

HETEROSKEDASTICITY – GLS (WLS) ESTIMATORS – WHITE CORRECTION

Maria Elena Bontempi [e.bontempi@economia.unife.it](mailto:e.bontempi@economia.unife.it)

Roberto Golinelli [roberto.golinelli@unibo.it](mailto:roberto.golinelli@unibo.it)

this version: 5 September 2007<sup>§</sup>

**1. Introduction**

The assumptions that makes OLS being the best (“B” of BLU) estimator are:

On average the regression line is correct	$E(\varepsilon_i) = 0, \forall i = 1, \dots, N$	} identically distributed	} $\varepsilon \sim iid(0, \sigma^2 I_N)$
Homoskedasticity	$E(\varepsilon_i^2   X) = Var(\varepsilon_i   X) = \sigma^2, \forall i = 1, \dots, N$		
Non cross-correlation	$E(\varepsilon_i \varepsilon_j   X) = Cov(\varepsilon_i, \varepsilon_j   X) = 0, \forall i \neq j$	} independently distributed	

where  $\sigma^2 I_N$  is the VCOV matrix of errors, equal to  $E(\varepsilon\varepsilon') = \sigma^2 \begin{bmatrix} 1 & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2 & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & \sigma^2 \end{bmatrix}$ .

In other terms, the VCOV matrix is a scalar matrix, i.e. a diagonal matrix whose diagonal elements are all equal. In cross-section and panel data the homoskedasticity assumption is rarely satisfied.

On the contrary, heteroskedasticity is highly probable, implying that  $Var(\varepsilon_i | X) = \sigma^2 \omega_i = \sigma_i^2$ . In matrix notation, we can write:

$$Var(\varepsilon|X) = E(\varepsilon\varepsilon' | X) = \sigma^2 \Omega = \sigma^2 Diag(\omega_i^2) = \sigma^2 \begin{bmatrix} \omega_1^2 & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & \omega_N^2 \end{bmatrix}$$

Consider the following example.

```
cd...
use GLS_data, clear
. descr

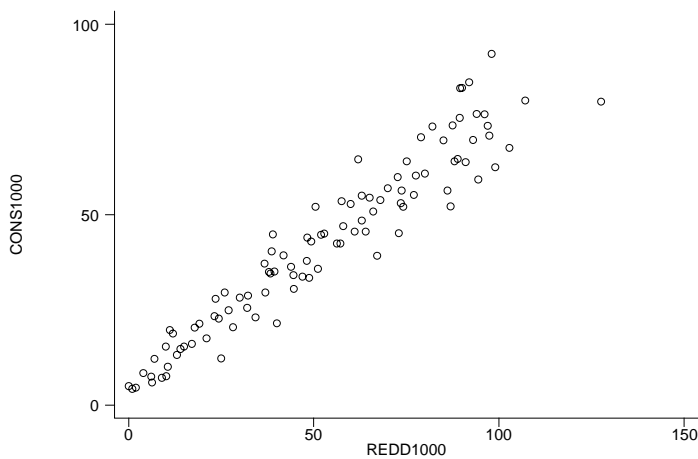
Contains data
obs:          100
vars:         3
size:        1,300 (99.9% of memory free)
-----
variable name  storage  display  value  variable label
              type    format   label
-----
obs            byte    %8.0g
cons1000       float   %9.0g   consumo nel 2003 a prezzi
              float   %9.0g   costanti
redd1000       float   %9.0g   reddito nel 2002 a prezzi
              float   %9.0g   costanti
-----
```

<sup>§</sup> Very preliminary. Comments welcome.

The idea of explaining the consumption with the income in the previous year predetermines the dynamic relationship in a quite restrictive way but at the advantage of avoiding consumption-income simultaneity.

