

SPHERICITY ESTIMATION BIAS FOR REPEATED MEASURES DESIGNS IN  
SIMULATION STUDIES

Roser Bono<sup>1,2</sup>, Jaume Arnau<sup>1</sup>, María J. Blanca<sup>3</sup>, and Rafael Alarcón<sup>3</sup>

<sup>1</sup> Department of Methodology of the Behavioral Sciences (Faculty of Psychology),  
University of Barcelona

<sup>2</sup> Institute for Brain, Cognition, and Behavior (IR3C), University of Barcelona

<sup>3</sup> Department of Psychobiology and Behavioral Sciences Methodology (Faculty of  
Psychology), University of Malaga

Address for correspondence:

Roser Bono

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](mailto:rbono@ub.edu)

Tel. +34 93 312 50 80 \* Fax +34 93 402 13 59

### Abstract

This study explores the accuracy of sphericity estimation and analyzes how the sphericity of covariance matrices may be affected when the latter are derived from simulated data. We analyzed the consequences that normal and non-normal data generated from an unstructured population covariance matrix — with low ( $\varepsilon = .57$ ) and high ( $\varepsilon = .75$ ) sphericity — can have on the sphericity of the matrix that is fitted to these data. To this end, data were generated for four types of distribution (normal, slightly skewed, moderately skewed, and severely skewed or lognormal), four sample sizes (very small, small, medium, and large), and four values of the within-subjects factor ( $K = 4, 6, 8, \text{ and } 10$ ). Normal data were generated using the Cholesky decomposition of the correlation matrix, while the Vale-Maurelli method was used to generate non-normal data.

The results indicate the extent to which sphericity is altered by recalculating the covariance matrix on the basis of simulated data. It is concluded that bias is greater with spherical covariance matrices, non-normal distributions, and small sample sizes, and that it increases in line with the value of  $K$ . An interaction was also observed between sample size and  $K$ : with very small samples the observed bias is greater as the value of  $K$  increases.

Keywords: Monte Carlo simulation, Cholesky decomposition, Vale-Maurelli method, sphericity estimation, covariance matrix.

In recent years, researchers have used Monte Carlo simulation methods to study the robustness and power of various analytic techniques. By means of simulation it is possible to generate not only normally distributed data but also data that reflect what is commonly found in real-world settings (Blanca, Arnau, López-Montiel, Bono, & Bendayan, 2013; Micceri, 1989). Thus, various Monte Carlo simulation studies have analyzed the fixed effects associated with time (repeated measures variable) using normally or non-normally distributed data (Arnau, Bono, Blanca, & Bendayan, 2012; Arnau, Bono, & Vallejo, 2009; Kowalchuk, Keselman, Algina, & Wolfinger, 2004; Vallejo & Ato, 2006, among others). Most of these studies analyzed mixed models by generating data from an unstructured (UN) population covariance matrix with sphericity values of .57 and .75. In simulation studies of this kind one would ideally know if the sphericity estimated from the simulated data is equivalent to the sphericity that was fixed initially. However, there are no published studies addressing this aspect.

In order to generate normally distributed data most simulation studies of repeated measures designs make use of the Cholesky decomposition of the correlation matrix (Lix, Algina, & Keselman, 2003). Among the various methods developed to generate non-normal data (Fleishman, 1978; Headrick, 2002, 2004; L'Ecuyer, 1990; Marsaglia, 2003; Ramberg, Tadikamalla, Dudewicz, & Mykytka, 1979; Tadikamalla, 1980, Vale & Maurelli, 1983, among others), the method of Vale and Maurelli (1983) is one of the most-widely used by simulation studies in the social sciences. According to Olvera-Astivia and Zumbo (2014), this method has more than 130 citation counts on the ISI Web of Knowledge, and over 230 on Google Scholar. The procedures used to generate data can alter the sphericity of the fixed covariance matrix, since the process of data simulation involves two steps: the generation of a population covariance matrix from sphericity values, and the generation of normal or non-normal data using this

covariance matrix. It is in this second step that the Cholesky decomposition or the method of Vale and Maurelli would be applied.

The aim of the present study was to examine how the type of distribution, the sample size, the number of repeated measures, and the sphericity value of the population covariance matrix affect the sphericity estimation of simulated repeated measures data. In other words, we sought to determine the extent to which the fixed sphericity (population sphericity) differs from the estimated sphericity (sample sphericity). To this end, data were generated for both the normal distribution and non-normal distributions commonly used in simulation studies. For each distribution we analyzed sphericity estimation bias in relation to different sample sizes, different numbers of repeated measures, and different sphericity values of the kind frequently found in simulation studies.

### **Vale-Maurelli Method**

The method of Vale and Maurelli (1983) is a multivariate extension of the method proposed by Fleishman (1978). The Fleishman method uses the polynomial transformation of normal variables:

$$Y = a + bX + cX^2 + dX^3 \quad (1)$$

where  $a$ ,  $b$ ,  $c$ , and  $d$  are the polynomial coefficients that control the first four moments of random variable  $Y$ , and  $X$  is a random variable distributed normally with mean zero and variance 1. The constant  $a$  is equal to  $-c$ .

The values of skewness ( $\gamma_1$ ) and kurtosis ( $\gamma_2$ ) are defined by:

$$\gamma_1 = 2c(b^2 + 24bd + 105d^2 + 2) \quad (2)$$

and

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

(3)

Vale and Maurelli (1983) extended this method to the generation of multivariate non-normal distributions. To this end, they defined the vectors  $\mathbf{x}$  and  $\mathbf{w}$ , and the variable  $Y$  as:

$$\mathbf{x}^T = [1, X, X^2, X^3] \quad (4)$$

$$\mathbf{w}^T = [a, b, c, d] \quad (5)$$

$$Y = \mathbf{w}^T \mathbf{x} \quad (6)$$

where  $X$  is specified as in Eq. 1, and  $\mathbf{w}^T$  is the vector of polynomial weights that control the first four moments of the new non-normal distribution  $Y$ .

Equation 7 represents the correlation coefficient of two non-normal variables  $Y_1$  and  $Y_2$  generated from two normal variables  $X_1$  and  $X_2$

$$r_{Y_1 Y_2} = E(Y_1 Y_2) = E(\mathbf{w}_1^T \mathbf{x}_1 \mathbf{x}_2^T \mathbf{w}_2) = \mathbf{w}_1^T \mathbf{R} \mathbf{w}_2 \quad (7)$$

where  $\mathbf{R} = E(\mathbf{x}_1 \mathbf{x}_2^T)$ .

