Heteroskedasticity

One assumption underlying the results established so far is that E(uu') = σ^2I

ie the residuals have constant variance.

Suppose relax this assumption such that

$$E(uu') = \sigma^2\Omega = \begin{bmatrix} \sigma_1^{\ 2} & & & \\ & \sigma_2^{\ 2} & & \\ & & : & \\ & & & \sigma_N^{\ 2} \end{bmatrix} = \sigma^2 \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & : & \\ & & & X_N \end{bmatrix}$$

ie residual variance somehow depends on level/values of the X variable(s)
This is known as heteroskedasticity

Consequences

Given

Y = XB + u

Now E(u) =0 as before but E(uu') = $\sigma^2\Omega$

Given $\hat{\beta} = (X'X)^{-1}X'y$ = $(X'X)^{-1}X'(X\beta + u)$ $= \beta + (X'X)^{-1}X'u$

So

$$E(\hat{\beta}) = \beta + (X'\hat{X})^{-1}X'E(u)$$

OLS estimators remain unbiased in the presence of heteroskedasticity

But..

$$Var(\hat{\beta}) = E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right]$$

$= E[\ (X'X)^{-1}X'uu'X(X'X)^{-1}\ ]$

$= (X'X)^{-1}X'E(uu')\ X(X'X)^{-1}$

$= (X'X)^{-1}X'\sigma^2\Omega\ X(X'X)^{-1}$

$\sigma^2(X'X)^{-1}\ X'\Omega\ X(X'X)^{-1}$ (1)

$\neq\ \sigma^2(X'X)^{-1}$ (2)

Without knowledge of form of Ω can't say whether variance is higher or lower in (1) than (2). However if use (2) this will give biased estimates of true ***OLS*** variance

Generalised Least Squares

To avoid this problem can try to transform the data by a matrix T such that in the transformed model the new residual variance = σ^2I

ie $Ty = TXB = Tu$

and $E(Tuu'T) = \sigma^2T\ \Omega T' = \sigma^2I$

If Ω is a positive definite matrix then can show (see Greene) that there will always exist a matrix T such that

a) $C'\Omega C' = \Lambda$ (diagonalisation)
b) $\Omega = C\ \Lambda\ C'$ (spectral decomposition)
c) $\Omega^{-1} = C\ \Lambda^{-1}\ C'$
d) $\Omega^{-1} = T'T$ (3)

where

C = matrix of characteristic vectors of Ω
Λ = diagonal matrix whose non-zero elements are the characteristic roots of Ω

$T = (C\ \Lambda^{-1/2})'$

From (3) $\Omega = (T'T)^{-1}$
$= T^{-1}\ (T')^{-1}$

So
$T\ \Omega T' = T\ (\ T^{-1}\ (T')^{-1}\ )\ T'$
$= I$

Follows that if the matrix T has this form then estimating the transformed model

Ty = TXB = Tu

or y* = X*B + u*

by OLS gives

$$\hat{\beta} = (X^{*}{'}X^{*})^{-1}X^{*}{'}y^{*}$$

$$= (X'T'TX)^{-1}X'T'Ty$$

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$ (from (3))

This is the Generalised Least Squares (GLS) estimator

which since it has the property that

$$E(u^{*}u^{*}{'}) = E(Tuu'T) = TE(uu')T'$$

$$= \sigma^{2}T\Omega T'$$

$$= \sigma^{2}I$$

with $E(\hat{\beta}_{GLS}) = E[\,(\mathrm{X}'\Omega^{-1}\mathrm{X})^{-1}\mathrm{X}'\Omega^{-1}(\mathrm{X}\beta + \mathrm{u})]$

$$= \beta + E[(\mathrm{X}'\Omega^{-1}\mathrm{X})^{-1}\mathrm{X}'\Omega^{-1}\mathrm{u})]$$

so $E(\hat{\beta}_{GLS}) = \beta$

ie GLS estimator is unbiased

Moreover it is this model and the resulting GLS estimate that now satisfies the criteria to be the Gauss-Markov minimum variance efficient estimator

With the variance of the GLS estimator give by

$$Var(\hat{\beta}_{GLS}) = E\left[(\hat{\beta}_{GLS} - \beta)(\hat{\beta}_{GLS} - \beta)'\right]$$

$$= E[\,(\mathrm{X}'\Omega^{-1}\mathrm{X})^{-1}\mathrm{X}'\Omega^{-1}\mathrm{uu}'\Omega^{-1}\mathrm{X}(\mathrm{X}'\Omega^{-1}\mathrm{X})^{-1}]$$

$$= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E[uu']\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$$

$$Var(\hat{\beta}_{GLS}) = \sigma^2 (X'\Omega^{-1}X)^{-1} = \sigma^2 (X^{*}{}'X^{*})^{-1}$$

with an unbiased estimate of σ^2 given by

$$s^2_{GLS} = \frac{(y^* - X^* \hat{\beta}_{GLS})'(y^* - X^* \hat{\beta}_{GLS})}{N-k} = \frac{\hat{u}^{*}{}'\hat{u}^{*}}{N-k}$$

