

# Heteroskedasticity in the Linear Model

## 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 (*OLS5a*) and shows how the parameters of the linear model are correctly estimated and tested when the error terms are heteroscedastic (*OLS5b*).

## 2 The Econometric Model

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

$$y_i = x_i' \beta + u_i$$

where  $y_i$  is the dependent variable,  $x_i'$  is a  $(K + 1)$ -dimensional row vector of  $K$  explanatory variables and a constant,  $\beta$  is a  $(K + 1)$ -dimensional column vector of parameters, and the scalar  $u_i$  is the error term.

Assume *OLS1*, *OLS2*, *OLS3c*, *OLS4*, *OLS6* and

*OLS5: Error Variance*

b)  $V(u_i|x_i) = \sigma_i^2 = g(x_i) < \infty$  (conditional heteroscedasticity)

Note that under *OLS2* (i.i.d. sample) the errors are *unconditionally* homoscedastic,  $V(u_i) = \sigma^2$  but allowed to be *conditionally* heteroscedastic,  $V(u_i|x_i) = \sigma_i^2$ . We still assume in *OLS2* that the error terms are *not autocorrelated*, i.e.  $\forall i \neq j : E(u_i u_j | x_i) = Cov(u_i, u_j | x_i) = 0$ .

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 variables (or other functions of variables) than  $y_i$ , we can include these variables in  $x_i$  and set the corresponding  $\beta$  parameters to zero.

The variance-covariance of the vector of error terms  $u = (u_1, u_2, \dots, u_N)'$  in the whole sample is therefore

$$\begin{aligned} V(u|X) &= E(uu'|X) = \sigma^2 \Omega \\ &= \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_N \end{pmatrix} \end{aligned}$$

where  $\sigma_i^2 = \sigma^2 \omega_i = \sigma^2 \omega(x_i)$  and  $\omega(\cdot)$  is a function which is constant across  $i$ . The decomposition of  $\sigma_i^2$  into  $\omega_i$  and  $\sigma^2$  is arbitrary but useful.

## 3 A Generic Case: Groupwise Heteroskedasticity

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

$$y_i = x_i' \beta + u_i \quad \text{with} \quad V(u_i|x_i) = V(u_i) = \sigma^2$$

Assume that instead of the individual observations  $y_i$  and  $x_i$  only the means  $\bar{y}_g$  and  $\bar{x}_g$  for  $g = 1, \dots, G$  groups are observed. The error term in the resulting regression model

$$\bar{y}_g = \bar{x}_g' \beta + \bar{u}_g$$

is now conditionally heteroskedastic with  $V(\bar{u}_g|N_g) = \sigma_g^2 = \sigma^2/N_g$ , where  $N_g$  is a random variable with the number of observations in group  $g$ .

## 4 Estimation with OLS

The parameter  $\beta$  can be estimated with OLS as

$$\widehat{\beta}_{OLS} = (X'X)^{-1} X'y$$

The OLS estimator of  $\beta$  remains unbiased under *OLS1*, *OLS2*, *OLS3c*, *OLS4*, *OLS5b* and *OLS6* in small samples. Additionally assuming *OLS3a*, it is normally distributed in small samples. It is consistent and approximately normally distributed under *OLS1*, *OLS2*, *OLS3d*, *OLS4*, *OLS5a* or *b* and *OLS6* 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 Covariance of the OLS Estimator

The small sample covariance matrix of  $\widehat{\beta}_{OLS}$  is under *OLS3c* and *OLS5b*

$$V(\widehat{\beta}_{OLS}|X) = (X'X)^{-1} [X'\sigma^2\Omega X] (X'X)^{-1}$$

and differs from usual OLS where  $V(\widehat{\beta}_{OLS}|X) = \sigma^2(X'X)^{-1}$ . Consequently, the usual estimator  $\widehat{V}(\widehat{\beta}_{OLS}|X) = \widehat{\sigma}^2(X'X)^{-1}$  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*, *OLS4*, and *OLS6*

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma)$$

where  $\Sigma = Q_{XX}^{-1} E(u_i^2 x_i x_i') Q_{XX}^{-1}$  and  $Q_{XX} = E(x_i x_i')$ . The OLS estimator is therefore approximately normally distributed in large samples as

$$\widehat{\beta} \overset{A}{\sim} N(\beta, Avar(\widehat{\beta}))$$

where  $Avar(\widehat{\beta}) = N^{-1}\Sigma$  can be consistently estimated as

$$\widehat{Avar}(\widehat{\beta}) = (X'X)^{-1} \left[ \sum_{i=1}^N \widehat{u}_i^2 x_i x_i' \right] (X'X)^{-1}$$

with some additional assumptions on higher order moments of  $x_i$  (see White 1980).

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 no unbiased. It is therefore more appropriate to use large sample tests (*z*, Wald).

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

## 6 Testing for Heteroskedasticity

There are several tests for the assumption that the residuals are 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  $\sigma_i^2$  is a linear (in parameters) function of explanatory variables, the Breusch-Pagan (1979) test is more powerful. Koenker (1981) proposes an 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 $\Omega$ is Known

When  $\Omega$  is known, the parameter  $\beta$  is efficiently estimated with generalized least squares (GLS)

$$\widehat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

The GLS estimator simplifies in the case of heteroskedasticity to

$$\widehat{\beta}_{WLS} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$$

where

$$\tilde{y} = \begin{pmatrix} y_1/\sqrt{\omega_1} \\ \vdots \\ y_T/\sqrt{\omega_T} \end{pmatrix} \quad \text{and} \quad \tilde{X} = \begin{pmatrix} x'_1/\sqrt{\omega_1} \\ \vdots \\ x'_N/\sqrt{\omega_N} \end{pmatrix}.$$

and is called weighted least squares (WLS) estimator. The WLS estimator of  $\beta$  can therefore be estimated by running OLS with the transformed variables.