The correlation between  $Y_1$  and  $Y_2$ , expressed with the weights, is

$$r_{Y_1 Y_2} = \rho_{X_1 X_2} (b_1 b_2 + 3b_1 d_2 + 3d_1 b_2 + 9d_1 d_2) + r_{X_1 X_2}^2 (2c_1 c_2) + r_{X_1 X_2}^3 (6d_1 d_2)$$

(8)

where  $\rho_{X_1 X_2}$  is the correlation between the normal variables  $X_1$  and  $X_2$ .

By solving for  $\rho_{X_1 X_2}$ , it is possible to find the intermediate correlation matrix and to specify all the elements that will serve to generate the data. In summary, the solution proposed by Vale and Maurelli (1983) calculates an intermediate correlation matrix. Its data are the same as the population correlation matrix and, given that one applies the Fleishman method to each marginal distribution, the correlation matrix is transformed to the desired one that is used to generate the data (Olvera-Astivia & Zumbo, 2014).

### A Monte Carlo Study

Data were generated using SAS/IML (version 9.4), since this software is one of the most suitable for simulating data (Kashyap, Butt, & Bhattacharjee, 2009), and is also one of the most popular for implementing the Vale and Maurelli method (Keselman & Lix, 1997; Lix et al., 2003; Vallejo, Arnau, & Ato, 2007; Vallejo & Livacic-Rojas, 2005).

The first step involved generating the UN population covariance matrices from variances and correlations with sphericity values of  $\varepsilon = .57$  and  $.75$  for the different values of the repeated measures,  $K = 4, 6, 8,$  and  $10$  (Table 1). The sphericity value of the population covariance matrices was calculated using the Greenhouse-Geisser epsilon (Greenhouse & Geisser, 1959).

[INSERT TABLE 1 ABOUT HERE]

In the next step, the RANNOR generator in SAS was used to obtain normally distributed multivariate pseudorandom observations by means of the Cholesky decomposition (Lix et al., 2003). The non-normal data distributions were generated using the method of Vale and Maurelli (1983). For each non-normal distribution the vector of Eq. 5 was obtained using Fleishman (1978) coefficients in order to provide the desired degree of multivariate skewness and kurtosis. Table 2 shows the Fleishman coefficients  $a, b, c,$  and  $d$  used to generate non-normal data. These coefficients

correspond to exponential distributions, with fixed skewness ( $\gamma_1 = 0.8$ ) and two values of kurtosis ( $\gamma_2 = 2.4$  and  $5.4$ ), and to the lognormal distribution ( $\gamma_1 = 1.75$  and  $\gamma_2 = 5.9$ ).

[INSERT TABLE 2 ABOUT HERE]

Finally, the average sphericity of the UN covariance matrices of the simulated data was estimated using the Greenhouse-Geisser epsilon (Greenhouse & Geisser, 1959), which was obtained through *proc glm* in SAS. We then calculated the empirical bias between the initially fixed sphericity (population sphericity) and the sphericity estimated on the basis of the simulations (sample sphericity).

### Study Variables

Four variables were manipulated in this study.

1. **Sample sizes.** The sample sizes chosen were the same as or similar to the cell sizes most widely used in the simulation studies of repeated measures designs published since 1990 (Arnau, Bendayan, Blanca, & Bono, 2013a, 2013b, 2014; Arnau et al., 2009; Keselman, Carriere, & Lix, 1993; Keselman, & Keselman, 1990; Kowalchuk et al., 2004, among many others). Based on these studies we chose to examine both very small ( $N = 5, 6, 7,$  and  $10$ ) and small ( $N = 12, 14, 15, 18,$  and  $21$ ) samples. In addition, and with the goal of determining the value of  $N$  at which sphericity estimation bias approaches zero, we also included the medium ( $N = 30, 45, 60,$  and  $75$ ) and large ( $N = 90, 100, 200, 300,$  and  $500$ ) sample sizes that have been used in other simulation studies of repeated measures designs. The studies by Arnau et al. (2013a, 2013b), Keselman, Algina, Kowalchuk, and Wolfinger (1998), Keselman et al. (1993), and Lix et al. (2003) examined medium group sizes. The study by Olvera-Astivia and Zumbo (2014)

examined large sample sizes with the aim of determining the properties of data-generation algorithms for multivariate non-normal data. In the study by Oberfeld and Franke (2013) both extremely small and larger sample sizes were examined with the aim of evaluating the robustness of repeated measures analyses.

**2. Degree of contamination of the distribution.** The distributions selected were the normal distribution and a series of non-normal distributions defined by the most common values of skewness and kurtosis, whether in simulation or empirical studies. In several simulation studies of repeated measures designs, the distributions are classified as either normal or as slightly, moderately, or strongly biased distributions (Berkovits, Hancock, & Nevitt, 2000; Vallejo et al., 2007). Among the strongly biased distributions, a number of simulation studies have analyzed the log-normal distribution (Algina & Keselman, 1998; Keselman, Kowalchuk, & Boik, 2000; Kowalchuk et al., 2004, among others).

The distributions used in the present study had positive values of skewness and kurtosis, given that such values are used in simulation studies and are also the most common found in distributions of psychological variables (Blanca et al., 2013). Regarding the degree of contamination, the extreme values chosen were  $\gamma_1 = 1.75$  and  $\gamma_2 = 5.9$ , which correspond to the log-normal distribution, one of the most widely studied. The other two distributions analyzed had a fixed skewness,  $\gamma_1 = 0.8$ , and two values of kurtosis,  $\gamma_2 = 2.4$  and  $\gamma_2 = 5.4$ . These values are well within the range of skewness and kurtosis that is observed in real-world settings (Blanca et al., 2013; Lei & Lomax, 2005), and they are also the values used in the study by Arnau et al. (2012).

**3. Sphericity values.** The sphericity indices used were  $\epsilon = .57$  and  $.75$ . The latter value was taken to be a good approximation to sphericity, while the former represented non-sphericity. Both values have been used in the majority of simulation studies of



repeated measures designs (Algina & Keselman, 1998; Arnau et al., 2012, 2013a, 2013b, 2014; Arnau et al., 2009; Berkovits et al., 2000; Keselman & Keselman, 1990, among many others).

