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 handout 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 \(i = 1, ..., N\)

\[ y_i = x'_i\beta + u_i \]

where \(x'_i\) 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 \(u_i\) is a scalar called the error term.

Assume \(OLS1, OLS2, OLS3\) and

\(OLS4: Error\ Variance\)

b) conditional heteroscedasticity:

\[ V[u_i|x_i] = \sigma^2_i = \sigma^2 \omega_i = \sigma^2 \omega(x_i) \]

and \(E[u^2_i|x_i,x'_i] = Q_X\sigma_H X\) is p.d. and finite

where \(\omega(\cdot)\) is a function constant across \(i\). The decomposition of \(\sigma^2_i\) into \(\omega_i\) and \(\sigma^2\) 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, \(V[u_i|x_i] = \sigma^2_i\). Assuming \(OLS2\) and \(OLS3c\) provides that the errors are also not conditionally autocorrelated, i.e. \(\forall i \neq j: Cov[u_i, u_j|x_i, x_j] = E[u_i u_j|x_i, x_j] = E[u_i|x_i] \cdot E[u_j|x_j] = 0\). 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 \(\beta\) parameters to zero.

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

\[ V[u|X] = E[uu'|X] = \sigma^2\Omega \]

\[
\begin{pmatrix}
\sigma_1^2 & 0 & \cdots & 0 \\
0 & \sigma_2^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma_N^2
\end{pmatrix} = \sigma^2
\begin{pmatrix}
\omega_1 & 0 & \cdots & 0 \\
0 & \omega_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \omega_N
\end{pmatrix}
\]

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

\[ y_i = x'_i\beta + u_i \]

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

\[ y_g = x'_g\beta + u_g \]
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 parameter \( \beta \) can be estimated with the usual OLS estimator

\[
\hat{\beta}_{OLS} = (X'X)^{-1} X'y
\]

The OLS estimator of \( \beta \) 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 covariance matrix of \( \hat{\beta}_{OLS} \) is under \( OLS4b \)

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

and differs from usual OLS where \( V[\hat{\beta}_{OLS}|X] = \sigma^2(X'X)^{-1} \). Consequently, the usual estimator \( \hat{V}[\hat{\beta}_{OLS}|X] = \hat{\sigma}^2(X'X)^{-1} \) 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 \( OLS1 \), \( OLS2 \), \( OLS3d \), and \( OLS5 \)

\[
\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N \left( 0, Q_{XX}^{-1} Q_{XNX} Q_{XX}^{-1} \right)
\]

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

\[
\hat{\beta} \overset{a}{\sim} N \left( \beta, Avar[\hat{\beta}] \right)
\]
where $\text{Avar}[\hat{\beta}] = N^{-1}Q^{-1}_{XX}Q_{X\Omega X}Q_{XX}^{-1}$ can be consistently estimated as

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

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 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 heteroskedasticity 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 $\sigma_i^2$ 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 $\Omega$ is Known

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

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

The GLS estimator simplifies in the case of heteroskedasticity to

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

where

$$\tilde{y} = \left( \begin{array}{c} y_1 / \sqrt{\omega_1} \\ \vdots \\ y_N / \sqrt{\omega_N} \end{array} \right), \tilde{X} = \left( \begin{array}{c} x_1' / \sqrt{\omega_1} \\ \vdots \\ x_N' / \sqrt{\omega_N} \end{array} \right)$$

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 the explanatory variables also applies to the constant, i.e. $\bar{x}_0 = 1/\sqrt{\omega_i}$. The OLS regression using the transformed variables does not include an additional constant.

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

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

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

The WLS estimator of $\beta$ is consistent, asymptotically efficient and approximately normally distributed under OLS4b (conditional heteroscedasticity) and OLS2, OLS1, OLS3d, and a modification of OLS5.
**OLS5: Identifiability**

\[ X'\Omega^{-1}X \text{ is p.d. and finite} \]

\[ E[\omega_i^{-1}x_i'x_i'] = Q_{X\Omega^{-1}X} \text{ is p.d. and finite} \]

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

The covariance matrix of \( \hat{\beta}_{WLS} \) can be estimated as

\[ \hat{\text{Avar}}[\hat{\beta}_{WLS}] = \hat{V}[\hat{\beta}_{WLS}] = \hat{\sigma}^2(\hat{X}'\hat{X})^{-1} \]

where \( \hat{\sigma}^2 = \hat{u}'\hat{u}/(N-K-1) \) in small samples and \( \hat{\sigma}^2 = \hat{u}'\hat{u}/N \) in large samples and \( \hat{u} = \hat{y} - \hat{X}\hat{\beta}_{GLS} \). Usual tests (\( t, F \)-) for small samples are valid (under OLS1, OLS2, OLS3a, OLS4b and OLS5; usual tests (\( z, \text{Wald} \)) for large samples are also valid (under OLS1, OLS2, OLS3d, OLS4b and OLS5).

**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 \( \hat{\omega}_i \) and calculates then \( \hat{\beta}_{FGLS} \) 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

\[ \hat{u}_i^2 = \exp(z_i'\delta') + \nu_i \]

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

\[ \hat{\omega}_i = \exp(z_i'\delta) \]
Implementation in Stata 15

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

```stata
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,

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

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

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,

```stata
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:

1. Estimate the regression model \( y_i = x'_i \beta + u_i \) by OLS and predict the residuals \( \hat{u}_i = y_i - x'_i \hat{\beta} \).
2. Use the residuals to estimate a regression model for the error variance, e.g. \( \log(\hat{\omega}_i^2) = z'_i \delta + \nu_i \), and to predict the individual error variance, e.g. \( \hat{\omega}_i = \exp(z'_i \delta) \).
3. perform a linear regression using the weights \( 1/\hat{\omega}_i \)

For example,

```stata
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]
```

Implementation in R

R reports the Eicker-Huber-White covariance after estimation using the two packages `sandwich` and `lmtest`

```r
> library(foreign)
> fm <- lm(price~mpg+weight, data=auto)
> library(sandwich)
> library(lmtest)
> coeftest(fm, vcov=sandwich)
```

`F`-tests for one or more restrictions are calculated with the command `waldtest` which also

```r
> waldtest(fm, "weight", vcov=sandwich)
```

tests \( H_0 : \beta_1 = 0 \) against \( H_A : \beta_1 \neq 0 \) with Eicker-Huber-White, and

```r
> waldtest(fm, .~.-weight-displacement, vcov=sandwich)
```

tests \( H_0 : \beta_1 = 0 \) and \( \beta_2 = 0 \) against \( H_A : \beta_1 \neq 0 \) or \( \beta_2 \neq 0 \).
References

Introductory textbooks

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

Companion textbooks

Articles