

TECHNIQUES FOR ANALYZING REPEATED MEASURES EXPERIMENTAL
DESIGNS WITH AUTOREGRESSIVE SUBPLOT ERRORS

by

DANIEL P. JOHNSON

B.S., Kansas State University, 1982
MBA, Kansas State University, 1984

A REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

STATISTICS

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1989

Approved by:



Dr. Mark S. McNulty

DEDICATION

A11208 301304

I dedicate this report to Dr. Mark S. McNulty,
to my parents Maynard and Micheline Johnson, to
my brothers and sister John, Lyndon, Donald, and
Kathy, and finally to Eileen Smith.

LD
21668
.R4
STA1
1989
J64
c. 2

ACKNOWLEDGEMENT

I owe Dr. Mark S. McNulty a debt of gratitude. This report would not have been possible without his encouragement, professional advice, and guidance.

Furthermore, I would like to thank Dr. Higgins and Dr. Yang for their encouragement and support.

TABLE OF CONTENTS

List of Tables	iv
I. Generalized Least Squares	1
Simple Linear Regression Model with Autoregressive Errors	3
Consequences of Autocorrelated Errors	5
GLS Estimates	6
II. Repeated Measures with Autoregressive Subplot Errors	8
Split-plot Analysis	9
Autoregressive Repeated Measures	10
Box's Adjusted Degrees of Freedom	11
GLS Estimates	15
GLM-GLS	19
Contrasts	24
Recommended Strategy	25
III. Example	26
Box's Adjusted Degrees of Freedom	26
GLS	31
GLM-GLS	33
Contrasts	35
Bibliography	40

LIST OF TABLES

Table 1: Split-plot Anova Table	10
Table 2: Actual and Adjusted Degrees of Freedom	13
Table 3: Values for Gamma When the Correlation Matrix has a Simplex Pattern	14
Table 4: Body Weights of Rats	26
Table 5: Split-plot Anova Table - Example	29
Table 6: Actual and adjusted Degrees of Freedom - Example	30
Table 7: GLS Anova Table	32
Table 8: GLM-GLS Anova Table	34
Table 9: GLS and GLM-GLS Parameter Estimates	36

I. GENERALIZED LEAST SQUARES

Consider the linear model:

$$y = X\beta + \varepsilon ,$$

where ε is $N(0, \sigma^2 \Sigma)$ and X is an $t \times p$ matrix of rank p . The ordinary least squares (OLS) estimate of β for this model is:

$$\hat{\beta} = (X'X)^{-1} X'y.$$

Furthermore, an estimate of $\text{Cov}(\hat{\beta})$ is given by:

$$\hat{\text{Cov}}(\hat{\beta}) = (X'X)^{-1} S^2 ,$$

where $S^2 = (n - p)^{-1} y'(I - X(X'X)^{-1} X')y$.

OLS is appropriate for the linear model presented above if the covariance matrix of the error term (ε) conforms to particular patterns. The necessary conditions are discussed in a paper published by Huynh and Feldt (1970). According to Milliken and Johnson (1984), the Huynh-Feldt (H-F) condition is equivalent to specifying that the variances of the differences between pairs of errors, such as $\varepsilon_i - \varepsilon_{i'}$, are equal for all

i and i' , $i \neq i'$. If a model has more than one error term, the covariance matrices of each error vector must satisfy the H-F condition for OLS to be valid.

There are essentially 3 consequences associated with using OLS when the H-F condition does not hold. They are:

1. $\hat{\beta}$ is unbiased. $E(\hat{\beta}) = \beta$.
2. $\hat{Cov}(\hat{\beta})$ will be biased.
3. Ordinary least-squares will give inefficient predictions.

Departures from the H-F condition can be remedied using generalized least squares (GLS). The GLS estimate of β is:

$$\hat{\beta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y,$$

which can be computed as a regression of Ay on AX , where

$$\Sigma^{-1} = A A'.$$

An estimate of $\hat{Cov}(\hat{\beta})$ is given by:

$$\hat{Cov}(\hat{\beta}) = (X' \Sigma^{-1} X)^{-1} S^{-2},$$

where $S^2 = (n - p)^{-1} y' (\Sigma^{-1} - \Sigma^{-1} X (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1}) y$.

According to Arnold (1981), the following are

properties of the GLS estimators $\hat{\mu}$ (where $\hat{\mu} = X\hat{\beta}$) and \hat{S}^2 :

1. $(\hat{\mu}, \hat{S}^2)$ is a complete sufficient statistic.
2. $\hat{\mu}$ and \hat{S}^2 are the minimum variance unbiased estimators of μ and σ^2 .
3. $\hat{\mu}$ and $[(n-p)/n]\hat{S}^2$ are the maximum likelihood estimators of μ and σ^2 .

SIMPLE LINEAR REGRESSION MODEL WITH AUTOREGRESSIVE ERRORS

To illustrate the problems that arise when serially correlated disturbances exist, let us postulate the following model:

$$Y_k = \beta_0 + \beta_1 X_k + \varepsilon_k, \quad k = 1, \dots, t,$$

where we assume the error term ε_k to be first-order autoregressive and of the form:

$$\varepsilon_k = \phi \varepsilon_{k-1} + e_k$$

where ϕ represents the coefficient of correlation and

hence $-1 < \phi < 1$. Furthermore, e_k satisfies the following assumptions for all k :

$$E(e_k) = 0$$

$$E(e_k e_{k+r}) = \sigma_e^2 \phi^r \quad r = 0$$

$$= 0 \quad r \neq 0$$

Given the above assumptions, the variance-covariance matrix for the serially correlated disturbances can be written as:

$$E(\epsilon\epsilon') = \sigma_\epsilon^2 \quad V = \sigma_e^2$$

$$\begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{t-1} \\ \phi & 1 & \phi & \dots & \phi^{t-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \phi^{t-1} & \phi^{t-2} & \phi^{t-3} & \dots & 1 \end{bmatrix}$$

txt

where $\sigma_\epsilon^2 = \sigma_e^2 / (1 - \phi^2)$ for all t .

