Kurt Schmidheiny
Unversity of Basel

Heteroskedasticity in the Linear Model

matrix-free

1 Introduction

This handout extends the handout on "The Multiple Linear Regression model" and refers to its definitions and assumptions in section 2.

This handouts relaxes the homoscedasticity assumption (*OLS4a*) and shows how the parameters of the linear model are correctly estimated and tested when the error terms are heteroscedastic (*OLS4b*).

2 The Econometric Model

Consider the multiple linear regression model for observation $i = 1, ..., N$

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + u_i$$

where x_{i0}, x_{i1}, ..., x_{iK} are K explanatory variables and a constant, $\beta_0, \beta_1, ..., \beta_K$ are $K + 1$ parameters and u_i is called the error term.

Assume *OLS1*, *OLS2*, *OLS3* and

OLS4: Error Variance

b) conditional heteroscedasticity:
$V[u_i|x_{i1}, ..., x_{iK}] = \sigma_i^2 = \sigma^2\omega_i = \sigma^2\omega(x_{i1}...x_{iK})$
and $(\sqrt{\omega_i}, \sqrt{\omega_i}\, x_{i1}, \cdots, \sqrt{\omega_i}\, x_{iK})$ are not linearly dependent
and $E[\omega_i x_{ik}^2] < \infty$ for all $k > 0$

where $\omega(.)$ is a function constant across i. The decomposition of σ_i^2 into ω_i and σ^2 is arbitrary but useful.

Note that under *OLS2* (i.i.d. sample) the errors are *unconditionally* homoscedastic, $V[u_i] = \sigma^2$ but allowed to be *conditionally* heteroscedastic, $V[u_i|x_i] = \sigma_i^2$. Assuming *OLS2* and *OLS3c* provides that the errors are also *not conditionally autocorrelated*, i.e. $\forall i \neq j : Cov[u_i, u_j|x_{i1}...x_{iK}, x_{j1}...x_{jK}] = 0$. Also note that the conditioning on x_i is less restrictive than it may seem: if the conditional variance $V[u_i|x_i]$ depends on other exogenous variables (or functions of them), we can include these variables in x_i and set the corresponding β parameters to zero.

3 A Generic Case: Groupewise Heteroskedasticity

Heteroskedasticity is sometimes a direct consequence of the construction of the data. Consider the following linear regression model with homoscedastic errors

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + u_i$$

with $V[u_i|x_i] = V[u_i] = \sigma^2$.

Assume that instead of the individual observations y_i and x_i only the mean values y_g and x_g for $g = 1, ..., G$ groups are observed. The error term in the resulting regression model

$$y_g = \beta_0 + \beta_1 x_{g1} + \cdots + \beta_K x_{gK} + u_g$$

is now conditionally heteroskedastic with $V[u_g|N_g] = \sigma_g^2 = \sigma^2/N_g$, where N_g is a random variable with the number of observations in group g.

4 Estimation with OLS

The parameters $\beta_0, \beta_1, ..., \beta_K$ can be estimated with the usual OLS estimator.

For the bivariate regression model, they are calculated as

$$\widehat{\beta}_1^{OLS} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

$$\widehat{\beta}_0^{OLS} = \overline{y} - \hat{\beta}_1 \overline{x}$$

The OLS estimator of β remains unbiased under *OLS1*, *OLS2*, *OLS3c*, *OLS4b*, and *OLS5* in small samples. Additionally assuming *OLS3a*, it is normally distributed in small samples. It is consistent and approximately normally distributed under *OLS1*, *OLS2*, *OLS3d*, *OLS4a* or *OLS4b* and *OLS5*, in large samples. However, the OLS estimator is not efficient any more. More importantly, the usual standard errors of the OLS estimator and tests (t-, F-, z-, Wald-) based on them are not valid any more.

5 Estimating the Variance of the OLS Estimator

The small sample variance $V[\widehat{\beta}_k^{OLS}|x_{11}, ..., x_{NK}]$ of $\widehat{\beta}_k^{OLS}$ differs from the usual OLS one under *OLS4b*. For the bivariate regression model, it is

$$V[\widehat{\beta}_1|x_{11}, ..., x_{NK}] = \frac{\sum_{i=1}^{N}\sigma^2\omega_i(x_i - \overline{x})^2}{[\sum_{i=1}^{N}(x_i - \overline{x})^2]^2} = \frac{\sum_{i=1}^{N}\sigma^2 \cdot \omega(x_i) \cdot (x_i - \overline{x})^2}{[\sum_{i=1}^{N}(x_i - \overline{x})^2]^2}$$

which differs from the usual OLS estimator

$$V[\widehat{\beta}_1|x_{11}, ..., x_{NK}] = \frac{\sigma^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

Consequently, the usual estimator $\widehat{V}[\widehat{\beta}_k^{OLS}|x_{11}, ..., x_{NK}]$ is biased. Usual small sample test procedures, such as the F- or t-Test, based on the usual estimator are therefore not valid.

