# ECON-122
# Introduction to Econometrics

Agnieszka Postepska

August 3rd, 2011

# Heteroskedasticity - introduction

- now we will go back to our assumption and see what are the implication of violating assumption A5: HOMOSKEDASTICITY OF THE ERROR TERM
- in other words we will see what happens if $Var(u|x) \neq \hat{\sigma}^2$ but the variance of the error term depends on where in the data are we, i.e. it depends on the value of x
- this situation is known as HETEROSKEDASTIC ERROR TERM

$$Var(u|x) = \sigma^2 h(x_i)$$

- recall that we didn't use A5 in the proof of unbiasedness - OLS is still unbiased under heteroskedasticity
- what is biased now is the estimator of the variance (and thus the estimator for the standard error)
- also, distributions of t-statistic as well as F-statistic are not the same anymore
- also, OLS is no longer BLUE (why?)

# Motivating example

Consider the following relationship between savings and income:

$$sav_i = \beta_0 + \beta_1 inc_i + \beta X_i + u_i$$

- Does the assumption of homoskedasticity seem reasonable here?
- $Var(u_i | inc_i, X_i) = \sigma^2$ implies that the level of savings does not depend on the level of income
- intuitively we would expect to see higher savings at higher income levels, i.e.:

$$Var(u_i | inc_i, X_i) = \sigma^2 h(inc_i)$$

- this formulation seems more reasonable
- how do we deal with this?

# Finding estimator for the variance of the estimator of the $\hat{\beta}$

- we will not prove anything here- we'll be introduced to the White robust standard errors estimator directly
- in a simple regression model

$$\hat{Var}(\hat{\beta}_1) = \frac{\sum_i (x_i - \bar{x})^2 \hat{u}_i^2}{SST_x}$$

  where $\hat{u}_i^2$ is the OLS residuals from the initial regression of y on x and $SST_x$ is the total sum of squares of the $x's$:
  $SST_x = \sum_i (x_i - \bar{x})^2$
- this formula is an unbiased estimator for any form of heteroskedasticity (even homoskedasticity) so we do not really need to know whether we are facing heteroskedasticity or not to use it

# Finding estimator for the variance of the estimator of the $\hat{\beta}$ cont.

- the analog in the multiple regression model:

$$\hat{Var}(\hat{\beta}_j) = \frac{\sum_i \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$$

  where $\hat{r}_{ij}^2$ is the $i^{th}$ OLS residual from regressing $x_j$ on all other independent variables and $SSR_j$ is the sum of squared residuals from this regression
- similarly as before, the square root is called heteroskedasticity-robust standard error
- no general claims can be made about the relationship between the OLS standard errors and the robust ones (i.e. we cannot predict the bias)

# Testing under heteroskedasticity: robust t-statistic and F statistic

- the only difference with the original case, with homoskedastic standard error, is how the standard error is computed
- recall the general form for t-statistic:

$$t = \frac{\text{estimate-hypothesized value}}{\text{standard error}}$$

- the problem with testing though is that under heteroskedasticity the t-statistic have the t distributions only in large samples
- within small samples the distribution might be far away from t distributions
- for F test - it is possible to obtain robust F-statistic (known as Wald test) but as derivations are difficult we will skip it in this course

# Testing for heteroskedasticity

- To summarize - the reasons why we do want a way to test for heteroskedasticity:
  - robust standard errors are only valid in large samples
  - t-statistic is only valid in large samples
  - F-test - more complicated
  - OLS is no longer BLUE under heteroskedasticity

  So we want to make sure we face heteroskedasticity before we use the heteroskedasticity-robust standard errors!
- there are two methods to test whether we have heteroskedasticity in the model:
  - BREUSCH-PAGAN TEST
  - WHITE TEST
- before we proceed to the formal tests let's see if we can gain something just by visual inspection (graph)

# Breusch Pagan test

- this test checks whether the variance is a function of some combination of known variables
- it is very general since it does not require us to know the form of heteroskedasticity but due to its generality we are more likely not to reject the null if it is not true (higher probability of type II error)
- so the model used to test for heteroskedasticity is as follows:

$$u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + .... + \delta_k x_k + error$$

- the null and the alternative are as follows:

$$\begin{aligned} H_0 &: \quad \delta_0 = \delta_1 = ... = \delta_k = 0 \text{ (homoskedastic error term)} \\ H_1 &: \quad \text{at least one of the } \delta' s \neq 0 \text{ (heteroskedastic error term)} \end{aligned}$$

# Breusch Pagan test cont.

- as errors are unknown, we estimate the following model:

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} + .... + \delta_k x_{ik} + error$$

- and the test statistic for Breusch Pagan test:

$$nR^2 \sim \chi_k^2$$

where $n$ is the sample size and $R^2$ comes from regression of $\hat{u}_i^2$ on a constant and all variables thought to affect the error variance

# Breusch Pagan test cook book procedure

1. estimate the model using OLS and get $\hat{u}_i^2$
2. run the regression of $\hat{u}_i^2$ on all $x's$ in the model and get $R^2$ from this regression
3. calculate the test statistic $nR^2$
4. pick significance level and get critical value from tables for $\chi^2_k$ distribution
5. reject the null if $nR^2 > c_\alpha$, fail to reject the null otherwise (if reject the null, conclude that there is enough evidence against homoskedasticity - error term is heteroskedastic)

# White test

- this test checks whether the error variance is affected by any of the variables, their squares and cross-products
- the strength of this test is that it checks for variety of possible functional form of heteroskedasticity
- so the model used to test for heteroskedasticity (for $k = 3$ independent variables) is as follows:

$$\begin{aligned} u^2 &= \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 \\ &+ \delta_4 x_1 x_2 + \delta_5 x_1 x_3 + \delta_6 x_2 x_3 + error \end{aligned}$$

- the null and the alternative are as follows:

$$\begin{aligned} H_0 &: \delta_0 = \delta_1 = ... = 0 \text{ all } \delta' s \text{ are zero (homoskedastic error term)} \\ H_1 &: \text{at least one of the } \delta' s \neq 0 \text{ (heteroskedastic error term)} \end{aligned}$$

# White test cont.

- so in our example with $k = 3$ we estimate the following model:

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

- and we use F-test to test the null hypothesis (we test for all joint restrictions; 9 in our example) at the same time)
- or we can compute the White statistic:

$$nR^2 \sim \chi^2_{l-1}$$

where $l$ is the number of parameters estimated in the regression involving residuals ($l = 9$ in our case)

# White test cook book procedure

1. estimate the model using OLS and get $\hat{u}_i^2$
2. run the regression of $\hat{u}_i^2$ on all $x's$, $(x's)^2$ and cross products of all $x's$ in the model and get $R^2$ from this regression
3. calculate the test statistic $nR^2$
4. pick significance level and get critical value from tables for $\chi^2_{l-1}$ distribution
5. reject the null if $nR^2 > c_\alpha$, fail to reject the null otherwise (if reject the null, conclude that there is enough evidence against homoskedasticity - error term is heteroskedastic)