#1680

November 25, 1985

Split-block Models with Time Series Components for Repeated Measurements

by

Sastry G. Pantula and Kenneth H. Pollock

Department of Statistics, North Carolina State University
Raleigh, North Carolina 27695-8203

SUMMARY

In this paper we consider the problem where repeated measurements are taken on each experimental unit of a randomized experimental design. If the design is a randomized complete block then traditionally a split-block analysis is used. Here we consider an extension of the traditional split-block analysis where we allow for several orthogonal autocorrelated error components depending on the definitions of fixed and random effects. Our approach is based on autoregressive processes applied to separate orthogonal components of the data and it can also be used for more complex designs such as Latin square.

Key Words: autoregressive processes; fixed and random effects; split-plot designs

1. INTRODUCTION

Frequently data arise from an experimental design where several successive measurements over time are made on each experimental unit. Such repeated measurements arise in plant and animal experiments, clinical trials, radiotelemetry experiments on wild animals and economic studies. In all cases, the observations on the same unit are correlated. Multivariate and univariate methods are commonly used to analyse the repeated measurements data.

Timm (1980) provides an excellent survey of the literature on the multivariate analysis of repeated measurements. We believe that the multivariate methods are usually not easy to compute or to interpret. Also, they are overparameterized and require an equal number of measurements on each individual.

One of the univariate analyses that is at times incorrectly used to analyse repeated measurements data is the split-plot method regarding time as the subplot treatment. (See, for example, Thomas & Wilkinson (1975)). The difference between the repeated measurements design and the split-plot design is that the subplot classification (time) is not randomized. It is likely that the observations on the same unit which are close together will be more highly correlated than those that are far apart in time. This is contrary to the assumption of constant correlation between subplots within a unit made by the traditional split-plot approach. In this respect, we feel that the split-plot analysis for repeated measurements may be inappropriate. Also, as stated earlier, we feel that the multivariate model, which allows for arbitrary correlations for observations on the same unit, is overparameterized. A time series modelling approach to the problem, which allows for decaying autocorrelations, seems to be an attractive compromise.

Box (1954) considered the effect of autocorrelation between the errors in the two-way classification and proposed an approximate F-test with degrees of freedom corrected by a number ϵ, to test the differences of the variable over time. Huynh (1978), Wallenstein & Fleiss (1979) and Anderson, Jensen & Schou (1981), among others, computed the ϵ-correction for different covariance structures, including the first order autoregressive structure over time. Danford, Hughes & McNee (1960) considered an arbitrary stationary covariance structure for the observations over time. Rowell & Walters (1976) in an excellent summary paper on split-plot analysis, recognize the need for a time series error covariance structure. For combining cross section and time series data (also known as panel data), Balestra & Nerlove (1966) and Wallace & Hussian (1969) considered a mixed-model with some lagged values of the dependent variable included in the collection of independent variables. Shumway (1970) and Bloomfield et al (1983) considered spectral density estimation for mixed-models with stationary time series components. As in traditional time series analysis, however, they require the number of observations over time on each individual to be large.

For the case where a moderate number of repeated measurements are taken on several units, Azzalini (1984) and Pantula & Pollock (1985) considered the following mixed-model:

$$Y_{ik} = \sum_{s=1}^{r} X_{iks}\beta_s + v_i + \eta_{ik} \qquad (1.1)$$

$$k = 1, \ldots, t_i; \; i = 1, \ldots, n,$$

where Y_{ik} denotes the value of the kth measurement on the ith unit; X_{iks}, $s = 1, \ldots, r$ denote the level of r control variables at which the observation Y_{ik} is obtained (the X_{iks} are assumed to be fixed); β_s, $s = 1, \ldots, r$ denote the unknown parameters to be estimated; v_i is the random effect associated with the ith sampled unit; η_{ik} is the random error associated with the kth measurement on the ith individual. They assumed that v_i are independent $N(0,\sigma_v^2)$ variables and that

$$\eta_{ik} = \begin{cases} (1-\alpha_1^2)^{-1/2} e_{i1}, & k = 1 \\ \alpha_1 \eta_{i,k-1} + e_{ik}, & k \geq 2 \end{cases} , \qquad (1.2)$$

where $|\alpha_1|<1$ and e_{ik} are independent $N(0,\sigma_e^2)$ variables that are independent of $\{v_i\}$. That is, $\{\eta_{ik}\}$ is a set of n independent first order autoregressive processes, AR(1), (Fuller (1976), p. 36). Note that if $\alpha_1 = 0$, then (1.1) reduces to the traditional split-plot model. Pantula & Pollock (1985) presented the following extension of the estimation procedure given in Fuller & Battese (1973): (a) obtain a consistent estimator for α_1, (b) obtain a linear transformation Z_{ik} of $Y_{i1},\ldots,Y_{it_i}$ involving α_1, so that the model for Z_{ik} is a sum of fixed effects and two independent random components, (c) obtain consistent estimates for the variance components σ_v^2 and σ_e^2, (d) obtain a linear transformation Z_{ik}^* involving α_1, σ_v^2 and σ_e^2 of $Z_{i1}, \ldots, Z_{it_i}$ so that Z_{ik}^* are uncorrelated and (e) regress Z_{ik}^* on the corresponding transformation X_{ik}^* of $X_{ik} = (X_{ik1},\ldots,X_{ikr})$ to obtain the (estimated) generalized least squares estimator of $\beta = (\beta_1,\ldots,\beta_r)'$. The estimation procedure can be extended to higher order autoregressive processes. Also, conceptually, the method can be extended to the case where η_{ik} is any general stationary process over time. However, it is not easy to obtain a simple algebraic expression for the transformation Z_{ik} in (b).