Scatterplot tells us that the consumption variability grows with the level of income: richer people behave in different ways. This fact *per se* implies the likely heteroschedasticity of the linear model residuals.

```
. graph7 cons1000 redd1000, ylabel xlabel
```



### Keynes's (linear) consumption function

```
. reg cons1000 redd1000
```

Source	SS	df	MS			
Model	46059.3208	1	46059.3208	Number of obs =	100	
Residual	4354.87802	98	44.4375308	F( 1, 98) =	1036.50	
Total	50414.1988	99	509.234332	Prob > F =	0.0000	
				R-squared =	0.9136	
				Adj R-squared =	0.9127	
				Root MSE =	6.6661	

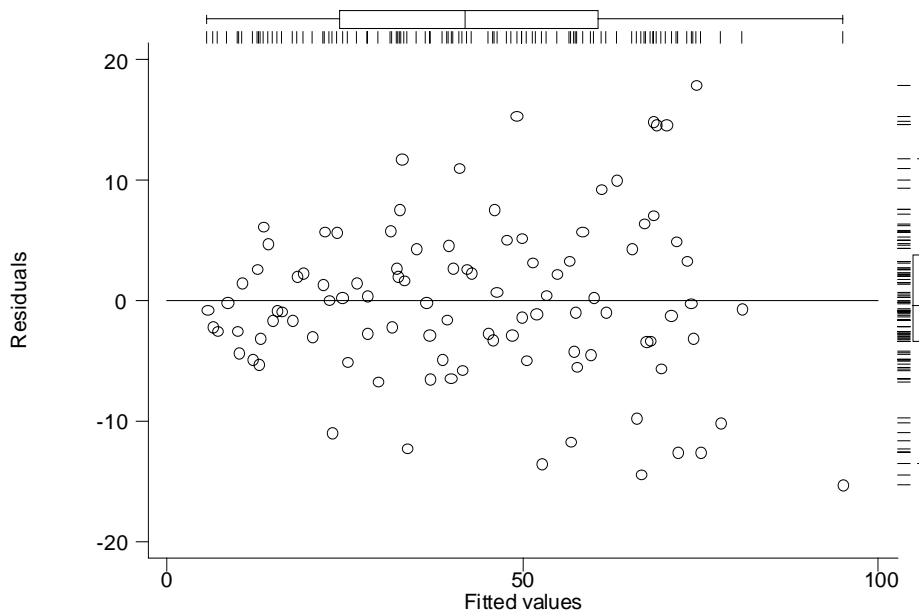
  

cons1000	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
redd1000	.7002875	.0217517	32.19	0.000	.6571221	.743453
_cons	5.668498	1.331578	4.26	0.000	3.026025	8.310971

## 2. Heteroskedasticity tests

Graphical analysis represents a first step towards discovering whether heteroskedasticity is present. We are supposing that the error variance is a function of income:

```
version 7: rvfplot, oneway twoway box ylabel xlabel yline(0)
```



Heteroskedasticity tests verify the hypothesis  $H_0: \text{VAR}(\varepsilon_i) = \sigma^2, \forall i = 1, \dots, N$ .

In general, the tests use auxiliary regressions in the form  $\hat{\varepsilon}_i^2 = f(\alpha'Z_i) + u_i$ , where  $u_i \sim iid(0, \sigma_u^2)$ ,  $\alpha$  and  $Z_i$  are  $V \times 1$  vectors, with  $V$  number of variables in  $Z$  (and of associated parameters  $\alpha$ s) used to explain the error variance; for this  $Z_i$  are called the variance indicator variables. The null hypothesis to be tested becomes  $H_0: \alpha=0$ . Under the alternative, the form of the detected heteroskedasticity depends on the choice of the explanatory indicators  $Z_i$ .

```
. hettest
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
```

```
Ho: Constant variance
```

```
Variables: redbl000
```

```
chi2(1)          =    21.50
```

```
Prob > chi2     =    0.0000
```

This command executes the **Breusch-Pagan (1979)**<sup>1</sup> Lagrange multiplier test for heteroskedasticity in the error distribution. It is the most general test, even if not powerful and sensitive to the original assumption of error normally distributed (see below for a change in this assumption). If you write `hettest` only, the residual variance is assumed to depend on the fitted values (*i.e.*  $Z_i \equiv \hat{y}_i$ , and  $V=1$ ); if you use the option `,rhs` the residual variance is assumed to depend on the explanatory variables of the model. In STATA 9 is also possible to test whether heteroskedasticity is conditional on any list of  $Z_i$  variables, which are presumed to influence the error variance (*i.e.* variance-indicators). The trade-off in the choice of indicator variables in these tests is that a smaller set of indicator variables will preserve degrees of freedom, at the cost of being unable to detect heteroskedasticity in certain directions.

