Econometria 1

División de Economía - CIDE

Dr. Francisco Cabrera

**Actividad 5. Heterocedasticidad**

*\*para una semana*

**Entregable**: Un archivo de Word contestando correctamente a las preguntas abajo presentadas y gravado con su nombre, así como el archivo “Script” de R. Si se le pide explícitamente que incluya el script de R en su respuesta, hágalo.

1. **Consider a linear model to explain monthly beer consumption:**

*Text

Description automatically generated*

1. Write the transformed equation that has a homoscedastic error term. Show this is homoscedastic.
2. **Using the data in GPA3, the following equation was estimated for the fall and second semester students:**

Graphical user interface, text, application

Description automatically generated

Here, *trmgpa* is term GPA, *crsgpa* is a weighted average of overall GPA in courses taken, *cumgpa* is GPA prior to the current semester, *tothrs* is total credit hours prior to the semester, *sat* is SAT score, *hsperc* is graduating percentile in high school class, *female* is a gender dummy, and *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. Do the variables *crsgpa*, *cumgpa*, and *tothrs* have the expected estimated effects? Which of these variables are statistically significant at the 5% level? Does it matter which standard errors are used?
2. Why does the hypothesis H0: b*crsgpa* = 1 make sense? Test this hypothesis against the two-sided alternative at the 5% level, using both standard errors. Describe your conclusions.
3. 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?
4. Consider a model at the employee level, where the unobserved variable *fi* is a “firm effect” to each employee at a given firm *i*. The error term *vi*,*e* is specific to employee *e* at firm *i*. The *composite error* is *ui*,*e* = *fi* + *vi*,*e*, such as in equation (8.28) in Wooldridge 7ed.

Text

Description automatically generated

1. **Use the data set wage1 for this exercise in R studio (pegue su script abajo).**
2. Use OLS to estimate the model relating Wage with education, experience, and tenure.
3. Compute in “manually” in R-studio the Breusch-Pagan Test, including the F-stat and LM.
4. Show that, what you get in b. equals: (where **est\_0** is the object where you save your first regression)

install.packages("lmtest")

library(lmtest)

bptest(**est\_0**)

1. Compute *manually* the alternative White test. In the regression of U-hat2 on the predicted values and the squared predicted values for wage coming from a.
2. Show that your results are the same using *lmtest* in R studio.
3. Use solo educación como predictor e interprete beta-hat ¿es heteroscedástica la relación? Interprete el coeficiente de educ.
4. Luego suponga que Var(Ui|Xk)=s2\*educi Obtenga los estimadores de bk con WLS ¿se soluciona el problema de heteroscedasticidad encontrado en inciso f. al 95% de confianza? Interprete el coeficiente de educ.
5. Verify that the fitted values from d. are all positive and obtain the weighted least squares after estimating hi (i.e. FGLS)
6. Conduct a white test for the FGLS estimation and report your results.
7. Estimate FGLS with heteroskedastic-robust white-Huber SEs. Report your results and discuss why you would like to estimate this? Are the SE too different from those obtained in h. why?
8. **Ejecute el siguiente Código en R Studio:**

**Hint**:

# generate heteroskedastic data

X <- 1:1000

Y <- rnorm(n = 1000, mean = X, sd = 0.6 \* X)

# plot the data

plot(x = X, y = Y,

pch = 19,

col = "steelblue",

cex = 0.8)

# add the regression line to the plot

abline(reg,

col = "darkred",

lwd = 1.5)

**Code:**

t <- c()

t.rob <- c()

# loop sampling and estimation

for (i in 1:10000) {

# sample data

X <- 1:1000

Y <- rnorm(n = 1000, mean = X, sd = 0.6 \* X)

# estimate regression model

reg <- lm(Y ~ X)

# homoskedasdicity-only significance test

t[i] <- linearHypothesis(reg, "X = 1")$'Pr(>F)'[2] < 0.05

# robust significance test

t.rob[i] <- linearHypothesis(reg, "X = 1", white.adjust = "hc1")$'Pr(>F)'[2] < 0.05

}

# compute the fraction of false rejections

round(cbind(t = mean(t), t.rob = mean(t.rob)), 3)

1. Describa brevemente qué es lo que está realizando el código
2. Cambie n a 100 ¿cuál es el resultado en la fracción de rechazos falsos aun con SE robustos? ¿Qué puede concluir sobre los errores estándar robustos?
3. Cambie SD a 1.6 ¿esto qué significa en términos de ajuste del modelo? ¿Qué efecto tiene en la fracción de rechazos falsos aun con SE robustos?
4. Describa formalmente el procedimiento para obtener WLS con estos datos.
5. **We estimate a linear probability model for whether a young man was arrested during 1986:**



1. Using the data in CRIME1, estimate this model by OLS and verify that all fitted values are strictly between zero and one. What are the smallest and largest fitted values?
2. Estimate the equation by weighted least squares, as discussed in Section 8-5.
3. Use the FGLS estimates to determine whether *avgsen* and *tottime* are jointly significant at the 5% level.