# On Monte Carlo Simulation Algorithms for Research in Psychometrics

by

OSCAR LORENZO OLVERA ASTIVIA

B.A., University of the Fraser Valley, 2008
M.A., The University of British Columbia, 2013

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Measurement, Evaluation, and Research Methodology)

The University of British Columbia
(Vancouver)
March 2017

## Abstract

Monte Carlo simulations have become the workhorse of the modern methodologist aimed at providing both novel statistical insights and to guide data analysis practice. In spite of its widespread use, familiarity with data-generating algorithms is rare among users and consumers of simulation-based research, making the process appear as a “black box” of sorts. Without a good understanding of these algorithms, design flaws can appear in Monte Carlo studies which can influence the recommendations offered to applied researchers. In order to address these potential problems, this dissertation will highlight three issues in three separate papers related to the process of simulation as well as potential recommendations to deal with them. The first paper (chapter 2) focuses on the importance of matching the population model with the simulation design underlying the researcher’s hypothesis. It takes the Spearman rank correlation as a case study and documents the impact that potential disparities between simulation design and methodology can have on the conclusions derived from computer studies. The second paper (chapter 3) investigates a popular data-generating method within the social sciences, the Vale-Maurelli algorithm, and compares its results to a second one, the Headrick method, in terms of the kind of data they can generate and how this influences simulation results within a Structural Equation Modelling framework. The third paper (chapter 4) takes a closer look at the both the univariate (Fleishman) and multivariate (Vale-Maurelli) versions of the 3rd-order polynomial transformation to generate correlated, nonnormal data and documents the impact that its multiplicity of solutions has on simulation study results. In conclusion, this dissertation has the ultimate goal to help illuminate the process of simulation to psychometricians and social scientists alike in order to help create better study designs and promote a critical evaluation of Monte Carlo studies among methodologists and applied researchers alike.

## Preface

Chapter 2. A version of chapter 2 of this dissertation has been accepted for publication in the British Journal of Mathematical and Statistical Psychology under the title “Population Models and Simulation Methods: The case of the Spearman Rank Correlation”. My supervisor, Dr. Bruno D. Zumbo helped me formulate the research question and edit it to satisfy the recommendations of the editor and reviewers. I conceptualized and wrote the manuscript, ran the computer simulation and worked out the mathematics included in Appendices A, B and C. Tables 2.1, 2.2 and 2.3 and Figures 2.1 and 2.2 are taken from this accepted manuscript with permission of the authors, which are both my supervisor and myself.

Chapter 3. A version of chapter 3 of this dissertation has been published as Astivia, O. L. O., & Zumbo, B. D. (2015). A Cautionary Note on the Use of the Vale and Maurelli Method to Generate Multivariate, Nonnormal Data for Simulation Purposes. *Educational and Psychological Measurement, 75*, 541-567. doi: 10.1177/0013164414548894. Tables 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 and Figures 3.1 and 3.2 were taken with permission from this article authored both by my supervisor and myself. My supervisor, Dr. Bruno D. Zumbo aided me with the planning of the simulation studies and editing. I wrote the computer code needed to run this simulation studies and wrote the manuscript.

## Table of contents

## List of tables

## List of figures

## List of symbols and abbreviations

**Symbols** (by alphabetic order)

English Letters

$A$: random variable A
$\{a, b, c, d\}$: polynomial coefficients from 3rd-order transformation
$B$: random variable B
$\mathbf{C}$: covariance matrix
$\{c_0, c_1, c_2, c_3, c_4, c_5\}$: polynomial coefficients from 5th -order transformation
$d$: rank difference
$F_i(\cdot)$: cumulative distribution function for random variable *i*
$\mathbf{I}$: identity matrix
$n$: sample size
$\mathbf{P}$: Cholesky decomposition matrix of $\mathbf{C}$
$\mathbf{R}$: intermediate correlation matrix
$r$: sample Pearson correlation
$r_{Y_1 Y_2}$: final correlation from Vale-Maurelli transformation
$r_{Z_1 Z_2}$: intermediate correlation from Vale-Maurelli transformation
$U$: uniformly-distributed random variable
$u_i$: realization of uniformly-distributed random variable $U$
$\mathbf{w}$: weight vector with 3rd-order polynomial transformation coefficients
$\mathbf{X}$: matrix X
$X$: random variable X
$Y$: random variable Y
$\mathbf{z}$: vector z of standard-normally distributed variables
$Z$: standard normally distributed random variable

Greek Letters

$\mathbf{\Gamma}$: gamma matrix
$\gamma_1$: skewness
$\gamma_2$: excess kurtosis
$\gamma_3$: standardized 5th-order moment
$\gamma_4$: standardized 6th-order moment
$\mu$: population mean
$\rho_S$: Spearman rank correlation
$\rho$: population Pearson correlation
$\sigma$: population standard deviation
$\sigma_{ijkl}$: sample 4th-order moment
$\Phi^{-1}(\cdot)$: cumulative distribution function for standard normal distribution

**Abbreviations** (by alphabetic order)

ADF: Asymptotic Distribution Free
CF: Curran and Flora
CWF: Curran, West and Finch
FWMcK: Finch, West and McKinnon
H: Headrick
M: sample mean
Max: maximum
Mdn: sample median
Min: minimum
Mku: Mardia's kurtosis
ML: Maximum Likelihood
SB: Satorra-Bentler
SD: sample standard deviation
ST: Skidmore and Thompson
VM: Vale and Maurelli

**Dedication**

Whether you live in the $\mathbb{C}, \mathbb{I}, or\ \mathbb{R}$ planes, thank you for helping me finish this.

## Chapter 1: Introduction

The method of Monte Carlo simulation has, for all intents and purposes, revolutionized the way in which both the natural and social sciences conceptualize data analysis and evaluate best practices related to it. With the advent of accessible computer power, quantitatively-oriented researchers have gained the ability to generate complex multivariate data to serve a wide variety of purposes, from approximating solutions to problems that would otherwise be intractable to mimicking real-life phenomena in an attempt to understand the inputs and outputs of non-deterministic physical, biological, or social systems (Beisbart & Norton, 2012). Underlying this panoply of applications, however, there exists one important assumption: that the researcher, at every point and every moment, has a formal understanding of the simulation process. This understanding of the process is of utmost importance if one is to implement robust, well-designed studies and the less the process is understood, the more researchers become susceptible to unwarranted assumptions that may or may not influence the results of their simulations. It is the purpose of this dissertation to explore some of these assumptions and document the type of impact they can have when researchers use Monte Carlo simulations.

There is not one specific definition of a Monte Carlo simulation. Since its early inception in the 1950s by Stanislaw Ulam, Nicholas Metropolis, and John Von Neumann, this family of methods is characterized, in its most general conception, by the implementation of multiple repeated instances of a controlled random process (Jones, Maillardet & Robinson, 2009). Sawilowsky and Fahoome (2003) emphasize that the main feature that separates a Monte Carlo simulation from other types of stochastic modelling is its reliance on a large number of finite samples. Rubinstein and Kroese (2007) mention that any recurrent process with calculated uncertainty that the researcher oversees can be called a Monte Carlo simulation. Overall, these differences of definitions reflect the variety of uses that of simulations can have. Beisbar and Norton (2012) list three broad applications of the Monte Carlo method, one of which will be the focus of the investigations present within this dissertation. The first one is to approach issues of tractability and estimation, such as within the context of Monte Carlo integration or Markov Chain Monte Carlo (MCMC) samplers. Many modern-day applications of physics, mathematics and statistics rely on complex sets of equations (such as in Bayesian analysis) for which analytic results are either unknown or too convoluted to be practically implemented (e.g., Fishman, 2013). Being able to approximate solutions through a random simulation is one of the ways in which these problems become solvable. The second application of Monte Carlo simulations is to mirror real-life phenomena, such as particle dispersions (e.g. Elishakoff, 2003) or the spread of disease (e.g. Martínez-López, Ivorra, Ngom, Ramos & Sánchez-Vizcaíno, 2012). In these cases, a deterministic system that is highly-sensitive to its initial conditions (usually referred to as chaotic system) is modelled as if it were a random process in an attempt to representatively sample a collection of all the instances the system can end up in (Hoover & Hoover, 2015). In these cases, Monte Carlo simulations are also referred to as computer experiments for they attempt to mimic the same process of a real-life experiment, where the researcher systematically manipulates the conditions of the system and uses the computer to calculate the impact that these conditions can have (Sacks, Welch, Mitchell & Wynn, 1989). The third application (which is the most popular use among social scientists and the focus of this dissertation) is a hybrid between the previous two, the use of computer simulations to design and implement robustness studies of statistical methods. Carsey and Hardsen (2013) point out that even though statistical theory provides the framework to evaluate the assumptions of the methods used in day-to-day data practice, there is a wide variety of specific instances where the theory may be somewhat unclear. Issues such as the presence of outliers, violation of distributional assumptions or even the influence of sample size cannot always be analytically derived but their presence can exert undue influence in the use of statistical methods for data analysis. Monte Carlo simulation plays a key role in these instances, helping applied researchers choose the best method to analyze their data. The use of Monte Carlo simulations within the social sciences (particularly psychology and sociology) began in the early 1960s and became widespread with the availability of computer power. Johnson (2013) notes that even though its early use permeated all fields of psychology, the types of motivating problems that popularized simulations were on the area of methodology, particularly in the tradition of factor analysis and best practices surrounding the number of factors problem (e.g. Horn, 1965). Although other areas in the social sciences have benefited from theorizing within a Monte Carlo framework (e.g. Johansen, Savage, Fouquet & Shanks, 2015; Szolnoki & Perc, 2015), its most widespread applications have been on investigating and comparing different types of statistical methodologies aimed at improving data analysis practice. This type of simulation studies can be described in four general steps:

(1) Decide the models and conditions from which the data will be generated (i.e. what "holds" in the population).

(2) Generate the data.

(3) Estimate parameters for the models being studied under step (1)'s conditions.

(4) Save the parameter estimates, standard errors, goodness-of-fit indices, etc. for later analyses and go back to step (2).

Steps (2)-(4) would be considered a replication within the framework of a Monte Carlo simulation and repeating them a large number of times shows the patterns of behaviour of the statistical methods under investigation that will result in further recommendations for users of these methods.

This type of simulation studies usually emphasizes the decisions made in step (1) because the selection of statistical methods to test and data conditions will guide the recommendations that will subsequently inform data practice. Most of the time, steps (2) through (4) are assumed to operate seamlessly either because the researcher has the sufficient technical expertise to program them in a computer or because it is just assumed that the subroutines and algorithms employed satisfy the requests of the researcher.

When step (2) pertains to the generation of multivariate, non-normal data, a great variety of algorithms exist to achieve this goal. Broadly speaking, these algorithms can be categorized as either based on transformations from normal to non-normal distributions and those defined by theoretical non-normal distributions. Ruscio and Kaczetow (2008) calls the first class of methods "Transform-Calculate" (TC) methods because they involve a series of transformations from normality to non-normality and calculation of intermediate correlation matrices until the final correlation is achieved. TC-type algorithms subsume the Fleishman (1978) and Vale and Maurelli (1983) $3^{rd}$-order polynomial method, the Hedrick (2002) $5^{th}$-order polynomial method, the Pearson distribution system (Nagahara, 2004) and Ruscio and Kaczetow (2008)'s method, among others. Examples of the second broad class of algorithms include the multivariate skew-normal distribution (Azzalini & Dalla Valle, 1996) or copula distributions (Mair, Satorra & Bentler, 2012). These algorithms have seen their uses and applications mostly determined by area of study and not necessarily by the relative merits they have. For instance, the multivariate skew-normal distribution has seen most of its citations within statistics and econometrics journals whereas the $3^{rd}$ and $5^{th}$-order polynomial transform methods are almost exclusively cited within the social sciences (Kraatz, 2011). Due to this fact, this dissertation will focus primarily on polynomial-based methods in order to further document and inform researchers who might use them in their Monte Carlo simulations.

A crucial aspect of the implementation of these algorithms and of the performance of the simulation in general is the ability of the researcher to ensure that the simulation design and the actual computer implementation of it are consistent with one another. If this consistency is not there then step (2) is brought into question and one, either as a producer or consumer of simulation research, needs to wonder whether or not the conclusions obtained from the Monte Carlo studies are reliable. This issue constitutes the cornerstone of the present dissertation and the papers presented herein attempt to document the consequences of violating these often-overlooked assumptions.

Kraatz (2011) is perhaps the one of the very few people ever to acknowledge that the simulation work within the social sciences is sometimes conducted in a “blind” fashion , where the researcher is, to a certain degree, disconnected from the intermediate process that happens between the design of the simulation study and the final analysis of the data. Instead of calling the process “blind”, this dissertation opts to borrow the engineering concept of the “black box” (see Karlsson, Nellore & Soderquist, 1998) because it mirrors the idea of software development testing where a process (such an algorithm or an application) is only understood in terms of the inputs it is fed and the outputs it produces. The user is, for the most part, completely unaware of any actions taking place inside the “black box” and considers the outputs as valid based on nothing more than naïve faith in the process. In contrast, the conceptual methodology of this dissertation is that of a “glass box” or “open box” engineering model (see du Boulay, O'Shea & Monk, 1981), where there is a direct connection and full awareness between inputs, intermediate processes, and outputs. If there is any uncertainty or ambiguity in the intermediate processes, there is awareness of it and it is considered in the final analysis of the outputs. The overall goal of this dissertation is then to offer a “glass box” account of the simulation process within the psychometrics and the quantitative social sciences.

Two guiding principles will be analyzed in the subsequent studies presented in this dissertation in order to make an explicit connection from simulation design (inputs) to data-generating algorithms (intermediate processes) and the final results (outputs). The first one is the issue of matching statistical models to the actual simulation design or, in other words, ensuring that the computer is simulating what the researcher actually hypothesized in her or his simulation. To this effect, the case of the Spearman correlation is presented in chapter 2 and the argument is made that a large number of Monte Carlo studies within the social sciences have either explicitly or implicitly made the unwarranted assumption that the sample rank correlation estimates the population Pearson product-moment correlation. This study derives the population model for the Spearman rank correlation by relying on copula distribution theory and shows instances and conditions where both types of sample correlations estimate the same or different parameter values. To inform the methodology of simulation studies, it presents the Conover and Iman (1981) algorithm as a potential alternative to simulate data where the population parameter that governs the process is the rank correlation (as opposed to the product-moment correlation) and implements a simulation study looking at the properties of bias and Type I error rate for the *t* test and *r-to-z* transformation of the correlation, analyzing the differences in conclusions that arise when an algorithm that accurately matches the simulation design is used.

The second issue to be investigated (and the focus of the remaining two papers of this dissertation) is the unexplored characteristics of data-generation algorithms that can potentially impact the results of a simulation process or, in other words, ensuring that the simulation conditions the researcher specify are the actual conditions that are being generated by the computer. To this purpose, the Fleishman (1978) algorithm and its multivariate extension proposed by Vale and Maruelli (1983) are explored in chapter 3 given their tremendous popularity among social scientists and the influence they have had on the methodology of Monte Carlo simulation in this area. The Fleishman (1978) method is based on the 3rd-degree polynomial transformation of a standard normal variable with a fixed mean of 0, variance of 1 and skewness and kurtosis defined by the researcher. Since this algorithm relies on the Fisher-Pearson conceptualization of higher-order moments, the definitions used throughout this dissertation are:

$$\widehat{\gamma_1} = \frac{(1/n)\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left[(1/n)\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{3/2}} \,.$$

$$\widehat{\gamma_2} = \frac{(1/n)\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left[(1/n)\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{2}} - 3 \,.$$

where $\gamma_1, \gamma_2$ are the skewness and kurtosis, $x_i$ is a sample unit, $\bar{x}$ is the mean of the sample and $n$ is the sample size. Notice that since 3 is being subtracted from $\gamma_2$, the term "kurtosis" being used here is the same as "excess kurtosis", given that these definitions take the normal distribution as reference and the kurtosis of the normal distribution is 3 (Headrick, 2002).

First, the Fleishman (1978) and Vale-Maurelli (1983) algorithms are studied descriptively in terms of the quality of the data they can generate. A thorough investigation of their performance at different values of skewness and kurtosis from previously-published articles is conducted by calculating average bias and variability within a simulation study. Next, a second simulation study is done comparing the characteristics of the data generated by the Vale-Maurelli (1983) algorithm to those generated by the Headrick (2002) algorithm which is conceptually similar but uses a $5^{th}$-order polynomial transformation as opposed to a $3^{rd}$-order order one. Lastly, in chapter 4 of this dissertation, a peculiarity of the Fleishman (1978) method is explored related to the type of data it can generate. At the crux of this method is the solution to a system of five non-linear equations that yield the polynomial coefficients needed to induce non-normality on the data. The solutions to this system are not unique and different types of solutions yield data that may have similar univariate characteristics, but different multivariate ones. These differences are explored further as well as the potential impact that they can have in applied conclusions when this method is employed in robustness-type simulation studies.

Due to the myriad of uses of Monte Carlo simulations and the influence they have on informing data practice, their popularity has exploded both within and outside the social sciences. Every day, more and more researchers are learning how to conduct their own simulations to test their own hypotheses and explore the properties of their methods. Although this is certainly commendable, it is important to point out that, to a certain extent, it has come at the price of divorcing the theory that substantiates Monte Carlo simulations as a methodology and the actual practice in the design of computer studies. Understanding the formalities associated with data generation and model specifications are crucial to develop a solid simulation design that will generate reliable conclusions. The more this issue is overlooked, the more one risks conducting simulations at a merely procedural level without acknowledging the influence that this type of “methods effects” can have in computer experiments. The challenge is, however, that these effects can be very subtle and hide deceivingly in plain sight, all within the clear, sharp logic of computer code. By turning the method of simulation to the simulation itself starting from first principles and building it up to the actual computer code implementation, my dissertation looks to turn some of the hidden assumptions visible and open the black box for all to see. We might be surprised by what we will find.

