Assignment 8

Chitresh Kumar

```
if(!require("pacman")) install.packages("pacman")
pacman::p_load(tidyverse, reshape, gplots, ggmap, RStata,haven,
                data.table,margins,pastecs,MASS,lmtest,broom,car,stargazer,sa
ndwich,knitr)
search()
theme_set(theme_classic())
```

```
pub<-read_dta('pubexp.dta')
head(pub)
```

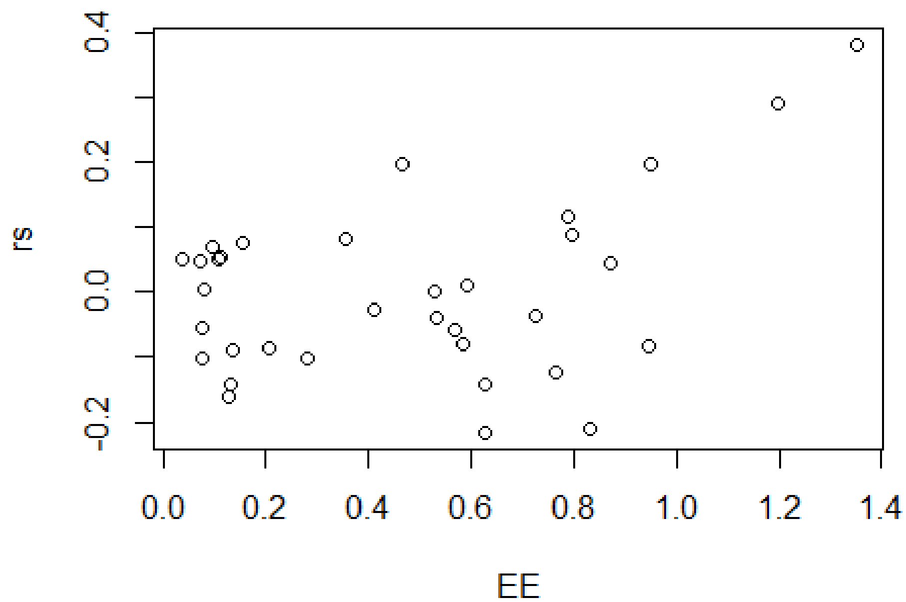
```
## # A tibble: 6 x 3
##      ee   gdp     p
##   <dbl> <dbl> <dbl>
## 1  0.34  5.67  0.36
## 2  0.22 10.1   2.9
## 3  0.32 11.3   2.39
## 4  1.23 18.9   3.44
## 5  1.81 20.9   3.87
## 6  1.02 22.2  10.7
```

