

## REGRESSION ANALYSIS OF DEPENDENT ERROR MODELS

C.A. MCGILCHRIST

A method of analysing the general linear regression model is described, for the case where the observations are correlated. For many applications the correlations are structured, with neighbouring observations being more strongly correlated than those some distance apart in time or space. Such correlation structures may often be assumed to belong to some class of models indexed by a small number of parameters. Estimation and inference procedures which are able to cope with a wide range of correlation models, are described and the methods are applied to problems which occur in biometry.

### 1. Introduction

Often in biometry it is required to fit a linear regression model in the presence of error which is correlated over successive observations. The model may be written

$$Y_t = \underline{x}'_t \beta + E_t$$

where  $Y$  is the dependent variable,  $\underline{x}$  is a vector of  $p$  regression

---

Received 11 November 1985. This is the text of an address given at a symposium which marked the seventy fifth anniversary of the University of Queensland.

---

Copyright Clearance Centre, Inc. Serial-fee code: 0004-9727/86  
\$A2.00 + 0.00.

variables and  $E$  is a residual or error term distributed with zero mean and constant variance. The ordering variable  $t$  is usually time or position in space so that neighbouring  $E$  terms are often correlated.

Let

$$\underline{y}_t = [y_1, y_2, \dots, y_t]', \quad X_t = [x_1, x_2, \dots, x_t]', \quad \underline{e}_t = [E_1, E_2, \dots, E_t]'$$

be accumulated arrays of variables up to  $t$  so that in matrix notation the model becomes

$$\underline{y}_t = X_t \underline{\beta} + \underline{e}_t, \quad \text{Var } \underline{e}_t = \sigma_e^2 \Sigma_t.$$

In the applications we consider, the variance matrix  $\Sigma_t$  is modelled in some way using a parsimonious vector of parameters  $\underline{\theta}$ . Thus  $\underline{\beta}$ ,  $\sigma_e^2$  and  $\underline{\theta}$  are the parameters of the model and the object of this paper is to present a method of inference which applies to a wide variety of such variance matrices.

### 2. Likelihood Estimation

If the errors are now assumed to be normally distributed, the likelihood function based on  $\underline{y}_t$ , denoted by  $L_t$ , may be written as

$$-2 \ln L_t = t \ln(2\pi\sigma_e^2) + \ln|\Sigma_t| + \sigma_e^{-2} (\underline{y}_t - X_t \underline{\beta})' \Sigma_t^{-1} (\underline{y}_t - X_t \underline{\beta})$$

For any given  $\underline{\theta}$  this function may easily be differentiated with respect to  $\sigma_e^2$  or  $\underline{\beta}$  to give estimators

$$\underline{\tilde{\beta}}_t = H_t^{-1} X_t' \Sigma_t^{-1} \underline{y}_t, \quad \tilde{\sigma}_{et}^2 = t^{-1} (\underline{y}_t - X_t \underline{\tilde{\beta}}_t)' \Sigma_t^{-1} (\underline{y}_t - X_t \underline{\tilde{\beta}}_t)$$

where  $H_t = X_t' \Sigma_t^{-1} X_t$ . While these estimators maximise  $L_t$  for a particular  $\underline{\theta}$ , there is no explicit expression for the value of  $\underline{\theta}$  which maximises  $L_t$  and such a value must be found by convergence techniques. To reduce the dimension of such search procedures, it is clearly better to find that value of  $\underline{\theta}$  which maximises  $L_t(\underline{\tilde{\beta}}_t, \tilde{\sigma}_{et}^2, \underline{\theta})$ ,

which is a function of  $\underline{\theta}$  alone. The equations for  $\tilde{\beta}_t$ , and  $\tilde{\sigma}_{et}^2$  then give the appropriate estimates for  $\underline{\beta}$  and  $\sigma_e^2$  respectively. The function to maximise is then

$$-2 \ln L_t(\tilde{\beta}_t, \tilde{\sigma}_{et}^2, \underline{\theta}) = t \ln(2\pi\tilde{\sigma}_{et}^2) + \ln|\Sigma_t| + t = \text{constant} + A_t(\underline{\theta}) ,$$

where  $A_t(\underline{\theta}) = \ln|\Sigma_t| + t \ln(\underline{y}_t - X_t \tilde{\beta}_t)' \Sigma_t^{-1} (\underline{y}_t - X_t \tilde{\beta}_t)$

and maximising  $L_t(\underline{\theta})$  is equivalent to minimising  $A_t(\underline{\theta})$  .

