Econometrics 1
Lecture 7: Heteroscedasticity
ENSAE 2014/2015

Michael Visser (CREST-ENSAE)

Outline of lecture

Today's lecture is devoted to:

- Consequences of heteroscedasticity
- Testing for the presence of heteroscedasticity
  - Breusch-Pagan test
  - White test
- Remedies for Heteroscedasticity
  - Weighted Least Squares
  - Feasible Generalized Least Squares
  - Robust Standard Errors

Motivation

Consider again the multiple linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i, \;\; i = 1, ..., n. \qquad (1)$$

Until now we studied the model under the homoscedasticity Assumption MLR.5.

We thus assumed that all error terms have the same conditional variance:

$$Var(u_i | x_{i1}, ..., x_{ik}) = \sigma^2$$

Recall that this implies that the conditional variance of the outcome variable does not vary with the regressors:

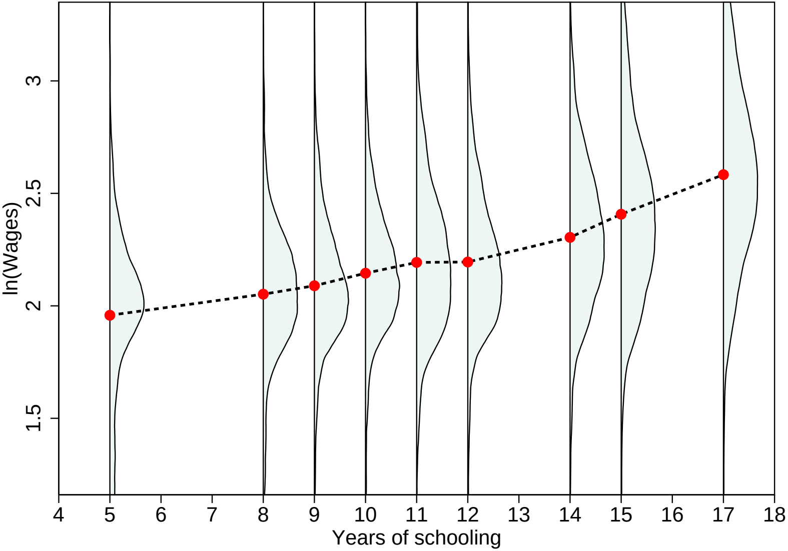
$$Var(y_i | x_{i1}, ..., x_{ik}) = \sigma^2$$

This may be a restrictive assumption in many settings. Some examples:

Motivation (ctnd)

- Cross-sectional data on units of different size (cities, states, countries). The order of magnitude of the omitted variables may be larger for the more populous cities, states, or countries. Consequently, the variance of the error term may vary with the size of cities, states, or countries.
- Data on individuals or households facing more or less strong restrictions on their behavior. For instance, in budget surveys, expenditure levels of high income households are typically more varied than of low income households. Similarly, saving rates of richer individuals typically show more variation than those of poorer individuals.
- In cross-sectional data on (hourly) wages, higher educated or more experienced individuals often earn wages that are more spread out compared to wages of persons with less education or experience (see illustration in next graph).

Example of heteroscedasticity



Another example

Suppose that in model (1) the outcome variable is binary (it takes only two values, say 0 and 1).

Since y is a 0-1 binary variable, we have $E[y|x] = \Pr(y = 1|x)$. Therefore:

$$p(x) \equiv \Pr(y = 1|x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

The model is therefore called the linear probability model. The parameter β_j measures how the probability of success changes when x_j changes

$$\Delta \Pr(y = 1|x) = \beta_j \Delta x_j$$

The model is inherently heteroscedastic since

$$Var(y|x) = p(x)(1 - p(x))$$

Consequences of heteroscedasticity

Let us now study what happens with OLS estimation when the homoscedasticity assumption is dropped.

For the moment we assume that the heteroscedasticity is of a very general form:

$$Var(u_i | x_{i1}, ..., x_{ik}) = \sigma_i^2. \quad (2)$$

Recall from previous lectures that unbiasedness and consistency of OLS estimators (of parameters in model (1)) hold under MLR.1-4. The homoscedasticity assumption MLR.5 is not used in the proofs.

Put in other words, OLS estimators are unbiased and consistent even under (2).

However, the usual expressions for the variance of OLS estimators are no longer valid (since they rely on MLR.5).