CONSEQUENCES OF AUTOCORRELATED ERRORS

In his book Econometric Methods (2nd ed.), Johnston describes three consequences of applying ordinary least-squares to a model with disturbances that are autocorrelated. They are:

1. the estimates of β are unbiased,
2. the sampling variances of the regression coefficients are likely to be seriously underestimated, and
3. ordinary least-squares will give inefficient predictions.

Because of the aforementioned concerns, any t or F test performed on the parameters of the model will be incorrect because $\hat{\text{cov}}(\hat{\beta})$ is biased and because SSE/σ^2 is not distributed as a chi-square nor is it independent of $(b - \beta)$, (where b is the OLS estimate of β).

Given that the matrix Σ is known, the generalized least squares estimator of β and its corresponding standard error can be computed so that the usual confidence interval and hypothesis test of β can be constructed. The minimum variance unbiased estimator of β in this situation is provided by generalized least-squares (GLS).

GLS ESTIMATES

Consider the matrix representation of the simple linear regression model:

$$y = X\beta + \varepsilon$$

$$\text{with } E(\varepsilon) = 0, \quad E(\varepsilon\varepsilon') = \sigma^2 \Sigma.$$

Given that the error term is known to be autoregressive and that Φ is known, GLS estimates for this model can be obtained either directly or by a two-stage procedure.

The direct GLS estimator of β is:

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}y.$$

The variance of this estimator is:

$$\text{VAR}(\hat{\beta}) = s^2 (X'\Sigma^{-1}X)^{-1},$$

where $s^2 = \delta'\Sigma^{-1}\delta / (n - 2)$ and $\delta = y - X\hat{\beta}$.

The two-stage procedure consists of:

1. transforming the elements of the Y vector and the X matrix by appropriate functions of Φ , and
2. applying OLS to the transformed data.

The model is transformed by premultiplying the Y vector, X matrix and error vector by a matrix A ($Ay = AX\beta + A\epsilon$)

such that $E(A\epsilon\epsilon'A') = \sigma^2 I$. The appropriate transform-

ation matrix in the context of a simple linear regression model with t observations and an AR(1) error structure is:

$$A = \begin{bmatrix} \text{SQRT}(1 - \phi^2) & 0 & 0 & \dots & 0 & 0 & 0 \\ -\phi & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\phi & 1 & \dots & 0 & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & \dots & -\phi & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -\phi & 1 \end{bmatrix}$$

txt

Ordinary least-squares is applied to the following transformed data:

$$Ay = \begin{bmatrix} \text{SQRT}(1 - \phi^2) Y_1 \\ Y_2 - \phi Y_1 \\ Y_3 - \phi Y_2 \\ \vdots \\ Y_t - \phi Y_{t-1} \end{bmatrix}$$

$$AX = \begin{bmatrix} \text{SQRT}(1 - \phi^2) & \text{SQRT}(1 - \phi^2)X_1 \\ 1 - \phi & X_2 - \phi X_1 \\ \vdots & \vdots \\ 1 - \phi & X_t - \phi X_{t-1} \end{bmatrix}$$

The same transformation applied to the simple linear regression can be extended to a repeated measures experimental design for which the subplot errors are serially correlated.

II. REPEATED MEASURES WITH AUTOREGRESSIVE SUBPLOT ERRORS

A repeated measures model is similar to a split-plot model, except for the assumptions inherent to the respective models. The model is:

$$Y_{ijk} = \mu + \alpha_i + \pi_{ij} + T_k + (\alpha T)_{ik} + \epsilon_{ijk} \quad (1),$$

$i = 1, \dots, a, j = 1, \dots, n, k = 1, \dots, t.$

Where μ is the overall mean.

α_i represents whole plot treatment effects.

- π_{ij} is the error term associated with the whole plots.
- $(\alpha T)_{ik}$ represents the whole plot treatment by subplot treatment interaction.
- T_j represents subplot treatment effects (usually time for a repeated measures experimental design).
- ϵ_{ijk} is the error term associated with the subplots.

The form of the covariance matrices associated with the error terms determines the appropriate analysis of a repeated measures experimental design. Two conditions will be discussed in particular:

1. the case for which a split-plot analysis is appropriate, and
2. the case where the covariance matrix of the subplot errors is assumed to be autoregressive.

SPLIT-PLOT ANALYSIS

A split-plot analysis of a repeated measures experimental design is appropriate if the covariance matrix for each error term in the model conforms to the Huynh-Feldt (H-F) condition. An Analysis of Variance table can be written as:

Table 1		ANOVA
Source of Variation	df	SS
α	$a - 1$	$SS(\alpha)$
Error(Whole Plot)	$a(n - 1)$	$SSE(\text{Whole Plot})$
T	$t - 1$	$SS(T)$
$\alpha * T$	$(a - 1)(t - 1)$	$SS(\alpha * T)$
Error(Subplot)	$a(n - 1)(t - 1)$	$SSE(\text{Subplot})$

		ANOVA
Source of Variation	MS	F
α	$SS(\alpha)/(a - 1)$	$\frac{MS(\alpha)}{MSE(\text{Whole Plot})}$
Error(Whole Plot)		
T	$SS(T)/(t - 1)$	$\frac{MS(T)}{MSE(\text{Subplot})}$
$\alpha * T$	$\frac{SS(\alpha * T)}{(a - 1)(t - 1)}$	$\frac{MS(\alpha * T)}{MSE(\text{Subplot})}$
Error(Subplot)		

AUTOREGRESSIVE REPEATED MEASURES

If the subplot error structure is autoregressive by nature, the test statistics computed from a split-plot analysis of the subplot treatment (T) and the whole plot treatment by subplot treatment interaction ($\alpha * T$) is

erroneous. However, the whole plot tests still remain valid even if the subplot error (ϵ_{ijk}) is not independent.

