Weighted Least Squares

Dr. Kiah Wah Ong

ILLINOIS TECH

Weighted Least Squares

Recall the major assumptions we have made in linear regression models

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i, \quad i = 1, 2, \cdots, n$$

are

- The relationship between the response and regressors is linear.
- The error terms ϵ_i have mean zero.
- The error terms ϵ_i have constant variance σ^2 (homoscedasticity)
- The error terms ϵ_i are normally distributed.
- The error terms ϵ_i and ϵ_j are uncorrelated for $i \neq j$.
- The regressors $x_1, \cdots, x_k$ are nonrandom.
- The regressors $x_1, \cdots, x_k$ are measured without error.
- The regressors are linearly independent.

Weighted Least Squares

When the errors are uncorrelated but have unequal variances, beside the variance-stabilizing transformation, **weighting** is another effective way of handling the non-constant variance.

This method is particularly useful when we know a priori how the variance looks like.

We will use the following example to illustrate this technique.

Consider the following model

$$y = \beta_0 + \beta_1 x + \epsilon, \quad \epsilon \sim \mathrm{N}(0, x\sigma^2)$$

where the predictor takes only positive values.

Weighted Least Squares

Taking a sample of n data points (x_i, y_i), we can write the model as

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim \mathrm{N}(0, x_i\sigma^2).$$

In this case, $\mathrm{Var}(y_i) = \mathrm{Var}(\epsilon_i) = x_i\sigma^2$.

To stabilize the variance, we let

$$\epsilon_i' = \frac{\epsilon_i}{\sqrt{x_i}} \sim \mathrm{N}(0, \sigma^2), \quad i = 1, \cdots, n$$

Weighted Least Squares

We then rewrite the original model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

as follow

$$\frac{y_i}{\sqrt{x_i}} = \beta_0 \frac{1}{\sqrt{x_i}} + \beta_1 \sqrt{x_i} + \epsilon_i', \quad i = 1, \cdots, n$$

where

$$\epsilon_i' = \frac{\epsilon_i}{\sqrt{x_i}} \sim \mathrm{N}(0, \sigma^2), \quad i = 1, \cdots, n$$

The sum of the weighted errors are then given by

$$\sum_{i=1}^{n} (\epsilon_i')^2 = \sum_{i=1}^{n} \left(\frac{y_i}{\sqrt{x_i}} - \beta_0 \frac{1}{\sqrt{x_i}} - \beta_1 \sqrt{x_i} \right)^2 = \sum_{i=1}^{n} \frac{1}{x_i} (y_i - \beta_0 - \beta_1 x_i)^2$$

Weighted Least Squares

The weighted least-squares function which we want to minimize is then given by

$$
\begin{aligned}
S(\beta_0, \beta_1) &= \sum_{i=1}^{n} \frac{1}{x_i}(y_i - \beta_0 - \beta_1 x_i)^2 \\
&= \sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i)^2
\end{aligned}
$$

with $w_i = 1/x_i$ as weight that can help stabilized the variance of the error terms.

Using the same method as laid out before, there is a way for us to compute $\hat{\beta}_0$ and $\hat{\beta}_1$ from the sample.

Weighted Least Squares

With $w_i = 1/x_i$ and writing

$$\overline{x}^{(w)} = \frac{\sum w_i x_i}{\sum w_i}, \quad \overline{y}^{(w)} = \frac{\sum w_i y_i}{\sum w_i}$$

and

$$S_{xx}^{(w)} = \sum_{i=1}^{n} w_i (x_i - \overline{x}^{(w)})^2$$

$$S_{xy}^{(w)} = \sum_{i=1}^{n} w_i (x_i - \overline{x}^{(w)})(y_i - \overline{y}^{(w)})$$