Consequences of heteroscedasticity (cntd)

This can be easily seen in the case $k = 1$. The OLS slope estimator can be written as (see Lecture 2):

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_{i1} - \overline{x}_1)u_i}{\sum_{i=1}^{n}(x_{i1} - \overline{x}_1)^2}.$$

Using (2), it follows that the conditional variance of the estimator is

$$Var(\hat{\beta}_1 | x_{11}, x_{21}, ..., x_{n1}) = \frac{\sum_{i=1}^{n}(x_{i1} - \overline{x}_1)^2 \sigma_i^2}{[\sum_{i=1}^{n}(x_{i1} - \overline{x}_1)^2]^2}. \tag{3}$$

This variance is generally not equal to the variance of $\hat{\beta}_1$ one gets when the heteroscedasticity problem is ignored (Lecture 2):

$$\frac{\sigma^2}{\sum_{i=1}^{n}(x_{i1} - \overline{x}_1)^2}$$

except in the case where $\sigma_i^2 = \sigma^2$ for all i.

Consequences of heteroscedasticity (cntd)

What this specific example shows is generally true: the variance formulas of OLS estimators that are calculated under the homoscedasticity assumption are not correct.

This implies that the estimated variances and the standard errors are generally also not valid.

This in turn implies that the usual t and F tests cannot be used. The usual confidence intervals are also no longer valid.

Another consequence of heteroscedasticity is that the OLS estimator is no longer BLUE, and no longer asymptotically efficient (in the class of estimators considered in Lecture 5).

Estimators with smaller (asymptotic) variance will be studied later on. First we turn to the issue of testing for the presence of heteroscedasticity.

Testing for heteroscedasticity

Since we are still assuming that u has zero conditional mean (MLR. 3), we have $Var(u|x_1, ..., x_k) = E(u^2|x_1, ..., x_k)$.

This suggests that testing for heteroscedasticity amounts to testing whether the conditional expectation of the squared error term depends in some way on the regressors.

When the homoscedasticity assumption is false, this conditional expectation can be any function of the regressors.

Suppose that it is a linear function of the regressors:

$$E(u^2|x_1, ..., x_k) = \delta_0 + \delta_1 x_1 + ... + \delta_k x_k.$$

This implies

$$u^2 = \delta_0 + \delta_1 x_1 + ... + \delta_k x_k + \nu$$

where ν is an error term with conditional mean zero:

$$E(\nu|x_1, ..., x_k) = 0$$

Testing for heteroscedasticity (cntd)

Under the null hypothesis of homoscedasticity, all parameters associated with the regressors equal zero:

$$H_0 : \delta_1 = 0, \delta_2 = 0, ..., \delta_k = 0. \tag{4}$$

If the u^2 were observed in the sample, we could regress u^2 on a constant and all regressors and test (4) using a F test.

The problem is that the error terms u_i are not observed. A natural idea is to replace them by the residuals $\hat{u}_i$ obtained after OLS regression of model (1), and then estimate the equation

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + ... + \delta_k x_k + \xi \tag{5}$$

and compute the F statistic for the joint significance of all regressors. Denoting the R-squared for the above model by $R^2_{\hat{u}^2}$, the F statistic is (see Lecture 4)

$$F = \frac{R^2_{\hat{u}^2}/k}{(1 - R^2_{\hat{u}^2})/(n-k-1)}. \tag{6}$$

Testing for heteroscedasticity (cntd)

Under the null hypothesis of homoscedasticity, this statistic approximately has a $F_{k,n-k-1}$ distribution in large samples. This result can be used to calculate the appropriate critical value, and then test (4) in the usual way.

The test just described is called the Breusch-Pagan test for heteroscedasticity of a particular form.

Remarks:

- The so called White test for heteroscedasticity is very similar. It consists in regressing $\hat{u}^2$ not only on the regressors x_j (as in model (5)), but also on the squared regressors x_j^2 and all cross products of the regressors ($x_j x_l$ for all $j \neq l$). The White test amounts to testing that all coefficients (except the intercept) are zero using the F test.
- The surprising result is that under the null the F statistic approximately has a $F_{k,n-k-1}$ distribution. Thus the large sample distribution of the F statistic is not affected by the fact that unobserved error terms are replaced by residuals!