Three methods are proposed to remedy the shortcomings of ordinary least squares (split-plot analysis) when the subplot error structure is autoregressive. They are:

1. Box's adjusted degrees of freedom associated with the subplot analysis,
2. generalized least squares (GLS), and
3. generalized linear model - generalized least squares (GLM - GLS).

Following the introduction of each methodology, an example will be presented to illustrate each concept.

BOX'S ADJUSTED DEGREES OF FREEDOM

According to Box [1954], an adjustment can be made to the degrees of freedom of an experimental design with an error structure of the form $\sigma^2 \Sigma$. The degrees of freedom are multiplied by a constant which is computed as follows:

$$\gamma = \frac{t(V - V)}{aa} \dots$$

$$(t - 1) \begin{pmatrix} a & b & & & \\ \Sigma & \Sigma & V & & \\ 1 & 1 & ab & & \end{pmatrix} - 2t \begin{pmatrix} a & & & & \\ \Sigma & V & & & \\ 1 & a & & & \end{pmatrix} + t \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{pmatrix}$$

where $(V_{ab}) = V$ is the covariance matrix of the t observations on a subject. \bar{V}_{aa} is the mean of the diagonal elements of V , $\bar{V}_{..}$ is the mean of all elements of V and \bar{V}_a is the mean of the elements in row "a" of V .

Wallenstein and Fleiss [1979] have extended Box's idea to encompass situations where the subplot errors are autoregressive by nature. For such a covariance structure, the degrees of freedom associated with the subplot treatment, the interaction between the whole plot treatment and the subplot treatment, and the subplot error are multiplied by the constant:

$$\gamma = R/B, \text{ where}$$

$$R = (1 + \phi)^2 \{t(1 - \phi)[t(1 - \phi) - (1 + \phi)] + 2\phi(1 - \phi)^2\}$$

and,

$$B = (t - 1)\{4\phi(1 - \phi)^t(1 + \phi)[\phi(1 + \phi)(1 - \phi)^t + t(1 - \phi)(1 + \phi)^2 - t\phi(1 - \phi)(1 + \phi)^t] - t(1 - \phi)^2 [2\phi(1 - \phi)^{2t} + (1 + \phi)^4 + 4(1 + \phi)\phi^{2t+1}] + t^3(1 - \phi)(1 + \phi)(1 + \phi)^2\}$$

The degrees of freedom to be adjusted can be obtained from the split-plot ANOVA table above. The appropriate adjustments are presented in Table 2 below.

Table - 2 Actual and Adjusted Degrees of Freedom

Source_of_Variation	Actual_df	Adjusted_df
α	$a - 1$	$(a - 1)^*$
Error(Whole Plot)	$a(n - 1)$	$a(n - 1)^*$
T	$t - 1$	$\gamma(t - 1)$
$\alpha * T$	$(a - 1)(t - 1)$	$\gamma[(a - 1)(t - 1)]$
Error(Subplot)	$a(n - 1)(t - 1)$	$\gamma[a(n - 1)(t - 1)]$

* Unchanged

The appropriate F - statistics can be computed once the adjusted degrees of freedom have been computed for the subplot. For example, assuming the subplot treatment is time (T), the appropriate computation is:

$$F = MS(\text{Time})/MS(\text{Error}(\text{Time}))$$

where $MS(\text{Time}) = SS(\text{Time})/\text{Adj. df Time}$, and
 $MS(\text{Error}(\text{Time})) = SS(\text{Error}(\text{Time}))/\text{Adj. df Error}(\text{Time})$.

Since the numerator and denominator sums of squares have been divided by degrees of freedom that have been adjusted by the same factor, γ , the F-statistic for the time effect will be exactly the same as the F-statistic given in the standard split-plot analysis above.

However, the critical region used in the analysis will differ from that of the standard split-plot analysis.

Table 3 below gives values of γ for $\phi = 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95$, and for $t = 3, 4, 5, 7, 10, 25, 50$, and infinity.

Table - 3: VALUES FOR GAMMA WHEN THE CORRELATION MATRIX HAS A SIMPLEX PATTERN

ϕ	NUMBER OF TIME POINTS							
	3	4	5	7	10	25	50	∞
0.05	0.999	0.998	0.997	0.997	0.996	0.995	0.995	0.995
0.10	0.996	0.992	0.990	0.987	0.985	0.982	0.981	0.980
0.15	0.991	0.983	0.977	0.971	0.966	0.960	0.958	0.956
0.20	0.985	0.971	0.961	0.949	0.940	0.929	0.926	0.923
0.25	0.977	0.956	0.941	0.923	0.909	0.892	0.887	0.882
0.30	0.968	0.940	0.919	0.892	0.872	0.848	0.841	0.835
0.35	0.958	0.921	0.893	0.858	0.831	0.799	0.790	0.782
0.40	0.948	0.901	0.866	0.822	0.787	0.745	0.734	0.724
0.45	0.936	0.880	0.838	0.783	0.740	0.688	0.674	0.663
0.50	0.925	0.859	0.809	0.743	0.692	0.692	0.613	0.600
0.55	0.912	0.837	0.779	0.703	0.643	0.568	0.550	0.536
0.60	0.900	0.814	0.749	0.663	0.594	0.507	0.486	0.471
0.65	0.887	0.791	0.719	0.623	0.546	0.446	0.423	0.406
0.70	0.875	0.769	0.689	0.584	0.449	0.387	0.360	0.342
0.75	0.862	0.747	0.660	0.546	0.454	0.330	0.299	0.280