For example, if we suppose that in our simple consumption model the levels of income and thier squares are both valid variance indicators, we can test for heteroskedasticity in the following way:

<sup>1</sup> Breusch-Pagan (1979), Godfrey (1978), and Cook-Weisberg (1983) separately derived the same test statistic. Breusch, T. and A. Pagan. "A Simple Test for Heteroskedasticity and Random Coefficient Variation." *Econometrica*, 47, 1979, 1287-1294; Godfrey, L. G. (1978), Testing for multiplicative heteroskedasticity, *Journal of Econometrics*, 8, 227-236; Cook, R. D. and S. Weisberg (1983), Diagnostics for heteroscedasticity in regression, *Biometrika*, 70, 1-10.

```
. g redd2=redd1000^2
. hettest redd1000 redd2
```

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance  
Variables: redd1000 redd2

```
chi2(2)      =    25.00
Prob > chi2  =    0.0000
```

Thus, in STATA 9 `hettest` replicates the same result before (in STATA 8 or earlier versions) obtainable only by applying a procedure, written by C. F. Baum and V. Wiggins, that specifically run the Breusch-Pagan (1979) test for heteroskedasticity conditional on a set of variables.

```
. bpagan redd1000 redd2
```

Breusch-Pagan LM statistic: 25.0018 Chi-sq( 2) P-value = 3.7e-06

In general, the Breusch and Pagan test-statistic is distributed as a chi-squared with  $V$  degrees of freedom (in the latter example  $V=2$ ). The statistic above may be replicated with the following steps.

1) run the model regression and define the dependent variable of the Breusch-Pagan auxiliary

regression  $g_i = \frac{\hat{\epsilon}_i^2}{\frac{1}{N} \sum_{i=1}^N \hat{\epsilon}_i^2}$  <sup>2</sup>:

```
. reg cons1000 redd1000
```

Source	SS	df	MS			
Model	46059.3208	1	46059.3208	Number of obs =	100	
Residual	4354.87802	98	44.4375308	F( 1, 98) =	1036.50	
				Prob > F =	0.0000	
				R-squared =	0.9136	
				Adj R-squared =	0.9127	
Total	50414.1988	99	509.234332	Root MSE =	6.6661	

	cons1000	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
redd1000		.7002875	.0217517	32.19	0.000	.6571221	.743453
_cons		5.668498	1.331578	4.26	0.000	3.026025	8.310971

```
. predict res, resid
. g BP_g= res^2/(e(rss)/e(N))
```

where  $e(rss)=4354.87802$  and  $e(N)=100$  are post estimation results corresponding, respectively, to the residual sum of squares and to the total number of observations.

2) run the auxiliary regression  $g_i = \alpha_0 + \alpha'Z_i + u_i$  and obtain the BP statistic as:  $BP = MSS/Mdf$

Source	SS	df	MS			
Model	<b>50.0036064</b>	2	25.0018032	Number of obs =	100	
Residual	188.030708	97	1.93846091	F( 2, 97) =	12.90	
				Prob > F =	0.0000	
				R-squared =	0.2101	
				Adj R-squared =	0.1938	
Total	238.034314	99	2.40438701	Root MSE =	1.3923	

<sup>2</sup> For this, Breusch-Pagan (1979, p. 1293) say: "... the quantity  $g_i$  is of some importance in tests of heteroskedasticity.