The OLS estimator is asymptotically normally distributed under *OLS1*, *OLS2*, *OLS3d*, and *OLS5*

$$\sqrt{N}(\widehat{\beta}_k - \beta_k) \xrightarrow{d} N\left(0, \varsigma^2\right)$$

where for the bivariate regression model, $\varsigma^2 = E[u_i^2(x_i - Ex_i)^2)]/[\sigma_x^2]^2$ and $\sigma_x^2 = V[x_i] = E[x_i - Ex_i]^2$. The OLS estimator is therefore approximately normally distributed in large samples as

$$\widehat{\beta}_k \overset{A}{\sim} N\left(\beta_k, Avar[\widehat{\beta}_k]\right)$$

where $Avar[\widehat{\beta}_k] = \varsigma^2/N$ can be consistently estimated with some additional assumptions on higher order moments of x_i (see White 1980). For the bivariate regression model, the White estimator is

$$\widehat{Avar}[\widehat{\beta}_1] = \frac{\sum_{i=1}^N \widehat{u}_i^2 (x_i - \overline{x})^2}{[\sum_{i=1}^N (x_i - \overline{x})^2]^2}$$

This so-called White or Eicker-Huber-White estimator of the covariance matrix is a *heteroskedasticity-consistent covariance matrix estimator* that does not require any assumptions on the form of heteroscedasticity (though we assumed independence of the error terms in *OLS2*). Standard errors based on the White estimator are often called *robust*. We can perform the usual z- and Wald-test for large samples using the White covariance estimator.

Note: t- and F-Tests using the White covariance estimator are only asymptotically valid because the White covariance estimator is consistent but not unbiased. It is therefore more appropriate to use large sample tests (z, Wald).

Bootstrapping (see the handout on "The Bootstrap") is an alternative method to estimate a heteroscedasticity robust covariance matrix.

6 Testing for Heteroskedasticity

There are several tests for the assumption that the error term is homoskedastic. White (1980)'s test is general and does not presume a particular form of heteroskedasticity. Unfortunately, little can be said about its power and it has poor small sample properties unless the number of regressors is very small. If we have prior knowledge that the variance σ_i^2 is a linear (in parameters) function of explanatory variables, the Breusch-Pagan (1979) test is more powerful. Koenker (1981) proposes a variant of the Breusch-Pagan test that does not assume normally distributed errors.

Note: In practice we often do not test for heteroskedasticity but directly report heteroskedasticity-robust standard errors.

7 Estimation with GLS/WLS when ω_i is Known

When ω_i is known, β is efficiently estimated with generalized least squares (GLS). The GLS estimator $\widehat{\beta}^{GLS}$ simplifies in the case of heteroskedasticity to the weighted least squares (WLS) estimator $\widehat{\beta}^{WLS}$ which is calculated as an OLS regression of a transformed dependent variable $\tilde{y}$ on transformed explanatory variables $\tilde{x}_{i0}, \tilde{x}_{i1}, ..., \tilde{x}_{iK}$ where

$$\tilde{y}_i = y_i/\sqrt{\omega_i} \quad \text{and} \quad \tilde{x}_{ik} = x_{ik}/\sqrt{\omega_i}$$

Note: the above transformation of the explanatory variables also applies to the constant, i.e. $\tilde{x}_{i0} = 1/\sqrt{\omega_i}$. The OLS regression using the transformed variables does not include an additional constant.

The WLS estimator minimizes the sum of squared residuals weighted by $1/\omega_i$:

$$S(\beta_0, ..., \beta_K) = \sum_{i=1}^{N} \frac{1}{\omega_i} [y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK})]^2 \rightarrow \min_{\beta_0,...,\beta_K}$$

The WLS estimator of β is unbiased and efficient (under *OLS1*, *OLS2*, *OLS3c*, *OLS4b*, and *OLS5*) and normally distributed additionally assuming *OLS3a* (normality) in small samples.

