11-1-2007

A Comparison of Procedures for the Analysis of Multivariate Repeated Measurements

Lisa M. Lix
University of Manitoba, lisa.lix@usask.ca

Anita M. Lloyd
University of Alberta

Regular Articles
A Comparison of Procedures for the Analysis of Multivariate Repeated Measurements

Lisa M. Lix
University of Manitoba

Anita M. Lloyd
University of Alberta

Three procedures for analyzing within-subjects effects in multivariate repeated measures designs are compared when group covariances are heterogeneous: the multiple regression model (MRM) with a structured covariance, Johansen's (1980) procedure, and the multivariate Brown and Forsythe (1974) procedure. A preliminary likelihood ratio test of a Kronecker product covariance structure is sensitive to sample size and derivational assumption violations. Error rates of the procedures are generally well-controlled except when the distribution is skewed. The MRM procedure displayed few power advantages over the other procedures.

Key words: doubly multivariate data; robustness; Kronecker product; assumption violations; general linear model; Kenward-Roger approximation.

Introduction

Multivariate repeated measures data arise when measurements are obtained from study participants on P dependent variables at each of T occasions. The choice of a procedure for testing multivariate within-subjects main and interaction effects depends, in part, on the assumptions made about $cov(\mathbf{Y}_{ij}) = \mathbf{\Omega}_{ij}$, where $\mathbf{Y}_{ij} = [Y_{ij11}\ Y_{ij12}\ \ldots\ Y_{ij1P}\ \ldots\ Y_{ijTP}]^{T}$, the vector of measurements for the ith subject $(i = 1\ ,\ldots,\ n_j)$ in the jth group $(j = 1\ ,\ldots,\ J)$, and ^{T} is the transpose operator.

Two procedures for testing within-subjects effects in multivariate repeated measures data are the doubly multivariate model (DMM) and multivariate mixed model (MMM) procedures (Boik, 1988, 1991; Crawford & Johnson, 1994; Naik & Rao, 2001; Thomas,

1983), which are extensions of multivariate analysis of variance (MANOVA) and analysis of variance (ANOVA) for repeated measurements, respectively, to the case of two or more dependent variables. Both procedures define the multivariate mean response as a function of the measurement occasions and the between-subjects (i.e., grouping) factor levels. The DMM makes no assumptions about the structure of $\mathbf{\Omega}_{ij} = \mathbf{\Omega}$, where $\mathbf{\Omega}$ is the pooled covariance, other than it is positive definite. The MMM assumes a multivariate spherical (M-spherical) structure for $\mathbf{\Omega}$, in which pairs of repeated measurements exhibit a common variance across the dependent variables. The MMM is more powerful than the DMM for testing multivariate within-subjects effects if the assumption of M-sphericity is satisfied and the data follow a multivariate normal distribution (Boik, 1988; 1991). However if M-sphericity is not a tenable assumption, Type I error rates of the MMM tests may be substantially inflated; the magnitude of the deviation from the nominal level of significance, α, will increase as the degree of departure from an M-spherical structure increases (Boik, 1988). Accordingly, Boik recommended the DMM over the MMM

Lisa Lix is an Associate Professor, Department of Community Health Sciences, and Director, Biostatistical Consulting Unit, University of Manitoba. Email her at: lisa_lix@cpe.umanitoba.ca. Anita Lloyd is a biostatistician.

provided that total sample size is sufficiently large.

The multiple regression model (MRM) with a structured covariance is one alternative to the ANOVA and MANOVA procedures for the analysis of repeated measures data. Jennrich and Schluchter (1986) and Zimmerman and Nunez-Anton (2001) (see also Fitzmaurice, Laird, & Ware, 2004; Littell, Pendergast, & Natarajan, 2000) have described this procedure for the case of a single dependent variable. The MRM procedure allows the researcher to specify a parametric form for both the mean and covariance of the repeated measurements. When a parsimonious structure is specified for the covariance, the MRM procedure should result in a more powerful test of within-subjects effects than the MANOVA procedure because there are fewer parameters to estimate and greater denominator degrees of freedom. However if the covariance structure is incorrectly specified, tests of within-subjects effects may be biased (Guerin & Stroup, 2000).

In multivariate repeated measures data, a parsimonious structure for $\mathbf{\Omega}$ in the MRM procedure is a Kronecker product structure (Galecki, 1994), that is, $\mathbf{\Omega} = \mathbf{\Sigma}_T \otimes \mathbf{\Sigma}_P$ (Chaganty & Naik, 2002; Mitchell, Genton, & Gumpertz, 2006; Naik & Rao, 2001), where $\mathbf{\Sigma}_T$ is the covariance of the repeated measurements, $\mathbf{\Sigma}_P$ is the covariance of the dependent variables and $\otimes$ is the Kronecker product operator. This structure is also referred to as a separable covariance structure (Mitchell et al., 2006). A likelihood ratio test (LRT) of a Kronecker product structure has been proposed for choosing between the MRM and DMM procedures (Naik & Rao, 2001; Roy & Khattree, 2005; Timm, 2002). If the MRM procedure is adopted, selection of the best-fitting model from the set of candidate models with different covariance structures is accomplished either by assessing the statistical significance of a LRT for two nested models, or by comparing the values of a penalized log likelihood-based information criterion, such as the Akaike criterion (Akaike, 1974), for these candidate models (Fitzmaurice et al., 2004; Littell et al., 2000).

There has been only limited investigation of the MRM with a structured covariance when $P \geq 2$ (e.g., Chinchilli & Carter, 1984; Reinsel, 1982), and not for the case when group covariances are heterogeneous. Previous research on methods for the analysis of multivariate repeated measures data when covariances are heterogeneous has focused on the properties of DMM tests of multivariate within-subjects main and interaction effects and robust alternatives to DMM tests. Robust alternatives include Johansen's (1980) approximate degrees of freedom (ADF) multivariate test and a multivariate extension of the Brown and Forsythe (1974) ADF test (Keselman & Lix, 1997; Lix, Algina, & Keselman, 2003; Vallejo, Fidalgo, & Fernandez, 2001). These ADF tests have been implemented with least-squares estimators when the data follow a multivariate normal distribution, as well as with trimmed estimators (i.e., trimmed means and Winsorized covariances) for the case when the data follow a multivariate heavy-tailed or skewed distribution. While the Johansen and Brown and Forsythe ADF tests are insensitive to covariance heterogeneity, they assume an unstructured form for $\mathbf{\Omega}$ and should, in theory, be less powerful than the MRM, provided that all procedures can control the rate of Type I errors to α. However, Johansen's procedure is also known to produce inflated error rates when sample size is small (Keselman & Lix, 1997). Thus, at present it is not clear which procedure(s) should be recommended for analyzing multivariate repeated measurements data when covariances are heterogeneous.

The objectives of this article are to: (a) examine the Type I error performance of a LRT of a Kronecker product structure for $\mathbf{\Omega}$, and (b) compare the Type I error and power of the MRM, Johansen (1980), and multivariate Brown and Forsythe (1974) procedures for testing multivariate within-subjects main and interaction effects when covariances are heterogeneous. As part of the second objective, several information criteria are investigated for selecting the best-fitting model from amongst candidate models with different covariance structures for the MRM in the presence of covariance heterogeneity.

Description of Procedures

Notation

The procedures are described for a multivariate design with T repeated measurements, P dependent variables, and J levels of a between-subjects factor. Consider the general linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{Y} = \left[\mathbf{Y}_{11}^{\mathrm{T}} \ldots \mathbf{Y}_{n_J J}^{\mathrm{T}}\right]^{\mathrm{T}}$ is the N x TP matrix of responses with $\mathbf{Y}_{ij}$ as defined previously, $\mathbf{X}$ is an N x q design matrix, $\boldsymbol{\beta}$ is the q x TP matrix of fixed effect parameters to be estimated, and $\boldsymbol{\varepsilon}$ is the N x TP matrix of residual errors. The rows of $\boldsymbol{\varepsilon}$ are assumed to be independent and to follow a normal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Omega}_{ij} = \boldsymbol{\Omega}_j$, the covariance for the jth group.

Likelihood Ratio Test of a Kronecker Product Structure for $\boldsymbol{\Omega}$

There are $T(T + 1)/2 + P(P + 1)/2$ parameters to estimate when $\boldsymbol{\Omega}$ has a Kronecker product structure, compared to $TP(TP + 1)/2$ parameters to estimate when $\boldsymbol{\Omega}$ is unstructured. For example, with $T = 4$ and $P = 2$, there are a total of $10 + 3 = 13$ parameters to be estimated in the former case, compared to $8(9)/2 = 36$ parameters to be estimated in the latter case.