A problem in carrying out such a convergence procedure is that  $\Sigma_t$  may be a large matrix and finding  $\Sigma_t^{-1}$  and  $|\Sigma_t|$  may become difficult. To overcome such difficulties we adopt recursive procedures. Let

$$\Sigma_{t+1} = \begin{pmatrix} \Sigma_t & \sigma_t \\ \sigma_t' & \sigma_{t+1,t+1} \end{pmatrix} , \quad \underline{q}_t = \Sigma_t^{-1} \sigma_t , \\ c_t = \sigma_{t+1,t+1} - \sigma_t' \underline{q}_t .$$

It is then easily shown that

$$\Sigma_{t+1}^{-1} = \begin{pmatrix} \Sigma_t^{-1} & \underline{0}_t \\ \underline{0}_t' & 0 \end{pmatrix} + c_t^{-1} \begin{pmatrix} -\underline{q}_t \\ 1 \end{pmatrix} [-\underline{q}_t', 1] ,$$

$$|\Sigma_{t+1}| = c_t |\Sigma_t| .$$

with initial conditions of  $\underline{q}_0$  being a null vector,  $c_0 = \sigma_{011}$  and the determinant of  $\Sigma_0$  being unity. Although the recurrence relationship for the determinant is directly usable, the recurrence for the inverse must be included into the quadratic form occurring in  $A_t(\underline{\theta})$ .. Letting

$$W_t = (c_t + z_t' H_t^{-1} z_t)^{-1/2} [y_{t+1} - \underline{y}_t' \underline{q}_t - z_t' \tilde{\beta}_t] ,$$

where  $\underline{z}_t = \underline{x}_{t+1} - X_t' \underline{q}_t$  , then it can be shown that

$$A_t(\underline{\theta}) = \sum_{i=0}^{t-1} \ln c_i + t \ln \sum_{i=0}^{t-1} W_i^2$$

Two updating formulae are

$$\tilde{\beta}_{t+1} = \tilde{\beta}_t + (c_t + z_t' H_t^- z_t)^{-1/2} W_t H_t^- z_t$$

and

$$H_{t+1}^- = H_t^- - (c_t + z_t' H_t^- z_t)^{-1} H_t^- z_t z_t' H_t^- ,$$

and these formulae apply provided the rank of  $H$  does not increase in going from  $t$  to  $t+1$ . The rank will increase by one if some variable whose coefficient has not been estimable becomes estimable for the first time at  $t+1$ . The updating formulae are then given in McGilchrist, Sandland and Hennessy [6]. Using such methods,  $\tilde{\beta}_0$  may be taken to be a vector of zeros, while  $H_0^-$  is a matrix of zeros.

The choice of search procedure for the minimum of  $A_t(\underline{\theta})$  over  $\underline{\theta}$  space, depends on how easily the derivatives of  $A_t(\underline{\theta})$  may be found. Usually there is a trade-off between extra complexity of programming, with associated increased storage of variables, and speed of convergence. It is, nevertheless, possible to write a program for estimating all parameters which calls on a subroutine to evaluate  $c$  and  $q$  for the particular variance matrix involved. The method hinges on whether or not such a subroutine can be efficiently written. The remaining part of this paper examines various problems occurring in biometry, outlines the structure of the variance matrix involved and gives methods for evaluating  $c$  and  $q$ .

An added bonus of this method is that, provided an appropriate model is being fitted, the  $W_t$  are distributed as independent  $N(0, \sigma_e^2)$  and are termed recursive residuals. They may be used in model diagnostics to test for residual correlation, constancy of variance and so on. Also very useful are the recursive estimates of  $\underline{\beta}$  give by  $\tilde{\beta}_t$ . These estimates may be graphed to check on constancy of regression over time. Such procedures are described in Brown et al. [1].

