

Section 8 Heteroskedasticity

Basics of heteroskedasticity

- We have assumed up to now (in our SR and MR assumptions) that the variance of the error term was constant across observations.
 - This is unrealistic in many or most econometric applications, especially in cross sections.
- *Heteroskedasticity* occurs when different observations have different error variance.
- Effects of heteroskedasticity
 - We will see that OLS estimators are unbiased and consistent in the presence of heteroskedasticity, but they are not efficient and the estimated standard errors are inconsistent, so test statistics using the standard error are not valid
- Detection of heteroskedasticity
 - At a visual level, we can look for heteroskedasticity by examining the plot of residuals against predicted values or individual explanatory variables to see if the spread of residuals seems to depend on these variables.
 - There are tests that formalize these visual descriptions, regressing the squared residuals on predicted values or explanatory variables.
- Dealing with heteroskedasticity: Two choices
 - Use inefficient OLS estimator but use “robust” standard errors that allow for the presence of heteroskedasticity
 - This is the easiest and most common solution
 - Use *weighted least squares (WLS)* to calculate efficient estimators, conditional on correct knowledge of the pattern of heteroskedasticity
 - This is the better solution *if* we know the pattern, which we usually don’t

Effects of heteroskedasticity

- Simple regression (multiple is similar) model with heteroskedasticity:
$$y_i = \beta_1 + \beta_2 x_i + e_i,$$
$$E(e_i) = 0,$$
$$\text{var}(e_i) = \sigma_i^2,$$
$$\text{cov}(e_i, e_j) = 0, i \neq j.$$
 - Covariance matrix of error vector is a diagonal matrix, but not a scalar matrix.

- We derived earlier that the OLS slope estimator could be written as

$$\begin{aligned}
 b_2 &= \beta_2 + \frac{\sum_{i=1}^N (x_i - \bar{x}) e_i}{\sum_{n=1}^N (x_n - \bar{x})^2} \\
 &= \beta_2 + \sum_{i=1}^N w_i e_i,
 \end{aligned}$$

with $w_i \equiv \frac{x_i - \bar{x}}{\sum_{n=1}^N (x_n - \bar{x})^2}$.

- **OLS is unbiased under heteroskedasticity:**

$$\begin{aligned}
 E(b_2) &= E\left[\beta_2 + \sum_{i=1}^N w_i e_i\right] \\
 &= \beta_2 + \sum_{i=1}^N w_i E(e_i) = \beta_2
 \end{aligned}$$

- This uses the assumption that the x values are fixed to allow the expectation of e to go inside the product.

- **Variance of OLS under heteroskedasticity is not the usual formula**

$$\begin{aligned}
 \text{var}(b_2) &= \text{var}\left(\sum_{i=1}^N w_i e_i\right) \\
 &= \sum_{i=1}^N w_i^2 \text{var}(e_i) + \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N w_i w_j \text{cov}(e_i, e_j) \\
 &= \sum_{i=1}^N w_i^2 \sigma_i^2 \\
 &= \frac{\sum_{i=1}^N [(x_i - \bar{x})^2 \sigma_i^2]}{\left[\sum_{n=1}^N (x_n - \bar{x})^2\right]^2}.
 \end{aligned}$$

- If σ_i^2 is constant, then we can take it out of the numerator summation and the numerator summation on deviations of x cancels one of the denominator

summations, leaving the usual formula: $\frac{\sigma^2}{\sum_{n=1}^N (x_n - \bar{x})^2}$.

- If the variance is not constant, we can't do this and the ordinary variance estimator is incorrect.

- **OLS is inefficient with heteroskedasticity**

- We don't prove this, but the Gauss-Markov Theorem requires homoskedasticity, so the OLS estimator is no longer BLUE.

