REPEATED MEASUREMENT ARMA-PROCESSES I

by

S. H. C. du Toit

Technical Report No. 276
August 1990

Prepared Under the Auspices of
National Science Foundation Grant
DMS 87-08083
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
REPEATED MEASUREMENT ARMA-PROCESSES I

by

Stephen H. C. du Toit
University of Pretoria and Stanford University

Technical Report No. 276

Prepared Under the Auspices
of
National Science Foundation Grant
DMS 87-08083
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
REPEATED MEASUREMENT ARMA-PROCESSES I

by

S. H. C. du Toit

University of Pretoria and Stanford University

ABSTRACT

Longitudinal studies are carried out in many fields of application, such as biology, psychology, agriculture, engineering, sociology, and medicine. In many such studies a response is measured $T$ times on each of $N$ experimental units. The concept "response on an experimental unit" is used here to cover a variety of situations. Examples of these are: the height of a child measured annually from one to twelve years, the output of an electronic filter measured at ten time intervals, the diastolic blood pressure of a person measured monthly over a period of several years and the intelligence factor of a personality questionnaire measured six times at equally spaced intervals.

Possible changes in the pattern of responses may occur over time and it is frequently possible to present these changes by means of a function which may be linear or nonlinear in the parameters. It is often also realistic to assume that the errors are generated by an autoregressive moving average (ARMA) process.

Since this paper deals with situations where the number of time points ($T$) are relatively small whereas the number of repetitions ($N$) are large, the usual stationarity assumptions necessary for estimation procedures when dealing with a single time series may be relaxed. As a result of this, a general closed form expression for the covariance structure of the errors is derived from which a number of special cases follow. This closed form expression is obtained as a combination of well-known definitions for ARMA($n,m$) processes and state-space equations for linear, discrete stochastic sequences.

Key words: Longitudinal analysis, time series, autoregressive moving average.

REPEATED MEASUREMENT ARMA-PROCESSES I

S. H. C. du Toit

1. Introduction.

Longitudinal studies are carried out in many fields of application, such as biology, psychology, agriculture, engineering and medicine. In many of these studies a response is measured $T$ times on each of $N$ experimental units. The concept “response on an experimental unit” is used here to cover a wide range of situations. Examples of these are: The height of a child measured annually from one to twelve years, the output of an electronic filter measured at ten time intervals, the diastolic blood pressure of a person measured monthly over a period of several years and the intelligence factor of a personality questionnaire measured six times at equally spaced intervals.

Changes in the pattern of responses may occur over time and it is frequently possible to represent these changes by means of a response function which may be linear or nonlinear in the parameters. When measurements over time are made on the same set of individuals, it is logical to assume that the measurement errors will be correlated.

The purpose of this report is to develop a framework for the analysis of repeated measurements when the change in time follows a deterministic trend which may be adequately described by a suitable mathematical function. In addition, it will be assumed that the errors are generated by an autoregressive moving average (ARMA) process. The mathematical model which fulfills the above requirements is given in Section 2. In Section 3 a general procedure for the analysis of structured mean and covariance matrices is considered.

It is shown in Section 4 that a closed form expression for the covariance matrix of the errors may be derived by using a combination of the well-known definition for ARMA $(n,m)$ processes (e.g. Box and Jenkins (1976), Anderson (1971, 1975, 1977)) and the so-called state-space equations for linear, discrete, stochastic sequences (e.g. Kuo (1970), Cadzow and Martens (1970), Myers and Tapley (1976) and Spall and Wall (1984)).

Since this report deals with situations where the number of time points are relatively small whereas the number of repetitions are large, the usual stationarity assumptions necessary for estimation procedures when dealing with a single time series, may be relaxed. As a result of this, a number of different types of ARMA models will be introduced in Section 5. In Section 6, practical applications are given.
2. The general model.

Suppose that the change in the response pattern over time may be represented by the following set of regression equations:

\[ y_i = f(\theta, t_i) + e_i, \quad i = 1, 2, \ldots, T, \]  

(2.1)

where \( f(\theta, t_i) \) is the response function, \( \theta \) is a \( r \)-dimensional vector of unknown but fixed parameters and the \( t_i, i = 1, 2, \ldots, T \) are the measurement times. It is further assumed that the unobservable error variates \( e_i \) are generated by an autoregressive moving average (ARMA \((n, m)\)) process defined by

\[ e_t + \sum_{j=1}^{n} \alpha_j e_{t-j} = u_t + \sum_{\ell=1}^{m} \beta_{\ell} u_{t-\ell}, \quad t = \ldots, -1, 0, 1, \ldots \]  

(2.2)

The sequence \( \{u_t\} \) consists of Gaussian distributed variates, with

\[ E(u_t) = 0 \]  

(2.3)

\[ E(u_t u_s) = \delta_{ts} \psi^2, \]  

(2.4)

where \( \delta_{ts} = 1, \ t = s \) and \( \delta_{ts} = 0 \) otherwise.

Let \( f(\theta) \) denote a \( T \)-component vector with typical element \( f(\theta, t_i) \). If the response function is linear in the parameters, e.g.

\[ f(\theta, t_i) = \theta_1 + \theta_2 t_i, \quad i = 1, 2, \ldots, T \]

then

\[ f(\theta) = B \theta, \]  

(2.5)

where \( B : T \times r \) is an appropriate design matrix. Growth over time is often described in terms of a response function which is nonlinear in the parameters, e.g.

\[ f(\theta, t_i) = \theta_1 + \theta_2 \theta_3^t_i, \quad i = 1, 2, \ldots, T, \]

so that

\[ f(\theta) = (f(\theta, t_1), f(\theta, t_2), \ldots, f(\theta, t_T))^t \]  

(2.6)

The set of regression equations (2.1) can therefore be written in matrix notation as

\[ y = f(\theta) + e \]  

(2.7)
Provided (see e.g. Box and Jenkins (1976)) that the conditions for invertibility are met, (2.2) may be expressed as

$$e_k = \sum_{j=0}^{\infty} \pi_j u_{k-j},$$

(2.8)

where \{\pi_j\} is a series of appropriate weights. Since the \{u_k\} is a series of normal random variables it follows that \(e\) has a \(N(O, \Sigma)\) distribution, with \(\Sigma\) yet to be determined. From (2.7) and (2.8) it therefore follows that

$$y \sim N(f(\theta), \Sigma).$$

(2.9)

Consider a \(T \times N\) matrix variate \(Y\), where

$$Y = (y_1, y_2, \ldots, y_N)$$

(2.10)

Suppose further that the first \(N_1\) columns of \(Y\) is a random sample from a \(N(f(\theta_1), \Sigma)\) population, the next \(N_2\) columns an independent random sample from a \(N(f(\theta_2), \Sigma)\) population, \ldots, and that the last \(N_g\) columns is an independent random sample from a \(N(f(\theta_g), \Sigma)\) population, where

$$N = N_1 + N_2 + \cdots + N_g$$

(2.11)

Let

$$\Xi : (T \times g) = (f(\theta_1), f(\theta_2), \ldots, f(\theta_g))$$

(2.12)

and

$$A : (N \times g) = \begin{pmatrix}
  j_1 & 0 & \cdots & 0 \\
  0 & j_2 & \cdots & 0 \\
  & & \vdots & \\
  0 & 0 & \cdots & j_g
\end{pmatrix},$$

(2.13)

where \(j_k\) denotes a unit vector of order \(N_k\), \(k = 1, \ldots, g\). From the above it then follows that \(Y\) has a matrix normal distribution with

$$E(Y) = \Xi A'$$

(2.14)

Example.