In this paper, we extend the results of Pantula & Pollock (1985) to more complex mixed-models with several random components that vary over time. We exploit the usual analysis of variance orthogonal components to simplify our time series approach. We model different orthogonal components over time as different sets of independent time series. Depending on the number of time series included in the model, we pool some of the orthogonal components. In sections 2 and 3 we develop estimation procedures for two "mixed-time series" models which can be viewed as extensions of the split-block analysis (Steel & Torrie (1980) p. 393). Our estimation procedure consists of two main septs: (i) estimate the variance components and the autocorrelation coefficients by analysing the orthogonal components separately, (ii) estimate the fixed effects by combining the different orthogonal components. In section 4 we present a discussion of our results and some possible extensions.

2. NESTED SPLIT-BLOCK MODELS

2.1 *Model presentation*

In this section we consider the situation where we have repeated measurements in a randomized block design. We consider a perennial crop experiment as an example. Suppose we wish to test the effect of a fixed set of treatments (fertilizer or variety, for example) on the yield. (The case where the set of treatments is selected at random from a collection of treatments will be considered in section 3). A set of a locations (or blocks) is selected at random. At each of the locations b plots are used. The treatments are assigned randomly to different plots. We consider two cases: (I) plots are random and (II) plots are fixed. Measurements are taken on each plot on t consecutive occasions. (Here, we are considering only the balanced case to keep the notation simple. The methods, however, can be extended to unequal numbers of repeated measurements as in Pantula & Pollock (1985)). We consider a linear model

$$Y_{ijk} = X_{ijk}\beta + \rho_i + \gamma^*_{ij} + \theta_{ik} + \xi_{ijk}, \tag{2.1}$$

$$\gamma^*_{ij} = \begin{cases} \gamma_{ij} & \text{case (I): plots random} \\ \gamma_{ij} - \bar{\gamma}_{i.} & \text{case (II): plots fixed} \end{cases}$$

where $i = 1, \ldots, a$ denotes the blocks; $j = 1, \ldots, b$ denotes the plots and $k = 1, 2, \ldots, t$ denotes the time points. In (2.1), $X_{ijk}\beta$ corresponds to the fixed effects (which includes the treatment and time effects), ρ_i corresponds to the random effect for the ith block, γ^*_{ij} corresponds to the random effect for the jth plot of the ith block, θ_{ik} corresponds to the block effect over time and ξ_{ijk} corresponds to the measurement error on the jth plot of the ith block at time k. When the plots are fixed, the average $\bar{\gamma}_{i.}$ is subtracted so that $\Sigma\gamma^*_{ij} = 0$ for all i where the sum is over j. We assume that the random components $\{\rho_i\}$, $\{\gamma_{ij}\}$, $\{\theta_{ik}\}$ and $\{\xi_{ijk}\}$ are independent. We also assume that ρ_i are independent $N(0,\sigma^2_\rho)$ and γ_{ij} are independent $N(0,\sigma^2_\gamma)$ variables. The random components θ_{ik} and ξ_{ijk} involve repeated measurements on the same location and the same plot and hence may be correlated over time. We assume that θ_{ik} and ξ_{ijk} are independent stationary time series over time. Consider the following orthogonal components over time:

1. Blocks: $\bar{Y}_{i.k} = \bar{X}_{i.k}\,\beta + v_i + \eta_{ik}$, (2.2)

and

2. Deviations from the blocks: $Y_{ijk} - \bar{Y}_{i.k} = (X_{ijk} - \bar{X}_{i.k})\beta + u_{ijk}$, (2.3)

where

$$\eta_{ik} = \theta_{ik} + \bar{\xi}_{i.k} ,$$

$$u_{ijk} = (\gamma_{ij} - \bar{\gamma}_{i.}) + (\xi_{ijk} - \bar{\xi}_{i.k}) ,$$

and

$$v_i = \begin{cases} \rho_i + \bar{\gamma}_{i.} & \text{, case (I): plots random} \\ \rho_i & \text{, case (II): plots fixed} \end{cases} .$$

Notice that $\{\bar{Y}_{i.k}\}$ and $\{Y_{ijk} - \bar{Y}_{i.k}\}$ are independent and we are modelling our data as two sets of independent time series.

For t moderate in size, we follow the approach of Azzalini (1984) and Pantula & Pollock (1985) to analyse the block means over time. We assume that η_{ik} is a first order autoregressive process, AR(1), given in (1.2). Similarly, we assume that, for fixed i and j, ξ_{ijk} is an AR(1) process given by,

$$\xi_{ijk} = \begin{cases} (1-\alpha_2^2)^{-1/2} \epsilon_{ijk} & , k = 1 \\ \alpha_2 \xi_{ij,k-1} + \epsilon_{ijk} & , k \geq 2 \end{cases} \tag{2.4}$$

where $|\alpha_2|<1$ and ϵ_{ijk} are independent $N(0,\sigma_\epsilon^2)$ variables.

