9.4 Weighted Least Squares

In Section 8.3 we considered the problem of *heteroscedastic errors* (i.e., regression errors that have a nonconstant variance). We learned that the problem is typically solved by applying a *variance-stabilizing transformation* (e.g., $\sqrt{y}$ or the natural log of y) to the dependent variable. In situations where these types of transformations are not effective in stabilizing the error variance, alternative methods are required.

In this section, we consider a technique called **weighted least squares.** Weighted least squares has applications in the following areas:

1. Stabilizing the variance of ε to satisfy the standard regression assumption of homoscedasticity.
2. Limiting the influence of outlying observations in the regression analysis.
3. Giving greater weight to more recent observations in time series analysis (the topic of Chapter 10).

Although the applications are related, our discussion of weighted least squares in this section is directed toward the first application.

Consider the general linear model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

To obtain the least squares estimates of the unknown β parameters, recall (from Section 4.3) that we minimize the quantity

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki})]^2$$

with respect to $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$. The least squares criterion weighs each observation equally in determining the estimates of the β's. With **weighted least squares** we want to weigh some observations more heavily than others. To do this we minimize

$$\begin{aligned}\text{WSSE} &= \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n} w_i [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki})]^2\end{aligned}$$

where w_i is the weight assigned to the ith observation. The resulting parameter estimates are called **weighted least squares estimates**. [Note that the ordinary least squares procedure assigns a weight of $w_i = 1$ to each observation.]

Definition 9.1 **Weighted least squares regression** is the procedure that obtains estimates of the β's by minimizing WSSE $= \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2$, where w_i is the weight assigned to the ith observation. The β-estimates are called **weighted least squares estimates**.

Definition 9.2 The **weighted least squares residuals** are obtained by computing the quantity

$$\sqrt{w_i}(y_i - \hat{y}_i)$$

for each observation, where $\hat{y}_i$ is the predicted value of y obtained using the weight w_i in a weighted least squares regression.

The regression routines of most statistical software packages have options for conducting a weighted least squares analysis. However, the weights w_i must be specified. When using weighted least squares as a variance-stabilizing technique, the weight for the ith observation should be the reciprocal of the variance of that observation's error term, σ_i^2, that is,

$$w_i = \frac{1}{\sigma_i^2}$$

In this manner, *observations with larger error variances will receive less weight (and hence have less influence on the analysis) than observations with smaller error variances.*

In practice, the actual variances σ_i^2 will usually be unknown. Fortunately, in many applications, the error variance σ_i^2 is proportional to one or more of the levels of the independent variables. This fact will allow us to determine the appropriate weights to use. For example, in a simple linear regression problem, suppose we know that the error variance σ_i^2 increases proportionally with the value of the independent variable x_i, that is,

$$\sigma_i^2 = kx_i$$

where k is some unknown constant. Then the appropriate (albeit unknown) weight to use is

$$w_i = \frac{1}{kx_i}$$

Fortunately, it can be shown (proof omitted) that k can be ignored and the weights can be assigned as follows:

$$w_i = \frac{1}{x_i}$$

If the functional relationship between σ_i^2 and x_i is not known prior to conducting the analysis, the weights can be estimated based on the results of an ordinary (unweighted) least squares fit. For example, in simple linear regression, one approach is to divide the regression residuals into several groups of approximately equal size based on the value of the independent variable x and calculate the variance of the observed residuals in each group. An examination of the relationship between the residual variances and several different functions of x (such as x, x^2, and $\sqrt{x}$) may reveal the appropriate weights to use.

Example 9.4

A Department of Transportation (DOT) official is investigating the possibility of collusive bidding among the state's road construction contractors. One aspect of the investigation involves a comparison of the winning (lowest) bid price y on a job with the length x of new road construction, a measure of job size. The data listed in Table 9.3 were supplied by the DOT for a sample of 11 new road construction jobs with approximately the same number of bidders.

(a) Use the method of least squares to fit the straight-line model

$$E(y) = \beta_0 + \beta_1 x$$

(b) Calculate and plot the regression residuals against x. Do you detect any evidence of heteroscedasticity?

(c) Use the method described in the preceding paragraph to find the approximate weights necessary to stabilize the error variances with weighted least squares.

Determining the Weights in Weighted Least Squares for Simple Linear Regression

1. Divide the data into several groups according to the values of the independent variable, x. The groups should have approximately equal sample sizes.
   (a) If the data is replicated (i.e., there are multiple observations for each value of x) and balanced (i.e., the same number of observations for each value of x), then create one group for each value of x.
   (b) If the data is not replicated, group the data according to ‘‘nearest neighbors,’’ that is, ranges of x (e.g., $0 \le x < 5$, $5 \le x < 10$, $10 \le x < 15$, etc.).
2. Determine the sample mean ($\bar{x}$) and sample variance (s^2) of the residuals in each group.
3. For each group, compare the residual variance, s^2, to different functions of $\bar{x}$ [e.g., $f(\bar{x}) = \bar{x}$, $f(\bar{x}) = \bar{x}^2$, and $f(\bar{x}) = \sqrt{\bar{x}}$], by calculating the ratio $s^2/f(\bar{x})$.
4. Find the function of $\bar{x}$ for which the ratio is nearly constant across groups.
5. The appropriate weights for the groups are $1/f(\bar{x})$.

