The Constant Variance Assumption

Our assumption about the covariance structure of the errors, $V(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}_n$, implies equal variances for all observations. When this assumption is violated, the least-squares estimator is inefficient, even though it remains unbiased and consistent. Common violations of this assumption lead to an increase in variance with increases in $E(Y)$ or to patterns relating error variance to a covariate. We can use plots of residuals against predicted values and covariate values to try to spot either problem. Variations on these types of plots use studentized residuals instead of raw residuals, or $|E_i^*|$ or E_i^{*2} against $\widehat{Y}_i$ values. We can enhance these plots by using a local regression or robust regression fit to help identify non-constant variance. For data with only positive $\widehat{Y}_i$ values, Tukey proposed a method for selecting a power transformation by fitting a line to a plot of $\log |E_i^*|$ versus $\log \widehat{Y}_i$.

Weighted-Least Squares Estimation

If we have heterogeneous variances but we know their values, weighted-least squares is a desirable alternative. Suppose the model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with $E(\boldsymbol{\varepsilon}) = 0$ but $V(\varepsilon_i) = \sigma_i^2$, and that the error variances are known up to a proportional constant, so $\sigma_i^2 = \sigma_\varepsilon^2 / w_i^2$. Then if we define the covariance matrix as $\boldsymbol{\Sigma} = \sigma_\varepsilon^2 \text{diag}(1/w_1^2, 1/w_2^2, \ldots, 1/w_n^2) \equiv \sigma_\varepsilon^2 \mathbf{W}^{-1}$, then the likelihood for the sample is

$$L(\boldsymbol{\beta}, \sigma_\varepsilon^2) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right].$$

The maximum-likelihood estimators for $\boldsymbol{\beta}$ and σ_ε^2 then are:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$$
$$\widehat{\sigma}_\varepsilon^2 = \frac{\sum (w_i E_i)^2}{n}$$

where the residuals E_i are as before, $\mathbf{e} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$. The maximum-likelihood estimators minimize the weighted sum of squares $\sum w_i^2 E_i^2$ and the estimated asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$ is:

$$\widehat{V}(\widehat{\boldsymbol{\beta}}) = \widehat{\sigma}_\varepsilon^2 (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}.$$

To implement weighted-least-squares methods we must either estimate the weights w_i or find a relationship between the variance and some variable.

The Sandwich Estimator for Correcting OLS Standard Errors

It is often difficult to find the information needed to implement weighted-least-squares procedures, but another approach proposed by White (1980) presents an alternative method. Recall that the covariance matrix of the least-squares estimator is:

$$V(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'V(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

which simplifies to $V(\mathbf{b}) = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}$ for the usual assumption of $V(\mathbf{y}) = \sigma_\varepsilon^2\mathbf{I}_n$. If the errors were instead independent but unequal with covariance matrix $\mathbf{\Sigma} \equiv$diag$(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$ then we have

$$V(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

Since $E(\varepsilon_i) = 0$ we have $\sigma_i^2 = E(\varepsilon_i^2)$, which suggests that we can estimate $\mathbf{\Sigma}$ with the E_i^2 values or similar statistics. This leads to different types of 'sandwich' estimators for $V(\mathbf{b})$, including:

$$\widetilde{V}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{\Sigma}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

where $\widehat{\mathbf{\Sigma}}\equiv$diag$(E_1^2, , \ldots, E_n^2)$ and

$$\widetilde{V}^*(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{\Sigma}}^*\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

where $\widehat{\mathbf{\Sigma}}^*\equiv$diag$(E_1^2/(1-h_1)^2, , \ldots, E_n^2/(1-h_n)^2)$. This method offers a more automatic way to adjust coefficient standard errors and tests for variance heterogeneity, but may be less efficient than transformations or WLS for cases of severe heterogeneity.

The Effect of Nonconstant Error Variance on the Least-Squares Estimator

The text presents calculations to suggest a rule of thumb, that the effect of heteroscedasticity on the least-squares estimator is serious if the ratio of largest to smallest variance is 4-10 or more.