3. Stationary Error Models

If the covariance structure is that of a stationary, zero mean, Gaussian process then  $\Sigma_t$  is Toeplitz of the form

$$\Sigma_t = \begin{pmatrix} \sigma_0 & \sigma_1 & \dots & \sigma_{t-1} \\ \sigma_1 & \sigma_0 & \dots & \sigma_{t-2} \\ \cdot & \cdot & \dots & \cdot \\ \sigma_{t-1} & \sigma_{t-2} & \dots & \sigma_0 \end{pmatrix}$$

For such a structure it is easy to show that

$$\begin{pmatrix} q_{t+1} \\ q_t \end{pmatrix} = \begin{pmatrix} 0 \\ q_t \end{pmatrix} + \rho_{t+1}^p \begin{pmatrix} 1 \\ -\text{rev } q_t \end{pmatrix}, \quad c_{t+1} = c_t [1 - (\rho_{t+1}^p)^2]$$

$$\rho_{t+1}^p = [\sigma_{t+1} - (\text{rev } \sigma_t^i) q_t] / c_t.$$

Here  $\rho_t^p$  is the partial autocorrelation coefficient of lag  $t$ , which is the correlation between  $E_u$  and  $E_{u+t}$  given the intervening values of  $E$ ; and  $\text{rev}$  is an operator which reverses the components of a vector. Thus

$$\text{rev } \sigma_t^i = [\sigma_1, \sigma_2, \dots, \sigma_t]$$

The vector  $q_t$  may be shown to have the property that  $q_t^i e_t$  is the minimum variance, unbiased, linear predictor of  $E_{t+1}$ , while  $\sigma_e^2 c_t^2$  is the variance of that linear predictor. The above is known as Levinson's algorithm and a survey of its rediscoveries and uses is given in Moretting [8].

We then have a simple set of recurrences which apply to any stationary covariance structure. It should be noticed that the covariance terms  $\sigma_t$  enter the computation only through the partial correlation  $\rho_t^p$ ; hence a subprogram to generate  $c$  and  $q$  may be written for any stationary model with only  $\sigma_t$ , or equivalently  $\rho_t^p$ , being changed in going from one model to another.

3.1 ARMA MODELS For each autoregressive, moving average model of the type

$$Y_t + \sigma_1 Y_{t-1} + \dots + \sigma_p Y_{t-p} = U_t + \theta_1 U_{t-1} + \dots + \theta_q U_{t-q}$$

the autocovariances may be generated by the Yule-Walker equations and it is often convenient to represent the error correlations by this class of models.

In biometry the above procedures have been used to model the growth of a single organism. The relative growth rate may change with time or respond to the supply of nutrients so that a regression model with time dependent errors will apply. Such methods are illustrated in Sandland and McGilchrist [10] and McGilchrist et al [7]. The same methods may be used in the study of population growth or annual deaths from specific causes such as motor vehicle accidents.

3.2 FRACTIONAL GAUSSIAN PROCESSES Such processes were introduced by Mandelbrot and van Ness [4] and have covariance function

$$\sigma_t = 1 \quad , \quad t=0, \\ (1/2) [(t+1)^{2\theta} + (t-1)^{2\theta} - 2t^{2\theta}], \quad 0 < \theta < 1, \quad t > 0.$$

The covariance functions, which have a long tail, were introduced by Mandelbrot to describe the long term dependence in share market movements and similar properties in the run-off of the Nile River at Aswan. The covariance model is one method of explaining the celebrated Hurst law in hydrology.