Tests of different forms of a Kronecker product structure have been described in the literature (e.g., Boik, 1991; Naik & Rao, 2001). Mitchell et al. (2006) derived a LRT for a general Kronecker product structure, which makes no assumptions about the form of either $\boldsymbol{\Sigma}_T$ or $\boldsymbol{\Sigma}_P$. To test the null hypothesis $H_{01} : \boldsymbol{\Omega} = \boldsymbol{\Sigma}_T \otimes \boldsymbol{\Sigma}_P$ against the alternative $H_{A1} : \boldsymbol{\Omega} \neq \boldsymbol{\Sigma}_T \otimes \boldsymbol{\Sigma}_P$ the test statistic is

$$\lambda = \frac{\left|\hat{\boldsymbol{\Sigma}}_T\right|^{NP/2} \left|\hat{\boldsymbol{\Sigma}}_P\right|^{NT/2}}{\left|\mathbf{E}\right|^{N/2}} \tag{2}$$

where

$$\mathbf{E} = \frac{1}{N} \mathbf{Y}^{\mathrm{T}} \left[\mathbf{I}_N - \mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X}\right]\mathbf{Y} \tag{3}$$

and $\mathbf{I}_N$ is an identity matrix of dimension N. The statistic $-2\ln\lambda$ asymptotically follows a χ_f^2 distribution, where $f = PT(PT + 1)/2 - P(P + 1)/2 - T(T + 1)/2$, under the assumptions of a multivariate normal distribution of responses and covariance homogeneity. Maximum likelihood (ML) estimates of $\boldsymbol{\Sigma}_P$ and $\boldsymbol{\Sigma}_T$ can be obtained via algorithms proposed by Boik (1991), Dutilleul (1999), or Mardia & Goodall (1993).

The Multiple Regression Model, Johansen (1980), and Multivariate Brown and Forsythe (1974) Procedures

For multivariate repeated measures data, the MRM procedure with a structured covariance is defined as

$$\mathrm{vec}(\mathbf{Y}^{\mathrm{T}}) = (\mathbf{X} \otimes \mathbf{I}_{TP})\mathrm{vec}(\boldsymbol{\beta}^{\mathrm{T}}) + \mathrm{vec}(\boldsymbol{\varepsilon}^{\mathrm{T}}) \tag{4}$$

where $\mathrm{vec}(\cdot)$ is the vec operator, and

$$\begin{aligned} \mathrm{cov}(\mathrm{vec}(\mathbf{Y}^{\mathrm{T}})) &= \mathbf{I}_N \otimes \boldsymbol{\Omega} \\ &= \mathbf{V}, \end{aligned} \tag{5}$$

when homogeneity of group covariances is assumed. If $\mathbf{V}$ is known, the least-squares estimator of $\boldsymbol{\beta}$ is

$$\begin{aligned} \mathrm{vec}(\hat{\boldsymbol{\beta}}^{\mathrm{T}}) = \\ \left[(\mathbf{X} \otimes \mathbf{I}_{TP})^{\mathrm{T}} \mathbf{V}^{-1}(\mathbf{X} \otimes \mathbf{I}_{TP})\right]^{-1}(\mathbf{X} \otimes \mathbf{I}_{TP})^{\mathrm{T}} \mathbf{V}^{-1}\mathrm{vec}(\mathbf{Y}^{\mathrm{T}}) \end{aligned} \tag{6}$$

When the data follow a multivariate normal distribution, $\hat{\boldsymbol{\beta}}$ also follows a multivariate normal distribution with mean $\boldsymbol{\beta}$ and covariance

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = \left[(\mathbf{X} \otimes \mathbf{I}_{TP})^{\mathrm{T}} \mathbf{V}^{-1}(\mathbf{X} \otimes \mathbf{I}_{TP})\right]^{-1}. \tag{7}$$

When $\mathbf{V}$ is unknown, covariance parameters (i.e., $\hat{\mathbf{V}}$) are estimated using ML or restricted maximum likelihood (REML). Then $\hat{\mathbf{V}}$ is substituted for $\mathbf{V}$ in equations 6 and 7.

When this substitution is made, $\hat{\boldsymbol{\beta}}$ is an unbiased estimate of $\boldsymbol{\beta}$ under asymptotic theory (Fitzmaurice et al., 2004). Kackar and Harville (1981) have also shown that, provided the population distribution is symmetric, $\hat{\boldsymbol{\beta}}$ is an asymptotically unbiased estimate of $\boldsymbol{\beta}$. However, for small sample sizes, the accuracy of the approximation may be poor, particularly when T and/or P are large. When covariances are not assumed homogeneous across groups, separate parameters are estimated for each level of the between-subjects factor, such that $\hat{\boldsymbol{\Omega}}_j$ denotes the estimated covariance for the jth group.

Hypotheses about multivariate within-subjects main and interaction effects are of the form H_{02}: $\mathbf{L}\text{vec}(\boldsymbol{\beta}^T) = \mathbf{0}$ where $\mathbf{L}$ of dimension r x qTP contains weights that define one or more linear contrasts among the elements of $\boldsymbol{\beta}$. To test hypotheses about the individual β_m ($m = 1$,…, qTP), a number of different test statistics may be adopted (Fouladi & Shieh, 2004), including an approximate t statistic. If the rank of $\mathbf{L}$ is greater than one, an F statistic to test the null hypothesis is,

$$F = \frac{1}{\text{rank}(\mathbf{L})}\left(\mathbf{L}\text{vec}(\hat{\boldsymbol{\beta}}^T)\right)^T\left[\mathbf{L}\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}})\mathbf{L}^T\right]^{-1}\left(\mathbf{L}\text{vec}(\hat{\boldsymbol{\beta}})\right) \quad (8)$$

where $\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}})$ estimates $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$. This statistic is compared to the critical value, $F[(1 - \alpha); v_1, v_2]$, where $v_1 = \text{rank}(\mathbf{L})$ and v_2 is approximated from the data (Kenward & Roger, 1997; Satterthwaite; 1946). Guerin and Stroup (2000) recommend adopting the Kenward-Roger approximation because the former can result in inflated error rates for small sample sizes.

As noted previously, for the MRM procedure applied to $P \geq 2$ dependent variables, $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_T \otimes \boldsymbol{\Sigma}_P$ defines a parsimonious structure for $\boldsymbol{\Omega}$. The matrix $\boldsymbol{\Sigma}_P$ is typically assumed to have an unstructured form with $P(P + 1)/2$ unique elements $\sigma_{ll'}$ ($l, l' = 1$,…, P), while the parameters in $\boldsymbol{\Sigma}_T$ are assumed to be a function of the measurement occasions u and v ($u, v = 1$,…, T) and the levels of one or more between-subjects factor(s) when covariance homogeneity is not assumed (Zimmerman & Nunez-Anton,

2001). Examples of possible structures for $\boldsymbol{\Sigma}_T$ have been enumerated in several sources, including Fitzmaurice et al. (2004), Littell et al. (2000), and Littell, Stroup, and Freund (2002). For example, the compound symmetric (CS) structure has the following variance and correlation specification: $\sigma_{kk} = \sigma^2$ and $\rho_{kk'} = \rho$ ($k = 1$,…, T ; $k \neq k'$). This parsimonious structure assumes constant variances and correlations across measurement occasions. Multivariate compound symmetry is a more restrictive assumption than that of M-sphericity (Crawford & Johnson, 1994). The variance and correlation specifications for the first order autoregressive (AR-1) structure is $\sigma_{kk} = \sigma^2$, for $k = 1$,…, T and $\rho_{kk'} = \rho^{k'-k}$ for $k' > k$ and $k' = 2, \ldots, T$. The unstructured (UN) covariance has $T(T + 1)/2$ unique elements denoted $\sigma_{kk'}$ ($k, k' = 1$,…, T).

