Introductory Econometrics

Based on the textbook by Wooldridge: *Introductory Econometrics: A Modern Approach*

Robert M. Kunst
robert.kunst@univie.ac.at

University of Vienna
and
Institute for Advanced Studies Vienna

December 11, 2012

Outline

Heteroskedasticity: the definition

Heteroskedasticity refers to the feature that the variance σ_i^2 of the regression errors u_i is not constant in i:

$$\mathrm{E}(u_i^2|x_{i1},\ldots,x_{ik}) = \sigma_i^2 \not\equiv \sigma^2$$

The variation of the unobserved influences depends on the observed explanatory variables. Example: Engel curves etc.

The consequences of heteroskedasticity

- OLS remains unbiased and consistent (good news);
- OLS becomes inefficient, as the conditions for the Gauss-Markov Theorem are no longer valid: there are better estimators (in large samples, inefficiency may not always be a problem);
- The standard errors are incorrect, inference based on t–statistics becomes misleading (bad news).

Three topics in heteroskedasticity

- Repair the incorrect standard errors, use inefficient least squares for estimation (White-Eicker method);
- Testing for heteroskedasticity (Breusch-Pagan test, White test);
- Efficient estimation with heteroskedasticity (weighted least squares, feasible generalized least squares).

The idea of White-Eicker in matrices

Presume $y = X\beta + u$ with $\mathrm{E}uu' = \Omega$, with Ω a diagonal matrix formed from $(\sigma_1^2, \ldots, \sigma_n^2)$. Then, the variance of the OLS estimate $\hat{\beta}$ is not $\sigma^2(X'X)^{-1}$ but [conditional on X where needed]

$$\mathrm{var}(\hat{\beta}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

Evaluation of this formula may need estimates for Ω, typically unavailable. It can be shown that replacing Ω by $\mathrm{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2)$ from OLS residuals $\hat{u}_i$ yields a consistent estimate of the variance:

$$\widehat{\mathrm{var}}(\hat{\beta})_r = (X'X)^{-1}X'\mathrm{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2)X(X'X)^{-1}.$$

We cannot estimate the big $n \times n$–matrix Ω but we can estimate the small $(k+1) \times (k+1)$–matrix.

White-Eicker in simple regression

In a simple regression $y_i = \beta_0 + \beta_1 x_i + u_i$, the OLS estimate for β_1 is

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

and hence

$$\mathrm{var}(\hat{\beta}_1) = \{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma_i^2\}/\mathrm{SST}_x^2,$$

which is estimated by

$$\widehat{\mathrm{var}}(\hat{\beta}_1)_r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \hat{u}_i^2}{\mathrm{SST}_x^2}.$$

White-Eicker in practice

The square roots of the estimates $\widehat{\mathrm{var}}(\hat{\beta}_j)$ are called *heteroskedasticity-robust standard errors*. The ratio

$$\frac{\hat{\beta}_j}{\sqrt{\widehat{\mathrm{var}}(\hat{\beta}_j)_r}}$$

is the *heteroskedasticity-robust t statistic*. Similarly, robust F–statistics (Wald statistics) can be computed.

The Breusch-Pagan test for heteroskedasticity

The way that σ_i^2 depends on the regressors $x_{i1}, \ldots, x_{ik}$ (the 'functional form') is unknown. Breusch & Pagan considered linear dependence

$$\sigma_i^2 = \delta_0 + \delta_1 x_{i1} + \ldots + \delta_k x_{ik},$$

and constructed a test based on the nR^2 approximation to the LM test. Such a test should have the homoskedastic null $\delta_1 = \ldots = \delta_k = 0$. The distribution of the statistic under the null should be χ^2 in large samples with k degrees of freedom.

Breusch-Pagan test: the construction

Like other LM tests, the Breusch-Pagan test can be implemented via an auxiliary regression and its R^2:

1. Estimate the main equation $y = X\beta + u$ by OLS. Save the OLS residuals $\hat{u}_i$;
2. Regress $\hat{u}_i^2, i = 1, \ldots, n$, on all original regressors (the auxiliary regression)

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \ldots + \delta_k x_{ik} + \nu_i,$$

   and keep the corresponding $R^2_{\hat{u}^2}$;
3. The statistic $nR^2_{\hat{u}^2}$ is the approximate LM test statistic. It is $\chi^2(k)$ distributed under the null in large samples. There is also an approximate F version available. For large values, reject homoskedasticity.

The White test for heteroskedasticity

Presume that σ_i^2 also depends on squares of explanatory variables and on cross-terms. This is a quite general form that may pick up heteroskedasticity that is missed by Breusch-Pagan:

$$\begin{aligned}\hat{u}^2 &= \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + \delta_{11} x_1^2 + \ldots + \delta_{kk} x_k^2 \\ &\quad + \delta_{12} x_1 x_2 + \ldots + \delta_{k-1,k} x_{k-1} x_k + \mathrm{error}\end{aligned}$$

A drawback is that there are $k(k+3)/2$ regressors (excluding the constant), which may be too many with large k and small n. An alternative suggestion is to use

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

with $\hat{y}$ the fitted values from the main original regression.

