula (1)?

- ***Heteroskedasticity-consistent* Wald F-statistics** are obtained by simply using one of the heteroskedasticity-consistent estimators of the OLS coefficient covariance matrix in place of $\hat{V}_{\hat{\beta}}$ in formula (1) for F_{WALD}.

Either

1. set $\hat{V}_{\hat{\beta}} = \hat{V}_{HC} = \left(X^T X\right)^{-1} X^T \hat{V} X \left(X^T X\right)^{-1}$ in formula (1) for F_{WALD},

or

2. set $\hat{V}_{\hat{\beta}} = \hat{V}_{HC1} = \dfrac{N}{N-K}\hat{V}_{HC} = \dfrac{N}{N-K}\left(X^T X\right)^{-1} X^T \hat{V} X \left(X^T X\right)^{-1}$ in formula (1) for F_{WALD}.

- **Two Heteroskedasticity-Consistent Wald F-Statistics**

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

where

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

$$F_{HC} = \frac{1}{q} W_{HC} = \frac{\left(R\hat{\beta} - r\right)^T \left(R\hat{V}_{HC} R^T\right)^{-1} \left(R\hat{\beta} - r\right)}{q} \overset{a}{\sim} F[q, N-K] \text{ under } H_0 \qquad (3)$$

2. Set $\hat{V}_{\hat{\beta}} = \hat{V}_{HC1} = \frac{N}{N-K}\hat{V}_{HC} = \frac{N}{N-K}\left(X^TX\right)^{-1}X^T\hat{V}X\left(X^TX\right)^{-1}$

$$F_{HC1} = \frac{1}{q}W_{HC1} = \frac{\left(R\hat{\beta}-r\right)^T\left(R\hat{V}_{HC1}R^T\right)^{-1}\left(R\hat{\beta}-r\right)}{q} \overset{a}{\sim} F[q, N-K] \text{ under } H_0 \qquad (4)$$

- ***Stata*** **test** commands compute **adjusted HC Wald F-statistics F_{HC1}** when used following a **regress** command with the **robust** option.

```
regress y x1 x2 x3 x4, robust
test x3 x4
test x2 = 1
test x1
```

- ***Stata*** **lincom** commands compute **adjusted HC t-statistics t_{HC1}** when used following a **regress** command with the **robust** option.

```
regress y x1 x2 x3 x4, robust
lincom _b[x1]
lincom _b[x1] - _b[x2]
```