ON THE INCLUSION OF ADDITIONAL DATA IN LINEAR REGRESSION WITH UNEQUAL VARIANCES

by

E. RAMISCAL[1]

1. **Introduction.** Often, the assumption of equality of variances of errors (homoscedasticity) in multiple linear regression misrepresents the real behaviour of the data concerned. It may be known from previous experience or through some reasonable deductions that the error variances are not equal and that they are approximately equal to known fractions of an unknown constant, say, σ^2. If these error variances are denoted by σ_i^2 (i = 1, 2, ..., n) where n is the number of observations, then we may write

$$\sigma_i^2 = (\frac{1}{w_i}) \ \sigma^2 \ (i = 1, 2, \ldots, n).$$

The w_i's are thus assumed to be known or estimable constants and their square roots are called the "weights" for each of the observations. These weights may be estimated by repeated trials. A method of estimating these constants from the repeated observations was given by Baker [2]. With these weights known, the estimates of the regression parameters and their corresponding variances could be found using the popular method of least squares. Throughout this paper, some of them problems involved in the inclusion of additional data will be considered assuming that w_i's are known or estimable.

2. **The Problems.** When the data obtained do not yield estimates of the regression parameters with the desired degree

---

[1] Instructor, Statistical Center, U.P.

of precision, a solution would be to take additional observations
to supplement the one already had. Consequently, two prob-
lems arise: (1) that of ascertaining the number of additional
observations to be taken and (2) that of making the adjust-
ments for the inclusion of additional observations with the
least labor. The second problem has already been considered
by Plackett in 1950 in the case where the assumption of equality
of variances is satisfied. Plackett's approach is used to develop
a solution for the same problem when the variances are equal
to $(1/w_i^2)\sigma^2$, where w_i $(i = 1, 2, \ldots, n)$ are known or estimable
constant and σ^2 is unknown. The first problem is also consi-
dered and a method presented. The method, however, applies
only to simple linear regression although it is conjectured that
the method may be extended to cover the general case with
some limiting conditions.

  3. **Review of Literature.** Let the observed values of a
variable y be represented by

$$y_i = \sum_{j=i}^{s} \beta_j x_{ij} + \epsilon_i \ (i = 1, 2, \ldots, n).$$

In matrix form,

$$Y = X\beta + \epsilon \tag{1},$$

where

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \ X = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1s} \\ x_{21} & x_{22} & \ldots & x_{2s} \\ \cdots\cdots\cdots\cdots \\ x_{n1} & x_{n2} & \ldots & x_{ns} \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}, \text{ and } \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

  The vector of random errors ϵ will be distributed with mean
0 and covariance matrix V, a positive definite matrix. A spe-
cial case of this is when $V = \sigma^2 W^{-1}$, where

$$W^{-1} = \begin{pmatrix} \dfrac{1}{w_1} & 0 & \ldots & 0 \\ 0 & \dfrac{1}{w_2} & \ldots & 0 \\ \multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots} \\ 0 & 0 & \ldots & \dfrac{1}{w_n} \end{pmatrix}$$

and σ^2 is unknown. In this case, the ϵ_i's are uncorrelated but with variances $\sigma_i^2 = (1/w_i)\sigma^2$ $(i = 1, 2, \ldots, n)$, where the w_i's are known constants. Since the variances are positive, it follows that the w_i's are all positive. Hence, the diagonal matrix W^{-1} is a positive definite matrix.

Before going further consider the matrices associated with W^{-1}. Let

$$w = \begin{pmatrix} w_1^{1/2} & 0 & \ldots & 0 \\ 0 & w_2^{1/2} & \ldots & 0 \\ \multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots\cdots} \\ 0 & 0 & \ldots & w_n^{1/2} \end{pmatrix}$$

It is clear that

$$w'w = W = \begin{pmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & \ldots & 0 \\ \multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots} \\ 0 & 0 & \ldots & w_n \end{pmatrix}$$

$$W^{-1} = \begin{pmatrix} \dfrac{1}{w_1} & 0 & \ldots & 0 \\ 0 & \dfrac{1}{w_2} & \ldots & 0 \\ \multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots} \\ 0 & 0 & \ldots & \dfrac{1}{w_n} \end{pmatrix}$$

4. **On the Estimation of Regression Parameters.** Since the parameters are unknown, they must be estimated and, likewise,

they must also have estimates of their standard errors. Markoff [5] had shown that the best linear unbiased estimator of β is the estimate of β which minimizes the sum of squares

$$\epsilon'W\epsilon = (Y - X\beta)\ 'W\ (Y - X\beta).$$

This method of estimation is known as the method of least squares. Since the time of Gauss, this method has been one of the most useful tools in statistical work. The early justification of this method was based on the assumption that all the variables y_i ($i = 1, 2, \ldots, n$) are independent and normally distributed. This assumption had been shown inessential from the point of view of least squares. Markoff made a big step forward by freeing the method from the assumption of normality and equality of variances. Neyman [6] generalized Markoff's theorem to multiple linear regression the proof of which was published by David and Neyman [3] in 1938. Aitken [1] made the last step in generalizing the theory of least squares by considering the case where the variables are both correlated and have unequal variances.

