ECON 3740: INTRODUCTION TO ECONOMETRICS

INSTRUCTOR: CHAOYI CHEN
Department of Economics and Finance, University of Guelph

Lecture 20

Lecture outline

Last lecture, we studied the definition of the heteroskedasticity for OLS. Today, we will

- Learn consequences with heteroskedasticity
- Suggest a heteroskedasticity-robust inference
- Test for heteroskedasticity
  - Motivation for the test and three test methods
  - The Graphical Method
  - The Breusch-Pagan (BP) Test

Heteroskedasticity for OLS: Consequences

- OLS is still unbiased under heteroskedasticity because Assumption MLR.4 $E[\mu_i|\mathbf{x}_i] = 0$ does **not** involve conditional variance.
- Also, interpretation of R^2 and adjusted R^2 is **not** changed because

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

  where σ_μ^2 is the unconditional variance of μ while heteroskedasticity is about the conditional variance of μ.
- However, heteroskedasticity invalidates variance formulas for OLS estimators.
- Hence, the usual F tests and t tests are **not** valid under heteroskedasticity because as mentioned before, normality assumption implies homoskedasticity.
- Under heteroskedasticity, OLS is **no longer** the best linear unbiased estimator (BLUE). There may be a more efficient linear estimator.

Heteroskedasticity for OLS: Heteroskedasticity-Robust Inference

- 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. (related to chapter five of the textbook, which will not be covered in this course)
- Formula for heteroskedasticity-robust OLS standard error is

$$\widehat{Var(\widehat{\beta}_j)} = \frac{\sum_{i=1}^{n} \widehat{r}_{ij}^2 \widehat{\mu}_i^2}{SSR_j^2} = SSR_j^{-1} [\sum_{i=1}^{n} \widehat{r}_{ij}^2 \widehat{\mu}_i^2] SSR_j^{-1}$$

  which is also called as Eicker/Huber/White standard errors or sandwich form standard errors. They involve the squared residuals from the regression, $\widehat{\mu}_i$ and from a regression of x_j on all other explanatory variables, $\widehat{r}_{ij}$.
- Using these formulas, the usual t-test is valid asymptotically ($n \longrightarrow \infty$)
- The usual F-statistic does not work under heteroskedasticity, but heteroskedasticity robust versions are available in most software (including *R* and *STATA*).

Heteroskedasticity for OLS: Heteroskedasticity-Robust Inference - An Example

- An Example To investigate the hourly wage, consider the following fitted regression line,

$$\widehat{log(wage)} = \underset{\substack{(0.105) \\ [0.107]}}{-0.128} \; \underset{\substack{(0.0075) \\ [0.0078]}}{+0.0904 educ} \; \underset{\substack{(0.0052) \\ [0.005]}}{+0.041 exper} \; \underset{\substack{(0.0001) \\ [0.0001]}}{-0.0007 exper^2}$$

  where (.) is the standard errors, and [.] is the heteroskedasticity-robust standard errors.
- Heteroskedasticity-robust standard errors may be larger or smaller than their non-robust counterparts. In most empirical applications, the heteroskedasticity-robust standard errors tend to be larger than the homoskedasticity-only standard errors. In other words, the t statistic using the heteroskedasticity-robust standard errors tend to be less significant (harder to reject the null).
- Consider the following null hypothesis,

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

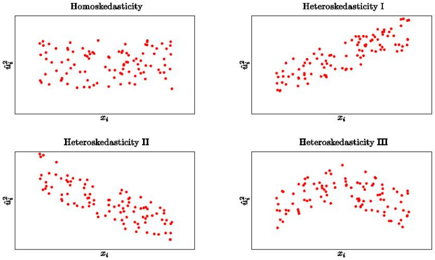
- The two F statistics are

$$F = 17.95 \text{ and } F_{robust} = 17.99$$

- Motivation: why we need to detect the heteroskedasticity? Although we can always use the robust standard errors regardless of homo/hetero, it may still be interesting whether there is heteroskedasticity because then OLS may not be the most efficient linear estimator anymore.
- Key idea behind all testing methods: σ_i^2 can be approximated by $\widehat{\mu}_i^2$ (we have already used this idea to compute the $\widehat{Var(\widehat{\beta}_j)}$). Note that μ_i is not observable, hence, μ_i^2 cannot be oberserved.
- We will introduce three methods to test for heteroskedasticity,
  - Method I: The Graphical Method
  - Method II: The Breusch-Pagan (BP) Test
  - Method III: The White Test

Heteroskedasticity for OLS: Testing for Heteroskedasticity - Method I: The Graphical Method

- The graphical method just need you to plot $\widehat{\mu}_i^2$ against x_i (or combination of $x_1, ..., x_k$) to check if there are some patterns.
- With homoskedasticity well see something like the first graph: no relationship between $\widehat{\mu}_i$ and x_i. Alternatively, with heteroskedasticity well see patterns like the other graphs: nonconstant variance (approximated by squared residuals).



