ECON471

Fall 2020

Problem Set 6

Due Wednesday December 9, by Midnight CST

Name: Albert Wiryawan

Section: B3

1. Consider the model $y_i = \beta x_i + u_i$ with $E(u_i|x_i) = 0$ and $Var(u_i|x_i) = \sigma^2 x_i^2$. An estimator of β is obtained as follows: $\widetilde{\beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}$.

(i) Derive the expected value of $\tilde{\beta}$ and show that it is unbiased.

*Written in the appendix below

(ii) Derive the weighted least squares estimator of β and show that it is identical to $\tilde{\beta}$. Is $\tilde{\beta}$ BLUE? Without any explicit derivations, compare the efficiency of $\tilde{\beta}$ to the OLS estimator of β.

*Written in the appendix below

2. There are different ways to combine features of the Breusch-Pagan and White tests for heteroskedasticity. One possibility not covered in the lectures is to run the regression

$$\hat{u}_i^2 \text{ on } x_{i1}, x_{i2}, \dots \dots .., x_{ik,}\ \hat{y}_i^2, \quad i = 1, . \dots \dots, n,$$

where the $\hat{u}_i$ are the OLS residuals and the $\hat{y}_i$are the OLS fitted values. Then, we would test joint significance of $x_{i1}, x_{i2}, \dots \dots .., x_{ik}$ and $\hat{y}_i^2$. (Of course, we always include an intercept in this regression.)

(i) What are the degrees of freedom associated with the proposed *F* test for heteroskedasticity?

The degrees of freedom associated with the proposed F-test for heteroskedasticity “k-1”.

(ii) Explain why the R-squared from the regression above will always be at least as large as the R-squared from the Breusch-Pagan regression and the special case of the White test.

The R-squared from the regression above will always be at least as large as the R-squared from the Breusch-Pagan Regression and special case of the white test. Breusch-Pagan regression checks for the linear form of heteroskedasticity while the white test is more of a generic test. As a result, the regression above relies on the assumption that there is no heteroskedasticity. As such, the classical error variance should at least bring us close to an estimator created by the robust estimators, if not larger due to the homoskedasticity assumption.

(iii) Does part (ii) imply that the new test always delivers a smaller *p*-value than either the Breusch-Pagan or special case of the White statistic? Explain.

Since all three tests depend on degree of freedom instead of R-squared, there is no relationship between R-squared and the p-value. As such, since each test has a different value for the degrees of freedom it cannot be said that the new test always delivers a smaller p-value than either Breusch-Pagan or special case of the White statistic.

(iv) Suppose someone suggests also adding $\hat{y}_i$ to the newly proposed test. What do you think of this idea?

This would not be a good idea because the proposed equation already includes $\hat{y}_i{}^2$and adding $\hat{y}_i$ would result in perfect collinearity.

3.

(i) Use the data in HPRICE1.TXT to obtain the heteroskedasticity-robust standard errors for the following model

$$\text{price} = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u.$$

Discuss any important differences with the usual standard errors.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.177e+01  2.948e+01  -0.739  0.46221
lotsize      2.068e-03  6.421e-04   3.220  0.00182 **
sqrft        1.228e-01  1.324e-02   9.275 1.66e-14 ***
bdrms        1.385e+01  9.010e+00   1.537  0.12795
---
```

Figure 1. Screen shot of standard error values without robust standard errors

```
               Estimate  Std. Error t value  Pr(>|t|)
(Intercept) -21.7703086  37.1382106 -0.5862    0.5593
lotsize       0.0020677   0.0012514  1.6523    0.1022
sqrft         0.1227782   0.0177253  6.9267 8.096e-10 ***
bdrms        13.8525219   8.4786250  1.6338    0.1060
---
```

Figure 2. Screen shot of standard error values of fit with robust standard errors

The robust standard error for lotsize is 0.0012514 while the usual standard error is 0.000642. In the ususal model the lotsize feature is significant at the 5% significance level while in the robust model it is not. For sqrft, the standard error in the usual model is 0.01324 while the standard error in the robust model is 0.0177253. However, the attributes are significant in both models. Finally, the standard error for the usual model for bdrms is 9.010 while it is 8.479 for the robust model. Both these features are insignificant in both models.

Code

```
```{r}
setwd("/Users/albertwiryawan/Code/Class_GitRepos/Fall_2020/Econometrics/Problem Set #6/data")
hprice1 =  read.table("hprice1.txt")

attach(hprice1)
lotsize = V4
sqrft = V5
bdrms = V3
price = V1

model = lm(price ~ lotsize + sqrft + bdrms)
summary(model)

model2 = hccm(model, type = "hc1")
mod.HC1 = coeftest(model, vcov.=model2)
mod.HC1