DOT11

Table 9.3 Sample data for new road construction jobs, Example 9.4

Job	Length of Road x, miles	Winning Bid Price y, \$ thousands	Job	Length of Road x, miles	Winning Bid Price y, \$ thousands
1	2.0	10.1	7	7.0	71.1
2	2.4	11.4	8	11.5	132.7
3	3.1	24.2	9	10.9	108.0
4	3.5	26.5	10	12.2	126.2
5	6.4	66.8	11	12.6	140.7
6	6.1	53.8			

(d) Carry out the weighted least squares analysis using the weights determined in part c.

(e) Plot the weighted least squares residuals (see Definition 9.2) against x to determine whether the variances have stabilized.

Solution

(a) The simple linear regression analysis was conducted using MINITAB; the resulting printout is given in Figure 9.11. The least squares line (shaded on the printout) is

$$\hat{y} = -15.112 + 12.0687x$$

Note that the model is statistically useful (reject $H_0\colon \beta_1 = 0$) at $p = .000$.

Figure 9.11 MINITAB printout of straight-line model, Example 9.4

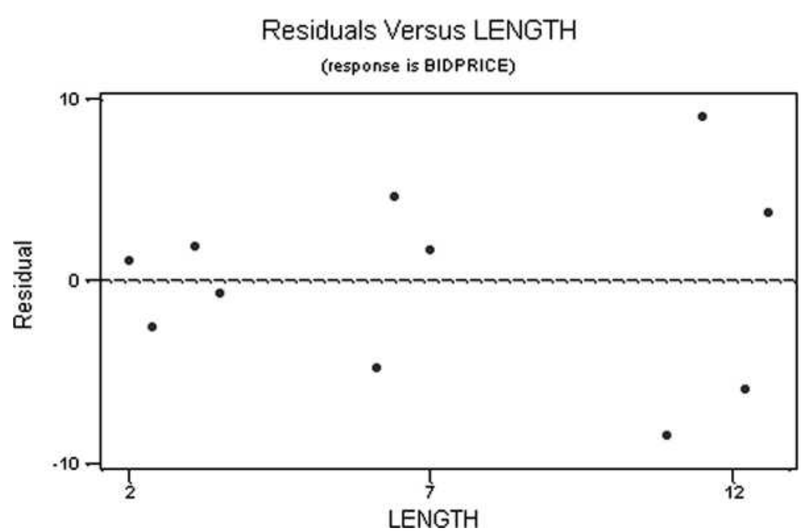
```
The regression equation is
BIDPRICE = - 15.1 + 12.1 LENGTH

Predictor     Coef     SE Coef       T       P
Constant    -15.112      3.342    -4.52   0.001
LENGTH      12.0687     0.4138    29.16   0.000

S = 5.374      R-Sq = 99.0%     R-Sq(adj) = 98.8%

Analysis of Variance

Source            DF        SS        MS        F        P
Regression         1     24558     24558   850.45    0.000
Residual Error     9       260        29
Total             10     24818
```

Obs	LENGTH	BIDPRICE	Fit	SE Fit	Residual	St Resid
1	2.0	10.10	9.02	2.65	1.08	0.23
2	2.4	11.40	13.85	2.52	-2.45	-0.52
3	3.1	24.20	22.30	2.31	1.90	0.39
4	3.5	26.50	27.13	2.19	-0.63	-0.13
5	6.4	66.80	62.13	1.64	4.67	0.91
6	6.1	53.80	58.51	1.67	-4.71	-0.92
7	7.0	71.10	69.37	1.62	1.73	0.34
8	11.5	132.70	123.68	2.45	9.02	1.89
9	10.9	108.00	116.44	2.27	-8.44	-1.73
10	12.2	126.20	132.13	2.67	-5.93	-1.27
11	12.6	140.70	136.95	2.81	3.75	0.82

(b) The regression residuals are calculated and reported in the bottom portion of the MINITAB printout. A plot of the residuals against the predictor variable x is shown in Figure 9.12. The residual plot clearly shows that the residual variance increases as length of road x increases, strongly suggesting the presence of heteroscedasticity. A procedure such as weighted least squares is needed to stabilize the variances.

(c) To apply weighted least squares, we must first determine the weights. Since it is not clear what function of x the error variance is proportional to, we will apply the procedure described in the box to estimate the weights.

Figure 9.12 MINITAB plot of residuals against road length, x



First, we must divide the data into several groups according to the value of the independent variable x. Ideally, we want to form one group of data points for each different value of x. However, unless each value of x is replicated, not all of the group residual variances can be calculated. Therefore, we resort to grouping the data according to "nearest neighbors." One choice would be to use three groups, $2 \le x \le 4$, $6 \le x \le 7$, and $10 \le x \le 13$. These groups have approximately the same numbers of observations (namely, 4, 3, and 4 observations, respectively).

