STAC67: Regression Analysis

Lecture 22

Sohee Kang

Apr. 1, 2021

Chapter 11

Building the Regression Model III: Remedial Measures

Unequal (Independent) Error Variances - Weighted Least Squares (WLS) (σ^2 not constant)

- Case 1 - Error Variances known exactly (VERY rare)
- Case 2 - Error Variances known up to a constant
  - Occasionally information known regarding experimental units regarding the relative magnitude (unusual)
  - If "observations" are means of different numbers of units (each with equal variance) at the various X levels, Variance of observation i is σ^2/n_i where n_i is known
- Case 3 - Variance (or Standard Deviation) is related to one or more predictors, and relation can be modeled (see Breusch-Pagan Test)
- Case 4 - Ordinary Least Squares with estimated variances

Checking Equal Variance

(Bartlette's Test, Levene's Test) can be used as well

- Two tests for equal variance are the Brown-Forsyth test and the Breusch-Pagan (aka Cook-Weisberg) test.
- Breusch-Pagan Test(aka Cook-Weisberg Test) - - Fits a regression of the squared residuals on X and tests whether the variance is related to X.

1) Fit model $Y \sim X$
2) Get residuals
3) Fit $e_i^2 \sim X$
4) Get SSR_{e^2}

$$H_0 : Var(\varepsilon_i) = \sigma^2 \qquad vs \qquad H_a : Var(\varepsilon_i) = \sigma^2 \cdot h(\gamma \cdot X_i)$$

- When the regression of the squared residuals is fit, we obtain SSR_{e2} , the regression sum of squares. The test is conducted as follows, where SSE is the Error Sum of Squares for the original regression of Y on X.

Test statistic $X^2 = \dfrac{SSR_{e_2}}{\left(\frac{SSE}{n}\right)^2} \sim \chi^2_1$

WLS - Case 1 Known Variances - I

$$Y_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma_i^2)$$

$$Cov(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j$$

$$Var(\varepsilon) = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{bmatrix}$$

MLE: $$L(\beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(Y_i - \beta_0 - \beta_1 x_1 - \cdots \beta_p x_p)^2\right)$$

Set $W_i = \frac{1}{\sigma_i^2}$

$$= \prod_{i=1}^{n} \sqrt{\frac{w_i}{2\pi}} \cdot \exp\left(-\frac{1}{2}\sum_{i=1}^{n} w_i \cdot (Y_i - \hat{Y}_i)^2\right)$$

Thus we need to max $Q(w) = \sum_{i=1}^{n} w_i \cdot (Y_i - \hat{Y}_i)^2$

WLS - Case 1 Known Variances - II

Since case 1, Variances are known.

- $W = \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{bmatrix} = W'$

- $Var(Y) = Var(\varepsilon) = W^{-1}$

- So normal equation can be written as:

$$(X'WX)\hat{B}_w = X'WY$$

$$\hat{B}_w = (X'WX)^{-1}X'WY$$

$$= AY, \quad A = (X'WX)^{-1}X'W$$

- $E(\hat{B}_w) = E(AY) = A\,E(Y) = (X'WX)^{-1}X'WXB = B$

Thus $\hat{B}_w$ is unbiased estimator of B.

- $Var(\hat{B}_w) = A \; Var(Y) A'$

$$= (X'WX)^{-1} X'WW^{-1} \left((X'WX)^{-1} X'W \right)'$$

$$= (X'WX)^{-1}$$

WLS - Case 2 - Variance Known up to Constant - I

- Each observation is not a simple measure, but an average of n_i actual measures and the orginal measure each have σ^2.

$$Var(\epsilon_i) = Var(\bar{Y}_i) = \frac{\sigma^2}{n_i} \Rightarrow Var(Y) = \sigma^2 \begin{bmatrix} \frac{1}{n_1} & & 0 \\ & \frac{1}{n_2} & \\ 0 & \ddots & \frac{1}{n_n} \end{bmatrix}$$