```
```

(ii) Repeat part (i) for the following model

$$\log\text{ (price)} = \beta_0 + \beta_1 \log\text{ }(lotsize) + \beta_2 \log\text{ }(sqrft) + \beta_3 bdrms + u.$$

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1.29704	0.65128	-1.992	0.0497	*
log(lotsize)	0.16797	0.03828	4.388	3.31e-05	***
log(sqrft)	0.70023	0.09287	7.540	5.01e-11	***
bdrms	0.03696	0.02753	1.342	0.1831	

Figure 3. Screen shot of standard error without robust standard error

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1.297042	0.781315	-1.6601	0.1006278	
log(lotsize)	0.167967	0.041473	4.0500	0.0001136	***
log(sqrft)	0.700232	0.103829	6.7441	1.835e-09	***
bdrms	0.036958	0.030601	1.2077	0.2305340	

Figure 4. Screen shot of standard error with robust standard error

The standard errors obtained from the robust model are slightly larger than that of the usual for every feature parameter estimate. The intercept, log(lotsize) and log(sqrft) seem to be significant at the 5% significance level for the usual model while only log(lotsize) and log(sqrft) are significant for the robust

Code

```{r}
setwd("/Users/albertwiryawan/Code/Class_GitRepos/Fall_2020/Econometrics/Problem Set #6/data")
hprice1 =  read.table("hprice1.txt")

attach(hprice1)
lotsize = V4
sqrft = V5
bdrms = V3
price = V1

model = lm(log(price) ~ log(lotsize) + log(sqrft) + bdrms)
summary(model)

model2 = hccm(model, type = "hc1")
mod.HC1 = coeftest(model, vcov.=model2)
mod.HC1

```

(iii) What does this example suggest about heteroskedasticity and the transformation used for the dependent variable?

Looking at the statistics gained from part (ii) of this problem, it can be seen that the use of logarithmic transformations reduces the heteroskedasticity such that the usual OLS standard error and heteroskedasticity-robust standard error across all estimated coefficients are close to one another

4. Use VOTE1.TXT for this exercise.

   (i) Estimate a model with *voteA* as the dependent variable and *prtystrA*, *democA*, log*(expendA)*, and log*(expendB)* as independent variables. Obtain the OLS residuals, $\hat{u}_i$, and regress this on all of the independent variables. Explain why you obtain $R^2 = 0$.

This is because the initial fit of *prtystrA*, *democA*, log*(expendA)*, and log*(expendB)* finds the OLS estimation of the coefficients of the independent variable such that the error term is uncorrelated with every single independent variable. Due to OLS, it is assumed that the sum of the squares of the error term is minimized to 0.

Code

```{r}
setwd("/Users/albertwiryawan/Code/Class_GitRepos/Fall_2020/Econometrics/Problem Set #6/data")
vote1 =  read.table("vote1.txt")

attach(vote1)
voteA = V4
prtystrA = V7
democA = V3
expendA = V5
expendB = V6

model = lm(voteA ~ prtystrA + democA + log(expendA) +log(expendB))
residuals = residuals(model)
model1 = lm(residuals ~ prtystrA + democA + log(expendA) +log(expendB))
summary(model1)
```

(ii) Now, compute the Breusch-Pagan test for heteroskedasticity. Use the *F*-statistic version and report the *p*-value.

The computed F-statistic = 2.330112 and p-value = 0.0581

(iii) Compute the special case of the White test for heteroskedasticity, again using the *F* statistic form. How strong is the evidence against heteroskedasticity now?

The p-value found for the f-statistic is 0.0645 which is higher than the p-value obtained from part (ii) above. This would mean that there is slightly less evidence of heteroskedasticity than that found in the Breusch-Pagan test

5. Use the data in MEAP00_01.TXT to answer this question.

(i) Estimate the model

$$math4 = \beta_0 + \beta_1 lunch + \beta_2 \log(enroll) + \beta_3 \log(exppp) + u$$

by OLS and obtain the usual standard errors and the fully robust standard errors. How do they generally compare?

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    46.1953    17.9148   2.579     0.01 **
lunch          -0.4714     0.0143 -32.953  < 2e-16 ***
log(enroll)   -13.3737     1.7997  -7.431 1.65e-13 ***
log(exppp)      8.5341     1.8365   4.647 3.61e-06 ***
```

Figure 5. Screen shot of standard error found from usual OLS model

```
               Estimate Std. Error  t value  Pr(>|t|)
(Intercept)   46.195329  21.406946   2.1580   0.03106 *
lunch         -0.471381   0.016519 -28.5359 < 2.2e-16 ***
log(enroll)  -13.373727   2.045273  -6.5388 8.031e-11 ***
log(exppp)     8.534120   2.153362   3.9632 7.682e-05 ***
```

Figure 6. Screen shot of standard error found with fully robust standard error

All features are significant at the 5% significance level. The standard errors obtained by the fully robust model is larger than those found in the usual model.