VALUES FOR GAMMA WHEN THE CORRELATION MATRIX
HAS A SIMPLEX PATTERN

ϕ	NUMBER OF TIME POINTS							
	3	4	5	7	10	25	50	∞
0.80	0.849	0.724	0.631	0.510	0.411	0.275	0.240	0.220
0.85	0.837	0.703	0.604	0.475	0.371	0.225	0.185	0.161
0.90	0.824	0.682	0.577	0.442	0.333	0.179	0.133	0.105
0.95	0.812	0.661	0.551	0.410	0.298	0.138	0.088	0.051
LOWER BOUND	0.800	0.641	0.526	0.381	0.266	0.103	0.051	
<u>1</u> (T-1)	0.500	0.333	0.250	0.167	0.111	0.042	0.020	

Source: REPEATED MEASUREMENTS ANALYSIS OF VARIANCE WHEN
THE CORRELATIONS HAVE A CERTAIN PATTERN By
Sylan Wallenstein & Joseph I. Fleiss (1979: p. 231)

Table 3 is useful in determining whether an adjustment is needed to the subplot analysis. Box's recommended adjustment to the degrees of freedom does not result in tests that are as powerful as tests generated by generalized least squares (GLS). However, GLS is time consuming. If the adjustment factor does not change the degrees of freedom much, one need not expend the time nor effort performing a GLS analysis on a repeated measures experiment with an autoregressive error structure.

GLS ESTIMATES

An alternative expression for the split-plot model

which emphasizes the nature of the whole plots is:

$$Y_{ijk} = \mu_{ij} + T_k + (\alpha T)_{ik} + \varepsilon_{ijk} \quad (2),$$

$$i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, t,$$

where μ_{ij} is a block effect for the whole plot

experimental units and is equal $\mu + \alpha_i + \pi_{ij}$. The

blocks average out the effect of the autoregressive subplot disturbances, validating the whole plot analysis of a split-plot model.

Model two above can be represented in matrix notation as:

$$y = X\beta + \varepsilon,$$

where y is an $(ant) \times 1$ response vector whose members consist of y_{ijk} ,

X is an $(ant) \times [a(n+t) + t]$ design matrix, and

β is an $[a(n+t) + t] \times 1$ vector of parameter estimates. $\beta' = [\mu', T', \alpha T']$,

$$\text{where } \mu' = [\mu_{11}, \mu_{12}, \dots, \mu_{an}]',$$

$$T' = [T_1, T_2, \dots, T_t]', \text{ and}$$

$$\alpha T' = [\alpha T_{11}, \alpha T_{12}, \dots, \alpha T_{at}]'.$$

ϵ is an $(ant) \times 1$ error vector whose members consist of ϵ_{ijk} .

To illustrate what the design matrix for such a model looks like, let us look at the simplest case ($a=n=t=2$). Assume an experiment was conducted to monitor the effect of a drug on the heart rate of a patient. The experiment was designed so that two patients ($n=2$) were randomly assigned to each of the two drugs ($a=2$) tested. The heart rates per minute were obtained for each of two weeks ($t=2$). The design matrix for this scenario is:

$$X = \left[\begin{array}{cccc|cc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{array} \right],$$

where the first partition consists of indicator variables representing the whole plots, the second partition represents time, and the final partition of indicator variables represents the drug by time interaction.

Since the design matrix for the split-plot model is not full column rank, a generalized inverse needs to be used to estimate β . For example, given that ϵ is

$$N(0, \sigma^2 I)$$

$$\hat{\beta} = (X'X)^{-1} X'y$$

Given that ε is autoregressive, the covariance matrix for the subplot error is:

$$E(\varepsilon'\varepsilon) = \sigma^2 \Sigma = \sigma^2 (I \otimes V)$$

The following steps should be taken to achieve an appropriate analysis of a repeated measures experimental design with autoregressive subplot disturbances:

1. Perform a split-plot analysis using OLS. The F-statistics for the whole-plot analysis will be correct, however, the subplot analysis will give erroneous F-statistics. The next two steps are needed to correct the subplot analysis.
2. The second step consists of premultiplying the design matrix, X , and the response vector, y , by the matrix Z , where $Z = I \otimes A$.
3. Because $\text{var}(Z\varepsilon) = \sigma^2 I$, one can obtain GLS estimators of β and σ^2 by running OLS on the

variables of the model:

$$y^* = X \beta^* + \epsilon^*$$

where $y^* = Zy$,

$$X^* = ZX,$$

$$\epsilon^* = Z\epsilon$$

The output from regressing Zy on ZX will provide appropriate F-statistics for the subplot part of the analysis.

The OLS estimator of β (GLM estimator of β for the split-plot model) is:

$$\hat{\beta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y = (X' Z' Z X)^{-1} X' Z' Z y$$

Recall that GLS produces minimum variance unbiased estimators of the mean vector and covariance matrix. However, the design matrix, X , needs to be constructed to implement this procedure. This is time consuming.

GLM-GLS

The procedures and properties for the generalized linear model - general least squares (GLM-GLS) method is similar to that of the regular GLS analysis presented

above. However, GLM-GLS is less time consuming.