The value of β that minimizes $\epsilon'W\epsilon$ is given by the solution to

$$\frac{\partial}{\partial\beta}\ (\epsilon'W\epsilon) = 0$$

from which is obtained the relation

$$X'WX\ \hat{\beta} = X'WY \qquad\qquad (2)$$

where $\hat{\beta}$ is the least estimate of $\hat{\beta}$. If we let $S = X'WX$, then (2) may be written as

$$\hat{\beta} = S^{-1}X'WY \qquad\qquad (3)$$

provided that S is non-singular. The covariance matrix of the estimates is given by

$$V(\hat{\beta}) = \sigma^2 S^{-1} \qquad\qquad (4)$$

and the unbiased estimate of σ^2 based on the least squares estimate of β is

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'W(Y - \hat{\beta})}{n-s}. \tag{5}$$

It has been shown by Markoff that among all linear unbiased estimates of β, the least squares estimate has the smallest variance.

5. **On the Adjustments for the Inclusion of Additional Data.** Plackett in 1950 gave a method of adjusting the estimates of the parameters and their covariance matrix when additional $m < n$ observations are included. One of his assumptions was that of homoscedasticity and this considerably limits the applicability of his results. Plackett's method should therefore be extended to a more general case.

6. **Extension of Plackett's Theorem to the Non-Homoscedastic Case.** For this, set

$$E(Z) = F\beta \tag{6}$$

where F is an m by s matrix of constants, the counterpart of X, and Z an m by 1 column matrix or vector, the counterpart of Y.

Let

$$P = \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \hdotsfor{4} \\ 0 & 0 & \dots & p_m \end{pmatrix}$$

be the matrix of weights for the new set of m observations which is also equal to p'p where

$$p = \begin{pmatrix} p_1^{1/2} & 0 & \dots & 0 \\ 0 & p_2^{1/2} & \dots & 0 \\ \hdotsfor{4} \\ 0 & 0 & \dots & p_m^{1/2} \end{pmatrix}$$

Denote the combined vector of observations

$$\begin{pmatrix} Y \\ \vdots \\ Z \end{pmatrix} \quad \text{by } Y* \text{ and} \quad \begin{pmatrix} X \\ \vdots \\ F \end{pmatrix} \quad \text{by } X* \tag{7}.$$

The model appropriate for this combined vector of observations will be

$$Y* = X*\beta + \epsilon* \tag{8}$$

and the combined set of weights will be

$$D = \begin{pmatrix} W & O \\ O & P \end{pmatrix} = d'd$$

where

$$d = \begin{pmatrix} w & o \\ o & p \end{pmatrix} .$$

The new estimate of β is $\hat{\beta}* = S*^{-1}X'* D Y*$ and its variance is $V(\hat{\beta}*) = \sigma^2 S*^{-1}$, where $| S* | = | X'*DX* | \neq 0$. Lastly, we define $R = pFS^{-1}F'p'$ and $R* = pFS*^{-1}F'p'$.

Adjustments. Consider

$$RR* = pFS^{-1}F'p'pFS*^{-1}F'p' = pFS^{-1}(F'PF) S*^{-1}F'p' \tag{9}.$$

Since $S* - S = F'PF$, equation (9) reduces to

$$RR* = R - R* \tag{10}.$$

Now consider $(I_m + R) (I_m - R*) = (I_m + R - R*) - RR* = I_m$ which gives $I_m - R* = (I_m + R)^{-1}$,   (11)

since $I_m + R$ evidently possesses an inverse. Postmultiplying (11) by pFS^{-1} we obtain

$$(I_m + R)^{-1} pFS^{-1} = pFS*^{-1}.$$

Consequently,

$$S^{-1}F'p' (I_m + R)^{-1} pFS^{-1} = S^{-1}F'p'pFS*^{-1} = S^{-1} - S*^{-1}, \text{ that}$$
is, $S^{-1} - S*^{-1} = S^{-1}F'p' (I_m + R)^{-1}pFS^{-1}$ (13)

which is a convenient computing formula for the necessary changes in the covariance matrix of $\hat{\beta}$.

