

## THE DISPERSION MATRIX OF RESIDUALS WHEN ALL VARIABLES IN A LINEAR REGRESSION FUNCTION ARE SUBJECT TO ERROR

By

CRISTINA P. PAREL<sup>1</sup>

### 1. Preliminary Discussion

Let  $x_1, x_2, x_3, \dots, x_s$  be  $s$  variables in the  $s$ -dimensional Euclidean space ( $s \geq 2$ ) which are assumed to be related by the functional relation  $f(x_1, x_2, \dots, x_s; \theta_0, \theta_1, \dots, \theta_s) = 0$ , where  $\theta_0, \theta_1, \dots, \theta_s$  are the parameters of the relationship. Let  $n$  multivariate observations be made of the variables  $x_1, x_2, x_3, \dots, x_s$  and assume that the expected value  $E(x_{jt}) = x_{j_0}$  ( $j = 1, 2, \dots, s; t = 1, 2, \dots, n$ ), which is the "true" value. If  $g(x_{jt})$  is the probability density function of the random variable  $x_{jt}$ , the expected value of  $x_{jt}$  is defined as

$$E(x_{jt}) = \int_M x_{jt} g(x_{jt}) dx_{jt}, \text{ where}$$

$M$  is the range of  $x_{jt}$  in a one dimensional space. It is assumed that the errors,  $(x_{jt} - x_{j_0})$ , have a joint multivariate distribution, their expected values zero and their variances finite. By assumption, the "true" values are related by the function  $f(x_{1_0}, x_{2_0}, \dots, x_{s_0}; \theta_0, \theta_1, \dots, \theta_s) = 0$ . Since the "true" values of  $x_1, x_2, \dots, x_s$  and the parameters are not known, it is necessary to specify the function relationship and the undetermined coefficients. Hence, let the approximating function be  $f(x'_1, x'_2, \dots, x'_s; \theta'_0, \theta'_1, \dots, \theta'_s) = 0$ .

The problem is to estimate the parameters of the functional relationship by the method of least squares from the observed values which may all be in error.

---

<sup>1</sup> Director, Statistical Center, University of the Philippines.

The least squares estimate of  $\theta$ , assuming a linear relationship, is

$$\hat{\theta} = (X^{0T}WX^0)^{-1} X^{0T}WY^0, \dots \quad (1)$$

where  $X^0$  is an  $n \times s$  non-singular matrix of observed values of  $x_i$ , ( $i = 1, 2, \dots, s$ ) and  $Y^0$  is an  $(n \times 1)$  matrix of observed values of the dependent variable. The weight matrix,  $W$ , is the identity matrix  $I_n$ , if the residuals are assumed uncorrelated and of equal weights.

The above estimate of  $\theta$  can be used somewhat less satisfactorily when the  $x$ 's are subject to error and no knowledge of the size of the error is available. If it is known that the errors in the  $x$ 's are relatively small with respect to the  $x$ 's, it may be quite satisfactory. More specifically, the first order error approximation to the error the  $\theta$  resulting from the consideration of  $dX^0$  in the normal equation is given by

$$d\hat{\theta} = - (X^{0T}X^0)^{-1} [(dX^0)^T X^0 + X^{0T} (dX^0)] \\ (X^{0T}X^0)^{-1} X^{0T}Y^0 + (X^{0T}X^0)^{-1} (dX^0)Y^0, \dots \quad (2)$$

where  $dX^0$  denotes the error in  $X^0$ . In general, these values of  $dX^0$  are not known, but in some cases, bounds for them are known so that some estimate of  $|d\hat{\theta}|$  can be computed. It can then be seen that if the errors of the observations on the  $x_{i1}$ 's ( $i = 1, 2, \dots, n; 1 = 1, 2, \dots, s$ ) are negligible with respect to these observations, the estimate  $\hat{\theta}$  is quite a satisfactory approximation.

However, in cases for which it cannot be assumed that the errors in the observations in the  $x_{i1}$ 's are trivial, these estimates may not be adequate, and even the use of (2) is not completely satisfactory since the values of  $dX^0$  are not commonly known though bounds for them may be available. It is necessary to take into account the errors of observations of the  $x_{i1}$ 's, as well as those of the  $Y_i$ 's. For this purpose, an

improved experimental design which can provide estimates of the variation of each measurement is suggested in this paper. Some earlier authors such as Deming, Acton, Brown, Kummel and Norton have indicated an appropriate experimental design with corresponding analysis in working out an appropriate theory for generalized regression coefficients appropriate for curve fitting. The experimental design proposed here is in more respects similar to the method proposed by Kummel in 1879, and the method for  $s = 1$  case presented by Norton, though the following presentation using matrices, is applicable to the general case with any  $s$  and with the possibility of correlated values, not only of the observations of the  $Y_i$ 's but also of the  $x_{ij}$ 's, and even of the  $Y_i$ 's and the  $x_{ij}$ 's.