Several information criteria for assessing model fit have been proposed, including the Akaike (AIC; Akaike, 1974), Bayesian-Schwarz (BIC: Schwarz, 1978), finite population-corrected AIC (CAIC; Bozdogan, 1987), and Hannan & Quinn (HQIC; Hannan & Quin, 1979) criteria. These are respectively defined as

$$\begin{aligned} \text{AIC} &= -2l + 2d, \\ \text{BIC} &= -2l + d\log(N^*), \\ \text{CAIC} &= -2l + d\log(N^*+1), \\ \text{HQIC} &= -2l + 2d\log\log(N^*), \end{aligned} \quad (9)$$

where l is the logarithm of the ML or REML function for the specified model, d is the number of covariance parameters to be estimated, $N^* = N$ for ML estimation and $N^* = N - q$ for REML estimation. Amongst candidate models with different covariance structures, the best-fitting model is the one with the smallest value for the selected information criterion. The criteria will not always select the same model. For example, the BIC penalizes the model more severely for the number of parameters than does the AIC, and therefore tends to choose less complex models than the AIC.

Computational formulae for the Johansen (1980) and multivariate Brown and Forsythe (1974) ADF procedures have been enumerated in a number of sources, and

therefore have not been repeated in this manuscript (e.g., Keselman & Lix, 1997; Lix et al., 2003). Vallejo, Fidalgo, and Fernandez (2001) extended the Brown and Forsythe (1974) procedure to the case of multivariate repeated measurements, but a recent modification proposed by Vallejo and Ato (2006), to address the conservative Type I error properties of this procedure for testing within-subjects effects, was adopted in the current study.

Methodology

Methods for Investigating the Properties of a Likelihood Ratio Test of a Kronecker Product Covariance Structure

Monte Carlo techniques were used to investigate the Type I error properties of the LRT for testing the null hypothesis that $\mathbf{\Omega}$ has a Kronecker product structure. The data were generated for a multivariate design containing a single between-subjects factor with two levels and a single within-subjects factor. The parameters manipulated in the study were: (a) total sample size, (b) number of repeated measurements, (c) number of dependent variables, (d) degree of covariance heterogeneity, and (e) degree of departure from a multivariate normal distribution. The value of the LRT statistic does not depend on the form of either $\mathbf{\Sigma}_T$ or $\mathbf{\Sigma}_P$ (Mitchell et al., 2003), so the data were generated from a population in which both $\mathbf{\Sigma}_T$ or $\mathbf{\Sigma}_P$ had CS covariance structures.

Dutilleul's (1999) algorithm (see also Dutilleul & Pinel-Alloul, 1996) was used to obtain ML estimates of $\mathbf{\Sigma}_T$ and $\mathbf{\Sigma}_P$. This algorithm finds solutions to the following system of equations,

$$\hat{\mathbf{\Sigma}}_P = \frac{1}{TN}\sum_{i=1}^{n_j}\sum_{j=1}^{J}(\mathbf{W}_{ij}-\bar{\mathbf{W}})\hat{\mathbf{\Sigma}}_T^{-1}(\mathbf{W}_{ij}-\bar{\mathbf{W}})^T \tag{10}$$

and

$$\hat{\mathbf{\Sigma}}_T = \frac{1}{PN}\sum_{i=1}^{n_j}\sum_{j=1}^{J}(\mathbf{W}_{ij}-\bar{\mathbf{W}})\hat{\mathbf{\Sigma}}_P^{-1}(\mathbf{W}_{ij}-\bar{\mathbf{W}})^T \tag{11}$$

where W_{ij} is the P x T matrix obtained by re-shaping Y_{ij}, and $\bar{\mathbf{W}}$ is the matrix of means obtained by averaging across all such observation matrices.

Three levels of total sample size were investigated: $N = 40$, 60, and 100. The number of repeated measurements was set at $T = 4$ and 6, while the number of dependent variables was set at $P = 2, 3, 4$, and 6. These conditions reflect the range of simulation parameters that have been investigated in previous research on methods for the analysis of multivariate repeated measures data (Boik, 1991; Lix et al., 2003; Vallejo et al., 2001).

The tests were investigated for homogeneous covariances (i.e., $\mathbf{\Omega}_1 = \mathbf{\Omega}_2$), as well as for two cases of covariance heterogeneity: $\Omega_1 = 5\Omega_2$, and $\Omega_1 = 9\Omega_2$.

Multivariate data were generated from both normal and non-normal distributions. Pseudorandom observation vectors Y_{ij} from a multivariate normal distribution with mean vector $\mathbf{\beta}_j$ and covariance matrix $\mathbf{\Omega}_j$ were obtained by the following method. A column vector of standard normal deviates (i.e., d_{ij}) was transformed to a vector of multivariate observations via $\mathbf{Y}_{ij} = \mathbf{\beta}_j + \mathbf{R}d_{ij}$ where R is an upper triangular matrix of dimension TP with the property $\mathbf{R}^T\mathbf{R} = \mathbf{\Omega}_j$.

Two multivariate non-normal distributions were investigated (Lix, Keselman, & Hinds, 2005). The first was a symmetric distribution with a mild degree of heavy-tailedness and skewness (γ_1) and kurtosis (γ_2) values of 0 and 1.7 respectively, (the normal distribution has $\gamma_1 = 0$ and $\gamma_2 = 0$) while the second distribution had $\gamma_1 = 2.0$ and $\gamma_2 = 6.0$, which are equivalent to the shape parameters of an exponential distribution. A vector of constants $\mathbf{w} = [a\ b\ c\ d]^T$ was obtained using Fleishman's (1978) method, to provide the desired degree of skewness and kurtosis for each of these distributions. An intermediate covariance matrix (i.e., ζ) was then computed so that $\mathbf{Y}_{ij}$ would have the desired final covariance structure. Elements of this intermediate matrix were computed using Vale and Maurelli's (1983) method. The vector of univariate deviates was transformed to a vector of multivariate

normal deviates via, $\mathbf{Z}(\zeta)_{ij} = \boldsymbol{\beta}_j + \mathbf{R}_\zeta \mathbf{d}_{ij}$, where $\mathbf{Z}(\zeta)_{ij}$ is the vector of transformed variates, and $\mathbf{R}_\zeta$ is an upper triangular matrix such that $\mathbf{R}_\zeta^T \mathbf{R}_\zeta = \zeta$. Next, each element of $\mathbf{Y}_{ij}$ was obtained by computing the zero through third powers of the corresponding elements of $\mathbf{Z}(\zeta)_{ij}$, so that $\mathbf{Z}(\zeta)_{ijm} = [1 \; \mathbf{Z}(\zeta)_{ijm} \; \mathbf{Z}(\zeta)_{ijm}^2 \; \mathbf{Z}(\zeta)_{ijm}^3]$ ($m = 1, \ldots, TP$) represents the vector of powers. From this, $Y_{ijkl} = \mathbf{Z}(\zeta)_{im}\mathbf{w}$ ($k = 1, \ldots, T; l = 1, \ldots, P$).

Five thousand replications of each combination of conditions were performed using $\alpha = .05$ as the criterion for assessing statistical significance. The simulation program was written in SAS/IML (SAS Institute Inc., 2004a). Descriptive techniques were used to summarize the Type I error rates.

Methods for Investigating the Properties of the Multiple Regression Model, Johansen, and Multivariate Brown and Forsythe Procedures

Monte Carlo techniques were used to evaluate the Type I error and power of the MRM, Johansen (1980), and multivariate Brown and Forsythe (1974) procedures for testing multivariate within-subjects main and interaction effects as well as to investigate the properties of the information criteria for assessing model fit. The data were generated for a design with a single between-subjects factor with two levels, a single within-subjects factor with four levels, and three dependent variables. The simulation parameters were: (a) total sample size, (b) degree of group size imbalance, (c) degree of covariance heterogeneity, (d) pairing of group sizes and covariances, (e) degree of departure from a multivariate normal distribution, (f) structure of $\boldsymbol{\Omega}$, and (g) configuration of the population means.

The analyses were conducted for total sample size conditions of 60 and 100. Group sizes were both equal and unequal. Table 1 lists the group sizes that were investigated for each value of N. The tests were investigated for homogeneous covariances (i.e., $\boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_2$), as well as for two cases of covariance heterogeneity: $\boldsymbol{\Omega}_1 = 5\boldsymbol{\Omega}_2$, and $\boldsymbol{\Omega}_1 = 9\boldsymbol{\Omega}_2$.

Positive and negative pairings of group sizes and group covariances were investigated.

A positive pairing refers to the case in which the largest n_j is associated with the covariance matrix containing the largest element values; a negative pairing refers to the case in which the largest n_j is associated with the covariance matrix with the smallest element values.