$n_i \equiv$ number of replicates at ith level of x

$$w_i = \frac{1}{\sigma_i^2} = \frac{n_i}{\sigma^2} \Rightarrow W = \frac{1}{\sigma^2} \begin{bmatrix} n_1 & & 0 \\ & \ddots & \\ 0 & & n_n \end{bmatrix} = \frac{1}{\sigma^2} W^*$$

$$X'WX = X' \frac{1}{\sigma^2} W^* X = \frac{1}{\sigma^2} (X'W^*X)$$

$$(X'WX)^{-1} = \sigma^2 (X'W^*X)^{-1}$$

$$X'WY = \frac{1}{\sigma^2} X'W^*Y$$

So $\hat{B}_W = (X'WX)^{-1} X'WY = (X'W^*X)^{-1} X'W^*Y$

$$Var(\hat{B}_W) = (X'WX)^{-1} = \sigma^2 (X'W^*X)^{-1}$$

$$\hat{var}(\hat{\beta}_w) = MSE_w \, (X' W^* X)^{-1}$$

$$MSE_w = \sum_i w_i^* \, (Y_i - \hat{Y}_i)^2 \Big/ (n-p')$$

Example: Cholesterol Drug Dose-Response Study

- Study of Drug Rosuvastin in Japanese Patients with high cholesterol
- 6 Doses - 1,2.5,5,10,20,40
- Response - % Change in LDL Cholesterol @ week 12
- Data Reported - Group Means by Dose
- Sample sizes varies by dose.
- Assuming equal variance among individual patients:

DoseGrp	ChgLDL	Dose	ln(Dose)	
1	-35.8	1	0.0000	15
2	-45.0	2.5	0.9163	17
3	-52.7	5	1.6094	12
4	-49.7	10	2.3026	14
5	-58.2	20	2.9957	18
6	-66.0	40	3.6889	13

R codes

```r
dLDL <- c(-35.8,-45.0,-52.7,-49.7,-58.2,-66.0)
DOSE <- c(1,2.5,5,10,20,40)
ni <- c(15,17,12,14,18,13)
lnDOSE <- log(DOSE)
cholest <- data.frame(dLDL,lnDOSE,ni)
attach(cholest)

# Usual linear Model
cholest.ols <- lm(dLDL ~ lnDOSE)
summary(cholest.ols)
```

```
##
## Call:
## lm(formula = dLDL ~ lnDOSE)
##
## Residuals:
##      1      2      3      4      5      6
##  1.190 -1.208 -3.763  4.382  1.027 -1.627
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -36.990      2.377 -15.560 9.96e-05 ***
## lnDOSE        -7.423      1.041  -7.134  0.00204 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.16 on 4 degrees of freedom
## Multiple R-squared:  0.9271, Adjusted R-squared:  0.9089
## F-statistic: 50.89 on 1 and 4 DF,  p-value: 0.002042
```

```r
confint(cholest.ols)
```

R codes

```r
# Weighted Least Squre
cholest.wls <- lm(dLDL ~ lnDOSE, weights=ni)
summary(cholest.wls)
```

```
##
## Call:
## lm(formula = dLDL ~ lnDOSE, weights = ni)
##
## Weighted Residuals:
##       1       2       3       4       5       6
##   4.488  -5.291 -13.410  15.869   3.620  -6.615
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -36.9588     2.2441 -16.469 7.96e-05 ***
## lnDOSE       -7.3753     0.9892  -7.456  0.00173 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.58 on 4 degrees of freedom
## Multiple R-squared:  0.9329, Adjusted R-squared:  0.9161
## F-statistic: 55.59 on 1 and 4 DF,  p-value: 0.001729
```

```r
confint(cholest.wls)
```

```
##                 2.5 %     97.5 %
## (Intercept) -43.18954 -30.728138
## lnDOSE      -10.12186  -4.628755
```

op!

```r
vcov(cholest.wls)
```

```
##             (Intercept)      lnDOSE
```

Models

$$\bar{Y}_j = \beta_0 + \beta_1 \ln X_j + \varepsilon_j \qquad \varepsilon_j \sim N\left(0, \frac{\sigma^2}{n_j}\right)$$

