The Annals of Mathematical Statistics
1968, Vol. 39, No. 5, 1638–1645

EFFICIENT DIFFERENCE EQUATION ESTIMATORS IN EXPONENTIAL REGRESSION

BY C. A. MCGILCHRIST[1]

University of New South Wales

1. Introduction. A multiple exponential regression curve is given by

$$\epsilon(Y_x) = \eta_x = \alpha - \textstyle\sum_{i=1}^{k} \beta_i \rho_i^{\,x}, \qquad 0 < \rho_1 < \cdots < \rho_k < 1, \tag{1}$$

where Y_x is the observation at x. We consider here the estimation of the ρ_i when the observations are independent and normally distributed with constant variance, σ^2, and are equally spaced at x values denoted by $x = 0, 1, 2, \cdots, n - 1$.

For single exponential regression ($k = 1$) a technique used by Hartley [3] and Patterson [6] is to replace the regression curve by a difference equation which generates it and to estimate the parameters of the difference equation. Using this technique Lipton and McGilchrist [5] have obtained a class of estimators for the ρ_i of the multiple exponential model. Denoting an estimator of ρ_i by r_i, this class is given by the solution of the equations,

$$\begin{aligned}(2)\quad [(-1)^k\hat\theta_k \mathbf{y}_0' + (-1)^{k-1}\hat\theta_{k-1}\mathbf{y}_1' + \cdots + \mathbf{y}_k']D[(-1)^k(\hat\theta_k^{\,j} b_j + \hat\theta_k)\mathbf{y}_0 \\ + (-1)^{k-1}(\hat\theta_{k-1}^{\,j} b_j + \hat\theta_{k-1})\mathbf{y}_1 + \cdots + \mathbf{y}_k] = 0, \quad j = 1, 2, \cdots k,\end{aligned}$$

where $\quad \mathbf{y}_p = \{Y_p, Y_{p+1}, \cdots, Y_{n+p-k-1}\}, \qquad p = 0, 1, 2, \cdots, k,$

$\hat\theta_p = p$th order symmetric function in $r_1, r_2, \cdots, r_k$,

$\hat\theta_p^{\,j} = \partial\hat\theta_p/\partial r_j$.

The above equations correspond to those given on p. 507 of [5]. The b_j are constants and D is a square matrix of order $n - k$, and the b_j and D are selected to satisfy suitable criteria. Except for the case of single exponential regression (already studied by Patterson [6]), Lipton and McGilchrist found the usual criteria of zero bias and minimum variance too difficult to apply to (2) in order to select the b_j and D, and were unable to proceed. In this paper alternative criteria are considered in Section 2 and these found much easier to apply in Section 4.

2. Estimating equations and criteria. The criteria now described are similar in principle to those suggested by Barnard and reported by Durbin [2]. Representing estimating equations (2) by

$$T_j(Y, r) = 0, \qquad j = 1, 2, \cdots, k, \tag{3}$$

we consider the equivalent estimating functions, $T_j(Y, \rho)$, which are obtained

Received 20 February 1967; revised 25 April 1968.
[1] Partially supported by Australian Research Grants Committee.

from the left hand side of (3) by replacing the r's with ρ's. The $T_j(Y, \rho)$, being functions of the observations, are random variables but considering the ρ's as variables the $T_j(Y, \rho)$ are random surfaces in $k + 1$ dimensions. The point on each surface corresponding to the true ρ values should be distributed close to zero. The first two criteria listed below require $T_j(Y, \rho)$ to be distributed about zero (for true ρ values) and the third that this distribution should have small variance.

The criteria are:

(a) If $T_j(\eta, \rho)$ is obtained from $T_j(Y, \rho)$ by replacing each Y_x with its expected value η_x, then we require $T_j(\eta, \rho)$ to be zero for all j.

(b) $\epsilon[T_j(Y, \rho)]$ should be zero for all j.

(c) Since Var $[T_j(Y, \rho)]$ may be made arbitrarily small by dividing (3) by a large number we minimise Var $[T_j(Y, \rho)]$ subject to the slope of $T_j(Y, \rho)$ considered as a random surface being standardised. It is standardised by holding

$$S = [\partial/\partial\rho_j T_j(Y, \rho)] \; Y_x = \eta_x$$

constant. The above procedure is then equivalent to minimising S^{-2} Var $[T_j(Y, \rho)]$.