PART b

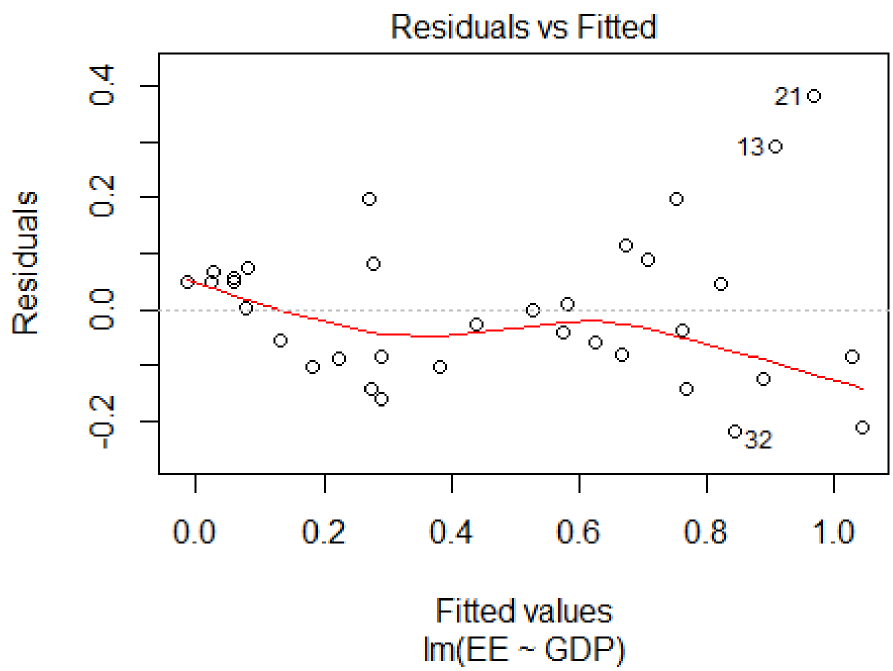
```
EE<-pub$ee/pub$p
GDP<-pub$gdp/pub$p
lm1<-lm(EE~GDP,data=pub)
summary(lm1)
```

```
##
## Call:
## lm(formula = EE ~ GDP, data = pub)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21682 -0.08804 -0.01401  0.06517  0.38156
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.124573   0.048523  -2.567   0.0151 *
## GDP          0.073173   0.005179  14.128 2.65e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1359 on 32 degrees of freedom
## Multiple R-squared:  0.8618, Adjusted R-squared:  0.8575
## F-statistic: 199.6 on 1 and 32 DF,  p-value: 2.65e-15
```