Suppose \(T = 3\) repeated measurements of a reaction time test are made and that the scores of \(N_1 = 3\) males and \(N_2 = 4\) females are recorded. Suppose further that the scores follow a linear trend over time. Then

$$Y = \begin{pmatrix}
  y_{11} & y_{12} & y_{13} & | & y_{14} & y_{15} & y_{16} & y_{17} \\
  y_{21} & y_{22} & y_{23} & | & y_{24} & y_{25} & y_{26} & y_{27} \\
  y_{31} & y_{32} & y_{33} & | & y_{34} & y_{35} & y_{36} & y_{37}
\end{pmatrix}$$
\[ f(\theta_k) = B\theta_k, \quad k = 1, 2, \text{ where} \]
\[
B = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{pmatrix} \quad \text{and} \quad \theta_k = \begin{pmatrix} \theta_{1k} \\ \theta_{2k} \end{pmatrix}.
\]

Therefore \( E(Y) = \Xi A' \), where
\[
A' = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \Xi = (B\theta_1 B\theta_2).
\]

Suppose \( \gamma_i \) denotes a typical element of the \( k \) component vector \( \gamma \) of unknown parameters and that \( \ln L \) denotes the logarithm of the likelihood function of the matrix variate \( Y \). A typical element of the information matrix is then given by \(-E \left\{ \frac{\partial^2 \ln L}{\partial \gamma_i \partial \gamma_j} \right\} \). It will be shown in the next section that the more difficult task of writing subroutines to evaluate \( \frac{\partial \ln L}{\partial \gamma_i} \) and \(-E \left\{ \frac{\partial^2 \ln L}{\partial \gamma_i \partial \gamma_j} \right\} \) may be avoided by expressing these derivatives in terms of \( \frac{\partial \Xi}{\partial \gamma_i} \) and \( \frac{\partial \Sigma}{\partial \gamma_i} \). An algorithm for the computation of the maximum likelihood estimates of the unknown parameters in models with structured mean and covariance matrices will then be given. This algorithm allows for the imposition of nonlinear equality or inequality constraints on the parameters and for the use of numerical derivatives.

3. A general procedure for the analysis of structured mean- and covariance matrices.

Suppose that the results obtained from a multivariate study can be adequately described by the model (cf. (2.10)-(2.14))

\[ E(Y) : (T \times N) = \Xi A', \quad (3.1) \]

where \( A \) is a known \( N \times g \) design matrix and \( \Xi \) is a \( T \times g \) matrix representing means or fixed effects. It is further assumed that the columns of \( Y - \Xi A' \) are independent \( N(O, \Sigma) \) variates.

Let
\[
\bar{Y} = YA(A' A)^{-1} \quad (3.2)
\]
\[
S = N^{-1}(Y - \bar{Y} A')(Y - \bar{Y} A')' \quad (3.3)
\]
\[
C = N^{-1}A' A \quad (3.4)
\]

For \( g = 1 \), \( Y \) represents a random sample from a single population and \( \Xi \) is the \( T \times 1 \) population mean vector while \( A \) is a \( N \times 1 \) vector with unit elements. \( \bar{Y} \) is the sample mean.
vector, $S$ is the sample covariance matrix and $C = 1$. For $g > 1$, $Y$ represents samples from $g$ populations with common covariance matrix $\Sigma$ and different mean vectors forming the columns of $\Xi$. Each row of $A$ then has $g - 1$ zero elements and one unit element.

Maximum normal likelihood estimates of the unknown parameters are obtained by minimizing $-\frac{2}{N} \ln L$ with the constant term omitted yielding the discrepancy function (Browne 1982)

$$F(\gamma) = \ln |\Sigma| - \ln |S| + \text{tr} \left[ S^* \Sigma^{-1} \right] - T$$

(3.5)

where

$$S^* = S + (\bar{Y} - \Xi)C(\bar{Y} - \Xi)'$$

(3.6)

Its minimum, $\hat{F} = F(\hat{\gamma})$, provides the likelihood ratio statistic, $N\hat{F}$, for testing the fit of the model.

Unless the model yields maximum likelihood estimates in closed form, it will be necessary to make use of an iterative procedure to minimize the discrepancy function. The optimization method which will be described next, was developed by Browne and du Toit (1987) and is based on the Fisher scoring algorithm, which in the case of structured mean and covariance matrices, may be regarded as a sequence of Gauss-Newton steps with quantities to be fitted as well as the weight matrix changing at each step. Fisher scoring algorithms require the gradient vector of the discrepancy function and use of the information matrix as an approximation to the Hessian matrix. Elements of the gradient vector, $g(\gamma)$, and approximate Hessian matrix $H(\gamma)$ of $\frac{1}{2}F(\gamma)$ are given by

$$\frac{1}{2} \frac{\partial F}{\partial \gamma_i} = [g(\gamma)]_i = - \left\{ \text{tr} Q \frac{\partial \Xi}{\partial \gamma_i} + \frac{1}{2} \text{tr} P \frac{\partial \Sigma}{\partial \gamma_i} \right\},$$

(3.7)

where

$$Q = C(\bar{Y} - \Xi)' \Sigma^{-1}$$

$$P = \Sigma^{-1}(S^* - \Sigma) \Sigma^{-1}$$

(3.8)

(3.9)

$$\frac{1}{2} \frac{\partial^2 F}{\partial \gamma_i \partial \gamma_j} \approx [H(\gamma)]_{i,j} = \text{tr} \left[ C \frac{\partial \Xi'}{\partial \gamma_i} \Sigma^{-1} \frac{\partial \Xi}{\partial \gamma_j} \right]$$

$$+ \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \gamma_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \gamma_j} \right]$$

(3.10)

Suppose that $\gamma_k$ is the $k$-th approximation to the $\hat{\gamma}$ which minimizes $F(\gamma)$. Let $g_k = g(\gamma_k)$, $H_k = H(\gamma_k)$ and $F_k = F(\gamma_k)$. The next approximation is obtained from

$$\hat{\gamma}_{k+1} = \hat{\gamma}_k + \alpha_k \delta_k,$$

(3.11)
where
\[
\delta_k = -H_k^{-1} G_k
\]  \hspace{1cm} (3.12)
and \(\alpha_k\) is a step size parameter chosen initially as 1 and then successively halved until \(F_{k+1} \leq F_k\).

A convenient feature of the Fisher scoring algorithm is that an estimate, \(N^{-1}\{H(\hat{\gamma})\}^{-1}\) of the asymptotic covariance matrix of estimator \(\gamma\) is available on convergence as a byproduct of the calculations.

It may be necessary to minimize \(F(\gamma)\) subject to \(r\) nonlinear constraints of the form
\[
c(\gamma) = 0, \hspace{1cm} (3.13)
\]
where \(c(\gamma)\) is a continuously differentiable \(T \times 1\) vector valued function of \(\gamma\). Let \(c_k = c(\gamma_k)\) and \(L_k = L(\gamma_k)\). Then the linear Taylor approximation for the constraint function is
\[
c(\gamma) \approx c_k + L_k \delta, \hspace{1cm} (3.14)
\]
where \(\delta = \gamma - \gamma_k\). A typical element of the \(\tau \times k\) Jacobian matrix \(L_k\) is given by
\[
[L_k]_{i,j} = \frac{\partial c_i}{\partial \gamma_j} \bigg|_{\gamma = \gamma_k},
\]
where \(c_i = [c(\gamma)]_i\). Consequently the nonlinear constraints (3.13) may be approximated by the linear constraints
\[
L_k \delta = -c_k \hspace{1cm} (3.16)
\]
The increment vector \(\delta_k\) is obtained (Browne and du Toit, 1987) as the solution of
\[
\begin{bmatrix}
\delta_k \\
\lambda_k
\end{bmatrix}
= \begin{bmatrix}
H_k + L_k' D_k L_k & L_k \\
L_k & 0
\end{bmatrix}^{-1} \begin{bmatrix}
-(G_k + L_k' D_k c_k) \\
-c_k
\end{bmatrix}
\]  \hspace{1cm} (3.17)
where \(\lambda_k\) is a \(r \times 1\) vector of Lagrange multipliers and \(D_k\) is an arbitrary nonnegative definite matrix. The sealing matrix \(D_k\) does not affect the solution and is often chosen to be the null matrix (Gill, Murray and Wright, 1981, Section 5.4). The next approximation \(\gamma_{k+1}\) for \(\hat{\gamma}\) is obtained from
\[
\gamma_{k+1} = \gamma_k + \alpha_n \delta_n,
\]
where \(\alpha_n\) is chosen initially as 1 and is halved successively until
\[
F_b + 2 \sum_{i=1}^r |[\lambda_k]_i[c_{k+1}]_i| < F_a + 2 \sum_{i=1}^r |[\lambda_k]_i[c_k]_i| \hspace{1cm} (3.18)
\]
where