4. *Levels of the within-subjects factor.* In the present study we decided to use  $K = 4, 6, 8,$  and  $10$ . It should be noted that the level of  $K = 4$  is the most commonly found in simulation studies (Berkovits et al., 2000; Keselman, et al., 2000; Kowalchuk et al., 2004; Lix et al., 2003; Tian, & Wilcox, 2007; Vallejo et al., 2007). Eight repeated measures were used in the studies by Keselman et al. (2000), Kowalchuk and Keselman (2001), and Vallejo and Ato (2006). The intermediate value of  $K = 6$  was also examined in the studies by Arnau et al. (2009, 2012), Padilla and Algina (2007), and Wilcox (2006). Finally, we also analyzed an extreme number of repeated measures ( $K = 10$ ), as used in the simulation study by Ahmad, Werner, and Brunner (2008).

Each combination of sample size, distribution shape, sphericity, and number of repeated measures was replicated 1000 times ( $18 \times 4 \times 2 \times 4 \times 1000 = 576,000$  simulations).

## Data Analysis

In order to simplify the statistical analysis the variable ‘sample size’ was recategorized into very small, small, medium, and large.

The univariate analyses of variance (ANOVAs) were performed using *proc glm* from SAS. Specifically, we conducted two separate  $4 \times 4 \times 4$  ( $N \times$  distribution  $\times K$ ) ANOVAs for each level of  $\epsilon$ , and eight separate  $4 \times 4$  ( $N \times$  distribution) ANOVAs for each level of  $\epsilon$  and  $K$ . Post hoc comparisons for each  $N$  and distribution were performed by means of the Bonferroni test. Polynomial contrasts using *proc glm*, and forward multiple regression analyses using *proc reg* from SAS, were performed for each level of

$K$  and sphericity with the aim of examining the impact of both distribution type and sample size on the empirical bias in sphericity. Finally, a  $4 \times 4 \times 4 \times 2$  ( $N \times$  distribution  $\times K \times \varepsilon$ ) ANOVA was conducted.

Values of empirical bias in the sphericity estimation were taken as the dependent variable, with sample size, distribution, sphericity, and the number of repeated measures being included as factors. Partial eta square  $\eta_p^2$  was calculated as a measure of effect size.

### Results

In this section we report the bias observed when estimating sphericity for the different sample sizes, numbers of repeated measures, and types of distribution when the sphericity of the population covariance matrix was .57 (Figure 1) and .75 (Figure 2). Bias was considered to be null when the deviation was close to zero, between -0.080 and 0.080. This interval, which was chosen arbitrarily by the authors, is shown shaded in both figures.

[INSERT FIGURE 1 ABOUT HERE]

[INSERT FIGURE 2 ABOUT HERE]

In Figure 1, which shows the empirical bias when  $\varepsilon = .57$ , it can be seen that with a small number of repeated measures ( $K = 4$ ) the sphericity estimation is not biased, regardless of the distribution and sample size. However, as the number of repeated measures increases, the sphericity estimation shows a negative bias with very small and small sample sizes. This bias then approaches the interval between -0.080 and

0.080 as sample size becomes medium or large. Thus, the interaction  $N \times K$  is statistically significant ( $F_{(9,224)} = 58.857$ ;  $p < .001$ ;  $\eta_p^2 = .703$ ; observed power = 1). There are also significant differences between sample sizes ( $F_{(3,224)} = 849.246$ ;  $p < .001$ ;  $\eta_p^2 = .910$ ; observed power = 1) and between distributions ( $F_{(3,224)} = 21.086$ ;  $p < .001$ ;  $\eta_p^2 = .220$ ; observed power = 1). The normal distribution is the least biased, followed by the slightly skewed ( $\gamma_1 = 0.8$  and  $\gamma_2 = 2.4$ ), the moderately skewed ( $\gamma_1 = 0.8$  and  $\gamma_2 = 5.4$ ), and the severely skewed or lognormal ( $\gamma_1 = 1.75$  and  $\gamma_2 = 5.9$ ) distributions. Table 3 shows the results of the ANOVAs and the multiple comparisons for each of the plots shown in Figure 1. The Bonferroni post hoc tests indicate that there are significant differences between all the sample sizes considered ( $p < .001$ ), except for the comparison of medium and large samples when  $K = 4$ . Regarding the distributions, significant differences are observed between the normal distribution and the lognormal distribution for any value of  $K$ , and also between the normal distribution and the moderately skewed distribution for  $K = 6$  and  $K = 8$ . Finally, none of the  $N \times K$  distribution interactions is statistically significant.

[INSERT TABLE 3 ABOUT HERE]

Figure 2, which depicts the empirical bias when  $\varepsilon = .75$ , shows a notable increase in bias in comparison with Figure 1. With  $K = 4$  the bias is negative for very small samples, irrespective of their distribution. For small samples, bias is observed with non-normal distributions. As the value of  $K$  increases so does the extent to which sphericity is underestimated, this being the case even for medium-sized samples. This is reflected in the analysis of the  $N \times K$  interaction ( $F_{(9,224)} = 33.711$ ;  $p < .001$ ;  $\eta_p^2 = .575$ ; observed power = 1). As in Figure 1 the effects of sample size and the type of

distribution are statistically significant:  $F_{(3,224)} = 1036.241$ ;  $p < .001$ ;  $\eta_p^2 = .933$ ; observed power = 1, and  $F_{(3,224)} = 23.879$ ;  $p < .001$ ;  $\eta_p^2 = .242$ ; observed power = 1. The multiple comparisons (Table 4) yield similar results to the previous analysis (Table 3). For  $K = 4$ , however, differences are now also observed between the slightly skewed and the lognormal distribution, and between the normal and the moderately skewed distribution, while for  $K = 6$  and  $K = 8$  there are differences between the slightly skewed and the lognormal distribution.

[INSERT TABLE 4 ABOUT HERE]

The polynomial coefficients for each level of  $K$  and  $\varepsilon$  are shown in Table 5. The linear and quadratic components are significant ( $p < .001$ ) in all the models of analysis. The weight of the linear component is greater than that of the quadratic one, and both increase in line with the value of  $K$  and  $\varepsilon$ . The linear contrast estimates increase in a positive direction, whereas the increase in the quadratic contrast estimates follows a negative direction.