3. Matrix notation. In this section we set up a notation and establish some results used in the following sections. Let

$$\mathbf{z}_1 = \sum_{p=0}^{k} (-1)^{k-p}\theta_{k-p}\mathbf{y}_p, \qquad \mathbf{z}_{2j} = \sum_{p=0}^{k} (-1)^{k-p}\theta_{k-p}^{j}\mathbf{y}_p,$$

then from (2) and (3) we have

$$T_j(Y, \rho) = \mathbf{z}_1'D(\mathbf{z}_1 + b_j\mathbf{z}_{2j}). \tag{4}$$

Let $\mathbf{1}$ be a vector with $n - k$ elements each of which is unity,

$$\boldsymbol{\varrho}_i = \{1, \rho_i, \rho_i^2, \cdots, \rho_i^{n-k-1}\},$$

and U_s be a square matrix of order $n - k$ with all elements zero except for a diagonal of 1's elevated s rows from the principal diagonal (note that U_0 is the identity matrix and if s is negative the diagonal of 1's is below the principal diagonal). Using this notation we have

$$\begin{aligned} \epsilon(\mathbf{z}_1) &= \alpha\sum_{p=0}^{k} (-1)^{k-p}\theta_{k-p}\mathbf{1} + \sum_{i=1}^{k} [\beta_i\boldsymbol{\varrho}_i(\sum_{p=0}^{k} (-1)^{k-p}\theta_{k-p}\rho_i^{p})] \\ &= \alpha\prod_{i=1}^{k} (1 - \rho_i)\mathbf{1}, \end{aligned} \tag{5}$$

since

$$\sum_{p=0}^{k} (-1)^{k-p}\theta_{k-p} = \prod_{i=1}^{k} (1 - \rho_i) \quad \text{and} \quad \sum_{p=0}^{k} (-1)^{k-p}\theta_{k-p}\rho_i^{p} = 0.$$

Similarly we find

$$\epsilon(\mathbf{z}_{2j}) = \alpha\prod_{i=1,(i\neq j)}^{k} (1 - \rho_i)\mathbf{1} + \beta_j\boldsymbol{\varrho}_j\prod_{i=1,(i\neq j)}^{k} (\rho_j - \rho_i). \tag{6}$$

Covariance matrices, $\Sigma_{11} = \text{Var}\, \mathbf{z}_1$ and $\Sigma_{12}^j = \text{Cov}\,(\mathbf{z}_1, \mathbf{z}_{2j})$ are then

$$\Sigma_{11} = \sigma^2 \textstyle\sum_{s=-k}^{k} (-1)^s U_s \sum_{h=0}^{k-|s|} \theta_h \theta_{h+|s|}, \tag{7}$$

$$\Sigma_{12}^j = \sigma^2 \textstyle\sum_{s=-k}^{k} (-1)^s U_s \sum_{h=0}^{k-|s|} \theta_h \theta_{h+|s|}^j .$$

4. Application of criteria.

4.1. *First criterion.* Using (5) we obtain

$$T_j(\eta, \rho) = \epsilon(\mathbf{z}_1')D\epsilon(\mathbf{z}_1 + b_j\mathbf{z}_{2j}) = \alpha \textstyle\prod_{i=1}^{k} (1 - \rho_i)\mathbf{1}'D\epsilon(\mathbf{z}_1 + b_j\mathbf{z}_{2j}),$$

and thus $T_j(\eta, \rho)$ is made zero by requiring

$$\mathbf{1}'D = 0. \tag{8}$$

4.2 *Second criterion.* Since

$$\epsilon[T(Y, \rho)] = \epsilon(\mathbf{z}_1')D\epsilon(\mathbf{z}_1 + b_j\mathbf{z}_{2j}) + \text{tr}\,[D\, \text{Cov}(\mathbf{z}_1 + b_j\mathbf{z}_{2j}, \mathbf{z}_1)]$$
$$= T_j(\eta, \rho) + \text{tr}\, D\Sigma_{11} + b_j\, \text{tr}\, D\Sigma_{12}^j,$$