# Chapter 2: Population models and simulation methods. The case of the Spearman rank correlation

## 2.1 Background

Spearman's rho is one of the first instances of non-parametric statistics that most researchers in the social sciences become familiar with. It has an intuitive conceptualization as the Pearson correlation between two ranked variables, it is readily available in most statistical software packages and provides an immediate alternative when the distributional assumption of normality is suspect. In spite of this seeming familiarity and common usage, there exist several important developments concerning this statistic which have yet to make their way into the quantitative behavioural sciences and which have contributed to generate some misconceptions within the methodological literature. As demonstrated in Hotelling and Pabst (1936), it can be shown that that when the definition of the Pearson correlation is applied to ranked data (expressed as positive integers *1, 2, 3, . . . , n*) with no ties, it results in

$$\hat{\rho}_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \tag{2.1}$$

where $d^2$ are the squared rank differences between variables and n is the total number of measured units (i.e., the sample size). This is the most commonly used computational formula to obtain the Spearman correlation and perhaps the most easily recognized one among introductory statistics textbooks in the social sciences. Its standard error and sampling distribution were extensively studied by Fieller, Hartley, and Pearson (1957) who found that, for the null case, $1/\sqrt{n-3}$ and the normal distribution are respectively good approximations as sample size grows arbitrarily large. For the non-null case, most work has relied on computer simulations which can be reviewed in Caruso and Cliff (1997), David and Mallows (1961) and Zimmerman, Zumbo, and Williams (2003).

An important issue that has been at least partially overlooked within the methodological literature on quantitative methods for psychology and education (and which serves as the guiding principle for this paper) is the question of what is known or understood of the Spearman rank correlation as a population parameter and not only as a descriptive sample statistic. Without a well-defined population model, it can be challenging to study the properties of statistics (such as bias) as well as the quality of the estimators used for them. The purpose of this paper is, therefore, to focus on the issue of matching the simulation design with the theoretical model implied by the researcher's hypotheses for the case of the Spearman rank correlation. Using analytic demonstrations and a simulation study, we highlight the impact of simulating Spearman's rho as if it were Pearson's r, and the potential shortcomings that arise from a mismatch between what is intended to be simulated and what is actually simulated.

This paper is organized as follows. Section 2.1.1 introduces the idea of a population definition for the Spearman correlation. Section 2.1.2 describes the parameterization of the rank correlation through a copula model and how it subsumes previous conceptualizations of this statistic. Section 2.1.3 introduces the Iman–Conover algorithm and posits it as one potential option (among others) of algorithms that can be selected to help match the simulation design with the population model of the Spearman rho. Section 2.2 presents a simulation study as a motivating empirical demonstration of the importance of having congruence between the population model and the simulation model employed by the researcher. Section 2.3 discusses the results from this simulation. Section 2.4 elaborates on recommendations for statistical researchers who use Monte Carlo simulation as a research methodology, as well as those readers who inform their data analysis practice based on simulation study outcomes.

#### 2.1.1 Towards a population model for the Spearman rank correlation

Generally speaking, the understanding that permeates the rank correlation is that it acts as either as an alternative estimator or a type of substitute for the Pearson product-moment correlation in the population (Christine & John, 2004; Hinton, McMurray, & Brownlow, 2014; Howitt & Cramer, 2005; Sapp, 2006). Spearman (1906) himself, for instance, mentioned that 'by using it [the rank correlation] we obtain a precise quantitative value which can be compared with that found for any other correlation under any other circumstances between any other things by the r method' (p. 104). He also comments on the fact that 'the values of these two symbols already, by definition, coincide with one another as regards their extreme upper and lower limits since both coefficients become 1 for perfect correlation and 0 for entire absence of correlation' (p. 101). It is, therefore, not surprising to find that a conflation between Spearman's rho and Pearson's r is prevalent both for methodological and applied researchers where the rank correlation is invoked every time the distributional assumptions needed for the product-moment correlation are suspect. The problem is that both types of correlations describe different types of relationships, rely on different assumptions and need not be the same even in the extreme cases of null or perfect correlation. For a simple example, assume $(X, Y)$ are sampled from a standard bivariate normal distribution such that $\mu_x = \mu_y = 0,\ \sigma_x = \sigma_y = 1,\ \ -1 \leq \rho \leq 1$. By letting $A = e^X$ and $B = e^Y$(where $e$ is the base of the natural logarithm) now $A$ and $B$ follow standard lognormal distributions. Because they are no longer normally distributed, the Pearson correlation between $A$ and $B$ is now restricted and no longer spans the $\pm 1$ range, regardless of what the initial population correlation $\rho$ is. To be more specific, this new standard lognormal distribution now has a correlation range restricted to $[-1/e, +1]$. Nevertheless, since exponentiation is a monotonic transformation, the Spearman rank correlation still spans the complete$\pm 1$ range, yielding a case where even through $\rho = -1$ is a perfectly plausible population value, the Pearson correlation is unable to capture this relationship due to the nature of the lower-dimensional marginal distributions. A proof of this result, as well as more detailed exploration of the boundaries of the product-moment correlation, is shown in Appendix A.

A population model for Spearman's rho was initially developed as part of the effort to theoretically substantiate non-parametric statistics and the rank transformation in general. Although early theoretical work explicitly assumes the existence of a population-level Spearman rho (Moran, 1948) and offers a proof of its existence (Hoeffding, 1948a,b), the implications of these results are rarely discussed within the quantitative social sciences. Kruskal (1958) was the first to explicitly define the population rank correlation in terms of a probability measure which he labelled q for quadrant association. This model, however, is not employed in the rest of the present paper. The copula-based model proposed in Schweizer and Wolff (1981) (which subsumes Kurskal's work on this statistic) was selected instead, given its flexibility of use, and ease of implementation for simulation studies.

### 2.1.2 A copula approach to the population the Spearman rank correlation

A thorough description of the theory behind copula distributions is beyond the scope of this paper[1], but a brief insight into these mathematical objects is needed in order to fully understand the population model of the rank correlation. Generally speaking, a copula is a type of multivariate distribution where all the marginals are uniformly distributed (Joe, 2014). An important advantage that copulas offer when modelling data is that they can capture the dependencies of any type of multivariate distribution by allowing the user to specify the dependence structure separately from its lower-dimensional marginal distributions (i.e., the user can couple marginal distributions in whichever way required). By employing the probability integral transformation, the marginal distributions can become any type of distribution (as long as its cumulative distribution function is well defined) and, as proved in Sklar's theorem (Sklar, 1959), they can be joined in whichever way is necessary to approximate any type of multivariate structure. In terms of more formal definition, assume $(X_1, X_2, X_3, \ldots, Xn)$ are jointly, continuously distributed random variables with well-defined cumulative density functions. By applying the probability integral transform one can obtain

[1] Interested readers can consult Nelsen (2010) or Joe (2014) for a more comprehensive introduction.

$$(U_1, U_2, U_3, \dots, U_n) = \big(F_1(X_1), F_2(X_2), F_3(X_3), \dots, F_n(X_n)\big), \tag{2.2}$$

where $U_i$ follows a standard uniform distribution and $F_i(\cdot)$ is a cumulative distribution function. The copula of $(X_1, X_2, X_3, \dots, Xn)$ is therefore defined as the joint cumulative distribution of $(U_1, U_2, U_3, \dots, U_n)$:

$$C(u_1, u_2, u_3, \dots, u_n) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2, U_3 \leq u_3, \dots, U_n \leq u_n), \tag{2.3}$$

so that $C(u_1, u_2, u_3, \dots, u_n)$ is the copula function, and $u_1, u_2, u_3, \dots, u_n$ are specific realizations of the random variables $(U_1, U_2, U_3, \dots, U_n)$. The copula function $C(\cdot)$ preserves the relationship among all the $(U_1, U_2, U_3, \dots, U_n)$ and by using the inverse of any suitable cumulative distribution function $F^{-1}(\cdot)$, the marginal distributions can obtain any arbitrary shape or property that may be needed.

For the purpose of obtaining the population model for the Spearman rank correlation, one last identity needs to be introduced. If the values of ranked variables (which are themselves natural numbers) are used in the sample definition of the Pearson product-moment correlation, it can be shown that

$$\widehat{\rho_s} = \frac{\sum_{i=1}^{n} r_{ix} r_{iy} - n[(n+1)/2]^2}{n(n^2-1)/12}, \tag{2.4}$$

where $r_{ix}, r_{iy}$ denote the ranks and *n* denotes the sample size. By letting $n \to \infty$, the sample rank correlation $\widehat{\rho_s}$ converges to

$$\rho_S = 12 \int C(u_1, u_2) du_1 du_2 - 3, \tag{2.5}$$

where $C(u_1, u_2)$ is the copula function as described as above. This same definition of the population Spearman rank correlation coefficient can be found in Schmid and Schmidt (2007), Nelsen (2010) and Joe (2014) and can be trivially derived by evaluating the double integral of equation (5) in Schweizer and Wolff (1981). The full derivation of equation (2.4) and a proof of convergence for the population Spearman rank correlation are shown in Appendix B.

If the Gaussian copula function (which in the two-variable case would be equivalent to the cumulative density function of the bivariate normal distribution) is used in (2.5) then the definition of Spearman's rho becomes the well-known identity that relates the rank correlation to the product-moment correlation:

$$\rho_S = \frac{6}{\pi} sin^{-1}\left(\frac{\rho}{2}\right), \tag{2.6}$$

where $\rho$ is the Pearson correlation parameter for the Gaussian copula. A proof of (2.6) is presented in Appendix C to help highlight the fact that for any given copula distribution function, a population version of the Spearman rank correlation can be obtained.

The use of a well-defined population model for the rank correlation not only allows researchers to explore the theoretical properties of this statistic, but also grants them the ability to conceive well-designed simulation studies such that issues like robustness or small-sample characteristics can be thoroughly investigated. In this particular case, being able to easily sample from copula distributions allows the user to choose simulation designs that actually match the intended research questions, where the data are generated using the Spearman rank correlation as a population parameter, and not just tangentially relying on the Pearson product-moment correlation. One such method that will be used throughout this paper is the Iman–Conover algorithm to generate rank-correlated data.

**2.1.3 Studying the properties of the estimator. A demonstration with the Iman–Conover algorithm**

There are several methods in the statistical literature for generating rank-correlated data. Although the methods recently introduced by Headrick et al. (Headrick, 2002, 2010; Headrick, Aman & Beasley, 2008; Headrick & Mugdadi, 2006; Headrick & Sawilowsky, 1999; Koran, Headrick & Kuo, 2015; Kowalchuk & Headrick, 2010; Pant & Headrick, 2015) show great potential, we have chosen to use the Conover and Iman (1981) approach as a way of demonstrating the point about the importance of population models.

This choice should not be interpreted as a recommendation of this method over any other, but rather as a reflection of the fact that the Conover and Iman (1981) method was developed earlier and, hence, is more widely used in the literature. It should be noted that our purpose is not to compare the various simulation methods, because that would take us away from the main point of the point of the paper: that a population model is important and that the simulation method needs to reflect the design implied by it.

Conover and Iman (1981) developed their approach after observing that the simulation methods available to researchers at the time were mostly limited to linear relationships and extensions of the multivariate normal distribution. They pointed out that the mathematics may become intractable for non-normal distributions and that if certain types of dependencies (such as those generated from stratified sampling) were to be induced, the current methods would simply not suffice. In order to address these issues, they derived a new method that allows for the correlation of any number of marginal distributions of any type by altering the rank order in which the elements from each distribution are being sampled. Altering the order of the elements does not change the properties of the distributions but it does influence the degree of dependency that they can have with reference to each other, making this method ideal to simulate variables with a given rank correlation. A description of the Iman–Conover algorithm follows.

Assume $\mathbf{X}$ to be an $n$-element sample from $r$-dimensional distribution. Consider the column vectors of $\mathbf{X}$ to be independently sampled so that its covariance matrix is an $r \times r$ identity matrix $\mathbf{I}$. Assume $\mathbf{C}$ to be the desired correlation matrix. Because $\mathbf{C}$ is positive definite, it has a Cholesky decomposition $\mathbf{C} = \mathbf{PP}'$(where $\mathbf{P}$ is some upper-level triangular matrix and $\mathbf{P}'$ is its transpose). The matrix product $\mathbf{XP}'$ will now have the desired correlation matrix $\mathbf{C}$ as a population parameter. Although this method is standard to simulate multivariate normal data with a given covariance matrix, the fact of the matter is that this procedure can be used with any marginal distribution, as long as the researcher keeps in mind that the properties of the marginal distributions (such as their moments) will change once the linear transformations on the last step are done (with the exception of the normal distribution, which is closed under linear transformations).

The issue becomes now how to ensure that $\mathbf{C}$ is a Spearman correlation matrix (and not just a Pearson correlation matrix) while keeping intact the properties of the intended marginal distributions. The solution provided by Conover and Iman is to find a set of numbers (referred by the authors as 'scores') which preserve a one-to-one relationship with the original elements sampled from the distributions and to assemble them in a different matrix with the same dimensions of the previously defined matrix $\mathbf{X}$ (the matrix of scores will be defined as $\mathbf{S}$ and is also $n \times r$). The said scores must have mean zero and standard deviation 1 (it suffices to ensure that for any collection of scores $s_i$, $\sum s_i = 0$ and $\sum s_i^2 = 1$) so that no further adjustments are needed for $\mathbf{C}$ to be a correlation and not a covariance matrix. There are various types of scoring schemes that can be implemented (such as scaled ranking or uniformly distributed $s_i$) but the method recommended by Conover and Iman is normal scoring (also known as van der Waerden scores) by using the cumulative distribution function of the Gaussian distribution $\Phi^{-1}(\cdot)$. For all the elements of the matrix $\mathbf{X}$, the elements in $\mathbf{S}$ are obtained by $s_i = \Phi^{-1}(i/(n+1))$ where $i$ is the respective rank of each $x_{ij}$. The algorithm then proceeds by randomly shuffling the columns of $\mathbf{S}$ to ensure that there are no correlations among the scores (in some implementations, both a shuffle of the columns and some matrix algebra are used so that the correlation matrix of $\mathbf{S}$ is an exact identity matrix). Once the score matrix has been generated and the necessary precautions have been taken so the correlations among the scores are 0, the standard Cholesky decomposition approach is taken to induce the desired correlations among the elements of $\mathbf{S}$ so that the matrix product $\mathbf{SP}'$ (with $\mathbf{P}$ defined as above) follows the desired correlation matrix $\mathbf{C}$. The last step (which induces the rank correlation) is then simply to rearrange the elements of $\mathbf{X}$ in the same rank order as the elements of $\mathbf{SP}'$. Once they are reordered, the matrix $\mathbf{C}$ (which would be a Pearson correlation matrix for the scores) becomes the population Spearman correlation for the originally intended variables in $\mathbf{X}$.

Although the authors themselves acknowledge that many of the properties of this method were unknown to them at the time of publishing, further research has demonstrated that, if normal scores are used, this algorithm becomes equivalent to sampling from a Gaussian copula (Mildenhall, 2006). Because copulas rely on measures of association which are invariant under monotonic transformations (which the Pearson correlation is not), they take in an association parameter (referred to as a 'generator') that can usually be transformed into either a Kendall's tau or a Spearman's rho (Joe, 2014).

There exist, however, a few limitations to this method that are important to point out. Since the scoring method is arbitrary, choosing different cumulative distribution functions (or a completely different scoring scheme) will result in different multivariate distributions, so researchers must be careful with their selection to ensure the scoring method matches their simulation design. A second issue is the use of an intermediate correlation matrix that needs to be calculated (the correlation matrix of $\mathbf{S}$ in the description above) before the final Spearman correlation intended by the researchers is attained. Headrick (2010) and Headrick and Pant (2012) have demonstrated that, since the Pearson correlation is not invariant under strictly increasing non-linear transformations, issues such as non-positive definiteness or out-of-bound correlations could occur.

Given the fact that the rank correlation does not always estimate the product-moment correlation (as illustrated in Section 2.1.1) and that, to the best of the authors' knowledge, very limited simulation work has been done where the value of Spearman's rho is set as the population parameter, the primary aim of this paper is to investigate the issue of what changes and what remains the same when the rank correlation is set as a parameter value within a broader simulation design. Rather than conduct a comprehensive study comparing the Pearson and Spearman correlation coefficients, the emphasis of this paper is on presenting brief examples from simulation conditions where the paradigm of the data being sampled from a bivariate normal distribution (where the population parameter is Pearson's $r$) is switched by sampling from a bivariate, Iman–Conover defined distribution (where the population parameter is Spearman's rho). Although exploring the small-sample properties of both types of correlations will be expanded upon, this goal is secondary to the primary aim of highlighting the impact that the theoretical simulation design has on the actual conclusions that can be derived from the study.

## 2.2 Method

Five papers published within social science journals were selected to evaluate the conclusions derived from simulation studies concerning the Spearman rank correlation. These articles were selected because of their focus on the performance of Spearman's rho under a variety of simulation conditions and because their high number of citations indicates that they have had an influence in shaping the understanding of this statistic within psychology and education. It is important to point out that, although Spearman's rho has been used for more than a century, the number of simulation studies focused on it within the social sciences is relatively limited, as pointed out in Bishara and Hittner (2012). Table 2.1 summarizes the simulation conditions found in each paper.

Based on this overview, the conditions chosen for this short study were sample sizes from 10 to 100 in increments of 10 ($n = 10, 20, 30, \ldots, 90, 100$) and population effect sizes from 0 to .9 in increments of .1 ($\rho = 0, .1, .2, .3, \ldots, .8, .9$). Some of the simulation studies in Table 2.1 also explored the impact that the non-normality of the unidimensional marginal distributions had on the correlation coefficients and, even though this is an important avenue of research, the present paper did not explore said conditions. A consistent conclusion from the papers reviewed here (and many others) is that nonnormality has an adverse effect on the Pearson correlation in terms of bias, Type I error rate and power. If non-normality had been used as a simulation condition, it could have masked the impact that the change in the data-generation paradigm had on the estimation of both types of correlations.

**Table 2.1** Summary chart of the simulation conditions per article

| Article | Population effect size | Sample size | Distributions |
|---|---|---|---|
| Rupinski & Dunlap (1996) | .1 to .9 | 5 to 100 | Normal |
| Caruso & Cliff (1997) | 0,.3, .45,.6,.75, .9 | 10, 50, 200 | Normal |
| Zimmerman, Zumbo & Williams (2003) | 0,.1,.3,.5,.7,.9 | 10, 20, 40 | Normal, Exponential, Log-normal, Uniform |
| Wilcox & Tian (2010) | .1,.3,.4,.6,.7 | 20, 50, 100 | Normal, Uniform, Exponential |
| Bishara & Hittner (2012) | 0,.1,.5 | 5, 10, 20 40, 80, 160 | Normal, Weibull, Chi-square, Uniform, Bi-modal, Long-tailed |

This creates a 10 (sample sizes) × 10 (effect sizes) × 2 (data-generation mechanisms) study design, where 10,000 replications per combination of conditions were implemented to ensure maximum stability. The process of simulation and analysis of the generated data were conducted in the R statistical programming environment (R Development Core Team, 2015), using the MASS package (Venables & Ripley, 2002) to generate bivariate normal data and the mc2d package (Pouillot & Delignette-Muller, 2010), which has a function to implement the Iman–Conover algorithm.

