USING THE LINEAR MIXED MODEL TO ANALYZE NON-NORMAL DATA DISTRIBUTIONS IN LONGITUDINAL DESIGNS

Jaume Arnau[1]

Roser Bono[1]

María J. Blanca[2]

Rebecca Bendayan[2]

1. University of Barcelona, Spain

2. University of Malaga, Spain

Address for correspondence:

Roser Bono Cabré

Department of Methodology of the Behavioral Sciences

Faculty of Psychology, University of Barcelona

Passeig de la Vall d'Hebron, 171

08035 Barcelona (Spain)

Email: rbono@ub.edu

Abstract

Using a Monte Carlo simulation and the Kenward-Roger (KR) correction for degrees of freedom this paper analyzes the application of the linear mixed model (LMM) to a mixed repeated measures design. The LMM was first used to select the covariance structure with three types of data distribution: normal, exponential and log-normal. This showed that with homogeneous between-groups covariance, and when the distribution was normal, the covariance structure with the best fit was the unstructured population matrix. With heterogeneous between-groups covariance and when the pairing between covariance matrices and group sizes was null the best fit was shown by the between-subjects heterogeneous unstructured population matrix, this being the case for all the distributions analyzed. By contrast, with positive or negative pairing the within-subjects and between-subjects heterogeneous first-order autoregressive structure produced the best fit. In the second stage of the study, the robustness of the LMM was tested. This showed that the KR method provided adequate control of Type I error rates for the time effect with normally distributed data. However, as skewness increased, as occurs, for example, in the log-normal distribution, robustness was null, especially when the assumption of sphericity was violated. As regards the influence of kurtosis the analysis showed that the degree of robustness increased in line with the amount of kurtosis.

Keywords: Longitudinal data; Linear mixed model; Kenward-Roger method; Robustness; Non-normal distributions.

Over the last three decades the analysis of repeated measures data has been centered round linear mixed modeling. Laird and Ware (1982) established the basis of the LMM with the incorporation of the within-subjects error correlation. Their work was subsequently extended by Cnaan, Laird, and Slasor (1997) and Verbeke and Molenberghs (2000), who applied the LMM to longitudinal data. In contrast to analyses that are based on variances (ANOVA and MANOVA) the LMM models the structure of the covariance matrix. This enables a more efficient estimation of the fixed effects and, consequently, yields more robust statistical tests. However, when the covariance structures are not properly fitted and the sample sizes are small, the Type I error rate tends to rise (Wright & Wolfinger, 1996).

With respect to the covariance structure Keselman, Algina, Kowalchuck, and Wolfinger (1999) demonstrated that when the covariance matrix is not spherical the degrees of freedom associated with the conventional F test are too large. One way of controlling this bias is to apply the procedure developed by Kenward and Roger (1977) to correct the degrees of freedom. Studies by Kowalchuk, Keselman, Algina, and Wolfinger (2004) and Vallejo and Ato (2006) showed that with an adequate covariance structure the KR method is able, in many cases, to control Type I error rates with small sample sizes and when the assumption of sphericity is violated. Thus, by applying the KR correction one can obtain more efficient estimates of the fixed effects associated with repeated measures (Arnau, Bono, & Vallejo, 2009; Kowalchuk et al., 2004; Schaalje, McBride, & Fellingham, 2001).

Generally speaking, studies based on the LMM assume that the data are normal. However, in applied contexts, data distributions tend to depart from normality. For example, in a study of 440 distributions corresponding to a range of achievement and psychometric data, Micceri (1989) found that only 28.4% of the distributions were relatively symmetrical, while 40.7% were moderately asymmetrical, 19.5% were extremely asymmetrical and 11.4% showed exponential asymmetry. As regards kurtosis, only 15.2% of the distributions

approached normality, whereas 49.1% of the distributions could be considered to show extreme kurtosis. According to Micceri (1989) the most common distributions within the psychometric, educational and psychological contexts are moderately asymmetrical and show kurtosis that deviates considerably from normal. These results are supported by research showing that in the social and health sciences many variables follow a log-normal distribution (Limpert, Stahel, & Abbt, 2001). Examples in the field of medicine include the latency period of infectious diseases (Kondo, 1977), survival times for certain types of cancer (Claret et al., 2009; Qazi, DuMez, & Uckun, 2007), and the age of onset of Alzheimer's disease (Horner, 1987). In the social sciences an example would be the age at which people first get married (Preston, 1981), while in psychology the log-normal distribution provides a good fit to data on reaction times or response latency (Shang-wen & Ming-hua, 2010; Ulrich & Miller, 1993; Van der Linden, 2006), as well as to data on attentional skills (Brown, Weatherholt, & Burns, 2010). With survival data a good fit is provided not only by the log-normal distribution but also by the exponential distribution and its extensions (Weibull, gamma and Gompertz), as well as by the generalized gamma and log-logistic distributions (Lee & Wang, 2003).

Given that the log-normal distribution is common with applied data, particularly in the behavioral sciences (Micceri, 1989; Wilcox, 1994), some authors have based their simulation studies on this distribution (Keselman et al., 1999; Keselman, Algina, Wilcox, & Kowalchuk, 2000). In this context Sawilowsky and Blair (1992) obtained non-robust results from the *t* test when the distributions showed extreme skewness (for example, $\gamma_1 = 1.64$). Algina and Oshima (1995) found that tests for the equivalence of means are affected when the distribution is log-normal and the assumption of homogeneity is not fulfilled. As regards the effect of kurtosis, research has shown that the Type I error rate tends to decrease as kurtosis increases (Wilcox, 1993). In light of these findings the present study considers the log-normal distribution as well as distributions involving slight skewness combined with different levels of kurtosis.

The aim of this research was to analyze longitudinal data by means of the LMM, comparing normal and non-normal distribution. Specifically, the study seeks to examine the robustness of the LMM when applied to data whose distribution is closer to that found in real life, i.e. the log-normal distribution and other non-normal distributions with slight skewness and extreme kurtosis according to the values described by Lei and Lomax (2005). According to these authors, when the absolute values of skewness and kurtosis are less than 1.0 the distribution deviates slightly from normality, values between 1.0 and 2.3 correspond to moderate deviation, and values above 2.3 indicate extreme deviation from normality.

The linear mixed model and longitudinal data

Application of the LMM to longitudinal data involves adapting the hierarchical structure to a repeated measures design. Thus, the observations or repeated measures are located on the first level and the subjects on the second level (Cnaan et al., 1997; Van der Leeden, Vrijburg, & De Leeuw, 1996; Van der Leeden, 1998; Wu, Clopper, & Wooldridge, 1999). Both levels are clearly represented by the LMM. This kind of model, in which repeated measures are nested within individuals, constitutes an extension of the multilevel methodology that is widely used to analyze longitudinal data. An illustration of the LMM as applied to longitudinal data can be found in Arnau, Balluerka, Bono, and Gorostiaga (2010).

This section examines, firstly, the regression model on two levels and, secondly, the integration of the two levels within the LMM.

First level: within-subjects model

With longitudinal data, and in accordance with the model proposed, the direct observations or repeated measurements that make up the first level are nested within subjects

or the second level. Since growth curves are processes over time they can be modeled as polynomial functions. On a first level (within-subjects model) the equation can be fitted to a polynomial function of order p, expressed by

$$y_{ti} = \beta_{0i} + \beta_{1i}T_{ti} + \beta_{2i}T_{ti}^2 + ... + \beta_{pi}T_{ti}^p + e_{ti} \qquad e_{ti} \cong N(0, \sigma_e^2) \qquad (1)$$

where y_{ti} is the measurement of the dependent variable for subject i on occasion t, β_{pi}s are the coefficients of the polynomial function of degree p, which represent the growth trajectory of the subject, and e_{ti} is the random error. It is assumed that e_{ti} has a normal and independent distribution with mean zero and constant variance. The T_{ti}s are the explanatory variables of subject i on occasion t. Note that the parameters β_{pi} and the residual variance σ_e^2 are specific to subject i. Thus, with the level-1 model the repeated measurements of a growth process are described as a polynomial function. In this model both the number of observations and the interval between observations may vary between subjects.