Multivariate normal and non-normal data were generated using the method described in the previous section. Two multivariate non-normal distributions were investigated. The first was a symmetric heavy-tailed distribution with shape parameters equivalent to those of a double exponential distribution (i.e., $\gamma_1 = 0; \gamma_2 = 3.0$.) The second was a skewed distribution that represented an extreme degree of departure from multivariate normality, with shape parameters equivalent to those of a multivariate lognormal distribution (i.e., $\gamma_1 = 6.2; \gamma_2 = 110.9$).

In this phase of the study, $\boldsymbol{\Sigma}_T$ had either a CS or AR-1 structure. Both structures had $\sigma^2 = 1$ and $\rho = 0.5$. $\boldsymbol{\Sigma}_P$ had a CS structure with $\sigma^2 = 1$ and $\rho = 0.4$.

All procedures were investigated when the configuration of population means was null and non-null. For the non-null case, the following configuration of means was investigated: $\boldsymbol{\beta}_1 = [.25\ 0\ 0\ -.25\ 0\ 0\ 0\ 0\ .25\ 0\ 0\ -.25]^T$ and $\boldsymbol{\beta}_2 = [.25\ 0\ 0\ .25\ 0\ 0\ 0\ 0\ .25\ 0\ 0\ .25]^T$.

Multivariate datasets were generated using a program written in SAS/IML (SAS Institute Inc., 2004a). A SAS/IML program was also written to analyze each dataset with the Johansen (1980) and multivariate Brown and Forsythe (1974) procedures. A PROC MIXED (SAS Institute Inc., 2004b) macro was written to analyze each dataset using the MRM procedure, and output the F statistics, p-values, and degrees of freedom for tests of the within-subjects main and interaction effects, as well as the numeric values for each of the four investigated information criterion. Only one thousand replications were performed for each combination of conditions because of the lengthy execution time required for PROC MIXED. The syntax to implement the MRM procedure is reported in Appendix A; it is the same as that reported by Timm (2002) and Thiebaut, Jacquim-Gadda, Chene, Leport and Commenges (2002). All parameters were estimated using REML. In the PROC MIXED macro, each dataset was analyzed using

Table 1. Group Sizes Investigated in the Simulation Study

N	n_1, n_2	Δn_j
60	30, 30	0.0
	24, 36	0.2
	20, 40	0.3
100	50, 50	0.0
	40, 60	0.2
	35, 65	0.3

Note. N = total sample size. Δn_j is the coefficient of variation for group sizes (see Lix & Keselman, 1997).

Three different models; each model had the same fixed effects, but a different Kronecker product covariance structure in the REPEATED statement (see Appendix A). The best-fitting model, among the three models that were fit to the data, was the model that resulted in the lowest numeric value of an information criterion. The percentage of times that the best-fitting model had the same covariance structure as the population covariance structure was recorded for each criterion; this is denoted as the percentage of correct model selection.

Type I error and power rates were calculated for the MRM, Johansen (1980), and multivariate Brown and Forsythe (1974) procedures. The percent bias in Type I error rates, $B = 100(\hat{\alpha} - \alpha)/\alpha$, where $\hat{\alpha}$ is the empirical error rate for a test, was also calculated. Type I error rates, percentages of bias, and power rates were summarized descriptively. For each of the three procedures, regression analyses were used to model the effect of the simulation parameters on the Type I error rates. For the MRM, the regression model had a random simulation effect, because there were repeated measurements on each model covariance structure. For each procedure, separate models were defined for equal and unequal group size cases for within-subjects main and interaction effects, respectively. All models included main effects as well as two-way interactions among the simulation parameters.

Results

Likelihood Ratio Test of a Kronecker Product Covariance Structure

Figure 1 contains the empirical Type I error rates for the LRT for each of the investigated values of P when $T = 4$ and $\Omega_1 = \Omega_2$. The sensitivity of the LRT to total sample size is apparent. When the data were multivariate normal and $P = 2$, the empirical error rate was 0.11 for $N = 40$, but quickly converged to 0.05 when $N = 100$. As the dimension of Ω increased from $TP = 8$ to 36, error rates also increased across the range of values of N. For example, with $P = 3$ and $N = 100$, the error rate was 0.09. As the dimension of the data increased, Type I error rates of the LRT rapidly approached the upper bound of 1.00.

Error rates were also highly sensitive to the presence of multivariate non-normality. For example, when the data were sampled from a heavy-tailed distribution when $P = 2$, the error rate was 0.18 for $N = 40$ and 0.12 for $N = 100$. Error rates were also more inflated for the skewed distribution than the heavy-tailed distribution. For the latter distribution, error rates attained or approached the upper bound of 1.00 for many of the investigated conditions.

When the number of repeated measurements increased to six and $\Omega_1 = \Omega_2$, the same pattern of results was observed, although error rates were even more inflated than for $T = 4$. For example, for $N = 100$, the empirical error
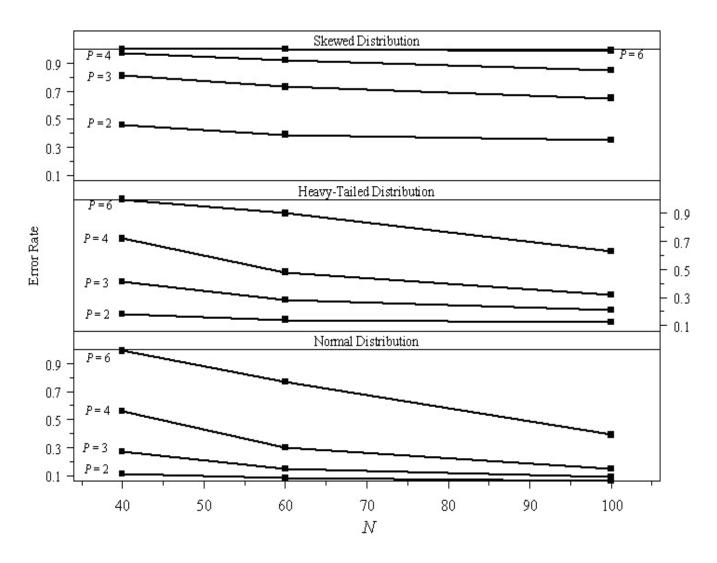
Figure 1. Type I Error Rates for the Likelihood Ratio Test (LRT) of a Kronecker Product Structure, $\Omega_1 = \Omega_2$, $T = 4$

Note. Heavy-tailed distribution has $\gamma_1 = 0$ and $\gamma_2 = 1.7$. Skewed distribution has $\gamma_1 = 2.0$ and $\gamma_2 = 6.0$.

rate was 0.07 for $P = 2$ but increased to 0.89 for $P = 6$ when the data were sampled from a multivariate normal distribution. When the data were from multivariate non-normal distributions, error rates always exceeded α and frequently attained the upper bound of 1.00.

When covariances were heterogeneous, the error rates of the LRT were always inflated regardless of the total sample size, dimension of Ω, or degree of departure from a multivariate normal distribution. For example, when the data

were multivariate normal with $T = 4$ and $P = 2$, the error rate was 0.52 for $N = 40$ and 0.37 for $N = 100$.

Information Criteria for Assessing Model Fit

The results for the four information criteria are reported in Table 2; they have been averaged over the conditions of total sample size, degree of covariance heterogeneity, and degree of group size imbalance because a similar

pattern of results was observed for these conditions.

When the data followed a multivariate normal distribution and $\Omega_1 = \Omega_2$, the differences among the four criteria were small and on average, the correct (i.e., population) covariance structure was selected in as many as 96.4% of the models. When covariances were heterogeneous and the distribution was multivariate normal, the percentages of correct model selection were lower when Σ_T had a CS structure (approximately 65% for all criteria) than when Σ_T had an AR-1 structure (approximately 90% for all criteria).

When $\Omega_1 = \Omega_2$ and the data were sampled from non-normal distributions, the average percentages of correct model selection were lower than when the data were sampled from a multivariate normal distribution. For the heavy-tailed distribution when Σ_T had a CS structure, correct model selection was observed, on average, for 64.8% of the models for the AIC and 75.2% of the models for the BIC. When Σ_T had an AR-1 structure, the average percentages for the AIC and BIC were 77.4% and 90.3%, respectively, for this distribution. For the skewed distribution, the percentages were substantially lower; for example, when Σ_T had a CS structure, correct model selection was observed for only 22.2% of the models for the AIC. Moreover, the AIC and HQIC were more sensitive to multivariate non-normality than the BIC and CAIC. The latter two procedures always resulted in higher average percentages of correct models than the former two procedures when the data were obtained from heavy-tailed or skewed distributions.