## 2. An Experimental Design Providing Estimates of Variation in the Measurements.

For each observation designed to enter the normal equations, a series of multivariate measurements in the vicinity of the desired  $x_{ij}$ 's ( $i = 1, 2 \dots n$ ;  $j = 1, 2 \dots s$ ) and  $Y_i$ , is made. The set of points corresponding to each of these multivariate points is called a "constellation". The collection of means is taken as the multivariate observation while the variances and covariances are used as measures of their errors.

Suppose  $n > s$  sets of observations are made on each of the variables  $x_1, x_2, x_3, \dots, x_s, y$  and that there are  $m_i$  individual observations in the  $i$ th set of the  $n$  constellations of observation points. There are, therefore  $n(s + 1)$  random variables with  $m_i$  ( $i = 1, 2, \dots, n$ ) observations in the  $i$ th set. Let it be assumed that  $E(x_{it}) = X_{i0}$ , ( $i = 1, 2, \dots, s$ ;  $t = 1, 2, \dots, m_i$ ;  $i = 1, 2 \dots n$ ) and  $E(Y_{it}) = Y_{i0}$ . The random variables  $(x_{it} - x_{i0})$  and  $(Y_{it} - Y_{i0})$  within the  $i$ th constellation are assumed to have the same multivariate distribution, expected value zero, and finite variances. Let the  $i$ th observation be

represented by the mean  $X_{i0} = \frac{1}{m_i} \sum_{t=1}^{m_i} x_{it}$

and  $y_i = \frac{1}{m_i} \sum_{t=1}^{m_i} y_{it}^0$ . Each residual  $e_j$  ( $i = 1, 2, \dots, n$ ),

then is associated with the means of the  $m_i$  observations of the variables  $x_{i1}, x_{i2}, \dots, x_{is}, y_i$  of the  $i$ th constellation, so that in the least squares estimate of  $\theta$ ,  $X^0$  and  $Y^0$  are now replaced by the  $n \times (s + 1)$  matrix  $A$  and the  $(n \times 1)$  vector  $Y$  of means. As the  $m_i$  of individual observations of the  $n$  sets of observations of  $x_1, x_2, x_3, \dots, x_s, y$  is increased indefinitely,  $x_{i1}, x_{i2}, x_{i3}, \dots, x_{is}, y_i$  converge stochastically to the true values and, therefore,  $\hat{\theta}$  is a consistent estimate of  $\theta$ .

In the classical regression theory, the weight matrix applied to the residuals is arbitrarily assigned, or, if no basis for arbitrary weighting is available, the dispersion matrix  $V_e$  of the residuals which is assumed to be known is used. The least squares estimates of  $\theta$  obtained for these cases may be, as pointed out in the above discussion very poor estimates of  $\theta$ , unless the errors of the observations on the  $x_{i1}$ 's ( $i = 1, 2, \dots, n$ ;  $1 = 1, 2, \dots, s$ ) are very small relative to the observed values. The estimates of  $\theta$  can be improved by taking into consideration the variances of the errors of observation and the fact that the residuals  $e_i$  ( $i = 1, 2, \dots, n$ ) are actually functions of the regression coefficients which are to be estimated.

### 3. The General Form of the Dispersion Matrix of Residuals.

Let the vector of residuals be denoted by

$$e = Y - X\theta, \dots \quad (3)$$

where  $\theta$  is the vector of parameters,  $Y$  and  $X$  are respectively  $n \times 1$  and  $n \times (s + 1)$  matrices of constellation mean values;  $e$  is the vector of residuals  $e_i$  associated with the "mean" point of the  $i$ th constellation. If the expected value of a matrix of random variables is defined as the matrix of expected values of its elements having the same row and column orders as the

matrix of elements, the dispersion matrix  $V_e$  of residuals,  $e$ , is

$$V_e = E (ee^T) = E (e) E (e^T) \dots \dots \quad (4)$$

Since the vector  $e$  of residuals  $e_i$  ( $i = 1, 2 \dots, n$ ) is a function of the vector  $Y$  and the matrix  $X$  of mean values, it seems necessary to express the dispersion matrix  $V_e$  in terms of the dispersion matrices of  $Y$  and of  $X$ , and the