Applying compact matrix notation the level-1 model for longitudinal data can be represented by:

$$\boldsymbol{y}_i = \mathbf{T}_i \boldsymbol{\beta}_i + \boldsymbol{e}_i \qquad e_i \cong N(0, \mathbf{R}) \qquad (2)$$

where $\boldsymbol{y}_i$ is the t x 1 repeated measure vector of the subject i, $\boldsymbol{\beta}_i$ is a p x 1 individual parameter vector that specifies the shape of the growth curve for subject i, and $\mathbf{T}_i$ is the t x p matrix of known variables and their polynomial transformations. Let us suppose that the response of the subject follows a quadratic function. $\mathbf{T}_i$ is then a t x 3 matrix, in which the first column is formed by ones, the second corresponds to the observation occasions, and the third to the occasions squared. The vector $\boldsymbol{e}_i$ is a t x 1 random error vector, and it is assumed that the errors are independent with normal multivariate distribution $N(0, \mathbf{R})$. Thus, $\mathbf{R} = $ var (e_i) is a positively defined covariance matrix. However, when the observations have a given

order or a specific structure it should be assumed that the between-error correlation is distinct from zero and that it varies systematically.

Second level: between-subjects model

In the longitudinal context the between-subjects model takes the individual growth parameters (β_{pi}s) to be random dependent variables which, in the simplest case, are expressed by the following equations,

$$\beta_{0i} = \gamma_{00} + u_{0i} \tag{3}$$

$$\beta_{1i} = \gamma_{10} + u_{1i} \tag{4}$$

$$\beta_{2i} = \gamma_{20} + u_{2i} \tag{5}$$

$$\dots$$

$$\beta_{pi} = \gamma_{p0} + u_{pi} \tag{6}$$

In equations (3), (4), (5) and (6) the between-subjects variation in the parameters that express the individual growth trajectories is modeled as a function of population averages and the deviations from these averages which subjects show.

If there is a predictive variable Z the level-2 model with $p+1$ parameters is expressed by

$$\beta_{pi} = \gamma_{p0} + \sum_{q=1}^{Q_p} \beta_{pq} Z_{qi} + u_{pi} \qquad u_{pi} \cong N(0, \mathbf{G}) \tag{7}$$

where γ_{p0} is the average intercept of the subjects, Z_{qi} represents a given level-2 predictive variable, β_{pq} expresses the effect of Z_{qi} on the growth variable p, and u_{pi} represents the random error term. The random effects $p+1$ associated with subject i are assumed to follow a normal multivariate distribution.

In matrix notation the level-2 model for longitudinal data would be

$$\boldsymbol{\beta}_i = \mathbf{Z}_i\boldsymbol{\gamma} + \boldsymbol{u}_i \qquad\qquad u_i \cong N\,(0,\,\mathbf{G}) \qquad\qquad (8)$$

where $\mathbf{Z}_i$ is the p x q between-subjects design matrix with known and fixed items. When only the random variation in individual growth parameters is modeled, $\mathbf{Z}_i$ takes the form of an identity matrix. More elaborate models can be formulated if $\mathbf{Z}_i$ contains dummy variables which codify subgroups of subjects or level-2 explanatory variables that enable the reasons for variability between growth parameters $\boldsymbol{\beta}_i$ to be examined. The covariables can be fixed or may vary over time. In equation (8), $\boldsymbol{\gamma}$ is a q x 1 fixed coefficient vector and $\boldsymbol{u}_i$ is a p x 1 random error vector with zero mean and variance $\mathbf{G}$.

Integrating the two levels in the linear mixed model

The LMM corresponds to the integration of the two levels described above. It can be used to analyze a wide variety of data structures which are commonly found in psychological, social, and health research and whose analysis with the classic linear model is problematic.

The complete LMM is obtained by inserting equation (8) into equation (2), as follows:

$$\mathbf{Y}_i = \mathbf{T}\mathbf{Z}_i\boldsymbol{\gamma} + \mathbf{T}\boldsymbol{u}_i + \boldsymbol{e}_i \qquad\qquad (9)$$

It is assumed that the elements of $\boldsymbol{e}_i$ are distributed independently and normally with constant variance, $\boldsymbol{e}_i \cong N\,(0,\,\mathbf{R})$, where $\mathbf{R} = \sigma_e^{2}\mathbf{I}$, that the level-2 random terms are distributed normally, i.e. $\boldsymbol{u}_i \cong N\,(0,\,\mathbf{G})$, and that the level-1 error terms ($\boldsymbol{e}_i$) are distributed independently of the level-2 ones ($\boldsymbol{u}_i$). In the complete model the term $\mathbf{T}\mathbf{Z}_i\boldsymbol{\gamma}$ is the fixed part and the term $\mathbf{T}\boldsymbol{u}_i + \boldsymbol{e}_i$ is the random part. The fixed effects define the expected values of the observations, while the random effects are variances and covariances. The covariance between the elements of $\mathbf{Y}_i$ or the longitudinal data consists of two parts, i.e. the between-subjects part and the within-subjects part, such that

$$\text{Var } (\mathbf{Y}_i) = \text{Var } (\mathbf{T}\boldsymbol{u}_i + \boldsymbol{e}_i) = \mathbf{TGT'} + \sigma_e^2 \mathbf{I} \qquad (10)$$

Note that in equation (10) the assumptions relating to errors, and especially to the term $\boldsymbol{e}_i$, lead to very simple covariance structures at the individual level of the model (constant and non-correlated errors across points in time). However, when there are many time points per subject the residual components often show a pattern of autocorrelation (Ware, 1985). With multilevel models a covariance matrix that reflects dependence between observations can be chosen. Thus, modeling the within-subjects covariance structure is particularly relevant, since the accuracy of the regression parameter estimates depends on the right choice (Littell, Pendergast, & Natarajan, 2000; Park & Lee, 2002).

One of the key advantages of the LMM is that it enables a choice to be made among the various covariance structures that could be used to model the data. This is of considerable importance because the better the fit of the covariance structure the greater the accuracy of the regression estimates.

KR method for correcting the degrees of freedom

In this study the degrees of freedom are corrected using the KR method, which yields more precise and efficient estimates of the fixed effects with small samples (Arnau et al., 2009; Kowalchuk et al., 2004; Schaalje et al., 2001).

If $\mathbf{C}$ is a contrast matrix with range q, the Wald F for the hypothesis H$_0$: $\mathbf{C}\boldsymbol{\beta} = 0$ is $F = $ W $/q$, where

$$W = (\mathbf{C}\hat{\boldsymbol{\beta}})'(\mathbf{C}(\mathbf{X'}\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{C'})^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}}) \qquad (11)$$

If we calculate a scale factor δ and an approximate value for the degrees of freedom v, then the F statistic for the KR method is given by

$$F^* = \quad \delta F_{KR} = \frac{\delta}{q} \quad (\mathbf{C}\hat{\beta})'(\mathbf{C}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\beta}) \qquad (12)$$

The moments of F^* are generated and matched to the moments of the distribution F so as to solve δ and v. Under the null hypothesis it is assumed that F^* is approximately distributed in the same way as F, with q degrees of freedom in the numerator and v degrees of freedom in the denominator. Therefore, it is necessary to calculate two values from the data: the degrees of freedom in the denominator v and a scale factor δ. Hence,

$$v = 4 + \frac{q+2}{qy-1} \qquad (13)$$

where

$$y = \frac{\mathrm{V}[F_{KR}]}{2\mathrm{E}[F_{KR}]^2} \qquad (14)$$

and

$$\delta = \frac{v}{\mathrm{E}[F_{KR}](v-2)} \qquad (15)$$

The inferences derived from simulation studies that use the KR method are usually more accurate, even with complex covariance structures (Arnau et al., 2009; Keselman, Algina, Kowalchuck, & Wolfinger, 1998; Schaalje, McBride, & Fellingham, 2002). It has been shown that under a normal distribution and with heterogeneous within-group covariance structures the KR procedure satisfies the criterion of robustness (Livacic-Rojas, Vallejo, & Fernández, 2006). Furthermore, Arnau et al. (2009) concluded that, compared with the Satterthwaite procedure (Satterthwaite, 1941), the KR correction offered better control of Type I error rates under most of the normal distribution conditions they studied, and particularly with unstructured and non-spherical covariance structures. A recent simulation study also showed that the KR method is better than the Satterthwaite approach when applying the multilevel model to data from multiple-baseline designs (Ferron, Farmer, & Owens, 2010).

