Ch 5: Transformations and Weighting to Correct Model Inadequacies

1. Introduction

- Plots of residuals are very powerful methods for detecting violations of error assumption. (See Ch. 4)
- Focus on methods and procedures for building regression models when some of the error assumptions are violated.
- Consider the use of transformation in this Chapter.
- Consider the use of more complex models in later Chapters.

2. Transformations to linearize to model

- It is for fixing nonlinearity.
- Consider this model

\[ y = \beta_0 x^{\beta_1} \varepsilon \]

With taking log transformation

\[ \log y = \log \beta_0 + \beta_1 \log x + \log \varepsilon \]
\[ y' = \beta'_0 + \beta_1 x' + \varepsilon' \]

where the error term \( \varepsilon' = \log \varepsilon \) are normally and independently distributed with mean zero and variance \( \sigma^2 \). (\( y' = \log y, \beta'_0 = \log \beta_0, x' = \log x \))

- Linearizable functions and corresponding linear form

1. \( f(x) = \beta_0 e^{\beta_1 x} \). Taking log transformation

\[ \log f(x) = \log \beta_0 + \beta_1 x \]
\[ f(x)' = \beta'_0 + \beta_1 x \]
2. \( f(x) = \beta_0 + \beta_1 (\log x) \). Letting \( x' = \log x \)

\[
f(x) = \beta_0 + \beta_1 x'
\]

3. \( f(x) = x / (\beta_0 x + \beta_1) \). Letting \( f(x)' = 1/f(x) \) and \( x' = 1/x \)

\[
f(x)' = \beta_0 + \beta_1 x'
\]

4. \( f(x) = \beta_0 + \beta_1 (1/x) \). Letting \( x' = 1/x \)

\[
f(x) = \beta_0 + \beta_1 x'
\]

Example with the Windmill data

- With the model \( y = \beta_0 + \beta_1 x + \varepsilon \) (\( \hat{\sigma} = 0.236 \) and \( R^2 = 0.8745 \))
• With the model \( y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \) (\( \sigma = 0.123 \) and \( R^2 = 0.9676 \))
• With the model \( y = \beta_0 + \beta_1(1/x) + \varepsilon (\hat{\sigma} = 0.09 \text{ and } R^2 = 0.98) \)

3. Variance-stabilizing transformation (for response \( y \))

• Useful transformations

1. \( \sigma_i^2 \propto \text{constant} \)

\[ y' = y \]
2. $\sigma_i^2 \propto E(y_i)$

\[ y' = \sqrt{y} \]

3. $\sigma_i^2 \propto E(y_i)(1 - E(y_i))$

\[ y' = \sin^{-1}(\sqrt{y}) \]

4. $\sigma_i^2 \propto \{E(y_i)\}^2$

\[ y' = \log(y) \]

5. $\sigma_i^2 \propto \{E(y_i)\}^4$

\[ y' = \frac{1}{y} \]

4. Generalized and Weighted Least Squares

- Ordinary Least Squares (OLS): consider the model

\[ y = X\beta + \varepsilon, \]

where $\varepsilon \sim N(0, \sigma^2I)$

- OLSE is

\[ \hat{\beta} = (X'X)^{-1}X'y \]

provided that $(X'X)^{-1}$ exists.

- Some properties of OLSE

\[ E(\hat{\beta}) = \beta \quad \text{and} \quad V(\hat{\beta}) = \sigma^2(X'X)^{-1} \]

- Consider the case

\[ \text{Var}(\varepsilon) = \sigma^2V, \]

where $V \neq I$

1. $V$ is diagonal but has unequal diagonal elements

\[ \Rightarrow \text{the observation } y \text{ are uncorrelated but have unequal variance} \]
2. Off-diagonal elements of \( V \) are not zero

\[ \Rightarrow \text{the observations are correlated} \]

- Consider the model
  \[ y = X\beta + \varepsilon, \]
  where \( \varepsilon \sim N(0, \sigma^2V) \)

- In this model, OLSE is no longer appropriate

- Need a modification of OLS

- Note that it can be shown that it is possible to find a unique nonsingular symmetric matrix \( K \) such that
  \[ K'K = KK = V \]

- Define the new variables
  \[ z = K^{-1}y, \quad B = K^{-1}X, \quad g = K^{-1}\varepsilon, \]

- With the new variables, the transformed model is
  \[ z = B\beta + g. \]

- For the error in the transformed model,
  \[ E(g) = K^{-1}E(\varepsilon) = 0 \quad Var(g) = \sigma^2I \]

- The generalized least squares estimator of \( \beta \) can be obtained using OLS
  \[ \hat{\beta}_G = (B'B)^{-1}B'z \]
  \[ = [X'(K^{-1})'K^{-1}X]^{-1}X'(K^{-1})'K^{-1}y \]
  \[ = (X'V^{-1}X)^{-1}X'V^{-1}y \]

- Some properties of GLSE
  \[ E(\hat{\beta}_G) = \beta \quad \text{and} \quad V(\hat{\beta}_G) = \sigma^2(X'V^{-1}X)^{-1} \]
Weighted Least Squares

• $V$ is diagonal but has unequal diagonal elements

• Consider a simple case of $V$ as

$$ V = \begin{pmatrix} \frac{1}{w_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{w_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{w_n} \end{pmatrix} $$