[INSERT TABLE 5 ABOUT HERE]

Equation (1) analyzes the regression model that includes  $N$  and the distribution for the different values of  $K$  and  $\varepsilon$ :

$$\text{Empirical bias} = b_0 + b_1N + b_2\text{Distribution} + e \quad (1)$$

where  $b_0$  is the constant,  $b_i$  are the unstandardized estimated coefficients in the regression analysis for each of the explanatory variables defined previously, and  $e$  is the error term. Unstandardized estimated coefficients represent the predicted change in empirical bias for a one unit change in the explanatory variable when all other explanatory variables are held constant. The  $b_1$  and  $b_2$  estimated coefficients using equation (1) are shown in Table 6. The results reveal a positive relationship between empirical bias and  $N$ , and a negative relationship between empirical bias and distribution. Note that sphericity is underestimated, such that the bias approaches zero as sample size increases, whereas the bias increases as the data deviate from the normal distribution. These effects are heightened as the value of  $K$  increases and when  $\varepsilon = .75$ .

[INSERT TABLE 6 ABOUT HERE]

If we compare the different plots shown in Figures 1 and 2, it can be seen that the profile of the sphericity estimation bias for spherical matrices ( $\varepsilon = .75$ ) and  $K = 4$  is similar to that for non-spherical matrices ( $\varepsilon = .57$ ) with  $K = 6$ , and that the profile of spherical matrices with  $K = 6$  is similar to that of non-spherical matrices with  $K = 10$ . In other words, the profile of estimation bias for spherical matrices approaches that of non-spherical matrices as the number of repeated measures increases. The  $K \times \varepsilon$  interaction is significant ( $F_{(3,448)} = 7.066$ ;  $p < .001$ ;  $\eta_p^2 = .045$ ; observed power = .981). As the value of  $K$  increases, so does the difference in bias between  $\varepsilon = .57$  and  $\varepsilon = .75$ . The  $N \times \varepsilon$  interaction is also significant ( $F_{(3,448)} = 121.861$ ;  $p < .001$ ;  $\eta_p^2 = .449$ ; observed power = 1). With very small and small sample sizes, bias is greater when  $\varepsilon = .75$ . Finally, the distribution  $\times \varepsilon$  interaction is not significant ( $F_{(3,448)} = 2.494$ ;  $p = .059$ ;  $\eta_p^2 = .016$ ;

observed power = .617), whereas the effect of the sphericity variable is statistically significant ( $F_{(1,448)} = 924.704$ ;  $p < .001$ ;  $\eta_p^2 = .674$ ; observed power = 1).

In conclusion, the results confirm that the underestimation of sphericity is greater with very small and small sample sizes, as the number of repeated measures increases, and as the distribution deviates from normality. These effects are observed to a greater extent when the covariance matrix is spherical. Note that a negative bias is produced even with normal distributions.

### **Discussion**

In this study the Cholesky decomposition of the correlation matrix was used to generate normally distributed data, while non-normal data were generated using the method of Vale and Maurelli. It is possible that these methods alter the covariance between the variables and, therefore, the value of sphericity. In addition, the covariance among the variables differed across the different distribution shapes, sample sizes, number of repeated measures, and sphericity.

We determined the range of estimated sphericity values of the covariance matrices that were generated. With non-normal data, the sample sphericity would tend to decrease as the population sphericity increases, and, therefore, the generated sphericity could be affected, especially for values of  $\varepsilon = .75$ . Thus, with spherical matrices the bias is greater with non-normal distributions, smaller sample sizes, and as the value of  $K$  increases. This effect is also observed with normal distributions, albeit to a lesser extent. It can be stated, therefore, that with simulated data there will always be a mismatch between population sphericity and sample sphericity.

The results of this study suggest that as the sphericity of the population covariance matrix approaches 1, the sphericity calculated on the basis of simulated data

tends to decrease. Furthermore, there is a certain equivalence between the profiles of sphericity estimation bias, since the sphericity estimation of spherical population matrices is similar to that of non-spherical matrices when the number of repeated measures in the latter increases. In other words, less bias is produced with non-spherical matrices, but it increases in line with the value of  $K$ , such that these matrices then behave as if they were spherical. These results are in line with what one would expect, as when estimating the error matrix for the calculation of the Greenhouse-Geisser epsilon, bias increases in line with the size of this matrix, and this bias is even greater when the population covariance matrix has a sphericity value close to 1.

The results also indicate that the population covariance matrix is transformed after generating non-normal data by means of the Vale-Maurelli method. The same occurs, albeit to a lesser extent, when using the Cholesky decomposition to generate normal data. With both methods, sphericity is underestimated, especially when  $N$  is very small or small. This is due to the direct relationship between sample size and the variance estimation.

In summary, the estimation of sample sphericity is influenced not only by the type of distribution and the population sphericity, but also — and notably — by the number of repeated measures and the sample size. An inverse relationship between  $N$  and  $K$  is clearly observed (Oberfeld & Franke, 2013). When  $N$  is very small, an increase in  $K$  leads to greater bias than is the case when these two conditions are not fulfilled. None of these aspects has been considered before, and we have followed the data generation procedures typically used in simulation studies. Consequently, researchers should exercise caution when interpreting the results of their simulations, especially when working with small sample sizes. At all events, we believe that these results highlight an interesting point that could be addressed in future studies. In the context of

such studies, the profiles of empirical bias presented here (Figures 1 and 2) could be used by researchers to identify the extent to which the sphericity estimation is biased. The obtained results can be extended to real data, with applications for applied research where it is necessary to know to what extent population sphericity and sample sphericity match and to ensure the power of the statistical model (e.g., Gracia, García, & Lila, 2008, 2014).

A final point to consider is that the results obtained here are limited to the conditions studied. Furthermore, the study has focused on the generation of unstructured population covariance matrices. In future studies it would therefore be interesting to determine profiles of sphericity estimation bias for other population matrices, such as the first-order autoregressive covariance matrix, which provides a good fit to repeated measures data (Arnau et al., 2012; Keselman et al., 1998). Another avenue of interest would be to generate data when sample sizes are not equal for each value of  $K$ , that is, when there are missing data.



### **Acknowledgements**

This research was supported by grant PSI2012-32662 from the Spanish Ministry of Economy and Competitiveness.

