Chapter 3 Other Issues in Multiple regression (Part 3: Weighted Least Squares Estimation)

1 Weighted Least Squares estimation

When the constant error variance assumption is violated, the estimator of LSE is unbiased (can be used). But it is not optimal, and $s(b)$ is inefficient, too large or small. We need to consider weighted least square estimation.

For a regression model

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + ... + \beta_p X_{1p} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_{21} + ... + \beta_p X_{2p} + \varepsilon_2 \\ ... & \\ Y_n &= \beta_0 + \beta_1 X_{n1} + ... + \beta_{1p} X_{np} + \varepsilon_n. \end{aligned}$$

or

$$\mathbf{Y} = \mathbf{X}\beta + \mathcal{E}$$

where

$$Var(\mathcal{E}) = Var(\mathbf{Y}) = \begin{pmatrix} \sigma_1^2 & 0 & ... & 0 \\ 0 & \sigma_2^2 & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & \sigma_n^2 \end{pmatrix} \text{ denoted by } \mathbf{\Sigma_0}$$

where $\sigma_1^2, ..., \sigma_n^2$ not all the same. The LSE is to minimize

$$Q(\beta_0, ..., \beta_p) = \sum_{i=1}^{n} [Y_i - (\beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip})]^2$$

with respect to $\beta_0, ..., \beta_p$.

We can also give a weight to the summation and estimate $\beta_0, ..., \beta_p$ by minimizing

$$Q_w(\beta_0, ..., \beta_p) = \sum_{i=1}^{n} w_i [Y_i - (\beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip})]^2$$

with respect to $\beta_0, ..., \beta_p$. where $w_i > 0$ are weights, if we believe we should rely more on subject i, w_i should be larger. The estimation is called WLSE

Connection with the weighted average? is it true that using more sample one can get better estimator?

If $\mathbf{Var}(\varepsilon_i) = \sigma_i^2 = \bar{\sigma}_i^2 \kappa^2$ (common part κ^2 and non-common part $\bar{\sigma}_i^2$) and $\varepsilon_i \sim N(0, \bar{\sigma}_i^2 \kappa^2)$, we can rewrite

$$\varepsilon_i = \bar{\sigma}_i \epsilon_i$$

with $Var(\epsilon_i) = \kappa^2$, where $\epsilon_i \sim N(0, \kappa^2)$. If $\bar{\sigma}_i$ is known, we can rewrite the model

$$\begin{aligned}
\frac{Y_1}{\bar{\sigma}_1} &= \beta_0 \frac{1}{\bar{\sigma}_1} + \beta_1 \frac{X_{11}}{\bar{\sigma}_1} + ... + \beta_p \frac{X_{1p}}{\bar{\sigma}_1} + \epsilon_1 \\
\frac{Y_2}{\bar{\sigma}_2} &= \beta_0 \frac{1}{\bar{\sigma}_2} + \beta_1 \frac{X_{21}}{\bar{\sigma}_2} + ... + \beta_p \frac{X_{2p}}{\bar{\sigma}_2} + \epsilon_2 \\
... & \\
\frac{Y_n}{\bar{\sigma}_n} &= \beta_0 \frac{1}{\bar{\sigma}_n} + \beta_1 \frac{X_{n1}}{\bar{\sigma}_n} + ... + \beta_p \frac{X_{np}}{\bar{\sigma}_n} + \epsilon_n
\end{aligned} \tag{1}$$

For this transformed model, it has constant error variance. The LSE of the transformed model is thus

$$\begin{aligned}
Q(\beta_0, ..., \beta_p) &= \sum_{i=1}^n [\frac{Y_i}{\bar{\sigma}_i} - (\frac{\beta_0}{\bar{\sigma}_i} + \beta_1 \frac{X_{i1}}{\bar{\sigma}_i} + ... + \beta_p \frac{X_{ip}}{\bar{\sigma}_i})]^2 \\
&= \sum_{i=1}^n \bar{\sigma}_i^{-2} [Y_i - (\beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip})]^2
\end{aligned}$$

which is a weighted LSE with

$$w_i = \bar{\sigma}_i^{-2} \propto \sigma_i^{-2}$$

Generally, the WLSE with weights w_i is equivalent to a transformed model

$$\begin{aligned}
w_1^{1/2} Y_1 &= \beta_0 w_1^{1/2} + \beta_1 w_1^{1/2} X_{11} + ... + \beta_p w_1^{1/2} X_{1p} + w_1^{1/2} \varepsilon_1 \\
w_2^{1/2} Y_2 &= \beta_0 w_2^{1/2} + \beta_1 w_2^{1/2} X_{21} + ... + \beta_p w_2^{1/2} X_{2p} + w_2^{1/2} \varepsilon_2 \\
... & \\
w_n^{1/2} Y_n &= \beta_0 w_n^{1/2} + \beta_1 w_n^{1/2} X_{n1} + ... + \beta_{1p} w_n^{1/2} X_{np} + w_n^{1/2} \varepsilon_n.
\end{aligned}$$