Consider the transformation:

$$\begin{aligned}
 Y_{ijk}^+ &= (1 - \phi) y_{ijk}^+ , & k = 1 \\
 &= y_{ijk}^+ - \phi y_{ij(k-1)}^+ , & k = 2, \dots, t.
 \end{aligned}$$

In terms of the split-plot model, this translates into:

$$Y_{ijk}^+ = \mu_{ij}^+ + T_k^+ + \alpha T_{ik}^+ + \varepsilon_{ijk}^+ \quad (3) ,$$

$$\text{where } \mu_{ij}^+ = (1 - \phi) \mu_{ij}^+ , \quad k = 1, \dots, t.$$

$$\begin{aligned}
 T_k^+ &= (1 - \phi) T_k^+ , & k = 1 \\
 &= T_k^+ - \phi T_{k-1}^+ , & k = 2, \dots, t.
 \end{aligned}$$

$$\begin{aligned}
 \alpha T_{ik}^+ &= (1 - \phi) \alpha T_{ik}^+ , & k = 1 \\
 &= \alpha T_{ik}^+ - \phi \alpha T_{i(k-1)}^+ , & k = 2, \dots, t.
 \end{aligned}$$

$$\begin{aligned}
 \varepsilon_{ijk}^+ &= (1 - \phi) \varepsilon_{ijk}^+ , & k = 1 \\
 &= e_{ijk}^+ , & k = 2, \dots, t.
 \end{aligned}$$

The errors from this transformed model are now

independent, but do not have constant variance.

$$\begin{aligned} \text{Var}(\varepsilon_{ijk}^+) &= \frac{(1 - \phi)^2}{1 - \phi^2}, \quad k = 1 \\ &= \sigma_e^2, \quad k = 2, \dots, t. \end{aligned}$$

However, a nonconstant variance can easily be dealt with by most computer packages.

Model three can be represented in matrix notation as:

$$y^+ = X\beta^+ + \varepsilon^+$$

where,

$$\beta^+ = \begin{bmatrix} \mu^+ \\ T^+ \\ \alpha T^+ \end{bmatrix}$$

$$\mu^+ = (1 - \phi) \mu'$$

$$T^+ = DT'$$

$$\alpha T^+ = (I \otimes D) \alpha T'$$

where,

$$D = \begin{bmatrix} (1 - \phi) & 0 & 0 & \dots & 0 & 0 & 0 \\ -\phi & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\phi & 1 & \dots & 0 & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & \dots & -\phi & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -\phi & 1 \end{bmatrix}$$

txt

The GLS estimator of β^+ is given by:

$$\hat{\beta}^+ = (X'WX)^{-1} X'Wy^+$$

where $W = I_n \otimes G$, and

$$G = \begin{bmatrix} \frac{1 - \phi^2}{(1 - \phi)} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}$$

txt

Now, D is a non-singular invertible matrix with inverse:

$$D^{-1} = \begin{bmatrix} \frac{1}{(1-\phi)} & 0 & 0 & 0 & \dots & 0 \\ \frac{\phi}{(1-\phi)} & 1 & 0 & 0 & \dots & 0 \\ \frac{\phi^2}{(1-\phi)} & \phi & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\phi^{t-1}}{(1-\phi)} & \phi & \phi & \phi & \dots & 1 \end{bmatrix}$$

txt

so that a test of $H : T = 0$ is the same as testing

$H : T^+ = 0$, and a test of $H : \alpha T = 0$ is the same as

testing $H : \alpha T^+ = 0$. As with GLS, the whole-plot

F-tests produced by a standard split-plot analysis will be correct.

A statistical package such as SAS will construct the design matrix, X , and the matrix of weights, W . One need only specify the weights along the diagonal of the matrix W and construct y^+ , which is relatively easy.

Presented below is an example of the SAS code needed to yield an appropriate analysis of the subplots for the simplest case presented above ($a=n=t=2$):

```

DATA EXAMPLE;
INPUT PERSON TIME DRUG Y;
/* COEFFICIENT OF CORRELATION IS ASSUMED*/;
/* FOR ILLUSTRATIVE PURPOSES TO EQUAL 0.6 */;
RHO = 0.6;
/* THE WEIGHTS FOR THE MATRIX W NEED TO BE */;
/* SPECIFIED AND ARE ASSUMED TO BE 4 FOR */;
/* TIME PERIOD 1 AND 1 FOR TIME PERIOD 2 */;
  IF TIME = 1 THEN
    W=(1 - RHO^2)/(1 - RHO)^2;
  ELSE W=1;
  LAGY = LAG(Y);
  IF TIME = 1 THEN
    YPLUS = (1 - RHO)*Y;
  ELSE YPLUS = Y - RHO*LAGY;
CARDS;
.
.
.
.
data
.
.
.
;
PROC GLM DATA=EXAMPLE;
CLASSES PERSON TIME DRUG;
  MODEL YPLUS = PERSON(DRUG) TIME DRUG*TIME/
    NOINT SOLUTION SS1;
WEIGHT W;

```

CONTRASTS

The GLM-GLS parameter estimates of the subplot treatment, usually time, are linear functions of the GLM parameter estimates of the subplot treatment. The relationship is given by:

$$T^+ = DT.$$

Conversely, T is a function of T^+ . The relationship is given by:

$$T = D^{-1} T^+.$$

Given the relationship that exists between T and T^+ , any contrasts of the form CT , where C is a matrix of desired contrasts, can be estimated alternatively as

$$CD^{-1} T^+.$$

RECOMMENDED STRATEGY

The following strategy is recommended for experimental designs with autoregressive subplot errors:

1. Fit the model by OLS. From this analysis:
 - test the whole plot,
 - estimate ϕ with the residuals obtained from OLS (Albohali, 1983), and
 - consult Table 3 above to see if a GLS or GLM-GLS analysis needs to be performed on the subplot. If the adjustment factor, γ , is close to one, a standard split-plot analysis will be appropriate for both the whole plot and the subplot.
2. Do the necessary transformations. Keep in mind that with a GLM-GLS analysis one need not produce the design matrix, X .

