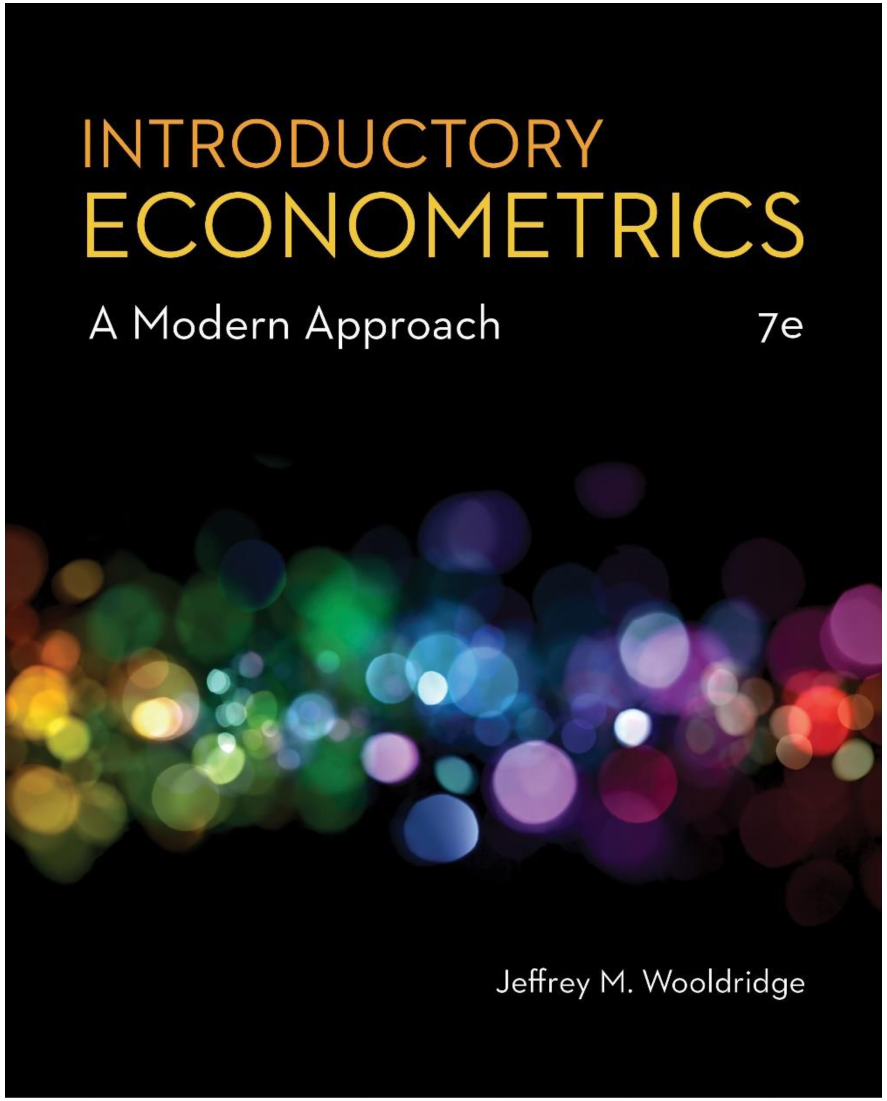
Chapter 8

Heteroskedasticity



Heteroskedasticity (1 of 18)

- **Consequences of heteroskedasticity for OLS**
  - OLS still unbiased and consistent under heteroskedastictiy!
  - Also, interpretation of R-squared is not changed

$$R^2 \approx 1 - \frac{\sigma_u^2}{\sigma_y^2}$$

Unconditional error variance is unaffected by heteroskedasticity (which refers to the conditional error variance)

  - **Heteroskedasticity invalidates variance formulas** for OLS estimators
  - The usual F tests and t tests are not valid under heteroskedasticity
  - Under heteroskedasticity, OLS is no longer the best linear unbiased estimator (BLUE); there may be more efficient linear estimators

Heteroskedasticity (2 of 18)

- **Heteroskedasticity-robust inference after OLS estimation**
  - Formulas for OLS standard errors and related statistics have been developed that are robust to heteroskedasticity of unknown form.
  - **All formulas are only valid in large samples**.
  - Formula for **heteroskedasticity-robust OLS standard error**.

$$\widehat{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$$

Also called White/Huber/Eicker standard errors. They involve the squared residuals from the regression and from a regression of x_j on all other explanatory variables.

  - Using these formulas, the usual t test is valid asymptotically.
  - The usual F statistic does not work under heteroskedasticity, but heteroskedasticity robust versions are available in most software.