\[ F_b = tr \left[ c (\bar{Y} - \Xi_{k+1})' \Sigma_{k+1}^{-1} (\bar{Y} - \Xi_{k+1}) \right] \]
\[ + \frac{1}{2} tr \left[ (\Sigma_{k+1}^{-1} (S^*_k - \Sigma_{k+1}))^2 \right] \]  
\[ F_a = tr \left[ c (\bar{Y} - \Xi_n)' \Sigma_k^{-1} (\bar{Y} - \Xi_k) \right] \]
\[ + \frac{1}{2} tr \left[ (\Sigma_k^{-1} (S^*_k - \Sigma_k))^2 \right] \]  
(3.19)

and \( \Xi_k = \Xi(\gamma_k) \), \( \Sigma_k = \Sigma(\gamma_k) \) and \( S^* = S^*(\gamma_k) \). If no constraints are imposed, all terms involving \( c_k \) and \( L_k \) are omitted.

It can happen that the matrix to be inverted in (3.17) is singular or near singular. An adaptation of the Jennrich and Sampson (1968) stepwise regression procedure may be used to obtain an appropriate conditional inverse. Their procedure for imposing bounds on the estimates may also be employed.

Du Toit and Browne (1982, 1989) developed the computer program AUFIT which implements the theoretical procedures described above. A modification of an algorithm described by Browne (1974, Section 4) is employed to handle nonlinear inequality constraints. AUFIT requires a subroutine to calculate the mean matrix and covariance matrix and optionally, a subroutine to calculate the constraints. Use is made of numerical derivatives, although the user may optionally provide code for the calculation of \( \frac{\partial \Xi}{\partial \gamma_j} \) and \( \frac{\partial \Sigma}{\partial \gamma_i} \). Practical experience has however shown that the convergence of the algorithm does not appear to be sensitive to whether exact derivatives are used or not. The approximation used for evaluating \( \frac{\partial \Sigma}{\partial \gamma_j} \) is

\[ \frac{\partial \Sigma}{\partial \gamma_j} \approx \frac{\Sigma(\gamma + \epsilon_j i_j) - \Sigma(\gamma)}{\epsilon_j}, \]  
(3.21)

where \( i_j \) is a column vector with zero elements except for the \( j \)-th element that is equal to unity and

\[ \epsilon_j = \epsilon \max \{1, |\gamma_j|\}, \]

where \( \epsilon \) is a small positive scalar (e.g. \( \epsilon = 10^{-5} \)).

Choice of \( \epsilon \) may affect the estimation of standard errors. The sensitivity to the choice of \( \epsilon \) will be greatest when the model is highly nonlinear in which case the asymptotic approximations for standard errors will tend to be inaccurate even if exact derivatives are used.
4. Time series covariance structures.

Consider the process (2.2) and let

\[ s = \max(n, m), \]  

(4.1)

then

\[ e_t + \sum_{j=1}^{s} \alpha_j e_{t-j} = u_t + \sum_{j=1}^{s} \beta_j u_{t-j}, \]

(4.2)

where

\[ \beta_{m+1} = \beta_{m+2} = \cdots = \beta_n = 0, \quad n > m \]  

(4.3)

or

\[ \alpha_{n+1} = \alpha_{n+2} = \cdots = \alpha_m = 0, \quad n < m \]  

(4.4)

Suppose \( T \) observations are made on \( \{e_t\} \) commencing at time \( t = 1 \) and that \( e_k \) denotes a typical element of the \( T \times 1 \) vector \( e \). A general closed form expression for the covariance matrix of \( e \) will be derived. The method of derivation is first illustrated for an ARMA(1,1) process and then generalized to an ARMA\((n, m)\) process.

For the ARMA(1,1) process, suppose \( T = 5 \), then from (4.2) and commencing at \( t = 1 \), it follows that

\[
\begin{align*}
    e_1 + \alpha_1 e_1 + e_2 &= u_1 + (\beta_1 u_0 - \alpha_1 e_0) \\
    \alpha_1 e_2 + e_3 &= \beta_1 u_1 + u_2 \\
    \alpha_1 e_3 + e_4 &= \beta_2 u_2 + u_3 \\
    \alpha_1 e_4 + e_5 &= \beta_3 u_3 + u_4 \\
    \alpha_1 e_5 &= \beta_4 u_4 + u_5
\end{align*}
\]

Let \( x_1 = (\beta_1 u_0 - \alpha_1 e_0) \), then the above equations may be expressed as

\[ T_\alpha e = T_\beta u + I_{5,1} x_1, \]

where \( I_{5,1} \) is formed from the first column of the identity matrix \( I \) of order 5,

\[ e = (e_1 \ e_2 \ e_3 \ e_4 \ e_5)' \]

\[ u = (u_1 \ u_2 \ u_3 \ u_4 \ u_5)', \]

\[ T_\alpha = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
\alpha_1 & 1 & 0 & 0 & 0 \\
0 & \alpha_1 & 1 & 0 & 0 \\
0 & 0 & \alpha_1 & 1 & 0 \\
0 & 0 & 0 & \alpha_1 & 1
\end{pmatrix} \quad \text{and} \quad T_\beta = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
\beta_1 & 1 & 0 & 0 & 0 \\
0 & \beta_1 & 1 & 0 & 0 \\
0 & 0 & \beta_1 & 1 & 0 \\
0 & 0 & 0 & \beta_1 & 1
\end{pmatrix} \]
In general, commencing at $t = 1$, the ARMA(1,1) process may therefore be expressed as

$$T_\alpha e = T_\beta u + I_{T,1}x_1,$$  \hspace{1cm} (4.5)

where

$$x_1 = (\beta_1 u_0 - \alpha_1 e_0).$$  \hspace{1cm} (4.6)

Since for $t = 1$

$$e_1 = u_1 + x_1$$

it follows that

$$e_0 = u_0 + x_0,$$  \hspace{1cm} (4.7)

with $x_0 = \beta_1 u_{-1} - \alpha e_{-1}$.

From (4.6) and (4.7) it follows that

$$x_1 = \beta_1 u_0 - \alpha_1 (u_0 + x_0)$$

$$= -\alpha_1 x_0 + (\beta_1 - \alpha_1)u_0,$$  \hspace{1cm} (4.8)

where $x_0$ (cf. (4.7)) and $u_0$ is independent. Let $P$ denote the variance of $x_1$, then for a stationary process and using (2.4) and (4.8)

$$P = \alpha_1^2 P + (\beta_1 - \alpha_1)^2 \psi^2,$$

so that

$$P = \frac{(\beta_1 - \alpha_1)^2}{(1 - \alpha_1^2)} \psi^2$$  \hspace{1cm} (4.9)

Since $T_\alpha$ is triangular, with unit diagonals, $|T_\alpha| = 1$. From (4.5) we have

$$e = T_\alpha^{-1} [I_{T,1} x_1 + T_\beta u]$$  \hspace{1cm} (4.10)

Let $\text{Cov}(e) = \Sigma$, then for an ARMA(1,1) process it follows from (2.4) and (4.10) that

$$\Sigma = T_\alpha^{-1} [I_{T,1} PP'I_{T,1} + \psi^2 T_\beta T_\beta'] T_\alpha^{-1}$$  \hspace{1cm} (4.11)

It will now be shown that the covariance matrix for an ARMA($n,m$)-process has the same general form (4.11).