### References

- Ahmad, M. R., Werner, C., & Brunner, E. (2008). Analysis of high dimensional repeated measures designs: The one sample case. *Computational Statistics and Data Analysis*, *53*, 416-427. doi: 10.1016/j.csda.2008.08.013
- Algina, J., & Keselman, H. J. (1998). A power comparison of the Welch-James and Improved General Approximation test in the split-plot design. *Journal of the Educational and Behavioral Statistics*, *23*, 152-159. doi: 10.3102/10769986023002152
- Arnau, J., Bendayan, R., Blanca, M. J., & Bono, R. (2013a). The effect of skewness and kurtosis on the robustness of linear mixed models. *Behavior Research Methods*, *45*, 873-879. doi: 10.3758/s13428-012-0306-x
- Arnau, J., Bendayan, R., Blanca, M. J., & Bono, R. (2013b). Should we rely on the Kenward-Roger approximation when using linear mixed models if the groups have different distributions? *British Journal of Mathematical and Statistical Psychology*, *67*, 408-429. doi: 10.1111/bmsp.12026
- Arnau, J., Bendayan, R., Blanca, M. J., & Bono, R. (2014). The effect of skewness and kurtosis on the Kenward-Roger approximation when group distributions differ. *Psicothema*, *26*, 279-285. doi: 10.7334/psicothema2013.174
- Arnau, J., Bono, R., Blanca, M. J., & Bendayan, R. (2012). Using the linear mixed model to analyze non-normal data distributions in longitudinal designs. *Behavior Research Methods*, *44*, 1224-1238. doi: 10.3758/s13428-012-0196-y
- 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 Computation*, *38*, 1083-1103. doi: 10.1080/03610910902785746

- Berkovits, I., Hancock, G. R., & Nevitt, J. (2000). Bootstrap resampling approaches for repeated measure designs: Relative robustness to sphericity and normality violations. *Educational and Psychological Measurement, 60*, 877-892. doi: 10.1177/00131640021970961
- Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 9*, 78-84. doi: 10.1027/1614-2241/a000057
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*, 521-532. doi: 10.1007/BF02293811
- Gracia, E., García, F., & Lila, M. (2008). Police involvement in cases of intimate partner violence against women: The influence of perceived severity and personal responsibility. *Violence against Women, 14*, 697-714. doi:10.1177/1077801208317288
- Gracia, E., García, F., & Lila, M. (2014). Male police officers' law enforcement preferences in cases of intimate partner violence versus non-intimate interpersonal violence: Do sexist attitudes and empathy matter? *Criminal Justice and Behavior, 41*, 1195-1213. doi:10.1177/0093854814541655
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 32*, 95-112. doi: 10.1007/BF02289823
- Headrick, T. C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics & Data Analysis, 40*, 685-711. doi: 10.1016/S0167-9473(02)00072-5
- Headrick, T. C. (2004). On polynomial transformations for simulating multivariate non-normal distributions. *Journal of Modern Applied Statistical Methods, 3*, 65-71.

- Kashyap, M. P., Butt, N. S., & Bhattacharjee, D. (2009). Simulations study to compare the random data generation from Bernoulli distribution in popular statistical packages. *Pakistan Journal of Statistics and Operation Research*, 5, 99-106.
- 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., Carriere, K. C., & Lix, L. M. (1993). Testing repeated measures hypotheses when covariance matrices are heterogeneous. *Journal of Educational Statistics*, 18, 305-319.
- Keselman, J. C., & Keselman, H. J. (1990). Analysis unbalanced repeated measures designs. *British Journal of Mathematical and Statistical Psychology*, 43, 265-282. doi: 10.1111/j.2044-8317.1990.tb00940.x
- Keselman, H. J., Kowalchuk, R. K., & Boik, R. J. (2000). An examination of the robustness of the empirical Bayes and other approaches for testing main and interaction effects in repeated measures designs. *British Journal of Mathematical and Statistical Psychology*, 53, 51-67. doi: 10.1348/000711000159178
- 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, 319-338. doi: 10.1111/j.2044-8317.1997.tb01148.x
- Kowalchuk, R. K., & Keselman, H. J. (2001). Mixed-model pairwise multiple comparisons of repeated measures means. *Psychological Bulletin*, 6, 282-296. doi: 10.1037/1082-989X.6.3.282

- Kowalchuk, 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
- L'Ecuyer, P. (1990). Random numbers for simulation. *Communications of the ACM, 33*, 85-97.
- 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
- 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. doi: 10.1207/s15327906mbr3804\_1
- Marsaglia, G. (2003). Random number generators. *Journal of Modern Applied Statistical Methods, 2*, 2-13.
- 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
- Oberfeld, D., & Franke, T. (2013). Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data. *Behavior Research Methods, 45*, 792-812. doi: 10.3758/s13428-012-0281-2.
- Olvera-Astivia, O., & Zumbo, B. D. (2014). A cautionary note on the use of the Vale and Maurelli method to generate multivariate, nonnormal data for simulation purposes. *Educational and Psychological Measurement, 1-27*. doi: 10.1177/0013164414548894.

- Padilla, M. A., & Algina, J. (2007). Type I error rates of the Kenward-Roger adjusted degree of freedom F test for a split plot design with missing values. *Journal of Modern Applied Statistical Methods*, *6*, 66-80.
- Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J., & Mykytka, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics*, *21*, 201-214. doi: 10.1080/00401706.1979.10489750
- Tadikamalla, P. (1980). On simulating non-normal distributions. *Psychometrika*, *45*, 273-279. doi: 10.1007/BF02294081
- Tian, T., & Wilcox, R. (2007). A comparison of two rank tests for repeated measures designs. *Journal of Modern Applied Statistical Methods*, *6*, 331-335.
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465-471. doi: 10.1007/BF02293687
- Vallejo, G., Arnau, J., & Ato, M. (2007). Comparative robustness of recent methods for analyzing multivariate repeated measures designs. *Educational and Psychological Measurement*, *67*, 410-432. doi: 10.1177/0013164406294777
- 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., & 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
- Wilcox, R. (2006). Pairwise comparisons of dependent groups based on medians. *Computational Statistics and Data Analysis*, *50*, 2933-2941. doi: 10.1016/j.csda.2005.04.017

Table 1. UN population covariance matrices.