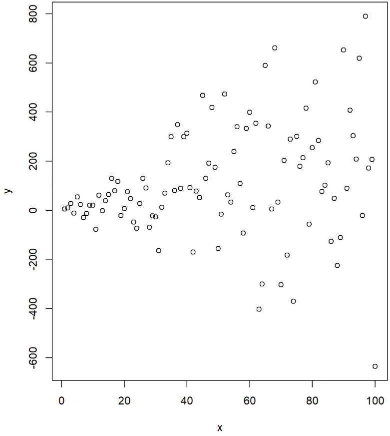
then the solution of the weighted least squares problem is given by

$$\hat{\beta}_1 = \frac{S_{xy}^{(w)}}{S_{xx}^{(w)}} \quad \text{and} \quad \hat{\beta}_0 = \overline{y}^{(w)} - \hat{\beta}_1 \overline{x}^{(w)}.$$

Weighted Least Squares

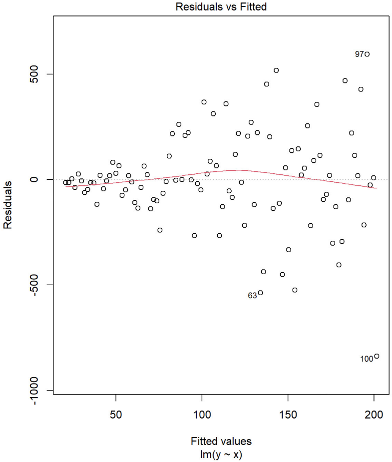
Example

Suppose we have the following data set with $\epsilon_i \sim \mathrm{N}(0, x_i\sigma^2)$, for some unknown σ^2.



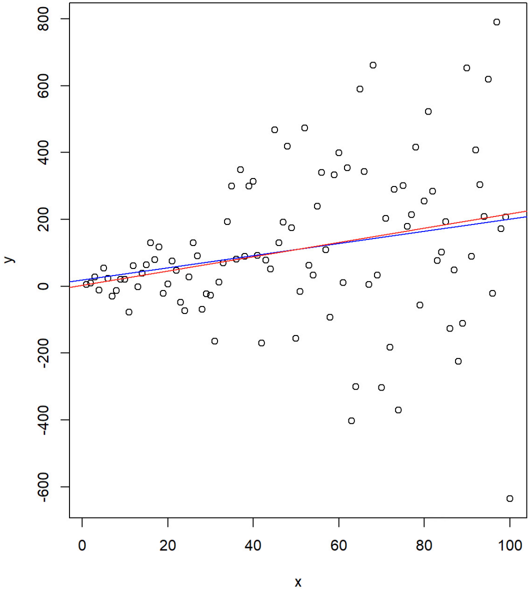

```r
WLSdata<-read.csv("WLS1.CSV", header=TRUE, sep=",")
x<-WLSdata$x
y<-WLSdata$y
plot(x,y)
```

Weighted Least Squares



The residuals vs fitted plot clearly indicate the "funnel pattern", opening up from left to right. This is in accordance to our a priori knowledge that $\epsilon_i \sim \mathrm{N}(0, x_i\sigma^2)$.

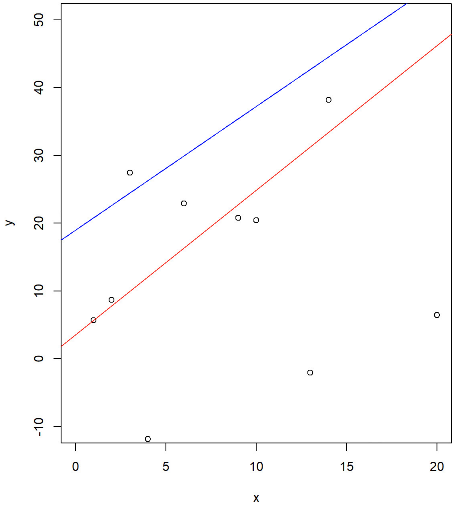
Weighted Least Squares




```
model1=lm(y~x)
abline(model1,col="blue")

w <- (1/x)
model2<-lm(y~x, weights=w)
abline(model2, col="red")
```