Suppose $T$ observations are made on \{$e_t$\} generated by the process (4.2) commencing at time $t = 1$, then

$$T_\alpha e = T_\beta u + I_{T,1} x_1,$$  \hspace{1cm} (4.12)
where $I_{T,s}$ is the first $s$ columns of the identity matrix of order $T$ and $\mathbf{x}_1$ is a $s$-component vector with typical element

$$x_{\ell 1} = \sum_{j=\ell}^{s} \beta_j u_{\ell-j} - \sum_{j=\ell}^{s} \alpha_j e_{\ell-j}, \quad \ell = 1, 2, \ldots, s. \tag{4.13}$$

In (4.13) the elements of $\mathbf{x}_1$ are unobserved values of the variables from previous unobserved trials. The $T \times T$ lower triangular matrices $T_\alpha$ and $T_\beta$ with unit diagonal elements are defined as follows:

$$[T_\alpha]_{i,j} = \begin{cases} \delta_{ij}, & i \leq j \\ \alpha(i-j), & 1 \leq i - j \leq n \\ 0, & i - j > n \end{cases} \tag{4.14}$$

$$[T_\beta]_{i,j} = \begin{cases} \delta_{ij}, & i \leq j \\ \beta(i-j), & 1 \leq i - j \leq m \\ 0, & i - j > m \end{cases} \tag{4.15}$$

Note that $T_\alpha = I$ for $n = 0$ and similarly $T_\beta = I$ for $m = 0$.

Example 1. Consider an ARMA(2, 0) process with $T = 5$, then

$$T_\alpha = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \alpha_1 & 1 & 0 & 0 & 0 \\ \alpha_2 & \alpha_1 & 1 & 0 & 0 \\ 0 & \alpha_2 & \alpha_1 & 1 & 0 \\ 0 & 0 & \alpha_2 & \alpha_1 & 1 \end{pmatrix}$$

Let $\mathbf{x}_0$ be the $s$-component vector whose elements $x_{t0}$ contain information about the process prior to $t = 0$. Since for $t = 1$

$$e_1 = u_1 + x_{11},$$

it follows that

$$e_0 = u_0 + x_{10}. \tag{4.16}$$

But

$$x_{11} = \beta_1 u_0 - \alpha_1 e_0 + \sum_{j=2}^{s} \beta_j u_{1-j} - \sum_{j=2}^{s} \alpha_j e_{1-j}$$

and therefore from (4.13) and (4.16)

$$x_{11} = \beta_1 u_0 - \alpha_1 (u_0 + x_{10}) + x_{20}$$

$$= -\alpha_1 x_{10} + x_{20} + (\beta_1 - \alpha_1) u_0. \tag{4.17}$$
Similarly
\[ x_{21} = \beta_2 u_0 - \alpha_2 e_0 + x_{30} = -\alpha_2 x_{10} + x_{30} + (\beta_2 - \alpha_2)u_0 \]
\[ x_{31} = -\alpha_3 x_{10} + x_{40} + (\beta_3 - \alpha_3)u_0 \]
\[ \vdots \]
\[ x_{s1} = -\alpha_s x_{10} + 0 + (\beta_s - \alpha_2)u_0 \] (4.18)

The equations (4.18) may be expressed as
\[ x_1 = \Phi x_0 + bu_0 \] (4.19)

where
\[ \Phi : (s \times s) = \begin{pmatrix} -\alpha_1 & 1 & 0 & \ldots & 0 \\ -\alpha_2 & 0 & 1 & \ldots & 0 \\ \vdots \\ -\alpha_{s-1} & 0 & 0 & \ldots & 1 \\ -\alpha_s & 0 & 0 & \ldots & 0 \end{pmatrix} \] (4.20)

\[ b' : (1 \times s) = (\beta_1 - \alpha_1, \beta_2 - \alpha_2, \ldots, \beta_s - \alpha_s) \] (4.21)

Similarly it follows from (4.16) to (4.21) for \( t = k + 1 \) that
\[ x_{k+1} = \Phi x_k + bu_0 \]
\[ e_k = c'x_k + u_k \] (4.22)

where
\[ c' : (1 \times s) = (1 \ 0 \ \ldots \ 0) \]

The equations (4.22) are known as the state variable representation of an ARMA process with \( \Phi \) the so-called state transition matrix. A state variable representation of an ARMA-process may be derived using the z-transform (cf. Cadzow and Martens (1970), Kuo (1970), du Toit (1979)).

Suppose that the covariance matrix of \( x_k \) is the \((s \times s)\) matrix \( P \). For a stationary process it follows from (4.22) and (2.4) that
\[ P = \Phi P \Phi' + b(\psi^2 I)b' \] (4.23)

Let \( B^* : (s \times s) = \psi^2 bb' \), then
\[ (I - \Phi \otimes \Phi) \text{vec}(P) = \text{vec}(B^*) \] (4.24)
where \([\Phi \otimes \Phi]_{ij,k\ell} = [\Phi]_{j,\ell}[\Phi]_{i,k}\) and the double subscript \((i, j, k\ell)\) represents a typical element of the matrix \(\Phi \otimes \Phi\) with \(ij = (i-1)s + j; \ i, j = 1, 2, \ldots, s\).

Since \(P\) is a symmetric matrix, the \((s^2 \times 1)\) vector \(\text{vec}(P)\) may be expressed in terms of the \((\frac{1}{2}p(p+1) \times 1)\) vector \(\text{svec}(P)\) by using the following result given by Browne (1974, cf. equation (7)). Let \(K^-\) be a matrix of order \((\frac{1}{2}s(s+1) \times s^2)\) with typical element

\[
[K^-]_{k\ell,ij} = 2^{-1}(2 - \delta_{k\ell})(\delta_{ij}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}),
\]

\[i \leq p, j \leq p; k \leq \ell \leq s. \tag{4.25}\]

Then

\[
\text{vec}(P) = (K^-)' \text{svec}(P) \tag{4.26}
\]

Premultiplication of (4.24) by \(K^-\) and use of (4.26) gives

\[
K^-(I - \Phi \otimes \Phi)(K^-)' \text{svec}(P) = K^- \text{vec}(B^*) \tag{4.27}
\]

Let

\[
C : \left(\frac{1}{2}s(s+1) \times \frac{1}{2}s(s+1)\right) \equiv K^-(I - \Phi \otimes \Phi)(K^-)', \tag{4.28}
\]

then

\[
\text{svec}(P) = C^{-1}K^- \text{vec}(B^*), \tag{4.29}
\]

where the non-symmetric matrix \(C\) has typical element

\[
[C]_{ij,k\ell} = \frac{2}{(1 + \delta_{ij})(1 + \delta_{k\ell})} \left\{\delta_{ij}\delta_{j\ell} + \delta_{i\ell}\delta_{jk} - [\Phi]_{ik}[\Phi]_{j\ell} - [\Phi]_{i\ell}[\Phi]_{jk}\right\},
\]

\[1 \leq j \leq i \leq s; 1 \leq \ell \leq k \leq s, \tag{4.30}\]

and the \((\frac{1}{2}s(s \times 1) \times 1)\) vector \(K^- \text{vec}(B^*)\) has typical element

\[
[K^- \text{vec}(B^*)]_{i,j,1} = (2 - \delta_{ij})\psi^2 b_i b_j, \quad 1 \leq j \leq i \leq s. \tag{4.31}
\]

**Example 2.** Consider an ARMA(3,3) process then from (4.28) it follows that

\[
C = \begin{pmatrix}
1 - \alpha_1^2 & 2\alpha_1 & -1 & 0 & 0 & 0 \\
-\alpha_1\alpha_2 & 1 + \alpha_2 & 0 & \alpha_1 & -1 & 0 \\
-\alpha_1^2 & 0 & 1 & 2\alpha_2 & 0 & -1 \\
-\alpha_1\alpha_3 & \alpha_3 & 0 & 1 & 0 & 0 \\
-\alpha_2\alpha_3 & 0 & 0 & \alpha_3 & 1 & 0 \\
-\alpha_3^2 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} \tag{4.32}
\]
Note that the $C$ matrices for ARMA(1, 1) and ARMA(2, 2) processes are the submatrices of order 1 and of order 3 respectively indicated by dotted lines in (4.32).