The present study focuses on analyzing the performance of the KR method with non-normal as opposed to normal distributions. To this end, the KR procedure is used to estimate the fixed effects associated with time and with the interaction time x group in small samples corresponding to normal, exponential and log-normal distributions. In addition, the Akaike information criterion (AIC; Akaike, 1974) is used to select the covariance structure that shows the best fit among eleven possible structures. These structures were: a) compound symmetry (CS); b) unstructured (UN); c) first-order autoregressive (AR); d) Huynh-Feldt spherical (HF); e) within-subjects heterogeneous compound symmetry (CSH); f) within-subjects heterogeneous first-order autoregressive (ARH); g) random coefficients (RC); h) between-subjects heterogeneous unstructured (UN_j); i) between-subjects heterogeneous Huynh-Feldt spherical (HF_j); j) within-subjects and between-subjects heterogeneous first-order autoregressive (ARH_j); and k) between-subjects heterogeneous random coefficients (RC_j). The subscript j indicates that the covariance matrices are not equal between the groups.

A Monte Carlo study

The data for the simulations were generated using a series of macros created *ad hoc* in SAS 9.2 (SAS Institute, 2008). The first step involved generating the covariance matrices from variances and correlations with sphericity values of $\varepsilon = 0.57$ and $\varepsilon = 0.75$. The RANNOR generator was then used to obtain normally distributed pseudorandom observations by applying the Cholesky factor of the covariance matrix **R**. The non-normal data distributions were generated through the same procedure but were transformed by the Fleishman coefficients (Fleishman, 1978) corresponding to each of the distributions studied. The Fleishman method calculates the coefficients *a*, *b*, *c* and *d* by means of a polynomial

transformation based on different values of skewness (γ_1) and kurtosis (γ_2). The restriction imposed on these four coefficients was mean zero and variance 1, such that:

$$\gamma_1 = 2c(b^2 + 24bd + 105d^2 + 2) \tag{16}$$

and

$$\gamma_2 = 24(bd + c^2 [1 + b^2 + 28bd] + d^2 [12 + 48bd + 141c^2 + 225d^2])$$

$$(17)$$

where the constant a is equal to $-c$.

Appendix 1 shows how the non-normal data matrices were generated by means of Fleishman coefficients, yielding both exponential distributions, with fixed skewness ($\gamma_1 = 0.8$) and two values of kurtosis ($\gamma_2 = 2.4$ and $\gamma_2 = 5.4$), and log-normal distributions ($\gamma_1 = 1.75$ and $\gamma_2 = 5.9$). These values are well within the range of skew and kurtosis that represent the real-world situation (Lei & Lomax, 2005).

Study variables

The robustness of the KR approximation with normal and non-normal distributions was examined by conducting a simulation study with split-plot designs. This study used three groups for the between-subjects factor ($J = 3$) and three levels for the within-subjects factor ($K = 4$, 6 and 8). The decision to investigate these levels was based on a review of 61 simulation studies of repeated measures designs that were published between 1967 and 2010. Of these 61 studies 62.3% used $J = 3$, 72.1% used $K = 4$ and 26.2% used $K = 8$. In the present study we decided to include an intermediate value ($K = 6$) as well (Arnau et al., 2009; Padilla & Algina, 2007). It should also be noted that the levels of J and K used in the present simulation study are the ones most commonly found in educational and psychological research. For each value of K, combinations of four variables were selected: a) homogeneous

and heterogeneous between-groups covariance structures; b) total sample size; c) equal and unequal group sizes; and d) pairings of the covariance matrices and group sizes. The indices of sphericity used were $\varepsilon = 0.57$ and 0.75. The latter value (0.75) was taken to be a good approximation to sphericity, while the former (0.57) represented non-sphericity.

The data were generated using the unstructured (UN) covariance structure, as this is the most typical with longitudinal data. The UN covariance matrix requires no assumptions as regards the error terms and allows any pattern of correlation between the observations. With this matrix the variances and correlations are assumed to be non-stationary, i.e. all the variances and covariances are different. Appendix 2 shows the values of the covariance matrices for the corresponding sphericity indices and levels of repeated measures.

From the set of UN covariance structures obtained, both the equal and unequal between-groups covariance matrices were analyzed. With heterogeneous matrices the inequality of groups was adjusted to the ratio 1:3:5. The analysis considered small total sample sizes of $N = 30$, 36 and 42, as well as equal and unequal group sizes. With unbalanced groups the variance coefficient of the group size, Δn_j, was 0.41, while for balanced groups $\Delta n_j = 0$. With $\Delta n_j = 0.41$ the group sizes were: a) 5, 10, 15 ($N = 30$); b) 6, 12, 18 ($N = 36$); and c) 7, 14, 21 ($N = 42$). With $\Delta n_j = 0$ the group sizes were: a) 10, 10, 10 ($N = 30$); b) 12, 12, 12 ($N = 36$); and c) 14, 14, 14 ($N = 42$). Finally, the type of pairing between group sizes and covariance matrices was defined as one of the following: null, positive or negative. In positive pairings the largest group is associated with the covariance matrix whose values are largest. Conversely, negative pairing associates the largest group with the covariance matrix containing the smallest element values. Pairing is null in the case of balanced groups.

Table 1 shows the different combinations of variables examined in this study. For each combination, 1000 replications were performed at a significance level of 0.05, for both normal and non-normal distributions.

[INSERT TABLE 1 ABOUT HERE]

Results

As indicated above, the first stage of this study consisted in generating the UN population covariance matrices for $\varepsilon = 0.57$ and $\varepsilon = 0.75$. In a second stage these matrices were then used to obtain normal and non-normal data distributions. In a third stage, 11 covariance structures were fitted to each data set by means of *Proc Mixed* from SAS, and according to the AIC. Finally, *Proc Mixed* was again used to calculate the Type I error rates for the effects of time and the interaction time x group. The next two sections report the fit percentages of the covariance matrices selected by means of the AIC, as well as the Type I error rates, specifying the covariance structure selected by the AIC.

Selecting the covariance structure

Given the presence of both within-subjects and between-subjects heterogeneity, 11 covariance structures were fitted according to the AIC of *Proc Mixed*. The present study used the AIC as it is more effective than the Bayesian information criterion (BIC; Schwarz, 1978). Keselman et al. (1998) demonstrated that the AIC selects the true population covariance structure on 47% of occasions, whereas the BIC only does so 35% of the time. A similar study by Ferron, Dailey, and Yi (2002) reported correct selection rates of 79% for the AIC and 66% for the BIC. Subsequently, Vallejo and Livacic-Rojas (2005) showed that the test based on the AIC outperformed that based on the BIC when it came to controlling the Type I error rate, especially when used in conjunction with the KR method. These authors found that with the

BIC, *Proc Mixed* offered poor control of the estimated probabilities of Type I error. However, in a study of the KR procedure Gomez, Schaalje, and Fellingham (2005) concluded that the AIC only performs better with complex covariance structures, such as the UN covariance matrix. This is consistent with the findings of Keselman et al. (1998). Recently Vallejo, Ato, and Valdés (2008) demonstrated that with different group sizes the AIC provides a better estimate of standard errors. In light of these findings the present simulation study made use of the AIC. Note, however, that the AIC does not always select the true structure, as other structures may provide more adequate approximations. In this regard, Vaida and Blanchard (2005) proposed the conditional AIC (CAIC) as an effective and useful alternative for selecting mixed-effects models, arguing that it provided a good approximation that was adequate to the amount of data and the amount of information it contains. More recently, Vallejo, Fernández, Livacic-Rojas, and Tuero-Herrero (2011) found that the shape of the distribution did not affect the correct decision rates of the AIC and the BIC with UN covariance patterns, and also showed that the AIC performed better with both normal and non-normal distributions.

The results of the present study are presented in Tables 2 and 3, which show the percentage fit of the most common covariance matrices in relation to the UN population matrix, and for homogeneous and heterogeneous between-groups covariances, respectively.

It can be seen in Table 2 that with a normal distribution the structure with the best fit (66.7%) is the UN population matrix. With $\varepsilon = 0.57$ it shows a good fit in all cases (for $K = 4$, 6 and 8). However, when $\varepsilon = 0.75$ it only shows the best fit for $K = 4$, and as the number of repeated measures increases ($K = 6$ and 8) it is the CSH matrix which offers the best fit (33.3%). When the distribution is exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 2.4$ the best fit (33.4%) is provided by the UN population matrix when $K = 4$. As the value of K increases, the best fit is shown by other structures: CSH with $\varepsilon = 0.75$, and ARH and UN$_j$ with $\varepsilon = 0.57$. When

kurtosis increases to $\gamma_2 = 5.4$ it is the UN_j matrix which provides the best fit (66.6%) when K = 4 and 6. However, when the number of repeated measures increases to $K = 8$ the best fit is offered by the UN population matrix with $\varepsilon = 0.57$, and by the CSH matrix with $\varepsilon = 0.75$. In the case of the log-normal distribution the population matrix never provides the best fit, and it can be seen that here the UN_j matrix has a fit percentage of 100.