ILLINOIS TECH

Weighted Least Squares

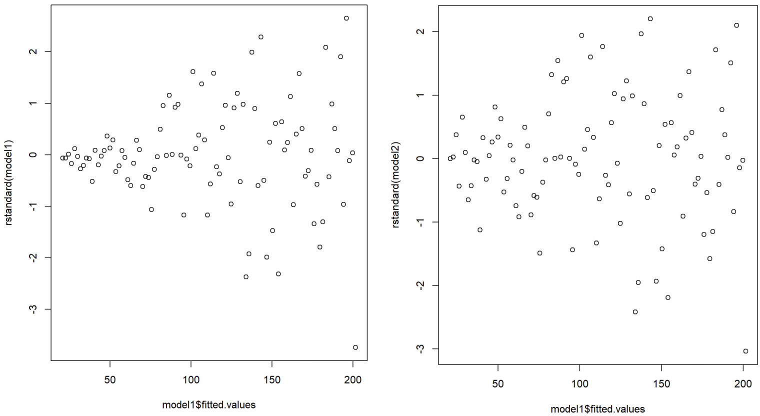



```
> summary(model1)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-836.97  -97.63   -6.72  112.16  594.39

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.9898    45.9893   0.413   0.6806
x             1.8255     0.7906   2.309   0.0231 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 228.2 on 98 degrees of freedom
Multiple R-squared:  0.05159,	Adjusted R-squared:  0.04191
F-statistic: 5.331 on 1 and 98 DF,  p-value: 0.02305

> summary(model2)

Call:
lm(formula = y ~ x, weights = w)

Weighted Residuals:
    Min      1Q  Median      3Q     Max
-85.208 -15.222   0.137  15.895  62.156

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.566     15.850   0.225    0.822
x              2.131      0.508   4.195    6e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 28.39 on 98 degrees of freedom
Multiple R-squared:  0.1522,	Adjusted R-squared:  0.1436
F-statistic:  17.6 on 1 and 98 DF,  p-value: 6.004e-05
```

Weighted Least Squares

Notice that how the weighting technique stabilized the variance of the errors.

```r
plot(x = model1$fitted.values, y=rstandard(model1))
plot(x = model1$fitted.values, y=rstandard(model2))
```



Note: The standardized residual is the residual divided by its standard deviation.

Weighted Least Squares

Statistical inference can be carried out just like the usual regression model.

Note that, when it comes to the confidence and prediction intervals for weighted least squares, we need to provide the weights information as shown below:

```
predict(model2, data.frame(x=16), weights=1/16, interval=
         'confidence', conf.level=0.95)

predict(model2, data.frame(x=16), weights=1/16, interval=
         'prediction', conf.level=0.95)
```

```
> predict(model2, data.frame(x=16), weights=1/16, interval=
+           'confidence', conf.level=0.95)
       fit      lwr      upr
1 37.66045 12.70846 62.61244
> predict(model2, data.frame(x=16), weights=1/16, interval=
+           'prediction', conf.level=0.95)
       fit       lwr      upr
1 37.66045 -189.0358 264.3567
```

```
confint(model2, weights=1/x)
```

```
                 2.5 %    97.5 %
(Intercept) -27.887534 35.020482
x             1.122772  3.138975
```

Weighted Least Squares

The method discussed previously can be easily generalized to a multiple linear regression model and with a more general assumptions on the variance terms σ_i^2. Suppose we have the following multiple regression model,

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$$

with ϵ_i having independent but unequal variance $\mathrm{N}(0, \sigma_i^2)$.

Let us suppose that σ_i^2 are inversely proportional to some wights w_i, that is,

$$\sigma_i^2 = \frac{\sigma^2}{w_i}, \quad i = 1, 2, \cdots, n$$

where $w_1, w_2, \cdots, w_n$ are weights that are positive and σ is unknown.

ILLINOIS TECH

Weighted Least Squares

Then the weighted least squares (WLS) function is then given by

$$S(\beta_0, \beta_1, \cdots, \beta_k) = \sum_{i=1}^{n} w_i \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2$$

As usual, we want to minimize S with respect to $\beta_0, \beta_1, \cdots, \beta_k$.

Weighted Least Squares

It can be shown that

$$\hat{\boldsymbol{\beta}}_{WLS} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y}$$

minimize the S, where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \mathbf{W} = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & & 0 & \cdots & w_n \end{bmatrix}$$

All the subsequent analysis can be done just like the analysis we have seen in the usual multiple regression model.