Note that we have assumed $\eta_{ik} = \theta_{ik} + \bar{\xi}_{i.k}$ and ξ_{ijk} to be two independent AR(1) processes and also that $\{\theta_{ik}\}$ and $\{\xi_{ijk}\}$ are independent. It then follows that, for a fixed i, $\{\theta_{ik}\}$ is a stationary ARMA (2,1) process (Fuller (1976), p. 66) with mean zero, variance $\sigma_\theta^2 = \sigma_\eta^2 - b^{-1}\sigma_\xi^2$, and

$$\text{cov}(\theta_{ik}, \theta_{ik'}) = \sigma_\eta^2 \alpha_1^{|k-k'|} - b^{-1} \sigma_\xi^2 \alpha_2^{|k-k'|} , \tag{2.5}$$

where $\sigma_\eta^2 = (1-\alpha_1^2)^{-1} \sigma_e^2$ and $\sigma_\xi^2 = (1-\alpha_2^2)^{-1} \sigma_\epsilon^2$. For our assumptions to be consistent, we require

$$\text{(i)} \quad \sigma_e^2 \geq b^{-1}\,\sigma_\epsilon^2 \;\; \max\{1,\ (1-\alpha_2^2)^{-1}(1-\alpha_1^2),\ (1-\alpha_2)^{-2}\;\;(1-\alpha_1)^2\}$$

and

$$\text{(ii)} \quad \sigma_e^2\,\alpha_2 \geq b^{-1}\,\alpha_1\,\sigma_\epsilon^2 \quad . \tag{2.6}$$

Note that if $\alpha_1 = \alpha_2 = 0$, then $\{\Theta_{ik}\}$ and $\{\xi_{ijk}\}$ are independent $N(0,\sigma_\Theta^2)$ and $N(0,\sigma_\xi^2)$ variables, respectively and the model (2.1) reduces to the traditional split-block model (Steel & Torrie (1980), p. 393). In this case, the sums of squares for blocks, time and blocks by time are obtained using the model (2.2). Similarly, the sums of squares for the plots within blocks and the residual are obtained from the model (2.3).

2.2 *Estimation procedure*

Our estimation procedure consists of several steps. First, we analyse the block means, $\bar{Y}_{i.k}$, over time to obtain estimates of α_1, σ_e^2 and σ_v^2 where $\sigma_v^2 = \sigma_\rho^2 + b^{-1}\,\sigma_\gamma^2$ for case (I) and $\sigma_v^2 = \sigma_\rho^2$ for case (II). Next we analyse the deviations, $Y_{ijk} - \bar{Y}_{i.k}$, to obtain estimates of α_2, σ_γ^2 and σ_ϵ^2. Finally, we obtain a linear transformation of Y_{ijk} involving α_1 and α_2 and the variance components σ_e^2, σ_ρ^2, σ_γ^2 and σ_ϵ^2 such that an ordinary least squares regression of the transformed Y-variable on the corresponding X-variables will give us the estimated generalized least squares (EGLS) estimate for β, the vector of fixed effects.

1. Analysis of the block means: Note that the model (2.2) for block means over time is the same as that of (1.1) considered by Pantula & Pollock (1985), with $\bar{Y}_{i.k}$ and $\bar{X}_{i.k}$ in place of Y_{ik} and X_{ik}, respectively. Therefore, we can obtain the estimates $\hat{\alpha}_1$, $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ for α_1, σ_v^2 and σ_e^2, respectively. Let Z_{ik}^* be the transformation of $\bar{Y}_{i.k}$ such that Z_{ik}^* are "uncorrelated." Let H_{ik}^* be the corresponding transformation of $\bar{X}_{i.k}$.

2. Analysis of the deviations: The model (2.3) for deviations, $Y_{ijk} - \bar{Y}_{i.k}$, is similar to the model (1.1) except for the fact that $\Sigma(Y_{ijk} - \bar{Y}_{i.k}) = 0$, for all i and k, where the sum is over j. The estimation procedure parallels that of Pantula & Pollock (1985). We include the main steps of the procedure for the sake of completeness.

(a) Regress $(Y_{ijk} - \bar{Y}_{i.k})$ on $(X_{ijk} - \bar{X}_{i.k})$ and let $\hat{u}_{ijk}$ denote the least squares residuals from this regression. A ratio-type (method of moments) estimator for α_2 is given by

$$\hat{\alpha}_2 = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{t-2} \hat{u}_{ijk}(\hat{u}_{ij,k+1} - \hat{u}_{ij,k+2})}{\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{t-2} \hat{u}_{ijk}(\hat{u}_{ijk} - \hat{u}_{ij,k+1})} \quad . \tag{2.7}$$

Under certain regularity conditions, it can be shown that

$$\mathrm{Var}(\hat{\alpha}_2) \doteq \frac{2(1+\alpha_2)}{a(b-1)(t-2)} + \frac{2\sigma_\gamma^2(1+\alpha_2)^2}{\sigma_\epsilon^2 a(b-1)(t-2)^2} + \frac{2\alpha_2(1+\alpha_2)}{a(b-1)(1-\alpha_2)(t-2)^2}$$

for fixed b and t and a tending to infinity.

(b) Obtain estimates of σ_ϵ^2 and σ_γ^2 by using the following transformed variables:

$$A_{ijk}^{(1)} = A_{ijk} - A_{ijk}^{(2)} \quad ,$$

$$A_{ijk}^{(2)} = \begin{cases} (1-\hat{\alpha}_2^2)^{1/2} c_2^{-1} d_{ij} , & k = 1 \\ (1-\hat{\alpha}_2) c_2^{-1} d_{ij} , & k \geq 2 \end{cases} ,$$

$$d_{ij} = (1-\hat{\alpha}_2^2)^{1/2} A_{ij1} + (1-\hat{\alpha}_2) \sum_{k=2}^{t} A_{ijk} ,$$

$$c_2 = (1-\hat{\alpha}_2)[t - (t-2)\hat{\alpha}_2] ,$$

and

$$A_{ijk} = \begin{cases} (1-\hat{\alpha}_2^2)^{1/2} (Y_{ij1} - \bar{Y}_{i.1}) , & k = 1 \\ Y_{ijk} - \bar{Y}_{i.k} - \hat{\alpha}_2 (Y_{ij,k-1} - \bar{Y}_{i.k-1}), & k \geq 2 \end{cases} .$$