WLS - Case 3 - Estimated Variances

- In most cases, the variances are unknown, and must be estimated.
- In this case, the squared residuals (variance) or absolute residuals (standard deviation) are regressed against one or more of the predictor variables or the mean (fitted values)

Estimation process:

1. Fit the regression model by unweighted least squres and analyze the residuals
2. Estimate the variance function or the standard deviation function by regressing either the squared residuals or the absolute residuals on the appropriate predictor(s)
3. Use the fitted values from the estimated variance or standard deviation function to obtain the weights w_i.
4. Estimate the regression coefficients using these weights
5. If the estimated coefficients differ substantially, then iterate the weighted least squares process

Models

① Variance case : $\hat{v}_i = \hat{\delta}_0 + \hat{\delta}_1 \hat{Y}_1$

② SD Case : $\hat{s}_i = \hat{\delta}_0 + \hat{\delta}_1 \hat{Y}_i$

①

- Regress $e_i^2 \sim \hat{Y}_i$ (or X)
  - fitted value $= \hat{v}_i$
- $\hat{w}_i = \frac{1}{\hat{v}_i}$

②

- Regress $|e_i| \sim \hat{Y}_i$ (or X)
  - fitted value $= \hat{s}_i$
- $\hat{w}_i = \frac{1}{\hat{s}_i^2}$

Then $\hat{B}_{w_0} = (X'WX)^{-1}X'WY$, $\hat{W} = \begin{bmatrix} \hat{w}_1 & & 0 \\ & \ddots & \\ 0 & & \hat{w}_n \end{bmatrix}$

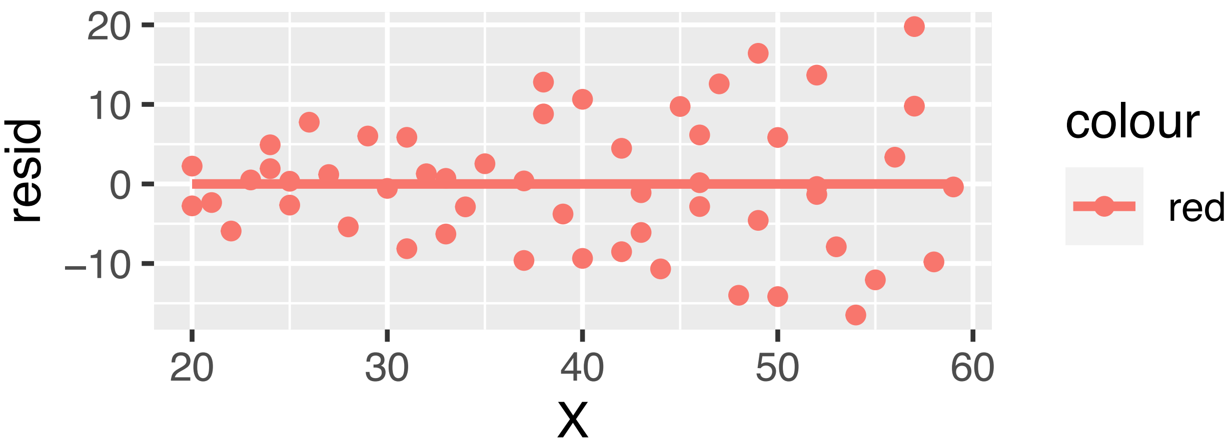
Repeat until $\hat{B}_{w_0}$ converge.

$\hat{V}(\hat{B}_w) = MSE_{\hat{w}}(X'\hat{W}X)^{-1}$ $MSE_{\hat{w}} = \frac{1}{n-p'}(Y - X\hat{B}_w)'(Y - X\hat{B}_w)$

Example: Textbook Table 11.1

A health researcher, interested in studying the relationship between diastolic blood pressure and age among healthy adult women 20-60 years old, collected data on 54 subjects.

```
df = read.table("Table11-1.txt", header=T)
fit = lm(Y~X, data=df)
df$resid = fit$residuals
library(ggplot2)
ggplot(data=df, aes(X, resid, col="red")) + geom_point() + geom_smooth(method="lm", se=FALSE)
```