then provided $T_j(\eta, \rho)$ is zero, this criterion is satisfied by requiring

$$b_j = -\text{tr}\,(D\Sigma_{11})/\text{tr}\,(D\Sigma_{12}^j) \tag{9}$$
$$= -\textstyle\sum_{s=-k}^{k} (-1)^s t_s \sum_{h=0}^{k-|s|} \theta_h\theta_{h+|s|} / \sum_{s=-k}^{k} (-1)^s t_s \sum_{h=0}^{k-|s|} \theta_h\theta_{h+|s|}^j$$

where $t_s = \text{tr}\, DU_s$ and we have used equations (7) in substituting for Σ_{11} and Σ_{12}^j.

4.3. *Third criterion.* Using (4) and results relating fourth order moments of a multinormal distribution to its second order moments,

$$\text{Var}\,[T_j(Y, \rho)] = \text{Var}\,[\mathbf{z}_1'D(\mathbf{z}_1 + b_j\mathbf{z}_{2j})]$$
$$= \epsilon(\mathbf{z}_1')D\,\text{Var}\,(\mathbf{z}_1' + b_j\mathbf{z}_{2j})D\epsilon(\mathbf{z}_1)$$
$$+ 2\epsilon(\mathbf{z}_1')D\,\text{Cov}\,(\mathbf{z}_1 + b_j\mathbf{z}_{2j}, \mathbf{z}_1)D\epsilon(\mathbf{z}_1 + b_j\mathbf{z}_{2j})$$
$$+ \epsilon(\mathbf{z}_1 + b_j\mathbf{z}_{2j})'D\Sigma_{11}D\epsilon(\mathbf{z}_1 + b_j\mathbf{z}_{2j}) + \text{tr}\,[D\Sigma_{11}D\,\text{Var}\,(\mathbf{z}_1 + b_j\mathbf{z}_{2j}]$$
$$+ \text{tr}\,[D\,\text{Cov}\,(\mathbf{z}_1, \mathbf{z}_1 + b_j\mathbf{z}_{2j})D\,\text{Cov}\,(\mathbf{z}_1 + b_j\mathbf{z}_{2j}, \mathbf{z}_1)].$$

The first three terms involve the square of means of the Y's and the first power of the variance while the last two do not depend on the mean values of the Y's but on the second power of their variance. Since the variance of the Y's must be small compared with their average mean value, particularly to be justified in fitting more than one exponential term we propose to neglect these last two terms. Using (8), (5) and (6) the first three terms reduce to

$$\text{Var}\,[T_j(Y, \rho)] = b_j^2\epsilon(\mathbf{z}_{2j})'D\Sigma_{11}D\epsilon(\mathbf{z}_{2j}) = b_j^2\beta_j^2 \textstyle\prod_{i=1,(i\neq j)}^{k} (\rho_j - \rho_i)^2 \boldsymbol{\varrho}_j'D\Sigma_{11}D\boldsymbol{\varrho}_j .$$

Since $\partial T_j(Y, \rho)/\partial \rho_j = b_j\mathbf{z}_{2j}'D\mathbf{z}_{2j}$, we have

$$S = b_j\epsilon(\mathbf{z}_{2j}')D\epsilon(\mathbf{z}_{2j}) = b_j\beta_j^2 \textstyle\prod_{i=1,(i\neq j)}^{k} (\rho_j - \rho_i)^2 \boldsymbol{\varrho}_j'D\boldsymbol{\varrho}_j$$

using (6).