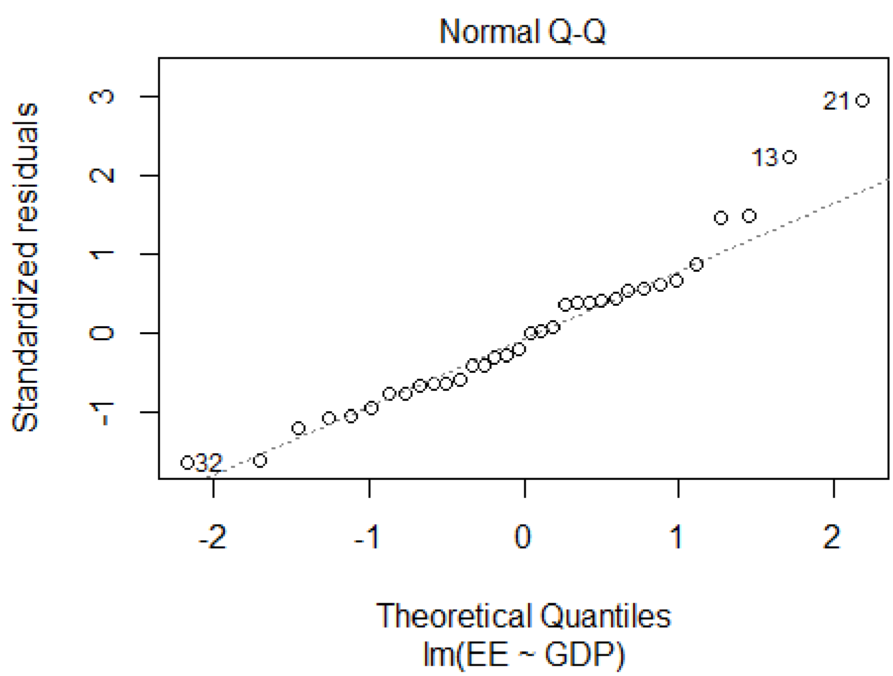
```
rs<-resid(lm1)
plot(EE,rs)
```



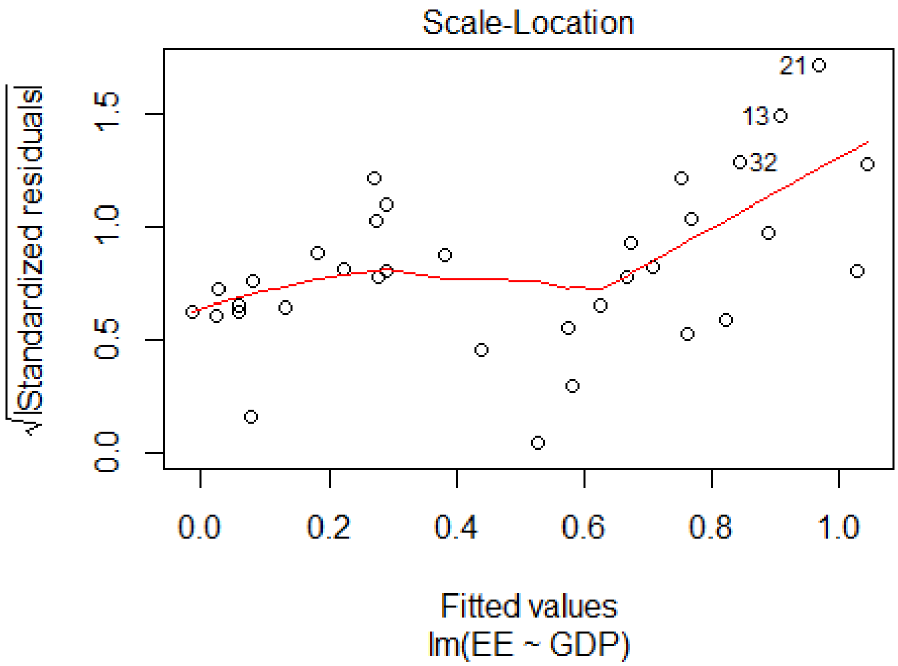

```
plot(lm1)
```

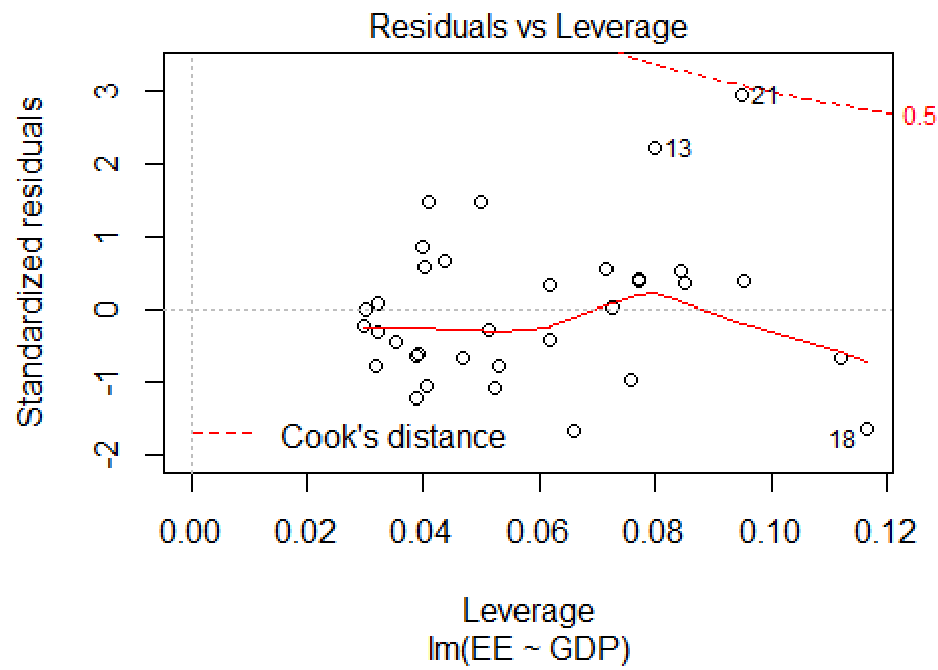


Residuals vs Fitted
Residuals
Fitted values
lm(EE ~ GDP)
0.4
0.2
0.0
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
21
13
32



Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(EE ~ GDP)
3
2
1
0
-1
-2
-1
0
1
2
21
13
32