3.3 VARIANCE COMPONENT MODELS Sometimes it may be convenient to think of the error term in the regression as being composed of more than one component, each with its own lagged correlation structure. For example, in an environmental study the distribution of the pollution variable  $Q$  follows the model

$$P(Q > q | R) = \exp -[(1+q)/\delta e^{\theta R}]^\gamma \quad , \quad q > 0 \quad ,$$

where  $R$  is an environmental risk variable which is made up of climatic

factors and soil factors which affect the chance of a high pollution level being recorded. Some of these factors which depend on variation of temperature, wind, etcetera contribute a correlation which rapidly decreases to zero as time lag increases. However, other factors such as general wetness of the area, vary slowly and contribute a relatively long term dependence pattern. Such correlations may be conveniently modelled as the sum of two component error models. Once the environmental risk is modelled, there is no difficulty in using the above theory to fit the model since the covariance structure is still stationary.

#### 4. Random Block Structures

In the random block structure the error component is composed of a random block effect plus the residual error term. If  $Y_{ij}$  is the  $j^{th}$  observation in block  $i$ , then the observation vector may be ordered

$$Y_{11}, Y_{12}, \dots, Y_{21}, Y_{22}, \dots, Y_{31}, \dots$$

The model is  $Y_{ij} = x_{ij}'\beta + B_i + E_{ij}$  and various types of correlation structure may be considered appropriate for the  $B$  and  $E$  terms.

4.1 INDEPENDENT BLOCKS, INDEPENDENT ERRORS If  $E_{ij}$  are distributed as independent  $N(0, \sigma_e^2)$  and  $B_i$  as independent  $N(0, \theta \sigma_e^2)$  then the variance matrix is of the type  $\sigma_e^2 \Sigma$  with

$$\Sigma = \begin{pmatrix} I + \theta \underline{1}\underline{1}' & 0 & \dots \\ 0 & I + \theta \underline{1}\underline{1}' & \dots \\ \cdot & \cdot & \dots \end{pmatrix},$$

where each diagonal block of  $I + \theta \underline{1}\underline{1}'$  is of dimension equal to the size of the block and  $\underline{1}$  is an appropriate vector of ones. The structure is no longer Toeplitz but suitable recurrences may be easily found to be:

(i) If  $t \rightarrow t + 1$  remains within a block

$$a_{t+1} = \begin{pmatrix} a_t / c_t \\ 1 - c_t^{-1} \end{pmatrix}, \quad c_{t+1} = 2 - c_t^{-1}.$$

(ii) If  $t \rightarrow t + 1$  goes to a new block

$$\underline{g}_{t+1} = \underline{0}_{t+1} \quad , \quad c_{t+1} = 1 + \theta .$$

There is no difficulty in analysing experiments with blocks containing different numbers of observations. Examples occur in experiments on litters of mice using different size litters as blocks; and observations on people where the person effect is the block effect.

4.2 DEPENDENT BLOCKS, INDEPENDENT ERRORS Sometimes a more appropriate model for blocks is that of an accumulated error process,

$$B_1 = \mu \quad , \quad B_{i+1} = B_i + U_i \quad , \quad i=1,2,\dots$$

where the  $U_i$  are independent  $N(0, \theta \sigma_e^2)$ . In this case the variance matrix is of the type

$$\Sigma_t = I_t + \theta V_t \quad , \quad V_t = \begin{pmatrix} 0_{11} & 0_{12} & 0_{13} & \dots \\ 0_{21} & 1_{22} & 1_{23} & \dots \\ 0_{31} & 1_{32} & 2_{33} & \dots \\ \cdot & \cdot & \cdot & \dots \end{pmatrix}$$

where  $i_{j'k}$  is a matrix of  $i$ 's with rows, columns equal in size to blocks  $j, k$  respectively. For such a variance structure

$$\underline{g}_{t+1} = \begin{pmatrix} c_t^{-1} \underline{g}_t \\ 1 - c_t^{-1} \end{pmatrix} \quad , \quad c_{t+1} = 2 - c_t^{-1} + \theta \delta_{t+1}$$

where  $\delta_{t+1}$  is 1 if  $t \rightarrow t + 1$  goes to a new block but is otherwise zero.

Such models are very similar to the Kalman filter models as discussed in Duncan and Horn [2] and may often be quite reasonable descriptions of block effects in agricultural field trials where blocks have been chosen perpendicularly to the fertility gradient and the fertility change up the gradient behaves like an integrated error process. Indeed, the success



of the Papadakis style analysis advocated by Wilkinson et al [11] depends on such integrated error processes applying.