White test: the construction

Again, the White test is implemented in the nR^2 form of the LM tests:

1. Estimate the main equation $y = X\beta + u$ by OLS. Save the OLS residuals $\hat{u}_i$;
2. Regress $\hat{u}_i^2, i = 1, \ldots, n$, on all original regressors and on their squares and cross-products (auxiliary regression). Keep the R^2 from this regression;
3. Compute $nR^2_{\hat{u}^2}$, which is asymptotically $\chi^2(\frac{k(k+3)}{2})$ distributed under the null. Reject homoskedasticity for large values of the test statistic.

With the alternative version $\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + \mathrm{error}$, the asymptotic null distribution of the test statistic is $\chi^2(2)$.

Suppose σ_i^2 are known

If the error variances σ_i^2 are known, it is fairly easy to construct an efficient estimator. Just divide the regression equation by σ_i:

$$\frac{y_i}{\sigma_i} = \frac{\beta_0}{\sigma_i} + \beta_1 \frac{x_{i,1}}{\sigma_i} + \ldots + \beta_k \frac{x_{i,k}}{\sigma_i} + \frac{u_i}{\sigma_i},$$

a homogeneous regression with homoskedastic errors. Gauss-Markov assumptions are fulfilled, and least squares is efficient. If $\sigma_i^2, i = 1, \ldots, n$, are known up to a constant factor, the same technique can be applied. If only the functional form of their dependence on the regressors is known, σ_i^2 can be estimated.

Weighted least squares in matrices

Presume the regression $y = X\beta + u$ has $\mathrm{E}uu' = \Omega$ with $\Omega = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, a diagonal matrix. Then, the regression

$$\tilde{y} = \tilde{X}\beta + \tilde{u}, \quad \tilde{y} = \Omega^{-1/2}y, \tilde{X} = \Omega^{-1/2}X, \tilde{u} = \Omega^{-1/2}u,$$

fulfils the conditions of Gauss-Markov, and OLS is efficient. In original variables, this OLS is written as

$$\hat{\beta}_w = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

It is called *weighted least squares* (WLS), as observations with strong variation in their errors now obtain smaller weights, observations with low variation obtain larger weights. It is also called *generalized least squares* (GLS), as it is an example for a more general technique that can always be used when Ω is known but not a scalar matrix, not only for heteroskedasticity with its diagonal Ω.

Standard errors for WLS

WLS is obtained by transforming the original regression model, re-scaling observations, and then applying OLS. Hence, the variance of the WLS estimate is obvious:

$$\mathrm{var}(\hat{\beta}_w) = (X'\Omega^{-1}X)^{-1}$$

The square roots of the diagonal elements are standard errors and can be used for t–statistics. Heuristically, observations with large variation in their errors contribute strongly to these standard errors. It can be shown that this variance is smaller than the variance of the OLS estimate

$$\mathrm{var}(\hat{\beta}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

Feasible GLS estimator

The variances σ_i^2 are usually unknown. So, they must be estimated. Functional forms that may yield negative variances must be discarded, such as any forms with linear terms. Consider the form

$$\mathrm{var}(u|x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k).$$

An obvious suggestion is to run a regression on OLS residuals $\hat{u}$,

$$\log \hat{u}_i^2 = \delta_0 + \delta_1 x_{i,1} + \ldots + \delta_k x_{i,k} + \mathrm{error},$$

and to plug in the corresponding fitted values for σ_i^2. This is an example for a *feasible GLS* (**fGLS**) method. These methods can actually be very unreliable in small samples, but as $n \to \infty$, they can be shown to achieve the efficiency of the (infeasible and efficient) GLS.

A feasible GLS estimator in heteroskedastic models

1. Regress y on $x_1, \dots, x_k$ using OLS. Keep residuals $\hat{u}$;
2. Create $\log(\hat{u}^2)$ from the OLS residuals of the main regression;
3. Regress $\log(\hat{u}^2)$ on $x_1, \dots, x_k$ using OLS. Keep fitted values $\hat{g}$;
4. Exponentiate the fitted values : $\hat{h} = \exp(\hat{g})$ to obtain variance estimates;
5. Regress y on $x_1, \dots, x_k$ using WLS with weights $\hat{h}_i^{-1/2}$, or transform all variables via division by $\sqrt{\hat{h}_i}$ and regress by OLS.

What if the specific functional form for heteroskedasticity is wrong?

It is preposterous to assume that there is an exact match of the entertained functional form and the true one, even when such a form exists. Nevertheless, as long as the technique captures the variation in dispersion across the sample well, fGLS will outperform OLS in larger samples.

The same remark holds for heteroskedasticity tests. As long as the functional form is reasonable, it will pick up existing heteroskedasticity, even when it is incorrect.

The logarithmic form is convenient for fGLS. However, it is not recommended for the construction of tests in analogy to White and Breusch/Pagan (Park test).