Thus

$$S^{-2}\operatorname{Var}[T_j(Y,\rho)] = \beta_j^{-2}\prod_{i=1,(i\neq j)}^{k}(\rho_j-\rho_i)^{-2}\boldsymbol{\rho}_j'D\Sigma_{11}D\boldsymbol{\rho}_j/(\boldsymbol{\rho}_j'D\boldsymbol{\rho}_j)^2,$$

and we apply the third criterion by choosing D to minimise $\boldsymbol{\rho}_j'D\Sigma_{11}D\boldsymbol{\rho}_j$ under the restrictions $\boldsymbol{\rho}_j'D\boldsymbol{\rho}_j$ = constant, $\mathbf{1}'D = 0$. This may be done using Lagrange multipliers in a similar way to that of Patterson [6] to obtain

$$D = \sigma^{-2}[\Sigma_{11}^{-1} - (\Sigma_{11}^{-1}\mathbf{1}\,\mathbf{1}'\Sigma_{11}^{-1}/\mathbf{1}'\Sigma_{11}^{-1}\mathbf{1})], \tag{10}$$

where Σ_{11} is given by (7). An important feature is that this D matrix is the same for all j.

5. Details of method. The D matrix to maximise the efficiency of the estimating equations, as measured by $S^{-2}\operatorname{Var}[T_j(Y,\rho)]$, depends on the true values of $\rho_1, \rho_2, \cdots, \rho_k$. We may proceed therefore as follows. We choose a D matrix with optimum efficiency for guessed ρ values denoted by $\rho_{10}, \rho_{20}, \cdots, \rho_{k0}$. These guessed values are used also to evaluate

$$K_{jp} = \theta_{k-p}^{j}b_j + \theta_{k-p}, \qquad p = 0, 1, 2, \cdots, k, \qquad j = 1, 2, \cdots, k,$$

where the guessed ρ values are used in the expressions for the θ's and b_j. Estimating equations (2) are now written as

$$\begin{aligned}&[(-1)^k\hat{\theta}_k\mathbf{y}_0' + (-1)^{k-1}\hat{\theta}_{k-1}\mathbf{y}_1'\\ &\quad + \cdots + \mathbf{y}_k']D[(-1)^kK_{j0}\mathbf{y}_0 + (-1)^{k-1}K_{j1}\mathbf{y}_1\\ &\quad + \cdots + K_{jk}\mathbf{y}_k] = 0, \qquad j = 1, 2, \cdots, k,\end{aligned} \tag{11}$$

and solved for $r_1, r_2, \cdots, r_k$. These new estimates may be used as initial values for a further iteration of the above procedure.

If reasonable efficiency over the whole range of ρ values can be obtained with just one D matrix for each sample size n some numerical labour is saved since then only one set of $\mathbf{y}_u'D\mathbf{y}_v$ values is needed for each problem. At least for small n it would be possible to tabulate such D matrices and this is the subject of further investigation.

6. Particular cases.

6.1. *Single exponential regression.* The estimator of ρ_1 is given by the solution for r_1 of the equation,

$$(-r_1\mathbf{y}_0' + \mathbf{y}_1')D(-K_{10}\mathbf{y}_0 + \mathbf{y}_1) = 0,$$

where $K_{10} = -[t_0 - \rho_{10}t_1](\rho_{10}t_0 - t_1)^{-1}$ and D is given by (10) where

$$\sigma^{-2}\Sigma_{11} = (1+\rho_{10}^2)U_0 - \rho_{10}(U_1 + U_{-1}).$$

This technique is very similar to that advocated by Patterson [6]. It differs only in the way the bias is treated.

6.2. *Double exponential regression.* Estimating equations for ρ_1 and ρ_2 are

$$\begin{aligned}&[r_1r_2\mathbf{y}_0' - (r_1+r_2)\mathbf{y}_1' + \mathbf{y}_2']D[K_{10}\mathbf{y}_0 - K_{11}\mathbf{y}_1 + \mathbf{y}_2] = 0,\\ &[r_1r_2\mathbf{y}_0' - (r_1+r_2)\mathbf{y}_1' + \mathbf{y}_2']D[K_{20}\mathbf{y}_0 - K_{21}\mathbf{y}_1 + \mathbf{y}_2] = 0,\end{aligned} \tag{12}$$