It can be shown that the changes in the estimate of β will be

$$\hat{\beta}* - \hat{\beta} = S^{-1}F'p' (I_m + R)^{-1} p(Z - F\hat{\beta}),$$ (14)

and that the new sum of squares of residuals will be

$$M* = M + (Z - F\hat{\beta}) 'p' (I_m + R)^{-1} p(Z - F\hat{\beta}),$$ (15)

where

$$M = (Y - X\hat{\beta})'W(Y - X\hat{\beta}).$$ (16).

The new estimate of σ^2 will therefore be

$$\hat{\sigma}_o^2 = M_o/(n + m - s)$$ (17).

7. **A Numerical Example.** A set of 30 observations was constructed and the resulting normal equations are

$$
\begin{pmatrix}
135 & 211.22 & 462.36 & 2,611.8 & 5,949.87 \\
 & 337.2 & 731.9 & 4,044.06 & 9,458.89 \\
 & & 1,859.91 & 8,922.66 & 20,321.99 \\
 & & & 54,641.48 & 115,455.46 \\
 & & & & 297,802.44
\end{pmatrix}
\begin{bmatrix}
\hat{\beta}_0 \\
\hat{\beta}_1 \\
\hat{\beta}_2 \\
\hat{\beta}_3 \\
\hat{\beta}_4
\end{bmatrix}
=
\begin{bmatrix}
37,194.97 \\
59,411.64 \\
129,121.03 \\
731,310.35 \\
1,678,572.64
\end{bmatrix}
$$

$$\quad\quad S \quad\quad\quad\quad\quad \hat{\beta} \quad\quad\quad\quad b$$

where $b = X'WY$. Using the square-root method, the estimates are

$$
\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{bmatrix} = \begin{bmatrix} 120.83581 \\ 20.31359 \\ 6.05208 \\ 2.99189 \\ 1.00541 \end{bmatrix} ,
$$

and

$$
\begin{bmatrix} .5985294 & -.2752585 & -.0045457 & -.0075708 & .0000360 \\ & .1875810 & -.0051431 & .0019827 & -.0008381 \\ & & .0037982 & -.0000413 & .0000300 \\ & & & .0002646 & -.0000116 \\ & & & & .0000317 \end{bmatrix}
$$

The estimate for σ^2 is given by

$$
\hat{\sigma}^2 = \frac{75.5234}{30 - 5} = 3.021
$$

and

$$
\hat{\sigma} = 1.738.
$$

Adjustments. An additional set of 6 observations was taken and given the following

TABLE 2. ADDITIONAL DATA

No.	x_1	x_2	x_3	x_4	Y	P
1	1.39	1.20	24.02	30.12	267.24	3
2	1.66	5.16	15.40	66.91	298.97	8
3	1.41	5.80	22.08	48.34	298.40	6
4.	1.76	4.97	27.66	55.93	325.79	7
5	1.77	3.54	14.29	49.26	270.37	8
6	1.70	5.95	28.10	61.46	337.84	1

Using equation (13), we get

$$\bar{S}^1 - \bar{S}^1_* = \begin{bmatrix} .0875804 & -.0326636 & -.002248 & -.0015548 & -.0000458 \\ & .02227 & -.0016173 & .0004143 & .0000989 \\ & & .0009092 & .0000138 & .0000434 \\ & & & .0000537 & -.0000035 \\ & & & & .0000036 \end{bmatrix}$$

Therefore

$$\bar{S}^1_* = \begin{bmatrix} .5109490 & -.2425949 & -.0022977 & -.006016 & .0000818 \\ & .165311 & -.0041258 & .0015684 & -.0007392 \\ & & .002889 & -.000551 & -.0000134 \\ & & & .0002109 & .0000081 \\ & & & & .0000281 \end{bmatrix}$$

Using (19), we obtain

$$\hat{\beta}_* - \hat{\beta} = \begin{bmatrix} -.20195 \\ .20098 \\ -.03384 \\ .00137 \\ -.00114 \end{bmatrix}$$

Therefore

$$\hat{\beta}_* = \begin{bmatrix} 120.63386 \\ 20.51457 \\ 6.01824 \\ 2.99326 \\ 1.00427 \end{bmatrix}$$

Using equation (25), we obtain

$$M_* = 75.5234 + 4.59016 = 80.11356, \quad \hat{\sigma}^2_* = \frac{80.11356}{30 + 6 - 5} = 2.5843.$$

Hence $\hat{\sigma}_* = 1.6076$.

8. **A Suggested Method of Estimating the Number of Additional Observations to Meet a Certain Degree of Precision for the Estimate of β.** For simple linear regression, it is known that when the errors are normally distributed with mean O and variance σ_i^2 (i = 1, 2, ..., n), the least squares estimate of β is distributed with mean β and variance $\sigma^2/(\Sigma w_i(X_i - \bar{X})^2$

where $\bar{X} = \Sigma X_i / \Sigma w_i$, i.e.

$$\hat{\beta} \sim N(\beta \; \sigma^2 / \Sigma w_i x_i^2) \tag{18}$$

where $x_i = X_i - X$. Also

$$t = (\hat{\beta} - \beta) / [\; \hat{\sigma} / (\Sigma w_i x_i)^{\frac{1}{2}}] \tag{19}$$

is distributed like a Student's t distribution with (n–2) degrees of freedom.