III. EXAMPLE

BOX'S ADJUSTED DEGREES OF FREEDOM

To illustrate Box's adjustment to the subplot degrees of freedom, consider an experiment in which the effect of a dose of a drug on the growth of rats is monitored. Presented below in Table 4 are growth measurements of 50 rats, where 10 rats were randomly assigned to each of the five doses. The weights were monitored and collected for 11 weeks.

Table - 4

Body Weights of Rats

DOSE	RAT	WEEK										
		1	2	3	4	5	6	7	8	9	10	11
0	1	54	60	63	74	77	89	93	100	108	114	124
	2	69	75	81	90	97	120	114	119	126	138	143
	3	77	81	87	94	101	110	117	124	134	141	151
	4	64	69	77	83	88	96	104	109	120	123	131
	5	51	58	62	71	74	81	88	93	99	103	113
	6	64	71	77	89	90	100	106	114	122	134	139
	7	80	91	97	101	111	119	129	131	137	147	154
	8	79	85	89	99	104	105	116	121	132	139	147
	9	77	82	88	92	101	109	119	127	135	144	158
	10	79	84	91	98	107	114	119	131	137	146	155

		Week										
DOSE	RAT	1	2	3	4	5	6	7	8	9	10	11
.5	1	62	71	75	79	87	91	100	105	111	121	124
	2	68	73	81	89	94	101	110	114	123	132	139
	3	94	102	109	110	128	133	147	151	153	171	184
	4	81	90	95	102	109	120	128	137	141	154	160
	5	64	69	72	76	84	89	97	103	108	114	124
	6	67	74	81	81	84	95	100	109	119	128	130
	7	73	80	86	89	97	101	110	116	117	135	141
	8	71	74	82	84	93	97	102	113	119	124	131
	9	69	74	79	89	94	100	107	113	124	134	139
	10	60	62	67	74	78	85	92	103	112	121	130
1	1	59	63	66	75	80	87	99	104	110	115	124
	2	56	66	70	81	77	88	96	100	113	120	130
	3	71	77	84	80	97	106	111	109	128	133	140
	4	59	64	69	76	85	88	96	104	110	119	126
	5	65	70	73	77	85	92	96	101	111	118	121
	6	61	69	77	81	89	92	107	111	118	127	132
	7	80	86	95	99	106	113	127	131	142	150	160
	8	74	80	84	90	99	101	108	117	126	133	140
	9	71	79	88	90	98	102	116	121	127	139	142
	10	69	75	80	86	96	97	104	113	122	129	138
4	1	64	71	79	82	85	94	103	110	113	122	125
	2	53	57	61	72	74	76	81	91	99	100	105
	3	64	69	76	85	89	96	104	108	116	120	128

		Week										
DOSE	RAT	1	2	3	4	5	6	7	8	9	10	11
	4	68	69	78	82	91	97	104	108	115	122	132
	5	69	74	80	85	90	99	104	114	123	129	133
	6	85	91	98	100	105	104	118	121	130	141	141
	7	75	82	85	92	99	107	112	125	130	137	146
	8	57	61	65	68	77	81	87	91	95	101	107
	9	69	72	77	80	84	91	96	103	109	116	125
	10	66	68	76	81	88	95	103	106	112	119	130
8	1	57	64	70	76	80	90	93	99	105	113	118
	2	62	67	74	83	87	93	104	108	114	124	129
	3	60	68	73	80	83	94	101	106	112	122	131
	4	64	66	76	81	91	100	102	111	120	128	136
	5	57	60	67	73	67	64	75	85	89	98	105
	6	78	83	89	99	105	113	117	128	132	139	149
	7	81	81	92	100	108	119	120	133	138	149	157
	8	46	47	51	55	63	65	68	74	78	85	90
	9	69	72	74	76	77	82	82	90	95	101	103
	10	67	77	83	83	92	99	104	108	114	118	129

Source: Analysis of Messy Data Volume I: Designed Experiments by Milliken & Johnson (1984: p. 371-372)

The appropriate model for this analysis is:

$$Y_{ijk} = \mu + \alpha_i + \pi_{ij} + T_k + (\alpha T)_{ik} + \epsilon_{ijk},$$

$$i = 1, \dots, 5, \quad j = 1, \dots, 10, \quad k = 1, \dots, 10.$$

Where μ is the overall mean.

α_i represents dose effect,

π_{ij} is the animal error term,

T_j represents time effect,

$(\alpha T)_{ik}$ represent the dose by time interaction,

ϵ_{ijk} is the time interval error term and is considered to be autoregressive.

A split-plot analysis of the data yields the following Analysis of Variance table:

Table 5		ANOVA		
Source of Var.	df	SS	MS	F
Dose	4	10295.72	2573.93	1.53
Error(Rat)	45	75668.30	1681.52	
Week	10	243381.13	24338.13	1783.51
Dose * Week	40	1517.88	37.95	2.78
Error(Week)	450	6140.80	13.65	

The adjustment in this case is made to the degrees of freedom associated with week, the interaction between dose and week, and the subplot error. Each is multiplied by the constant:

$$\gamma = 467.6421/811.1840 = 0.57649 (\hat{\phi} = 0.60).$$

The adjusted degrees of freedom are given below:

Table - 6 Actual and Adjusted Degrees of Freedom

Source of Var.	Actual df	Adjusted df
Dose	4	4*
Error(Rat)	45	45*
Week	10	5.7649
Dose * Week	40	23.0596
Error(Week)	450	259.4205

* Unchanged (between rat comparison).

