

Lecture 30: Estimation under heteroskedasticity: WLS, transforming the model, feasible GLS

Econometrics ■ Unit II: Multicollinearity and heteroskedasticity

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

B.Sc. (Hons) Statistics, Semester V
Academic Session 2026–2027

Previous lecture

Lecture 29: The Goldfeld–Quandt test; comparison of the three tests.

Today

- Estimation procedures under heteroskedastic disturbances
 - Variances known: weighted least squares
 - Variances known up to a constant: transforming the model
 - Variances unknown: feasible GLS
- Summary of key formulas

Variances known: weighted least squares

If $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ is known, GLS (Chapter 2) applies. Since the inverse of a diagonal matrix is the diagonal matrix of reciprocals,

$$\Sigma^{-1} = \text{diag}(w_1, \dots, w_n), \quad w_i = \frac{1}{\sigma_i^2},$$

and the GLS criterion becomes

$$(y - Xb)' \Sigma^{-1} (y - Xb) = \sum_{i=1}^n w_i (y_i - x_i' b)^2.$$

This is a *weighted* sum of squares: observations with large variance get small weight. The method is called **weighted least squares (WLS)**, and the estimator is

$$\hat{\beta}_W = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y, \quad \text{Var}(\hat{\beta}_W) = (X' \Sigma^{-1} X)^{-1}.$$

By Aitken's theorem it is BLUE.

Minimise

$$Q(\beta_0, \beta_1) = \sum_i w_i (y_i - \beta_0 - \beta_1 x_i)^2.$$

Step 1 (differentiate).

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_i w_i (y_i - \beta_0 - \beta_1 x_i),$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_i w_i x_i (y_i - \beta_0 - \beta_1 x_i).$$

Step 2 (weighted normal equations). Equating to zero,

$$\begin{aligned} \beta_0 \sum w_i + \beta_1 \sum w_i x_i &= \sum w_i y_i, \\ \beta_0 \sum w_i x_i + \beta_1 \sum w_i x_i^2 &= \sum w_i x_i y_i. \end{aligned}$$

Step 3 (solve). Define the weighted means

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}, \quad \bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}.$$

Dividing the first equation by $\sum w_i$ gives $\hat{\beta}_0 = \bar{y}_w - \hat{\beta}_1 \bar{x}_w$. Substituting into the second and simplifying,

$$\hat{\beta}_1 = \frac{\sum_i w_i (x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sum_i w_i (x_i - \bar{x}_w)^2}.$$

With all w_i equal, these become the OLS formulas.

Remember

WLS: $\hat{\beta}_W = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y$, with weights $w_i = 1/\sigma_i^2$.

Usually only the *form* of the variance is known, for example $\sigma_i^2 = \sigma^2 f(x_i)$ for a known function f . Since a constant factor does not change GLS (Remark in Section 2.4 of the notes), we divide each equation by $\sqrt{f(x_i)}$ and apply OLS.

Case 1: $\sigma_i^2 = \sigma^2 x_i^2$ ($x_i > 0$). Divide $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ by x_i :

$$\frac{y_i}{x_i} = \beta_0 \frac{1}{x_i} + \beta_1 + \frac{\varepsilon_i}{x_i}.$$

The new error has variance

$$\text{Var}\left(\frac{\varepsilon_i}{x_i}\right) = \frac{\sigma^2 x_i^2}{x_i^2} = \sigma^2,$$

which is constant.

Case 1 (cont.) So regress y_i/x_i on $1/x_i$ by OLS. Note the change of roles: the *intercept* of the transformed equation estimates the original slope β_1 , and the *slope* on $1/x_i$ estimates the original intercept β_0 .

Case 2: $\sigma_i^2 = \sigma^2 x_i$ ($x_i > 0$). Divide by $\sqrt{x_i}$:

$$\frac{y_i}{\sqrt{x_i}} = \beta_0 \frac{1}{\sqrt{x_i}} + \beta_1 \sqrt{x_i} + \frac{\varepsilon_i}{\sqrt{x_i}}, \quad \text{Var} \left(\frac{\varepsilon_i}{\sqrt{x_i}} \right) = \frac{\sigma^2 x_i}{x_i} = \sigma^2.$$

Regress $y_i/\sqrt{x_i}$ on $1/\sqrt{x_i}$ and $\sqrt{x_i}$ *without* an intercept.

Case 3: $\sigma_i^2 = \sigma^2 \{E(y_i)\}^2$. Divide by $\hat{y}_i$ from a first OLS fit; this is a feasible version of Case 1.

In every case the transformed model satisfies the OLS assumptions, so the usual t and F tests can be applied to it.

Example 4.4 (WLS with variance proportional to x^2)

Data: $x = (1, 2, 4)$, $y = (3, 6, 10)$, and suppose $\sigma_i^2 = \sigma^2 x_i^2$.

Transformed data. $Y_i^* = y_i/x_i = (3, 3, 2.5)$ and $X_i^* = 1/x_i = (1, 0.5, 0.25)$.