Heteroskedasticity for OLS: Testing for Heteroskedasticity - Method II: The Breusch-Pagan (BP) Test

- Lets consider the following null hypothesis test,

$$H_0 : Var(\mu|x_1, ..., x_k) = Var(\mu|\mathbf{x}) = \sigma^2$$

  That is, under the null, we have homoskedastic error.
- Recall that, under assumption MLR.4, $Var(\mu|\mathbf{x}) = E[\mu^2|\mathbf{x}]$, hence, we can rewrite the above null hypothesis as

$$E[\mu^2|x_1, ..., x_k] = E[\mu^2] = \sigma^2,$$

  This shows that, in order to test for violation of the homoskedasticity assumption, we want to test whether μ^2 is related (in expected value) to one or more of the explanatory variables $(x_1, ..., x_k)$.
- The simplest way to model the possible relationship between μ^2 and the explanatory variables is to assume they have a linear relationship,

$$\widehat{\mu}_i^2 = \delta_0 + \delta_1 x_1 + ... + \delta_k x_k + error$$

  As a result, we just need to test

$$H_0 : \delta_1 = ... = \delta_k = 0,$$

  that is, regress squared residuals on all explanatory variables and test whether this regression has explanatory power.

Heteroskedasticity for OLS: Testing for Heteroskedasticity - Method II: The Breusch-Pagan (BP) Test Continue

- Naturally, we can use an F test. The resulting F staistic is,

$$F = \frac{R^2_{\widehat{\mu}^2}/k}{(1 - R^2_{\widehat{\mu}^2})/(n-k-1)} \sim F_{k,n-k-1}$$

  Question: why we ignored the R^2 from the restricted model here?
- A large test statistic (= a high R^2) is evidence against the null hypothesis.
- Alternatively, we can use the Lagrange multiplier (LM) statistic,

$$LM = nR^2_{\widehat{\mu}^2} \sim \chi^2_k$$

- Again, high R-squared leads to rejection of the null hypothesis.
- The LM version of the test is typically called the Breusch-Pagan (BP) test for heteroskedasticity (Breusch and Pagan (1979)).

- (Proof not required) Why $LM \sim \chi^2_k$? Recall from the lecture 2, $F \sim \chi^2_k / k$ when $n \longrightarrow \infty$. Also, we can rewrite LM as a function of F,

$$LM = kF * \frac{n}{n-k-1}(1 - R^2_{\widehat{\mu}^2}),$$

where, as $n \longrightarrow \infty$, $\frac{n}{n-k-1} \longrightarrow 1$ and $R^2_{\widehat{\mu}^2} \longrightarrow 0$ under the null. Hence,

$$LM = kF \sim kF_{k,n-k-1} \longrightarrow \chi^2_k$$

Heteroskedasticity for OLS: Testing for Heteroskedasticity - Method II: The Breusch-Pagan (BP) Test Continue

- A generalized steps to do the Breusch-Pagan test for heteroskedasticity are as follows:
  - Step 1: Estimate the econometric model by OLS. Then obtain the OLS residual $\widehat{\mu}_i$ and calculate the $\widehat{\mu}_i^2$.
  - Step 2: Regress $\widehat{\mu}_i^2$ on all explanatory variables $(x_1, ..., x_k)$. Keep the R^2 from the regression, $R^2_{\widehat{\mu}^2}$.
  - Step 3: Form either the F statistic or the LM statistic and complete the F test and the LM test.

Heteroskedasticity for OLS: Testing for Heteroskedasticity - Method II: The Breusch-Pagan (BP) Test - An Example

- Consider the following fitted regression line,

$\widehat{price}=$	-21.77	$+0.00207 lotsize$	$+0.123 sqrft$	$13.85 bdrms$
	(29.48)	(0.00064)	(0.013)	(9.01)

  where $n = 88$, $R^2 = 0.672$.
- Then, we regress $\widehat{\mu}_i^2$ on *lotsize*, *sqrft*, and *bdrms* and obtain the $R^2_{\widehat{\mu}^2} = 0.1601$.
- Therefore,

$$F = \frac{0.1601/3}{(1 - 0.1601)/(88 - 3 - 1)} \approx 5.34 \text{ with } pvalue = 0.002$$

$$LM = 88 * 0.1601 \approx 14.09 \text{ with } pvalue = 0.0028$$

- So both tests reject the null, and we conclude that the model is heteroskedastic when the dependent variable is *price*.

Heteroskedasticity for OLS: Testing for Heteroskedasticity - Method II: The Breusch-Pagan (BP) Test - An Example Continue

- Consider the following fitted regression line after the logarithmic transformation,

$$\widehat{log(price)} = \underset{(0.65)}{-1.3} + \underset{(0.038)}{0.168 log(lotsize)} + \underset{(0.093)}{0.7 log(sqrft)} \quad \underset{(0.028)}{0.037 bdrms}$$

where $n = 88$, $R^2 = 0.643$.
- Then, we regress $\widehat{\mu}_i^2$ on $log(lotsize)$, $log(sqrft)$, and $bdrms$ and obtain the $R^2_{\widehat{\mu}^2} = 0.048 < 0.1601$.
- Therefore,

$$F = \frac{0.048/3}{(1-0.048)/(88-3-1)} \approx 1.41 \text{ with } pvalue = 0.245$$

$$LM = 88 * 0.048 \approx 4.22 \text{ with } pvalue = 0.239$$

- So neither test can reject the null, and we conclude that the model is homoskedastic when the dependent variable is $log(price)$
- Taking logs on the dependent variable often helps to secure