covariances of vectors  $Y$  and  $\begin{bmatrix} X_{0j} \\ X_{j1} \\ \dots \\ X_{jn} \end{bmatrix}$ . After algebraic manipulations, the general form for  $V_e$  is obtained as:

$$(5) \dots V_e = \text{var } Y - \text{cov} (X\theta, Y) - \text{cov} (Y, X\theta) + \text{var} (X\theta);$$

where

$$\left. \begin{aligned} \text{var } Y &\equiv E (\bar{Y}\bar{Y}^T); \\ \text{cov} (Y, \theta X) &\equiv E (\bar{Y}\theta^T\bar{X}^T); \\ \text{cov} (X\theta, Y) &\equiv E (\bar{X}\theta\bar{Y}^T); \\ \text{var} (X\theta) &\equiv E (\bar{X}\theta\theta^T\bar{X}^T); \end{aligned} \right\} \dots \dots (6)$$

$Y \equiv [Y_{ie} - E (y_i)]$  is an  $n \times 1$  vector of deviates.

Further reduction gives

$$(7) \dots V_e = V_y + \sum_{i=1}^n \sum_{j=1}^n B_i (\theta^T V_{x_{ij}} \theta - \theta^T V_{y_i x_j} - V_{x_i y_j} \theta) B_i^T;$$

where  $V_y$  is the dispersion matrix of the vector of  $Y_i$ 's;  $V_{x_i x_j}$  and  $V_{y_i x_j}$  are respectively column and row covariance vectors of the  $y$  in the  $i$ th constellation and the  $x$ 's in the  $j$ th constellation;  $V_{x_{ij}}$  is the dispersion matrix of  $x$ 's in the  $i$ th and  $j$ th constellations. It should be noted that  $V_{x_i y_j} = V_{y_i x_j}^T$ .  $B_i$  is a column vector with unity in the  $i$ th position and zero elsewhere.

The matrix,  $V_e$ , is in general non-diagonal. However, if the observations of the  $x_{i1}$ 's and the  $y_i$ , ( $i = 1, 2, 3 \dots n$ ) are independent, then the residuals are uncorrelated. In this case, the dispersion matrices  $V_{x_{ij}}$  and the covariance vectors,  $V_{x_i y_j}$  ( $i \equiv j$ ) are, therefore, all zero. The dispersion matrix of residuals then becomes a diagonal matrix, i.e.,

$$V_e = V_y + \sum_{i=1}^n B_i (\theta^T V_{x_{i1}} \theta - \theta^T V_{x_i y_i} - V_{x_i y_i}^T \theta) B_i^T \dots \quad (8)$$

where  $V_y$  is a diagonal matrix. However,  $V_{x_{ij}}$  is non-diagonal unless the  $x$ 's in the  $i$ th constellation are mutually stochastically independent.

4. Example.

To illustrate, suppose the simple case of two variables  $y$  and  $x_1$  where the regression function of  $y$  or  $x_1$  is given by

$$E(y) = x_0 \theta_0 + x_1 \theta_1 \quad (\text{where } x_0 \equiv 1)$$

is considered. Suppose that three sets of observations are made of  $y$  and  $x_1$ , with the  $i$ th set consisting of  $m_i$  observations ( $i = 1, 2, 3$ ), and that all these observations are in error. The means of these  $n$  sets of observations are calculated. Further, assume that the observations of  $x_{i1}$  and  $y_i$  are not independent. Then, the residuals  $e_i = y_i - (x_{i0} \theta_0 + x_{i1} \theta_1)$ ;  $i = 1, 2, 3$ , are not independent. The dispersion matrix  $V_e$  of residuals is, therefore, a non-diagonal matrix, and is a function of  $\theta_0$  and  $\theta_1$ . Let the calculated variances of the  $y_i$ 's and the  $x_{i1}$ 's ( $i = 1, 2, 3$ ;  $1 = 0, 1$ ) and the covariances of the  $y_i$ 's and the  $x_{i1}$ 's be given as:

$i$	$\sigma^2_{y_i}$	$\sigma^2_{x_0}$	$\sigma^2_{x_1}$	$\sigma_{y_i x_1}$
1	1	0	2	.98
2	2	0	1	.98
3	4	0	3	2.38

$\sigma_{x_1x_2} = 1.12$	$\sigma_{y_1y_2} = 1.12$	$\sigma_{x_1y_2} = 1.57$	$\sigma_{x_3y_1} = 1.36$
$\sigma_{x_1x_3} = 1.90$	$\sigma_{y_1y_3} = 1.60$	$\sigma_{x_2x_1} = .80$	$\sigma_{x_2y_3} = 1.60$
$\sigma_{x_2x_3} = 1.36$	$\sigma_{y_2y_3} = 2.24$	$\sigma_{x_1y_3} = 2.24$	$\sigma_{x_3y_2} = 1.90$