The outcomes of this simulation were the average bias of the sample estimates of the Spearman and Pearson correlations, the Type I error rate and power of Pearson's t-test and Fisher's r-to-z transformation. All of the articles in Table 2.1 explored at least one of these two classical significance tests and only the Zimmerman et al. (2003) study focused on the small sample bias of the correlation estimates. It is important to keep in mind, though, that the ultimate goal of the present paper is not to redo published simulation work or document the potential impact that the choice of data-generation algorithm has on the quality of the data, but to understand what types of claims regarding each type of correlation still hold when one uses the theoretically appropriate model to simulate data that matches the study design. For instance, the Zimmerman et al. (2003) study on small-sample bias comments on the fact that the sample Spearman rank correlation coefficient is a more biased estimate than the Pearson product-moment correlation. But the parameter that defined the population of the bivariate distributions in their study (and against which bias is evaluated) is the Pearson correlation coefficient, not Spearman's rho. Would it be sensible to assume, then, that if bias is calculated with respect to a population version of the Spearman rank correlation, the sample estimate of the Pearson correlation is actually more biased? Similar claims were made in the other four articles, where the performance of the Spearman rank correlation is evaluated without having a population-defined model from which the data were sampled.

## 2.3 Results

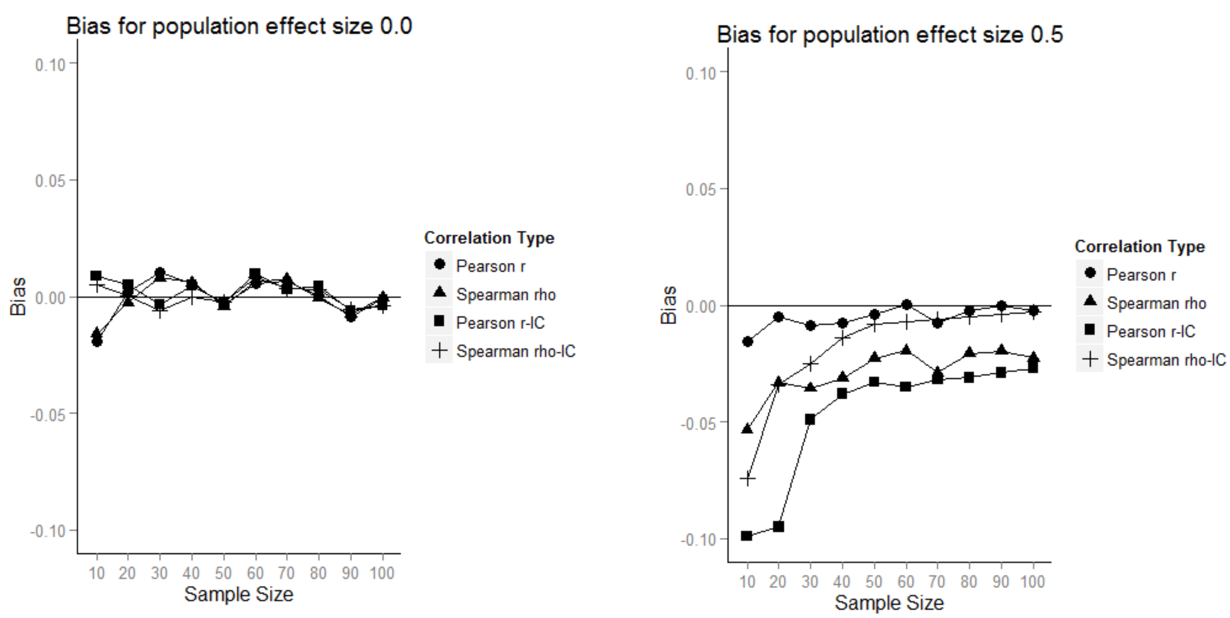
Simulation results in Table 2.2 show evidence of the small-sample bias of the correlation coefficient present both in the product-moment correlation coefficient and the rank correlation coefficient. Generally speaking, both types of correlations underestimate the population parameter value, albeit that the size of the bias becomes smaller as the sample size increases. For the largest effect size, however, the rank correlation coefficient estimated on Iman–Conover sampled data overestimated its population value. It is important to highlight that the cases where the data-generation process matches the type of correlation being estimated (so the Norm columns with the r rows, or the rho columns with the IC rows) showed both the smallest bias and the fastest convergence towards the population effect size.

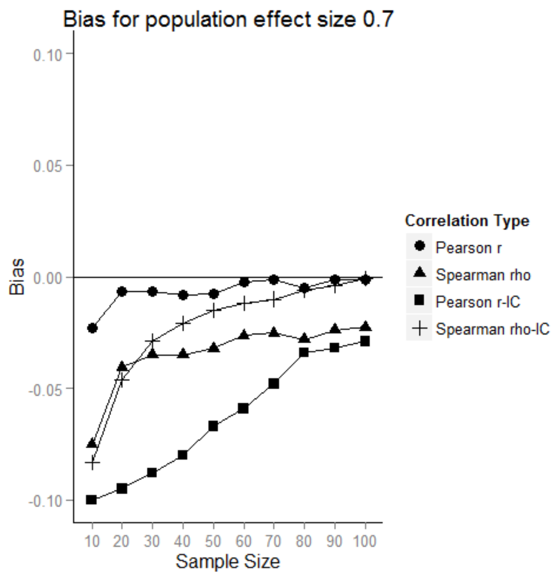
Figure 2.1 shows the mean bias for four effect sizes across all sample size conditions in the simulation. Although biased estimates were found across all different population effect sizes, this bias became especially pronounced when the effect size was large, particularly for the conditions where the data-generation mechanism did not match the type of correlation being calculated. When either the

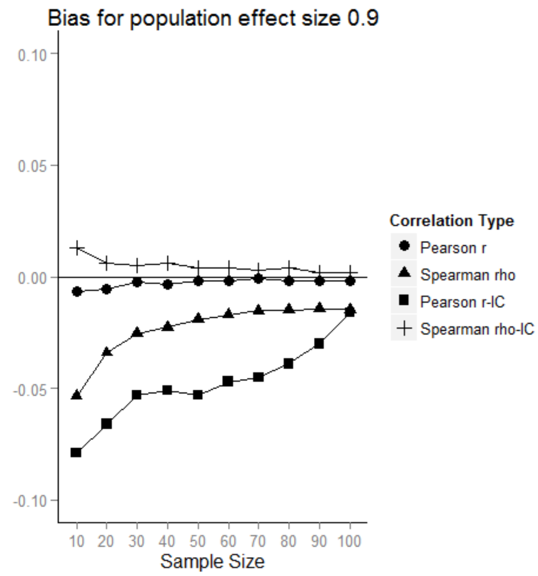
**Table 2.2** Mean correlation estimates for the Pearson ($r$) product-moment and Spearman ($\rho_S$) rank correlation across 10,000 replications per condition. Data were sampled from either a bivariate normal (Norm) or Iman-Conover (I-C) distribution using a subset of population effect sizes (ES) and sample sizes ($n$) from the simulation conditions.

| | ES | 0 | | 0.1 | | 0.5 | | 0.7 | | 0.8 | | 0.9 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| n | | Norm | I-C | Norm | I-C | Norm | I-C | Norm | I-C | Norm | I-C | Norm | I-C |
| | $r$ | -.019 | .009 | .102 | .009 | .484 | .401 | .677 | .600 | .782 | .712 | .894 | .821 |
| 10 | $\rho_S$ | -.016 | .005 | .089 | .076 | .397 | .426 | .625 | .617 | .729 | .782 | .847 | .913 |
| | $r$ | .002 | .005 | .096 | .019 | .495 | .405 | .694 | .605 | .793 | .739 | .894 | .834 |
| 20 | $\rho_S$ | -.002 | -.004 | .087 | .057 | .467 | .466 | .660 | .654 | .763 | .794 | .866 | .906 |
| | $r$ | .011 | -.006 | .097 | .061 | .492 | .451 | .694 | .612 | .792 | .751 | .897 | .847 |
| 30 | $\rho_S$ | .008 | -.003 | .091 | .088 | .464 | .475 | .665 | .671 | .762 | .797 | .875 | .905 |
| | $r$ | -.003 | -.003 | .106 | .076 | .496 | .467 | .693 | .633 | .798 | .761 | .898 | .857 |
| 50 | $\rho_S$ | -.004 | -.003 | .094 | .104 | .477 | .492 | .668 | .685 | .777 | .798 | .881 | .904 |
| | $r$ | .000 | .002 | .103 | .089 | .497 | .469 | .695 | .666 | .797 | .769 | .899 | .861 |
| 80 | $\rho_S$ | -.001 | .002 | .105 | .092 | .479 | .495 | .672 | .694 | .776 | .802 | .885 | .904 |
| | $r$ | -.001 | -.001 | .100 | .094 | .498 | .473 | .699 | .671 | .798 | .771 | .900 | .884 |
| 100 | $\rho_S$ | .000 | .000 | .097 | .101 | .477 | .497 | .677 | .699 | .781 | .802 | .885 | .902 |

**Figure 2.1** Mean bias across sample sizes for population effect sizes of 0.0, 0.5, 0.7 and 0.9. The "IC" abbreviation implies the data was sampled from an Iman-Conover-generated distribution. No abbreviation implies the data was sampled from a normal distribution





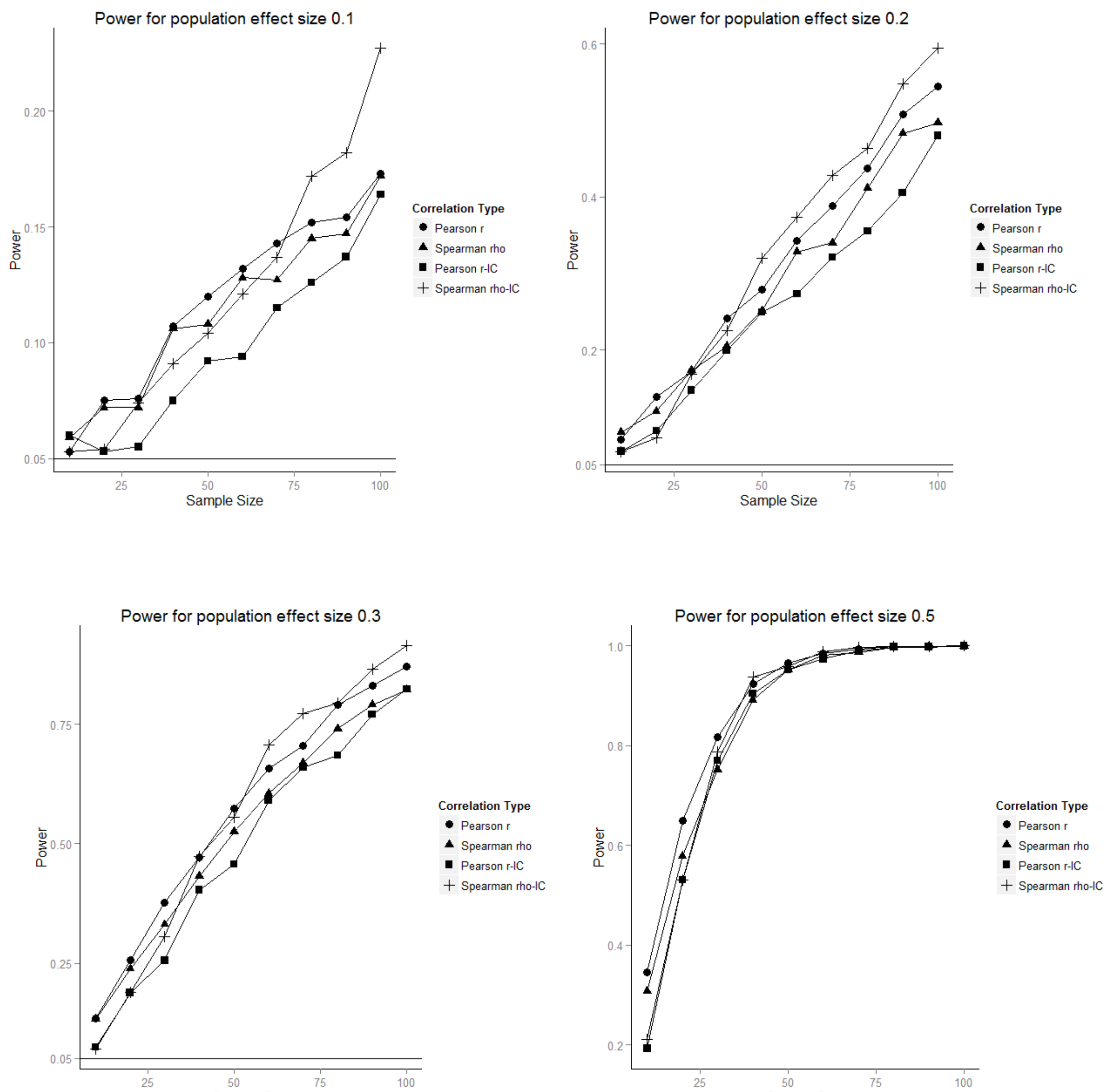

and the estimates reached an upper asymptote, not converging to the population effect size even at the largest sample size of 100. Overall, the Pearson correlation calculated on bivariate normal data showed the least amount of bias and fastest convergence towards its parameter value, but it also showed the largest bias and upper asymptote when calculated on Iman–Conover sampled data.

The Type I error rate of both data-generation mechanisms is presented in Table 2.3, calculated on both the standard t-test and the r-to-z Fisher transformation. Overall, it can be said that when the data-generation mechanism matches the statistic being calculated, the empirical error rate converges faster to the nominal 5% than when there is no match. The most extreme cases of Type I error rate inflation happened at the smallest sample size of 10, but it quickly reached its nominal rejection rate, particularly at the largest sample sizes. In spite of this, when these results are evaluated according to the criterion presented in Bradley (1978), one can point out that no considerable inflation or deflation of the error rate was found, regardless of the simulation conditions or which test statistic was being calculated. The power of the t-test and r-to-z transformation shows a remarkably similar pattern, with almost identical results to those found previously in the literature (e.g., Bishara & Hittner, 2012). Because of this, the comments on this section apply to both tests. Generally speaking, trends in empirical power were also contingent on whether the type of correlation being calculated matched the data-generation mechanism. As shown in Figure 2.2, the Spearman rank correlation computed on Iman–Conover sampled data outperformed all other correlations/data mechanisms for small population effect sizes at larger sample sizes. But for both small effect and small samples, the Pearson product-moment correlation yielded the highest power in every case. As the effect size in the population becomes larger, the power curves become closer and closer together (particularly for large sample sizes), to the point that they become indistinguishable. It was found that for population effect sizes of .5 and larger, the differences between estimated powers are sufficiently small to be considered negligible.

**Table 2.3** Empirical type 1 error rates (α=.05) for both bivariate normal (Norm) and Iman-Conover algorithm (I-C) simulated data at different sample sizes (*n*) both for the Pearson (*r*) and Spearman (*rho*) correlation coefficients.

| | | *t* | | *r-to-z* | |
|---|---|---|---|---|---|
| n | | Norm | I-C | Norm | I-C |
| | $r$ | .045 | .062 | .045 | .065 |
| 10 | $\rho_S$ | .048 | .060 | .058 | .059 |
| | $r$ | .046 | .048 | .048 | .044 |
| 20 | $\rho_S$ | .047 | .044 | .056 | .047 |
| | $r$ | .048 | .046 | .047 | .046 |
| 30 | $\rho_S$ | .050 | .045 | .057 | .048 |
| | $r$ | .049 | .057 | .048 | .057 |
| 50 | $\rho_S$ | .052 | .056 | .053 | .055 |
| | $r$ | .048 | .054 | .049 | .052 |
| 80 | $\rho_S$ | .051 | .051 | .051 | .052 |
| | $r$ | .050 | .052 | .051 | .054 |
| 100 | $\rho_S$ | .051 | .050 | .052 | .050 |

**Figure 2.2** Empirical power across sample sizes for population effect sizes of 0.1, 0.2, 0.3 and 0.5. The “IC” abbreviation implies the data was sampled from an Iman-Conover-generated distribution. No abbreviation implies the data was sampled from a bivariate normal distribution. A horizontal line at the .05 level is included as reference



## 2.4 Discussion

The main purpose of this paper was to address two goals, one pertaining specifically to the simulation studies of the Spearman rank correlation coefficient, and a more general one aimed at informing the methodology of Monte Carlo simulations more broadly within the behavioural sciences. The theoretical derivations and simulation study results presented here were aimed at reaching both goals, helping psychologists and educational researchers with a qualitative bent to become familiar with some of the formalism regarding the rank correlation and how this formalism helps guide the design and implement simulation studies.

Regarding Spearman's rho, the formulation of its population model, as shown in equation (2.5), helps highlight the idea that this type of correlation estimates a parameter with properties that may or may not necessarily align themselves with those of the Pearson product-moment correlation coefficient (Borkowf, 2002). For instance, even though identity (2.6) explicitly relates the Pearson and Spearman correlations, this only happens under the assumption of bivariate normality (i.e., choosing a Gaussian copula as the integrand in (2.5)). Changing the copula distribution function to a different one could either result in a different identity or no identity at all, if a closed-form expression does not exist. Under this paradigm, it makes sense to bring forward the idea that treating the rank correlation as a 'robust' estimator of the Pearson correlation – as sometimes presented or implied in the literature (e.g., Hinton et al., 2014) – is not always warranted, particularly for cases where the Pearson correlation may not even be theoretically feasible, as shown in Appendix A. In spite of this, it is possible to find cases in the literature where the Pearson and Spearman correlations are treated as if they always referred to the same parameter in the population. For instance, even though it was not explicitly stated, the simulation methodologies present in the papers summarized in Table 2.1 make the implicit equivalence between the rank correlation and the product-moment correlation at the population level, which is not always true since, in many instances, non-normal marginal distributions were used as simulation conditions. To the authors' knowledge, the present paper is one of the very few instances within the quantitative behavioural sciences literature where theoretical model underlying the Spearman rank correlation matches the data-generation mechanism and, as shown in Section 2.3, simulation study results can change depending on which simulation design is used. Claims such as 'even for larger sample sizes, the correlation based on ranks remains biased and apparently does not approach zero asymptotically as does the correlation based on scores' (Zimmerman et al., 2003, p. 144) would need further re-evaluation since, as demonstrated here, the nature of the small-sample bias changes depending on which population effect size is used to generate the synthetic data.