When covariances were heterogeneous and the data were obtained from multivariate non-normal distributions, a similar pattern of results was observed. All four information criteria produced similar average percentages of correct model selection regardless of whether group sizes were equal or unequal. The values obtained when Σ_T had an AR-1 structure were higher than those obtained when Σ_T had a CS structure. The AIC and HQIC were more sensitive to multivariate non-normality than the BIC and CAIC.

Tests of Multivariate Within-Subjects Main and Interaction Effects

Type I error rates

Percentages of bias in Type I error rates are reported in Tables 3 and 4; the results are averaged across the two values of total sample size because a similar pattern of results was observed. For both the main and interaction effect tests when group sizes were equal (Table 3), the average bias was small for Johansen's (1980) procedure as well as for the multivariate Brown and Forsythe (1974) procedure when the data were from symmetric distributions. For Johansen's test, average bias ranged from 10.0% to 19.5% for the within-subjects main effect and from -1.5% to 22.0% for the within-subjects interaction effect. For the Brown and Forsythe (1974) test, average bias ranged from -2.0% to 13.0% for the within-subjects main effect and from -11.0% to 7.5% for the within-subjects interaction. Overall, the Type I error rates of the MRM procedure were more biased than error rates of either ADF procedure when the distribution was symmetric. Average bias ranged from -28.0% to 29.0% for the within-subjects main effect and from -15.0% to 23.0% for the within-subjects interaction effect. However, the results for symmetric distributions reveal that the magnitude of bias varied across the three model covariance structures. As expected, when Σ_T had a CS structure, there was generally less bias in the error rates when a model with a CS or UN covariance structure for the repeated measurements was adopted than when a model with an AR-1 structure was adopted. Similarly, when Σ_T had an AR-1 structure, there was less bias when a model with either an AR-1 or UN structure was selected than when a model with a CS structure was selected. However, for the former case, the AR-1 structure tended to result in negatively biased error rates for the within-subjects main effect and positively biased results for the within-subjects interaction effect, while in the latter case the CS structure results in

Table 2. Average Percentages of Correct Model Selection for Four Information Criteria

		AIC	HQIC	BIC	CAIC
		Σ_T = CS			
Nor	$=n_j/=\Omega_j$	86.4	87.5	87.5	87.5
	$=n_j/\neq\Omega_j$	63.0	63.2	63.3	63.3
	+ pair	65.9	66.3	66.3	66.3
	- pair	64.9	65.3	65.3	65.3
HT	$=n_j/=\Omega_j$	64.8	73.7	75.2	75.3
	$=n_j/\neq\Omega_j$	59.2	63.8	64.5	64.5
	+ pair	58.9	64.5	65.2	65.2
	- pair	58.9	64.5	65.3	65.3
SK	$=n_j/=\Omega_j$	22.2	24.3	30.6	38.4
	$=n_j/\neq\Omega_j$	20.6	22.4	29.2	36.5
	+ pair	21.6	23.4	30.4	37.9
	- pair	21.5	23.6	31.0	38.8
		Σ_T = AR-1			
Nor	$=n_j/=\Omega_j$	95.6	96.4	96.4	96.4
	$=n_j/\neq\Omega_j$	89.9	90.4	90.4	90.4
	+ pair	92.1	92.8	92.8	92.8
	- pair	90.5	91.1	91.1	91.1
HT	$=n_j/=\Omega_j$	77.4	88.5	90.3	90.4
	$=n_j/\neq\Omega_j$	75.5	83.5	84.6	84.6
	+ pair	75.4	85.8	87.3	87.4
	- pair	75.6	83.0	84.1	84.2
SK	$=n_j/=\Omega_j$	20.4	23.1	30.5	38.4
	$=n_j/\neq\Omega_j$	19.8	22.0	30.0	37.8
	+ pair	20.0	22.9	31.9	40.1
	- pair	20.8	23.7	31.9	40.7

Note. CS = compound symmetric; AR-1 = first-order autoregressive. Nor = multivariate normal distribution with $\gamma_1 = 0$ and $\gamma_2 = 0$; HT = multivariate heavy-tailed distribution with $\gamma_1 = 0$ and $\gamma_2 = 3$; SK= multivariate skewed distribution with $\gamma_1 = 6.2$ and $\gamma_2 = 110.9$. $=n_j/=\Omega_j$ = equal group sizes and equal group covariances; $=n_j/\neq\Omega_j$ = equal group sizes and unequal covariances; + pair = positive pairing of group sizes and covariances; - pair = negative pairing of group sizes and covariances.

positively biased error rates for both main and interaction tests. For the multivariate skewed distribution, Type I error rates were almost always negatively biased for all procedures when group sizes were equal. This finding was observed for both main and interaction effects. For example, for the multivariate within-subjects interaction, average bias for Johansen's (1980) procedure ranged from -23.0% to -48.0% and for the multivariate Brown and Forsythe (1974) procedure average bias ranged from -31.0% to

-51.0%. For the MRM procedure, average bias was similar, and ranged from -13.5% to -45.0% across the three models for the interaction effect.

Percentages of bias for unequal group sizes are reported in Table 4; separate summaries are given for positive and negative pairings of group sizes and covariances and the results are averaged across conditions of total sample size, degree of covariance heterogeneity, and degree of group size imbalance because

Table 4. Average Percentages of Bias in Type I Error Rates for Multivariate Within-Subjects Effects when Group Sizes are Unequal

| | | Σ_T = CS | | | | | Σ_T = AR-1 | | | | |
		MRM CS	MRM AR-1	MRM UN	J	BF	MRM CS	MRM AR-1	MRM UN	J	BF
						Main Effect					
Nor	$\Omega_1 = \Omega_2$	7.0	-4.0	4.0	18.0	13.0	24.0	7.0	16.0	6.0	4.0
	$\Omega_1 \neq \Omega_2$	9.7	-23.2	6.6	11.5	4.0	29.0	4.0	17.0	19.5	7.5
HT	$\Omega_1 = \Omega_2$	-5.0	-28.0	-22.0	11.0	7.0	27.0	10.0	-1.0	10.0	9.0
	$\Omega_1 \neq \Omega_2$	-11.9	-23.4	-11.4	10.0	-1.5	19.0	-4.0	8.0	11.0	-2.0
SK	$\Omega_1 = \Omega_2$	-17.0	-47.0	-24.0	-39.0	-39.0	-16.0	-29.0	-32.0	-44.0	-46.0
	$\Omega_1 \neq \Omega_2$	-36.0	-54.0	-40.5	-22.0	-29.0	4.5	-12.0	-31.5	-26.5	-35.5
						Interaction Effect					
Nor	$\Omega_1 = \Omega_2$	10.0	13.0	-11.0	5.0	1.0	13.0	-4.0	-14.0	10.0	6.0
	$\Omega_1 \neq \Omega_2$	-3.6	9.6	-3.1	9.0	-1.5	21.5	2.0	10.0	22.0	7.5
HT	$\Omega_1 = \Omega_2$	-15.0	12.0	-10.0	8.0	4.0	8.0	-11.0	-4.0	3.0	0.0
	$\Omega_1 \neq \Omega_2$	-7.6	-7.7	-9.1	-1.5	-11.0	23.0	-0.5	8.5	8.5	-2.5
SK	$\Omega_1 = \Omega_2$	-23.0	-14.0	-36.0	-46.0	-50.0	-19.0	-44.0	-45.0	-48.0	-51.0
	$\Omega_1 \neq \Omega_2$	-29.5	-31.5	-39.0	-24.5	-33.0	-13.5	-20.5	-32.0	-23.0	-31.0

Note. CS = compound symmetric; AR-1 = first-order autoregressive; UN = unstructured. MRM = multiple regression model; J = Johansen's (1980) procedure; BF = multivariate Brown and Forsythe (1974) procedure; Nor = multivariate normal distribution with $\gamma_1 = 0$ and $\gamma_2 = 0$; HT = multivariate heavy-tailed distribution with $\gamma_1 = 0$ and $\gamma_2 = 3$; SK = multivariate skewed distribution with $\gamma_1 = 6.2$ and $\gamma_2 = 110.9$. + pair = positive pairing of group sizes and covariances; - pair = negative pairing of group sizes and covariances.