2 matrix version of weighted least squares estimation

- Minimize $Q_w(b_0, ..., b_p) = \sum_{i=1}^n w_i (Y_i - b_0 - b_1 X_{i1} - ... - b_p X_{i,p})^2$

- by calculus, we have the following **(p+1)** Normal equations:

$$\begin{aligned}\sum_{i=1}^{n} w_i(Y_i - b_0 - b_1 X_{i1} - ... - b_p X_{i,p}) &= 0\\ \sum_{i=1}^{n} w_i(Y_i - b_0 - b_1 X_{i1} - ... - b_p X_{i,p})X_{i1} &= 0\\ &\vdots\\ \sum_{i=1}^{n} w_i(Y_i - b_0 - b_1 X_{i1} - ... - b_p X_{i,p})X_{ip} &= 0\end{aligned}$$

- let $b = (b_0, b_1, ..., b_p)'$ be the solution (estimator). Then the Normal equations can be written as

$$\mathbf{X}'\mathbf{W}\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{W}\mathbf{Y}$$

where $W : n \times n$ and

$$W = \begin{pmatrix} w_1 & 0 & ... & 0\\ 0 & w_2 & ... & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & ... & w_n \end{pmatrix}$$

- The solution, i.e. the estimator of the coefficient vector, is

$$\mathbf{b} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$$

how to "derive" it? the expression for the special simple linear regression model?

- The estimated model is

$$\hat{Y} = b_0 + b_1 X_1 + ... + b_p X_p$$

- Fitted values

$$\hat{Y}_i = b_0 + b_1 X_{i1} + ... + b_p X_{ip}, \qquad i = 1, ..., n$$

- (Fitted) residuals

$$e_i = Y_i - \hat{Y}_i = Y_i - (b_0 + b_1 X_{i1} + ... + b_p X_{ip}), \qquad i = 1, ..., n$$

But, for the transformed model

$$\tilde{e}_i = w_i^{1/2} Y_i - w_i^{1/2} \hat{Y}_i$$

- Estimator of κ^2 (for the transformed model), denoted by $\hat{\kappa}^2$,

$$\widetilde{MSE} = \sum_{i=1}^{n} \tilde{e}_i^2 / \{n - (p+1)\}$$

3 Unbias of the estimators of coefficients

The estimator of coefficient vector is unbiased, i.e.

$$E(\mathbf{b}) = (X^{'}WX)^{-1}X^{'}WEY = (X^{'}WX)^{-1}X^{'}WX\beta = \beta$$

and

$$\mathbf{Var}(\mathbf{b}) = (\mathbf{X}'W\mathbf{X})^{-1}(\mathbf{X}'W\Sigma_0 W\mathbf{X})(\mathbf{X}'W\mathbf{X})^{-1}$$

If $w_i = 1/\bar{\sigma}_i^2$, then

$$\mathbf{Var}(\mathbf{b}) = \kappa^2(\mathbf{X}'W\mathbf{X})^{-1}$$

In details

$$E(\mathbf{b}_k) = \beta_k$$

and

$$\mathbf{Var}(\mathbf{b}_k) = \kappa^2 c_{k+1,k+1}, \qquad k = 0, 1, ..., p-1$$

where c_{kk} is the (k,k)th entry in $(\mathbf{X}'W\mathbf{X})^{-1}$.

4 The distribution of estimators

If $\mathcal{E} \sim \mathbf{N}(\mathbf{0}, \mathbf{\Sigma_0})$ (i.e. ε_i are IID $N(0, \sigma_i^2)$) and $w_i = \kappa^2/\sigma_i^2$, then

- The estimated coefficients

$$\mathbf{b} \sim \mathbf{N}(\beta, \boldsymbol{\kappa}^{\mathbf{2}}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-\mathbf{1}})$$

  Denote the $(i,j)th$ entry of $(\mathbf{X}'W\mathbf{X})^{-1}$ by c_{ij}, then

$$b_k \sim N(\beta_k, \kappa^2 c_{k+1,k+1}), \qquad k = 0, 1, ..., p-1$$

  (where $\mathbf{b} = (b_0, b_1, ..., b_p)'$)

- Let $s(b_k) = \sqrt{\widetilde{MSE} * c_{k+1,k+1}}$, called **Standard Error (S.E.)** for b_k (which can be found in the output of R), then

$$\frac{b_k - \beta_k}{s(b_k)} \sim t(n-p-1)$$

- t-value

$$t^* = \frac{b_k}{s(b_k)}$$

5 Confidence interval for β_k

with $1-\alpha$ confidence, the Confidence interval for β_k is

$$[b_k - s(b_k) * t(1-\alpha/2, n-p-1), \quad b_k - s(b_k) * t(1-\alpha/2, n-p-1)]$$