Thus, if one is going to plot any quantity, it would seem more reasonable to plot  $g_i$  than  $\hat{\epsilon}_i^2$ ."

BP_g	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
redd1000	-.0092459	.0167538	-0.55	0.582	-.0424975	.0240058
redd2	.0002848	.0001499	1.90	0.060	-.0000127	.0005823
_cons	.4226007	.4038909	1.05	0.298	-.3790109	1.224212

```
. di e(mss)/e(df_m)
25.001803
```

where  $e(mss)=50.0036064$  and  $e(df_m)=2$  are post estimation results corresponding, respectively, to the model sum of squares and to the model degrees of freedom of the auxiliary regression. The p-value of the test is obtained as:

```
. display chi2tail(2,e(mss)/e(df_m))
3.723e-06
```

Besides Breusch-Pagan test, the very often reported **White (1980)**<sup>3</sup> test for heteroskedasticity is based on a different auxiliary regression where the squared residuals are regressed on the model regressors, all their squares, and all their possible (nonredundant) cross products. The asymptotic chi-squared White test-statistic is obtained by the product of the number of observations times the R-squared of the auxiliary regression. The F-version for small samples is obtained by setting to zero all the explanatory variables of the auxiliary regression (*i.e.* by looking at the F-test for the overall significance of the auxiliary regression). Baum and Cox have written a procedure that automatically computes the asymptotic version of the White test. In our simple one-regressor consumption model, we have:

```
. qui reg cons1000 redd1000
. whitetst
```

White's general test statistic : 21.00689 Chi-sq( 2) P-value = 2.7e-05

This result may be replicated with the following steps.

```
. g res2=res^2
```

The corresponding auxiliary regression is (remember that we have only one explanatory variable):

```
. reg res2 redd1000 redd2
```

Source	SS	df	MS	Number of obs = 100		
Model	94831.6529	2	47415.8264	F( 2, 97) =	12.90	
Residual	356599.534	97	3676.28386	Prob > F =	0.0000	
				R-squared =	0.2101	
				Adj R-squared =	0.1938	
Total	451431.187	99	4559.91098	Root MSE =	60.632	

res2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
redd1000	-.4026459	.7296083	-0.55	0.582	-1.850716	1.045424
redd2	.0124035	.0065273	1.90	0.060	-.0005513	.0253583
_cons	18.40375	17.58896	1.05	0.298	-16.50546	53.31296

<sup>3</sup> White, H. (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, 817-838.

### Asymptotic version

The LM test-statistic for heteroskedasticity is just the sample size N times the R-squared of the auxiliary regression:

```
. di e(N)*e(r2)
21.00689
```

where  $e(N)=100$  and  $e(r2)=0.2101$  are post estimation results corresponding, respectively, to the total number of observations and to the R-squared of the auxiliary regression. The p-value of the test is obtained as:

```
. display chi2tail(2,e(N)*e(r2))
.00002744
```

### F version for small samples

```
. testparm redd1000 redd2

( 1) redd1000 = 0.0
( 2) redd2 = 0.0

F( 2, 97) = 12.90
Prob > F = 0.0000
```

The Breusch-Pagan (1979) test from the `hettest` command is numerically equal to the White (1980) test for heteroskedasticity, if the same White's auxiliary regression is specified and the option `iid` is used. Differently from the default of `hettest` and from `bpagan`, that compute the original Breusch-Pagan test assuming that the regression disturbances are normally distributed, the option `iid` causes `hettest` to compute the  $NR^2$  version of the score test, which drops the normality assumption.<sup>4</sup>

```
. qui reg cons1000 redd1000
. hettest redd1000 redd2, iid
```