It can easily be shown that for an ARMA(0, $m$) process, the covariance matrix of $x_k$ is given by

$$ P : (m \times m) = \psi^2 L'_\beta L_\beta $$

(4.33)

where $L_\beta$ is the lower triangular matrix

$$ L_\beta = \begin{cases} 
\beta_m, & i = j, \\
\beta_{m-(i-j)}, & 1 \leq i - j < m, \\
0, & i - j \geq m. 
\end{cases} $$

(4.34)

From (4.12) it follows that

$$ e : (T \times 1) = T^{-1}_\alpha (I_{T,s} x_1 + T_\beta u) $$

(4.35)

and since $x_1$ and $u$ are independent the covariance matrix of $e$ is given by

$$ \Sigma = T^{-1}_\alpha (I_{T,s} P I'_{T,s} + T_\beta \psi^2 T'_\beta) T^{-1}_\alpha $$

(4.36)

Let $\Gamma = I_{T,s}$ and

$$ P^* = \psi^{-2} P $$

(4.37)

then it follows from (4.36) that

$$ |\Sigma| = |I_{T,s} P I'_{T,s} + T_\beta \psi^2 T'_\beta| $$

$$ = \psi^{2T} |\Gamma'(T_\beta T'_\beta)^{-1} P^* + I_s| $$

(4.38)

Specifically, for an AR($n$) process

$$ |\Sigma| = \psi^{2T} |P^* + I| $$

(4.39)

**Example 3.** Consider an AR(1) process then (cf (4.9)) from (4.29) and (4.32)

$$ P = \frac{\psi^2 \alpha_1^2}{(1 - \alpha_1^2)} $$

$$ |\Sigma| = \frac{\psi^{2T}}{(1 - \alpha_1^2)} $$

$$ \psi^2 \Sigma^{-1} = L'L, $$

(4.40)
where

\[
L = \begin{pmatrix}
\sqrt{1 - \alpha_1^2} & 0 & 0 & \ldots & 0 \\
\alpha_1 & 1 & 0 & \ldots & 0 \\
0 & \alpha_1 & 1 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & 1
\end{pmatrix}
\] (4.41)

5. Relaxation of stationarity assumptions.

The covariance structure (4.36) is generated by an ARMA(n, m) process with homogeneous AR and MA parameters as well as homogeneous white noise variances and will be referred to as Model 1.

An alternative model may be obtained if it is assumed that the initial state vector is the null vector, i.e.

\[
x_1 = 0
\] (5.1)

This assumption implies that information about the behavior of the process, prior to \( t = 1 \), may be neglected. From (5.1) and (4.36) it follows that

\[
\Sigma = T_\alpha^{-1}(T_\beta \psi^2 T_\beta')T_\alpha'^{-1}
\] (5.2J)

The covariance structure (5.2) with a null initial state covariance matrix will be referred to as Model 2. The stochastic process that generates this structure is identical to the process considered by Anderson (1975, 1977). If assumption (2.4) is replaced by an assumption of non-homogeneous white noise variances, then

\[
\begin{align*}
\text{var}(u_t) &= \Psi_t^2, & t = 1, 2, \ldots, T \\
\text{cov}(u_t, u_s) &= 0, & t \neq s
\end{align*}
\] (5.3)

Therefore, using (4.35), (5.1) and (5.3)

\[
\Sigma = T_\alpha^{-1}(T_\beta D_\psi T_\beta')T_\alpha'^{-1},
\] (5.4)

where \( D_\psi \) is a \( T \times T \) diagonal matrix with typical element \( \psi^2_t \). In a number of practical applications it was found that the series of white noise variances

\[
\psi_{11}, \psi_{22}, \ldots, \psi_{TT}
\]

follow a smooth trend over time which may be adequately described by a nonlinear 2 or 3 parameter function such as, for example,

\[
\psi_t = \lambda_1 + \lambda_2 \psi_t^l, \quad 0 \leq \lambda_3 \leq 1
\] (5.5)
Models 1 and 2 may be interpreted in the following manner: In Model 1, the process has started in the distant past and continues indefinitely. Model 2 is defined in terms of a process which has started at the time of the first observation (i.e. \( t = 1 \)). It is our opinion that generally, neither of these models will adequately describe the results of a repeated measurement experiment.

An assumption which seems to be more realistic is to suppose that the process has "started" at some definite time in the past, e.g. at the birth or conception of an individual. However, the first observation took place on a more recent occasion. It is clear that in this case one cannot assume that the initial state vector \( x_1 \) is the null vector, since it accounts for the behavior of the process before it was observed.

For a stationary process it follows from (4.23) that the elements of \( \text{cov}(x_1) \) are functions of the AR and MA parameters and the white noise variance, \( \psi^2 \). An alternative way to account for the past behavior of the process, is to regard the non-duplicated elements of \( P \) as additional unknown parameters of the time series process. Suppose therefore that

\[
\text{cov}(x_1) = P_1,
\]

then (cf. (4.35) and (5.3))

\[
\Sigma = T_{t,t}^{-1}(I_{t,t} + P_1 T_{T,s} + T_{s} \Sigma T_{s}^T) T_{t,t}^{-1},
\]

where the covariance matrix \( D_{\psi} \) of the \( T \times 1 \) vector \( u \) may be structured in different ways, for example

\[
D_{\psi} = \psi^2 I
\]

or

\[
[D_{\psi}]_{t,t} = \lambda_1 \lambda_2 t, \quad \lambda_1, \lambda_2 \geq 0.
\]

For some ARMA processes and choices of structure for \( D_{\psi} \), parameter identification problems may arise. As an illustration, suppose the process is ARMA(1,0) and that \( [D_{\psi}]_{t,t} = \psi_{tt}, \ t = 1, 2, \ldots, T \), then it follows from (5.7) that

\[
\Sigma = T_{t,t}^{-1} \begin{pmatrix} P_1 + \psi_{11} & 0 & 0 & \cdots & 0 \\ 0 & \psi_{22} & 0 & \cdots & 0 \\ 0 & 0 & \psi_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \psi_{TT} \end{pmatrix} T_{t,t}^{-1},
\]

showing that \( P_1 \) and \( \psi_{11} \) cannot be estimated separately. The structure (5.7) will be referred to as Model 3.