4.3 STRUCTURED BLOCKS AND DEPENDENT ERRORS It is possible to carry on the study to blocks which have a nested or crossed structure and errors with a Toeplitz variance structure. An example of such a structure is discussed in Sallas and Harville [9]. Each such structure has a different set of  $c$ ,  $g$  recurrence relations.

### 5. Field Experiments

In field experiments observations are often made at points of a regularly spaced grid and

$$Y_{i,j} = \text{observation at } (i,j), \quad i=1,2,\dots,I; \quad j=1,2,\dots,J$$

obeys the linear regression model  $Y_{i,j} = \underline{x}'_{i,j}\beta + E_{i,j}$  with

$$\text{Cov}(E_{i,j}, E_{i+r,j+s}) = \sigma_{rs}$$

The symmetry of the two dimensional stationary covariance structure may be exploited to obtain recurrence relations for  $c_{t+I}$  and  $g_{t+I}$  in terms of  $c_t$  and  $g_t$ . Such recurrences are given in McGilchrist and Knudsen [5] or Gleeson and McGilchrist [3].

It has been noted that the residual component  $E_{i,j}$  of the model often does not appear to be stationary. One method of handling this problem is to treat  $E_{i,j}$  as an integrated error process and taking row and column differences produces a stationary residual term. The regression model still applies but the differenced variable now becomes the dependent variable and a similarly treated  $\underline{x}_{i,j}$  term becomes the vector of regression variables. For many agricultural field experiments, such a method is appropriate and is at the heart of the Papadakis style analyses. The factor which is differenced out is, almost certainly, a slowly varying fertility component.

If the experiment covers a larger geographical scale, the slowly varying component may be more appropriately modelled by a separate variance component with a faster varying component representing the more local variation.

### References

- [1] R.L. Brown, J. Durbin and J.M. Evans, "Techniques for testing the constancy of regression relationships over time", *J.R. Statist. Soc. Ser. B.* 37 (1975) 149-163.
- [2] D.B. Duncan and S.D. Horn, "Linear dynamic recursive estimation from the viewpoint of regression analysis", *J. Amer. Statist. Assoc.* 67 (1972), 815-821.
- [3] A.C. Gleeson and C.A. McGilchrist, "Recursive residuals on a rectangular lattice", *Biometrical* 26 (1984), 675-681.
- [4] B.B. Mandelbrot and J.W. Van Ness, "Fractional Brownian motions, fractional noises and applications", *SIAM Rev.* 10 (1968), 422-437.
- [5] C.A. McGilchrist and G.J. Knudsen, "Estimation of regression models in a field with stationary error structure", *Contributions to Statistics. Essays in Honour of Normal L. Johnson.* (1983), North-Holland Publishing Co.
- [6] C.A. McGilchrist, R.L. Sandland and J.L. Hennessy, "Generalised inverses used in recursive estimation of the general linear model", *Austral. J. Statist.* 25 (1983), 321-328.
- [7] C.A. McGilchrist, R.L. Sandland and L.J. Hills, "Estimation in regression models with stationary, dependent errors", *Comm. Statist. A Theory Methods A10*, (1981) 2563-2580.
- [8] P.A. Morettin, "The Levinson algorithm and its applications in time series analysis", *Internat. Statist. Rev.* 52 (1984), 83-92.
- [9] W.M. Sallas, and D.A. Harville, "Best linear recursive estimation for mixed linear models", *J. Amer. Statist. Assoc.* 76 (1982), 860-869.

- [10] R.L. Sandland and C.A. McGilchrist, "Stochastic growth curve analysis", *Biometrics* 35 (1979), 255-271.
- [11] G.N. Wilkinson, S.R. Eckert, T.W. Hancock and O. Mayo, "Nearest neighbour (NN) analysis of field experiments (with discussion)", *J.R. Statist. Soc. Ser. B.* 45 (1983), 151-211.

Department of Statistics  
University of New South Wales  
P.O. Box 1  
Kensington, N.S.W. 2033  
Australia.