6 test for $\beta_k = 0$

Our hypothesis is

$$H_0 : \beta_k = 0, \quad H_a : \beta_k \neq 0$$

under H_0,

$$t = \frac{b_k - \beta_k}{s(b_k)} = \frac{b_k}{s(b_k)} \sim t(n-p-1)$$

For significant level α, our criterion is

If the calculated $|t^*| > t(1-\alpha/2, n-p-1)$, reject H_0

If the calculated $|t^*| \leq t(1-\alpha/2, n-p-1)$, accept H_0

Similarly, we can do the test based on the p-value

If p-value $< \alpha$, reject H_0

If p-value $\geq \alpha$, accept H_0

7 Prediction

For any new individual with $X_{new} = (x_1, ..., x_p)^\top$, the predict mean response is

$$\hat{Y}_{new} = \mathcal{X}'\mathbf{b}$$

where

$$\mathcal{X} = (1, x_1, ..., x_p)'$$

We have

$$E\hat{Y}_{new} = EY_{new}$$

Note that if normal errors are assumed, i.e. ϵ_i are IID $N(0, \kappa^2)$, then

$$\hat{Y}_{new} \sim N(\mathbf{E}Y_{new}, \mathcal{X}'(\mathbf{X}'W\mathbf{X})^{-1}\mathcal{X}\kappa^2)$$

Let

$$s^2(\hat{Y}_{new}) = \mathcal{X}'(\mathbf{X}'W\mathbf{X})^{-1}\mathcal{X}\hat{\kappa}^2 = \mathcal{X}'(\mathbf{X}'W\mathbf{X})^{-1}\mathcal{X} * \widetilde{MSE}$$

We have

$$\frac{\hat{Y}_{new} - \mathbf{E}Y_{new}}{s(\hat{Y}_{new})} \sim t(n-p-1)$$

With confidence $100(1-\alpha)\%$, the C.I. for $\mathbf{E}(Y_{new})$ is

$$[\hat{Y}_{new} - s(\hat{Y}_{new}) * t(1-\alpha/2, n-p-1), \quad \hat{Y}_{new} + s(\hat{Y}_{new}) * t(1-\alpha/2, n-p-1)]$$

What about the prediction interval (P.I.) for the value Y_{new}? With confidence $100(1-\alpha)\%$, the P.I. for Y_{new} is

$$[\hat{Y}_{new} - s(pred) * t(1-\alpha/2, n-p-1), \quad \hat{Y}_{new} + s(pred) * t(1-\alpha/2, n-p-1)]$$

where

$$s^2(pred) = \widetilde{MSE} + s^2(\hat{Y}_{new}) = \widetilde{MSE}\{1 + \mathcal{X}'(\mathbf{X}'W\mathbf{X})^{-1}\mathcal{X}\}$$

8 R code for the calculation

```
lm(y ~ X1+...+XP, weights = ??)
```

9 Selection of the weight

The best choice for w_i is

$$w_i = \frac{1}{\bar{\sigma}_i^2}.$$

Even better if $w_i = \frac{1}{\sigma_i^2}$

- if there are repeated observations for each X_i, we can calculate the variance directly. At X_i, we have observations Y_{ij}:

  $$Y_{ij} = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip} + \varepsilon_{ij}, j = 1, ..., n_i$$

  or

  $$Y_{ij} \sim N(\mu_i, \sigma_i^2)$$

  where $\mu_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip}$. We can estimate σ_i^2 by

  $$\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{n_i}(Y_{ij} - Yi.)^2}{n_i - 1}$$

  where $Y_{i.} = \sum_{j=1}^{n_i} Y_{ij}/n_i$.

Example 9.1 *we have data*

X	Y	Estimated $\hat{\sigma}_i$
2.5	2.1299	0.3857
2.5	2.5297	0.3857
2.5	2.7529	0.3857
2.5	1.8971	0.3857
3.0	4.2780	0.8896
3.0	3.1008	0.8896
3.0	2.1252	0.8896
3.0	2.9095	0.8896
4.1	4.6243	1.8249
4.1	3.3015	1.8249
4.1	3.5811	1.8249
4.1	1.7543	1.8249
4.1	6.6880	1.8249

Then we can get w_i easily.

- We use simple LSE (or called ordinary least squares), calculate its fitted residuals e_i. Plot of e_i against X_i and $\hat{Y}_i$, we hope to find the trend in σ_i with X_i or $\hat{Y}_i$.