[INSERT TABLE 2 ABOUT HERE]

Table 3 shows the percentage fit for heterogeneous between-groups covariance matrices. As regards the type of pairing it can be seen that with normal distributions it is only in the case of null pairing that the best fit (22.2%) is shown by the UN_j population matrix. This is also the case for exponential distributions (16.7% when $\gamma_2 = 2.4$ and 22.2% when $\gamma_2 = 5.4$). With the log-normal distribution the UN_j population matrix provides the best fit (33.4%) for all values of K. When the distribution is exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 2.4$, the ARH_j structure also shows a correct fit (16.7%), as it does, to a slightly lesser extent (11.2%), when the distribution is normal or exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 5.4$. With positive and negative pairings the UN_j population matrix only provides the best fit when $K = 4$, this being the case for both normal and non-normal distributions. As the number of repeated measures increases it is the ARH_j matrix which shows the best fit for all the distributions, especially when $\varepsilon = 0.75$. Note, however, that when $\varepsilon = 0.57$ and $K = 6$ the UN matrix offers a good fit for all distributions except the exponential distribution with $\gamma_1 = 0.8$ and $\gamma_2 = 2.4$. The reason why the ARH_j matrix provides the best fit with positive and negative pairing is probably that the dependency structure of the data becomes more apparent as the number of repeated measures increases.

[INSERT TABLE 3 ABOUT HERE]

Type I error rates

The simulated data were then analyzed with *Proc Mixed*, specifying the covariance structure with the best fit according to the AIC (Tables 2 and 3). The analysis involved estimating the *p* values associated with the fixed effects, via the KR approximation, and the empirical Type I error rates for each combination of the different study variables (Table 1).

Robustness was determined according to Bradley's criterion, whereby the effect estimate is robust when the empirical Type I error rate is between 0.025 and 0.075 for $\alpha = 0.05$. A test is considered to be liberal when the empirical Type I error rate is above the upper limit, and conservative when it is below the lower limit.

Table 4 shows the empirical Type I error rates for the time effect, as well as the percentages of robustness. With $K = 4$ and $\varepsilon = 0.57$ the KR method shows 66.7% robustness when the distribution is normal, and when it is exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 2.4$. The percentage of robustness falls slightly to 58.3% when kurtosis increases to $\gamma_2 = 5.4$, and the test is never robust when the distribution is log-normal ($\gamma_1 = 1.75$ and $\gamma_2 = 5.9$). With $K = 4$ and $\varepsilon = 0.75$ the difference between normal and log-normal distributions is not so marked: when the distribution is normal the robustness percentage is the same as that for $\varepsilon = 0.57$, while with the log-normal distribution the test achieves 41.7% robustness. In this case, and in contrast to what occurs with the normal distribution, the test is liberal with homogeneous between-groups covariances. The highest robustness percentage for $\varepsilon = 0.75$ is the 75% corresponding to exponential distributions with $\gamma_1 = 0.8$ and $\gamma_2 = 5.4$, where the test is robust for both null and positive pairings.

With $K = 6$ and $\varepsilon = 0.57$, robustness is 41.7% with the normal distribution, but only 8.3% with the log-normal distribution. With $\varepsilon = 0.75$, robustness is 75% with the normal distribution and 25% with the log-normal distribution, where the test is liberal with null pairings. Note, however, that with positive pairings and $\varepsilon = 0.57$ the test is conservative for both normal and log-normal distributions. As regards the exponential distributions, robustness ranged between 50% and 83.3%, it being greater with $\varepsilon = 0.75$. When the distribution was exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 5.4$ the test is robust even with negative pairings.

With $K = 8$ and $\varepsilon = 0.57$ robustness is 50% with the normal distribution, but 0% for the log-normal distribution. An even greater difference can be seen with $\varepsilon = 0.75$, where robustness was again 0% with the log-normal distribution but 75% in the case of the normal distribution. As regards exponential distributions, the performance is similar when $K = 6$, although the test is slightly less robust.

[INSERT TABLE 4 ABOUT HERE]

Table 5 shows the empirical Type I error rates and percentages of robustness for the effect of the interaction time x group. It can be seen that with $K = 4$ there are hardly any differences between normal and non-normal (exponential and log-normal) distributions. When $\varepsilon = 0.57$, robustness is 58.3% with the normal distribution, 41.7% for the exponential distribution with $\gamma_1 = 0.8$ and $\gamma_2 = 2.4$, 58.3% for the exponential distribution with $\gamma_1 = 0.8$ and $\gamma_2 = 5.4$, and 50% with the log-normal distribution. These percentages are more or less maintained when $\varepsilon = 0.75$, where robustness is 41.7% for the normal distribution, for the exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 2.4$ and for the log-normal distribution, and 66.7% for the exponential distribution with $\gamma_1 = 0.8$ and $\gamma_2 = 5.4$. Similarly, with K = 6 and $\varepsilon = 0.75$ there are no differences between the normal distribution, the exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 2.4$

and the log-normal distribution. In these cases, robustness ranges between 41.7% and 50%, although it rises to 75% for the exponential distribution with $\gamma_1 = 0.8$ and $\gamma_2 = 5.4$, the test being robust with negative pairings. With $\varepsilon = 0.57$ the test is again more robust (41.7%) when the distribution is exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 5.4$. When the distribution is log-normal and $\varepsilon = 0.57$, robustness reaches 33.3%, compared to 25% for the normal distribution. It can be seen that the normal distribution and the exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 2.4$ have the same percentage of robustness for both sphericity indices.

With K = 8 the test is more robust when the distribution is exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 5.4$, followed by the log-normal and finally the normal distribution. The greatest difference between the normal and log-normal distributions is observed with positive pairings, where the test is robust with the log-normal distribution but not with normally distributed data.

[INSERT TABLE 5 ABOUT HERE]

Overall, Tables 4 and 5 show that *Proc Mixed* in combination with the KR method is unable to control the Type I error rate when pairing is negative. This finding is corroborated by previous analyses of UN and UN$_j$ population matrices. Vallejo and Ato (2006) concluded that *Proc Mixed* in conjunction with the KR procedure based on the AIC tends to inflate Type I error rates for the interaction effect when pairing is negative, this being the case for both normal and non-normal distributions. Moreover, this tendency of the test to be liberal increases as the sample size gets smaller. The present results also illustrate this. As regards the main time effect, Vallejo and Livacic-Rojas (2005) found that when the condition of sphericity was fulfilled the test was robust even with distributions that deviated slightly from

normality, and with small sample sizes and positive pairing. The present study reached the same conclusion with respect to exponential distributions.

Discussion

Moderate non-normality has a minimal effect on the standard errors of estimation methods. However, standard error bias tends to increase in line with the degree of non-normality (Lei & Lomax, 2005). Hence the focus of the present analysis of longitudinal data was to examine the extent to which non-normality influenced the estimation of fixed effects. Specifically, the aim of the study was to determine the robustness of the LMM in mixed longitudinal designs when the data are not normally distributed. A previous study with normally distributed data showed that the LMM in combination with the KR method is more robust than the between-within and Satterthwaite approximations, particularly when the population covariance matrices are unstructured or heterogeneous first-order autoregressive (Arnau et al., 2009). The main contribution of this study by Arnau et al. (2009) was to demonstrate that the KR procedure corrects the liberal Type I error rates obtained through the between-within and Satterthwaite methods, especially when there are positive pairings between group sizes and covariance matrices. Vallejo and Ato (2006) concluded that in split-plot designs the KR correction may be a viable alternative for estimating the interaction effect. In their analysis of the estimation of time and interaction effects Vallejo and Livacic-Rojas (2005) found that the KR correction based on the AIC was able to control the Type I error rates in most of the distributions studied. Hence the present study examined the Type I error rates for time and interaction effects in order to determine the influence of skewness and kurtosis on the estimation of fixed effects when distributions are non-normal.

An interesting result of this study, in relation to selecting the covariance structure with heterogeneous between-groups covariance matrices, is that the ARH_j matrix showed the best fit, regardless of the type of distribution and with both positive and negative pairings. This ARH_j structure of dependency in repeated measures data was previously reported by Keselman et al. (1998). In their simulation study they found that with normal distributions and $\varepsilon = 0.75$ the ARH_j matrix showed the best fit to the UN_j population matrix, this being the case for both positive and negative pairings. By contrast, with log-normal distributions and $\varepsilon = 0.75$ there were no differences between the ARH_j and UN_j matrices, with both showing a correct fit to the UN_j population matrix.

