



## Exercise Class - Econometrics Class 3

**Instructor:** Irene Iodice

**Email:** irene.iodice@malix.univ-paris1.fr

### Ex.1: Recall what we have already done on heteroskedasticity..

Consider the consumption function:

$$\text{cons}_i = \beta_0 + \beta_1 \text{inc}_i + u_i \quad u_i = e_i \sqrt{\text{inc}_i} \quad (1)$$

where  $e_i$  is a random variable with  $E(e_i) = 0$  and  $\text{Var}(e) = \sigma_e^2$ . Assume that  $e_i$  is independent of  $\text{inc}_i$ .

1. Explain the intuition behind the structure of the error.
2. Show that HP3 is satisfied, which is that the zero conditional mean assumption holds.
3. Show that HP5 is violated, which is that the error terms and the regressor are not independent.

### Ex.2: Consequences of heteroskedasticity

Discuss which of the following are consequences of heteroskedasticity?

1. The OLS estimators are inconsistent.
2. The usual F statistic no longer has an F distribution.
3. The OLS estimators are no longer BLUE.

### Ex.3: Example on inference with different standard errors

The following equation was estimated for the fall and second semester students:

$$\begin{aligned} \text{trmgpa} = & -2.12 + .900 \text{ crsgpa} + .193 \text{ cumgpa} + .0014 \text{ tothrs} \\ & (.55) \quad (.175) \quad (.064) \quad (.0012) \\ & [.55] \quad [.166] \quad [.074] \quad [.0012] \\ + & .0018 \text{ sat} - .0039 \text{ hsperc} + .351 \text{ female} - .157 \text{ season} \\ & (.0002) \quad (.0018) \quad (.085) \quad (.098) \\ & [.0002] \quad [.0019] \quad [.079] \quad [.080] \\ & n = 269, R^2 = .465. \end{aligned}$$

Here,  $\text{trmgpa}$  is term GPA,  $\text{crsgpa}$  is a weighted average of overall GPA in courses taken,  $\text{tothrs}$  is total credit hours prior to the semester,  $\text{sat}$  is SAT score,  $\text{hsperc}$  is graduating percentile in high school class,  $\text{female}$  is a gender dummy, and  $\text{season}$  is a dummy variable equal to unity if the student's sport is in season during the fall. The usual and heteroskedasticity-robust standard errors are reported in parentheses and brackets, respectively.

1. Test whether there is an in-season effect on term GPA, using both standard errors. Does the significance level at which the null can be rejected depend on the standard error used?
2. Comment the differences between the standard errors and the robust ones.

### Ex.4: Introduction to GLS

Consider a linear model to explain monthly beer consumption

$$\text{beer} = \beta_0 + \beta_1 \text{inc} + \beta_2 \text{age} + \beta_3 \text{educ} + \beta_4 \text{female} + u$$

with  $E(u|\text{inc}, \text{age}, \text{educ}, \text{female}) = 0$  and  $\text{Var}(u|\text{inc}, \text{price}, \text{educ}, \text{female}) = \sigma^2 \text{inc}^2$ . Write the transformed equation that has a homoskedastic error term and comment.



### Ex.5: Implication of GLS

Consider the model

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (2)$$

with heteroskedastic variance  $Var(u_i) = \sigma_i^2$  and its transformed homoskedastic version  $y_i^* = b_0 \sigma_i^{-1} + b_1 x_i^* + u_i^*$  where  $y_i^* = \sigma_i^{-1} y_i$ ,  $x_i^* = \sigma_i^{-1} x_i$  and  $u_i^* = \sigma_i^{-1} u_i$ .

1. Write the formula for the estimator of  $b_1$  and  $b_0$
2. Show that  $b_0$  and  $b_1$  are equal to the least squares estimators  $\beta_0$  and  $\beta_1$  when  $Var(u_i|x_i) = \sigma^2$ . That is, the error variances are constant.
3. Does a comparison of the formulas for  $b_0$  and  $b_1$  with those for  $\beta_0$  and  $\beta_1$  suggest an interpretation for the new estimators?

### Ex.6: Example of GLS

Consider the model:

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (3)$$

where the  $u_i$  are independent errors with  $E(u_i) = 0$  and  $var(u_i) = \sigma^2 x_i^2$ . Suppose that you have the following five observations  $y = (4, 3, 1, 0, 2)$  and  $x = (1, 2, 1, 3, 4)$ . Find the generalized least squares estimates of  $\beta_0$  and  $\beta_1$ .