• Define $W = V^{-1}$

• The weighted least squares estimator (WLSE) of $\beta$ is

$$ \hat{\beta}_W = (X'W X)^{-1} X'Wy $$

• Recall that $KK = V$, where

$$ K = \begin{pmatrix} \frac{1}{\sqrt{w_1}} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sqrt{w_2}} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{\sqrt{w_n}} \end{pmatrix} $$

• Obtain the matrix $B = K^{-1}X$ as

$$ \begin{pmatrix} \sqrt{w_1} & 0 & \ldots & 0 \\ 0 & \sqrt{w_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{w_n} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix} = \begin{pmatrix} \sqrt{w_1} & x_{11}\sqrt{w_1} & \ldots & x_{1k}\sqrt{w_1} \\ \sqrt{w_2} & x_{21}\sqrt{w_2} & \ldots & x_{2k}\sqrt{w_2} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{w_n} & x_{n1}\sqrt{w_n} & \ldots & x_{nk}\sqrt{w_n} \end{pmatrix} $$

• For the vector $z = K^{-1}y$,

$$ \begin{pmatrix} \sqrt{w_1} & 0 & \ldots & 0 \\ 0 & \sqrt{w_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{w_n} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} y_1\sqrt{w_1} \\ y_2\sqrt{w_2} \\ \vdots \\ y_n\sqrt{w_n} \end{pmatrix} $$
• Example:

\[
\begin{array}{cc}
x & y \\
-3 & 1 \\
1 & 5 \\
2 & 8
\end{array}
\]

- Find OLSE of the slope

\[
X = \begin{pmatrix}
1 & -3 \\
1 & 1 \\
1 & 2
\end{pmatrix}, \quad y = \begin{pmatrix}
1 \\
5 \\
8
\end{pmatrix}, \quad X'X = \begin{pmatrix}
3 & 0 \\
0 & 14
\end{pmatrix}
\]

- The OLSE of slope is

\[
(X'X)^{-1}X'y = \begin{pmatrix}
14/3 \\
18/14
\end{pmatrix}
\]

\[
\begin{array}{cc}
x & y \\
-3 & 1 \\
1 & 5 \\
2 & 8
\end{array}
\]

- Find WLSE of the slope under the assumption that the variance of \(y_i\) is proportional to \(1 + x_i^2\)

\[
V = \begin{pmatrix}
10 & 0 & 0 \\
0 & 2 & 0 \\
0 & 0 & 5
\end{pmatrix}, \quad V^{-1} = W = \frac{1}{10} \begin{pmatrix}
1 & 0 & 0 \\
0 & 5 & 0 \\
0 & 0 & 2
\end{pmatrix}
\]

- The OLSE of slope is

\[
(X'WX)^{-1}X'Wy = \frac{1}{14} \begin{pmatrix}
60 \\
18
\end{pmatrix}
\]
1. Introduction

- Leverage point

  1. A leverage point has an unusual $x$-value.
  2. A leverage point does not affect the estimates of regression coefficients, but it has a dramatic affect on the model summary statistic such as $R^2$, $se(\hat{\beta})$.

- Influence point

  1. An influence point has an unusual $x$-value as well as $y$-value.
  2. An influence point affects the estimates of regression coefficients.
2. Hat Matrix

- The hat matrix \( H = X(X'X)^{-1}X' \) plays an important role in identifying influential observations.

- \( \text{Var}(\hat{y}) = \sigma^2H \) and \( \text{Var}(e) = \sigma^2(I - H) \).

- The diagonal elements \( h_{ii} \) of the hat matrix is a measure of the distance of the \( i \)th observation from the center of the \( x \)-space.

- A decision rule: if \( h_{ii} > 2p/n \), the \( i \)th observation can be considered as leverage point.

- Example (the delivery time data): \( 2p/n = 2 \times 3/25 = 0.24 \).

3. Cook’s D

- Cook’s D is a measure of influence

\[
D_i = \frac{r_i^2 h_{ii}}{p(1 - h_{ii})}.
\]
- A decision rule: if $D_i > 1$, the $i$th observation can be considered as an influence point.

<table>
<thead>
<tr>
<th>Obs</th>
<th>h.ii</th>
<th>Cook’s D</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.06373</td>
<td>0.00305</td>
</tr>
<tr>
<td>9</td>
<td>0.49829</td>
<td>3.42031 ****</td>
</tr>
<tr>
<td>10</td>
<td>0.19630</td>
<td>0.05386</td>
</tr>
<tr>
<td>11</td>
<td>0.08613</td>
<td>0.01620</td>
</tr>
<tr>
<td>12</td>
<td>0.11366</td>
<td>0.00160</td>
</tr>
<tr>
<td>20</td>
<td>0.10168</td>
<td>0.13248</td>
</tr>
<tr>
<td>21</td>
<td>0.16528</td>
<td>0.05088</td>
</tr>
<tr>
<td>22</td>
<td>0.39158</td>
<td>0.45118 **</td>
</tr>
<tr>
<td>23</td>
<td>0.04126</td>
<td>0.02991</td>
</tr>
<tr>
<td>24</td>
<td>0.12061</td>
<td>0.10235</td>
</tr>
<tr>
<td>25</td>
<td>0.06664</td>
<td>0.00011</td>
</tr>
</tbody>
</table>