Note: the above transformation of  $x'_i$  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 vector.

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

$$S(\alpha, \beta) = \sum_{i=1}^N (\tilde{y}_i - \tilde{x}'_i\beta)^2 = \sum_{i=1}^N \frac{1}{\omega_i} (y_i - x'_i\beta)^2 \rightarrow \min_{\beta}$$

The WLS estimator of  $\beta$  is unbiased and efficient (under *OLS1*, *OLS2*, *OLS3c*, *OLS4*, *OLS5b* and *OLS6*) and normally distributed additionally assuming *OLS3a* (normality) in small samples.

The WLS estimator of  $\beta$  is consistent, asymptotically efficient and approximately normally distributed under *OLS5b* (conditional heteroscedasticity) and *OLS2* (implying conditional homoscedasticity), *OLS1*, *OLS3d*, *OLS4*, and

*OLS6: moments of explanatory variables*

d)  $E(x_ix'_i) = Q$  and  $E(\omega_i^{-1}x_ix'_i) = S$  are both p.d. and finite

where  $\omega_i = \omega(x_i)$  is a known function of  $x_i$ .

The covariance matrix of  $\widehat{\beta}_{WLS}$  can be estimated as

$$\widehat{Avar}(\widehat{\beta}_{WLS}) = \widehat{V}(\widehat{\beta}_{WLS}) = \widehat{\sigma}^2(\tilde{X}'\tilde{X})^{-1}$$

where  $\widehat{\sigma}^2 = \widehat{u}'\widehat{u}/(N - K - 1)$  in small samples and  $\widehat{\sigma}^2 = \widehat{u}'\widehat{u}/N$  in large samples and  $\widehat{u} = \tilde{y} - \tilde{X}\widehat{\beta}_{GLS}$ . Usual tests (*t*-, *F*-) for small samples are valid (under *OLS1*, *OLS2*, *OLS3a*, *OLS4*, *OLS5b* and *OLS6*); usual tests (*z*-, Wald-) for large samples are also valid (under *OLS1*, *OLS2*, *OLS3d*, *OLS4*, *OLS5b* and *OLS6*).

## 8 Estimation with FGLS/FWLS when $\Omega$ is Unknown

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

A useful model for the error variance is

$$\sigma_i^2 = V(u_i|z_i) = \exp(z'_i\delta)$$

where  $z_i$  are  $L + 1$  variables that may belong to  $x_i$  including a constant and  $\delta$  is a vector of parameters. We can estimate the auxiliary regression

$$\widehat{u}_i^2 = \exp(z'_i\delta) + \nu_i$$

by nonlinear least squares (NLLS) where  $\hat{u}_i = y_i - x_i' \hat{\beta}_{OLS}$  or alternatively,

$$\log(\hat{u}_i^2) = z_i' \delta + \nu_i$$

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

$$\hat{\omega}_i = \exp(z_i' \hat{\delta})$$

in the calculations for  $\hat{\beta}_{FGLS}$  and  $\widehat{Avar}(\hat{\beta}_{FGLS})$ .

The FGLS estimator of  $\beta$  is unbiased under *OLS1*, *OLS2*, *OLS3c*, *OLS4*, *OLS5b* and *OLS6* and normally distributed (additionally assuming *OLS3a*) in small samples. 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*, *OLS4*, *OLS5b*, *OLS6* and some additional more technical assumptions. If  $\sigma_i^2$  is correctly specified,  $\beta_{FGLS}$  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  $\sigma_i^2$  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.

## 9 Implementation in Stata 10.0

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)
```

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-estimation` command

```
estat imtest, white
```

The original Breusch-Pagan (1979) test is implemented in the `post-estimation` command `estat hettest`. For example,

```
estat hettest weight foreign
```

assumes  $\sigma_i^2 = \delta_0 + \delta_1 \text{weight}_i + \delta_2 \text{foreign}_i$  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  $\omega_i$  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  $\hat{\sigma}^2 = (1/N) \sum \hat{\sigma}_i^2$ .

FGLS/FWLS is carried out step by step. First, we estimate the regression model  $y_i = x_i' \beta + u_i$  by OLS and predict the residuals  $\hat{u}_i = y_i - x_i' \hat{\beta}$ . Second, the residuals are used to estimate a regression model for the error variance, e.g.  $\log(\hat{u}_i^2) = z_i' \delta + \nu_i$ , and to predict the individual error variance, e.g.  $\hat{\omega}_i = \exp(z_i' \hat{\delta})$ . Third, we perform a linear regression using

the weights  $1/\hat{\omega}_i$ . For example,

```
regress price mpg weight
predict e, residual
generate loge2 = log(e^2)
regress loge2 weight foreign
predict zd
generate w=exp(zd)
regress price mpg weight [aweight = 1/w]
```

## References

### Introductory textbooks

- Stock, James H. and Mark W. Watson (2007), *Introduction to Econometrics*, 2nd ed., Pearson Addison-Wesley. 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.
- White, H. (1980), A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica* 48, 817-838.