With respect to Type I error rates the present results show that for the time effect the test is more robust with normal as opposed to log-normal distributions. However, the difference was not so marked between normal and exponential distributions. With a log-normal distribution and $\varepsilon = 0.57$ the test is not robust: the percentages are null or close to zero. With a normal distribution and $\varepsilon = 0.75$ the test is robust for all values of K. However, with a log-normal distribution and $\varepsilon = 0.75$ the test becomes less robust as the number of repeated measures increases, and ceases to be robust when $K = 8$. It should be noted, therefore, that the test is more robust with a normal as opposed to a log-normal distribution, whereas, overall, there are no significant differences between normal and exponential distributions. In addition, when the covariance matrix is spherical the test tends to become more robust with normal and exponential distributions, especially when the number of observations increases, this being contrary to what occurs with the log-normal distribution. Finally, comparison of the two exponential distributions shows that the test becomes more robust as kurtosis increases, regardless of whether the assumption of sphericity is fulfilled or not.

As compared to the estimation of the time effect alone, the interaction between time and group leads to a considerable increase in the test's robustness when the distribution is log-normal. Indeed, with a log-normal distribution and $\varepsilon = 0.57$ the test has zero robustness in relation to the time effect, but is much better when estimating the interaction effect. Conversely, robustness decreases when the distribution is normal, particularly as the number of repeated measures increases. In this regard, Padilla and Algina (2007) showed that the Type I error rate tends to be higher as the value of K rises. When $\varepsilon = 0.75$ no differences are observed between the normal distribution, the exponential with $\gamma_1 = 0.8$ and $\gamma_2 = 2.4$, and the log-normal with $K = 4$ and 6. By contrast, with $K = 8$ the test is more robust with the log-normal distribution. As occurs when estimating the time effect the highest percentage of robustness in relation to the interaction effect corresponds to the sphericity condition $\varepsilon = 0.75$ and the exponential distribution with $\gamma_2 = 5.4$.

To summarize, the KR approximation is least robust when the distribution is log-normal, where robustness is null when estimating the time effect and with non-spherical covariance matrices ($\varepsilon = 0.57$). In our view this is largely due to the increase in skewness ($\gamma_1 = 1.75$), as this lack of robustness is not observed for exponential distributions with $\gamma_1 = 0.8$. Robustness is high when the exponential distribution has a degree of kurtosis that is very similar to the log-normal distribution, i.e. $\gamma_2 = 5.4$. Therefore, we conclude that the KR procedure is compromised with log-normal distributions, which show moderate skewness, especially as regards the estimation of the time effect. By contrast, when distributions are normal or have slight skewness ($\gamma_1 = 0.8$) the test is robust even with extreme kurtosis ($\gamma_2 = 2.4$ and $\gamma_2 = 5.4$). In fact, the percentage of robustness was high for the exponential distribution with $\gamma_1 = 0.8$ and $\gamma_2 = 5.4$, and when estimating both the time and interaction effects.

In conclusion, two effects are revealed by the present results, one due to skewness and the other to kurtosis. The effect of skewness is detected when comparing exponential with log-normal distributions, where robustness decreases as skewness increases. On the other hand, the effect of kurtosis is revealed when comparing two exponential distributions, where robustness increases in line with the degree of kurtosis.

Two final reflections can be made. Firstly, while acknowledging that the results are limited to the conditions examined in this study we believe they could in fact be generalized to a wide variety of conditions. However, the results obtained are difficult to compare with other simulation studies, since the sample sizes and values of skewness and kurtosis differ from one study to another. Secondly, it would be interesting in future research to study the functioning of generalized linear mixed models. These techniques do not require the error terms to be normally distributed, and they are well-suited to most of the distributions found with real-life data.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control, AC-19*, 716-723. doi:10.1109/TAC.1974.1100705

Algina, J., & Oshima, T. C. (1995). An improved general approximation test for the main effect in a split-plot design. *British Journal of Mathematical and Statistical Psychology, 48*, 149-160. doi: 10.1111/j.2044-8317.1995.tb01055.x

Arnau, J., Balluerka, N., Bono, R., & Gorostiaga, A. (2010). General linear mixed model for analysing longitudinal data in developmental research. *Perceptual and Motor Skills, 110*, 547-566. doi:10.2466/PMS.110.2.547-566

Arnau, J., Bono, R., & Vallejo, G. (2009) Analyzing small samples of repeated measures data with the mixed-model adjusted F test. *Communications in Statistics: Simulation and Computations, 38*, 1083-1103. doi:10.1080/03610910902785746

Brown, D. D., Weatherholt, T. N., & Burns, B. M. (2010). Attention skills and looking to television in children from low income families. *Journal of Applied Developmental Psychology, 31*, 330-338. doi:10.1016/j.appdev.2010.04.003

Claret, L., Girard, P., Hoff, P. M., Van Cutsem, E., Zuideveld, K. P., Jorga, K., Fagerberg, J., & Bruno, R. (2009). Model based-prediction of phase III overall survival in colorectal cancer on de basis of phase II tumor dynamics. *Journal of Clinical Oncology, 27*, 4103-4108. doi:10.1200/JCO.2008.21.0807

Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyze unbalanced repeated measures and longitudinal data. *Statistics in Medicine, 16*, 2349-2380.

Ferron, J., Dailey, R., & Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavior Research, 37*, 379-403. doi:10.1207/S15327906MBR3703_4

Ferron, J. M., Farmer, J. L., & Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study of multilevel-modeling approaches. *Behavior Research Methods, 42*, 930-943. doi:10.3758/BRM.42.4.930

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*, 521-532. doi:10.1007/BF02293811

Gomez, E. V., Schaalje, G. B., Fellingham, G. W. (2005). Performance of the KR method when the covariance structure is selected using AIC and BIC. *Communication in Statistics: Theory and Methods, 34*, 377-392. doi:10.1081/SAC-200055719

Horner, R. D. (1987). Age at onset of Alzheimer's disease: Clue to the relative importance of etiologic factors?. *American Journal of Epidemiology, 126*, 409-414.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983-997. doi:10.2037/2533558

Keselman, H. J., Algina, J., Kowalchuk R. K., & Wolfinger, R. D. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics: Simulation and Computation, 27,* 591-604. doi:10.1080/03610919808813497

Keselman, H. J., Algina, J., Kowalchuk R. K., & Wolfinger, R. D. (1999). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, *52*, 63-78. doi:10.1348/000711099158964

Keselman, H. J., Algina, J., Wilcox, R. R., & Kowalchuk, R. K. (2000). Testing repeated measures hypothesis when covariance matrices are heterogeneous: Revisiting the

robustness of the Welch-James test again. *Educational and Psychological Measurement, 60*, 925-938. doi:10.1177/00131640021970998

Kondo, K. (1977). The log-normal distribution of the incubation time of exogenous diseases. *Japanese Journal of Human Genetics, 21*, 217-237.

Kowalchuck, R. K., Keselman, H. J., Algina, J., & Wolfinger, R. D. (2004). The analysis of repeated measurements with mixed-model adjusted F tests. *Educational and Psychological Measurement, 64*, 224-242. doi:10.1177/0013164403260196

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38*, 963-974. doi:10.2307/2529876

Lee, E. T., & Wang, J. W. (2003). *Statistical methods for survival data analysis* (3a ed.). New Jersey: John Wiley and Sons

Lei, M., & Lomax, R. G. (2005). The effect of varying degrees on nonnormality in structural equation modeling. *Structural Equation Modeling, 12*, 1-27. doi:10.1207/s15328007sem1201_1

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues. *BioScience, 51*, 341-352. doi:10.1641/006-3568

Littell, R. C., Pendergast, J. P., & Natarajan, R. (2000) Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine, 19,* 1793-1819. doi:10.1002/1097-0258(20000715)

Livacic-Rojas, P., Vallejo, G., & Fernández, P. (2006). Procedimientos estadísticos alternativos para evaluar la robustez mediante diseños de medidas repetidas. *Revista Latinoamericana de Psicología, 38,* 579-599.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166. doi:10.1037/0033-2909.105.1.156

Padilla, M. A., & Algina, J. (2007). Type I error rates of the KR adjusted degree of freedom *F*-test for a split-plot design with missing values. *Journal of Modern Applied Statistical Methods, 6*, 66-80.