Testing for heteroscedasticity (cntd)

To summarize the Breusch-Pagan test, here are its different steps:

1. Estimate model (1) by OLS, and obtain the residuals $\hat{u}_i$.
2. Estimate the regression of $\hat{u}_i^2$ on a constant and $x_{i1}, ..., x_{ik}$, and compute the R-squared of this regression: $R^2_{\hat{u}^2}$.
3. Calculate the F statistic (6) and find the α% critical value of the $F_{k,n-k-1}$ distribution. Reject the null if the F statistic is larger than the critical value, accept otherwise.

Let us reconsider the crime/enrollment example to illustrate the Breusch-Pagan test.

The results of the regression of $log(crime)$ on a constant and $log(enroll)$ were reported in Lecture 4. From this regression we determine the residuals

$$\begin{aligned}\hat{u}_i &= log(crime)_i - \hat{\beta}_0 - \hat{\beta}_1 log(enroll)_i \\ &= log(crime)_i + 6.63 - 1.27 log(enroll)_i.\end{aligned}$$

Testing for heteroscedasticity (cntd)

The R-squared from the regression of $\hat{u}^2$ on a constant and $log(enroll)$, $R^2_{\hat{u}^2}$, equals 0.06.

The F statistic (6) equals $(0.06/1)/(0.94/95) = 6.06$. The 5% critical value of the $F_{1,95}$ being approximately equal to 3.95, we reject the null hypothesis of homoscedasticity at the 5% level.

Generalized least squares

Now that we have described how (a certain form of) heteroscedasticity can be detected, we turn to estimation methods that explicitly take heteroscedasticity into account.

As mentioned above, when the error terms are not homoscedastic, the OLS estimator is no longer BLUE or asymptotically efficient.

The objective is now to discuss modified OLS methods that have these optimal properties in the presence of heteroscedasticity.

First we study the case where the heteroscedasticity is *known up to a multiplicative constant*:

$$Var(u_i|x_{i1}, ..., x_{ik}) = \sigma^2 h(x_{i1}, ..., x_{ik})$$

where h is some *known function* of the regressors. Since a variance is by definition positive, we must have $h(x_{i1}, ..., x_{ik}) > 0$ for all possible values of the regressors.

Generalized least squares (cntd)

In the crime/enrollment model, for example, the assumption might be that $Var(u_i|log(enroll)_i) = \sigma^2(log(enroll)_i)^2$. This corresponds to the case $h(x_{i1}) = x_{i1}^2$.

Let us write $h_i \equiv h(x_{i1}, ..., x_{ik})$.

The idea is now to transform the original model (1) such that the transformed model has homoscedastic error terms.

Since $Var(u_i|x_{i1}, ..., x_{ik}) = \sigma^2 h_i$, we have

$$Var(u_i/\sqrt{h_i}|x_{i1}, ..., x_{ik}) = \sigma^2$$

Now divide all variables appearing in (1) by $\sqrt{h_i}$ to get the transformed model:

$$y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + ... + \beta_k x_{ik}^* + u_i^*, \;\; i = 1, ..., n. \quad (7)$$

where $y_i^* = y_i/\sqrt{h_i}$, $x_{i0}^* = 1/\sqrt{h_i}$, $x_{ij}^* = x_{ij}/\sqrt{h_i}$ for $j > 1$, and $u_i^* = u_i/\sqrt{h_i}$.

Generalized least squares (cntd)

Note that there is *no constant* in the transformed model. Note also that the parameters appearing in the transformed model are the same as those in the initial untransformed model. Thus the parameters in the transformed model *have the same interpretation* as in the initial model.

The error term in model (7) is homoscedastic. This means that if the original model (1) satisfies MLR.1-4, then model (7) satisfies MLR.1-5.

Therefore, estimate the parameters in model (7) by OLS: β_0^*, β_1^*,..., β_k^*, are called the *generalized least squares (GLS) estimators* of β_0, β_1,..., β_k.

Remarks:

- Since model (7) satisfies MLR.1-5, the GLS estimator is BLUE. The conditional variance of β_j^* is thus smaller than the conditional variance of the OLS estimator $\hat{\beta}_j$ (the estimator obtained from the untransformed equation (1)).

Generalized least squares (cntd)

- If it is assumed that u_i follows a normal distribution $N(0, \sigma^2 h_i)$, then u_i^* follows a $N(0, \sigma^2)$. Model (7) then satisfies MLR.1-6. Consequently, the usual t and F tests can be applied using the GLS method.

The GLS estimators correcting for heteroscedasticity are also called *weighted least squares (WLS)* estimators.

This is because the GLS estimates β_j^* minimize the weighted sum of squared residuals, where each contribution i is weighted by $1/h_i$.