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: redd1000 red2

chi2(2) = 21.01
Prob > chi2 = 0.0000
```

### **3. How accounting for heteroskedasticity?**

Heteroskedasticity implies that  $Var(\varepsilon | X) = \sigma^2 \Omega$ , where  $\Omega$  is a positive definite matrix, not necessarily scalar; thus, N additional parameters (the diagonal terms) may need to be estimated.

As a consequence, the OLS parameter estimate is unbiased but not efficient because:

$$\begin{aligned} Var(\hat{\beta} | X) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = (X'X)^{-1} X'Var(\varepsilon | X)X(X'X)^{-1} = (X'X)^{-1} X'\sigma^2 \Omega X(X'X)^{-1} = \\ &= (X'X)^{-1} X'Diag(\sigma_i^2)X(X'X)^{-1} \neq \sigma^2 (X'X)^{-1} \end{aligned}$$

---

<sup>4</sup> Koenker (1981) showed that when the assumption of normality is removed, a version of the test is available that can be calculated as the sample size N times the centered R-squared from an artificial regression of the squared residuals from the original regression on the indicator variables. Koenker, R. (1981), A note on studentizing a test for heteroskedasticity, *Journal of Econometrics*, 17, 107-112.

In particular, the variance of  $\hat{\beta}$  is higher than  $\sigma^2(X'X)^{-1}$  (homoskedastic case) by the positive definite matrix  $(X'X)^{-1}X'\Omega X$ .

Moreover,  $s^2$  is a biased estimator of  $\sigma^2$ , the error variance:

$$\begin{aligned} E(s^2) &= E\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{N-K}\right) = E\left(\frac{\varepsilon'M\varepsilon}{N-K}\right) = E\left[\frac{tr(M\varepsilon\varepsilon')}{N-K}\right] = \frac{1}{N-K}tr[E(M\varepsilon\varepsilon')] = \frac{1}{N-K}tr[ME(\varepsilon\varepsilon')] = \\ &= \frac{1}{N-K}tr(M\sigma^2\Omega) = \frac{\sigma^2}{N-K}tr(M\Omega) \neq \sigma^2 \end{aligned}$$

where  $M=I-X(X'X)^{-1}X'$  is the matrix projecting  $Y$  upon the space orthogonal to the one spanned by the columns of  $X$ . The matrix  $M$  is symmetric ( $M'=M$ ), idempotent ( $MM=M$ ), with  $\text{rank}(M)=N-K$ .

Hence, the variance of  $\hat{\beta}_{OLS}$  is biased because the weighing matrix is no more  $\sigma^2(X'X)^{-1}$  and because  $s^2$  is a biased estimator of  $\sigma^2$ . As a consequence, inference (test t and F) is not correct: statistic-test do not have the standard distribution; standard confidence regions are no more valid.

### 3.1. Heteroskedasticity-consistent estimates of the standard errors

A first way to account for heteroskedasticity is that of estimating model's parameters by OLS (if the Keynesian model is correctly specified, OLS estimator is unbiased and consistent, even if not efficient due to heteroskedasticity), and of correcting the OLS estimates of the standard errors (biased). To do so, consistent standard errors are needed.

The robust option of the regress Stata command specifies that the Eicker/Huber/White sandwich estimator<sup>5</sup> of variance is used instead of the traditional OLS error variance estimator. In particular, White (1980) argues that it is not necessary to estimate all  $\sigma_i^2$ s, but that we simply need a consistent estimator of the  $(K \times K)$  matrix  $X'E(\varepsilon\varepsilon')X = X'\sigma^2\Omega X = X'Diag(\sigma_i^2)X = \sum_{i=1}^N \sigma_i^2 X_i X_i'$ .

If we define as  $X_i$  the  $(K \times 1)$  vector of explanatory variables for observation  $i$ , a consistent estimator can be obtained as  $\frac{X'\hat{\varepsilon}\hat{\varepsilon}'X}{N} = \frac{1}{N} \sum_{i=1}^N \hat{\varepsilon}_i^2 X_i X_i'$ , where  $\hat{\varepsilon}_i$  is the OLS residual and

$$plim \frac{X'\hat{\varepsilon}\hat{\varepsilon}'X}{N} = X'\sigma^2\Omega X.$$

Thus, the “sandwich”:

$$Var(\hat{\beta}) = (X'X)^{-1} \sum_{i=1}^N \hat{\varepsilon}_i^2 X_i X_i' (X'X)^{-1} = \left( \sum_{i=1}^N X_i X_i' \right)^{-1} \sum_{i=1}^N \hat{\varepsilon}_i^2 X_i X_i' \left( \sum_{i=1}^N X_i X_i' \right)^{-1}$$

can be used as an estimate of the true variance of the OLS estimator.

<sup>5</sup> Eicker F. (1967) “Limit Theorems for Regressions with Unequal and Dependent Errors”, in L. Le Cam and J. Neyman (eds.) Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 1, 59-82; Huber P. J. (1973) “Robust Regression: Asymptotics, Conjectures, and Monte Carlo”, The Annals of Statistics, 1, 799-821; White, H. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.”, Econometrica, 48, 1980, 817-838.

In our case above, after we detected residuals heteroskedasticity, under the assumption that the other assumptions about our keynesian model hold, we can obtain consistent standard errors using a very simple option:

```
. reg cons1000 redd1000, robust
```

```
Regression with robust standard errors                                Number of obs =      100
                                                                    F(   1,   98) =   799.78
                                                                    Prob > F      =   0.0000
                                                                    R-squared     =   0.9136
                                                                    Root MSE     =   6.6661
```

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
cons1000						
redd1000	.7002875	.0247622	28.28	0.000	.6511477	.7494274
_cons	5.668498	1.076363	5.27	0.000	3.532492	7.804505

NOTE: parameter estimates (with and without standard errors correction) are identical: the White correction does not modify the parameters' estimates.

### 3.2. Feasible generalised least squares (FGLS)

If we have some idea about the heteroskedasticity determinants, we can introduce a different estimator: WLS (weighted least squares). WLS is a specific case of FGLS (feasible generalised least squares), the efficient estimator in the context of heteroskedastic errors (remember: OLS is only consistent but inefficient because does not account for the heteroskedastic behaviour of errors).

The general principle at the basis of FGLS is the following. Suppose we dispose of a consistent estimate  $\hat{\Omega}$  (or we know  $\Omega$ ). In addition,  $\hat{\Omega}$  is not singular, and it is possible to find a  $(K \times N)$  matrix  $L$  such that  $L\hat{\Omega}L' = I_N$  and  $L'L = \hat{\Omega}^{-1}$ .

Pre-multiply it by  $L$  the heteroskedastic model:  $y = X\beta + \varepsilon$  and obtain

$$y^* = X^*\beta + \varepsilon^*$$

where  $y^* = Ly$ ,  $X^* = LX$  and  $\varepsilon^* = L\varepsilon$

Now it is true that:

$$E(\varepsilon^*) = E(L\varepsilon) = LE(\varepsilon) = 0$$

$$E(\varepsilon^*\varepsilon^{*\prime}) = E(L\varepsilon\varepsilon'L') = LE(\varepsilon\varepsilon')L' = \sigma^2 L\hat{\Omega}L' = \sigma^2 I_N.$$

Hence, the OLS estimator of the transformed model is best (minimum variance) and corresponds to the FGLS estimator:

$$\hat{\beta}_{FGLS} = (X^{*\prime}X^*)^{-1}X^{*\prime}y^* = (X'L' LX)^{-1}X'L'Ly = (X\hat{\Omega}^{-1}X)^{-1}X\hat{\Omega}^{-1}y.$$

The specific form of the  $L$  matrix depends on the problem one has to tackle. The general principle is that of minimise an appropriately weighted average of squared errors, with lower weights to the observations characterised by the higher residual variance. The FGLS are BLUE despite the presence of heteroskedasticity (and/or autocorrelation); in other terms, the Aitken Theorem applied to transformed data substitutes for the Gauss-Markov Theorem, and, in particular, the Gauss-Markov theorem is a special case of the Aitken theorem for  $\Omega = I_N$ .

*Behavioural assumption:* the error variance is a linear function of income (redd1000) because wealthy people have a larger set of consumption options. If this is true, then it is reasonable use such information in the estimation phase down-weighting the observation corresponding to higher incomes because less informative about the regression line. In fact, they are assumed to be more dispersed (higher variance) than those of poorer people.

From the model  $C_i = \alpha + \beta R_i + \varepsilon_i$ , where  $Var(\varepsilon_i) = \sigma_i^2 = \sigma^2 R_i$  if we scale all the variables by the root square of income, we obtain the transformed model:

$$\frac{C_i}{\sqrt{R_i}} = \frac{\alpha}{\sqrt{R_i}} + \beta \frac{R_i}{\sqrt{R_i}} + \frac{\varepsilon_i}{\sqrt{R_i}} = \frac{\alpha}{\sqrt{R_i}} + \beta \frac{R_i}{\sqrt{R_i}} + u_i,$$

where  $Var(u_i) = Var\left(\frac{\varepsilon_i}{\sqrt{R_i}}\right) = \frac{1}{R_i} Var(\varepsilon_i) = \frac{1}{R_i} \sigma_i^2 = \frac{1}{R_i} \sigma^2 R_i = \sigma^2$ , i.e. errors  $u_i$  are homoskedastic.

WLS are efficient just because the higher-variance observations (i.e. those corresponding to richer people) have less weight.