Obtain, similarly, G_{ijk}, $G_{ijk}^{(1)}$ and $G_{ijk}^{(2)}$ using Z_{ijk} in place of Y_{ijk}. (Note that, if $\hat{\alpha}_2 = 0$, then

$$A_{ijk}^{(2)} = \bar{Y}_{ij.} - \bar{Y}_{i..} \text{ and } A_{ijk}^{(1)} = Y_{ijk} - \bar{Y}_{i.k} - \bar{Y}_{ij.} + \bar{Y}_{i..}) .$$

Regress $A_{ijk}^{(1)}$ on $G_{ijk}^{(1)}$ and $A_{ijk}^{(2)}$ on $G_{ijk}^{(2)}$. Let $\hat{\epsilon}^{(1)\prime}\hat{\epsilon}^{(1)}$ and $\hat{\epsilon}^{(2)\prime}\hat{\epsilon}^{(2)}$ denote the residual sums of squares from these regressions, respectively. Compute

$$\hat{\sigma}_\epsilon^2 = \nu_1^{-1} \hat{\epsilon}^{(1)\prime}\hat{\epsilon}^{(1)} , \tag{2.8}$$

and

$$\hat{\sigma}_{\gamma}^2 = (c_2\nu_2)^{-1} \hat{\epsilon}^{(2)'} \hat{\epsilon}^{(2)} - c_2^{-1} \hat{\sigma}_{\epsilon}^2 \tag{2.9}$$

where

$$\nu_1 = a(b-1)(t-1) - \text{rank } (G^{(1)})$$

and

$$\nu_2 = a(b-1) - \text{rank } (G^{(2)})$$

If $\hat{\sigma}_{\gamma}^2$ is negative, then we set $\hat{\sigma}_{\gamma}^2 = 0$. It can be shown that

$$\text{var}(\hat{\sigma}_{\epsilon}^2) \doteq 2\, \nu_1^{-1}\, \sigma_{\epsilon}^4 \quad ,$$

and

$$\text{var}(\hat{\sigma}_{\gamma}^2) \doteq 2(\nu_2 c_2^2)^{-1}\, (\sigma_{\epsilon}^2 + c_2\sigma_{\gamma}^2)^2 + 2(\nu_1 c_2^2)^{-1}\, \sigma_{\epsilon}^4 \ .$$

(c) Obtain the following transformations, which will be needed in the combined analysis to obtain the EGLS estimate of the fixed effects β:

$$A_{ijk}^{*} = A_{ijk}^{(1)} + \left[(\hat{\sigma}_{\epsilon}^2 + c_2 \hat{\sigma}_{\gamma}^2)^{-1}\, \hat{\sigma}_{\epsilon}^2 \right]^{1/2} A_{ijk}^{(2)} \quad ,$$

and

$$G^{*}_{ijk} = G^{(1)}_{ijk} + \left[(\hat{\sigma}^2_{\epsilon} + c_2\hat{\sigma}^2_{\gamma})^{-1}\hat{\sigma}^2_{\epsilon}\right]^{1/2} G^{(2)}_{ijk} . \tag{2.10}$$

3. Combined analysis: Obtain

$$W_{ijk} = A^{*}_{ijk} + \left[(b\hat{\sigma}^2_{e})^{-1}\hat{\sigma}^2_{\epsilon}\right]^{1/2} Z^{*}_{ik}$$

and

$$B_{ijk} = G^{*}_{ijk} + \left[(b\hat{\sigma}^2_{e})^{-1}\hat{\sigma}^2_{\epsilon}\right]^{1/2} H^{*}_{ik} . \tag{2.11}$$

Regress W_{ijk} on B_{ijk} to obtain the EGLS estimate $\hat{\beta}$ for β. It can be shown that W, the vector consisting of W_{ijk} is $\hat{\sigma}_{\epsilon}\hat{V}^{-1/2}Y$ where $\hat{V}$ is an estimate of V, the variance covariance matrix of Y, the vector of observations. Similarly, the matrix B, consisting of B_{ijk} is $\hat{\sigma}_{\epsilon}\hat{V}^{-1/2}X$. Therefore, the standard errors of the estimator $\hat{\beta}$ in the regression of W on B are asymptotically valid. Also, the matrix $\hat{\sigma}^2_{\epsilon}(B'B)^{-1}$ obtained in this regression is an asymptotically consistent estimate of var($\hat{\beta}$), under certain regularity conditions on X.

Note that, for the traditional split-block model, the parameters α_1 and α_2 are both zero. If we set $\hat{\alpha}_1 = \hat{\alpha}_2 = 0$ in the above estimation procedure, then the estimates of the variance components and of β coincide with those proposed by Pantula, Nelson & Anderson (1985).

2.3 *Hypothesis Testing*

Under certain regularity conditions on X_{ijk}, as in Fuller & Battese (1973), it can be shown that $\hat{\beta}$ is asymptotically normal for fixed b and t and a tending to infinity. Therefore, any hypothesis regarding the fixed effects β can be tested from the regression of the combined analysis. However, care must be used when a is not large. In small samples, the EGLS estimator $\hat{\beta}$ may not be better than the ordinary least squares (OLS) estimator $\tilde{\beta}$ of β obtained by regressing Y_{ijk} on X_{ijk}. Also, the distribution of $\hat{\beta}$ may not be close to normal. If a is not very large, one may use the OLS estimate $\tilde{\beta}$ (instead of $\hat{\beta}$) with an estimated covariance matrix of $\tilde{\beta}$, $(X'X)^{-1} X'\hat{V} X(X'X)^{-1}$, where $\hat{V}$ is the estimate of V obtained above. The method of using the OLS estimate for testing of hypotheses regarding the fixed effects is similar to the tests of contrasts suggested by Rowell & Walters (1976).