Regarding the practice and use of simulation methodology within the behavioural sciences, the present authors believe that the population model for the rank correlation offers a good and easy-to-understand opportunity to help remind quantitative researchers who both conduct and make use of simulation studies to always consider the type of Monte Carlo design that underlies the conclusions present in this type of literature. No simulation study (or scientific study in general) exists without its limitations, and being aware of them from a technical and theoretical standpoint helps create better simulation studies to inform day-to-day data analysis practice. For instance, Astivia and Zumbo (2015) showed empirically, through simulation results, that choosing different data-generating algorithms for multivariate, non-normal data can lead to different conclusions when re-examining previously published papers, even if all conditions are kept the same and only the algorithm changes. It is important to point out that this is not a claim to consider previously published simulation results as wrong or invalid. It is merely an attempt to highlight the possibility that other (perhaps non-standard) simulation designs exist and that they could provide a more comprehensive understanding of the conditions in which certain statistical methodologies perform. In this specific simulation demonstration, for example, hypothesis tests based on normal theory performed relatively well (in terms of power) when the Spearman rank correlation was calculated on Iman–Conover sampled data. But they did not perform as well when calculated on bivariate normal data, as shown here, on Zimmerman et al. (2003) and Bishara and Hittner (2012). So if one is applying normal-theory tests on the rank correlation when the inference pertains only to the population rank correlation, their use is not entirely unwarranted (given the simulation conditions studied). But when the inferences go from the sample rank correlation to the population product-moment correlation coefficient, the reduction in power (which has been well documented in the literature) becomes apparent. Therefore, previously published simulation results and the conclusions presented here are consistent with each other (as expected), with the added benefit that new information was discovered when extending the original theoretical simulation designs by using algorithms that match the proper population model for the parameter being investigated (i.e., the bivariate normal data for the Pearson correlation and the Iman–Conover data for the Spearman correlation).

There still exist many future avenues of research to pursue within the theoretical framework of the Spearman correlation that have been left unexplored in this paper. Very little is known within the behavioural sciences about the properties of the Spearman correlation matrix, even though multivariate extensions to this coefficient have been already developed (see Schmid & Schmidt, 2007). Other types of algorithms (such as the ones developed by Headrick et al. that were mentioned previously) could also be employed to explore other types of correlational structures that still take the rank correlation as a population parameter, in order to see which conclusions from the published literature could be changed or expanded upon. Generally speaking, not much research is devoted to understanding the Spearman rank correlation (or other types of non-parametric measures of association), but with a clear understanding of its mathematical formalism and the algorithms that exist to design proper simulation studies it will possible to expand the research surrounding this statistic as well as making more comprehensive recommendations to practitioners and applied researchers.

# Chapter 3: A cautionary note on the use of the Vale and Maurelli method to generate multivariate, nonnormal data for simulation purposes

## 3.1 Background

With the increasing availability of computer power, Monte Carlo simulations have become one of the most popular methods to explore the robustness of statistical procedures to the less-than-ideal conditions found in applied research (Bandalos & Leite, 2013). In order to be able to properly conduct simulation studies, it is important for the researcher to use algorithms that accurately reflect the conditions being studied or the conclusions from said simulation results would be suspect. To investigate the effect of violating distributional assumptions (mostly, normality) for multivariate methods, simulation studies usually require the implementation of a data-generating procedures that allow the researcher to control the correlation or covariance structure of the data as well as the degree of nonnormality (Mattson, 1997; Ruscio & Kaczetow, 2008; Vale & Maurelli, 1983). Various methods have been proposed to this effect (e.g., Azzalini & Dalla Valle, 1996; Headrick, 2002, etc.) but, arguably, none have been as widely used within the methodological literature as the Vale and Maurelli (1983) multivariate extension of the Fleishman (1978) power method.

### 3.1.1 Overview of the Fleishman–Vale–Maurelli method (third-order polynomial)

The Fleishman method or (third-order polynomial approach) employs a polynomial transformation of standard, normal variables where the polynomial coefficients are used to specify the first four moments of a distribution. Let $Z \sim N(0,1)$ and define

$$Y = a + bZ + cZ^2 + dZ^3 \tag{3.1}$$

where $\{a, b, c, d\}$ are the polynomial coefficients that will control the first four order moments of the new random variable $Y$. For appropriately chosen coefficients, the researcher can control the skewness and kurtosis of $Y$ taking advantage of the following relationships.

Let us define the first four moments of $Y$ as $\mathbb{E}[Y] = 0$, $\mathbb{E}[Y^2] = 1, \mathbb{E}[Y^3] = \gamma_1, \mathbb{E}[Y^4] = \gamma_2$ (where $(\gamma_1, \gamma_2)$ are the skewness and kurtosis of $Y$). Then it follows that

$$a = -c \tag{3.2}$$

$$b^2 + 6bd + 2c^2 + 15d^2 = 1 \tag{3.3}$$

$$2c(b^2 + 24bd + 105d^2 + 2) = \gamma_1 \tag{3.4}$$

$$24(bd + c^2[1 + b^2 + 28bd] + d^2[12 + 48bd + 141c^2 + 225d^2]) = \gamma_2 \tag{3.5}$$

where equations (3.2) to (3.5) correspond to Fleishman's equations (5), (11), (17), and (18), respectively.

Because of its relative ease of implementation, the Fleishman (1978) method was extended by Vale and Maurelli (1983) to the multivariate case by applying the algorithm to each one-dimensional marginal in the multivariate distribution being defined. An undesirable aspect of this process is that it changes the correlation structure intended by the user, so a correction needs to be implemented before proceeding. The solution proposed in Vale and Murelli (1983) is to calculate an intermediate correlation matrix so that the data takes it as the initial ''population'' correlation matrix and, as one applies the Fleishman (1978) method to each marginal distribution, the correlation matrix is transformed to the originally intended one. Define the vectors $\mathbf{z}$ and $\mathbf{w}$ and the new variable $Y$as

$$\mathbf{z}' = [1, Z, Z^2, Z^3] \tag{3.6}$$

$$\mathbf{w}' = [a, b, c, d] \tag{3.7}$$

$$Y = \mathbf{w}'\mathbf{z} \tag{3.8}$$

where $Z$ is defined as above and the vector $\mathbf{w}$ contains the polynomial weights needed to control the moments of the nonnormal distribution $Y$. Letting $r_{Y_1 Y_2}$ be the correlation coefficient of two nonnormal variables $Y_1$ and $Y_2$ generated from the two normally distributed variables $Z_1$ and $Z_2$ (with a correlation coefficient $r_{Z_1 Z_2}$) it follows that

$$r_{Y_1Y_2} = E(Y_1Y_2) = E(\mathbf{w'}_1\mathbf{z}_1\mathbf{z'}_2\mathbf{w}_2) = \mathbf{w'}_1E(\mathbf{z}_1\mathbf{z'}_2)\mathbf{w}_2 = \mathbf{w'}_1\mathbf{R}\mathbf{w}_2 \tag{3.9}$$

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & r_{Z_1Z_2} & 0 & 3r_{Z_1Z_2} \\ 1 & 0 & 2r_{Z_1Z_2}^2 + 1 & 0 \\ 0 & 3r_{Z_1Z_2} & 0 & 6r_{Z_1Z_2}^3 + 9r_{Z_1Z_2} \end{bmatrix} \tag{3.10}$$

And by collecting the appropriate terms from the matrix $\mathbf{R}$ in equation (3.10) it can be shown that the correlation between $Y_1$ and $Y_2$ is

$$r_{Y_1Y_2} = r_{Z_1Z_2}(b_1b_2 + 3b_1d_2 + 3d_1b_2 + 9d_1d_2) + r_{Z_1Z_2}^2(2c_1c_2) + r_{Z_1Z_2}^3(6d_1d_2) \tag{3.11}$$

By solving for $r_{Z_1Z_2}$, it is possible to find the pairwise elements of the intermediate correlation matrix so that the researcher can specify all the elements $r_{Y_1Y_2}$ that will act as the desired and final correlation matrix used to generate the data.

### 3.1.2 Overview of the Headrick method (fifth-order polynomial)

Tadikamalla (1980) published one of the first overviews of algorithms to generate nonnormal data that let the user specify the higher order moments of a distribution. As a criticism, he pointed out that the probability density function that arises from implementing the Fleishman (1978) method is not known and that various combinations of skewness and kurtosis are not possible (a boundary exists that relates the values of kurtosis and skewness as $\gamma_2 \geq \gamma_1^2 - 2$ and Fleishman's method can only obtain a subset of the values in it). Nevertheless, he did recommend the third-order polynomial method because of its ease of implementation.

To address some of the previously mentioned limitations, Headrick (2002) proposed an extension of the power method where polynomials of Order 5 (instead of those of Order 3) are implemented to obtain a wider range of values in the skewness–kurtosis parabola. Again, taking $Z \sim N(0,1)$ the Headrick (2002) approach defines the new, nonnormal variable $Y$ as

$$Y = c_0 + c_1Z + c_2Z^2 + c_3Z^3 + c_4Z^4 + c_5Z^5 \tag{3.12}$$

As expected, the increase in the number of polynomial coefficients also increases both the number and the complexity of the equations that need to be solved in order to obtain the estimates of $\{c_0, c_1, c_2, c_3, c_4, c_5\}$. The general equations (which are quite lengthy) as well as the technical details can be found in Headrick (2002), but the general logic rests on the idea of moment-matching the new distribution of $Y$ to the skewness/kurtosis values defined by the user and needed to solve Equation (3.12). The more moments one can control, the better the approximation of $Y$ will be. By using the fifth- and sixth-order moments, the user now has the ability to choose from a wider range of skewness and kurtosis values and can better reproduce well-known distributions whose theoretical higher order moments are known.

The fifth-order polynomial approach has been extended to the multivariate case as well, following the same premise of finding an intermediate correlation matrix for the data before the nonnormal transformations are carried out (Headrick, 2002, 2004; Headrick & Kowalchuk, 2007). For two nonnormal variables $Y_1$ and $Y_2$ generated from two standard normally distributed variables $Z_1$ and $Z_2$ with a correlation coefficient $r_{Z_1Z_2}$, the resulting correlation of the $Y$ variables is

$$
\begin{aligned}
r_{Y_1Y_2} = {} & 3c_{4(1)}c_{0(2)} + 3c_{4(1)}c_{2(2)} + 9c_{4(1)}\, c_{4(2)} + \; c_{0(1)}\big(\, c_{0(2)} + c_{2(2)} + 3\, c_{4(2)}\big) \\
& + c_{1(1)}\, c_{1(2)}\, r_{Z_1Z_2} + 3c_{3(1)}c_{1(2)}r_{Z_1Z_2} + 15c_{5(1)}c_{1(2)}r_{Z_1Z_2} \\
& + 3c_{1(1)}c_{3(2)}r_{Z_1Z_2} + 9c_{3(1)}c_{3(2)}r_{Z_1Z_2} + 45c_{5(1)}c_{3(2)}r_{Z_1Z_2} \\
& + 15c_{1(1)}c_{5(2)}r_{Z_1Z_2} + 45c_{3(1)}c_{5(2)}r_{Z_1Z_2} + 225c_{5(1)}c_{5(2)}r_{Z_1Z_2} \\
& + 12c_{4(1)}c_{2(2)}r_{Z_1Z_2}^2 + 72c_{4(1)}c_{4(2)}r_{Z_1Z_2}^2 + 6c_{3(1)}c_{3(2)}r_{Z_1Z_2}^3 \\
& + 60c_{5(1)}c_{3(2)}r_{Z_1Z_2}^3 + 60c_{3(1)}c_{5(2)}r_{Z_1Z_2}^3 + 600c_{5(1)}c_{5(2)}r_{Z_1Z_2}^3 \\
& + 24c_{4(1)}c_{4(2)}r_{Z_1Z_2}^4 + 120c_{5(1)}c_{5(2)}r_{Z_1Z_2}^5 + c_{2(1)}(c_{0(2)} + c_{2(2)} + 3c_{4(2)} \\
& + 2c_2r_{Z_1Z_2}^2 + 12c_4r_{Z_1Z_2}^2)
\end{aligned}
\tag{3.13}
$$

So by substituting the desired correlation in $r_{Y_1\,Y_2}$ and obtaining the appropriate polynomial coefficients, it is possible to solve for $r_{Z_1Z_2}$ and obtain the intermediate correlation matrix.

Both theoretical and simulation results have shown that the fifth-order polynomial method can, indeed, obtain estimates of skewness and kurtosis that cannot be achieved through the implementation of third-order polynomials (Headrick 2002, 2004). It has also been shown that the range of possible correlation values allowed by using the Headrick (2002) method is much wider than when using the Vale and Maurelli's (1983) method (Headrick, 2004). The polynomial coefficients place boundaries on the correlational structure that each method can generate, so it is not unusual to find that certain correlations (particularly on the higher or lower ranges) cannot be reproduced when certain combinations of skewness or kurtosis are present, particularly if one variable is severely skewed and the other one is symmetric (Mair, Satorra, & Bentler, 2012). It is also relevant to point out that since the intermediate correlation matrix is estimated in a pairwise fashion, there is no guarantee that it will be positive definite once it is fully assembled (Fan, Sivo, & Keenan, 2002; Li & Hammond, 1975). Both polynomial methods have this problem, but Headrick (2002) has shown that his method is much more flexible in terms of the values that these correlations may have.

### 3.1.3 Issues on the implementation of the polynomial methods in simulation studies

Because of its ease of implementation, the Vale and Maurelli (1983) multivariate extension of Fleishman's (1978) method is arguably the most-widely used method in simulation studies for the social sciences. It has more than 130 citation counts on ISI's Web of Knowledge (and more than 230 on Google Scholar) and has been the default in popular software programs such as EQS and the *lavaan* and *semTools* packages in *R*. In spite of its widespread use, there is not much research documenting its implementation, outside of the fact that other methods can cover a wider range of skewness/kurtosis combinations and ranges of correlation values.

Only two series of studies were found to specifically address the issue of empirically assessing the quality of the nonnormal data generated by both the third-order polynomial and the fifth-order polynomial methods (Kraatz, 2011; Luo, 2011). Kraatz (2011) evaluates the third- and fifth-order polynomial methods (alongside the g- and h- distributions) in terms of the theory behind the method and the quality of the estimates they generate. In her evaluation of the Vale and Maurelli (1983) method (which is directly relevant to the purpose of this article), Kraatz (2011) presents eight combinations of skewness/kurtosis from published articles (0/25, 0/3, 1/1, 1.75/3.75, 2/6, 3/21, 21.25/3.75, and 2/40) at two different sample sizes (40 and 100) and 100,000 replications per condition to empirically estimate their expected values. Overall, she found that the simulation-derived expected values are almost never close to the values specified in the population (and intended by the researcher), sometimes underestimating in them by a substantial amount. She concluded that skewness tends to be better reproduced than kurtosis and that, overall, larger sample sizes were needed to obtain better parameter estimates of skewness and kurtosis, even though they also resulted in larger variability of the estimates.

Luo (2011) focuses on ordinal data in structural equation models but contains two sets of studies where the Fleishman (1978) method, the Vale and Maurelli (1983) extension, the Headrick and Sawilowsky (1999) modification of Vale and Maurelli, the Headrick (2002), and Ruscio and Kaczetow (2008) methods are investigated. In her simulation study, she chose distributions of Dimension 2 (with a skewness of $\sqrt{8}$ and a kurtosis of 12), 3 (with skewness/kurtosis combinations of 2/6, 0/3, and 0/1.2), and 4 (with skewness/kurtosis combinations of 2/6, 0/0, $\sqrt{8}$/3, and 0/3) at sample sizes of 10, 20, 100, and 1,000 and 50,000 replications per condition. In her assessment of these methods, she concludes that the Ruscio and Kaczetow (2008) method is preferred in the cases of large sample sizes and that Headrick and Sawilowsky's (1999) modification is to be preferred at small samples.

One relevant aspect of both Luo and Kraatz's work touches on the fact that the estimates of higher order moments can be biased and highly variable if the sample sizes are small or moderate. In some of the simulation conditions studied by Kraatz and Luo, the mean of the empirical distributions of the kurtosis values were several units lower than the population-defined ones, sometimes even less than half of their intended value.

Another issue that is raised in both studies is the fact that both the third and fifth order polynomial methods can have more than one set of solutions for the exact same values of skewness and kurtosis. The fact that there is more than one solution can be predicted by the fundamental theorem of algebra so that nonconstant polynomials of degree higher than one are susceptible to have many solutions, especially if nonreal solutions are also taken into account. Aside from the work of Headrick and Kowalchuk (2007) there are no guidelines in the literature that help researchers choose which set of polynomial coefficients are appropriate for data-generation purposes and no studies have yet been done to see whether the conclusions from simulation studies change depending on the use of different sets of polynomial coefficients.

Because very limited literature exists that evaluates the use and quality of the data generated through the previously discussed methods, the series of studies herein aims to address two specific goals. The first goal, and hence the first study, is meant to assess the performance of the third-order polynomial method with high-dimensional data. Although the studies conducted both by Kraatz and Luo investigate correlated data, Kraatz only worked with bivariate distributions and Luo did not consider distributions with more than 4 dimensions. Currently, to our knowledge, no published studies exist where the quality of the data generated by the Vale and Maurelli (1983) method is analyzed in models where the factor structure of the covariance is known in the population. There are also, to our knowledge, no published studies answering the question of whether or not typical factor models used in simulation studies imply nonpositive definite intermediate correlation matrices. This study will look at both issues.

The second goal, and hence the second study, aims to redo the first simulation study in Curran, West, and Finch (1996) using Headrick’s fifth-order polynomial method, in order to discover whether or not some of the conclusions from this published article would change if a different data-generation method had been used. Some of the computational difficulties of implementing the Headrick (2002) method will be discussed as well as general recommendations for quantitative analysts for the use of each algorithm.

## 3.2 Method

### 3.2.1 Study 1

Four articles were chosen to exemplify the use of the Vale and Maurelli (1983) method in the robustness literature: Curran et al. (1996); Finch, West, and MacKinnon (1997); Flora and Curran (2004); and Skidmore and Thompson (2011). The first three were chosen because of their prominence in the structural equation modeling (SEM) literature, as evidenced by their citation count. The Skidmore and Thompson (2011) article was selected both because it is outside of the field of SEM (hence illustrating the use of the Vale & Maurelli, 1983, method in other types of robustness studies) and because of its relatively recent publication date, hinting toward the fact that the third-order polynomial approach is still very much in vogue.