The F-statistics in the split-plot Anova table should be compared to the critical regions corresponding to the F distribution with the appropriately adjusted degrees of freedom.

The appropriate critical regions to use to test for a significant week effect and to test for a significant dose by week interaction, respectively, are:

$$F_{\alpha = 0.05} = 2.15917$$

$$(5.77, 259.42)$$

$$F_{\alpha = 0.05} = 1.57059$$

$$(23.06, 259.42)$$

For the rat data above, the adjustment factor (γ)

is substantially below an adjustment factor of one (see table 3). Hence a GLS analysis or a variant thereof is in order. The magnitude of the adjustment factor needed before a GLS analysis be done is a subjective question.

GLS

To illustrate GLS, let us take another look at the rat data presented above. The first step involves running a split-plot analysis of the data using a procedure such as SAS's PROC GLM. The F-statistic from this analysis will be correct for the whole plot treatment, dose (see split-plot table above).

The next step requires the creation of a split-plot design matrix (X). The design matrix for the rat data contains 121 variables, five for dose, fifty for whole plot error - which is a nested error structure consisting of rat nested within dose, eleven for week, and fifty-five for the dose by week interaction. A no intercept model was used.

The design matrix (X) and response vector (y) is premultiplied by the matrix $Z = I \otimes A$, where,

$$A = \begin{bmatrix} 0.8 & 0 & 0 & \dots & 0 & 0 & 0 \\ -0.6 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -0.6 & 1 & \dots & 0 & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & \dots & -0.6 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -0.6 & 1 \end{bmatrix}$$

11x11

Note: The transformation matrix, Z, is a 550 x 550 matrix.

The final step involves running OLS on the 121 transformed variables. The F-statistic from this analysis will be correct for the subplot treatment, week, and for the interaction term, dose * week. However, the whole plot part of the analysis is erroneous. Combining the whole plot part of the split-plot analysis with the subplot part of the GLS analysis produces the following ANOVA table:

Table 7		ANOVA		
Source of Var.	df	SS	MS	F
Dose	4	10295.72	2573.93	1.53
Error(Rat)	45	75668.30	1681.52	
Week	10	78626.58	7862.66	846.04
Dose * Week	40	787.88	19.70	2.12
Error(Week)	450	4182.06	9.29	

The conclusions supported by the GLS analysis in this case do not differ from the conclusions supported by a standard split-plot analysis. In either case there is a significant dose by week interaction at a 0.05 level of significance. However, it is worth noting that the week sums of squares dropped substantially from 1783.51 to 846.04 .

To do this analysis a 550 x 121 design matrix had to be produced. A laborious task to say the least. The final proposed method eliminates the need to produce the design matrix and is therefore advocated. SAS's generalized linear model (GLM), for example, will produce the design matrix.

GLM-GLS

To illustrate GLM-GLS, let us review the rat data once more. As with GLS, the first step involves running a split-plot analysis of the data using a procedure such as SAS's PROC GLM. The F-statistic from this analysis will be correct for the whole plot treatment, dose (see table 5 above).

The next step involves converting y_{ijk} to y_{ijk}^+ by a direct transformation.

The final step involves running a weighted least squares (WLS) on the 121 variables of the split-plot

design matrix (X) created, for example, by SAS's GLM procedure and on the variable of the transformed response vector, y^+ . The matrix of weights to apply to WLS is of the form $W = I \otimes G$, where,

$$G = \begin{bmatrix} 4 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix} \quad 11 \times 11$$

The F-statistics from this analysis will be correct for the subplot treatment, week, and for the interaction term, dose * week. However, the whole plot part of the analysis is erroneous. Combining the whole plot part of the split-plot analysis with the subplot part of the GLS analysis produces the following ANOVA table:

Table 8		ANOVA		
Source of Var.	df	SS	MS	F
Dose	4	10295.72	2573.93	1.53
Error(Rat)	45	75668.30	1681.52	
Week	10	78626.58	7862.66	846.04
Dose * Week	40	787.88	19.70	2.12
Error(Week)	450	4182.06	9.29	

Note: there is no difference in the ANOVA table presented from the GLS analysis versus the table presented from the GLM-GLS analysis. However, the parameters (β) that are estimated by the respective methods differ drastically.

CONTRASTS

SAS uses set-to-zero restrictions to compute the GLM and GLM-GLS parameter estimates. Therefore, the GLS parameter estimates outputted by SAS is of the following form:

$$CT = \begin{bmatrix} T_1 - T_t \\ T_2 - T_t \\ T_3 - T_t \\ \vdots \\ T_{t-1} - T_t \\ 0 \end{bmatrix}$$

To illustrate the procedure for obtaining GLS contrasts from the GLM-GLS parameter estimates, let us take one last look at the rat data. Table 9 below shows

the parameter estimates for both GLS and GLM-GLS for the rat data. These are estimates for the subplot treatment, week.