PART c

```
ressq<-rs^2
GDP_S<-GDP^2
```

```r
lm2<-lm(ressq~GDP+GDP_S,data=pub)
glm2<-glance(lm2)
Rsq<-glm2$r.squared
chisq<-34*Rsq
pval<-1-pchisq(chisq,1)
print(chisq)
```

```
## [1] 9.961449
```

```r
print(pval)
```

```
## [1] 0.001598522
```

PART d

```r
cov1 <- hccm(lm1, type="hc1")
pub.HC1 <- coeftest(lm1, vcov.=cov1)
kable(tidy(pub.HC1))
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.1245728	0.0404140	-3.08242	0.0042041
GDP	0.0731732	0.0062116	11.78005	0.0000000

PART e

```r
w<-1/GDP
lmwls<-lm(EE~GDP,weights=w,data=pub)
summary(lmwls)
```

```
##
## Call:
## lm(formula = EE ~ GDP, data = pub, weights = w)
##
## Weighted Residuals:
##       Min        1Q    Median        3Q       Max
## -0.072028 -0.038561 -0.008488  0.027706  0.105415
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.092921   0.028904  -3.215  0.00298 **
## GDP          0.069321   0.004412  15.713  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04451 on 32 degrees of freedom
## Multiple R-squared:  0.8853, Adjusted R-squared:  0.8817
## F-statistic: 246.9 on 1 and 32 DF,  p-value: < 2.2e-16
```

8.4 a) Income vs Residual plot shows heteroskedasticity. As income increases the variances of residual also increases

Age vs Residual plot does not show any pattern. Hence, the 'Age' is not causing heteroskedasticity

(c) Distance in miles increase with increase in income. β_2 values = +ve
Similarly, the distance in miles increase with increase in age. β_3 values = +ve

The distance in miles decreases with increase in number of kids. β_4 = -ve

8.12 $$y_i = \beta_1 + \beta_2 x_i + e_i$$

$$y_i = \left(\frac{EE_i}{P_i}\right)$$

$$x_i = \left(\frac{GDP_i}{P_i}\right)$$

(a) GDP is related the money generated in the economy. The countries having higher GDP will have more flexibility of spending because the amount of spending on education (EE) will be larger when compared to countries with low gdp

(b) $$\hat{EE} = -0.125 + 0.0732\, GDP$$

from the plot of EE vs residual, we cannot see any pattern. Hence we conclude that there is no heteroskedasticity.

(c) with white test,

$EE = \alpha_1 + \alpha_2 \, GDP + \alpha_3 \, GDP^2 + \gamma_i$

where $GDP = GDP_i / P_i$

$H_0 : \alpha_2 = \alpha_3 = 0$

$H_1 : \alpha_2 \neq 0 , \alpha_3 \neq 0$

$R^2 = 0.29298$

$\chi^2_{stat} = 9.962$

$\chi^2_{critical} = \chi^2_{(0.95, 2)} = 5.99$

$\chi^2_{stat} > \chi^2_{critical}$

∴ We reject the null, hence the heteroskedasticity exists.

(d) White's robust standard errors

Intercept = 0.0404
GDP = 0.0062

CI with OLS se.

$b_2 \pm t_{critical} \times se(b_2)$
$= 0.0732 \pm 2.0369 \times 0.0052$
$= 0.0838 , 0.0626$

CI with White robust SE

$b_2 \pm t_{critical} \times se(b_2)$
$= 0.0732 \pm 2.0369 \times 0.0062$
$= 0.0858, 0.0606$

The CI for white robust SE is more than OLS se. Therefore, the precision of estimate from OLS is overstated.

(e) Assumption : $var(e_i) = \sigma^2 x_i$

$\widehat{EE} = -0.0929 + 0.0693 \ GDP$

95% CI for β_2 :

$b_2 \pm t_{critical} \ se(b_2) = 0.0693 \pm$

$= 0.0783, 0.0603$

CI is smaller than white robust standard errors. The precision of estimate has improved when we use weighted least squares