For the first study, the models and simulation conditions investigated by the authors in the previously mentioned articles were redone, but instead of looking at the impact that nonnormality had on the parameter estimates, the sample estimates of skewness and kurtosis were recorded. All simulations and analyses were done in the *R* (Version 3.0.3) programming environment, using the *lavaan* package (Rosseel, 2012, Version 0.5-15), which implements the Vale and Maurelli (1983) method to generate nonnormal data through the *simulateData()* function. Each combination of sample size and nonnormality condition was replicated 10,000 times. Table 3.1 summarizes the sample sizes and skewness/kurtosis population values used in the articles. Path diagrams of the models can be found in the Appendix D.

**Table 3.1** Sample sizes and skewness/kurtosis combinations studied in each article

| Article | Sample size | (Skewness, Kurtosis) |
|---|---|---|
| Curran, West & Finch (1996) | 100, 200, 500, 1000 | (0,0) (2,7) (3,21) |
| Finch, West & MacKinnon (1997) | 150, 250, 500, 1000 | (0,0) (2,7) (3,21) |
| Flora & Curran (2004) | 100, 200, 500, 1000 | (0,0) (.75,1.75) (1.25, 3.75) |
| Skidmore & Thompson (2011) | 10,20,40,60,100,200 | (0,0), (1,1), (-1.5,3.5) |

In all four articles, the large-sample properties of the skewness and kurtosis estimates were verified by means of generating ''empirical'' populations of size 10,000 (Curran et al., 1996; Finch et al., 1997); 50,000 (Flora & Curran, 2004); and 100,000 (Skidmore & Thompson, 2011). The same empirical populations were reproduced using *lavaan* to ensure that the data-generation method was comparable across the different studies (*EQS* was used in the first three articles to generate data and the *SAS* macro described in Fan et al., 2002, was used for the last article). Descriptive statistics were calculated to investigate any potential biases of the estimates as well as their variability. Intermediate-correlation matrices were also calculated in each case to verify their positive-definiteness.

**3.2.2 Study 2**

For the second study, the fifth-order polynomial method as described in Headrick (2002) was implemented in *R* to compare its sample estimates of skewness and kurtosis to the ones generated by using Vale and Maurelli (1983). To investigate this method further, the first simulation study described in Curran et al. (1996) (''Model 1'' on p. 19) was rerun, under the same conditions, but using the newer data-generation algorithm to discover the impact (if any) that it could have on the conclusions from the study.

Headrick, Sheng, and Hodis (2007) provide a *Mathematica* script that needed to be adapted in *R* before the main simulation studies could be set up. The *nlimnb()* function was used to find the roots of the fifth-order polynomial equations instead of the *FindRoot* routine from *Mathematica*. To verify the validity of the *R* implementation of the *Mathematica* script, some of the simulations done in Headrick (2002) and Headrick and Kowalchuck (2007) were run and results were compared with the tables published in said articles. All answers were within four or five decimal points showing that the *R* script was accurately implementing the fifth-order polynomial method as it was originally intended.

An initial problem that had to be overcome was the fact that no ranges of potential values for the standardized higher order moments are known if the skewness and kurtosis values are set in advance. Even though Headrick (2002, 2004) provides certain values for the fifth- and sixth-order moments for certain combinations of skewness and kurtosis, no values are available for the combinations of skewness/kurtosis used in the published articles being analyzed. To obtain a solution, a grid search was programmed where the known values shown in Headrick (2002) are used as starting values and then *R* would attempt different solutions until estimates of the fifth and sixth standardized higher order moments were found. The grid search starts by randomly generating two lists of 50 candidate values (each for one higher order moment) and inspects all first $10 \times 10$ combinations in search of a solution. If no solution is found, the index for the fifth-order moment is increased to 15 so 5 new combination of values is explored. If all those values are traversed and no solution is found, the index for the sixth-order moment is increased then to allow new candidates and the process is repeated. This method of exhaustively exploring candidate solutions is standard practice for grid search—the interested reader can consult Kruschke (2010, Chap. 6), for a more in-depth explanation. A list of these values was kept in order to use them either to calculate future values of skewness and kurtosis or to go through the data-generation process more efficiently. Once the values for the fifth and sixth standardized order moments are found, the data-generation process is relatively straightforward.

For the purposes of comparing methods, descriptive statistics of the estimated skewness and kurtosis values of the fifth-order polynomial were computed. Headrick (2002) showed in simulation studies that estimates of skewness and kurtosis from his method are superior in terms of less variability and smaller bias, but Luo's (2011) results seem to contradict this finding. In her simulation studies, she found that the improvement provided by the Headrick (2002) method over Vale and Maurelli (1983) was marginal at the cost of great computational complexity. She also found that in the cases of three- and four-dimensional distributions, the sample estimates of the correlations were biased (even at the sample size of 1,000) making it the least efficient one among the ones she studied. Kraatz (2011) also speculates that the fifth-order polynomial method could become ''unpredictable'' (p. 61) if used to simulate data in higher dimensional settings (much like the ones being studied here) but provides no evidence of this.

The main purpose of Study 2, however, is to answer the question of whether conclusions derived from simulations using the Vale and Maurelli (1983) method hold when the Headrick (2002) method is chosen instead. Because of the exploratory nature of this study, only the simulations done for Model 1 (properly specified model) described in Curran et al. (1996) were redone to exemplify the uses of the fifth-order polynomial method. The same simulation conditions were implemented with the only exception that instead of 500 replications (as in the original study), 10,000 replications were used to ensure the stability of the results. Portions from Table 1 (p. 22) from the Curran et al. (1996) article were reproduced including the estimated chi-square values obtained when the Headrick (2002) method was used. Although not investigated in the original study, the average standard errors in each simulation condition will also be reported to further investigate any differences that may arise by using each data-generation method. To ensure comparability between both simulation studies, the *mimic=''EQS''* option was specified within the *cfa()* function of *lavaan*

## 3.3 Results

### 3.3.1 Study 1

Descriptive statistics of the estimates across the 10,000 replications for each sample size are shown in Table 3.2. Because of the large number of simulation conditions studied in each article, only one SEM model was chosen for Curran et al. (1996); Finch et al. (1997); and the Flora and Curran (2004) to report in the table. For Skidmore and Thompson (2011), only results with population correlation of 0.5 are depicted. For these particular articles, the various types of SEM models and ranges of correlations did not seem to have an impact on the quality of the sample skewness and kurtosis estimates generated by the Vale and Maurelli (1983) algorithm. There was also no evidence found that any of the models or or correlational structures implied nonpositive definite intermediate correlation matrices.

There is a consistent downward bias of the skewness/kurtosis estimates across all sample size conditions and across all different studies. In general, sample skewness is estimated much more accurately than sample kurtosis and higher levels of kurtosis at smaller sample sizes were generated with a heavily downward bias, which is consistent with the findings from the Luo (2011) and Kraatz (2011). From inspecting the minimum and maximum values, it is possible to see that there is much variability in the estimates, particularly for sample kurtosis. For the population kurtosis value of 21, there were instances where the Vale and Maurelli (1983) algorithm generated estimates more than 10 times larger than what was specified in the population. It is important to point out that the variability in the estimates of skewness and kurtosis (measured by the standard deviation of the empirical distributions) seems to follow different patterns as sample size becomes larger. For skewness estimates, the standard deviations are progressively reduced but for kurtosis estimates, they tend to increase.

To further analyze the nature of the bias in the skewness and kurtosis estimates, Table 3.3 shows the percentage of the values generated by Vale and Maurelli (1983), which fall below the population values specified by the authors. In general, the majority of the estimates are below the the population parameter, particularly for small sample sizes at higher levels of skewness/kurtosis. Even at the largest sample size condition of 1,000, a substantial amount of sample skewness and kurtosis are still below the values intended by the simulation conditions.

**Table 3.2** Mean (M), median (Mdn), standard deviation (SD), minimum (Min) and maximum (Max) of the skewness and kurtosis estimates generated by Vale & Maurelli (1983) at various sample sizes (n).

| | n= | 100 | | | 200 | | | 500 | | | 1000 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M / Mdn | SD | Min / Max | M / Mdn | SD | Min / Max | M / Mdn | SD | Min / Max | M / Mdn | SD | Min / Max |
| C,W &F (1996) | | | | | | | | | | | | |
| Skewness = 2 | 1.70 / 1.59 | 0.62 | 0.62 / 7.43 | 1.83 / 1.72 | 0.55 | 0.57 / 7.58 | 1.97 / 1.91 | 0.33 | 1.22 / 7.21 | 1.99 / 1.94 | 0.31 | 1.35 / 7.93 |
| Kurtosis = 7 | 4.26 / 3.04 | 4.45 | -1.01/ 63.11 | 5.32 / 4.09 | 4.63 | -0.42/ 102.8 | 6.57 / 5.72 | 3.73 | 1.57/ 114.2 | 6.60/ 5.74 | 3.66 | 1.77 / 137.21 |
| Skewness = 3 | 2.21 / 2.09 | 1.21 | -4.63 / 8.09 | 2.48 / 2.33 | 1.15 | -4.56 / 11.04 | 2.74 / 2.60 | 1.02 | -3.82/ 13.26 | 2.86 / 2.72 | 0.83 | -1.89 / 14.66 |
| Kurtosis = 21 | 9.65 / 7.11 | 8.51 | -0.53/ 75.68 | 12.88 / 9.71 | 10.6 | 1.20 /136.55 | 16.44/ 12.77 | 12.40 | 3.12/ 211.7 | 18.38 /14.83 | 12.5 | 4.69 / 359.73 |
| | n= | 150 | | | 250 | | | 500 | | | 1000 | |
| F, W & McK (1997) | M / Mdn | SD | Min / Max | M / Med | SD | Min / Max | M / Med | SD | Min / Max | M / Med | SD | Min / Max |
| Skewness = 2 | 1.78 / 1.67 | 0.58 | 0.35 / 7.95 | 1.86 / 1.77 | 0.51 | 0.78 / 8.08 | 1.93 / 1.85 | 0.42 | 1.00 / 7.70 | 1.96 / 1.90 | 0.32 | 1.20 / 4.02 |
| Kurtosis = 7 | 4.86 / 3.59 | 4.70 | -0.61/ 80.22 | 5.61 / 4.35 | 4.57 | -0.12 / 95.56 | 6.26 / 5.16 | 4.01 | 0.75/ 107.7 | 6.56 / 5.71 | 3.65 | 1.51 / 132.12 |
| Skewness = 3 | 2.25 / 2.12 | 1.43 | -4.29 / 9.11 | 2.55 / 2.42 | 1.11 | -8.36/12.02 | 2.76 / 2.58 | 0.92 | -4.33/ 11.62 | 2.86 / 2.73 | 0.84 | -4.78 / 16.66 |
| Kurtosis= 21 | 9.74/7.44 | 9.12 | -0.32/ 93.86 | 13.71/ 10.46 | 11.6 | 1.36/ 175.92 | 16.74/ 13.11 | 13.64 | 4.32/ 208.9 | 18.20/ 14.90 | 12.1 | 4.82 / 390.06 |
| | n= | 100 | | | 200 | | | 500 | | | 1000 | |
| C & F (2004) | M / Mdn | SD | Min / Max | M / Mdn | SD | Min / Max | M / Mdn | SD | Min / Max | M / Mdn | SD | Min / Max |
| Skewness = 0.75 | 0.56 / 0.41 | 0.36 | -1.12 / 7.32 | 0.64 / 0.59 | 0.28 | -1.06 / 6.33 | 0.71 / 0.68 | 0.16 | -0.98 / 4.12 | 0.74 / 0.71 | 0.08 | 0.01 / 3.12 |
| Kurtosis = 1.25 | 0.98 / 0.82 | 1.13 | -0.16/ 23.18 | 1.11 / 0.96 | 1.64 | -0.04/ 17.44 | 1.16 / 1.04 | 1.78 | 0.06/ 19.26 | 1.22 / 1.17 | 2.03 | 0.09 / 13.11 |
| Skewness = 1.25 | 1.06 / 0.99 | 0.55 | -1.13 / 6.11 | 1.14 / 1.08 | 0.43 | -0.99 / 5.56 | 1.21 / 1.16 | 0.33 | 0.13 / 4.43 | 1.23 / 1.20 | 0.21 | 0.55 / 3.04 |
| Kurtosis = 3.75 | 2.34 / 1.56 | 2.93 | -0.83/ 46.72 | 2.96 / 2.16 | 3.11 | -0.71 / 38.83 | 3.40 / 2.77 | 3.96 | 0.11 / 81.35 | 3.59 / 3.05 | 4.22 | 0.65 / 77.58 |
| | n= | 10 | | | 40 | | | 100 | | | 200 | |
| S & T (2012) | M / Mdn | SD | Min / Max | M / Mdn | SD | Min / Max | M / Mdn | SD | Min / Max | M / Mdn | SD | Min / Max |
| Skewness = 1 | 0.18 / 0.09 | 0.21 | -1.65 / 1.98 | 0.36 / 0.21 | 0.16 | -1.08 / 1.33 | 0.78 / 0.52 | 0.12 | -0.96 / 1.29 | 0.91 / 0.76 | 0.08 | -0.10 /1.06 |
| Kurtosis = 1 | 0.12 / 0.06 | 0.33 | -2.34 / 3.02 | 0.22 / 0.16 | 1.62 | -1.04 / 7.88 | 0.31 / 0.25 | 2.02 | -0.86/ 14.61 | 0.72 / 0.64 | 2.14 | -0.88 / 19.06 |
| Skewness = -1.5 | -0.63/ 0.59 | 0.52 | -2.22 / 1.36 | -1.15 / -1.08 | 0.50 | -4.96 / 0.08 | -1.33 / -1.26 | 0.44 | -5.20/ -.031 | -1.41 / -1.36 | 0.37 | -4.38 / -0.50 |
| Kurtosis = 3.5 | -0.68/ -0.99 | 1.00 | -2.15 / 3.39 | 1.36 / 0.71 | 2.31 | -1.59 / 20.44 | 2.35 / 1.64 | 2.58 | -1.08/ 37.93 | 2.87 / 2.25 | 2.48 | -0.59 / 59.67 |

**Table 3.3** Percentage of sample estimates lower than the population estimates across 10,000 replications.

| | Curran, West & Finch (1996) | | | |
|---|---|---|---|---|
| | N = 100 | N=200 | N=500 | N=1000 |
| Skewness = 2 | 75.66% | 71.36% | 62.06% | 58.51% |
| Kurtosis = 7 | 83.35% | 78.08% | 68.29% | 58.61% |
| Skewness = 3 | 79.81% | 74.88% | 72.77% | 66.54% |
| Kurtosis = 21 | 91.24% | 86.40% | 84.75% | 76.98% |
| | Finch, West & MacKinnon (1997) | | | |
| | N = 150 | N=250 | N=500 | N=1000 |
| Skewness = 2 | 72.85% | 69.82% | 65.58% | 59.82% |
| Kurtosis = 7 | 80.92% | 77.68% | 66.93% | 60.11% |
| Skewness = 3 | 73.69% | 72.04% | 69.23% | 65.38% |
| Kurtosis = 21 | 88.74% | 85.82% | 81.93% | 76.04% |
| | Flora & Curran (2004) | | | |
| | N = 100 | N=200 | N=500 | N=1000 |
| Skewness = 0.75 | 62.13% | 59.48% | 50.12% | 46.88% |
| Kurtosis = 1.25 | 68.11% | 65.72% | 61.66% | 58.16% |
| Skewness = 1.25 | 70.33% | 66.34% | 61.46% | 60.32% |
| Kurtosis = 3.75 | 81.54% | 77.11% | 70.06% | 67.82% |
| | Skidmore & Thompson 2011 | | | |
| | N=10 | N=40 | N=100 | N=200 |
| Skewness = 1 | 96.11% | 88.50% | 67.95% | 59.12% |
| Kurtosis = 1 | 98.12% | 90.42% | 74.57% | 69.16% |
| Skewness = -1.5 | 6.24% | 20.85% | 28.42% | 34.07% |
| Kurtosis = 3.5 | 100% | 87.28% | 78.85% | 73.49% |

### 3.3.2 Study 2

For Study 2, the same values of skewness and kurtosis used in the Curran et al. (1996) simulation studies were re-calculated using the Headrick (2002) method. Only the correctly specified model condition (‘’Model Specification 1’’ in their article) was run. Table 3.4 shows the average across 10,000 replications for the differing sample sizes and higher order moment conditions.