Next, we calculate the sample variance s_j^2 of the residuals included in each group. The three residual variances are given in Table 9.4. These variances are compared to three different functions of $\bar{x}$ ($\bar{x}$, $\bar{x}^2$, and $\sqrt{\bar{x}}$), as shown in Table 9.4, where $\bar{x}_j$ is the mean road length x for group j, $j = 1, 2, 3$.

Table 9.4 Comparison of residual variances to three functions of $\bar{x}$, Example 9.4

Group	Range of x	$\bar{x}_j$	s_j^2	$s_j^2/\bar{x}_j$	$s_j^2/\bar{x}_j^2$	$s_j^2/\sqrt{\bar{x}_j}$
1	$2 \le x \le 4$	2.75	3.722	1.353	.492	2.244
2	$6 \le x \le 7$	6.5	23.016	3.541	.545	9.028
3	$10 \le x \le 13$	11.8	67.031	5.681	.481	19.514

Note that the ratio $s_j^2/\bar{x}_j^2$ yields a value near .5 for each of the three groups. This result suggests that the residual variance of each group is proportional to $\bar{x}^2$, that is,

$$\sigma_j^2 = k\bar{x}_j^2, \quad j = 1, 2, 3$$

where k is approximately .5. Thus, a reasonable approximation to the weight for each group is

$$w_j = \frac{1}{\bar{x}_j^2}$$

With this weighting scheme, observations associated with large values of length of road x will have less influence on the regression residuals than observations associated with smaller values of x.

(d) A weighted least squares analysis was conducted on the data in Table 9.3 using the weights

$$w_{ij} = \frac{1}{\bar{x}_j^2}$$

where w_{ij} is the weight for observation i in group j. The weighted least squares estimates are shown in the MINITAB printout reproduced in Figure 9.13. The prediction equation (shaded) is

$$\hat{y} = -15.274 + 12.1204x$$

Note that the test of model adequacy, H_0: $\beta_1 = 0$, is significant at $p = .000$. Also, the standard error of the model, s, is significantly smaller than the value of s for the unweighted least squares analysis (.669 compared to 5.37). This last result is expected because, in the presence of heteroscedasticity, the

Figure 9.13 MINITAB printout of weighted least squares fit, Example 9.4

```
Weighted analysis using weights in WEIGHT

The regression equation is
BIDPRICE = - 15.3 + 12.1 LENGTH

Predictor        Coef    SE Coef        T       P
Constant      -15.274      1.601    -9.54   0.000
LENGTH        12.1204     0.3792    31.97   0.000

S = 0.6691      R-Sq = 99.1%     R-Sq(adj) = 99.0%

Analysis of Variance

Source            DF        SS        MS        F       P
Regression         1    457.48    457.48  1021.77   0.000
Residual Error     9      4.03      0.45
Total             10    461.51
```

unweighted least squares estimates are subject to greater sampling error than the weighted least squares estimates.

(e) A MINITAB plot of the weighted least squares residuals against x is shown in Figure 9.14. The lack of a discernible pattern in the residual plot suggests that the weighted least squares procedure has corrected the problem of unequal variances. ∎

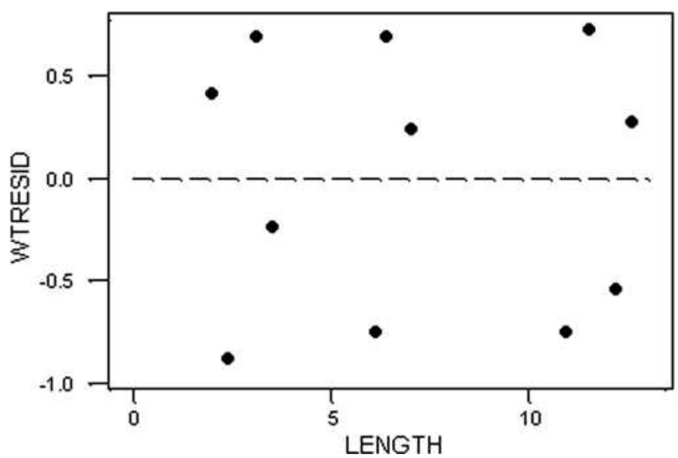


Figure 9.14 MINITAB plot of weighted residuals against road length, x

Before concluding this section, we mention that the "nearest neighbor" technique, illustrated in Example 9.4, will not always be successful in finding the optimal or near-optimal weights in weighted least squares. First, it may not be easy to identify the appropriate groupings of data points, especially if more than one independent variable is included in the regression. Second, the relationship between the residual variance and some preselected function of the independent variables may not reveal a consistent pattern over groups. In other words, unless the right function (or approximate function) of x is examined, the weights will be difficult to determine. More sophisticated techniques for choosing the weights in weighted least squares are available. Consult the references given at the end of this chapter for details on how to use these techniques.