Hypotheses regarding the variance components may be tested using F-statistics that are similar to the traditional analysis of variance. For example, to test H_o: $\sigma_\gamma^2 = 0$, we may use the statistic

$$F = \frac{\hat{\epsilon}^{(2)'}\hat{\epsilon}^{(2)}/\nu_2}{\hat{\epsilon}^{(1)'}\hat{\epsilon}^{(1)}/\nu_1} \tag{2.12}$$

where $\hat{\epsilon}^{(1)'}\ \hat{\epsilon}^{(1)}$, $\hat{\epsilon}^{(2)'}\ \hat{\epsilon}^{(2)}$, ν_1 and ν_2 are defined in (2.8) and (2.9). If α_1 and α_2 are known and are used in the computations then the F-statistic in (2.12) has, under H_o, Fisher's F-distribution with degrees of freedom ν_2 and ν_1. When α_1 and α_2 are estimated the tests are approximately valid.

The hypothesis H_o: $\alpha_1 = \alpha_2 = 0$, which means that the traditional split-block model is adequate, can be tested using the independence and the asymptotic normality of $\hat{\alpha}_1$ and $\hat{\alpha}_2$. To test the adequacy of two first order autoregressive models, likelihood ratio type test statistics may be considered. We will discuss this problem in section 4.

2.4 *Large number of repeated measurements*

In some experiments and in some economic situations it is possible to obtain a large number of repeated measurements. In such cases, traditional time series identification procedures (Fuller (1976)) may be used for the orthogonal components, $\bar{Y}_{i.k}$ and $Y_{ijk} - \bar{Y}_{i.k}$, to obtain parsimonious models and efficient estimates for the time series θ_{ik} and ξ_{ijk}. If $\eta_{ik} = \theta_{ik} + \bar{\xi}_{i.k}$ and ξ_{ijk} are identified to be autoregressive processes then the above transformations may be used with minor modifications to obtain the estimates. However, if they are identified to be ARMA type models then it is not easy to obtain simple algebraic expressions for the transformations Z_{ik} and A_{ijk}. In this case, one may use iterative procedures and obtain the maximum likelihood estimates.

3. CROSSED SPLIT-BLOCK MODELS

3.1 *Model presentation*

In this section we consider the example of section 2 with the treatments selected at random. Suppose a set of a locations (or blocks) and a set of b treatments are selected at random. The same set of treatments are used in the different locations selected. Measurements are taken on t consecutive occasions. We consider a linear model,

$$Y_{ijk} = X_{ijk}\beta + \rho_i + \tau_j + \gamma_{ij} + \theta_{ik} + \delta_{jk} + \xi_{ijk} \quad , \qquad (3.1)$$

$$i=1,\ldots,a; \; j=1,\ldots,b; \; k=1,\ldots,t,$$

where $X_{ijk}\beta$ represents the fixed effects and ρ_i, τ_j, γ_{ij}, θ_{ik}, δ_{jk} and ξ_{ijk} represent the block, treatment, block by treatment, block by time, treatment by time and block by treatment by time random effects, respectively. We assume that the random effects are independent of each other; ρ_i are independent $N(0,\sigma_\rho^2)$ variables; τ_j are independent $N(0,\sigma_\tau^2)$ variables and γ_{ij} are independent $N(0,\sigma_\gamma^2)$ variables. We also assume that $\{\theta_{ik}\}$, $\{\delta_{jk}\}$ and $\{\xi_{ijk}\}$ are independent stationary time series. Note that in model (3.1) the random effect associated with the subscripts i and j are crossed with each other, whereas in model (2.1), the random effect associated with the subscript j is nested within i.

We now consider models for the following orthogonal components over time:

1. Blocks: $\bar{Y}_{i.k} - \bar{Y}_{..k} = (\bar{X}_{i.k} - \bar{X}_{..k})\beta + (v_i - \bar{v}_.) + (\eta_{ik} - \bar{\eta}_{.k})$, (3.2)

2. Treatments: $\bar{Y}_{.jk} - \bar{Y}_{..k} = (\bar{X}_{.jk} - \bar{X}_{..k})\beta + (v_j^* - \bar{v}_.^*) + (\eta_{jk}^* - \bar{\eta}_{.k}^*)$, (3.3)

3. Residuals:

$$Y_{ijk} - \bar{Y}_{i.k} - \bar{Y}_{.jk} + \bar{Y}_{..k} = (X_{ijk} - \bar{X}_{i.k} - \bar{X}_{.jk} + \bar{X}_{..k})\beta + u_{ijk}, \quad (3.4)$$

and

4. Overall mean: $\bar{Y}_{..k} = \bar{X}_{..k}\beta + a_k$ (3.5)

where

$$v_i = \rho_i + \bar{\gamma}_{i.} ,$$

$$v_j^* = \tau_j + \bar{\gamma}_{.j} ,$$

$$\eta_{ik} = \theta_{ik} + \bar{\xi}_{i.k} ,$$

$$\eta_{jk}^* = \delta_{jk} + \bar{\xi}_{.jk} ,$$

$$u_{ijk} = (\gamma_{ij} - \bar{\gamma}_{i.} - \bar{\gamma}_{.j} + \bar{\gamma}_{...}) + (\xi_{ijk} - \bar{\xi}_{i.k} - \bar{\xi}_{.jk} + \bar{\xi}_{..k}) ,$$

and

$$a_k = \bar{\rho}_{.} + \bar{\tau}_{.} + \bar{\gamma}_{..} + \bar{\theta}_{.k} + \bar{\delta}_{.k} + \bar{\xi}_{..k} .$$