where $$K_{10} = \rho_{20}(b_1 + \rho_{10}), \qquad K_{11} = (b_1 + \rho_{10} + \rho_{20}),$$
$$K_{20} = \rho_{10}(b_2 + \rho_{20}), \qquad K_{21} = (b_2 + \rho_{10} + \rho_{20});$$
$$b_1 = -[t_0a_0 + 2t_1a_1 + 2t_2a_2][t_0(a_1 + \rho_{10}a_2) + t_1(1 + \rho_{20}^2 + 2a_2) - t_2\rho_{20}]^{-1},$$
$$b_2 = -[t_0a_0 + 2t_1a_1 + 2t_2a_2][t_0(a_1 + \rho_{20}a_2) + t_1(1 + \rho_{10}^2 + 2a_2) - t_2\rho_{10}]^{-1};$$
$$a_0 = 1 + (\rho_{10} + \rho_{20})^2 + \rho_{10}^2\rho_{20}^2, \qquad a_1 = -(\rho_{10} + \rho_{20})(1 + \rho_{10}\rho_{20}),$$
$$a_2 = \rho_{10}\rho_{20}.$$

The D matrix is given by (10) where Σ_{11} follows from
$$\sigma^{-2}\Sigma_{11} = a_0U_0 + a_1(U_1 + U_{-1}) + a_2(U_2 + U_{-2}).$$
Equations (12) may be written
$$F_1r_1r_2 - G_1(r_1 + r_2) + H_1 = 0,$$
$$F_2r_1r_2 - G_2(r_1 + r_2) + H_2 = 0,$$
where
$$F_1 = K_{10}y_0'Dy_0 - K_{11}y_0'Dy_1 + y_0'Dy_2, \qquad F_2 = K_{20}y_0'Dy_0 - K_{21}y_0'Dy_1 + y_0'Dy_2,$$
$$G_1 = K_{10}y_1'Dy_0 - K_{11}y_1'Dy_1 + y_1'Dy_2, \qquad G_2 = K_{20}y_1'Dy_0 - K_{21}y_1'Dy_1 + y_1'Dy_2,$$
$$H_1 = K_{10}y_2'Dy_0 - K_{11}y_2'Dy_1 + y_2'Dy_2, \qquad H_2 = K_{20}y_2'Dy_0 - K_{21}y_2'Dy_1 + y_2'Dy_2,$$
whence the solution of the quadratic equation
$$(F_1G_2 - F_2G_1)r^2 - (F_1H_2 - F_2H_1)r + (G_1H_2 - G_2H_1) = 0,$$
will give r_1 and r_2.

7. Example. The observations used in this example are those used by Lipton and McGilchrist [4] in obtaining maximum likelihood estimates in double exponential regression and came originally from [1] where a full description of them may be found.
They are

x	0	1	2	3	4	5	6	7
Y	10.430	4.703	2.327	1.140	0.615	2.325	0.170	0.117
x	8	9	10	11	12	13	14	15
Y	0.050	0.040	0.046	0.022	0.036	0.021	0.018	0.016

An initial value of $(\rho_{10}, \rho_{20}) = (0.3, 0.7)$ was taken simply because this is about the centre of the region $0 < \rho_1 < \rho_2 < 1$. The D matrix for these values is symmetrical about each diagonal so that it is merely necessary to tabulate (as is

TABLE I

D Matrix for (0.3, 0.7)

+0.88667	+0.77411	+0.47643	+0.20394	−0.00381	−0.15116	−0.25036	−0.31215	−0.34390	−0.34960	−0.33023	−0.28428	−0.20919	−0.10645
	+1.54972	+1.16493	+0.62035	+0.16097	−0.17775	−0.41103	−0.56036	−0.64246	−0.66727	−0.63836	−0.55394	−0.40972	
		+1.75636	+1.20744	+0.54039	−0.00273	−0.39341	−0.65240	−0.80422	−0.86596	−0.84560	−0.74299		
			+1.71212	+1.10069	+0.39353	−0.17058	−0.56529	−0.81177	−0.93309	−0.94314			
				+1.58915	+0.97113	+0.26652	−0.28550	−0.65706	−0.86657				
					+1.47708	+0.87823	+0.19666	−0.32559					
						+1.41438	+0.84529						

$t_0 = +20.7710$ $t_1 = +13.0384$ $t_2 = +4.9878$

D Matrix for (0.161, 0.509)