```
. reg cons1000 redd1000 [aweight=1/redd1000]
(sum of wgt is 1.0121e+02)
```

Source	SS	df	MS			
Model	2623.25305	1	2623.25305	Number of obs =	100	
Residual	84.2000896	98	.859184587	F( 1, 98) =	3053.19	
Total	2707.45314	99	27.3480115	Prob > F =	0.0000	
				R-squared =	0.9689	
				Adj R-squared =	0.9686	
				Root MSE =	.92692	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
redd1000	.7145188	.0129311	55.26	0.000	.6888574	.7401803
_cons	4.914329	.0935686	52.52	0.000	4.728645	5.100013

FGLS (WLS) can be reproduced by the following steps:

```
. g peso=1/redd1000^0.5
. g consp=cons1000*peso
. g reddp=redd1000*peso
. reg consp reddp peso, noconst
```

Source	SS	df	MS			
Model	5852.17626	2	2926.08813	Number of obs =	100	
Residual	85.2218967	98	.869611191	F( 2, 98) =	3364.82	
Total	5937.39815	100	59.3739815	Prob > F =	0.0000	
				R-squared =	0.9856	
				Adj R-squared =	0.9854	
				Root MSE =	.93253	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
reddp	.7145188	.0129311	55.26	0.000	.6888574	.7401803
peso	4.914329	.0935686	52.52	0.000	4.728645	5.100013

```
. whitetst
```

White's general test statistic : 5.086438 Chi-sq( 5) P-value = .4054

```
. g reddp2=reddp^2
. bpagan reddp reddp2
```

Breusch-Pagan LM statistic: 1.265137 Chi-sq( 2) P-value = .5312

Opposite to the previous two heteroskedasticity tests, note that after a regression without the constant term we cannot run the `hettest` command:

```
. hettest
not appropriate after regress, nocons
r(301);
```

All the issues raised above can be summarised in a single table in order to fix ideas. In doing so, we use the previous (we checked it is heteroskedastic) consumption function. To do so, we can quietly run three regressions of interest, namely: (1) heteroskedastic OLS without White's standard errors correction; (2) heteroskedastic OLS with White's standard errors correction; (3) WLS assuming that the errors variance is a linear function of incomes:

```
. qui reg cons1000 redd1000
. est store OLS

. qui reg cons1000 redd1000, robust
. est store white

. qui reg cons1000 redd1000 [aweight=1/redd1000]
. est store WLS

. est table OLS white WLS , b(%6.3f) se(%6.3f) t(%6.2f) /*
*/                                stats(N df_f df_m r2 r2_a rmse F)
```

Variable	OLS	white	WLS
redd1000	0.700	0.700	0.715
	0.022	0.025	0.013
	32.19	28.28	55.26
_cons	5.668	5.668	4.914
	1.332	1.076	0.094
	4.26	5.27	52.52
N	100.000	100.000	100.000
df_f			
df_m	1.000	1.000	1.000
r2	0.914	0.914	0.969
r2_a	0.913	0.913	0.969
rmse	6.666	6.666	0.927
F	1036.496	799.785	3053.189

Legend: b/se/t

*Discussion.* In the context of a model with heteroskedastic errors, both OLS and WLS estimators are unbiased and consistent, therefore all the estimates are fairly close each other. The parameters' standard errors estimated by OLS in the first column are biased (because of heteroskedastic errors), while those in the second column are robust to heteroskedasticity (hence, reliable). However, being WLS also efficient, the standard errors reported in the third column are remarkably lower than those in the second column.



## 4. Conclusions

To conclude this lecture and to add useful information especially in the panel data context, we compare three OLS estimators with different correction of the standard errors that are available in the `regress` command.

(1) No correction of the standard errors or homoskedastic estimator (`regress`)

$$\text{Var}(\hat{\beta}_{OLS}) = s^2(X'X)^{-1}, \text{ where } s^2 = \frac{1}{N-K} \sum_{i=1}^N \hat{\varepsilon}_i^2.$$

(2) Heteroskedastic-consistent estimator (`regress, robust`)

$$\text{Var}(\hat{\beta}_{robust}) = (X'X)^{-1} \sum_{i=1}^N (\hat{\varepsilon}_i X_i)(\hat{\varepsilon}_i X_i)'(X'X)^{-1}.$$

Sometimes, this matrix is multiplied by  $N/(N-K)$  as a degree-of-freedom adjustment.

(3) Estimator that accounts for clustering into groups, with observations correlated within groups, but independent between groups [`regress, cluster(name_groups)`]

$$\text{Var}(\hat{\beta}_{cluster}) = (X'X)^{-1} \sum_{j=1}^{N_c} \hat{u}_j \hat{u}_j'(X'X)^{-1}, \text{ where } \hat{u}_j = \sum_{i \in j} \varepsilon_i X_i \text{ and } N_c \text{ is the total number of}$$

clusters. The finite sample adjustment is  $(N-1)/(N-K) \times N_c / (N_c - 1)$ .

Note that `cluster` implies `robust` option. The formula for the clustered estimator is simply that of the robust (unclustered) estimator with the individual  $\hat{\varepsilon}_i X_i$  replaced by their sums over each cluster. In other terms, the standard errors are computed based on aggregate  $y$  for the  $N_c$  independent groups.

If the variance of the clustered estimator (3) is smaller than that of the robust (unclustered) estimator (2), it means that the cluster sums of  $\hat{\varepsilon}_i X_i$  have less variability than the individual  $\hat{\varepsilon}_i X_i$ . That is, when you sum the  $\hat{\varepsilon}_i X_i$  within a cluster, some of the variation gets cancelled out, and the total variation is smaller. This means that a big positive is summed with a big negative to produce something small; in other words, there is negative correlation within cluster.

If the number of clusters is very small compared to the overall sample size, it could be that the clustered standard errors (3) are quite larger than the homoskedastic ones (1), because they are computed on aggregate data for few groups.

Interpreting a difference between (1) the OLS estimator and (2) or (3) is trickier. In (1) the squared residuals are summed, but in (2) and (3) the residuals are multiplied by the X's (then for (3) summed within cluster) and then "squared" and summed. Hence, any difference between them has to do with very complicated relationships between the residuals and the X's. If big (in absolute value)  $\hat{\varepsilon}_i$  are paired with big  $X_i$ , then the robust variance estimate will be bigger than the OLS estimate. On the other hand, if the robust variance estimate is smaller than the OLS estimate, it is not clear at all what is happening (in any case, it has to do with some odd correlations between the residuals and the X's).