$\varepsilon = .57$	$\varepsilon = .75$
$K = 4$	
$\begin{bmatrix} 1 & 1.60 & 1.80 & 2.32 \\ & 4 & 4.80 & 4.65 \\ & & 9 & 9.30 \\ & & & 15 \end{bmatrix}$	$\begin{bmatrix} 1 & 1.40 & 2.40 & 3.37 \\ & 4 & 4.20 & 6.20 \\ & & 9 & 8.13 \\ & & & 15 \end{bmatrix}$
$K = 6$	
$\begin{bmatrix} 1 & 1.63 & 1.50 & 1.70 & 1.90 & 2.08 \\ & 4 & 3.99 & 3.39 & 3.79 & 4.16 \\ & & 6 & 5.65 & 4.65 & 5.09 \\ & & & 8 & 7.29 & 5.88 \\ & & & & 10 & 8.93 \\ & & & & & 12 \end{bmatrix}$	$\begin{bmatrix} 1 & 1.41 & 1.47 & 1.70 & 1.90 & 2.08 \\ & 4 & 3.45 & 3.39 & 3.79 & 4.16 \\ & & 6 & 4.88 & 4.65 & 5.09 \\ & & & 8 & 6.31 & 5.88 \\ & & & & 10 & 7.72 \\ & & & & & 12 \end{bmatrix}$
$K = 8$	
$\begin{bmatrix} 1 & 1.58 & 1.47 & 1.70 & 1.90 & 2.08 & 2.24 & 2.40 \\ & 4 & 3.88 & 3.39 & 3.79 & 4.16 & 4.49 & 4.80 \\ & & 6 & 5.49 & 4.65 & 5.09 & 5.50 & 5.88 \\ & & & 8 & 7.08 & 5.88 & 6.35 & 6.79 \\ & & & & 10 & 8.68 & 7.10 & 7.59 \\ & & & & & 12 & 10.27 & 8.31 \\ & & & & & & 14 & 11.85 \\ & & & & & & & 16 \end{bmatrix}$	$\begin{bmatrix} 1 & 1.34 & 1.47 & 1.70 & 1.90 & 2.08 & 2.24 & 2.40 \\ & 4 & 3.28 & 3.39 & 3.79 & 4.16 & 4.49 & 4.80 \\ & & 6 & 4.64 & 4.65 & 5.09 & 5.50 & 5.88 \\ & & & 8 & 5.99 & 5.88 & 6.35 & 6.79 \\ & & & & 10 & 7.34 & 7.10 & 7.59 \\ & & & & & 12 & 8.68 & 8.31 \\ & & & & & & 14 & 10.03 \\ & & & & & & & 16 \end{bmatrix}$
$K = 10$	
$\begin{bmatrix} 1 & 0.93 & 0.64 & 0.41 & 0.25 & 0.15 & 0.08 & 0.05 & 0.03 & 0.01 \\ & 3 & 2.08 & 1.32 & 0.80 & 0.47 & 0.26 & 0.15 & 0.08 & 0.05 \\ & & 5 & 3.17 & 1.93 & 1.14 & 0.64 & 0.36 & 0.20 & 0.12 \\ & & & 7 & 4.25 & 2.52 & 1.41 & 0.79 & 0.45 & 0.26 \\ & & & & 9 & 5.33 & 2.99 & 1.67 & 0.96 & 0.55 \\ & & & & & 11 & 6.16 & 3.44 & 1.98 & 1.13 \\ & & & & & & 12 & 6.69 & 3.85 & 2.20 \\ & & & & & & & 13 & 7.48 & 4.27 \\ & & & & & & & & 15 & 8.56 \\ & & & & & & & & & 17 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.42 & 0.13 & 0.04 & 0.01 & 0 & 0 & 0 & 0 & 0 \\ & 3 & 0.93 & 0.26 & 0.07 & 0.02 & 0 & 0 & 0 & 0 \\ & & 5 & 1.42 & 0.39 & 0.10 & 0.03 & 0.01 & 0 & 0 \\ & & & 7 & 1.90 & 0.50 & 0.13 & 0.03 & 0.01 & 0 \\ & & & & 9 & 2.39 & 0.60 & 0.15 & 0.04 & 0.01 \\ & & & & & 11 & 2.76 & 0.69 & 0.18 & 0.04 \\ & & & & & & 12 & 3.00 & 0.77 & 0.20 \\ & & & & & & & 13 & 3.35 & 0.86 \\ & & & & & & & & 15 & 3.83 \\ & & & & & & & & & 17 \end{bmatrix}$

UN, unstructured model.

Table 2. Values of Fleishman's (1978)  $a$ ,  $b$ ,  $c$ , and  $d$  coefficients for each value of skewness and kurtosis for the distributions generated in the present study.

$\gamma_1$	$\gamma_2$	$a$	$b$	$c$	$d$
0.8	2.4	-.104049451	.848445836	.104049451	.044849610
0.8	5.4	-.082964688	.702207971	.082964688	.088904623
1.75	5.9	-.222093456	.774926306	.222093456	.054958336

$\gamma_1$ : value of the skewness coefficient;  $\gamma_2$ : value of the kurtosis coefficient.



Table 3. *F*-tests and Bonferroni post hoc tests for  $\epsilon = .57$ .

	<i>F</i>	<i>p</i>	$\eta_p^2$	Observed power	Post-hoc contrasts	<i>p</i>
<i>K</i> = 4						
<i>N</i>	114.785	<.001	.860	1	M=L	1
Distribution	3.926	.013	.174	.803	VS>S>M	<.001
					ND=SSD	1
					ND=MSD	.237
					SSD=MSD	.993
					SSD=LD	.136
					MSD=LD	1
					ND<LD	.021
<i>N</i> x Distribution	0.871	.556	.123	.385		
<i>K</i> = 6						
<i>N</i>	224.408	<.001	.923	1	M>L	.003
Distribution	7.405	<.001	.284	.979	VS>S>M	<.001
					ND=SSD	.486
					SSD=MSD	.645
					SSD=LD	.077
					MSD=LD	1
					ND<MSD	.007
					ND<LD	<.001
<i>N</i> x Distribution	0.537	.842	.079	.236		
<i>K</i> = 8						
<i>N</i>	259.989	<.001	.933	1	VS>S>M>L	<.001
Distribution	7.871	<.001	.297	.985	ND=SSD	.438
					SSD=MSD	.656
					SSD=LD	.053
					MSD=LD	1
					ND<MSD	.006
					ND<LD	<.001
<i>N</i> x Distribution	0.457	.897	.068	.202		
<i>K</i> = 10						
<i>N</i>	283.694	<.001	.938	1	VS>S>M>L	<.001
Distribution	4.284	.009	.187	.840	ND=SSD	.811
					ND=MSD	.053
					SSD=MSD	1
					SSD=LD	.475
					MSD=LD	1
					ND<LD	.010
<i>N</i> x Distribution	0.207	.992	.032	.109		

VS = very small sample size (*N* = 5, 6, 7, and 10), S = small sample size (*N* = 12, 14, 15, 18, and 21), M = medium sample size (*N* = 30, 45, 60, and 75), L = large sample size (*N* = 90, 100, 200, 300, and 500), ND = normal distribution, SSD = slightly skewed distribution, MSD = moderately skewed distribution, LD = lognormal distribution.