It is further possible to relax the assumption of homogeneous AR and MA parameters by defining a general ARMA(\( n, m \)) process

\[
e_k + \sum_{j=1}^{n} a_{kj} e_{k-j} = u_k + \sum_{\ell=1}^{m} \beta_{k\ell} u_{k-\ell}, \quad k = \ldots, -1, 0, 1, \ldots
\]

(5.8)
where the sequence \( \{u_k\} \) consists of independently distributed \( N(0, \psi_{kk}) \) variates. Suppose the process (5.8) is ARMA(1, 1) and that \( T = 4 \), then

\[
\begin{align*}
e_1 + e_2 & = \beta_{21} u_1 + u_2 \\
\alpha_{31} e_2 + e_3 & = \beta_{31} u_2 + u_3 \\
\alpha_{41} + e_4 & = \beta_{41} u_3 + u_4
\end{align*}
\]

These equations may, in general, be written in the form

\[
T_{\alpha}^* e = T_{\beta}^* u + I_T x_1,
\]

where \( x_1 = (\beta_{11} u_0 - \alpha_{11} e_0) \) and

\[
T_{\alpha}^* = \begin{pmatrix}
1 & 0 & 0 & 0 \\
\alpha_{21} & 1 & 0 & 0 \\
0 & \alpha_{31} & 1 & 0 \\
0 & 0 & \alpha_{41} & 1
\end{pmatrix}, \quad T_{\beta}^* = \begin{pmatrix}
1 & 0 & 0 & 0 \\
\alpha_{21} & 1 & 0 & 0 \\
0 & \alpha_{31} & 1 & 0 \\
0 & 0 & \alpha_{41} & 1
\end{pmatrix}
\]

Neglecting any information of the behavior of the process before \( t = 1 \), it follows that, for the general ARMA(\( n, m \)) process

\[
e = T_{\alpha}^{*-1} T_{\beta}^* u,
\]

where

\[
[T_{\alpha}^*]_{i,j} = \begin{cases}
\delta_{i,j}, & i \leq j \\
\alpha_{(i-j)}, & 1 \leq i - j \leq n \\
0, & i - j > n
\end{cases}
\]

\[
[T_{\beta}^*]_{i,j} = \begin{cases}
\delta_{i,j}, & i \leq j \\
\beta_{(i-j)}, & 1 \leq i - j \leq n \\
0, & i - j > n
\end{cases}
\]

Since (cf. (5.8)) \( u \sim N(0, D_{\psi}) \), it follows from (5.10) that

\[
\Sigma = \text{cov}(e) = T_{\alpha}^{*-1} T_{\beta}^* D_{\psi} T_{\beta}^* (T_{\alpha}^*)^{-1}
\]

The covariance structure (5.13) for the non-stationary ARMA(\( n, m \)) process will be referred to as Model 4.

Consider a non-stationary ARMA(1, 1) with \( T = 5 \). Then

\[
T_{\alpha}^* = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
\alpha_{21} & 1 & 0 & 0 & 0 \\
0 & \alpha_{31} & 1 & 0 & 0 \\
0 & 0 & \alpha_{41} & 1 & 0 \\
0 & 0 & 0 & \alpha_{51} & 1
\end{pmatrix}
\]

16
\[ T^{-1}_\alpha = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -\alpha_{21} & 1 & 0 & 0 & 0 \\ \alpha_{21}\alpha_{31} & -\alpha_{31} & 1 & 0 & 0 \\ -\alpha_{21}\alpha_{31}\alpha_{41} & \alpha_{31}\alpha_{41} & -\alpha_{41} & 1 & 0 \\ \alpha_{21}\alpha_{31}\alpha_{41}\alpha_{51} & -\alpha_{31}\alpha_{41}\alpha_{51} & \alpha_{41}\alpha_{51} & -\alpha_{51} & 1 \end{pmatrix}, \quad (5.15) \]

and

\[ T^*_\beta = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \beta_{21} & 1 & 0 & 0 & 0 \\ 0 & \beta_{31} & 1 & 0 & 0 \\ 0 & 0 & \beta_{41} & 1 & 0 \\ 0 & 0 & 0 & \beta_{51} & 1 \end{pmatrix}. \quad (5.16) \]

Let

\[ \eta_i = \beta_{i,1} - \alpha_{i,1}; \quad i = 2, 3, 4, 5, \]

then

\[ T^{-1}_\alpha T^*_\beta u = \begin{pmatrix} u_1 \\ \eta_2 u_1 + u_2 \\ -\alpha_{31}\alpha_{21} u_1 + \eta_2 u_2 + u_3 \\ \alpha_{31}\alpha_{41}\eta_2 u_1 - \alpha_{41}\eta_3 u_2 + \eta_4 u_3 + u_4 \\ -\alpha_{31}\alpha_{41}\alpha_{51}\eta_2 u_1 + \alpha_{41}\alpha_{51}\eta_3 u_2 - \alpha_{51}\eta_4 u_3 + \eta_5 u_4 + u_5 \end{pmatrix}. \quad (5.18) \]

The covariance matrix of \( e : (5 \times 1) \) is given by (cf. (5.13) and (5.18))

\[ \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{22} & \sigma_{33} & \sigma_{44} & \sigma_{55} \\ \eta_2 \sigma_{11} & -\alpha_{31}\sigma_{22} + \beta_{31}\psi_{22} & -\alpha_{41}\sigma_{33} + \beta_{41}\psi_{33} & -\alpha_{51}\sigma_{44} + \beta_{51}\psi_{44} \\ -\alpha_{21}\sigma_{22} & \sigma_{11} & \sigma_{33} & \sigma_{44} & \sigma_{55} \\ -\alpha_{41}\sigma_{33} & \sigma_{22} & \sigma_{11} & \sigma_{44} & \sigma_{55} \\ -\alpha_{51}\sigma_{44} & \sigma_{33} & \sigma_{22} & \sigma_{11} & \sigma_{55} \end{pmatrix}. \quad (5.19) \]

where

\[ \sigma_{11} = \psi_{11} \]

\[ \sigma_{22} = \eta_2^2 \sigma_{11} + \psi_{22} \]

\[ \sigma_{33} = \alpha_{31}^2 \sigma_{22} + \beta_{31}\psi_{22}(\beta_{31} + 2\alpha_{31}) + \psi_{33} \]

\[ \sigma_{44} = \alpha_{41}^2 \sigma_{33} + \beta_{41}\psi_{33}(\beta_{41} + 2\alpha_{41}) + \psi_{44} \]

\[ \sigma_{55} = \alpha_{51}^2 \sigma_{44} + \beta_{51}\psi_{44}(\beta_{51} + 2\alpha_{51}) + \psi_{55} \]

Since \( \eta_2 \) (cf. (5.17)) appears only in the expressions for \( \sigma_{12} \) and \( \sigma_{22} \), the parameters \( \alpha_{21} \) and \( \beta_{21} \) are not separately identified. To eliminate this indeterminacy, let

\[ \beta_{21} = 0 \quad (5.20) \]

Under condition (5.20) a one-to-one correspondence exists between the remaining \( 3T - 3 \) ARMA parameters and the elements of the covariance matrix \( \Sigma \).
It can be shown that in general,

\[
\psi_{11} = \sigma_{11}, \\
\alpha_{i+1,1} = \frac{-\sigma_{1,i+1}}{\sigma_{1,i}}, \quad i = 1, 2, \ldots, T - 1, \\
\psi_{ii} = \sigma_{ii} - \alpha_{i,1}^2 \sigma_{i-1,i-1} + (\beta_{i,1}^2 + 2\alpha_{i,1}\beta_{i,1})\psi_{i-1,i-1}, \quad i = 2, \ldots, T, \\
\beta_{i+1,1} = \frac{\alpha_{i+1,1}\sigma_{ii} - \sigma_{i,i+1}}{\psi_{ii}}, \quad i = 2, 3, \ldots, T - 1,
\]

(5.21)

where \(\beta_{21} = 0\).

Initial estimates are obtained for the unknown parameters by substituting the elements of \(\Sigma\) by the corresponding elements of the sample covariance matrix \(S\) in the recursion formulae (5.21).

The estimates can be employed to obtain initial estimates for the unknown parameters in the stationary ARMA(1, 1) model, as follows

\[
\hat{\psi}_1 = \frac{1}{p} \sum_{i=1}^{p} \psi_{ii}, \quad (5.22)
\]

\[
\hat{\alpha}_1 = \frac{1}{p - 1} \sum_{i=1}^{p-1} \hat{\alpha}_{i+1,1}, \quad (5.23)
\]

and

\[
\hat{\beta}_1 = \frac{1}{p - 2} \sum_{i=2}^{p-2} \hat{\beta}_{i+1,1}. \quad (5.24)
\]

Du Toit (1979) showed that specific LISREL and Simplex models are mathematically equivalent to specific ARMA(\(n, m\)) models. This was accomplished by comparing the state-space equations for stationary and non-stationary ARMA models with the structural equations of the LISREL models. For example, the non-stationary Markov Simplex is equivalent to the non-stationary ARMA(1,0) model and the Quasi Markov Simplex is equivalent to the non-stationary ARMA(1, 1) process.