Table 9: GLS and GLM-GLS Parameter Estimates

GLS	GLM-GLS
$T_1 - T_{11} = -66.20$	$T_1^+ - T_{11}^+ = -30.788$
$T_2 - T_{11} = -60.72$	$T_2^+ - T_{11}^+ = -25.308$
$T_3 - T_{11} = -54.80$	$T_3^+ - T_{11}^+ = -22.676$
$T_4 - T_{11} = -49.14$	$T_4^+ - T_{11}^+ = -20.568$
$T_5 - T_{11} = -42.88$	$T_5^+ - T_{11}^+ = -17.704$
$T_6 - T_{11} = -36.18$	$T_6^+ - T_{11}^+ = -14.760$
$T_7 - T_{11} = -29.06$	$T_7^+ - T_{11}^+ = -11.660$
$T_8 - T_{11} = -22.52$	$T_8^+ - T_{11}^+ = -9.392$

GLS and GLM-GLS Parameter Estimates

GLS	GLM-GLS
$T_{9}^{\hat{}} - T_{11}^{\hat{}} = - 15.32$	$T_{9}^{+\hat{}} - T_{11}^{+\hat{}} = - 6.116$
$T_{10}^{\hat{}} - T_{11}^{\hat{}} = - 7.18$	$T_{10}^{+\hat{}} - T_{11}^{+\hat{}} = - 2.296$
0.00	0.00

The desired contrasts, $CT^{\hat{}}$, for the subplot treatment, week, can be obtained by premultiplying the GLM-GLS estimate, $T^{\hat{}}$, by CD^{-1} . For example, the contrast $T_{1}^{\hat{}} - T_{2}^{\hat{}}$ can be obtained by multiplying:

$$CT^{\hat{}} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} - 66.21 \\ - 60.72 \\ - 54.80 \\ - 49.14 \\ - 42.88 \\ - 36.18 \\ - 29.06 \\ - 22.52 \\ - 15.32 \\ - 7.18 \\ 0 \end{bmatrix}$$

$$= CD^{-1} \hat{T}^+ = -5.49 \quad \text{where,}$$

$$D^{-1} = \begin{bmatrix} 2.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1.5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.9 & 0.6 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.54 & 0.36 & 0.6 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.32 & 0.22 & 0.36 & 0.6 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.19 & 0.13 & 0.22 & 0.36 & 0.6 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0.12 & 0.08 & 0.13 & 0.22 & 0.36 & 0.6 & 1 & 0 & 0 & 0 & 0 \\ 0.07 & 0.05 & 0.08 & 0.13 & 0.22 & 0.36 & 0.6 & 1 & 0 & 0 & 0 \\ 0.04 & 0.03 & 0.05 & 0.08 & 0.13 & 0.22 & 0.36 & 0.6 & 1 & 0 & 0 \\ 0.03 & 0.02 & 0.03 & 0.05 & 0.08 & 0.13 & 0.22 & 0.36 & 0.6 & 1 & 0 \\ 0.02 & 0.01 & 0.02 & 0.03 & 0.05 & 0.08 & 0.13 & 0.22 & 0.36 & 0.6 & 1 \end{bmatrix}$$

$$\hat{T}^+ =$$

$$\begin{bmatrix} -30.788 \\ -25.308 \\ -22.676 \\ -20.568 \\ -17.704 \\ -14.760 \\ -11.660 \\ -9.392 \\ -6.116 \\ -2.296 \\ 0 \end{bmatrix}$$

Note: \hat{CT} may not exactly equal $CD^{-1} \hat{T}^+$ due to rounding error.

In summary, three techniques were presented for analyzing the subplot of a repeated measures experimental design with an autoregressive error structure. The first technique, proposed by Box (1954), involved adjusting the degrees of freedom associated with the subplot treatment, whole plot treatment by subplot treatment interaction, and the subplot error.

The second technique, GLS, involved transforming the design matrix and the response vector. The transformed response vector is then regressed on the transformed design matrix.

The third and final technique, GLM-GLS, requires regressing a transformed response vector on the original design matrix.

In deciding on which technique to use, it is helpful to note that GLS provides tests that are more powerful than Box's proposed technique. However, the design matrix needs to be constructed to do a GLS analysis. This is time consuming. However, one can avoid the laborious task of creating the design matrix and still retain the nice properties of GLS by performing a GLM-GLS analysis.

BIBLIOGRAPHY

- Albohali, M. N. 1983. A time series approach to the analysis of a repeated measures designs. Ph.D dissertation, Kansas State University.
- Arnold, S. 1981. The theory of linear models and multivariate analysis. New York: John Wiley & Sons.
- Box, G.E.P. 1954. Some theorems on quadratic forms applied in the study of analysis of variance problems. *Annals of Mathematical Statistics* 25: 290-302.
- Huynh, H., and Feldt, L. S. 1970. Conditions under which mean square ratios in repeated measures designs have exact F-distributions. *Journal of the American Statistical Association* 65: 1582-89.
- Johnston, J. 1972. Econometric methods. 2nd ed. New York: McGraw - Hill.
- Milliken, G.; and Johnson, D. 1984. Analysis of messy data volume 1: designed experiments. New York: Van Nostrand Rein Hold Company.
- Wallenstein, S.; and Fleiss, J. 1979. Repeated measurements analysis of variance when the correlations have a certain pattern. *Psychometrika* 44: No. 2: 229-232.

TECHNIQUES FOR ANALYZING REPEATED MEASURES EXPERIMENTAL
DESIGNS WITH AUTOREGRESSIVE SUBPLOT ERRORS

by

DANIEL P. JOHNSON

B.S., Kansas State University, 1982
MBA, Kansas State University, 1984

AN ABSTRACT OF A REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

STATISTICS

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1989

Several statistical computer packages exist that have a generalized linear model procedure with a weighted least squares option. Such packages can be used to perform a modified generalized least squares subplot analysis of a repeated measures experimental design with autoregressive subplot disturbances. One need only regress an appropriately transformed response vector on the repeated measures design matrix using the weighted least squares option. The weights need to be specified.

The modified generalized least squares estimators retain all the properties of the standard generalized least squares estimators. However, a standard generalized least squares analysis requires the specification of a design matrix. Computer packages with the generalized linear model-weighted least squares option produce the design matrix at a considerable time savings to the analyst.

Contrasts can be constructed for the modified generalized least squares technique by taking advantage of the linear relationship that exist between the modified generalized least squares estimators and the standard generalized least squares estimators.