The WLS estimator of β is consistent, asymptotically efficient and approximately normally distributed under *OLS4b* (conditional heteroscedasticity) and *OLS2* , *OLS1*, *OLS3d*, and a modification of *OLS5*

OLS5: Identifiability

$(\tilde{x}_{i0}, \tilde{x}_{i1}, \cdots, \tilde{x}_{iK})$ are not linearly dependent and
$0 < V[\tilde{x}_{ik}] < \infty$ and $0 < \widehat{V}[\tilde{x}_{ik}] < \infty$ for all $k > 0$

The variance of $\widehat{\beta}_K^{WLS}$ is estimated as the usual OLS estimator in the transformed variables $\tilde{y}$ and $\tilde{x}_k$. For the bivariate regression model, it is

$$\widehat{V}[\widehat{\beta}_1^{WLS}|x_{11}, ..., x_{NK}] = \frac{\widehat{\sigma^2}}{\sum_{i=1}^{N}(\tilde{x}_i - \overline{\tilde{x}})^2}$$

where in small samples

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{N} \widehat{\tilde{u}}^2}{N - K - 1} = \frac{\sum_{i=1}^{N} \widehat{\tilde{u}}^2}{N - 2}$$

and in large samples

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{N} \widehat{\tilde{u}}^2}{N}$$

and $\widehat{\tilde{u}}_i = \tilde{y}_i - (\widehat{\beta}_0^{GLS} + \widehat{\beta}_1^{GLS}\tilde{x}_{i1} + ... + \widehat{\beta}_K^{GLS}\tilde{x}_{iK})$. Usual tests ($t$-, F-) for small samples are valid (under *OLS1*, *OLS2*, *OLS3a*, *OLS4b* and *OLS5*; usual tests (z, Wald) for large samples are also valid (under *OLS1*, *OLS2*, *OLS3d*, *OLS4b* and *OLS5*).

8 Estimation with GLS/WLS when ω_i is Unknown

In practice, ω_i is typically unknown. However, we can model the ω_i's as a function of the data and estimate this relationship. Feasible generalized least squares (FGLS) replaces ω_i by their predicted values $\widehat{\omega}_i$ and calculates then $\widehat{\beta}_{FGLS}$ as if ω_i were known.

A useful model for the error variance is

$$\sigma_i^2 = V[u_i|z_{i1}, ..., z_{iL}] = e^{\delta_0 + \delta_1 z_{i1} + ... + \delta_L x_{iL}}$$

where $z_{i0}, ..., z_{iL}$ are $L+1$ variables that may belong to $x_{i0}, ..., x_{iK}$ including a constant and $\delta_0, ..., \delta_L$ are parameters. We can estimate the auxiliary regression

$$\widehat{u}_i^2 = e^{\delta_0 + \delta_1 z_{i1} + ... + \delta_L x_{iL}} + \nu_i$$

by nonlinear least squares (NLLS) where $\widehat{u}_i = y_i - (\widehat{\beta}_0^{OLS} + \widehat{\beta}_1^{OLS} x_{i1} + ... + \widehat{\beta}_K^{OLS} x_{iK})$ or alternatively,

$$\log(\widehat{u}_i^2) = \delta_0 + \delta_1 z_{i1} + ... + \delta_L x_{iL} + \nu_i$$

by ordinary least squares (OLS). In both cases, we use the predictions

$$\widehat{\omega}_i = e^{\widehat{\delta}_0 + \widehat{\delta}_1 z_{i1} + ... + \widehat{\delta}_L x_{iL}}$$

in the calculations for $\widehat{\beta}_{FGLS}$ and $\widehat{Avar}[\widehat{\beta}_{FGLS}]$.

The FGLS estimator is consistent and approximately normally distributed in large samples under *OLS1*, *OLS2* ($\{x_i, z_i, y_i\}$ i.i.d.), *OLS3d*, *OLS4b*, *OLS5* and some additional more technical assumptions. If σ_i^2 is correctly specified, β_{FGLS} is asymptotically efficient and the usual tests (z, Wald) for large samples are valid; small samples tests are only asymptotically valid and nothing is gained from using them. If σ_i^2 is not correctly specified, the usual covariance matrix is inconsistent and tests (z, Wald) invalid. In this case, the White covariance estimator used after FGLS provides consistent standard errors and valid large sample tests (z, Wald).