Park, T., & Lee, J. (2002) Covariance models for nested repeated measures data: analysis of ovarian steroid secretion data. *Statistics in Medicine, 21,* 143-164. doi:10.1002/sim.949

Preston, F. W. (1981). Pseudo-log-normal distributions. *Ecology, 62*, 355-364. doi:10.2307/1936710

Qazi, S., DuMez, D., & Uckun, F. M. (2007). Meta analysis of advanced cancer survival data using log-normal parametric fitting: A statistical method to identify effective treatment protocols. *Current Pharmaceutical Design, 13*, 1533-1544. doi:10.2174/138161207780765882

SAS Institute Inc. (2008). *SAS/STAT 9.2 user's guide*. Cary, NC: Author.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika 6*, 309-316. doi:10.1007/BF02288586

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and type II error properties of the *t* test to departures from population normality. *Psychological Bulletin, 111*, 353-360. doi:10.1037/0033-2909.111.2.352

Schaalje, G. B., McBride, J. B., & Fellingham, G. W. (2001). Approximations to distributions of test statistics in complex mixed linear models using SAS Proc MIXED, *SUGI 26 Proceedings, Statistics, Data Analysis and Data Mining Section,* Paper 262-26. Consulted on 15 September 2011 at http://www2.sas.com/proceedings/sugi26/p262-26.pdf

Schaalje, G. B., McBride, J. B., & Fellingham, G. W. (2002). Adequacy of approximations to distributions of test statistics in complex mixed linear models, *Journal of Agricultural Biological and Environmental Statistics, 7*, 512-524. doi:10.1198/108571102726

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464. doi:10.1214/aos/1176344136.

Shang-wen., Y., & Ming-hua, H. (2010). Estimation of air traffic longitudinal conflict probability based on the reaction time of controllers. *Safety Science, 48*, 926-930. doi:10.1016/j.ssci.2010.03.016

Ulrich, R., & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology, 37*, 513-525. doi:10.1006/jmps.1993.1032

Vaida, F., & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika, 92*, 351-370. doi:10.1093/biomet/92.2.351

Vallejo, G., & Ato, M. (2006). Modified Brown-Forsythe procedure for testing interaction effects in split-plot designs. *Multivariate Behavioral Research, 41*, 549-578. doi:10.1207/s15327906mbr4104_6

Vallejo, G., Ato, M., Valdés, T. (2008). Consequences of misspecifying the error covariance structure in linear mixed models for longitudinal data. *Methodology: European Journal of Research for the Behavioral and Social Sciences, 4*, 10-21. doi:10.1027/1016-9040.12.1.10

Vallejo, G., Fernández, M. P., Livaci-Rojas, P. E, & Tuero-Herrero, E. (2011). Selecting the best unbalanced repeated measures model. *Behavior Research Methods, 43*, 18-36. doi:10.3758/s13428-010-0040-1

Vallejo, G., & Livacic-Rojas, P. (2005). Comparison of two procedures for analyzing small sets of repeated measures data. *Multivariate Behavioral Research, 40*, 179-205. doi:10.1207/s15327906mbr4002_2

Van der Leeden, R. (1998) Multilevel analysis of repeated measures data. *Quality and Quantity, 32*, 15-29. doi:10.1023/A:1004233225855

Van der Leeden, R., Vrijburg, K. E., & De Leeuw, J. (1996) A review of two different approaches for the analysis of growth data using longitudinal mixed linear models: comparing hierarchical linear regression (ML3, HLM) and repeated measures designs with structured covariance matrices (BMDP5V). *Computational Statistics and Data Analysis, 21*, 583-605. doi:10.1016/0167-9473(96)82296-1

Van der Linden, W. J. (2006). A log-normal model for response times on test items. *Journal of Educational and Behavioral Statistics, 31*, 181-204. doi:10.3102/10769986031002181

Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

Ware, J. H. (1985) Linear models for the analysis of longitudinal studies. *Journal of the American Statistical Association, 80,* 95-101.

Wilcox, R. R. (1993). Analysing repeated measures or randomized block designs using trimmed means. *British Journal of Mathematical and Statistical Psychology, 46*, 63-76. doi:10.1111/j.2044-8317.1993.tb01002.x

Wilcox, R. R. (1994). A one-way random effects model for trimmed means. *Psychometrika, 59*, 289-306. doi:10.1007/BF02296126

Wright, S. P., & Wolfinger, R. D. (1996, October). *Repeated measures analysis using mixed models: Some simulation results*. Poster presented at the Conference on Modelling

Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions, Nantucket, MA.

Wu, Y. B., Clopper, R. R., & Wooldridge, P. J. (1999). A comparison of traditional approaches to hierarchical linear modelling when analysing longitudinal data. *Research in Nursing and Health, 22*, 421-432. doi:10.1002/(SICI)1098-240X(199910)

Table 1. Group sizes for balanced and unbalanced designs with $J = 3$, $K = 4$, 6 and 8, the UN population covariance matrix, and $\varepsilon = 0.57$ and $\varepsilon = 0.75$

N	N_1	n_2	n_3	Δn_j	Between-groups covariance	Pairing
30	10	10	10	0	=	Null
36	12	12	12	0	=	Null
42	14	14	14	0	=	Null
30	10	10	10	0	$\neq$	Null
36	12	12	12	0	$\neq$	Null
42	14	14	14	0	$\neq$	Null
30	5	10	15	0.41	$\neq$	+
36	6	12	18	0.41	$\neq$	+
42	7	14	21	0.41	$\neq$	+
30	15	10	5	0.41	$\neq$	-
36	18	12	6	0.41	$\neq$	-
42	21	14	7	0.41	$\neq$	-

Note. J: groups; K: number of repeated measures; ε: sphericity; N: total sample size; n_1, n_2 and n_3: group sizes; Δn_j: variance coefficient of the group size; =/$\neq$: homogeneity/heterogeneity of between-groups covariance matrices; +/-: positive/negative pairing of group sizes and covariance matrices.

Table 2. Percentage fit of the most common covariance structures (UN, CSH, ARH and UN$_j$) in relation to the UN population covariance matrix. The percentages correspond to homogeneous between-groups covariance.

	Covariance matrices that were fitted					
	UN		CSH	ARH	UN$_j$	
Distributions	ε = 0.57	ε = 0.75	ε = 0.75	ε = 0.57	ε = 0.57	ε = 0.75
Normal	K = 4, 6, 8	K = 4	K = 6, 8			
	50	16.7				
Totals	66.7		33.3			
Exponential	K = 4	K = 4	K = 6, 8	K = 6	K = 8	
(γ_1=0.8 γ_2=2.4)	16.7	16.7				
Totals	33.4		33.3	16.7	16.7	
Exponential	K = 8		K = 8		K = 4, 6	K = 4, 6
(γ_1=0.8 γ_2=5.4)					33.3	33.3
Totals	16.7		16.7		66.6	
Log-normal					K = 4, 6, 8	K = 4, 6, 8
(γ_1=1.75 γ_2=5.9)					50	50
Totals					100	

Table 3. Percentage fit of the most common covariance structures (UN, UN$_j$ and ARH$_j$) in relation to the UN$_j$ population matrix. The percentages correspond to heterogeneous between-groups covariance.

| | Pairing | | | | | |
| | Null | | + | | - | |
Distributions	ε=0.57	ε=0.75	ε=0.57	ε=0.75	ε=0.57	ε=0.75
UN						
Normal			K=6		K=6	
			5.6		5.6	
Exponential (γ_1=0.8 γ_2=2.4)						
Exponential (γ_1=0.8 γ_2=5.4)			K=6		K=6	
			5.6		5.6	
Log-normal (γ_1=1.75 γ_2=5.9)			K=6		K=6	
			5.6		5.6	
UN$_j$						
Normal	K=6,8	K=4,8	K=4	K=4	K=4	K=4
	11.1	11.1	5.6	5.6	5.6	5.6
Totals	22.2		11.2		11.2	
Exponential (γ_1=0.8 γ_2=2.4)	K=4	K=4,8	K=4	K=4	K=4	K=4
	5.6	11.1	5.6	5.6	5.6	5.6
Totals	16.7		11.2		11.2	
Exponential (γ_1=0.8 γ_2=5.4)	K=4,6	K=4,6	K=4	K=4	K=4	K=4
	11.1	11.1	5.6	5.6	5.6	5.6
Totals	22.2		11.2		11.2	
Log-normal (γ_1=1.75 γ_2=5.9)	K=4,6,8	K=4,6,8	K=4	K=4	K=4	K=4
Totals	16.7	16.7	5.6	5.6	5.6	5.6
	33.4		11.2		11.2	
ARH$_j$						
Normal	K=4	K=6	K=8	K=6,8	K=8	K=6,8
	5.6	5.6	5.6	11.1	5.6	11.1
Totals	11.2		16.7		16.7	
Exponential (γ_1=0.8 γ_2=2.4)	K=6,8	K=6	K=6,8	K=6,8	K=6,8	K=6,8
	11.1	5.6	11.1	11.1	11.1	11.1
Totals	16.7		22.2		22.2	
Exponential (γ_1=0.8 γ_2=5.4)	K=8	K=8	K=8	K=6,8	K=8	K=6,8
	5.6	5.6	5.6	11.1	5.6	11.1
Totals	11.2		16.7		16.7	
Log-normal (γ_1=1.75 γ_2=5.9)			K=8	K=6,8	K=8	K=6,8
			5.6	11.1	5.6	11.1
Totals			16.7		16.7	

Table 4. Empirical rates of Type I error for the time effect (nominal value 0.05).