**Table 3.4** Comparison of the mean skewness and kurtosis estimates between the Vale & Maurelli (1983) method (VM) and the Headrcik (2002) method (H) for the Curran, West & Finch (1996) study across 10,000 replications. Standard deviations appear between parentheses

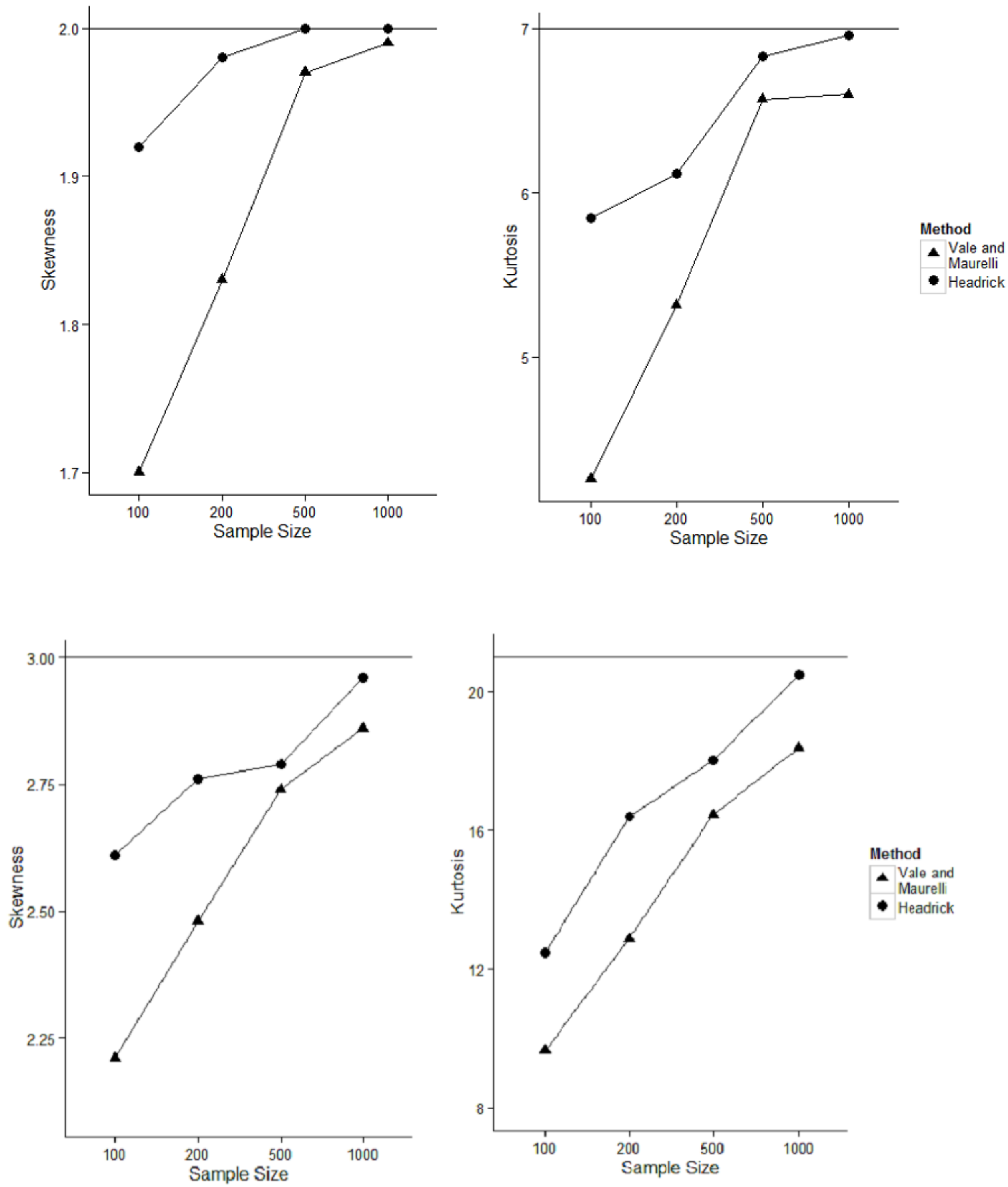
| | n=100 | | n=200 | |
|---|---|---|---|---|
| | VM | H | VM | H |
| Skewness= 2 | 1.70(0.62) | 1.92(0.51) | 1.83(0.55) | 1.98(0.42) |
| Kurtosis = 7 | 4.26(4.45) | 5.85(2.42) | 5.32(4.63) | 6.12(3.44) |
| Skewnes = 3 | 2.21(1.21) | 2.61(1.19) | 2.48(1.15) | 2.76(1.06) |
| Kurtosis = 21 | 9.65(8.51) | 12.46(7.90) | 12.88(10.63) | 16.39(9.58) |
| | n=500 | | n=1000 | |
| | VM | H | VM | H |
| Skewness= 2 | 1.97(0.33) | 2.00(0.31) | 1.99(0.31) | 2.00(0.30) |
| Kurtosis = 7 | 6.57(3.73) | 6.83(3.11) | 6.60(3.66) | 6.96(3.50) |
| Skewnes = 3 | 2.74(1.02) | 2.79(0.91) | 2.86(0.83) | 2.96(0.67) |
| Kurtosis = 21 | 16.44(12.40) | 18.02(10.88) | 18.38(12.55) | 20.46(11.05) |

It can be seen that the empirical averages generated by the Headrick (2002) method, although still downward biased, are closer to those specified in the population. The standard deviations for each condition are also consistently lower when compared with the ones calculated from the values generated via the Vale and Maurelli (1983) algorithm, which helps show that the fifth-order polynomial method also generates estimates that are more consistent.

Just as with the estimates shown in Table 3.2, the kurtosis values obtained via the Headrick (2002) method exhibit an increasing trend in their standard deviations as the sample size grows larger, the opposite of what happens to the skewness estimates, where larger sample sizes are associated with less variability.

In Figure 3.1, it is possible to see the mean values across replications for each condition of

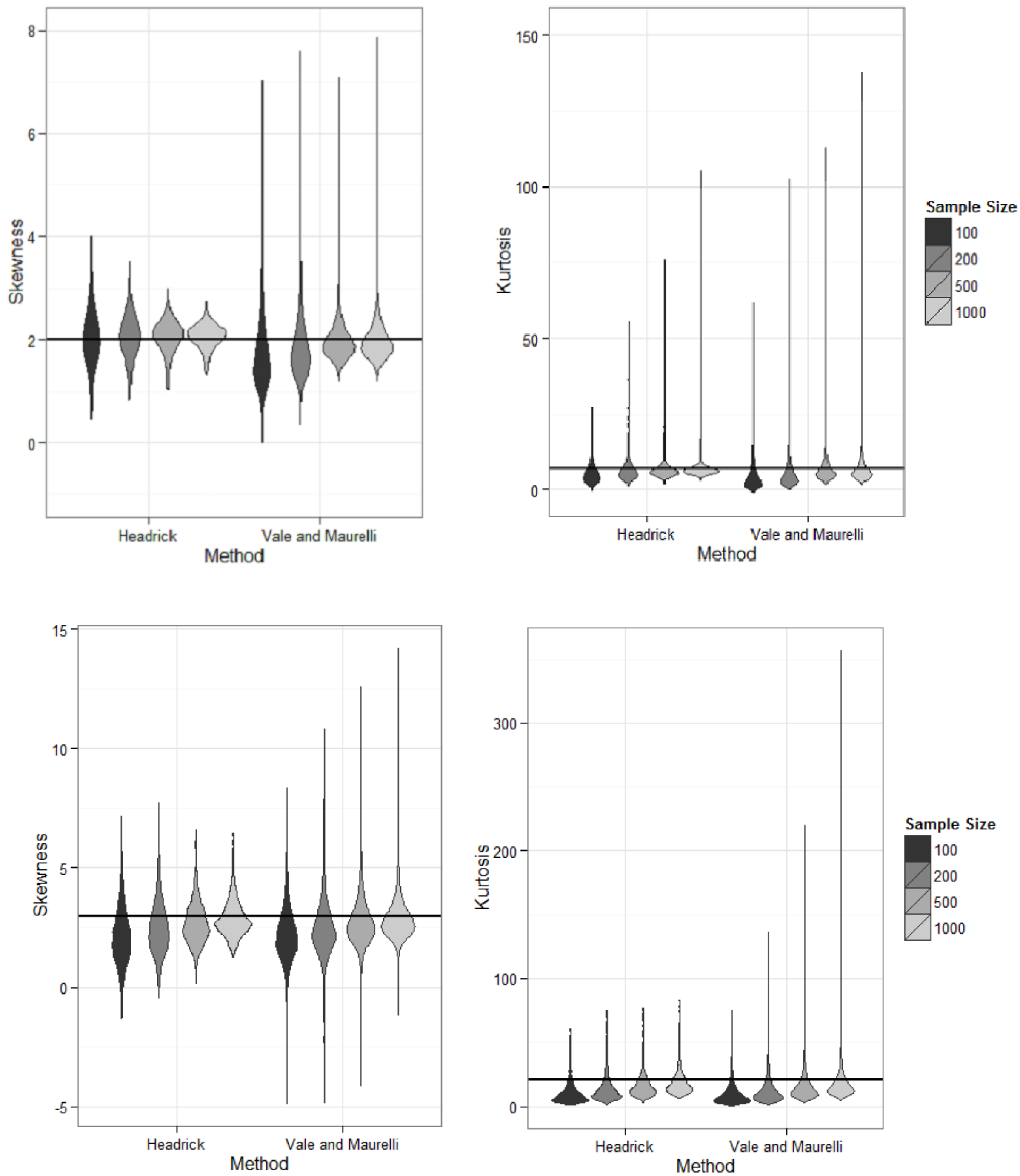
**Figure 3.1** Plots of means for skewness and kurtosis values across simulation replications as a function of sample size



sample size and population skewness and kurtosis. The Headrick (2002) method consistently outperforms the Vale and Maurelli (1983) method in terms of less bias in the parameter estimates, but its overall slope is smaller so it approaches the true population values at a slower rate, particularly for small samples. With the exception of the population skewness of 2, neither method ever reaches their intended value, although the fifth-order polynomial approach is always the closest.

To visually investigate the variability of the estimates, violin plots for each skewness/ kurtosis combination and sample size condition are presented in Figure 3.2. As a hybrid between a boxplot and an empirical density plot, violin plots highlight the areas of highest mass in the distribution of the data as well as the behavior of its furthermost points. It is possible to see that the Vale and Maurelli (1983) method is susceptible to generating extreme points, particularly when the population values of skewness or kurtosis are high. The Headrick (2002) method is much more consistent in generating sample skewness and kurtosis values closer to the true value intended in the population. By comparing the areas of highest density of the data (where the ‘‘shoulders’’ of the violin plot begin to appear) across methods, it becomes apparent that the fifth-order polynomial algorithm consistently outperforms the third-order polynomial approach, particularly with increasing sample size. This fact is most evident in the condition of population kurtosis of 21 where, as sample size increases, the curves of the violin plot for the Headrick (2002) method become wider and closer around the true population value. Although the Vale and Maurelli’s (1983) plot also approaches the intended value, it can be readily seen that the highest density of the data is still below it. Contrary to the plots depicting skewness, where larger sample sizes imply that the largest density of the data concentrates around the population values, kurtosis plots show the same trend of more extreme values being generated at higher sample sizes, reflecting the increases in standard deviation of the estimates shown in Tables 3.2 and 3.4.

**Figure 3.2** Violin plots showing the empirical density of the values of skewness and kurtosis generated by each method. The intended population value is highlighted with a black line. The top two plots correspond to the skewness of 2 and a kurtosis of 7 condition. The bottom two plots show the skewness of 3 and kurtosis of 21 condition.



In terms of whether or not using Headrick's (2002) method would have influenced the conclusions found in Curran et al.'s (1996) first study on the robustness of the chi-square test of fit, Table 3.5 compares the values obtained by the authors with the values obtained in this simulation when using both the Vale and Maurelli (1983) approach and the Headrick (2002) approach. The values reported by Curran et.al. (1996) differ little from what was found by fitting the model in *lavaan*. Only the

**Table 3.5** Chi-square values obtained by the methods of maximum likelihood (ML), Satorra-Bentler (SB) correction or asymptotic distribution free (ADF).

| | | | Moderately nonnormal (skewness =2, kurtosis =7) | | | | |
|---|---|---|---|---|---|---|---|
| n | Method | Expected | C,W&F (1996) | VM (1983) | Min/Max VM | H(2002) | Min/Max H |
| 100 | ML | 24 | 29.35 | 29.46 | 8.59 /72.58 | 36.19 | 7.24 / 61.98 |
| | SB | 24 | 26.06 | 26.38 | 8.13/58.76 | 28.83 | 6.68/41.51 |
| | ADF | 24 | 38.04 | 36.85 | 9.01/64.94 | 36.16 | 11.10/47.2 |
| 200 | ML | 24 | 30.15 | 29.96 | 8.93/71.08 | 34.78 | 7.41/62.47 |
| | SB | 24 | 25.44 | 25.18 | 9.71/47.29 | 26.36 | 7.23/40.83 |
| | ADF | 24 | 29.27 | 28.53 | 10.11/57.98 | 31.32 | 7.36/42.27 |
| 500 | ML | 24 | 31.26 | 30.66 | 10.40/61.01 | 36.23 | 10.06/50.76 |
| | SB | 24 | 25.44 | 24.28 | 8.33/53.43 | 25.92 | 8.11/43.84 |
| | ADF | 24 | 26.42 | 25.79 | 10.19/53.73 | 26.95 | 8.29/44.48 |
| 1000 | ML | 24 | 30.78 | 30.06 | 9.94/66.84 | 36.82 | 11.77/55.79 |
| | SB | 24 | 24.77 | 24.32 | 8.96/55.09 | 24.16 | 8.71/42.46 |
| | ADF | 24 | 25.36 | 25.11 | 8.79/53.42 | 24.85 | 8.04/46.33 |
| | | | Severely nonnormal (skewness =3, kurtosis =21) | | | | |
| n | Method | Expected | C,W&F (1996) | VM (1983) | Min/Max VM | H(2002) | Min/Max H |
| 100 | ML | 24 | 33.54 | 33.46 | 10.65/91.09 | 30.82 | 7.84/74.38 |
| | SB | 24 | 27.26 | 27.41 | 12.17/53.20 | 28.19 | 8.31/44.19 |
| | ADF | 24 | 44.82 | 40.56 | 11.24/64.14 | 34.16 | 11.35/59.43 |
| 200 | ML | 24 | 34.40 | 34.78 | 11.33/104.39 | 29.86 | 9.82/97.83 |
| | SB | 24 | 25.80 | 26.17 | 11.09/68.59 | 25.94 | 9.12/48.94 |
| | ADF | 24 | 31.29 | 28.66 | 11.52/52.61 | 26.80 | 8.92/41.94 |
| 500 | ML | 24 | 35.55 | 34.95 | 9.46/102.21 | 31.12 | 8.57/94.06 |
| | SB | 24 | 24.85 | 24.62 | 10.15/52.21 | 23.62 | 9.01/46.28 |
| | ADF | 24 | 26.83 | 25.31 | 9.37/50.56 | 25.11 | 10.09/47.72 |
| 1000 | ML | 24 | 37.40 | 37.10 | 10.50/94.89 | 29.73 | 9.38/95.47 |
| | SB | 24 | 25.01 | 24.84 | 8.43/50.16 | 24.60 | 8.05/52.96 |
| | ADF | 24 | 25.47 | 25.20 | 8.88/48.7 | 25.12 | 8.61/49.52 |

*Note.* The values reported by Curran, West and Finch (1996) (C, W &F) are presented alongside with those obtained using the Vale and Maurelli (1983) (VM) method and the Headrick (H) method at the various sample size (N) conditions. Minimum and maximum (Min/Max) chi-square values per method and sample size conditions are also presented.

results of the ‘‘moderately nonnormal’’ and ‘‘severely nonnormal’’ conditions are shown, but the normal distribution condition was run to ensure *lavaan*’s results were comparable.

Overall, both the Headrick (2002) and the Vale and Maurelli (1983) methods generated mean chi-square values which were relatively comparable to each other. For the cases of the ‘‘moderately nonnormal’’ condition, the fifth-order polynomial method was consistently associated with higher chi-square values when the model was fitted using likelihood theory. This situation seemed to reverse itself for the ‘‘severely nonnormal’’ condition though, in which the Headrick (2002) method was associated with lower chi-square values (albeit still higher than the expected value for the chi-square distribution). The Satorra–Bentler correction and asymptotic distribution free (ADF) chi-square estimates followed the overall pattern reported by Curran et al. (1996) where the Satorra–Bentler correction yielded the closest values to the expected model chi-square when calculated with nonnormal data, particularly for small sample sizes. No considerable differences were found in the mean chi-square values obtained for the Satorra–Bentler and ADF chi-squares, regardless of whether the Headrick (2002) method or the Vale and Maurelli (1983) method were used. It is important to point out, however, that the Vale Maurelli (1983) method was associated with higher chi-square values, particularly with higher maximums than the Headrick (2002) method.

Table 3.6 shows the empirical rejection rates published by Curran et al. (1996) and those obtained by generating data through the Vale and Maurelli (1983) and Headrick (2002) method. In general, the Vale and Maurelli (1983) rejection rates appear to be somewhat lower than those reported by Curran et al. (1996; with the exception of chi-squares obtained from the normal theory maximum likelihood), although they are still reasonably close to the ones published in the original study. The empirical rejection rates from the Headrick (2002) method, however, do exhibit some important changes, particularly for the ADF estimator. The Headrick (2002) algorithm yielded empirical rejection rates that were overall higher than those obtained through the Vale and Maurelli (1983) method for the case of the normal-theory maximum likelihood, but they seemed to converge to the nominal alpha of 5% much faster, that is at smaller sample sizes, for the Satorra–Bentler correction and ADF estimator. Although the ADF chi-square still required larger sample sizes to approach the theoretical rejection rate, it did so at smaller sample sizes when the data were generated by the Headrick (2002) method than when the data generated by the Vale and Maurelli (1983) method. For instance, the ADF estimator at a

**Table 3.6** Empirical rejection rates obtained by the methods of maximum likelihood (ML), Satorra-Bentler (SB) correction or asymptotic distribution Free (ADF) across 10,000 replications.

| | | | Moderately nonnormal (skewness =2, kurtosis =7) | | |
|---|---|---|---|---|---|
| n | Method | Expected | C,W&F (1996) | VM (1983) | H (2002) |
| 100 | ML | 5% | 20% | 21.7% | 30.8% |
| | SB | 5% | 8.5% | 8.6% | 9.5% |
| | ADF | 5% | 49% | 46.45% | 26.67% |
| 200 | ML | 5% | 25% | 20.5% | 27.9% |
| | SB | 5% | 8% | 7.7% | 6.7% |
| | ADF | 5% | 19% | 16.1% | 7.2% |
| 500 | ML | 5% | 24% | 21.3% | 29.4% |
| | SB | 5% | 6.9% | 6.5% | 6.4% |
| | ADF | 5% | 6.7% | 6.5% | 6% |
| 1000 | ML | 5% | 24% | 23.3% | 25.6% |
| | SB | 5% | 7.5% | 6.8% | 5.8% |
| | ADF | 5% | 7.5% | 6.6% | 4.8% |
| | | | Severely nonnormal (skewness =3, kurtosis =21) | | |
| n | Method | Expected | C,W&F (1996) | VM (1983) | H (2002) |
| 100 | ML | 5% | 30.0% | 33.8% | 37.9% |
| | SB | 5% | 13.0% | 12.4% | 9.3% |
| | ADF | 5% | 68.0% | 59.4% | 46.6% |
| 200 | ML | 5% | 36.0% | 34.7% | 38.3% |
| | SB | 5% | 6.5% | 7.0% | 5.2% |
| | ADF | 5% | 25.0% | 19.2% | 10.3% |
| 500 | ML | 5% | 40.0% | 40.0% | 40.7% |
| | SB | 5% | 8.5% | 8.4% | 6.5% |
| | ADF | 5% | 8.5% | 7.9% | 4.6% |
| 1000 | ML | 5% | 48.0% | 42.4% | 42.5% |
| | SB | 5% | 7.0% | 6.4% | 4.6% |
| | ADF | 5% | 7.2% | 6.1% | 5.9% |

Note: The rejection rates reported by Curran, West and Finch (1996) (C, W &F) are presented alongside with those obtained using the Vale and Maurelli (1983) (VM) method and the Headrick (H) method at the various sample size (N) conditions.

sample size of 200 for the Headrick (2002) method had a better rejection rate than the Satorra–Bentler chi-square calculated from data generated by the Vale and Maurelli (1983) method.