Heteroskedasticity (3 of 18)

- **Example: Hourly wage equation**

$$\widehat{\log}(wage) = -\ .128 + .0904\ educ + .0410\ exper - .0007\ exper^2$$
$$\begin{matrix} (.105) & (.0075) & (.0052) & (.0001) \\ [.107] & [.0078] & [.0050] & [.0001] \end{matrix}$$

Heteroskedasticity robust standard errors may be larger or smaller than their nonrobust counterparts. The differences are often small in practice.

$$H_0 : \beta_{exper} = \beta_{exper^2} = 0$$

$$F = 17.95$$

F statistics are also often not too different.

$$F_{robust} = 17.99$$

If there is strong heteroskedasticity, differences may be larger. To be on the safe side, it is advisable to always compute robust standard errors.

Heteroskedasticity (4 of 18)

- **Testing for heteroskedasticity**
  - It may still be interesting whether there is heteroskedasticity because then OLS may not be the most efficient linear estimator anymore.

- Breusch-Pagan test for heteroskedasticity

$$H_0 : Var(u|x_1, x_2, \ldots, x_k) = Var(u|\mathbf{x}) = \sigma^2$$

$$Var(u|\mathbf{x}) = E(u^2|\mathbf{x}) - [E(u|\mathbf{x})]^2 = E(u^2|\mathbf{x})$$

Under MLR.4

$$\Rightarrow \quad E(u^2|x_1, \ldots, x_k) = E(u^2) = \sigma^2$$

The mean of u^2 must not vary with x_1, x_2, …, x_k

Heteroskedasticity (5 of 18)

- **Breusch-Pagan test for heteroskedasticity (cont.)**

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + error$$

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_k = 0$$

Regress squared residuals on all explanatory variables and test whether this regression has explanatory power.

$$F = \frac{R^2_{\hat{u}^2}/k}{1 - R^2_{\hat{u}^2}/(n-k-1)}$$

A large test statistic (= a high R-squared) is evidence against the null hypothesis.

$$LM = n \cdot R^2_{\hat{u}^2} \sim \chi^2_k$$

Alternative test statistic (= Lagrange multiplier statistic, LM). Again, high values of the test statistic (= high R-squared) lead to rejection of the null hypothesis that the expected value of u^2 is unrelated to the explanatory variables.

Heteroskedasticity (6 of 18)

Hprice1.dta

- **Example: Heteroskedasticity in housing price equations**

$$\widehat{price} = -\underset{(29.48)}{21.77} + \underset{(.0006)}{.0021}\ lotsize + \underset{(.013)}{.123}\ sqrft + \underset{(9.01)}{13.85}\ bdrms$$

$$\Rightarrow R^2_{\hat{u}^2} = .1601,\ p\text{-}value_F = .002,\ p\text{-}value_{LM} = .0028$$

Heteroskedasticity

$$\widehat{\log}(price) = -\underset{(.65)}{1.30} + \underset{(.038)}{.168}\ \log(lotsize) + \underset{(.093)}{.700}\ \log(sqrft) + \underset{(.028)}{.037}\ bdrms$$

$$\Rightarrow R^2_{\hat{u}^2} = .0480,\ p\text{-}value_F = .245,\ p\text{-}value_{LM} = .2390$$

In the logarithmic specification, homoskedasticity cannot be rejected

Heteroskedasticity (7 of 18)

- **The White test for heteroskedasticity**

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 + \delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + error$$

Regress squared residuals on all explanatory variables, their squares, and in-teractions (here: example for k=3)

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_9 = 0$$

$$LM = n \cdot R^2_{\hat{u}^2} \sim \chi^2_9$$

The White test detects more general deviations from heteroskedasticity than the Breusch-Pagan test

- Disadvantage of this form of the White test
  - Including all squares and interactions leads to a large number of estimated parameters (e.g. k=6 leads to 27 parameters to be estimated).

Heteroskedasticity (8 of 18)

- **Alternative form of the White test**

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + error$$

This regression indirectly tests the dependence of the squared residuals on the explanatory variables, their squares, and interactions, because the predicted value of y and its square implicitly contain all of these terms.

- Example: Heteroskedasticity in (log) housing price equations

$$H_0 : \delta_1 = \delta_2 = 0,\ \ LM = n \cdot R^2_{\hat{u}^2} \sim \chi^2_2$$

$$R^2_{\hat{u}^2} = .0392, LM = 88(.0392) \approx 3.45, p{-}value_{LM} = .178$$

Heteroskedasticity (9 of 18)

- **Weighted least squares estimation**
- Heteroskedasticity is known up to a multiplicative constant

$$Var(u_i|\mathbf{x}_i) = \sigma^2 h(\mathbf{x}_i), \;\; h(\mathbf{x}_i) = h_i > 0$$ ← The functional form of the heteroskedasticity is known