Detecting heteroskedasticity

- The eye-ball test is a simple but casual way to look for heteroskedasticity
 - Plot the residuals (or the squared residuals) against the explanatory variables or the predicted values of the dependent variable
 - If there is an apparent pattern, then there is heteroskedasticity of the type that the variance is related to x or \mathbf{x}^2 .
- The **Breusch-Pagan test** is a formal way to test whether the error variance depends on anything observable.
 - Suppose that $\text{var}(e_i) = \sigma_i^2 = E(e_i^2) = h(\alpha_1 + \alpha_2 z_{i,2} + \dots + \alpha_S z_{i,S})$, where the z variables may be the same or different from the x variables in the regression, and h may be any kind of function.
 - To test this, we regression the squared residuals on the z variables and test the hypothesis that $\alpha_2 \dots \alpha_S$ are all zero:
 - $\hat{e}_i^2 = \alpha_1 + \alpha_2 z_{i,2} + \dots + \alpha_S z_{i,S} + v_i$
 - Several possible tests of $H_0 : \alpha_2 = \alpha_3 = \dots = \alpha_S = 0$:
 - Lagrange multiplier test is $NR^2 \sim \chi_{(S-1)}^2$
 - Reject homoskedasticity if test statistic $>$ critical value
 - This is asymptotic test
 - Can use **estat hettest, iid** in Stata if the z variables are just the set of x variables
 - You can include a variable list after the word hettest in this command to make the z variables differ from the x variables in the regression
- The **White test** is a test that is similar to the Breusch-Pagan test, using as the z variables
 - All of the x variables in the original equation
 - The squares of all of the x variables
 - Optionally, the cross-products of the x variables.
 - This leads to lots of variables if K is large
 - Dummies can't be included as squares
- The **Goldfeld-Quandt** test is suitable for samples in which the data can be divided into two groups and with variance differing only between the groups.
 - Suppose that the groups are A and B with variances σ_A^2 and σ_B^2 .
 - Run separate regressions for the two sub-samples A and B and calculate the estimated error variances from the residuals
 - $$F = \frac{\hat{\sigma}_A^2 / \sigma_A^2}{\hat{\sigma}_B^2 / \sigma_B^2} \sim F_{(N_A - K, N_B - K)}$$

- If the null hypothesis is that $\sigma_A^2 = \sigma_B^2$, then the ratio of the estimated variances is the F statistic and we can do a one-tailed or two-tailed test.
 - Must be careful with the two-tailed F test, though, because F tables only report the right-hand tail area and critical values.
 - Make A the sample with the larger variance so that all of the critical area is on the right.
 - The one-tailed test with alternative hypothesis $\sigma_A^2 > \sigma_B^2$ is just the ordinary F test with the usual critical value.
 - For the two-tailed test, a 5% critical value becomes a 10% critical value because of the possibility that the variance of A is smaller than the variance of B .

Heteroskedasticity-consistent standard errors

- The first, and most common, strategy for dealing with the possibility of heteroskedasticity is **heteroskedasticity-consistent standard errors** (or **robust errors**) developed by White.
- We use OLS (inefficient but) consistent estimators, and calculate an alternative (“robust”) standard error that allows for the possibility of heteroskedasticity.

- From above, $\text{var}(b_2) = \frac{\sum_{i=1}^N [(x_i - \bar{x})^2 \sigma_i^2]}{\left[\sum_{n=1}^N (x_n - \bar{x})^2 \right]^2}$ under heteroskedasticity.

- Logical estimator for variance is

$$\widehat{\text{var}}(b_2)_{\text{robust}} = \frac{\sum_{i=1}^N [(x_i - \bar{x})^2 \hat{e}_i^2] / (N - K)}{\left[\sum_{n=1}^N (x_n - \bar{x})^2 \right]^2 / N}$$

- This compares with the homoskedasticity-only estimator of

$$\widehat{\text{var}}(b_2)_{\text{homoskedastic}} = \frac{\sum_{i=1}^N \hat{e}_i^2 / (N - K)}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

- In matrix terms, the covariance matrix of the coefficient vector is

$$\widehat{\text{var}}(\mathbf{b})_{\text{robust}} = \frac{N}{N - K} (\mathbf{X}'\mathbf{X})^{-1} \hat{\Sigma} (\mathbf{X}'\mathbf{X})^{-1}, \text{ with } \hat{\Sigma} = \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i' \hat{e}_i^2$$

- Stata calculates the White heteroskedasticity-consistent standard errors with the option “robust” in most regression commands.