+0.91088	+0.52119	+0.18548	+0.02773	−0.10912	−0.16121	−0.18739	−0.19987	−0.20455	−0.20364	−0.16152	−0.18062	−0.14841	−0.08896
	+1.20041	+0.61282	+0.22233	−0.08911	−0.22125	−0.28841	−0.32127	−0.33521	−0.33681	−0.26808	−0.30080	−0.24742	
		+1.21400	+0.65354	+0.10930	−0.15511	−0.29273	−0.36167	−0.39345	−0.40294	−0.52298	−0.36485		
			+1.32106	−0.33299	−0.15362	−0.11010	−0.24668	−0.31377	−0.34201	−0.28015			
				+1.14883	+0.51910	+0.04851	−0.21336	−0.34669	−0.40757				
					+1.13184	+0.50653	+0.03921	−0.21905					
						+1.12466	+0.50277						

$t_0 = +16.103$ $t_1 = +6.422$ $t_2 = +1.517$

D Matrices

done in Table I) as follows:

$$\begin{matrix} d_{11} & d_{12} & \cdot & \cdot & \cdot & d_{1,13} & d_{1,14} \\ & d_{22} & \cdot & \cdot & \cdot & d_{2,13} & \\ & & \cdot & & & \cdot & \\ & & & \cdot & \cdot & & \\ & & & & \cdot \quad \cdot & & \\ & & & & d_{77} \quad d_{78} & & \end{matrix}$$

The full matrix is obtained by symmetry.
From this the following results may be found on a desk calculator:

$$\begin{aligned} \mathbf{y}_0'D\mathbf{y}_0 &= 281.387, & \mathbf{y}_1'D\mathbf{y}_1 &= 62.387, \\ \mathbf{y}_0'D\mathbf{y}_1 &= 132.443, & \mathbf{y}_1'D\mathbf{y}_2 &= 30.942, \\ \mathbf{y}_0'D\mathbf{y}_2 &= 65.671, & \mathbf{y}_2'D\mathbf{y}_2 &= 15.349, \\ K_{10} &= -3.562, & K_{20} &= -0.793, \\ K_{11} &= -4.389, & K_{21} &= -2.344. \end{aligned}$$

These lead to the quadratic equation

$$89.360r^2 - 59.996r + 7.383 = 0$$

giving

$$r_1 = 0.161 \quad \text{and} \quad r_2 = 0.509,$$

and these two roots are then treated as the initial values for a second iteration. Matrix D is given as above in Table I, and corresponding results to those given above are

$$\begin{aligned} \mathbf{y}_0'D\mathbf{y}_0 &= 207.105, & \mathbf{y}_1'D\mathbf{y}_1 &= 45.041, \\ \mathbf{y}_0'D\mathbf{y}_1 &= 96.538, & \mathbf{y}_1'D\mathbf{y}_2 &= 22.270, \\ \mathbf{y}_0'D\mathbf{y}_2 &= 47.718, & \mathbf{y}_2'D\mathbf{y}_2 &= 11.014, \\ K_{10} &= -2.283, & K_{20} &= -0.560, \\ K_{11} &= -3.976, & K_{21} &= -3.318, \end{aligned}$$

$$59.549r^2 - 38.366r + 4.164 = 0,$$

$$r_1 = 0.138, \qquad r_2 = 0.506.$$

These results agree closely with those obtained in [4].

REFERENCES

[1] CORNELL, R. G. (1962). A method for fitting linear combinations of exponentials. *Biometrics* **18** 104–113.

[2] Durbin, J. (1960). Estimation of parameters in time-series regression models. *J. Roy. Statist. Soc. Ser. B.* **22** 139–153.

[3] Hartley, H. O. (1948). The estimation of non-linear parameters by internal least squares. *Biometrika* **35** 32–45.

[4] Lipton, S. and McGilchrist, C. A. (1963). Maximum likelihood estimators of parameter in double exponential regression. *Biometrics* **19** 144–151.

[5] Lipton, S. and McGilchrist, C. A. (1964). The derivation of methods for fitting exponential regression curves. *Biometrika* **51** 504–508.

[6] Patterson, H. D. (1958). The use of autoregression in fitting an exponential curve. *Biometrika* **45** 389–400.