Following the approach of section 2, we assume that η_{ik}, η^*_{jk} and ξ_{ijk} are independent AR(1) processes given by

$$\eta_{ij} = \begin{cases} (1-\alpha_1^2)^{-1/2} e_{i1} & , k = 1 \\ \alpha_1 \eta_{i,k-1} + e_{ik} & , k \geq 2 \end{cases} ,$$

$$\eta^*_{jk} = \begin{cases} (1-\alpha_1^{*2})^{-1/2} e^*_{j1} & , k = 1 \\ \alpha^*_1 \eta^*_{j,k-1} + e^*_{jk} & , k \geq 2 \end{cases} ,$$

and

$$\xi_{ijk} = \begin{cases} (1-\alpha_2^2)^{-1/2} \epsilon_{ij1} & , k = 1 \\ \alpha_2 \xi_{ij,k-1} + \epsilon_{ijk} & , k \geq 2 \end{cases} ,$$

where $|\alpha_1|<1$, $|\alpha^*_1|<1$, $|\alpha_2|<1$, e_{ik} are independent $N(0,\sigma_e^2)$ variables, e^*_{jk} are independent $N(0,\sigma^2_{e*})$ variables and ϵ_{ijk} are independent $N(0,\sigma^2_\epsilon)$ variables.

Note that the above assumptions imply that for a fixed i, θ_{ik} is an ARMA(2,1) process with covariance structure given in (2.5), provided (2.6) holds. Similarly, for a fixed j, δ_{jk} is an ARMA(2,1) process. If $\alpha_1=\alpha^*_1=\alpha_2=0$, then $\{\theta_{ik}\}$, $\{\delta_{jk}\}$ and $\{\xi_{ijk}\}$ are independent $N(0,\sigma^2_\theta)$, $N(0,\sigma^2_\delta)$ and

$N(0,\sigma_\xi^2)$ variables, respectively and model (3.1) reduces to the traditional split-block model (Steel & Torrie (1980), p. 393). In the traditional analysis, the model (3.2) is used to obtain the sums of squares for blocks and blocks by time; the model (3.3) is used to obtain the sums of squares for treatments and treatments by time; the model (3.4) is used to obtain the sums of squares for blocks by treatment and blocks by treatment by time and the model (3.5) is used to obtain the sums of squares for time. In section 2, since the subscript j is nested within i we could pool the models (3.2) and (3.5); and (3.3) and (3.4) respectively.

3.2 *Estimation procedure*

Our estimation procedure consists of analysing the four orthogonal components separately and then combining the analyses to get the EGLS estimate $\hat\beta$ of β. We now present the main steps of the estimation procedure.

1. From model (3.2) for block deviations obtain the estimates $\hat\alpha_1$, $\hat\sigma_e^2$ and $\hat\sigma_v^2$ of α_1, σ_e^2 and $\sigma_v^2 = \sigma_\rho^2 + b^{-1}\sigma_\gamma^2$. The estimation procedure parallels that of block deviations in section 2, with $\bar Y_{i.k} - \bar Y_{..k}$ in place of $\bar Y_{i.k}$ and (a-1) in place of a in the expressions for $\hat\sigma_e^2$, $\hat\sigma_v^2$, $\mathrm{var}(\hat\sigma_e^2)$, $\mathrm{var}(\hat\sigma_v^2)$ and $\mathrm{var}(\hat\alpha_1)$. Let Z_{ik}^* and H_{ik}^* be the transformations of $\bar Y_{i.k} - \bar Y_{..k}$ and $\bar X_{i.k} - \bar X_{..k}$ as given in section 2.

2. From model (3.3) for treatment deviations obtain, similarly, the estimates $\hat\alpha_1^*$, $\hat\sigma_{e*}^2$ and $\hat\sigma_{v*}^2$ of α_1^*, σ_{e*}^2 and $\sigma_{v*}^2 = \sigma_\tau^2 + a^{-1}\sigma_\gamma^2$. Let Z_{jk}^{**} and H_{jk}^{**} be the transformations of $\bar Y_{.jk} - \bar Y_{..k}$ and $\bar X_{.jk} - \bar X_{..k}$ similar to Z_{ik}^* and H_{ik}^*.

3. From model (3.4) for the residuals obtain, as in section 2, the estimates $\hat{\alpha}_2$, $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_\gamma^2$ of α_2, σ_ϵ^2 and σ_γ^2. Also, obtain A_{ijk}^* and G_{ijk}^* the transformations of $Y_{ijk} - \bar{Y}_{i.k} - \bar{Y}_{.jk} + \bar{Y}_{..k}$ and $X_{ijk} - \bar{X}_{i.k} - \bar{X}_{.jk} + \bar{X}_{..k}$ similar to the transformations in section 2.