More precisely, the estimates β_0^*, β_1^*,..., β_k^* minimize

$$\sum_{i=1}^{n} (y_i - \beta_0^* - \beta_1^* x_{i1} - ... - \beta_k^* x_{ik})^2 / h_i.$$

WLS example

Consider the regression model

$$y_{ig} = \beta_0 + \beta_1 x_{ig1} + \ldots + \beta_k x_{igk} + u_{ig}$$

where y_{ig} is the outcome variable of observation i in group g (i is for example an employee in firm g, or i is a city in region g), x_{igj} the value of the j-th regressor for the i-th observation in group g, and u_{ig} the error term. Suppose that u_{ig} is homoscedastic.

Suppose also that the data we observe are averages at a higher level of aggregation (firm, region):

$$\overline{y}_g = \beta_0 + \beta_1 \overline{x}_{g1} + \ldots + \beta_k \overline{x}_{gk} + \overline{u}_g, \qquad g = 1, \ldots, G \tag{8}$$

where G is the number of groups, and $\overline{y}_g$, $\overline{x}_{gj}$ and $\overline{u}_g$ are averages in group g of the outcome variable, the j-th regressor, and the error term, respectively. Then $Var(\overline{u}_g) = Var(n_g^{-1} \sum_{i=1}^{n_g} u_i) = \sigma^2 / n_g$ where n_g is the number of observations in group g.

Model (8) can be estimated by WLS, where each observation g is weighted by n_g.

Feasible generalized least squares

A drawback of GLS is that, in practice, one rarely knows the exact form of the conditional variance $Var(u_i|x_{i1},...,x_{ik})$, and hence the method is not directly applicable in most applications.

A more common approach is to *build a flexible model* for the heteroscedasticity function, then *estimate* the model using the data, and implement the GLS method.

A popular heteroscedasticity model is the following model:

$$\begin{aligned} Var(u_i|x_{i1},...,x_{ik}) &= E(u_i^2|x_{i1},...,x_{ik}) \\ &= \sigma^2 exp(\delta_0 + \delta_1 x_{i1} + ... + \delta_k x_{ik}) \end{aligned} \tag{9}$$

where δ_0, δ_1,...,δ_k are unknown parameters.

We have $h_i = h(x_{i1},...,x_{ik}) = exp(\delta_0 + \delta_1 x_{i1} + ... + \delta_k x_{ik})$.

The model can be rewritten as

$$u_i^2 = \sigma^2 exp(\delta_0 + \delta_1 x_{i1} + ... + \delta_k x_{ik})\nu_i$$

with $E(\nu_i|x_{i1},...,x_{ik}) = 1$.

Feasible generalized least squares (cntd)

Assume now in addition that ν_i is independent of the regressors. It follows then that

$$log(u_i^2) = \alpha_0 + \delta_1 x_{i1} + ... + \delta_k x_{ik} + \xi_i$$

where

$$\alpha_0 = E(log(\nu_i)) + log(\sigma^2) + \delta_0$$

and

$$\xi_i = log(\nu_i) - E(log(\nu_i))$$

The error term ξ_i also has mean zero and is independent of the regressors.

The idea is now to run the regression of $log(\hat{u}^2)$ on a constant and $x_1, ..., x_k$, obtain the OLS estimates $\hat{\alpha}_0$, $\hat{\delta}_1$,..., $\hat{\delta}_k$, and estimate each h_i by

Feasible generalized least squares (cntd)

$$\hat{h}_i = exp(\hat{\alpha}_0 + \hat{\delta}_1 x_{i1} + ... + \hat{\delta}_k x_{ik}).$$

Next estimate the parameters of interest β_0,..., β_k, by applying OLS to the transformed model

$$y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + ... + \beta_k x_{ik}^* + u_i^*, \ \ i = 1, ..., n$$

where $y_i^* = y_i/\sqrt{\hat{h}_i}$, $x_{i0}^* = 1/\sqrt{\hat{h}_i}$, $x_{ij}^* = x_{ij}/\sqrt{\hat{h}_i}$ for $j > 1$, and $u_i^* = u_i/\sqrt{\hat{h}_i}$.

The resulting estimators, β_j^* for all $j > 0$, are the *feasible GLS (FGLS)* estimators.

The FGLS estimator β_j^* is not an unbiased estimator of β_j (this comes from the fact that h_i is not known but must be estimated), but it is consistent.