Even though it was not initially investigated by Curran et al. (1996), it was of interest to also document whether the choice of data-generation method affects the estimates of standard errors for parameter estimates or not. Percentage reduction of the standard errors was calculated the same way Finch et al. (1997) did by subtracting the estimated standard error from the true parameter and dividing the difference by the true parameter. Table 3.7 summarizes these results, where it is possible to see that, under the Headrick (2002) method, the shrinkage of the standard errors is higher than when data are generated by the Vale and Maurelli (1983) method.

**Table 3.7** Percentage reduction of the standard error of the parameter estimates at different sample sizes (n) for the Vale and Maurelli (VM) and Headrick (H) methods.

| Skewness = 2 and Kurtosis = 7 | | |
|---|---|---|
| n | % bias (VM) | % bias (H) |
| 100 | 35% | 43% |
| 200 | 31% | 39% |
| 500 | 33% | 39% |
| 1000 | 34% | 41% |
| Skewness = 3 and Kurtosis = 21 | | |
| n | % bias (VM) | % bias (H) |
| 100 | 52% | 61% |
| 200 | 51% | 58% |
| 500 | 53% | 58% |
| 1000 | 52% | 57% |

## 3.4 Discussion

As it can be readily seen from the results in Study 1, the quality of the estimates generated by the Vale and Maurelli (1983) algorithm is extremely susceptible to the sample size being chosen, yielding estimates of skewness and kurtosis that can be both moderately to severely downwardly biased and extremely variable. These findings are in line with the ones of Luo (2011) and Kraatz (2011), so it is reasonable to assume that regardless of the dimensionality of the data, the third-order polynomial method can result in suboptimal estimates, unless the sample sizes are moderately large (likely larger than 200, although even at 1,000 one can see some bias when the population kurtosis is large in the population). The estimates of kurtosis were much more biased and variable than those of skewness, possibly because larger sample sizes are needed to estimate more accurately the higher moments of a distribution. In spite of this, even after 10,000 replications the bias was still present. Moderate estimates were generated with more precision than higher estimates of kurtosis, so the Vale and Maurelli (1983) method is still ideal for simulations where either larger sample sizes are used or if lower values of skewness and kurtosis are being inspected. Of particular interest is the seemingly contradictory finding that larger sample sizes are accompanied by increases in the variability of the estimates of kurtosis, as measured by the standard deviation of the sampling distribution. The same trend can be observed in Kraatz's (2011) work in her Tables 20 to 27 (pp. 105-112). A potential explanation for this result is an interplay between the upper bound that the sample size places on the estimates of kurtosis and the tendency of the Vale and Maurelli (1983) algorithm to generate extreme values of it. Dalen (1987) provides the most up to date upper bound to the sample kurtosis as a quadratic function of the sample size. If the sample size grows larger, the bound also becomes higher and the third-order polynomial method has more opportunity to generate values that concentrate toward the extreme of the distribution. Clearly, more research is needed that helps evaluate the quality of the data generated by these algorithms not only in terms of bias but also in terms of the variability of these higher order moments.

Although Li and Hammond (1975) and Headrick (2002) present examples where nonpositive intermediate correlation matrices arose while using the third-order polynomial method, the inspection of the intermediate correlation matrices implied by the models specified in each one of the four articles were all positive definite. A potential explanation as for why this could be the case relies on the fact that the range of correlation values is most restricted when the intended population correlation is high (0.7

and greater) and the variables are either highly skewed in opposite directions or if one variable is skewed and the other one is symmetric (Mair et al., 2012). In all the simulation studies considered here, all the variables had skewness and kurtosis values specified in the same direction, so further research is needed to understand other potential factors that may influence the quality of the data obtained via the Vale and Maurelli (1983) method. Preliminary inspections, for instance, show that if the population correlation coefficients are negative, it is much more likely to obtain a nonpositive definite intermediate correlation matrix.

Study 2 helped understand the use of the fifth-order polynomial algorithm and highlighted many of the advantages and potential problems one may encounter with it. In terms of the quality of the generated data, the results were consistent with those found in Headrick (2002), Kraatz (2011), and Luo (2011), where the newer method outperformed the third-order polynomial method both in terms of less bias and variability. This is, of course, a reasonable expectation because of the fact that the addition of the two higher order moments provides the algorithm with more information to better reproduce the values intended in the population. The use of higher dimensional data (9 dimensions) did not result in unstable estimates and the data-generation procedure was, overall, relatively straightforward once suitable values for the $\gamma_3$ and $\gamma_4$ values were found. This is, perhaps, one of the biggest drawbacks associated with the Headrick (2002) method. The optimization of these high-dimensional polynomials is not a simple task and there is currently no theory to help potential users choose suitable values of $\gamma_3$ and $\gamma_4$ for any arbitrary case. Table 2 from Headrick (2002; p. 698) is a good starting point to properly choose some of these values, but most authors use values of skewness and kurtosis well above the range of what is presented there. Experimentation during the process of coding the grid search that currently finds these values in *R* suggests that there are no specific values but ranges of values of $\gamma_3$ and $\gamma_4$, which yield acceptable solutions for the polynomial coefficients. Deriving the boundaries of these ranges for given population skewness and kurtosis could be a potential way to expand the results found here.

The comparison of the simulation results from the first model specification in Curran et al. (1996) yielded some very interesting results. Overall, whether one uses the Vale and Maurelli (1983) or the Headrick (2002) method, the results are consistent ith the statistical theory surrounding SEM: higher levels of kurtosis inflate the value of the chi-square test of fit (Bollen, 1989). It also supported the general conclusions found in Curran et al. (1996), where the Satorra–Bentler correction is favoured over the ADF correction when nonnormality is present. In spite of this, a more nuanced analysis of the results highlights the fact that some of the conclusions the authors arrived at would have changed had the Headrick (2002) method been available to them.

The inflation of the chi-square appeared to be higher with the Headrick (2002) method for the case of moderately nonnormal distributions but somewhat lower for severe nonnormality. This result is slightly counterintuitive given than the third-order polynomial method not only generates values that had lower kurtosis but also that the bulk of these values is on the lower end of the distribution. Further analysis, however, helped show that some of the datasets generated by the Vale and Maurelli (1983) algorithm included pockets of very extreme values of kurtosis, with some having kurtosis values more than 100 for the cases of population kurtosis of 7 and even close to 400 in the cases of population kurtosis of 21, as can be seen in Figure 3.2. These extreme values of kurtosis generated extreme chi-square values which raised the overall average across simulation repetitions. Because the Headrick (2002) method yields estimates that are more consistent, the inflation of chi-square values is not as severe. Table 3.5 helps exemplify this by showing the minimum and maximum chi square values from the 10,000 replications across conditions. In general, the Vale and Maurelli (1983) method had maximum chi-square values that were 10 times higher in the moderately nonnormal condition and around 20 times higher in the severe nonnormal condition, when compared with the Headrick (2002) method.

Table 3.6 highlights some of the important differences of using the Headrick (2002) versus the Vale and Maurelli (1983) procedures. Even though under normal likelihood theory the Headrick (2002) method resulted in higher rejection rates, it seemed to reach the nominal rate of 5% faster. This case is particularly noticeable for the ADF estimator, where it becomes almost as good as the Satorra–Bentler correction even at samples of 200. This is goes against some of the recommendations found in Curran et al. (1996), where they suggest this estimator should be used in samples of 500 or higher. A working hypothesis of why these differences may have arisen has to do with the fact that because the Vale and Maurelli (1983) method is prone to generating data with very large kurtosis values, the demands it places on the data to estimate the weight matrix needed by the ADF estimator are considerably higher than when the data is generated using the Headrick (2002) method. Under the Headrick (2002) method, the ADF estimator reaches its asymptotic chi-square distribution much faster and obtains a better empirical rejection rate. Regardless of whether one uses the third- or the fifth-order polynomial method, both the Satorra–Bentler correction and the ADF chi-square statistic tended to behave similarly when compared with normal-likelihood chi-square. This comes from the fact that both approaches are specifically designed to handle excess kurtosis and, as sample size grows larger, it is expected that they would behave more and more similarly by returning the Type I error rate to its nominal value. It is important to highlight the fact, however, that when the data is generated by the Headrick (2002) method it appears that the asymptotic properties of the ADF estimator manifest themselves faster. The fact that normal likelihood chi-squares are more different between both data-generating methods could point toward the fact that the multivariate structure implied by both them is not the same. It is important to keep in mind that, when using the power polynomial methods, the multivariate nonnormality is indirectly attained by modifying all the one-dimensional marginal distributions. Neither method offers the researcher any control over what the shape of the joint distribution looks like. An interesting avenue for future research could be to explore different methods that allow the researcher to control the multivariate nonnormality (perhaps by setting a population value of Mardia's kurtosis or another measure of multivariate nonnormality) and not only the lower dimensional moments.

Table 3.7 also helps highlight the fact that the shrinkage of standard errors changes depending on whether one uses the Vale and Maurelli (1983) method or the Headrick (2002) method. Future research could look at whether the robust corrections to standard errors require larger sample sizes than what is usually recommended in the published literature, given that standard errors of the parameter estimates tend to be smaller when the data is generated via fifth-order polynomials.

### 3.5 Conclusions and recommendations

There are two sets of concluding remarks to be made from this article. First, it was the purpose of these studies to start a conversation about the algorithms being used in simulations by highlighting some of the advantages and drawbacks of the Vale and Maurelli (1983) method and looking at the differences in terms of results that can be encountered when another method is used. The third-order polynomial is the status quo in terms of nonnormal data-generation procedures for Monte Carlo studies in the social sciences and, even though it used to be the best alternative available, advances in both computational power and statistical theory have given the quantitative researcher a much wider range of choices, many of which could be better suited to investigate the simulation conditions intended by the researchers. Just a small, and by no means exhaustive list of these methods include the multivariate skew-normal distribution (Azzalini & Dalla Valle, 1996), the multivariate g-and-h distribution (Kowalchuk & Headrick, 2010), Gaussian mixtures (Muthén & Muthén, 2002), copula distributions (Mair et al., 2012), and iterative approaches such as the one described in Ruscio and Kaczetow (2008).

Still, the majority of these methods have received little (if any) attention from researchers in psychometrics and the social sciences, in general. The Headrick (2002) method, for instance, has only been used once within the published SEM literature as the data-generating procedure used in the simulation studies of Tong and Bentler (2013). Most of the citations of the fifth-order polynomial in the social sciences come from researchers commenting on the improvements done by Headrick (2002) andHeadrick and Kowalchuk (2007) on the power method, but not many actually use the algorithm itself for the purpose it was intended to (particularly in the multivariate case). Even the overall lack of literature on the empirical properties of the estimates of these algorithms points to the fact that a wide area of study exists which is not seeing much incursion from quantitatively oriented social scientists.

The second concluding remark is that the validity of conclusions in simulation studies depends heavily on the quality of the data-generating procedures, and if these procedures are suspect, the conclusions from simulation studies are suspect as well. For example, the simulation studies done in Lix and Fouladi (2007); Lix, Keselman, and Hinds (2005); and Weathers, Sharma, and Niedrich (2005) use skewness–kurtosis values that imply nonreal polynomial coefficients as solutions to the Fleishman (1978) equations, and a variety of combination of skewness and kurtosis conditions used commonly in the literature can imply odd-shaped (and even bi-modal) distributions on the generated datasets (Kraatz, 2011). There is no published research to document whether different sets of solutions to the Fleishman (1978) and Vale and Maurelli (1983) polynomials influence the results intended by the researchers, yet most people who rely on simulations used this method unsuspectingly.

In conclusion, it is important to raise awareness of a type of ‘‘black box’’ approach to simulation that exists among many quantitative social scientists. Although exceptions do exist where researchers create empirical populations to study the large sample properties of the data-generating methods they use (all the articles studied here reported doing this as a manner of checking whether the data they generated matches their simulation conditions), there seems to be a general lack of concern or awareness about how exactly many of these algorithms work and, more importantly, their limitations. Being the field traditionally associated with fighting against the ‘‘black box’’ approach to understanding statistics and data analysis, it is crucial that we start taking the necessary steps to not only fully understand the tools of our trade, but also to know which tools work better for which task.

# Chapter 4: Solution multiplicity of the Fleishman method and its impact in simulation studies. Is your non-normality the same as mine?

## 4.1 Background

Within the social sciences, Monte Carlo simulations are routinely used to investigate the small sample properties of statistical tests and methods, as well as their potential robustness to the assumptions required to use them. Conducting these simulations implies the researcher has the ability to generate data with the conditions needed to reliably assess their impact on the methods being studied. To simulate correlated non-normal data, several algorithms have been proposed in the literature (e.g. Azzalini & Dalla Valle (1996); Headrick (2002); Foldnes & Olsson (2016)) but few have been used with such frequency as Fleishman (1978) 3rd-order polynomial method and its multivariate extension developed by Vale and Maurelli (1983).

The Fleishman (1978) method relies on a non-linear transformation of standard normal random variates where the first four-order moments of the distribution are parameterized as a 3rd degree polynomial. Assume the random variable $Z$ is normally distributed, $Z \sim N(0,1)$. Define a new, non-normal, variable $Y$ as:

$$Y = a + bZ + cZ^2 + dZ^3 \,, \tag{4.1}$$

where $a, b, c$ and $d$ are real-valued polynomial coefficients. By taking powers of the expected value of $Y$, $E[Y], E[Y^2], E[Y^3], E[Y^4]$, and algebraically expanding them, Fleishman (1978) was able to derive the following system of non-linear equations to solve for the polynomial coefficients needed to obtain the non-normal distribution:

$$a = -c. \tag{4.2}$$

$$b^2 + 6bd + 2c^2 + 15d^2 = 1. \tag{4.3}$$

$$2c(b^2 + 24bd + 105d^2 + 2) = \gamma_1. \tag{4.4}$$

$$24(bd + c^2[1 + b^2 + 28bd] + d^2[12 + 48bd + 141c^2 + 225d^2]) = \gamma_2, \tag{4.5}$$

where $\gamma_1, \gamma_2$ are the respective values of skewness and kurtosis selected by the user.

Although this method allows the user to control the 3rd and 4th central moments for the case of univariate data, one further step needs to be taken in order to expand it to a truly multivariate setting, allowing for the control of the skewness, kurtosis, and correlation structure. Vale and Maurelli (1983) noticed that naively applying the Fleishman transformation to correlated data alters the correlation matrix of the variables. In a similar manner, if uncorrelated data are transformed to non-normality first and then a correlation structure is imposed, the values of skewness and kurtosis would end up being different than those originally intended. In order to tackle this issue, they begin by re-expressing $Y$ as the vector-product of the powers of the standard normally-distributed variate $Z$ and its respective polynomial coefficients:

$$Y = \mathbf{w}'\mathbf{z} = [a, b, c, d]\begin{bmatrix} 1 \\ Z \\ Z^2 \\ Z^3 \end{bmatrix}. \tag{4.6}$$

Because the correlation of two standardized variables can be defined as their cross-product, they proceeded by conceiving two non-normal variables $(Y_1, Y_2)$ that have been Fleishman-transformed and obtain the expected value of their cross-product. It then becomes apparent that:

$$r_{Y_1Y_2} = E[Y_1Y_2] = E[\mathbf{w_1}'\mathbf{z_1}\mathbf{z}'_\mathbf{2}\mathbf{w_2}] = \mathbf{w_1}'E[\mathbf{z_1}\mathbf{z}'_\mathbf{2}]\mathbf{w_2} = \mathbf{w_1}'\mathbf{R}\mathbf{w_2}, \tag{4.7}$$

where $\mathbf{R}$ is the expectation of:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & r_{Z_1Z_2} & 0 & 3r_{Z_1Z_2} \\ 1 & 0 & 2r_{Z_1Z_2}^2 + 1 & 0 \\ 0 & 3r_{Z_1Z_2} & 0 & 6r_{Z_1Z_2}^3 + 9r_{Z_1Z_2} \end{bmatrix}. \tag{4.8}$$

Finally, Vale and Maurelli (1983) used the relationships between the coefficients of the Fleishman transformation and the elements of $\mathbf{R}$ to arrive at the following expression:

$$r_{Y_1Y_2} = r_{Z_1Z_2}(b_1b_2 + 3b_1d_2 + 3d_1b_2 + 9d_1d_2) + r_{Z_1Z_2}^2(2c_1c_2) + r_{Z_1Z_2}^3(6d_1d_2), \quad (4.9)$$

so that for a user-specified correlation $r_{Y_1Y_2}$ , one needs to solve (4.9) in order to obtain $r_{Z_1Z_2}$ first, the "intermediate correlation". This intermediate correlation is then imposed on uncorrelated, standard-normal variates before they are transformed to non-normality via the 3rd order polynomial method and that yields the full multivariate extension of this algorithm.

#### 4.1.1 Limitations of the 3rd-order polynomial transform

In spite of its popularity and ease of implementation, there exist various issues with both the univariate and multivariate versions of this method of which researchers need to be aware. The earliest and most-thoroughly studied limitation refers to the range of possible skewness/kurtosis combinations that the 3rd-order polynomial algorithm can generate (Headrick, 2004; Tadikamalla, 1980). There exists a quadratic relationship between the 3rd and 4th central moments, defined by $\gamma_2 \geq \gamma_1^2 - 2$ , which places a theoretical bound on the possible values that the moments of any probability distribution can take. The wider the region within this parabola an algorithm can cover, the wider the variety of theoretical distributions it can approximate. The Fleishman transformation covers an approximate range of $\gamma_2 \geq 1.588\gamma_1^2 - 1.139$ with an upper bound to the kurtosis set at around 101 (Kraatz, 2011). This implies that it cannot accurately simulate distributions such as the uniform distribution ($\gamma_1 = 0, \gamma_2 = -1.2$) or the standard log-normal distribution ($\gamma_2 \approx 110.936$). Other methods such as Headrick (2002)'s 5th-order polynomial approach or the use of copulas (Mair, Satorra & Bentler, 2012) can be used as an alternative since they allow for a wider choice of more extreme values of higher-order moments.