OLS on the transformed model $Y^* = \beta_1 + \beta_0 X^*$. Means:

$\bar{X}^* = 1.75/3 = 7/12$, $\bar{Y}^* = 8.5/3 = 17/6$. Deviations of X^* :

$(5/12, -1/12, -4/12)$; deviations of Y^* : $(1/6, 1/6, -1/3)$. Then

$$S_{X^*X^*} = \frac{25 + 1 + 16}{144} = \frac{42}{144} = \frac{7}{24}, \quad S_{X^*Y^*} = \frac{5}{72} - \frac{1}{72} + \frac{8}{72} = \frac{12}{72} = \frac{1}{6}.$$

So $\hat{\beta}_0 = \frac{1/6}{7/24} = \frac{4}{7} \approx 0.571$ and

$$\hat{\beta}_1 = \bar{Y}^* - \hat{\beta}_0 \bar{X}^* = \frac{17}{6} - \frac{4}{7} \cdot \frac{7}{12} = \frac{17}{6} - \frac{1}{3} = \frac{5}{2}.$$

Example 4.4 (cont.)

Check with the weighted normal equations ($w_i = 1/x_i^2 = 1, 1/4, 1/16$):

$$\sum w_i = \frac{21}{16}, \quad \sum w_i x_i = \frac{7}{4}, \quad \sum w_i x_i^2 = 3, \quad \sum w_i y_i = \frac{41}{8}, \quad \sum w_i x_i y_i = \frac{17}{2}.$$

Then

$$\frac{21}{16} \cdot \frac{4}{7} + \frac{7}{4} \cdot \frac{5}{2} = \frac{3}{4} + \frac{35}{8} = \frac{41}{8}, \quad \frac{7}{4} \cdot \frac{4}{7} + 3 \cdot \frac{5}{2} = 1 + \frac{15}{2} = \frac{17}{2}. \quad \checkmark$$

Comparison with OLS. OLS on the original data gives $\bar{x} = 7/3$, $\bar{y} = 19/3$, $S_{xx} = 14/3$, $S_{xy} = 32/3$, so $b_1 = 16/7 \approx 2.286$ and $b_0 = 1$.

WLS gives $\hat{y} = 0.571 + 2.5x$, because it gives less weight to the observation at $x = 4$, where the variance is largest.

When the σ_i^2 are unknown, they must be estimated; with n observations we cannot estimate n separate variances, so some structure is assumed.

(a) Grouped variances. Suppose the data form groups, with the same variance within each group. Fit the model in each group and compute s_j^2 . Then build $\hat{\Sigma}$ by putting s_j^2 in the positions of group j . For example, with two groups of sizes n_1, n_2 and estimates $s_1^2 = 4$, $s_2^2 = 9$,

$$\hat{\Sigma} = \text{diag}(4I_{n_1}, 9I_{n_2}),$$

and $\hat{\beta}_F = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y$. Equivalently, divide the observations of group j by s_j and apply OLS.

(b) A variance function. Suppose $\sigma_i^2 = \exp(\alpha_1 + \alpha_2 z_i)$, as in the Breusch–Pagan model. Then:

1. fit OLS and obtain e_i ;
2. regress $\log e_i^2$ on a constant and z_i and obtain fitted values $\hat{g}_i$;
3. set $\hat{\sigma}_i^2 = \exp(\hat{g}_i)$ and apply WLS with weights $1/\hat{\sigma}_i^2$.

As noted in Section 2.8 of the notes, FGLS is only approximately efficient, and its properties hold in large samples.

- $\text{Var}(\epsilon) = \Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$; OLS unbiased,
 $\text{Var}(b) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$; simple regression
 $\text{Var}(b_1) = \sum (x_i - \bar{x})^2 \sigma_i^2 / S_{xx}^2$.
- Bartlett: $B = \{\nu \log s_p^2 - \sum \nu_j \log s_j^2\} / C \sim \chi_{m-1}^2$,
 $C = 1 + \frac{1}{3(m-1)} \left(\sum \frac{1}{\nu_j} - \frac{1}{\nu} \right)$.
- Breusch-Pagan: $p_i = e_i^2 / \tilde{\sigma}^2$, $\tilde{\sigma}^2 = \sum e_i^2 / n$; regress p_i on z 's;
 $\Theta = \text{ESS} / 2 \sim \chi_{p-1}^2$.
- Goldfeld-Quandt: $F_{GQ} = (\text{RSS}_H / \nu_H) / (\text{RSS}_L / \nu_L) \sim F_{\nu_H, \nu_L}$.
- WLS: $\hat{\beta}_W = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$, weights $w_i = 1/\sigma_i^2$.
- $\sigma_i^2 = \sigma^2 x_i^2$: divide by x_i . $\sigma_i^2 = \sigma^2 x_i$: divide by $\sqrt{x_i}$.

Practice

- Exercises 4.6 and 4.7 (Exercises and Solutions booklet).
- R practical P03: Weighted least squares for county cancer counts.
- R practical P08: Robust inference, transformation and FGLS.

Next lecture

Lecture 31 begins Unit III: Autocorrelation and instrumental variables.