Table 4. *F*-tests and Bonferroni post hoc tests for  $\epsilon = .75$ .

	<i>F</i>	<i>p</i>	$\eta_p^2$	Observed power	Post-hoc contrasts	<i>p</i>
<i>K</i> = 4						
<i>N</i>	202.310	<.001	.916	1	M>L	.011
Distribution	8.627	<.001	.316	.991	VS>S>M	<.001
					ND=SSD	.352
					SSD=MSD	.757
					MSD=LD	1
					SSD<LD	.036
					ND<MSD	.006
					ND<LD	<.001
<i>N</i> x Distribution	0.677	.726	.098	.297		
<i>K</i> = 6						
<i>N</i>	257.837	<.001	.932	1	VS>S>M>L	<.001
Distribution	8.178	<.001	.305	.988	ND=SSD	.392
					SSD=MSD	.623
					MSD=LD	1
					SSD<LD	.046
					ND<MSD	.005
					ND<LD	<.001
<i>N</i> x Distribution	0.435	.910	.065	.193		
<i>K</i> = 8						
<i>N</i>	289.887	<.001	.940	1	VS>S>M>L	<.001
Distribution	8.153	<.001	.304	.988	ND=SSD	.410
					SSD=MSD	.594
					MSD=LD	1
					SSD<LD	.047
					ND<MSD	.005
					ND<LD	<.001
<i>N</i> x Distribution	0.373	.943	.057	.169		
<i>K</i> = 10						
<i>N</i>	315.138	<.001	.944	1	VS>S>M>L	<.001
Distribution	3.012	.038	.139	.679	ND=SSD	1
					ND=MSD	.157
					SSD=MSD	1
					SSD=LD	.756
					MSD=LD	1
					ND<LD	.046
<i>N</i> x Distribution	0.151	.998	.024	.091		

VS = very small sample size (*N* = 5, 6, 7, and 10), S = small sample size (*N* = 12, 14, 15, 18, and 21), M = medium sample size (*N* = 30, 45, 60, and 75), L = large sample size (*N* = 90, 100, 200, 300, and 500), ND = normal distribution, SSD = slightly skewed distribution, MSD = moderately skewed distribution, LD = lognormal distribution.

Table 5. Polynomial coefficients of the empirical bias for each level of  $K$  and  $\varepsilon$ .

$K$	$\varepsilon$	Linear		Quadratic		Cubic	
		Contrast estimate	$p$	Contrast estimate	$p$	Contrast estimate	$p$
4	.57	0.036	<.001	-0.015	<.001	0.002	.413
6	.57	0.116	<.001	-0.026	<.001	-0.002	.656
8	.57	0.159	<.001	-0.030	<.001	-0.004	.506
10	.57	0.186	<.001	-0.033	<.001	-0.007	.304
4	.75	0.110	<.001	-0.028	<.001	0	.944
6	.75	0.199	<.001	-0.037	<.001	-0.005	.501
8	.75	0.250	<.001	-0.037	<.001	-0.009	.280
10	.75	0.284	<.001	-0.038	<.001	-0.013	.159

Table 6. Forward multiple regression of the empirical bias using equation 1.

<i>K</i>	$\epsilon$	Regression coefficients					
		Constant		$b_1$	<i>N</i>		Distribution
		$b_0$	<i>p</i>		<i>p</i>	$b_2$	<i>p</i>
4	.57	-0.045	<.001	0.015	<.001	-0.003	.014
6	.57	-0.176	<.001	0.051	<.001	-0.009	<.001
8	.57	-0.251	<.001	0.070	<.001	-0.012	<.001
10	.57	-0.307	<.001	0.082	<.001	-0.010	.002
4	.75	-0.163	<.001	0.048	<.001	-0.010	<.001
6	.75	-0.315	<.001	0.087	<.001	-0.016	<.001
8	.75	-0.413	<.001	0.110	<.001	-0.019	<.001
10	.75	-0.491	<.001	0.125	<.001	-0.012	.006

$b_i$  = unstandardized estimated coefficients.