4. The analysis for the means over time, however, is different. From model (3.5) we know that,

$$\mathrm{cov}(\bar{Y}_{..k}, \bar{Y}_{..k'}) = (ab)^{-1}(\sigma_\gamma^2 + b\sigma_\rho^2 + a\sigma_\tau^2) \qquad (3.6)$$

$$+ (ab)^{-1}[b\sigma_\eta^2 \alpha_1^{|k-k'|} + a\sigma_{\eta*}^2 \alpha_1^{*|k-k'|} - \sigma_\xi^2 \alpha_2^{|k-k'|}].$$

Since t is only moderate in size, we can compute the txt covariance matrix $C = \mathrm{var}(\bar{Y}_{..1}, \ldots, \bar{Y}_{..t})$ and the positive definite square root $C^{-1/2}$ of C^{-1} . Let

$$g = (g_1, \ldots, g_t)' = C^{-1/2}\ (\bar{Y}_{..1}, \ldots, \bar{Y}_{..t})' \quad ,$$

and

$$P = C^{-1/2}\ (\bar{X}'_{..1}, \ldots, \bar{X}'_{..t})' \quad . \qquad (3.7)$$

We use the estimates of the variance components, α_1, α_1^* and α_2 to compute an estimate of C.

5. Now, to obtain the EGLS estimate $\hat{\beta}$ of β we compute

$$W_{ijk} = A^*_{ijk} + \left[\frac{\hat{\sigma}^2_\epsilon}{b\hat{\sigma}^2_e}\right]^{1/2} Z^*_{ik} + \left[\frac{\hat{\sigma}^2_\epsilon}{a\hat{\sigma}^2_{e*}}\right]^{1/2} Z^{**}_{jk} + \left[\frac{\hat{\sigma}^2_\epsilon}{ab}\right]^{1/2} g_k$$

and

$$B_{ijk} = G^*_{ijk} + \left[\frac{\hat{\sigma}^2_\epsilon}{b\hat{\sigma}^2_e}\right]^{1/2} H^*_{ik} + \left[\frac{\hat{\sigma}^2_\epsilon}{a\hat{\sigma}^2_{e*}}\right]^{1/2} H^{**}_{jk} + \left[\frac{\hat{\sigma}^2_\epsilon}{ab}\right]^{1/2} P_k .$$

The OLS estimate obtained by regressing W_{ijk} on B_{ijk} is the EGLS estimate $\hat{\beta}$ of β. Also, the matrix $\hat{\sigma}^2_\epsilon(B'B)^{-1}$ obtained from this regression, under certain regularity conditions, will be a consistent estimate for the asymptotic variance covariance matrix of $\hat{\beta}$. Any hypothesis regarding the parameter can be tested as in section 2.

3.3 *Large number of repeated measurements*

If the number of repeated measurements, t, is large, then traditional time series (frequency and time domain) identifying procedures may be used for the orthogonal partitions, $\bar{Y}_{i.k} - \bar{Y}_{..k}$, $\bar{Y}_{.jk} - \bar{Y}_{..k}$ and $Y_{ijk} - \bar{Y}_{i.k} - \bar{Y}_{.jk} + \bar{Y}_{..k}$, to obtain parsimonious models for the time series θ_{ik}, δ_{jk} and ξ_{ijk}. Also, with minor modifications, transformations of section 3.2 may be used to obtain the EGLS estimate of β.

Another model that is sometimes used is the symmetric random effects model given by

$$Y_{ijk} = X_{ijk}\beta + \rho_i + \tau_j + \ell_k + \gamma_{ij} + \theta_{ik} + \delta_{jk} + \xi_{ijk} , \tag{3.8}$$

where the component ℓ_k is included to reflect the global (time) effect. The time series ℓ_k is assumed to be independent of the remaining random components. We can again use the orthogonal components to obtain parsimonious models for the time series θ_{ik}, δ_{jk} and ξ_{ijk} and then use the model for $\bar{Y}_{..k}$ to identify a model for the time series ℓ_k.

Suppose the model (3.8) holds with ρ_i, τ_j, γ_{ij}, θ_{ik}, δ_{jk} and ξ_{ijk} as defined in section 3.2. Note that,

$$\bar{Y}_{..k} = \bar{X}_{..k}\beta + \bar{\rho}_{.} + \bar{\tau}_{.} + \bar{\gamma}_{..} + b_k \tag{3.9}$$

where

$$b_k = \ell_k + \bar{\theta}_{.k} + \bar{\delta}_{.k} + \bar{\xi}_{..k} \quad .$$

If b_k is an AR(1) process over time with parameters α and σ^2, then one can estimate α and σ^2 as in section 2. Also, simple algebraic expressions can be given for the transformations in (3.6) unlike in model (3.1). The estimates of the remaining parameters can be obtained using the procedures in section 3.2.

4. CONCLUDING REMARKS

In this paper we have considered simple extensions of traditional split-block models. We believe that the paper presents two important general concepts regarding the analysis of repeated measurements data from some randomized experimental design. First, we regard the different orthogonal components of the design over time as independent sets of time series. In some situations, as in section 2, some of the orthogonal components may be pooled. The estimation of several time series components is thus simplified. Second, if autoregressive models are considered for the time series in the orthogonal components, explicit transformations are available that can be used to obtain consistent estimates of the autocorrelations, variance components and the fixed effects. Some readers may find one or the other or both of these aspects useful in other settings.