Note: In practice, we often choose a simple model for heteroscedasticity using only one or two regressors and use robust standard errors.

Implementation in Stata 17

Stata reports the White covariance estimator with the `robust` option, e.g.

```
webuse auto.dta
regress price mpg weight, vce(robust)
matrix list e(V)
```

Stata reports the same robust covariance correcting for degrees of freedom in small samples by multiplying the variance by $N/(N-K-1)$. For small sample sizes ($N < 250$), the following version of robust standard errors should be used (Long and Ervin, 2000):

```
regress price mpg weight, vce(hc3)
```

Alternatively, Stata estimates a heteroscedasticity robust covariance using a nonparametric bootstrap. For example,

```
regress price mpg weight, vce(bootstrap, rep(100))
matrix list e(V)
```

The White (1980) test for heteroskedasticity is implemented in the post*t*-estimation command

```
estat imtest, white
```

The Koenker (1981) version of the Breusch-Pagan (1979) test is implemented in the postestimation command `estat hettest`. For example,

```
estat hettest weight foreign, iid
```

assumes $\sigma_i^2 = \delta_0 + \delta_1 weight_i + \delta_2 foreign_i$ and tests $H_0 : \delta_1 = \delta_2 = 0$.

WLS is estimated in Stata using analytic weights. For example,

```
regress depvar indepvars [aweight = 1/w]
```

calculates the WLS estimator assuming ω_i is provided in the variable w. Recall that we defined $\sigma_i^2 = \sigma^2 \omega_i$ (mind the squares). The analytic weight is proportional to the inverse variance of the error term. Stata internally scales the weights s.t. $\sum 1/\omega_i = N$. The reported Root MSE therefore reports $\overline{\widehat{\sigma}}^2 = (1/N) \sum \widehat{\sigma}_i^2$.

Implementation in R 4.2.0

R reports the Eicker-Huber-White covariance after estimation

```
library(foreign)
auto <- read.dta("http://www.stata-press.com/data/r11/auto.dta")
ols <- lm(price ~ mpg + weight, data = auto)
```

using the two packages `sandwich` and `lmtest`

```
library(sandwich)
library(lmtest)
coeftest(ols, vcov = sandwich)
```

The following two commands are equivalent

```
coeftest(ols, vcov = vcovHC, type="HC0")
coeftest(ols, vcov = vcovHC(ols, type="HC0"))
```

The type `HC1` reports the same robust covariance as Stata correcting for degrees of freedom in small samples by multiplying the variance by $N/(N - K - 1)$.

```
coeftest(ols, vcov = vcovHC, type="HC1")
```

For small sample sizes ($N < 250$), the type `HC3` should be used (Long and Ervin, 2000).

```
coeftest(ols, vcov = vcovHC, type="HC3")
```

A heteroskedasticity robust F-test for $H_0 : \beta_1 = 0$ is called by

```
waldtest(ols, "weight", vcov = vcovHC(ols, type="HC3"))
```

and for $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ against $H_A : \beta_1 \neq 0$ or $\beta_2 \neq 0$ by

```
waldtest(ols, .~.-weight - displacement, vcov = vcovHC(ols, type="HC3"))
```

References

Introductory textbooks

Stock, James H. and Mark W. Watson (2020), Introduction to Econometrics, 4th Global ed., Pearson. Section 5.4.

Wooldridge, Jeffrey M. (2009), Introductory Econometrics: A Modern Approach, 4th ed., South-Western Cengage Learning. Chapter 8.

Advanced textbooks

Cameron, A. Colin and Pravin K. Trivedi (2005), Microeconometrics: Methods and Applications, Cambridge University Press. Sections 4.5.

Hayashi, Fumio (2000), Econometrics, Princeton University Press. Sections 1.6 and 2.8.

Wooldridge, Jeffrey M. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press. Sections 4.23 and 6.24.

Companion textbooks

Kennedy, Peter (2008), A Guide to Econometrics, 6th ed., Blackwell Publishing. Chapters 8.1-8.3.

Articles

Breusch, T.S. and A.R. Pagan (1979), A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* 47, 987–1007.

Koenker, R. (1981), A Note on Studentizing a Test for Heteroskedasticity. *Journal of Econometrics* 29, 305–326.

Long, J. S., and Ervin, L. H. (2000), Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3), 217-224.

White, H. (1980), A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica* 48, 817–838.