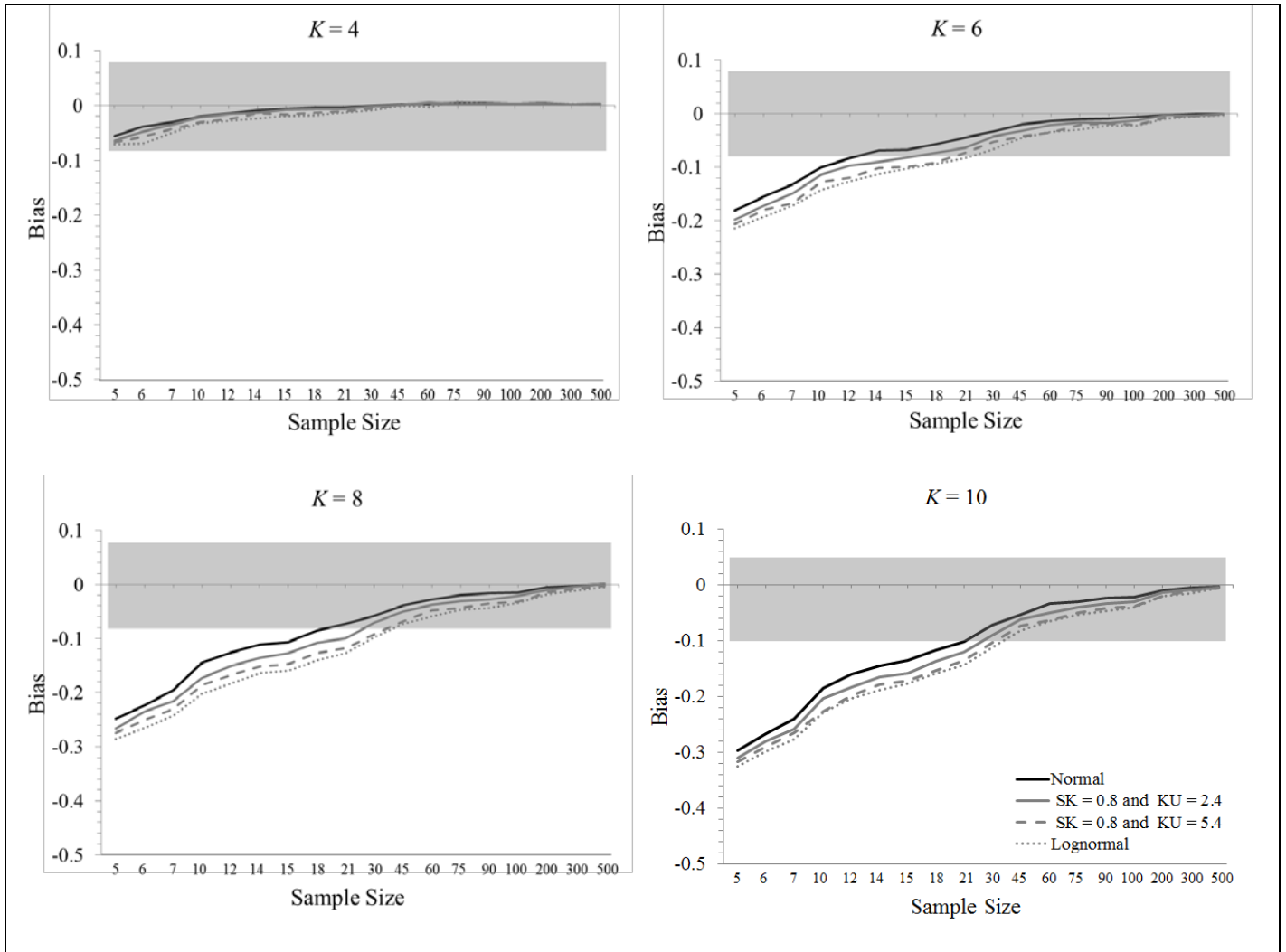


Figure 1. Empirical bias with  $\varepsilon = .57$  across the different  $N$ , distributions, and  $K$  (SK = Skewness and KU = Kurtosis).

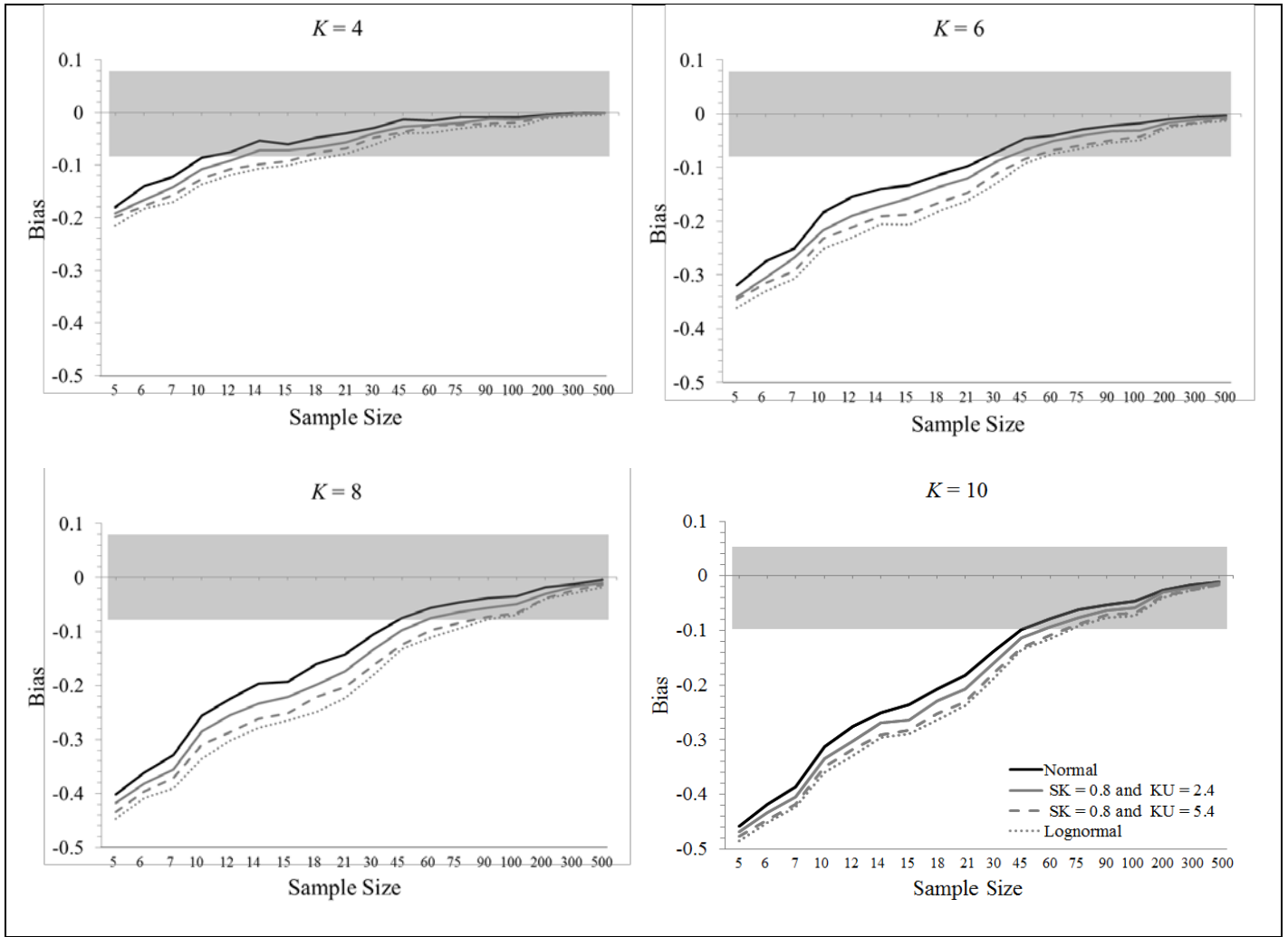


Figure 2. Empirical bias with  $\varepsilon = .75$  across the different  $N$ , distributions, and  $K$  (SK = Skewness and KU = Kurtosis).