In the next section, the results derived will be applied to analyze three data sets.

6. Some applications.

We shall consider some examples of models with nonstandard features that illustrate the capabilities of the methods described in preceding sections. All results to be presented were computed using AUFIT without providing subroutines to evaluate partial derivatives. In all cases normal theory maximum likelihood results are reported.
The first example illustrates a situation where both means and covariances have "non-standard" (cf. Section 2) structures.

Example 1. A repeated time series model with nonlinear trends for means and white noise variances.

In a longitudinal study carried out by Dr. J.J. de Beer of the Institute for Communications Research, Human Sciences Research Council, S.A., to investigate the effect of television, five repeated measurements were made on the Intelligence Factor of the High School Personality Questionnaire (Cattell, 1973) on school children. The first measurement was obtained during the first year of high school and measurements were repeated at yearly intervals over the next four years. Subjects were divided into three groups according to their television viewing habits: light viewers (L), moderate viewers (M) and heavy viewers (H). Sample sizes were $N_L = 196$, $N_M = 675$, and $N_H = 254$ respectively, with $N = N_L + N_M + N_H = 1125$. The three mean vectors are shown in Table 1 and the within groups covariance matrix in Table 2.

<table>
<thead>
<tr>
<th>Year</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light viewers</td>
<td>7.388</td>
<td>8.031</td>
<td>8.082</td>
<td>8.398</td>
<td>8.408</td>
</tr>
<tr>
<td>Moderate viewers</td>
<td>7.305</td>
<td>7.850</td>
<td>7.970</td>
<td>8.138</td>
<td>8.237</td>
</tr>
<tr>
<td>Heavy viewers</td>
<td>7.028</td>
<td>7.587</td>
<td>7.768</td>
<td>8.004</td>
<td>8.020</td>
</tr>
</tbody>
</table>

A typical element $\xi_{ij}$ of the $5 \times 3$ mean matrix $\Xi$ represents the population mean for the $i$-th repeated measurement, $i = 1, \ldots, 5$, in the $j$-th group, $j = 1, 2, 3$. It can be seen from Table 1 that the sample intelligence score means give the impression of increasing with age towards an asymptote, as may be expected. Consequently the model incorporates the following logistic trend for the $j$-th column of $\Xi$, $j = 1, 2, 3$:

$$
\xi_{ij} = \frac{\theta_1}{(1 + \theta_2 \theta_3^i)} , \quad i = 1, \ldots, 5 ,
$$

where $\theta_{1j} (\theta_{1j} \geq 0)$ represents the asymptote, $\theta_{2j} (\theta_{2j} \geq 0)$ governs the amount of change and $\theta_{3j} (0 \leq \theta_{3j} \leq 1)$ the rate of change.
Table 2.
Television viewers: Within groups covariance matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.683</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.973</td>
<td>2.238</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.894</td>
<td>0.996</td>
<td>2.133</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.925</td>
<td>0.928</td>
<td>0.983</td>
<td>2.086</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.795</td>
<td>0.856</td>
<td>0.933</td>
<td>1.015</td>
<td>2.134</td>
</tr>
</tbody>
</table>

Deviations from the group mean curve for an individual over the five measurements are postulated to follow an ARMA(1, 1) time series commencing with the first measurement (cf. Anderson, 1975, p. 1284.) We assume that this process has homogeneous autoregressive weights and homogeneous moving average weights but non-homogeneous white noise variances. This implies that the covariance matrix has the structure (cf. (5.4))

$$\Sigma = T^{-1}_{\alpha} T_{\beta} D_\psi T_{\beta} T^{-1}_{\alpha} \tag{6.2}$$

Because fluctuations in consecutive intelligence scores for an individual tend to become smaller as adulthood is approached, it is realistic to require that disturbances induced by white noise should decrease with time. White noise variances are therefore assumed to have the decreasing exponential trend

$$\psi_{ii} = \lambda_1 + \lambda_2 \lambda_3^i, \quad i = 1, \ldots, 5. \tag{6.3}$$

where \(\lambda_1 (\lambda_2 \geq 0)\) is the asymptote, \(\lambda_2 (\lambda_2 \geq 0)\) governs the amount of decrease and \(\lambda_3 (0 \leq \lambda_3 \leq 1)\) the rate of change.

The model given by (6.2) with (6.3) is a modification of a model suggested by Anderson (1975, equation (5.2)) in which white noise variances were homogeneous so that \(D_\psi\) was a diagonal matrix with equal diagonal elements. Although (6.2) fits into the standard LISREL format, our model is nonstandard because of the nonlinear component equations (6.1) and (6.3).

Maximum likelihood estimates and associated standard errors are shown in Table 3.
Table 3.
Television viewers: Maximum likelihood estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Light viewers</th>
<th>Moderate viewers</th>
<th>Heavy viewers</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_{1} )</td>
<td>8.46</td>
<td>8.26</td>
<td>8.09</td>
</tr>
<tr>
<td>( \theta_{2} )</td>
<td>0.30</td>
<td>0.27</td>
<td>0.31</td>
</tr>
<tr>
<td>( \theta_{3} )</td>
<td>0.46</td>
<td>0.47</td>
<td>0.47</td>
</tr>
<tr>
<td>St. error</td>
<td>0.14</td>
<td>0.07</td>
<td>0.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \lambda_{1} )</th>
<th>( \lambda_{2} )</th>
<th>( \lambda_{3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-0.96</td>
<td>-0.63</td>
<td>1.42</td>
<td>3.42</td>
<td>0.37</td>
</tr>
<tr>
<td>Std. error</td>
<td>0.02</td>
<td>0.03</td>
<td>0.06</td>
<td>0.80</td>
<td>0.08</td>
</tr>
</tbody>
</table>

The minimum of the discrepancy function \( F \) is 0.0181 yielding a value of \( 1125 \times 0.0181 = 20.33 \) for the likelihood ratio test statistic. The model therefore appears to fit the data quite well and yields an exceedence probability of 0.21 for the associated asymptotic chi-squared distribution with 16 degrees of freedom. Equation (6.3) shows that \( \lambda_{2} = 0 \) implies that the white noise variances are homogeneous. The ratio of the estimate of \( \lambda_{2} \) to its standard error is 4.28 and this yields an exceedence probability of less than 0.001 for its asymptotic standard normal distribution. Consequently the assumption of non-homogeneous white noise variances seems plausible.

There are noticeable and systematic differences in the asymptote parameter estimates \( \hat{\theta}_{ij} \) between the three groups. A likelihood ratio test of the null hypothesis that the three groups have the same logistic growth curve for means against the \textit{restricted alternative} that the model holds with different logistic function parameter values for different groups gave a test statistic value of \( 1125 \times (0.0320 - 0.0181) = 15.64 \). This yields an exceedence probability of 0.02 for the associated asymptotic chi-squared distribution with 6 degrees of freedom.

\textit{Example 2. Non-stationary ARMA(1,1) model with Gompertz trend for the mean applied to the weights of mice.}

A data set, comprising the weights of 42 male mice was obtained from the Department of Zoology, Pretoria University, South Africa.

A brief background to the origin of the data is the following. In February 1970, 7 female and 6 male striped mice were released in a 36.34m\(^2\) outdoor enclosure with 8 nest boxes, food and water in excess and were allowed to multiply freely. Occurance of birth was recorded daily and the mice were weighed weekly. After two years the experiment
was concluded. Physical maturity was reached approximately 9 weeks after birth. The mean vector and sample covariance matrix are given in Table 4 and are based on weekly measurements for a period of 9 weeks.