The dispersion matrix of the  $y_i$ 's is then

$$V_y = \begin{pmatrix} 1 & 1.12 & 1.60 \\ 1.12 & 2 & 2.24 \\ 1.60 & 2.24 & 4 \end{pmatrix}$$

The dispersion matrices of the  $x_{ij}$ 's are:

$$V_{x_{11}} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}; \quad V_{x_{12}} = \begin{pmatrix} 0 & 0 \\ 0 & 1.12 \end{pmatrix}; \quad V_{x_{13}} = \begin{pmatrix} 0 & 0 \\ 0 & 1.90 \end{pmatrix}$$

$$V_{x_{22}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}; \quad V_{x_{23}} = \begin{pmatrix} 0 & 0 \\ 0 & 1.36 \end{pmatrix}; \quad V_{x_{33}} = \begin{pmatrix} 0 & 0 \\ 0 & 3 \end{pmatrix}$$

The covariance vectors of the  $y_i$ 's and the  $x_i$ 's are:

$$V_{x_1y_1} = \begin{pmatrix} 0 \\ .98 \end{pmatrix}; \quad V_{x_1y_2} = \begin{pmatrix} 0 \\ 1.50 \end{pmatrix}; \quad V_{x_2y_1} = \begin{pmatrix} 0 \\ .80 \end{pmatrix}$$

$$V_{x_2y_2} = \begin{pmatrix} 0 \\ .98 \end{pmatrix}; \quad V_{x_1y_3} = \begin{pmatrix} 0 \\ 2.24 \end{pmatrix}; \quad V_{x_3y_1} = \begin{pmatrix} 0 \\ 1.36 \end{pmatrix}$$

$$V_{x_3y_3} = \begin{pmatrix} 0 \\ 2.38 \end{pmatrix}; \quad V_{x_2y_3} = \begin{pmatrix} 0 \\ 1.60 \end{pmatrix}; \quad V_{x_3y_2} = \begin{pmatrix} 0 \\ 1.90 \end{pmatrix}$$

The dispersion matrix of the residuals is therefore,

$$V_e = \begin{pmatrix} 1 & 1.12 & 1.60 \\ 1.12 & 2 & 2.24 \\ 1.60 & 2.24 & 4 \end{pmatrix} + \sum_{i=1}^3 \sum_{j=1}^3 B_i (\theta^T V_{x_j} \theta - \theta^T V_{x_j} \theta) - V_{y_1x_j} (\theta) B_j^2$$





b) If the observations of the  $y_i$ 's are not in error,  $V_y = O$ . It also follows that  $V_{x_i y_i} = O = V_{y_i x_i}$ . Hence,

$V_e$  becomes

$$V_e = \sum_{i=1}^n B_i (\theta^T V_{x_{ii}} \theta) B_i^T$$

This is not a common case.

2. When  $V_e$  is a non-diagonal matrix; i.e., when the residuals are correlated.

a) If the observations on the  $x_{ij}$ 's are free of error ( $i = 1, 2 \dots n; j = 1, 2 \dots s$ ), then  $V_{x_{ij}} = O$  for every  $i, j = 1, 2 \dots n$ .

Also  $V_{x_i y_i} = O = V_{y_i x_j}$ . Hence, the matrix of residuals be-

comes  $V_e = V_y$ , a constant non-diagonal matrix. This is the classical case of correlated residuals treated by Aitken, Duyer, and Brown.

b) If the observations of the  $y_i$ 's are free of error,  $V_y = O$ , and  $V_{x_i y_j} = O = V_{y_i x_j}$ . Then

$$V_e = \sum_{i=1}^n \sum_{j=1}^n B_i (\theta^T V_{x_{ij}} \theta) B_j^T$$

As in 1(a), this is not a common case.

In many problems, the dispersion matrix of residuals  $V_e$  must be estimated from the observed data. In cases where there are correlated residuals, there must be an equal number of observation points in each constellation and the order of observation is essential. The individual observations are not replicate in the strict sense.

6. **Concluding Remarks.** In the given form for  $V_e$ , an estimate of  $V_e$  can be obtained if good estimates of  $\theta$  are available. A useful estimate of  $\theta$  can be obtained by minimizing the diagonal of  $V_e$  with respect to  $\theta$ . As the diagonal terms of  $V_e$  are quadratic functions of  $\theta$  and  $V_{x_{ij}}$  is assumed positive definite, the diagonal terms have a minimum.