Suppose that we want the new estimate $\hat{\beta}$ to have a 90% probability of being different from β by no more than a. Then t will be approximately 1.70 and from (19), we get

$$t = \frac{a}{\hat{\sigma}} (\Sigma w_i x_i^2)^{\frac{1}{2}}$$

or

$$\Sigma w_i x_i^2 = (t/a)^2 \hat{\sigma}^2 \tag{20}$$

Since

$$\Sigma w_i x_i^2 = \Sigma w_i X_i^2 = (\Sigma w_i X_i)^2 / \Sigma w_i,$$

for a given n,

$$\overset{n}{\Sigma} w_i X_i^2 - (\overset{n}{\Sigma} w_i X_i)^2 / \overset{n}{\Sigma} w_i = (t/a)^2 \hat{\sigma}^2 \tag{21}$$

If n_1 is the number of original observations and n_2 is the required number of observations for the given precision, then, for $n = n_2$, (21) becomes

$$\overset{n_2}{\Sigma} w_i X_i^2 - (\overset{n_2}{\Sigma} w_i X_i)^2 / \overset{n}{\Sigma} w_i = (t/a)^2 \hat{\sigma}^2 \tag{22}$$

The terms $\overset{n_2}{\Sigma} w_i X_i^2$ and $\overset{n_2}{\Sigma} w_i X_i$ are not known. We might use, however, the approximate relations

$$\overset{n_2}{\Sigma} w_i X_i = (\overset{n_2}{\Sigma} w_i / \overset{n_1}{\Sigma} w_i) \overset{n_1}{\Sigma} w_i X_i \tag{23}$$

and

$$\sum_{}^{n_2} w_i X_i^2 = (\sum_{}^{n_2} w_i / \sum_{}^{n_1} w_i) \sum_{}^{n_1} w_i X_i^2 \tag{23}.$$

By substituting (23) and (24) in (22) and simplifying, we obtain the relation

$$\sum_{}^{n_2} w_i = (\sum_{}^{n_1} w_i / \sum_{}^{n_1} w_i x_i^2)(t/a)^2 \hat{\sigma}^2 \tag{25}.$$

Thus, the additional number of observations is given as the sum of their weights

$$\triangle w = \sum_{}^{n_2} w_i - \sum_{}^{n_1} w_i. \tag{26}$$

The additional observations should therefore be chosen such that the sum of their weights will be equal to $\triangle w$.

If the error variances are equal, the weights can be taken as unity and (25) becomes

$$n_2 = (n_1 / \sum_{}^{n} x_i^2)(t/a)^2 \hat{\sigma}^2$$

and the number of additional observations will be

$$\triangle n = n_2 - n_1. \tag{28}$$

9. **Remarks.** We notice that the value of n_2 or $\sum_{}^{n_2} w_i$ that we get from (27) or (25) is based on the assumptions that equations (23) and (24) are true. It follows that if these stringent assumptions are not satisfied, the true probability will, in general, not be equal to the one imposed. To get the true probability, we substitute the observed values in (22) and solve for t. The computed t, together with the known $(n_2 - 2)$ degrees of freedom, upon looking into a table of Student's t distribution, shall give the true probability.

BIBLIOGRAPHY

[1]. AITKEN, A. C.   "On least squares and linear combinations of observations", *Proceedings of the Royal Society of Edinburgh* (A), Vol. 55 (1935) p. 24.

[2]. BAKER, G. A.   "Linear regression when the standard deviation of arrays are not equal", *Journal of the American Statistical Association*, Vol. 36 (1941) p. 213.

[3]. DAVID and NEYMAN.   "Extension of Markoff's theorem on least squares", *Stat. Res. Mem.*, Volume 2 (1938).

[4]. DWYER, P. S. "The square root method and its use in correlation and regression", *Journal of the American Statistical Association*, Vol. 40 (1945) p. 493.

[5]. MARKOFF, A. A.   *Calculus of Probability.*   Russian Edition Il and IV.   St. Petersburg, Moscow, 1908.

[6]. NEYMAN, J.   *Journal of the Royal Statistical Society*, Vol. 97 (1937) p. 558.

[7]. RAMISCAL, E. R.   *An Extension of Plackett's Approach on the Estimation of Regression Parameters.*   Unpublished M. S. Thesis, U.P. Statistical Training Center, 1967.