similar patterns of results were observed. For Johansen's (1980) procedure, bias was almost always positive for the multivariate normal and heavy-tailed distributions for both within-subjects main and interaction effects; it was highest for negative pairings, where average bias ranged from 31.8% to 51.5%. For the multivariate Brown and Forsythe (1974) procedure, bias was small for the normal and heavy-tailed distributions and ranged from -9.0% to 8.5% across the main and interaction effects. For the MRM procedure, average bias ranged from -29.0% to 23.5% for the main effect and from -14.4% to 28.3% for the interaction effect for symmetric distributions across the three model covariance structures. The same pattern of results was observed for unequal group sizes as was observed for equal group

sizes. When Σ_T = AR-1 and a model with a CS structure was adopted, positive bias was observed for both the multivariate normal and heavy-tailed distributions. However, when Σ_T = AR-1 and a model with either an AR-1 or UN covariance structure was adopted, error rates were less biased and ranged from -6.0% to 10.8% for the main effect and from 1.5% to 14.8% for the interaction effect. When Σ_T = CS and a model with an AR-1 structure was adopted, bias ranged from -29.0% to -19.9% for the main effect and from -5.9% to 7.1% for the interaction effect.

When the distribution was multivariate skewed and group sizes were unequal, error rates were almost always negatively biased for the three multivariate procedures. For example, for the MRM procedure, bias ranged from -50.8% to -5.6% for the main effect and from

-86.5% to -19.0% for the interaction effect. The range of values for average bias was similar for equal and unequal group sizes.

As noted previously, for both balanced and unbalanced designs, it was generally the case that less bias was observed in Type I error rates when the model covariance structure corresponded to the population covariance structure. For MRM tests of the multivariate within-subjects main effect, the average bias across equal group sizes for Σ_T = CS when the data were multivariate normal was 8.8% for the model with the CS covariance structure, -16.8% for the AR-1 model, and 5.8% for the UN structure. When Σ_T = AR-1 under a normal multivariate distribution, the average bias across equal group size conditions was 5.0% for the model with the AR-1 covariance structure, 27.3% for the model with the CS structure, and 16.7% for the model with the UN covariance structure. The average percentages of bias were similar for multivariate within-subjects main and interaction effects. As sample size increased from N = 60 to 100 when group sizes were equal, bias tended to decrease for both the normal and skewed distributions under either population covariance structure while bias tended to increase when data were from a heavy-tailed distribution. When group sizes were unequal and sample size increased from N = 60 to 100, bias tended to increase when data was from a normal distribution while bias tended to decrease when data was from a skewed distribution. For data from a heavy-tailed distribution, a trend in average bias values was not evident.

When group sizes were equal, the regression model for Johansen's (1980) procedure accounted for 87.7% of the variation in Type I error rates for the within-subjects main effect and 90.5% of the variation in error rates for the within-subjects interaction effect. For the multivariate Brown and Forsythe (1974) procedure, the model accounted for 87.5% of the variation in Type I error rates for the within-subjects main effect and 86.2% of the variation in Type I error rates for the interaction effect. For both procedures, the majority of this variation was attributed to the main effect of population distribution (i.e., between 85.5% and 99.0% of the total explained variation). For

Johansen's procedure for the test of the within-subjects interaction, total sample size and the two-way interaction of population distribution and degree of covariance heterogeneity accounted for 6.1% and 5.1% of the variation in error rates, respectively. Other main and two-way interaction effects in the models accounted for a small percentage of the explained variation.

For the MRM when group sizes were equal, the regression model that contained main effects and two-way interactions accounted for 92.8% of the variation in Type I error rates for the within-subjects main effect test and 91.5% of the variation in error rates for the within-subjects interaction effect test. For the main effect test, population distribution, population covariance structure for Σ_T, and model covariance structure respectively accounted for 39.1%, 20.1%, and 12.1% of the explained variation. None of the other model effects individually accounted for more than 5% of the variation. For the within-subjects interaction effect test, the type of population distribution, two-way interaction between total sample size and degree of covariance heterogeneity, two-way interaction between population covariance structure and model covariance structure, and two-way interaction of total sample size and type of population distribution accounted for 47.5%, 6.6% and 5.7%, and 5.4% of the explained variance, respectively. None of the other model effects accounted for more than 5% of the variation.

When group sizes were unequal, the regression analyses for Johansen's (1980) procedure revealed that the model containing main effects and two-way interactions accounted for 90.2% of the variation in Type I error rates for the within-subjects main effect and 89.3% of the variation for the within-subjects interaction. For both tests, the majority of the explained variation was due to the main effects of population distribution, total sample size, and pairing of group sizes and covariances, and to the two-way interaction of total sample size and pairing of group sizes and covariances for the within-subjects interaction effect tests. For the multivariate Brown and Forsythe (1974) procedure, the regression model accounted for 80.3% and 84.3% of the variation in Type I error rates for the multivariate within-subjects main

and interaction effects, respectively. Almost all of this explained variation (i.e., > 90%) was due to the main effect of population distribution, although the two-way interaction of degree of covariance heterogeneity and degree of group size imbalance also accounted for slightly more than 5% of the explained variation.

For the MRM procedure, the regression model accounted for 91.5% of the variation in Type I error rates for the test of the within-subjects main effect and 93.9% of the variation in error rates for the within-subjects interaction effect. The model effects that accounted for the most explained variation for the within-subjects main effect were population type (34.1%), population covariance structure (21.2%), and model covariance structure (10.8%). For the within-subjects interaction, variables that accounted for most of the explained variance were population type (55.2%) and population covariance structure (6.8%).

Power

Percentages of power are reported in Table 5. They are summarized separately for equal and unequal group sizes for multivariate normal and heavy-tailed distributions. Power results are not reported for the skewed distribution because the three procedures could not control the Type I error rate for this condition. Only the results for N = 60 are reported because the pattern of results was similar for N = 100, and because power approached its upper bound for several of the simulation conditions for this latter value.

The Johansen (1980) and multivariate Brown and Forsythe (1974) procedures produced similar percentages of power for all conditions, except when group sizes and covariances were negatively paired. In this case, the Johansen (1980) procedure was more powerful than the multivariate Brown and Forsythe (1974) procedure. This is likely a result of the slightly liberal Type I error rates that were observed for the former procedure for negative pairing conditions. Average power was 56.9% and 56.2% for Johansen's procedure for the within-subjects main and interaction effects, respectively, and the corresponding values for the multivariate Brown and Forsythe (1974) procedure were 53.3% and 51.9%.

In general, the MRM procedure resulted in lower power than either the Johansen (1980) or multivariate Brown and Forsythe (1974) procedures for the within-subjects main effect; the average differences ranged between 5% and 20% for most of the investigated conditions although the difference was occasionally greater than this. For the interaction effect, the MRM procedure was often more powerful than either of the ADF tests, although the differences were never more than 10%. However, Johansen's procedure for testing the interaction effect was, on average, more powerful than the MRM procedure when group sizes and covariances were negatively paired.

When Σ_T = CS, the MRM procedure with either a CS or UN model covariance structure resulted in substantially higher power than the MRM with an AR-1 model structure. For example, when the data were normally distributed and group sizes were equal, the average power was 41.1% and 40.2%, respectively for the CS and UN model covariance structures, and only 26.5% for the AR-1 structure. This large difference in power was observed for both equal and unequal group sizes regardless of the shape of the population distribution. In contrast, when Σ_T = AR-1, the MRM procedure resulted in similar percentages of power for the CS, UN, and AR-1 structures. For example, when the data were from a heavy-tailed distribution and group sizes and covariances were negatively paired, the average power was 33.1%, 32.7%, and 33.2% for the multivariate interaction effect under CS, AR-1 and UN structures respectively.