We have considered particular (ARMA) models for the time series components in the model. If the number of time points (repeated measurements), t, is large then traditional time series methods can be used to identify and test the adequacy of the models. When t is small or moderate in size, we should use likelihood ratio type statistics for testing the adequacy of the hypothesized models. Olkin & Vaeth (1981) and Jennrich & Schluchter (1985) present iterative procedures to obtain the maximum likelihood estimates for models with structured covariance matrices. Using these maximum likelihood estimates one may test the adequacy of our autoregressive models against stationary and arbitrary (multivariate) covariance structures. The estimates presented in this paper are not the most efficient estimates of the parameters. Our estimation procedure is, however, consistent, simple and similar to the traditional analysis of split-block models. For example, when $\alpha_1 = \alpha_2 = 0$ in model (2.1) our estimates coincide with those of Pantula, Nelson & Anderson (1985). If, in addition, the covariables are simple dummy variables, our estimates coincide with those of the regular analysis of variance estimates. Also, our estimation procedure can be used to test whether the traditional split-block model is adequate as compared to our model.

The results presented in this paper are asymptotic in nature. We assume the number of blocks, a, to be large. If a is not very large, the EGLS estimator $\hat{\beta}$ may or may not be better than the OLS estimator $\tilde{\beta}$ of β. Also, the small sample distribution of $\hat{\beta}$ may not be well approximated by a normal distribution. On the other hand, the distribution of the OLS estimate $\tilde{\beta}$ is $N(0,var(\tilde{\beta}))$. If a is not very large one may use the OLS estimate $\tilde{\beta}$ with an estimated covariance matrix $\hat{var}(\tilde{\beta}) = (X'X)^{-1} X'\hat{V} X(X'X)^{-1}$, where $\hat{V}$ is the estimate of $V = var(Y)$ obtained in our procedure. Small sample properties of all our estimators need to be studied. Also, the effect of autocorrelations on the properties of the OLS estimator $\tilde{\beta}$ and the variance components needs to be investigated.

The modelling approaches and the estimation procedures presented in this paper can be used to analyse other balanced orthogonal designs with repeated time measurements. For example, with the Latin square design, there can be orthogonal components to model as time series for the overall mean, row component, column component, treatment component and residual component. The exact form of the analysis will depend on which effects are classified as fixed and random. (See also Kunert (1985).)

In this paper, we have considered only linear models with random and autocorrelated error components. In addition to presenting consistent estimates for the autocorrelations and variance components, we presented a transformation to obtain the weighted least squares estimate for the fixed effects. This transformation and the estimates of the autocovariance structure can also be used in estimating nonlinear (growth curve) models. (See Gallant (1975).)

ACKNOWLEDGEMENTS

We wish to thank Professors Peter Bloomfield and Cavell Brownie for their comments. This work was partially funded by NSF Grant DMS-8406372 A01.

REFERENCES

Anderson, A. H., Jensen, E. B. & Schou, G. (1981). Two-way analysis of variance with correlated errors. International Statistical Review **49**, 153-67.

Azzalini, A. (1984). Estimation and hypothesis testing for collections of autoregressive time series. Biometrika **71**, 85-90.

Balestra, P. & Nerlove, M. (1966). Pooling cross-section and time series data in the estimation of a dynamic model; the demand for natural gas. Econometrica **34**, 585-612.

Bloomfield, P., Oehlert, G., Thompson, M. L. & Zeger, S. (1983). A frequency domain analysis of trends in Dobson total ozone records. Journal of Geophysical Research **88**, 8512-22.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and of correlation between errors in the two-way classification. The Annals of Mathematical Statistics **25**, 484-98.

Danford, M. B., Hughes, H. M., & McNee, R. C. (1960). On the analysis of repeated measurements experiments. Biometrics **16**, 547-65.

Fuller, W. A. (1976). Introduction to Statistical Time Series. Wiley, NY.

Fuller, W. A. & Battese, G. E. (1973). Transformations for the estimation of linear models with nested-error structure. Journal of the American Statistical Association **68**, 626-32.

Gallant, A. R. (1975). Seemingly unrelated nonlinear regressions. Journal of Econometrics **3**, 35-50.

Huynh, H. (1978). Some approximate tests for repeated measurement designs. Psychometrika **43**, 161-75.

Jennrich, R. I. & Schluchter (1985). Incomplete repeated measures models with structured covariance matrices. Paper presented at the Joint Statistical Meetings of the American Statistical Association. Las Vegas, Nevada, USA.

Kunert, J. (1985). Optimal repeated measurements designs for correlated observations and analysis by weighted least squares. Biometrika **72**, 375-90.

Olkin, I. & Vaeth, M. (1981). Maximum likelihood estimation in a two-way analysis of variance with correlated errors in one classification. Biometrika **68**, 653-60.

Pantula, S. G. & Pollock, K. H. (1985). Nested analysis of variance with autocorrelated errors. To appear in Biometrics (December).

Pantula, S. G., Nelson, L. A. & Anderson, R. L. (1985). Estimation of linear models for field experiments. Communications in Statistics - Theory and Methods **14**, 2199-217.

Rowell, J. G. & Walters, D. E. (1976). Analysing data with repeated observations on each experimental unit. Journal of Agricultural Science **87**, 423-32.

Shumway, R. H. (1970). Applied regression and analysis of variance for stationary time series. Journal of American Statistical Association **65**, 1527-36.

Steel, R. G. D. & Torrie, J. H. (1980). Principles and Procedures of Statistics: A Biometric Approach. McGraw Hill, New York.

Thomas, C. & Wilkinson, J. M. (1975). The utilization of maize silage for intensive beef production. 3. Nitrogen and acidity as factors affecting the nutrition value of ensiled maize. Journal of Agricultural Science **85**, 255-61.

Wallace, T. D. & Hussian, A. (1969). The use of error component models in combining cross-section with time series data. Econometrica **37**, 55-72.

Wallenstein, S. & Fleiss, J. L. (1979). Repeated measurements analysis of variance where the correlations have a certain pattern. Psychometrika **44**, 229-33.