(ii) Apply the special case of the White test for heteroskedasticity. What is the value of the *F* test. What do you conclude?

The test statistic obtained for the F-test was 280 which obtained a p-value of 2e-57 which is extremely close to 0. Due to this low p-value, the null hypothesis that the model has homoscedastic variance is rejected.

Code:

````r
```{r}

setwd("/Users/albertwiryawan/Code/Class_GitRepos/Fall_2020/Econometrics/Problem Set #6/data")
meap01 =  read.table("meap01.txt")

attach(meap01)

math4 = V3
lunch = V5
enroll = V6
exppp = V7
model = lm(math4 ~ lunch + log(enroll) + log(exppp))
summary(model)

model2 = hccm(model, type = "hc1")
mod.HC1 = coeftest(model, vcov.=model2)
mod.HC1


white_lm(model)
```
````

1) $Y_i = \beta x_i + u_i$ $\qquad E(u_i | x_i) = 0$

$\qquad\qquad\qquad\qquad Var(u_i | x_i) = \sigma^2 x_i^2$

$$\beta = \frac{1}{n} \sum_{i=1}^{n} y_i / x_i$$

$$E(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^{n} E(y_i)/x_i = \frac{1}{n} \sum_{i=1}^{n} \frac{E(\beta x_i + u_i)}{x_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{\beta x_i + 0}{x_i} = \beta$$

2) $SS = \sum_{i=1}^{n} w_i u_i^2$

$$= \sum_{i=1}^{n} w_i (y_i - \beta x_i)^2$$

$$\frac{\partial SS}{\partial \beta} = 2 \sum_{i=1}^{n} w_i (y_i - \beta x_i)(-x_i)$$

$$\frac{\partial SS}{\partial \beta} = 0 \Rightarrow \sum_{i=1}^{n} w_i (y_i - \beta x_i) x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} w_i y_i x_i = \tilde{\beta} \sum w_i x_i^2$$

$$\Rightarrow \tilde{\beta} = \sum_{i=1}^{n} w_i y_i x_i / \sum_{i=1}^{n} w_i x_i^2$$

$$E(\tilde{\beta}) = \sum_{i=1}^{n} w_i x_i E(y_i) / \sum x_i^2 = \beta \sum_{i=n} w_i x_i^2 / \sum_{i=1}^{n} w_i x_i^2$$

$E(\hat{\beta}) = \beta$ $\quad \tilde{\beta}$ is BLUE of β

eff $\quad V(\tilde{\beta}) = \sum_{i=1}^{n} x_i^2 V(y_i) / (\sum x_i)^2 = \dfrac{\sum_{i=1} x_i^2 (0 + \sigma^2 x_i^2)}{(\sum x_i^2)^2}$

$$= \frac{\sum_{i=1}^{n} x_i^4 \sigma^2}{(\sum x_i^2)^2}$$

greenroom®

$$V(\hat{\beta}) = \frac{\sum_{i=1}^{n} w_i^2 x_i^2 V(y_i)}{(\sum x_i^2)^2} = \frac{\sigma^2 \sum_{i=1}^{n} w_i x_i^4}{(\sum x_i^2)^2}$$

$$\frac{V(\hat{\beta})}{V(\tilde{\beta})} = \frac{\sigma^2 \sum_{i=1}^{n} x_i^4 / (\sum x_i^2)^2}{\sigma^2 \sum_{i=1}^{n} w_i x_i^4 / (\sum x_i^2)} = \frac{\sum_{i=1}^{n} x_i^4}{\sum_{i=1}^{n} w_i x_i^4} < 0$$

if $\frac{V(\hat{\beta})}{V(\tilde{\beta})} < 0$, $\hat{\beta}$ is better then $\tilde{\beta}$

if $\frac{V(\hat{\beta})}{V(\tilde{\beta})} > 0$, $\tilde{\beta}$ is better then $\hat{\beta}$