The FGLS estimator is asymptotically more efficient than the OLS estimator (applied to the untransformed model (1)).

Feasible generalized least squares (cntd)

Asymptotically, the t and F statistics associated with the FGLS estimates have t and F distributions. This implies that (at least in large samples), hypotheses testing can be done as in the standard OLS setting.

A drawback of the FGLS method is that it relies on the assumption that the *actual* heteroscedasticity in the data is indeed of the form (9).

If the true form of the heteroscedasticity differs from specification (9), the variance of the FGLS estimator is incorrect. Hence the standard errors of the FGLS estimators are not valid (and thus the t and F tests would be misleading).

Robust standard errors

A popular approach nowadays is simply to estimate the untransformed model (1) by OLS, and report the so-called *heteroscedasticity-robust standard errors*.

These are correct standard errors (at least in large samples) under any form of the heteroscedasticity. Regression packages typically allow you to report robust standard errors.

Let us first consider the case of a single regressor (i.e., $k = 1$).

It is assumed that the conditional variance of u_i has the very general form (2) (except that the conditioning is only on x_{i1}).

The conditional variance of the OLS estimator $\hat{\beta}_1$ is given in (3).

The idea is to estimate this variance by replacing each σ_i by the residual $\hat{u}_i$:

$$\widehat{Var}(\hat{\beta}_1 | x_{11}, x_{21}, ..., x_{n1}) = \frac{\sum_{i=1}^{n} (x_{i1} - \overline{x}_1)^2 \hat{u}_i^2}{[\sum_{i=1}^{n} (x_{i1} - \overline{x}_1)^2]^2}. \tag{10}$$

Robust standard errors (cntd)

This is a sensible estimator of $Var(\hat{\beta}_1|x_1, x_2, ..., x_n)$ in the sense that the probability limit of n times (3) equals the probability limit of n times (10).

It can be shown that they both converge to

$$E[(x_{i1} - \mu_{x_1})^2 u_i^2]/(\sigma_{x_1}^2)^2$$

where μ_{x_1} is the mean of x_{i1} and $\sigma_{x_1}^2$ its variance.

The heteroscedasticity-robust standard error of $\hat{\beta}_1$ is simply the square root of (10).

Next consider the general case $k > 1$. Adopting a similar proof as in Lecture 3, it can be verified that under (2), the conditional variance of $\hat{\beta}_j$ is (to simplify notations we do not indicate that we are conditioning on all values of all regressors):

$$Var(\hat{\beta}_j) = \frac{\sum_{i=1}^n \hat{r}_{ij}^2 \sigma_i^2}{\left[(1 - R_j^2) \sum_{i=1}^n (x_{ij} - \overline{x}_j)^2\right]^2}.$$

Robust standard errors (cntd)

For the same reason as above, a valid estimator of this variance is

$$\widehat{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^n \hat{r}_{ij}^2 \hat{u}_i^2}{\left[(1 - R_j^2) \sum_{i=1}^n (x_{ij} - \overline{x}_j)^2\right]^2}.$$

The heteroscedasticity-robust standard error of $\hat{\beta}_j$ is the square root of the above expression.

The *heteroscedasticity-robust t statistic* is the usual t statistic where the usual standard error is replaced with the heteroscedasticity-robust standard error.

It can be shown that asymptotically the heteroscedasticity-robust t statistic follows a t distribution.

Let us illustrate with the crime/enrollment data.

Below is the estimated crime equation with the heteroscedasticity-robust standard errors in brackets and, for comparison, the usual standard errors in parentheses:

Robust standard errors (cntd)

$$\widehat{log(crime)} = \underset{(1.03)[1.39]}{-6.63} + \underset{(0.11)[0.14]}{1.27} log(enroll).$$

$$n = 97, R^2 = 0.585.$$

The robust standard errors are slightly larger than the usual standard errors. The variable $log(enroll)$ is still statistically different from one at the 10% level, but no longer at the 5% level.

It is also possible to obtain F statistics that are robust to heteroscedasticity of any arbitary form.

Most econometric software packages nowadays compute these statistics, which allow you to perform tests of hypotheses involving multiple restrictions that are valid under arbitrary heteroscedasticity.

I understand/can apply...

- Consequences of heteroscedasticity
- Detecting heteroscedasticity (Breusch-Pagan test, White test)
- Estimation in the presence of heteroscedasticity (WLS,FGLS)
- Heteroscedasticity robust standard errors