P4&5

November 6, 2021

[1]:
```
#4
```

[2]:
```
data_SV <- read.table("Table6.9.txt",
                      head = TRUE,
                      sep = "\t")
#data_SV
```

[3]:
```
mod_sv1 <- lm(Y ~ X, data = data_SV, weights = 1/(X^2))
print(summary(mod_sv1))
```

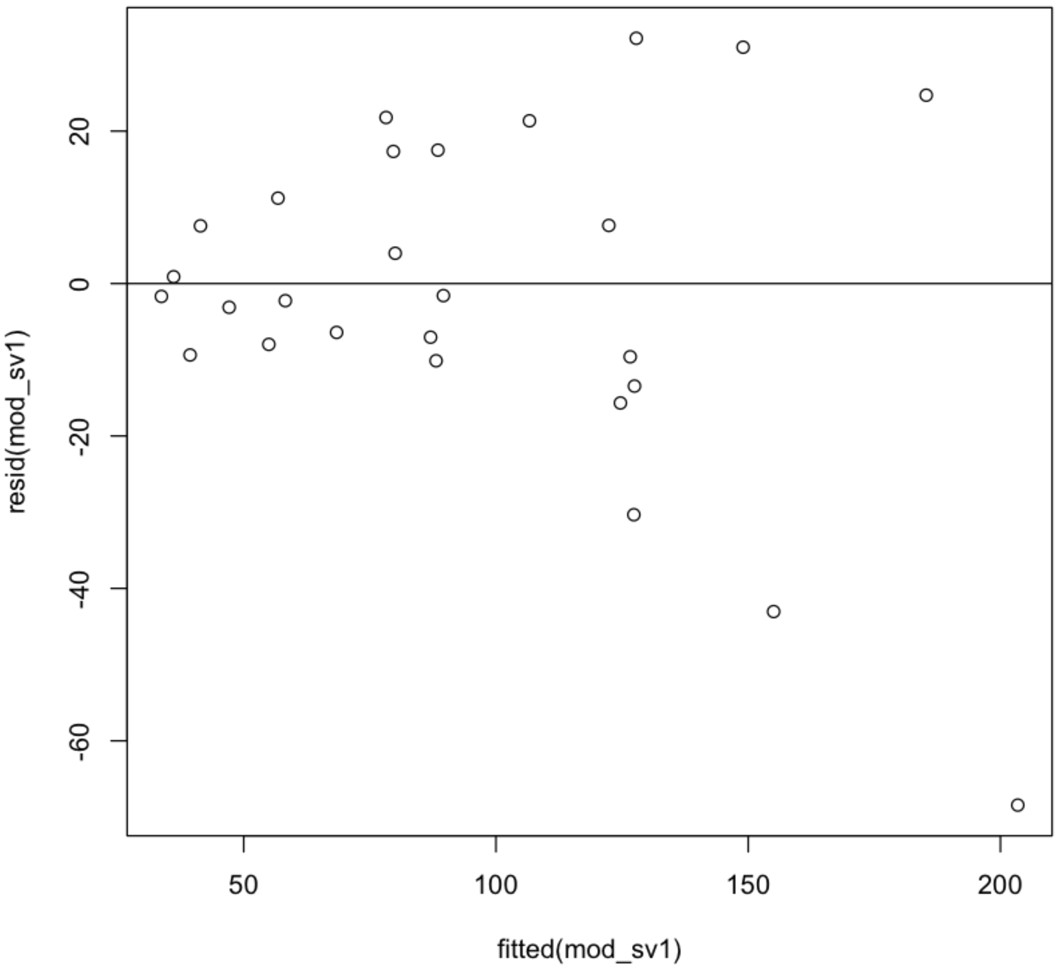
```
Call:
lm(formula = Y ~ X, data = data_SV, weights = 1/(X^2))

Weighted Residuals:
      Min        1Q    Median        3Q       Max
-0.041477 -0.013852 -0.004998  0.024671  0.035427

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.803296   4.569745   0.832    0.413
X           0.120990   0.008999  13.445 6.04e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02266 on 25 degrees of freedom
Multiple R-squared:  0.8785,    Adjusted R-squared:  0.8737
F-statistic: 180.8 on 1 and 25 DF,  p-value: 6.044e-13
```

[4]:
```
plot(fitted(mod_sv1), resid(mod_sv1)) + abline(0,0)
```



Regression equation: $Y = 3.80 + 0.12X + \varepsilon$

[5]:
```
data_SV$Y_R <- data_SV$Y/data_SV$X
data_SV$x_R <- 1/data_SV$X
mod_sv2 <- lm(Y_R ~ x_R, data = data_SV) #WLS
print(summary(mod_sv2))
```