R codes

```
df$s.hat= abs(df$resid)
fit2 = lm(s.hat~X, data=df)
df$var.s = (predict(fit2))^2

fit3 = lm(Y~X, weights =1/var.s,data=df)
summary(fit3)
```

```
##
## Call:
## lm(formula = Y ~ X, data = df, weights = 1/var.s)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0230 -0.9939 -0.0327  0.9250  2.2008
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 55.56577    2.52092  22.042  < 2e-16 ***
## X            0.59634    0.07924   7.526 7.19e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.213 on 52 degrees of freedom
## Multiple R-squared:  0.5214, Adjusted R-squared:  0.5122
## F-statistic: 56.64 on 1 and 52 DF,  p-value: 7.187e-10
```

```
df$resid.2 = residuals(fit3)
```

Comparison with OLS

```
# OLS coefficients
coefficients(fit)
```

```
## (Intercept)           X
##  56.1569294   0.5800308
```

```
# Weighted Least Squre coefficients
coefficients(fit3)
```

```
## (Intercept)           X
##  55.5657664   0.5963417
```

Example: Construction Plant Maintenance Costs

- Edwards, Holt, and Harris (2000), studied the relationship between Maintenance Costs (Y) and $p = 4$ predictors:
- Machine Weight (X_1), and indicators for Industry Type ($X_2 = 1$=opencast coal, 0 if slate), Machine Type ($X_3 = 1$ if front shovel, 0 if backacter), and Company attitude to Oil Analysis ($X_4 = 1$, if regular use, 0 if not).

```
cmc <- read.table("const_maint.txt",header=F,col.names=c("mach_id","mach_cost","coal","front_shov","use_oil","m
attach(cmc)

mod1 <- lm(mach_cost ~ mach_wt + coal + front_shov + use_oil)
yhat.1 <- predict(mod1); e.1 <- resid(mod1)

head(cmc)
```

```
##   mach_id mach_cost coal front_shov use_oil mach_wt
## 1       1     6.068    1          0       0   16.60
## 2       2     4.602    1          0       0   20.37
## 3       3     3.282    1          0       0   20.37
## 4       4     2.192    1          1       0   20.37
## 5       5     2.572    1          1       0   20.37
## 6       6     4.142    1          0       0   20.37
```

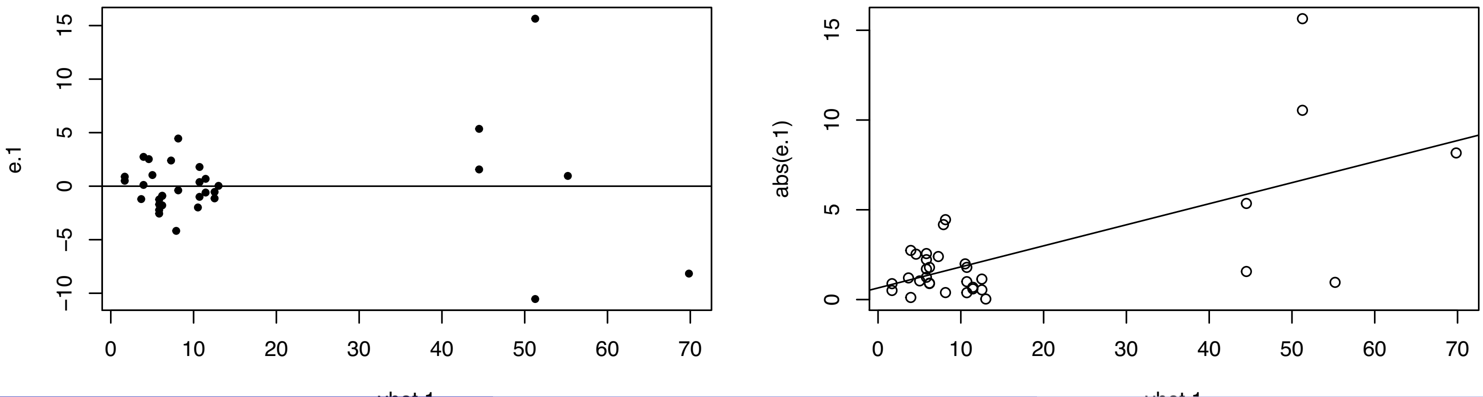
Example

```
par(mfrow=c(2,2))
plot(yhat.1, e.1, pch=20)
abline(h=0)
plot(yhat.1, abs(e.1))
mod2 = lm(abs(e.1) ~ yhat.1)
abline(mod2)

library(lmtest)
bptest(mod1, studentize = F)
```