Conclusion

Multivariate repeated measurements arise in the social, behavioral, and health sciences when researchers collect data on multiple psychological or physiological characteristics of study participants over time or across multiple experimental conditions. Global tests of hypotheses for multivariate within-subjects main or interaction effects take account of the correlation that exists among the repeated measurements and dependent variables. These tests may be conducted within the context of the

Table 5. Average Percentages of Power for Multivariate Within-Subjects Effects when Group Sizes are Equal and Unequal, $N = 60$

| | | $\Sigma_T = CS$ | | | | | $\Sigma_T = AR\text{-}1$ | | | | |
		MRM CS	MRM AR-1	MRM UN	J	BF	MRM CS	MRM AR-1	MRM UN	J	BF
						Multivariate Main Effect					
Nor	$= n_j$	41.1	26.5	40.2	50.2	47.3	30.9	30.7	30.1	38.2	35.5
	+ pair	47.3	31.0	46.1	57.5	55.8	35.5	34.2	33.5	43.9	42.2
	- pair	31.3	19.5	30.0	41.2	32.6	24.1	22.5	22.0	32.2	24.3
HT	$= n_j$	39.1	25.5	38.6	52.2	49.4	31.5	30.2	29.9	40.1	37.2
	+ pair	47.1	32.3	46.6	59.3	57.9	36.8	35.0	35.8	44.7	42.9
	- pair	30.8	19.7	30.6	42.5	34.5	24.9	23.1	23.3	33.2	25.8
						Multivariate Interaction Effect					
Nor	$= n_j$	55.5	42.2	53.6	50.7	48.3	42.5	43.9	41.5	37.23	34.8
	+ pair	61.8	48.7	60.2	58.5	57.2	50.4	51.2	48.0	43.8	42.5
	- pair	39.4	29.3	38.0	42.3	31.1	31.7	32.0	29.6	31.5	22.7
HT	$= n_j$	54.0	42.2	54.8	53.0	50.4	42.5	42.4	42.2	39.8	36.9
	+ pair	61.5	47.9	60.2	59.8	58.6	50.2	51.7	49.6	45.0	43.9
	- pair	40.7	31.5	38.8	43.9	33.8	33.1	32.7	33.2	33.5	24.6

Note. CS = compound symmetric; AR-1 = first-order autoregressive; UN = unstructured. MRM = multiple regression model; J = Johansen's(1980) procedure; BF = multivariate Brown and Forsythe (1974) procedure; Nor = multivariate normal distribution with $\gamma_1 = 0$ and $\gamma_2 = 0$; HT = multivariate heavy-tailed distribution with $\gamma_1 = 0$ and $\gamma_2 = 3$. $=n_j =$ equal group sizes; + pair = positive pairing of group sizes and covariances; -pair= negative pairing of group sizes and covariances.

general linear model using one of several procedures, including multivariate extensions of ANOVA and MANOVA. The choice of a procedure depends, in part, on the assumptions the researcher is willing to make about the covariance structure of the data.

A likelihood ratio test of a Kronecker product covariance structure, which might be used as a preliminary test to choose between the multiple regression model procedure that assumes a structured covariance and a procedure that makes no assumptions about the structure of the covariance matrix, requires a large sample size, relative to the dimension of the data, to control the rate of Type I errors to the nominal level of significance when the data are sampled from a multivariate normal distribution and covariances are homogeneous. When covariances are heterogeneous or the data are sampled from multivariate non-normal

distributions, this test can result in severely inflated Type I error rates. Thus, the likelihood ratio test is not useful as a preliminary test because it will almost always reject the null hypothesis of a Kronecker product structure in favor of the alternative hypothesis of an unstructured covariance.

Consistent with the results of previous research, the Johansen (1980) and multivariate Brown and Forsythe (1974) procedures provided good control of the Type I error rate across the majority of the investigated conditions when the data were sampled from multivariate normal and heavy-tailed distributions. When group sizes and covariance matrices were negatively paired, Johansen's (1980) test produced slightly inflated Type I error rates, although the magnitude of this positive bias decreased as total sample size increased. Both procedures produced conservative Type I error rates when the data

were sampled from a multivariate skewed distribution.

The multiple regression model procedure also provided good control of Type I error rates across the majority of the investigated conditions when the data were sampled from multivariate normal and heavy-tailed distributions. Like the other two procedures, it resulted in conservative error rates when the data were sampled from a multivariate skewed distribution. As expected, Type I error rates deviated less from the nominal level of significance when the selected model covariance structure corresponded to the population covariance structure, or when an unstructured covariance was selected.

For tests of the within-subjects main effect, the Johansen (1980) and multivariate Brown and Forsythe (1974) procedures were more powerful than the multiple regression model procedure regardless of which model covariance structure was selected for the latter. The differences in power were moderate to large. For the multiple regression model, power was higher when the selected model covariance structure was the same as the population covariance structure, than when an unstructured covariance model was selected, but the differences were small (i.e., less than five percentage points). For tests of the within-subjects interaction effect, the multiple regression model procedure was often more powerful than the other two procedures, but the differences were modest. Moreover, the multiple regression model procedure could also be less powerful than the Johansen or the multivariate Brown and Forsythe procedures if the covariance structure was misspecified.

For both within-subjects main and interaction effects, the magnitude of power differences among the multiple regression model procedures when the three model covariance structures for the repeated measurements were compared indicated that the power advantages gained by correctly specifying the covariance structure varies as a function of the form of the population covariance. When the population covariance structure of the repeated measurements was compound symmetric, the model with a compound symmetric structure was much more powerful than the model with an autoregressive structure. However, when the population covariance structure of the repeated measurements was autoregressive, there was only a small difference in power between the models with compound symmetric and autoregressive covariance structures.

Comparison of four penalized log-likelihood information criteria for assessing model fit revealed that all of the criterion performed better when group covariances were homogeneous than when they were heterogeneous, and when the data were sampled from symmetric distributions than when they were sampled from skewed distributions. The BIC and CAIC more often selected the model with the correct covariance structure than the AIC and HQIC.

Given these results, there appear to be limited benefits associated with adopting a multiple regression model procedure for testing multivariate within-subjects main and interaction effects in multivariate repeated measurement designs when covariances are heterogeneous and sample size is small or moderate. The Johansen (1980) and multivariate Brown and Forsythe (1974) approximate degrees of freedom procedures controlled the Type I error rates and were often more powerful than the multiple regression model procedures. Moreover, previous research has demonstrated that when the distribution is non-normal, robust versions of both procedures can control the rate of Type I errors to the nominal level of significance and can result in increased power to detect within-subjects effects.

While the results of this study suggest that either of the Johansen (1980) or multivariate Brown and Forsythe (1974) procedures could be recommended for analyzing within-subjects effects, researchers should be cautious in generalizing these results to all data-analytic conditions encountered in the analysis of multivariate repeated measures data. First, the properties of the three test procedures were only examined when the covariance of the repeated measurements and dependent variables had a Kronecker product structure. There have been no studies of the degree to which data encountered in the social, behavioral, and health sciences conform to a Kronecker product structure, nor of the magnitude of positive or negative bias in

error rates of the multiple regression model procedure when the data do not conform to a Kronecker product structure. Second, the test procedures were compared only for datasets with no missing observations. Unbiased estimates of regression parameters can be obtained for the multiple regression model procedure provided the observations are either missing completely at random or missing at random (Little & Rubin, 1987). The other two procedures investigated in this research do not accommodate study participants with missing observations; rather, incomplete cases are removed from the analysis, which can result in reduced power to detect within-subjects effects. Finally, power analyses were conducted only for a single effect size and a single configuration of the population means.

A number of opportunities for further research arise from this study. A likelihood ratio test of a Kronecker product structure that is less sensitive to sample size and is robust to multivariate non-normality and/or covariance heterogeneity requires investigation. Boik (1991) proposed an approximate likelihood ratio test of multivariate sphericity for small or moderate sample sizes under the assumption of a multivariate normal distribution. The approximation is based on the work of Box (1949), who proposed finding the moments of the likelihood statistics to derive the approximation. Mitchell et al. (2006) proposed a bootstrap likelihood ratio test that is less sensitive to sample size, but did not investigate its properties in the presence of covariance heterogeneity or multivariate non-normality. Zhu, Ng, and Jing (2002) compared likelihood ratio tests based on bootstrap and permutation re-sampling methods to test equality of covariances in the presence of multivariate non-normality. They found that the permutation test performed better than the bootstrap test.

Graphic techniques and statistical tests to assess model fit and select among candidate model covariance structures for the multiple regression model also need to be investigated and described for the case of multivariate repeated measures data. Littell et al. (2000) and Zimmerman and Nunez-Anton (2001) provide a thorough discussion of graphic and descriptive techniques for the case of a single dependent

variable, but they have not been extended to multivariate data. Techniques for assessing model fit in the presence of multivariate non-normality include bias-corrected versions of the AIC and empirical cross-validation techniques, have been proposed (e.g., Yanagihara, 2006), and could be investigated in the context of multivariate repeated measurements.