So GLS is the more efficient estimator than OLS in either (1) or (2) and the inefficiency of OLS rises with the dispersion n the residuals (If the assumption about the nature of the residual variance ie the form of Ω is correct)

Weighted Least Squares

One area in which can make reasonable assumptions about the functional form of Ω is when the data are ***grouped*** ie averaged up from a lower level (eg regional averages of individual or firm-level responses)

$$\bar{y}_g = \bar{X}_g \beta + \bar{u}_g \qquad \text{g = 1, 2 … G groups}$$

In this case

$$Var(\bar{u}_g) = Var[\frac{1}{N_g}\sum_i u_i] = \frac{1}{N_g{}^2} Var[\sum_i u_i]$$

$$= \sigma^2/N_g$$

and the variance/covariance matrix of residuals is given by

$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} 1/N_1 & & & \\ & 1/N_2 & & \\ & & \vdots & \\ & & & 1/N_G \end{bmatrix}$$

so the residual variance **falls** as the size of the group on which the average is based **rises**

It follows that

$$\Omega^{-1} = \begin{bmatrix} N_1 & & & \\ & N_2 & & \\ & & : & \\ & & & N_G \end{bmatrix}$$

Since can always write

$$X'X = \sum_{j=1}^{G} x_j x_j'$$

where x_j' is the j^{th} row of X

then

$$X'\Omega^{-1}X = \sum_{j=1}^{G} N_g x_j x_j' \qquad X'\Omega^{-1}y = \sum_{j=1}^{G} N_g x_j y_j$$

and

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}y = \left[\sum_{j=1}^{G} N_g x_j x_j'\right]^{-1} \left[\sum_{j=1}^{G} N_g x_j y_j\right]$$

so if each element of the x_j and y_j row are multiplied by the **square root** of N_j and the resulting transformed variable model estimated by OLS this weighted least squares estimator will be equal to the GLS estimator

Observations based on a larger sample have smaller variances and are therefore perceived to be more accurate and therefore receive more weight in the computations

Generalising

If the elements of Ω are known then the **Feasible GLS** procedure is to transform the data by the *reciprocal* of the **square root** of the appropriate element in Ω

ie if $\quad Var(u_i) = \sigma_i^2 = \sigma^2 w_i$

where w_i is a function of one or more of the X variables in the model

then the GLS estimator becomes

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{N} \frac{1}{w_i} x_i x_i ' \right]^{-1} \left[\sum_{j=1}^{G} \frac{1}{w_i} x_i y_i \right]$$

with

$$Ty = \begin{bmatrix} y_1 / \sqrt{w_1} \\ y_2 / \sqrt{w_2} \\ \vdots \\ y_N / \sqrt{w_N} \end{bmatrix} \qquad \text{and} \qquad TX = \begin{bmatrix} X_{11} / \sqrt{w_1} & .. & X_{1k} / \sqrt{w_1} \\ X_{21} / \sqrt{w_2} & & X_{2k} / \sqrt{w_2} \\ \vdots & & \vdots \\ X_{N1} / \sqrt{w_N} & .. & X_{Nk} / \sqrt{w_N} \end{bmatrix}$$

Again observations with smaller variances receive larger weight in the computations

2-Step Feasible GLS when Ω unknown

More likely that will not know exact form of heteroskedasticity.

Since

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1} y = \left[\sum_{i=1}^{N} \left(\frac{1}{\sigma_i^2} \right) x_i x_i ' \right]^{-1} \left[\sum_{i=1}^{N} \left(\frac{1}{\sigma_i^2} \right) x_i y_i \right]$$

If could estimate σ^2_i then could again use some form of feasible GLS, replacing σ^2_i with a consistent estimator $\hat{\sigma}_i^2$

One solution is to parameterise the form of heteroskedasticity so that

$$\sigma_i^2 = \exp[Z\gamma]$$

or

$$\log(\sigma_i^2) = \sigma^2 + Z\gamma$$

where the Z are a subset of some or all of the X variables in the model
(the exponent ensures the transformation is always positive)

Can show if regress the log of the square of the OLS residuals $\hat{u}_i$ from the original model y = XB + u on the set of variables in Z

$$\log(\hat{u}_i^2) = Z_i\gamma + v$$

and use the estimates of γ to calculate then can estimate $\hat{\sigma}_i^2 = \exp[Z\hat{\gamma}]$

Can show that **if** the assumption about the functional from of Ω is correct then feasible GLS is consistent asymptotically more efficient tan OLS (but will be biased in small samples)

If assumption about functional form is wrong then feasible GLS is inconsistent

Testing for Heteroskedasticity

a) Goldfeld-Quandt

Assume you know which variable, X, may be causing the problem then under null of homoskedasticity if split sample into 2 groups according to the value of X then the residual variances should be the same in the X

i) Sort data by size of variable of interest.
ii) Split sample into 2 and run separate regressions for the two sub-samples. Save the RSS
iii) Under the null of no difference then

$$F = \frac{RSS_{high\,variance} / N_2 - k}{RSS_{low\,variance} / N_1 - k} \sim \mathrm{F}[\mathrm{N}_2 - k, N_1 - k]$$

Reject null of homoskedasticity if $\hat{F} > F_{critical}$

In order to increase the ***power*** of the test, can drop the middle c observations

$$F = \frac{RSS_{high\ variance} / N_2 - c/2 - k}{RSS_{low\ variance} / N_1 - c/2 - k} \sim \mathrm{F}[\mathrm{N}_2 - c/2 - k, N_1 - c/2 - k]$$

In general it is unlikely that you will know the variable causing heteroskedasticity or there may be more than one variable responsible.