					Btw. group cov.	Pairing	Distributions							
							Normal		$\gamma_1 = 0.8$ $\gamma_2 = 2.4$		$\gamma_1 = 0.8$ $\gamma_2 = 5.4$		$\gamma_1 = 1.75$ $\gamma_2 = 5.9$	
							Sphericity							
N	N_1	n_2	n_3	Δn_j			0.57	0.75	0.57	0.75	0.57	0.75	0.57	0.75
						$K = 4$								
30	10	10	10	0.00	=	Null	0.075	0.061	**0.086**	**0.080**	0.069	0.068	**0.086**	**0.084**
36	12	12	12	0.00	=	Null	0.061	0.061	0.071	**0.077**	0.070	0.064	**0.085**	**0.087**
42	14	14	14	0.00	=	Null	0.066	0.047	0.068	**0.081**	0.066	0.062	**0.083**	**0.092**
30	10	10	10	0.00	≠	Null	0.067	**0.079**	0.062	0.072	0.069	0.069	**0.107**	0.068
36	12	12	12	0.00	≠	Null	0.064	0.064	0.063	0.069	**0.078**	0.060	**0.084**	0.068
42	14	14	14	0.00	≠	Null	0.068	0.067	0.075	0.063	0.063	0.061	**0.104**	0.064
30	5	10	15	0.41	≠	+	**0.079**	0.074	0.068	**0.077**	0.072	0.073	**0.091**	0.078
36	6	12	18	0.41	≠	+	0.072	0.073	0.073	0.066	**0.077**	0.065	**0.112**	0.075
42	7	14	21	0.41	≠	+	0.063	0.068	0.067	0.069	0.060	0.068	**0.107**	0.065
30	15	10	5	0.41	≠	-	**0.127**	**0.130**	**0.125**	**0.102**	**0.106**	**0.115**	**0.135**	**0.114**
36	18	12	6	0.41	≠	-	**0.092**	**0.096**	**0.106**	**0.097**	**0.088**	**0.082**	**0.126**	**0.090**
42	21	14	7	0.41	≠	-	**0.094**	**0.087**	**0.079**	**0.076**	**0.089**	**0.083**	**0.155**	**0.080**
Percentages of robustness							66.7	66.7	66.7	41.7	58.3	75.0	0.0	41.7
						$K = 6$								
30	10	10	10	0.00	=	Null	0.075	0.070	**0.087**	**0.083**	0.069	0.070	**0.100**	**0.103**
36	12	12	12	0.00	=	Null	**0.079**	0.058	**0.078**	0.063	0.074	0.061	**0.111**	0.075
42	14	14	14	0.00	=	Null	0.064	0.073	0.066	0.073	0.062	0.059	**0.090**	**0.102**
30	10	10	10	0.00	≠	Null	0.054	0.065	0.074	0.071	**0.082**	**0.088**	**0.103**	**0.104**
36	12	12	12	0.00	≠	Null	0.054	0.070	0.075	0.068	0.067	**0.078**	**0.109**	**0.102**
42	14	14	14	0.00	≠	Null	0.068	0.060	**0.081**	0.063	0.073	0.072	**0.090**	**0.107**
30	5	10	15	0.41	≠	+	*0.017*	**0.079**	0.057	0.069	**0.079**	0.055	0.038	**0.086**
36	6	12	18	0.41	≠	+	*0.011*	0.062	0.073	0.066	**0.079**	0.058	*0.021*	0.067
42	7	14	21	0.41	≠	+	*0.011*	0.054	0.061	0.070	0.064	0.066	*0.017*	0.069
30	15	10	5	0.41	≠	-	**0.319**	**0.083**	**0.085**	**0.092**	0.071	0.071	**0.345**	**0.113**
36	18	12	6	0.41	≠	-	**0.277**	0.069	**0.091**	**0.087**	0.066	0.073	**0.301**	**0.125**
42	21	14	7	0.41	≠	-	**0.272**	**0.088**	**0.080**	**0.082**	**0.082**	0.070	**0.320**	**0.076**
Percentages of robustness							41.7	75.0	50.0	66.7	66.7	83.3	8.3	25.0
						$K = 8$								
30	10	10	10	0.00	=	Null	**0.094**	0.072	**0.078**	**0.091**	0.071	**0.079**	**0.129**	**0.108**
36	12	12	12	0.00	=	Null	0.073	0.059	**0.084**	0.073	0.075	0.064	**0.115**	**0.102**
42	14	14	14	0.00	=	Null	0.073	0.057	**0.084**	0.073	0.075	0.064	**0.115**	0.098
30	10	10	10	0.00	≠	Null	0.073	**0.078**	0.071	**0.103**	**0.078**	0.063	**0.132**	**0.139**
36	12	12	12	0.00	≠	Null	0.073	0.055	**0.078**	0.075	0.075	0.051	**0.115**	**0.127**
42	14	14	14	0.00	≠	Null	0.069	0.075	**0.079**	**0.082**	0.063	0.059	**0.132**	0.098
30	5	10	15	0.41	≠	+	**0.085**	**0.081**	**0.081**	0.074	0.067	0.068	**0.081**	**0.090**
36	6	12	18	0.41	≠	+	0.066	0.071	**0.084**	0.067	**0.076**	**0.080**	**0.099**	**0.090**
42	7	14	21	0.41	≠	+	**0.087**	0.071	0.069	0.064	0.060	0.053	**0.080**	**0.079**
30	15	10	5	0.41	≠	-	**0.094**	**0.103**	**0.089**	**0.078**	**0.089**	**0.090**	**0.123**	**0.131**
36	18	12	6	0.41	≠	-	**0.092**	0.072	**0.088**	**0.085**	**0.083**	0.073	**0.124**	**0.119**
42	21	14	7	0.41	≠	-	**0.076**	0.075	0.066	**0.089**	**0.078**	0.060	**0.097**	**0.101**
Percentages of robustness							50.0	75.0	25.0	50.0	50.0	75.0	0.0	0.0

Note. In bold = liberal; in italics = conservative

Table 5. Empirical rates of Type I error for the interaction effect (nominal value 0.05).