< 0.05 Reject that $Var(\varepsilon_i) = \sigma^2$

```
##
##  Breusch-Pagan test
##
## data:  mod1
## BP = 51.742, df = 4, p-value = 1.562e-10
```



Example

```
var.s = (mod2$fitted.values)^2

mod3 = lm(mach_cost ~ mach_wt + coal + front_shov + use_oil, weight=1/var.s)

e.2 = mod3$residuals
yhat.2 = predict(mod3)
mod4 = lm(abs(e.2) ~ yhat.2)
var.s2 = (mod4$fitted.values)^2

mod5 = lm(mach_cost ~ mach_wt + coal + front_shov + use_oil, weight=1/var.s2)

cbind(coefficients(mod1),coefficients(mod3), coefficients(mod5))
```

```
##                   [,1]       [,2]       [,3]
## (Intercept) -6.0698953 -5.9341947 -5.9325775
## mach_wt      0.2166055  0.2201925  0.2205674
## coal         7.5019251  7.3464820  7.3281107
## front_shov  -4.1582322 -3.9818716 -3.9524789
## use_oil      4.3910957  3.5340154  3.4948147
```

very similar

Case 4 - OLS with Estimated Variances

$E(Y) = XB \quad Var(Y) = Var(\varepsilon) = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$

$\hat{B} = (X'X)^{-1}X'Y = AY$

$\hat{Var}(\varepsilon) = S_0 = \begin{bmatrix} e_1^2 & & & 0 \\ & e_2^2 & & \\ & & \ddots & \\ 0 & & & e_n^2 \end{bmatrix}$

$S_0 \; \hat{Var}(\hat{B}) = A\,Var(Y)\,A' = A\,S_0\,A' \Rightarrow$ white's estimator

Weighted Least Squares from the view of transformation of OLS

1. WLS is a transformation of OLS

Ex. $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, $\quad Var(\varepsilon_i) = X_i^2 \cdot \sigma^2$

So let $\varepsilon_i^* = \dfrac{\varepsilon_i}{X_i}$, $Y_i^* = \dfrac{Y_i}{X_i}$, $X_i^* = \dfrac{1}{X_i}$

Then $Y_i^* = \beta_1 + \beta_0 X_i^* + \varepsilon_i^*$ $\quad Var(\varepsilon_i^*) = \sigma^2$

The normal equation of this transformed OLS

$$S(\beta_0, \beta_1) = \sum (Y_i^* - \beta_1 - \beta_0 X_i^*)^2$$
$$= \sum \left(\frac{1}{X_i}\right)^2 (Y_i - \beta_0 - \beta_1 X_i)^2$$

This is the normal equation of a WLS with $w_i = \left(\dfrac{1}{X_i}\right)^2$

Weighted Least Squares from the view of transformation of OLS

In general for $Y = XB + \varepsilon$ where $Var(\varepsilon) = W^{-1}\sigma^2$

Let $W^{\frac{1}{2}}$ be a diagonal matrix with elements : $\sqrt{w_i}$.

Then $Var(W^{\frac{1}{2}}\varepsilon) = W^{\frac{1}{2}} Var(\varepsilon) W^{\frac{1}{2}}$

$= \sigma^2 W^{\frac{1}{2}} W^{-1} W^{\frac{1}{2}}$

$= \sigma^2 I$

So the transformation $Y^* = W^{\frac{1}{2}} Y$

$X^* = W^{\frac{1}{2}} X$

$\varepsilon^* = W^{\frac{1}{2}} \varepsilon$

gives the usual least square model.