Table 4.
Average weights (in grams) and sample covariance matrix for 42 male mice

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means:</td>
<td>11.33</td>
<td>17.00</td>
<td>23.00</td>
<td>28.52</td>
<td>33.33</td>
<td>37.90</td>
<td>40.48</td>
<td>42.40</td>
<td>44.43</td>
</tr>
<tr>
<td>Covariances:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7.57</td>
<td>16.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5.98</td>
<td>13.40</td>
<td>16.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4.94</td>
<td>13.19</td>
<td>16.88</td>
<td>25.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.39</td>
<td>16.88</td>
<td>21.12</td>
<td>30.83</td>
<td>44.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.87</td>
<td>11.00</td>
<td>16.48</td>
<td>26.69</td>
<td>40.72</td>
<td>46.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.75</td>
<td>8.07</td>
<td>12.57</td>
<td>21.16</td>
<td>36.06</td>
<td>41.45</td>
<td>47.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-0.56</td>
<td>6.33</td>
<td>9.76</td>
<td>18.67</td>
<td>35.08</td>
<td>42.80</td>
<td>51.19</td>
<td>62.72</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-0.76</td>
<td>4.57</td>
<td>8.38</td>
<td>15.87</td>
<td>31.90</td>
<td>40.59</td>
<td>50.22</td>
<td>62.45</td>
<td>66.96</td>
</tr>
</tbody>
</table>

A typical element $\xi_{i1}$ of the $9 \times 1$ mean vector $\Xi$ represents the population mean for the $i$-th repeated measurement, $i = 1, 2, \ldots, 9$. The model incorporates a Gompertz trend for $\Xi$,

$$
\xi_{i1} = \theta_1 \exp(-\theta_2 \theta_3^i), \quad i = 1, 2, \ldots, 9
$$

where $\theta_1$ represents the asymptote, $\theta_2$ ($\theta_1 \geq 0$) governs the amount of change and $\theta_3$ ($0 \leq \theta_3 \leq 1$) the rate of change. Deviations from the group mean curve for an individual are postulated to follow a nonstationary ARMA(1, 1) time series commencing with the first measurement (cf. Section 5, (5.13)). Initial estimates for the unknown ARMA parameters were obtained using (5.21) with $\beta_{21} = 0$. Initial estimates for the growth curve parameters may be obtained using a nonlinear regression analysis computer program. Results of the analysis are summarized in Table 5.
Table 5.
Summary of analysis: Non-stationary ARMA(1, 1) deviations and Gompertz trend, Male Mice data (n = 42)

<table>
<thead>
<tr>
<th>Reproduced</th>
<th>Means:</th>
<th>11.41</th>
<th>17.28</th>
<th>28.88</th>
<th>33.73</th>
<th>37.71</th>
<th>40.87</th>
<th>43.31</th>
<th>45.16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reproduced Covariance Matrix:</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>7.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7.59</td>
<td>17.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6.13</td>
<td>13.48</td>
<td>16.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6.14</td>
<td>13.49</td>
<td>17.01</td>
<td>25.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7.82</td>
<td>17.18</td>
<td>21.67</td>
<td>30.90</td>
<td>44.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6.53</td>
<td>14.36</td>
<td>18.11</td>
<td>25.81</td>
<td>40.40</td>
<td>46.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6.10</td>
<td>13.41</td>
<td>16.92</td>
<td>24.12</td>
<td>37.75</td>
<td>41.32</td>
<td>47.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6.29</td>
<td>13.81</td>
<td>17.43</td>
<td>24.84</td>
<td>38.88</td>
<td>42.55</td>
<td>51.62</td>
<td>63.59</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>6.21</td>
<td>13.63</td>
<td>17.20</td>
<td>24.51</td>
<td>38.37</td>
<td>42.00</td>
<td>50.94</td>
<td>63.16</td>
<td>67.54</td>
</tr>
</tbody>
</table>

Estimates of the unknown parameters:

\[
\hat{\psi}_{ii}, \ i = 1, 2, \ldots, 9:
\]

<table>
<thead>
<tr>
<th></th>
<th>7.04</th>
<th>8.84</th>
<th>5.41</th>
<th>7.04</th>
<th>6.46</th>
<th>8.45</th>
<th>10.11</th>
<th>6.60</th>
<th>4.79</th>
</tr>
</thead>
</table>

\[
\hat{\alpha}_{j1}, \ j = 2, 3, \ldots, 9:
\]

<table>
<thead>
<tr>
<th>-1.08</th>
<th>-0.81</th>
<th>-1.00</th>
<th>-1.27</th>
<th>-0.84</th>
<th>0.93</th>
<th>-1.03</th>
<th>-0.99</th>
</tr>
</thead>
</table>

\[
\hat{\beta}_{j1}, \ j = 2, 3, \ldots, 9:
\]

<table>
<thead>
<tr>
<th>0.00</th>
<th>-0.03</th>
<th>0.17</th>
<th>-0.16</th>
<th>0.49</th>
<th>0.24</th>
<th>0.29</th>
<th>0.06</th>
</tr>
</thead>
</table>

\[
\hat{\theta}, \ j = 1, 2, 3:
\]

<table>
<thead>
<tr>
<th>3.9176</th>
<th>-6.185</th>
<th>-0.9459</th>
</tr>
</thead>
</table>

The \( \chi^2 \)-goodness of fit statistic is equal to 35.73. Since the critical value of \( \chi^2 \) with 27 degrees of freedom is 40.11 for \( \alpha = 0.05 \), we conclude that the fit of the model is acceptable.
Example 3. Bilodeau’s data: Proficiency measures in six trials \((N = 152)\).

This data was previously analyzed by Bock and Bargman (1966), Jöreskog (1970) and Browne (1974). The sample covariance matrix for the six measurements is shown in Table 6.

\[
\begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
1 & 521.00 & & & & \\
2 & 477.00 & 576.00 & & & \\
3 & 484.00 & 536.00 & 601.00 & & \\
4 & 510.00 & 575.00 & 593.00 & 755.00 & \\
5 & 523.00 & 580.00 & 598.00 & 718.00 & 797.00 & \\
6 & 528.00 & 584.00 & 613.00 & 722.00 & 751.00 & 802.00 \\
\end{array}
\]

Assuming homogeneous AR and MA parameters and non-homogeneous white noise variances \(\psi_i, i = 1, 2, \ldots, 6\), and an arbitrary initial state variance \(P_1\), Model 3 (cf. (5.7)) was fitted to the data. Table 7 is a summary of the results.

\[
\begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
1 & 522.05 & & & & \\
2 & 482.55 & 602.05 & & & \\
3 & 493.49 & 556.61 & 623.30 & & \\
4 & 504.68 & 569.22 & 599.49 & 744.81 & \\
5 & 516.12 & 582.13 & 613.08 & 690.17 & 770.64 & \\
6 & 527.82 & 595.32 & 626.98 & 711.96 & 746.58 & 807.23 \\
\end{array}
\]
Estimates of the unknown parameters:

Initial State variance $\hat{P}_1$: 401.05

White noise variances $\hat{\psi}_{ii}$: 121.00 139.28 89.44 154.44 97.89 68.58

AR parameter $\hat{\alpha}_1$: -1.02

MA parameter $\hat{\beta}_1$: -0.42

To obtain initial estimates for the parameters, one may use initial estimates for the non-stationary ARMA(1,1) model (cf. (5.21)) and calculate the averages of the estimated non-homogeneous AR and MA parameters (cf. (5.23) and (5.24)). An initial estimate for the variance of $x_1$ is obtained from

$$\hat{P}_1 = s_{11} = \hat{\psi}^2,$$

where $\hat{\psi}^2$ is defined by (5.22) and $s_{11}$ denotes the sample variance for the first proficiency measures. The $\chi^2$-goodness of fit statistic was found to be 11.63 with 12 degrees of freedom indicating that Model 3 provides a good fit to the data.
References