- Many econometricians argue that one should pretty much always use robust standard errors because one never can count on homoskedasticity

Weighted least squares

- If one wants to correct for heteroskedasticity by using a fully efficient estimator rather than accepting inefficient OLS and correcting the standard errors, the appropriate estimator is **weight least squares**, which is an application of the more general concept of **generalized least squares**.
- The GLS estimator applies to the least-squares model when the covariance matrix of \mathbf{e} is a general (symmetric, positive definite) matrix $\mathbf{\Omega}$ rather than $\sigma^2\mathbf{I}_N$.

- $\hat{\beta}_{GLS} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y})$

- The most intuitive approach to GLS is to find the “Cholesky root” matrix \mathbf{P} such that $\mathbf{P}'\mathbf{P}$ is equal to $\mathbf{\Omega}^{-1}$.

- This gives us $\mathbf{Py} = \mathbf{PX}\beta + \mathbf{Pe}$, for which the OLS estimator is

$$\begin{aligned} \mathbf{b}_{transformed} &= \left((\mathbf{PX})' (\mathbf{PX}) \right)^{-1} (\mathbf{PX})' \mathbf{Py} \\ &= (\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X})^{-1} \mathbf{X}'\mathbf{P}'\mathbf{Py} \\ &= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y} = \hat{\beta}_{GLS} \end{aligned}$$

- Thus we can use the usual OLS procedure on the transformed model to get the efficient GLS estimator
 - This estimator is sometimes called “infeasible” GLS because it requires that we know $\mathbf{\Omega}$, which we usually don’t.
 - “Feasible” GLS is when we use an estimator for $\mathbf{\Omega}$ rather than the actual value.
- For the case of heteroskedasticity,

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \vdots \\ \vdots & 0 & \ddots & \\ 0 & \dots & 0 & \sigma_N^2 \end{bmatrix}, \text{ and the corresponding } \mathbf{F} \text{ matrix is}$$

$$\mathbf{P} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sigma_N} \end{bmatrix}$$

- Multiplying the \mathbf{X} , \mathbf{y} , and \mathbf{e} matrices by \mathbf{P} transforms each observation by dividing x , y , and e by σ_i

- Thus, WLS consists of OLS on the transformed model

$$y_i^* = \frac{y_i}{\sigma_i}, x_{i,j}^* = \frac{x_{i,j}}{\sigma_i}, e_i^* = \frac{e_i}{\sigma_i}$$

- Note that the constant term is no longer just ones, because each x_1 is divided by a different σ .
- Because $\text{var}(e_i) = \sigma_i^2$, $\text{var}(e_i^*) = 1$.
 - Thus dividing each observation by something *proportional to* the error standard deviation for the observation converts the model to a homoskedastic one.
- HGL example: $\text{var}(e_i) = x_i \sigma^2$.
 - If we divide by $\sqrt{x_i}$ then the variance of the transformed model is σ^2 for all observations.
 - This is why we call it “weighted” least squares: we weight each observation by the reciprocal of its standard deviation, giving greatest weight to the observations for which the error variance is smallest.
- If pattern of heteroskedasticity is unknown, we may estimate a model such as $\ln(\hat{e}_i^2) = \alpha_1 + \alpha_2 z_{i,2} + \dots + \alpha_S z_{i,S} + \nu_i$ and use fitted values as estimates of the error variance of each observation.
 - This feasible GLS is consistent and asymptotically efficient as long as we have the pattern of heteroskedasticity specified correctly.