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$$

$$\Rightarrow \left[\frac{y_i}{\sqrt{h_i}}\right] = \beta_0 \left[\frac{1}{\sqrt{h_i}}\right] + \beta_1 \left[\frac{x_{i1}}{\sqrt{h_i}}\right] + \cdots + \beta_k \left[\frac{x_{ik}}{\sqrt{h_i}}\right] + \left[\frac{u_i}{\sqrt{h_i}}\right]$$

$$\Leftrightarrow \;\; y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \cdots + \beta_k x_{ik}^* + u_i^*$$ ← Transformed model

Heteroskedasticity (10 of 18)

- **Example: Savings and income**

$$sav_i = \beta_0 + \beta_1 inc_i + u_i, \ \ Var(u_i|inc_i) = \sigma^2 inc_i$$

$$\left[\frac{sav_i}{\sqrt{inc_i}}\right] = \beta_0 \left[\frac{1}{\sqrt{inc_i}}\right] + \beta_1 \left[\frac{inc_i}{\sqrt{inc_i}}\right] + u_i^* \leftarrow$$

Note that this regression model has no intercept

- The transformed model is homoskedastic

$$E(u_i^{*2}|\mathbf{x}_i) = E\left[\left(\frac{u_i}{\sqrt{h_i}}\right)^2 |\mathbf{x}_i\right] = \frac{E(u_i^2|\mathbf{x})}{h_i} = \frac{\sigma^2 h_i}{h_i} = \sigma^2$$

- If the other Gauss-Markov assumptions hold as well, OLS applied to the transformed model is the best linear unbiased estimator.

Heteroskedasticity (11 of 18)

- **OLS in the transformed model is weighted least squares (WLS)**

$$\min \sum_{i=1}^{n} \left(\left[\frac{y_i}{\sqrt{h_i}} \right] - b_0 \left[\frac{1}{\sqrt{h_i}} \right] - b_1 \left[\frac{x_{i1}}{\sqrt{h_i}} \right] - \cdots - b_k \left[\frac{x_{ik}}{\sqrt{h_i}} \right] \right)^2$$

$$\Leftrightarrow \quad \min \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - \cdots - b_k x_{ik})^2 / h_i$$

Observations with a large variance get a smaller weight in the optimization problem

- Why is WLS more efficient than OLS in the original model?
  - Observations with a large variance are less informative than observations with small variance and therefore should get less weight.
- WLS is a special case of generalized least squares (GLS)

Heteroskedasticity (12 of 18)

- **Example: Financial wealth equation**

Net financial wealth

TABLE 8.1 Dependent Variable: *nettfa*

Independent Variables	(1) OLS	(2) WLS	(3) OLS	(4) WLS
inc	.821 (.104)	.787 (.063)	.771 (.100)	.740 (.064)
$(age - 25)^2$	—	—	.0251 (.0043)	.0175 (.0019)
male	—	—	2.48 (2.06)	1.84 (1.56)
e401k	—	—	6.89 (2.29)	5.19 (1.70)
intercept	−10.57 (2.53)	−9.58 (1.65)	−20.98 (3.50)	−16.70 (1.96)
Observations	2,017	2,017	2,017	2,017
R-squared	.0827	.0709	.1279	.1115

© Cengage Learning, 2016

Assumed form of heteroskedasticity

WLS estimates have considerably smaller standard errors (which is line with the expectation that they are more efficient).

Participation in 401K pension plan

Heteroskedasticity (13 of 18)

- Important special case of heteroskedasticity
  - If the observations are reported as averages at the city/county/state/-country/firm level, they should be weighted by the size of the unit

Average contribution to pension plan in firm i

Average earnings and age in firm i

Percentage firm contributes to plan

Heteroskedastic error term

$$\overline{contrib}_i = \beta_0 + \beta_1 \overline{earns}_i + \beta_2 \overline{age}_i + \beta_3 mrate_i + \overline{u}_i$$

$$\Rightarrow \quad Var(\overline{u}_i) = Var\left(\frac{1}{m_i}\sum_{i=1}^{m_i} u_{i,e}\right) = \sigma^2/m_i$$

Error variance if errors are homoskedastic at the individual-level

- If errors are homoskedastic at the individual-level, WLS with weights equal to firm size m_i should be used. If the assumption of homoskedasticity at the individual-level is not exactly right, one can calculate robust standard errors after WLS (i.e. for the transformed model).