Instead we would like to test whether the residual variance depends on a set of variables, Z

$$u_i^2 = d_0 + Z_i\delta + e_i \qquad (1)$$

Since u_i^2 is unobserved replace with OLS residuals (which are consistent estimates of u_i if the model is correctly specified)

$$\hat{u}_i^2 = d_0 + Z_i\delta + e_i \qquad (2)$$

and estimate (2) by OLS

Can show that

$$N^*R^2 \overset{asy}{\sim} \chi_r^2$$

where r = number of right hand side variables in Z (ie excluding constant)

[this is the **Breusch-Pagan** test for heteroskedasticity and belongs to the set of test known as LM tests – see Johnston & DiNardo ch. 5 for derivation]

If $N^*R^2 > \chi_{critical}^2$ then reject null of homoskedasticity

The choice of what variables to include in Z is a matter of choice. The variant known as the **White test** takes the set of variables in the original X vector adds their squares (note the square of a dummy variable does not exist) and the interaction of all the variables as the elements of the Z vector

Note that this can involve lots of parameters when the X matrix is large. As a result the degrees of freedom in this version of the test are quite large and this tends to reduce the power of the test, particularly when N is small.

If don't know the exact form of heteroskedasticity may be useful instead to use an estimator that is consistent no matter the form of heteroskedasticity.

Since the true OLS variance/covariance matrix of the parameters

$$Var(\hat{\beta}_{OLS}) = \sigma^2(X'X)^{-1} X'\Omega X(X'X)^{-1} \neq \sigma^2(X'X)^{-1}$$

Again Ω is unknown, but we can replace the middle (sandwich) term, X'Ω X, using the fact that this will always be a $k_x k$ matrix regardless of the sample size and so as the sample size gets larger, the sample average

$$\frac{\sum_{i=1}^{N} u_i^2 x_i' x_i}{N} \xrightarrow{p} E(u^2 X' X) = \sigma^2 X' \Omega X$$

(this result holds even if the estimate of Ω is inconsistent)

Once again since the true residuals are never observed, use instead the observed OLS residuals squared

White (1980) showed that a consistent estimate of the true OLS variance is given by

$$Var(\hat{\beta}_{OLS}) = (X' X)^{-1} S (X' X)^{-1}$$

where $S = \frac{\sum_{i=1}^{N} \hat{u}_i^2 x_i' x_i}{N} \xrightarrow{p} E(u^2 X' X) = \sigma^2 X' \Omega X$

This variance/covariance matrix is said to be **robust** to the presence of unknown heteroskedasticity

Note 1 that these resulting robust standard errors are only valid asymptotically.

Note 2 also that this is not the minimum variance efficient estimator (the GLS estimator is), but often this is the best that can be done in the presence of unknown heteroskedasticty and so it is good practice when working with cross-section data to present estimates based on the robust standard errors as a matter of routine.

Note 3. The R^2 and the $\bar{R}^2$ are unaffected by the presence of heteroskedasticity since the RSS does not depend on whether E[uu'/X} is constant or not

Clustering

A related issue arises when trying to estimate models of the form

$$y_{is} = X_{is}\beta + Z_s\gamma + u_{is}$$

where i = individual
s = stratum

and the strata are a higher level of aggregation than the individual (eg region, industry, country)

Suppose that the error term is ***composite*** , ie consists of more than one component such that

$$u_{is} = q_s + e_{is}$$

where q_s varies across strata but not individuals (ie an unobserved effect common to all individuals in the same strata) and e_{is} is a residual that varies over both individuals and strata

The common strata component introduces an element of correlation between the residuals of individuals in the same strata

Moulton (1990) shows that in this case the residual variance/covariance matrix is given by

$$E(uu') = \sigma^2\Omega = \sigma^2\left[(1-\rho)I_N + \rho DD'\right]$$

where ρ is the correlation coefficient between residuals in the same strata (assumed constant across all strata)

and D is an N_xs matrix of 0,1 indicators for membership of one of the s groups

Moulton shows that the difference between the unadjusted OLS and the adjusted OLS clustered variance/covariance matrix

$$\sigma^2(X'X)^{-1} X'\Omega X(X'X)^{-1} = \sigma^2(X'X)^{-1}[I + \rho(W-I)]$$

where $W = X'DD'X\,(X'X)^{-1}$

so when ρ = 0 the variances are identical. As ρ rises the corrected clustered variance also rises. So unadjusted OLS variances and standard errors are ***downward biased and t and F values upward biased***

This can make a big difference when the number of strata are small. Again there is a danger of Type I error (rejecting a true null)