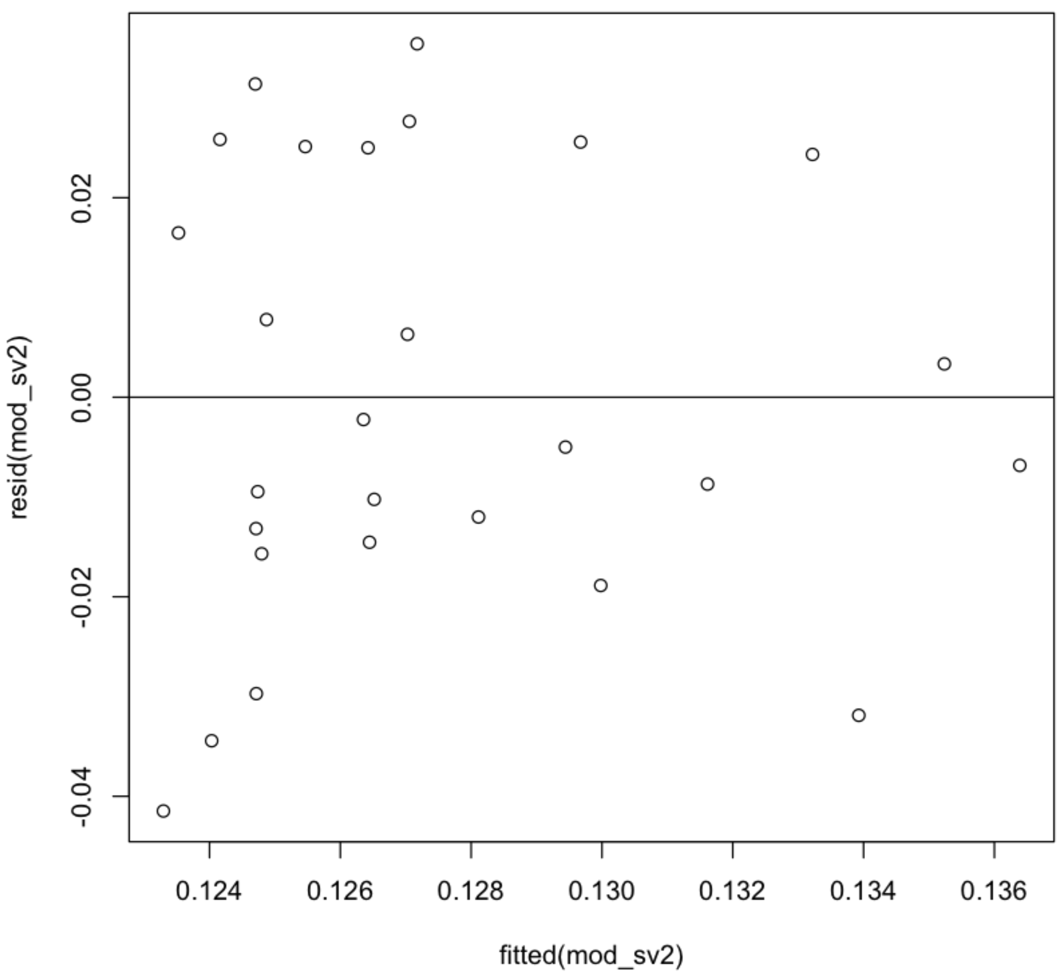
```
Call:
lm(formula = Y_R ~ x_R, data = data_SV)

Residuals:
      Min        1Q    Median        3Q       Max
-0.041477 -0.013852 -0.004998  0.024671  0.035427
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.120990   0.008999  13.445 6.04e-13 ***
x_R         3.803296   4.569745   0.832    0.413
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02266 on 25 degrees of freedom
Multiple R-squared:  0.02696,	Adjusted R-squared:  -0.01196
F-statistic: 0.6927 on 1 and 25 DF,  p-value: 0.4131
```

[6]:
```r
plot(fitted(mod_sv2), resid(mod_sv2)) + abline(0,0)
```



Regression equation: $Y/X = 0.12 + 3.80/X + \varepsilon$

(c)

As shown above, the two methods are equivalent. However, they don't have the same effect in terms of removing heteroscedasticity. It may be that the relationship between the residual variance and X is incorrectly assumed, which leads to incorrect weights. We can use other methods (e.g. Glejser test) to find an adequate weight.

[7]:
```
#5
```

[8]:
```
data_SV <- read.table("Table6.9.txt",
                      head = TRUE,
                      sep = "\t")
mod_sv1 <- lm(Y ~ X, data = data_SV, weights = 1/(X^2))

mod_sv3 <- lm(Y ~ X,data = data_SV)#ols regression
print(summary(mod_sv3))# results of ols with out data transformation
```

```
Call:
lm(formula = Y ~ X, data = data_SV)

Residuals:
    Min      1Q  Median      3Q     Max
-53.294  -9.298  -5.579  14.394  39.119

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.44806    9.56201   1.511    0.143
X            0.10536    0.01133   9.303 1.35e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21.73 on 25 degrees of freedom
Multiple R-squared:  0.7759,    Adjusted R-squared:  0.7669
F-statistic: 86.54 on 1 and 25 DF,  p-value: 1.35e-09
```

[9]:
```
print(summary(mod_sv1))# results of wls
```

```
Call:
lm(formula = Y ~ X, data = data_SV, weights = 1/(X^2))

Weighted Residuals:
      Min        1Q    Median        3Q       Max
```

```
-0.041477 -0.013852 -0.004998  0.024671  0.035427

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.803296   4.569745   0.832    0.413
X           0.120990   0.008999  13.445 6.04e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02266 on 25 degrees of freedom
Multiple R-squared:  0.8785,    Adjusted R-squared:  0.8737
F-statistic: 180.8 on 1 and 25 DF,  p-value: 6.044e-13
```

As shown above, the wls method yields smaller variances.