Weighted Least Squares: Why and How

- In survey data, individuals are often sampled from population with different probabilities
  - For example, the most poor or rich individuals often has a higher probability to be sampled.
  - If we are computing the mean income, an unbiased estimator is the Horvitz-Thompson estimator:

$$\bar{x}^w = \frac{1}{M}\sum_{i=1}^{N}\frac{1}{\pi_i}x_i$$

where π_i is the sampling probability, and $M = \sum_i \frac{1}{\pi_i}$

- Principle: weight = inverse probability, i.e. Inverse Probability Weighting(IPW)

Weighted Least Squares: Why and How

- Similar weighting strategy can be used for survey data in regression analysis
  - Inverse of probability can be used as the weights in regression so that the regression is more representative.
  - Note that the inverse of probability is equivalent to number of individuals represented by a specific sample.
  - For example, if we have 1000 samples in Shanghai, then every single individual in Shanghai is sampled with probability 1000/24million=1/2400, which means each sample represents 2400 individuals in Shanghai.
  - In Stata: [w=swgt]

Weighted Least Squares: Why and How

- If the observations are reported as averages at the city/county/state/-country/firm level, each observation can be view as a single sample from the city/county/state/-country/firm
- So according to the IPW principle, the weights = size of the unit, which is equivalent to the weights above.
  - But here, the motivation is different: representative v.s. Heteroscedasticity
  - Since we don‘t konw whether the homoscedasticity of individuals‘ error term is satisfied, robust standard errors should be use.
  - For example, if the dependent variable if averages at the city level, then weight=city population.

Heteroskedasticity (14 of 18)

- **Unknown heteroskedasticity function (feasible GLS)**

$$Var(u|\mathbf{x}) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k) = \sigma^2 h(\mathbf{x})$$

Assumed general form of heteroskedasticity; exp-function is used to ensure positivity

$$u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k) \cdot v$$

$$\Rightarrow \quad \log(u^2) = \alpha_0 + \delta_1 x_1 + \cdots + \delta_k x_k + e$$

Multiplicative error (assumption: independent of the explanatory variables)

$$\log(\hat{u}^2) = \hat{\alpha}_0 + \hat{\delta}_1 x_1 + \cdots + \hat{\delta}_k x_k + error$$

$$\Rightarrow \quad \hat{h}_i = \exp(\hat{\alpha}_0 + \hat{\delta}_1 x_1 + \cdots + \hat{\delta}_k x_k)$$

Use inverse values of the estimated heteroskedasticity funtion as weights in WLS

Heteroskedasticity (15 of 18)

- **Example: Demand for cigarettes**
- Estimation by OLS

SMOKE.dta

Cigarettes smoked per day

Logged income and cigarette price

$$\widehat{cigs} = -\underset{(24.08)}{3.64} + \underset{(.728)}{.880}\ \log(income) - \underset{(5.773)}{.751}\ \log(cigpric)$$

$$- \underset{(.167)}{.501}\ educ - \underset{(.160)}{.771}\ age - \underset{(.0017)}{.0090}\ age^2 - \underset{(1.11)}{2.83}\ restaurn$$

Smoking restrictions in restaurants

$$n = 807, R^2 = .0526, p\text{-}value_{Breusch-Pagan} = .000$$

Reject homo-skedasticity

Heteroskedasticity (16 of 18)

- **Estimation by FGLS**

Now statistically significant

$$\widehat{cigs} = -\underset{(17.80)}{5.64} + \underset{(.44)}{1.30}\ \log(income) - \underset{(4.46)}{2.94}\ \log(cigpric)$$

$$-\underset{(.120)}{.463}\ educ + \underset{(.097)}{.482}\ age - \underset{(.0009)}{.0056}\ age^2 - \underset{(.80)}{3.46}\ restaurn$$

$$n = 807, R^2 = .1134$$

- Discussion
  - The income elasticity is now statistically significant; other coefficients are also more precisely estimated (without changing qualitative results).

Heteroskedasticity (17 of 18)

- **What if the assumed heteroskedasticity function is wrong?**
  - If the heteroskedasticity function is misspecified, WLS is still consistent under MLR.1 – MLR.4, but robust standard errors should be computed.
  - WLS is consistent under MLR.4 but not necessarily under MLR.4‘
  - If OLS and WLS produce very different estimates, this typically indicates that some other assumptions (e.g. MLR.4) are wrong.
  - If there is strong heteroskedasticity, it is still often better to use a wrong form of heteroskedasticity in order to increase efficiency.

Heteroskedasticity (18 of 18)

- **WLS in the linear probability model**

$$P(y=1|\mathbf{x}) = p(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

$$\Rightarrow \quad Var(y|\mathbf{x}) = p(\mathbf{x})\,[1 - p(\mathbf{x})]$$

In the LPM, the exact form of heteroskedasticity is known

$$\Rightarrow \quad \widehat{h}_i = \widehat{y}_i(1 - \widehat{y}_i)$$

Use inverse values as weights in WLS

- Discussion
  - Infeasible if LPM predictions are below zero or greater than one.
  - If such cases are rare, they may be adjusted to values such as .01/.99.
  - Otherwise, it is probably better to use OLS with robust standard errors.