							Distributions							
							Normal		$\gamma_1 = 0.8$ $\gamma_2 = 2.4$		$\gamma_1 = 0.8$ $\gamma_2 = 5.4$		$\gamma_1 = 1.75$ $\gamma_2 = 5.9$	
							Sphericity							
N	N_1	n_2	n_3	Δn_j	Btw. group cov.	Pairing	0.57	0.75	0.57	0.75	0.57	0.75	0.57	0.75
									$K = 4$					
30	10	10	10	0.00	=	Null	**0.077**	**0.083**	**0.077**	**0.081**	0.057	0.050	0.063	0.065
36	12	12	12	0.00	=	Null	0.066	0.065	0.058	0.063	0.059	0.070	0.069	**0.076**
42	14	14	14	0.00	=	Null	0.073	0.066	0.064	0.065	0.068	0.051	0.071	0.075
30	10	10	10	0.00	≠	Null	0.073	**0.086**	0.063	**0.088**	**0.077**	0.072	**0.088**	**0.089**
36	12	12	12	0.00	≠	Null	0.063	**0.080**	**0.084**	**0.076**	0.059	0.062	0.075	0.075
42	14	14	14	0.00	≠	Null	0.064	0.069	0.068	0.072	0.055	0.058	**0.089**	0.072
30	5	10	15	0.41	≠	+	**0.088**	**0.094**	**0.089**	**0.088**	0.074	**0.094**	0.060	**0.093**
36	6	12	18	0.41	≠	+	0.072	0.069	0.061	0.068	0.064	0.069	**0.079**	0.074
42	7	14	21	0.41	≠	+	0.058	0.064	**0.082**	0.062	**0.077**	0.053	0.055	**0.081**
30	15	10	5	0.41	≠	-	**0.171**	**0.198**	**0.170**	**0.153**	**0.161**	**0.147**	**0.189**	**0.187**
36	18	12	6	0.41	≠	-	**0.118**	**0.129**	**0.157**	**0.111**	**0.111**	**0.103**	**0.181**	**0.132**
42	21	14	7	0.41	≠	-	**0.111**	**0.111**	**0.106**	**0.105**	**0.106**	**0.105**	**0.191**	**0.121**
					Percentages of robustness		58.3	41.7	41.7	41.7	58.3	66.7	50.0	41.7
									$K = 6$					
30	10	10	10	0.00	=	Null	**0.094**	**0.083**	**0.082**	**0.087**	0.071	0.061	**0.081**	**0.083**
36	12	12	12	0.00	=	Null	**0.082**	0.066	**0.081**	0.067	**0.082**	0.063	0.069	0.075
42	14	14	14	0.00	=	Null	0.063	0.061	**0.076**	0.059	0.059	0.065	**0.080**	0.064
30	10	10	10	0.00	≠	Null	**0.093**	0.059	0.075	**0.079**	**0.087**	**0.092**	**0.111**	**0.108**
36	12	12	12	0.00	≠	Null	**0.083**	**0.083**	0.074	**0.084**	**0.091**	0.070	**0.100**	**0.097**
42	14	14	14	0.00	≠	Null	**0.084**	0.065	**0.092**	0.068	0.069	**0.088**	0.076	**0.107**
30	5	10	15	0.41	≠	+	0.028	**0.092**	**0.077**	0.072	**0.080**	0.057	0.029	0.062
36	6	12	18	0.41	≠	+	0.025	0.068	0.068	0.069	**0.079**	0.057	0.030	0.062
42	7	14	21	0.41	≠	+	*0.020*	0.059	**0.076**	0.062	0.068	0.062	0.031	0.060
30	15	10	5	0.41	≠	-	**0.350**	**0.116**	**0.109**	0.097	**0.091**	**0.103**	**0.357**	**0.119**
36	18	12	6	0.41	≠	-	**0.306**	**0.096**	**0.090**	0.083	**0.079**	0.073	**0.306**	**0.114**
42	21	14	7	0.41	≠	-	**0.308**	**0.095**	**0.081**	0.083	0.070	0.072	**0.310**	0.087
					Percentages of robustness		25.0	50.0	25.0	50.0	41.7	75.0	33.3	41.7
									$K = 8$					
30	10	10	10	0.00	=	Null	**0.093**	**0.076**	**0.094**	**0.079**	**0.080**	**0.083**	0.075	**0.089**
36	12	12	12	0.00	=	Null	**0.085**	**0.076**	**0.084**	**0.081**	0.075	0.068	**0.076**	0.075
42	14	14	14	0.00	=	Null	0.068	0.071	0.068	**0.077**	0.073	**0.080**	**0.078**	**0.080**
30	10	10	10	0.00	≠	Null	**0.083**	**0.119**	**0.082**	**0.105**	0.065	0.075	**0.115**	**0.103**
36	12	12	12	0.00	≠	Null	**0.103**	**0.082**	**0.084**	**0.097**	**0.077**	0.066	**0.095**	**0.123**
42	14	14	14	0.00	≠	Null	**0.086**	**0.080**	**0.080**	**0.100**	0.066	0.063	**0.096**	**0.105**
30	5	10	15	0.41	≠	+	**0.094**	0.075	**0.079**	**0.090**	**0.091**	**0.076**	0.068	0.072
36	6	12	18	0.41	≠	+	**0.085**	**0.093**	0.074	0.070	**0.076**	0.069	0.068	0.051
42	7	14	21	0.41	≠	+	**0.086**	**0.083**	**0.076**	0.069	0.060	0.055	0.063	0.061
30	15	10	5	0.41	≠	-	**0.104**	**0.116**	**0.116**	**0.120**	**0.104**	**0.086**	**0.128**	**0.130**
36	18	12	6	0.41	≠	-	**0.105**	**0.112**	**0.097**	**0.114**	**0.097**	**0.076**	**0.126**	**0.106**
42	21	14	7	0.41	≠	-	**0.082**	**0.105**	**0.091**	**0.096**	**0.089**	**0.084**	**0.108**	**0.096**
					Percentages of robustness		8.3	16.7	16.7	16.7	41.7	50.0	33.3	33.3

Note. In bold = liberal; in italics = conservative

Appendix 1: Generation of non-normal data matrices by means of Fleishman coefficients

```
/*…*/
/*Distribution with skewness 0.8 and kurtosis 2.4*/

    b=0.848445836;
    c=0.104049451;
    d=0.044849610;
    a=-c;


      z1=j(n1,q,.);
      do i=1 to n1;
            do j=1 to q;
      z1[i,j]=rannor(0);
      z1[i,j]=a+b*z1[i,j]+c*z1[i,j]**2+d*z1[i,j]**3;
      end;
            end;

/*…*/


/*…*/
/*Distribution with skewness 0.8 and kurtosis 5.4*/

    b=0.702207971;
    c=0.082964688;
    d=0.088904623;
    a=-c;

      z1=j(n1,q,.);
      do i=1 to n1;
            do j=1 to q;
      z1[i,j]=rannor(0);
      z1[i,j]=a+b*z1[i,j]+c*z1[i,j]**2+d*z1[i,j]**3;
      end;
            end;
/*…*/


/*…*/
/*Distribution with skewness 1.75 and kurtosis 5.9*/

    b=0.774926306;
    c=0.222093456;
    d=0.054958336;
    a=-c;


      z1=j(n1,q,.);
      do i=1 to n1;
            do j=1 to q;
      z1[i,j]=rannor(0);
      z1[i,j]=a+b*z1[i,j]+c*z1[i,j]**2+d*z1[i,j]**3;
      end;
            end;
/*…*/
```

Appendix 2: UN population covariance matrices.

	ε = 0.57			ε = 0.75	

$K = 4$

$$\begin{bmatrix} 1 & 1.58 & 1.87 & 1.91 \\ & 4 & 4.74 & 4.83 \\ & & 9 & 9.18 \\ & & & 15 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1.40 & 2.40 & 3.37 \\ & 4 & 4.20 & 6.20 \\ & & 9 & 8.13 \\ & & & 15 \end{bmatrix}$$

$K = 6$

$$\begin{bmatrix} 1 & 1.63 & 1.50 & 1.70 & 1.90 & 2.08 \\ & 4 & 3.99 & 3.39 & 3.79 & 4.16 \\ & & 6 & 5.65 & 4.65 & 5.09 \\ & & & 8 & 7.29 & 5.88 \\ & & & & 10 & 8.93 \\ & & & & & 12 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1.41 & 1.47 & 1.70 & 1.90 & 2.08 \\ & 4 & 3.45 & 3.39 & 3.79 & 4.16 \\ & & 6 & 4.88 & 4.65 & 5.09 \\ & & & 8 & 6.31 & 5.88 \\ & & & & 10 & 7.72 \\ & & & & & 12 \end{bmatrix}$$

$K = 8$

$$\begin{bmatrix} 1 & 1.58 & 1.47 & 1.70 & 1.90 & 2.08 & 2.24 & 2.40 \\ & 4 & 3.88 & 3.39 & 3.79 & 4.16 & 4.49 & 4.80 \\ & & 6 & 5.49 & 4.65 & 5.09 & 5.50 & 5.88 \\ & & & 8 & 7.08 & 5.88 & 6.35 & 6.79 \\ & & & & 10 & 8.68 & 7.10 & 7.59 \\ & & & & & 12 & 10.27 & 8.31 \\ & & & & & & 14 & 11.85 \\ & & & & & & & 16 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1.34 & 1.47 & 1.70 & 1.90 & 2.08 & 2.24 & 2.40 \\ & 4 & 3.28 & 3.39 & 3.79 & 4.16 & 4.49 & 4.80 \\ & & 6 & 4.64 & 4.65 & 5.09 & 5.50 & 5.88 \\ & & & 8 & 5.99 & 5.88 & 6.35 & 6.79 \\ & & & & 10 & 7.34 & 7.10 & 7.59 \\ & & & & & 12 & 8.68 & 8.31 \\ & & & & & & 14 & 10.03 \\ & & & & & & & 16 \end{bmatrix}$$

Note: UN = unstructured model.