Parametric and non-parametric procedures for the analysis of multivariate repeated measurements with structured covariances which are robust to the presence of non-normal distributions and covariance heterogeneity require development and evaluation (Wang & Zhu, 2006; Reilly, 2005). Furthermore, comparisons among these procedures under the types of data-analytic conditions that may be encountered in practice are necessary to develop recommendations on choosing a statistical procedure.

Finally, models with structured covariances that bridge the gap between the restrictive compound symmetric Kronecker product structure and the less efficient unstructured Kronecker product structure require further development for the multivariate case. One flexible covariance structure described by Zimmerman and Nunez-Anton (2001) for models with a single dependent variable is the antedependence structure. It allows for a pattern of monotonically decreasing correlation among the repeated measurements, which is common in repeated measurements, as well as for non-constant variances of the repeated measurements. The authors describe software to implement a multiple regression model with the antedependence covariance structure for the case of a single dependent variable, and this could be investigated for possible extension to the multivariate case.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC -19, 716 -723.

Boik, R. J. (1988). The mixed model for multivariate repeated measures: Validity conditions and an approximate test. *Psychometrika*, *53*, 469-486.

Boik, R. J. (1991). Scheffe's mixed model for multivariate repeated measures: A relative efficiency evaluation. *Communications in Statistics –Theory and Methods*, *20*, 1233-1255.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, *36*, 317-346.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345 -370.

Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics,* 16, 129-132.

Chaganty, N. R., & Naik, D. N. (2002). Analysis of multivariate longitudinal data using quasi-least squares. *Journal of Statistical Planning and Inference*, *103*, 421-436.

Chinchilli, V. M., & Carter, W. H., Jr. (1984). A likelihood ratio test for a patterned covariance matrix in a multivariate growth curve model. *Biometrics*, *40*, 151-156.

Crawford, A. M. K., & Johnson, W. D. (1994). Statistical considerations for multivariate repeated measures with patterned covariance. *Proceedings of the 1993 Biopharmaceutical Section of the American Statistical Association,* 61-66.

Dutilleul, P. (1999). The MLE algorithm for the matrix normal distribution. *Journal of Statistical Computation and Simulation*, *64*, 105-123.

Dutilleul, P., & Pinel-Alloul, B. (1996). A doubly multivariate model for statistical analysis of spatio-temporal environmental data. *Environmetrics*, *7*, 551-565.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. Hoboken, NJ: Wiley.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, *43*, 521-532.

Fouladi, R. T., & Shieh, Y-Y. (2004). A comparison of two general approaches to mixed model longitudinal analyses under small sample size conditions. *Communications in Statistics - Simulation and Computation*, *33*, 807-824.

Galecki, A. T. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics –Theory and Methods*, *23*, 3105-3119.

Guerin, L., & Stroup, W. W. (2000). A simulation study to evaluate PROC MIXED analysis of repeated measures data. *Proceedings of the 12th Annual Conference on Applied Statistics in Agriculture*. Manhattan, KS: Kansas State University.

Hannan, E. J., & Quinn, B.G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society* – Series B, *41*, 190 -195.

Jennrich, R. I., & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, *42*, 805-820.

Johansen, S. (1980). The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika*, *67*, 85-92.

Kackar, R. N., & Harville, D. A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics – Theory and Methods*, *10*, 1249-1261.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*, 983-997.

Keselman, H. J., & Lix, L. M. (1997). Analysing multivariate repeated measures designs when covariance matrices are heterogeneous. *British Journal of Mathematical and Statistical Psychology*, *50*, 319-339.

Littell, R. C., Pendergast, J., & Natarajan, R. (2000). Tutorial in biostatistics: Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine*, *19*, 1793-1819.

Littell, R. C., Stroup, W. W., & Freund, R. J. (2002). *SAS for linear models, Fourth edition*. Cary, NC: SAS Institute, Inc.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Lix, L. M., Algina, J., & Keselman, H. J. (2003). Analyzing multivariate repeated measures designs: A comparison of two approximate degrees of freedom procedures. *Multivariate Behavioral Research*, *38*, 403-431.

Lix, L. M., Keselman, H. J., & Hinds, A. (2005). Robust tests for the multivariate Behrens-Fisher problem. *Computer Methods and Programs in Biomedicine*, *77*, 129-139.

Mardia, K. V., & Goodall, C. R. (1993). Spatial-temporal analysis of multivariate environmental monitoring data. In G. P. Patil, & C. R. Rao (Eds.), *Multivariate Environmental Statistics* (pp. 347-386). Amsterdam: Elsevier Science.

Mitchell, M. W., Genton, M. G., & Gumpertz, M. L. (2006). A likelihood ratio test for separability of covariances. *Journal of Multivariate Analysis, 97*, 1025-1043.

Naik, D. N., & Rao, S. S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *Journal of Applied Statistics*, *28*, 91-105.

Reilly, C. (2005). A nonparametric approach to the analysis of longitudinal data via a set of level crossing problems with application to the analysis of microarray time course experiments. *Biostatistics*, *6*, 271-278.

Reinsel, G. (1982). Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure. *Journal of the American Statistical Association*, *77*, 190-195.

Roy, A. & Khattree, R. (2005). On implementations of a test for Kronecker product covariance structure for the multivariate repeated measures data. *Statistical Methodology, 2*, 297-306.

SAS Institute, Inc. (2004a). *SAS/IML user's guide, version 9.1*. Cary, NC: Author.

SAS Institute, Inc. (2004b). *SAS/STAT user's guide, version 9.1*. Cary, NC: Author.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*, 110-114.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461 -464.

Thiebaut, R., Jacquim-Gadda, H., Chene, G., Leport, C., Commenges, D. (2002). Bivariate linear mixed models using SAS Proc MIXED. *Computer Methods and Programs in Biomedicine, 69,* 249-256.

Thomas, D. R. (1983). Univariate repeated measures techniques applied to multivariate data. *Psychometrika*, *48*, 451-464.

Timm, N. H. (2002). *Applied multivariate analysis*. New York: Springer.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions, *Psychometrika*, *48*, 465 - 471.

Vallejo, G., & Ato, M. (2006). Modified Brown-Forsythe test for testing interaction effects in split-plot designs. *Multivariate Behavioral Research, 41*, 549-578.

Vallejo, G., Fidalgo, A., & Fernandez, P. (2001). Effects of covariance heterogeneity on three procedures for analyzing multivariate repeated measures designs. *Multivariate Behavioral Research*, *36*, 1-27.

Wang, Y-G., Zhu, M. (2006). Rank-based regression for analysis of repeated measures. *Biometrika*, *93*, 459-464.

Yanagihara, H. (2006). Corrected version of AIC for selecting multivariate normal linear regression models in a general nonnormal case. *Journal of Multivariate Analysis, 97*, 1070-1089.

Zhu, L.-X., Ng, K. W., & Jing, P. (2002). Resampling methods for homogeneity tests of covariance matrices. *Statistica Sinica*, *12*, 769-783.

Zimmerman, D. L., & Nunez-Anton, V. (2001). Parametric modeling of growth curve data: An overview. *Test*, *10*, 1-73.

Appendix A. PROC MIXED Code to Implement the MRM

This is the syntax that was used to implement the MRM when the data are assumed to have a Kronecker product covariance structure, with an unstructured model for Σ_T as well as an unstructured model for Σ_P.

```
proc sort data=datafin;
by id;
run;

proc mixed data=datafin method=reml ic;
class group id dv rm;
model val=group dv rm dv*rm group*dv*rm /noint
ddfm=kenwardroger;
repeated dv rm / type=un@un subject=id(group) group=group;
run;
```

Where:
id = subject identification variable
group = variable to identify levels of between-subjects grouping factor
dv = variable to identify "levels" of the dependent variable factor
rm = variable to identify levels of within-subjects factor

      The data are assumed to be arranged in a "long" structure, with one value of the dependent variable *val* per line. Accordingly, each row of *datafin* contains a single observation and the corresponding values of the variables id, group, dv, and rm.

      A model with a compound symmetric form for Σ_T is obtained by specifying *type = cs@un* in the *repeated* statement. A model with a first-order autoregressive form for Σ_T is obtained by specifying *type = ar(1)@un* in the *repeated* statement. The only available model for Σ_P is unstructured (UN).