### Ex.7: Clusters, robust and standard errors

You want to estimate the following model:

$$sales_{it} = \beta_0 + \beta_1 solvency\_rate_{it} + u_{it} \quad (4)$$

where  $i$  is a firm identifier and  $t$  represent the year. Now you estimate the model with artificial data provided by Petersen (2009) through the 'Sandwich' package of R, making different assumption on the standard errors. Below the graph of data and the estimated regression line and the table summarizing the results.

1. Compare the estimated coefficients of running OLS with classical and robust standard errors, have they changed?
2. Compare the robust standard error with the classical ones (do they increase or decrease?). By looking at the graph try to comment whether you would expect that direction of change.
3. Compare the clustered standard errors, which type of effect is predominant in the individual error term?



Figure 1: Scatterplot with regression line

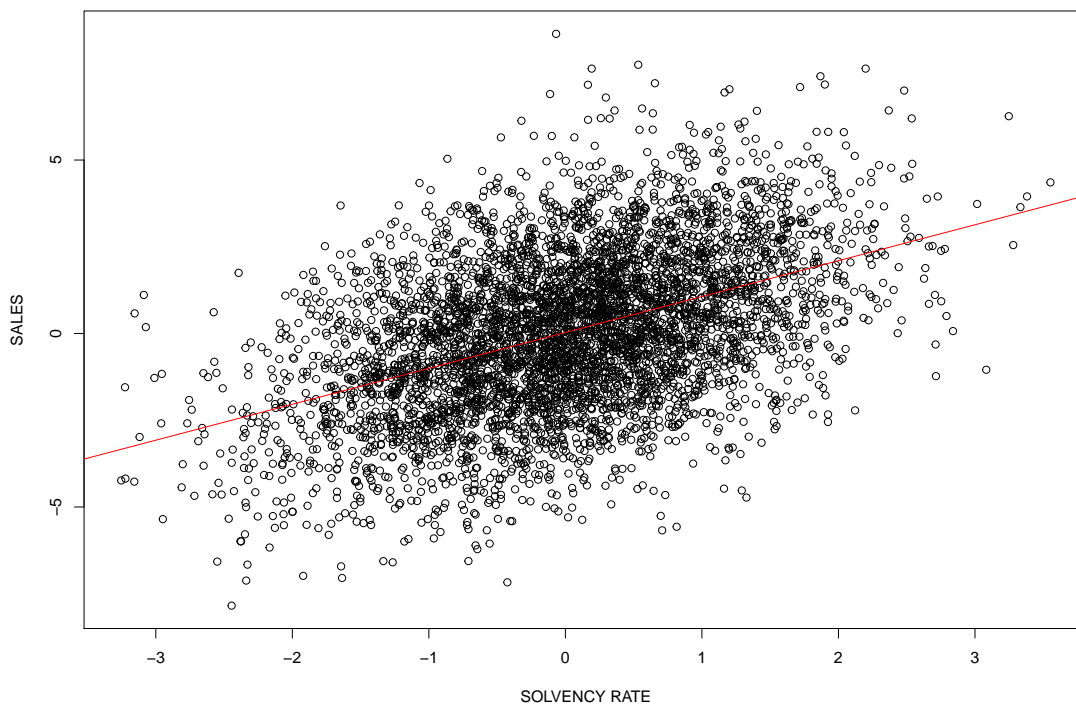


Figure 2: OLS results with classical standard errors

```
Call:
lm(formula = sandw$y ~ sandw$x)

Residuals:
    Min       1Q   Median       3Q      Max
-6.7611 -1.3680 -0.0166  1.3387  8.6779

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02968   0.02836   1.047   0.295
sandw$x      1.03483   0.02858  36.204 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.005 on 4998 degrees of freedom
Multiple R-squared:  0.2078,    Adjusted R-squared:  0.2076
F-statistic: 1311 on 1 and 4998 DF,  p-value: < 2.2e-16
```

Figure 3: OLS results with robust standard errors

```
> coeftest(est1, vcov = vcovHC(est1, "HC1"))

t test of coefficients:

            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.029680   0.028361   1.0465  0.2954
sandw$x      1.034833   0.028395  36.4440 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: Standard errors clustered by groups: (i) firm, (ii) years and (ii) firm\*year

	classical	Firm-cluster	Year-cluster	Firm*Year-cluster
(Intercept)	0.02835932	0.06701270	0.02338672	0.06506392
x	0.02858329	0.05059573	0.03338891	0.05355802