Example 9.2 *consider Y with X_1 and X_2* **(data)**

The plot of residuals against X are shown in Fig 1 the top panels **(code)**

There is clear trend of σ_i with X_{i1}, a guess of the relation is

$$\sigma_i = \alpha_0' + \alpha_1' X_{i1} + \alpha_2' X_{i2}$$

Note

$$\varepsilon_i = \sigma_i u_i$$

where $u_i \sim N(0,1)$, and that e_i is an "observed" value of ε_i. Therefore

$$e_i \approx \sigma_i u_i$$

and

$$|e_i| \approx c\sigma_i + \sigma_i(|u_i| - c)$$

where $c = E|u_i|$. Thus, we consider model

$$|e_i| \approx \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \text{random error}_i$$

After we estimate the model, we can estimate σ_i

Or similarly, we consider $\sigma_i^2 = \beta_0' + \beta_1' X_{i1} + \beta_2' X_{i2}$ and

$$e_i^2 \approx \sigma_i^2 u_i^2$$

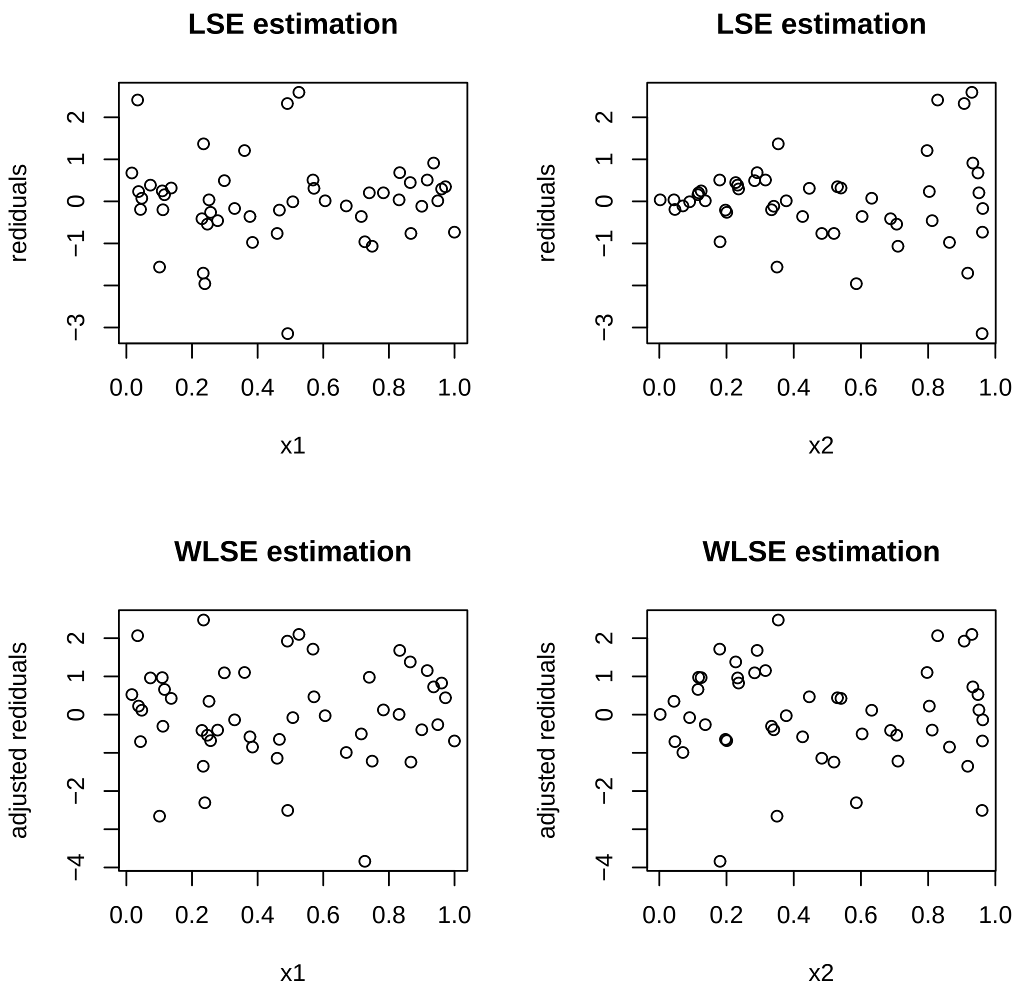

Figure 1:

and

$$|e_i|^2 \approx \sigma_i^2 + \sigma_i^2(|u_i| - 1)$$

Thus, we consider model

$$|e_i|^2 \approx \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + random\ error_i$$

After we estimate the model, we can estimate σ_i^2.

Thus, we set weight $w_i = 1/\sigma_i^2$. *By using the WLSE, the adjusted fitted residuals,* $e_i' = e_i * w_i^{1/2}$, *shows no relation with all* X_i; *see Fig 1 the bottom panels*