A second limitation that pertains to the Vale-Maurelli (1983) multivariate extension, concerns the calculation of the intermediate correlation and whether or not there are restrictions on the type of relationships among variables that can be generated through this method (Headrick & Kowalchuk, 2007; Li & Hammond, 1975). From (4.9) it is possible to see that, although $r_{Z_1Z_2}$ spans the complete [-1, +1] range, its interplay with the values of the polynomial coefficients restricts the choices the researchers can have for $r_{Y_1Y_2}$. For instance, assume that a researcher is interested in correlating two variables, one normally-distributed and another non-normally distributed. For the purposes of this example, the values of skewness and kurtosis of the chi-square distribution with one degree of freedom will be selected, $\gamma_1 = \sqrt{8}, \gamma_2 = 12$. The Fleishman transformation for the non-normally distributed variable $Y_1$ and the normally-distributed variable $Y_2$ are:

$$Y_1 = (-0.52067576) + (0.61459822\ )Z + (0.52067576)Z^2 + (0.02007246)Z^3$$

$$Y_2 = (0) + (1)Z + (0)Z^2 + (0)Z^3.$$

Since the mean, skewness and excess kurtosis of the standard normal distribution are defined to be 0, those polynomial coefficients of $Y_2$ are also 0. One can substitute the values of these coefficients in (4.9) to find:

$$r_{Y_1Y_2} = r_{Z_1Z_2}(0.61459822 \times 1) + 3(0.61459822 \times 0) + 3(0.02007246 \times 1) + 9(0.02007246 \times 0) + 2r^2_{Z_1Z_2}(0.52067576 \times 0) + 6r^3_{Z_1Z_2}(0.02007246 \times 0)$$

$$r_{Y_1Y_2} = r_{Z_1Z_2}(0.61459822) + (0.06021738).$$

By letting $r_{Z_1Z_2} = 1$ and $r_{Z_1Z_2} = -1$ one can obtain the bounds of $r_{Y_1Y_2}$ which are [-0.5543808, 0.6748156]. Other choices of higher-order moments can result in a narrower range of correlations, placing restrictions on the population effect sizes that researchers can potentially study. In addition to this, it is important to emphasize that when moving to a higher number of dimensions, the elements of the intermediate correlation matrix have to be found in a *pairwise* fashion (Vale & Maurelli, 1983). Each intermediate correlation is calculated for its own set of polynomial coefficients and plugged into a correlation matrix. Because of this, there is no guarantee that the intermediate correlation matrix will even be positive definite and very little research exists documenting the potential issues that may arise from this.

When compared to other data-generation algorithms, the 3rd-order polynomial method also displays certain features that require caution from researchers. From an empirical standpoint, Astivia and Zumbo (2015) demonstrated through simulation studies that the Fleishman (1978) algorithm generates downward-biased estimates of skewness and kurtosis even at sample sizes of 500. Few but very extreme values of kurtosis tended to be generated as well, particularly when compared with the Headrick (2002) 5th-order polynomial approach that offers better control of the higher moments of the data. In their simulation study, these issues ended up being sufficiently influential to the point that recommendations from previously-published literature would have changed if a different data-generating algorithm had been used. From a more formal, mathematical perspective, Foldnes and Grønneberg (2015) derived the probability distribution of the Vale-Maurelli (1983) transformation as an extension of the Gaussian copula and demonstrate that it shares many similarities with the multivariate normal distribution, particularly the absence of tail dependence. This is a potential issue for both previously-published studies and future simulations that rely on the 3rd-order polynomial transformation since it restricts the types of distributions from which researchers can actually generate data.

**4.1.2 Multiplicity of solutions to the Fleishman (1978) polynomials**

There exists one more property of the 3rd-order polynomial that has remained virtually unexplored in the literature with a few exceptions (e.g., Kraatz (2011), Luo (2011)). This property alludes to the fact that the system of equations found in (4.2)-(4.5) consists of high-degree polynomials with potentially more than one distinct solution.

Appendix E provides the full mathematical derivation to show that four sets of real-valued solutions exist for the system of equations described in (4.2)-(4.5). Of these four sets of solutions, two sets share the same $b$ and $d$ coefficients, albeit with opposite signs, because both coefficients in the equations appear either squared or multiplying one another, so that they end up being positive by themselves or when multiplied together. The effect that this switch of signs has on the potential distributions is merely a reflection alongside the y-axis so the variables end up being either skewed to the left or skewed to the right by the same amount. Coefficients $a$ and $c$, however, do change in absolute value for two of the sets so that, in total, there are two out of four completely distinct sets of solutions for each combination of skewness and kurtosis handled by the Fleishman method. To clarify, take the previous example of generating a variable that matches the skewness and kurtosis of the chi-square distribution with one degree of freedom, $\gamma_1 = \sqrt{8}, \gamma_2 = 12$. When applying the 3rd-order polynomial transformation, the potential coefficients for (4.1) that achieve these high-order moments could be:

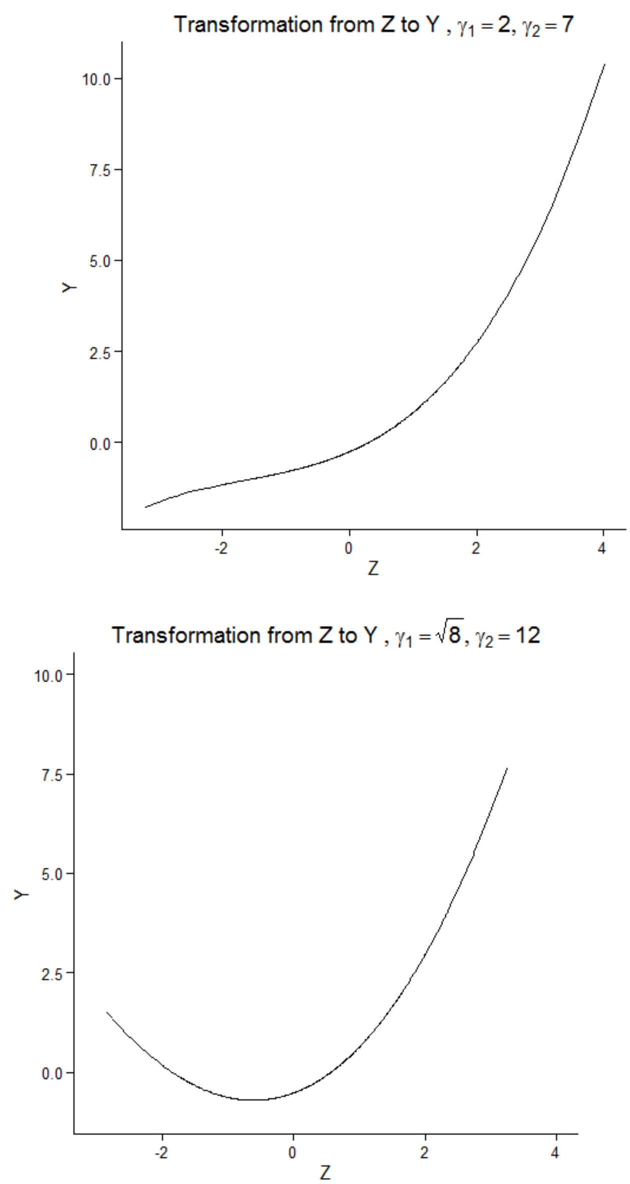
| | $a$ | $b$ | $c$ | $d$ |
|---|---|---|---|---|
| Set 1: | -0.5206759 | 0.614598122 | 0.5206759 | 0.020072438 |
| Set 2: | -0.5206759 | -0.614598122 | 0.5206759 | -0.020072438 |
| Set 3: | -0.7066366 | 0.036456329 | 0.7066366 | 0.000002020 |
| Set 4: | -0.7066366 | -0.036456329 | 0.7066366 | -0.000002020 |

Set 1 and Set 3 are the two completely distinct solutions and Set 2 and Set 4 are the y-axis reflected ones. If these polynomial coefficients were extended to the Vale-Maurelli (1983) multivariate setting, plugging them in equation (4.9) would also give different intermediate correlations, each with a potentially different set of bounds for $r_{Y_1Y_2}$.

Headrick and Kowalchuk (2007) provide one of the very few recommendations within the published literature regarding which set of solutions should be preferred over others by making the distinction between monotonic and non-monotonic polynomial transformations. Although their characterization of polynomial transformations is based on the ratio of the normal probability density function and the derivative of the polynomial transformation, a more intuitive way to distinguish them is to see whether or not the values of the standard normal variable $Z$ have a monotonic relationship with those of the non-normally transformed variable $Y$.

The top panel of Figure 4.1 shows the 3rd-order polynomial transformation corresponding to a

**Figure 4.1** Monotonic (top) and non-monotonic (bottom) transformations from the standard normal variate (*Z*) to the non-normal variate (*Y*)



population skewness of 2 and kurtosis of 7. The relationship between the original standard normal variable and the new non-normal variable is clearly monotonic and strictly increasing. Each and every value of $Z$ is paired with only one value of $Y$ and the curve does not switch direction or changes its trend. Compare this transformation with the chi-square distribution with one degree of freedom example, using the values $\gamma_1 = \sqrt{8}, \gamma_2 = 12$ on the bottom panel in Figure 4.1. This relationship is neither monotonic nor strictly increasing, since there exists a clear quadratic relationship between $Z$ and $Y$. Given that the chi-square distribution with one degree of freedom is defined as a squared standard normal variable, it should not be surprising that a similar type of relationship arises whenever the Fleishman transformation is used to mimic this distribution.

Headrick and Kowalchuk (2007) were able to derive the analytic probability density function (PDF) and cumulative density function (CDF) of the general case of the power polynomial transformation (which encompasses the 3rd-order polynomial transformation) under the assumption of monotonic transformations. They refer to non-monotonic transformations as generating "invalid" PDFs and CDFs given that their analytic form is not known. Kraatz (2011) questions the use of this term since the distribution functions of non-monotonic transformations do not have anything intrinsically wrong or "invalid" about them, they simply cannot be described under the framework developed by Headrick and Kowalchuk (2007). As shown in the previous example, "invalid" transformations can encompass well-known theoretical distributions (such as the chi-square) whose theoretical polynomial transform would be $Y = a + bZ^2$, with all the other powers of $Z$ beyond two having coefficients of 0.

Not all combinations of skewness / kurtosis allow for monotonic transformations and, in cases where they do, there is no guarantee that the multiple sets of solutions will only generate one type of them. Chen and Tung (2003) and Kraatz (2011) show that for the 3rd-order polynomial to be strictly monotonic, the inequality $c^2 - 3bd < 0$ among the polynomial coefficients must hold, further reducing the range of available skewness / kurtosis combinations within the approximate Fleishman parabola $\gamma_2 \geq 1.588\gamma_1^2 - 1.139$.

There is not much known about how the issue of multiple solutions impacts simulation practice among quantitative researchers in the social sciences. No information exists concerning whether certain sets of solutions are more prevalent than others and even less is known about the type of data that each type of solution generates. Because every software package that implements the Fleishman (1978) or the Vale-Maurelli (1983) algorithms uses its own specific optimization routines, it is currently unknown whether or not the synthetic datasets generated in each simulation design are even comparable or what type of distributions they can generate. Because of this, the studies presented here attempt to meet two specific goals. The first one is concerned with the type of data itself that each type of solution generates to see whether or not different sets of polynomial coefficients for the same combinations of skewness and kurtosis generate similar univariate and multivariate structures. The second goal is aimed at understanding the impact that this multiplicity of coefficients has on previously-published results. Since there is no way of knowing which polynomial solutions were being used in the literature, three published articles within the social sciences that have used this method will be chosen and sections of their simulation studies will be redone under both polynomial conditions to see whether or not their recommendations generalize beyond the type of solution selected. Using the guiding principle described in Headrick and Kowalchuk (2007), monotonic and non-monotonic transformations will be compared whenever available to see what kind of differences arise when one set of coefficients is used instead of the other one. Recommendations will be presented at the end of this article to help inform researchers who plan on using this method so they can be aware of potential issues that may influence the results of their simulation designs.

## 4.2 Method

The focus of these studies is solely on the Vale-Maurelli (1983) generalization of the Fleishman (1978) approach because the issue of solution multiplicity influences the multivariate extension just as much as it does the univariate case, with the added benefit that the potential impact it has on correlational structures can also be explored. All simulations and analyses were carried out in the *R* programming language version 3.3.1 (*R* Core Team, 2015), using the *lavaan* (Rosseel, 2012) and *psych* (Revelle, 2016) packages when conducting latent variable models. Whenever needed, the number of replications per condition was set at 10,000. The code to estimate the Fleishman polynomial coefficients and the intermediate correlation matrices was modified from the *simulateData()* function present in *lavaan* in order to have better control over the type of solutions generated. The three articles selected to guide the simulation designs are shown in Table 4.1.

**Table 4.1** Sample sizes and skewness / kurtosis combinations for each article.

| Article | Sample Size | (Skewness, Kurtosis ) |
|---|---|---|
| Curran, West and Finch (1996) | 100, 200, 500, 1,000 | (2,7) (3,21) |
| Finch, West, and MacKinnon (1997) | 150, 250, 500, 1,000 | (2,7) (3,21) |
| Skidmore and Thompson (2011) | 10, 20, 40, 60, 100, 200 | (1,1) (-1.5, 3.5) |

The first two articles were selected both because of the potential impact they have had on data analysis practice (the Curran, West and Finch (1996) article, for instance, has been cited more than 2000 times) and because of their use of popular simulation designs within Structural Equation Models (SEM): the correlated three factor model and the latent mediation model. The skewness and kurtosis combinations used by the authors are also common in the literature, with published articles implementing either the exact same values (e.g. Enders, 2001; Nevitt & Hancock, 2001; Rhemtulla, Brosseau-Liard & Savalei, 2012; Seco, Gras & Garcia, 2007) or values close to them (e.g. Kelley & Pornprasertmanit, 2016). The third article was selected to highlight the use of the Vale-Maurelli (1983) method outside of SEM and because it allows to explore the impact of the multiplicity of solutions issue on lower-dimensional (i.e. bivariate) settings. Only one full simulation study was reproduced for reasons of space.

**4.2.1 Study 1**

Study 1 aimed to investigate the properties of each set of solutions to the Fleishman system of polynomial equations. A grid search over the potential parameter space was conducted by setting the starting values of the parameters at 1 and moving up to 20 in incremental steps of .0005. Preliminary simulations had shown that, when the starting values were greater than 20, the optimization process settled in too many bad solutions and non-convergences, therefore, no values beyond 20 were explored. Once the solutions were found, they were categorized as yielding either monotonic or non-monotonic transformations and the bounds to the final correlation $r_{Y_1 Y_2}$ were calculated. These bounds were compared to the bounds obtained by the generate-sort-correlate (GSC) method described in Demritas and Hedeker (2011) where they use the Fréchet-Hoeffding theorem to approximate the boundaries of correlation coefficients under non-normal conditions. Ideally, the closer the range of the final correlation $r_{Y_1 Y_2}$ is to that specified by the GSC method, the better these intermediate correlations will be able to approximate theoretical distributions.

Once the polynomial coefficients and intermediate correlations were generated, the full Fleishman (1978) transformation was done in a simulation study, each time using the two distinct sets of solutions (i.e., solutions defined by different sets of $a$ and $c$ coefficients) in order to document which one better approximated the population values of univariate skewness/kurtosis at the sample sizes reported in Curran, West and Finch (1996). The sample sizes of this article were the only ones used both because they are very similar to those used in Finch, West, and MacKinnon (1997) and because previous simulation work (Astivia and Zumbo, 2015) has shown that, at smaller sample sizes, the estimates of higher order moments are simply too variable to be able to appreciate any clear differences due to the choice of polynomial solution. Even though the 3rd-order polynomial method induces non-normality through the one-dimensional marginal distributions, the Vale-Maurelli (1983) method and the articles shown in Table 4.1 pertain to the study of multivariate structures. Because of this, Mardia's (1970) multivariate measure of kurtosis was calculated, as well as the asymptotic variance-covariance matrix of the sample estimates, $\Gamma$. Bentler (1995) defines this matrix as

$$[\Gamma]_{ij,kl} = \sigma_{ijkl} - \sigma_{ij}\sigma_{kl}, \tag{4.10}$$

where $\sigma_{ij}$ and $\sigma_{kl}$ are the covariances of observed variables $x_i, x_j, x_k, x_l$ and $\sigma_{ijkl} = E[(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l)]$. Previous results (e.g. Astivia & Zumbo, 2015) had indicated that different types of data-generation mechanisms can result in different estimates of $\Gamma$, but this hypothesis was never explicitly tested. In order to measure the differences between matrices, a small simulation with 100 replications at sample sizes of 10 million was conducted. In each replication, the $\hat{\Gamma}$ matrix was calculated and the average was taken among those 100 replications to yield a pseudo-population matrix $\Gamma$. During the simulation study comparing polynomial coefficient solutions, the Frobenius distance between the sample $\hat{\Gamma}$ and its pseudo-population analog $\Gamma$ was calculated, where the Frobenius distance is defined as $\sqrt{Tr((\hat{\Gamma} - \Gamma)(\hat{\Gamma} - \Gamma)')}$ in order to understand the ways in which each polynomial solution influences the properties of this matrix.

**4.2.2 Study 2**

Study 2 focused on documenting the impact that the choice of polynomial solution had on the conclusions obtained by previously-published simulation studies. As there is no way to know which polynomial coefficients were used in the Table 4.1 articles (or any other articles for that matter), it would be of interest to see whether or not any of the recommendations suggested by the authors change depending on which coefficients are used to induce the non-normality of the data.

Because Study 2's purpose is mostly illustrative (and to economize the space in this chapter), only the first simulation of Curran, West and Finch (1996) was implemented. A path diagram of the population factor model can be found in Appendix G. It is the same one found in the appendix of Astivia & Zumbo (2015) or Figure 1 of Curran, West and Finch (1996). It consists of a 3-factor model with 3 indicators per factor (for a total of 9 indicators). Factors are correlated at 0.3 in the population, with population loadings of 0.7 and error variances of 0.51. For its simulation conditions, Study 2 uses the same sample sizes and levels of moderate and severe skewness/kurtosis combinations as those found in Table 4.1 of the present article (i.e. $n$ = 100, 200, 500 and 1000 and skewness/kurtosis of 2/7 and 3/21). The outcomes of the study are the observed $\chi^2$ statistic of model fit, average bias, and empirical rejection rates when two different estimation methods, normal theory Maximum Likelihood (ML), and Asymptotic Distribution Free estimation (ADF) and one correction, Satorra-Bentler (SB), were implemented. The only three differences between this present study and the original study design are the number of replications per condition (200 for the original study and 10,000 for the present study to ensure maximum stability), the software where the simulations were implemented (EQS for the original study and the *lavaan* package for the current study), and differences in polynomial solutions for data-generation purposes. Because the EQS software was used originally, the *mimic="EQS"* setting was defined in *lavaan* to ensure maximum comparability between the two.

## 4.3 Results

### 4.3.1 Study 1

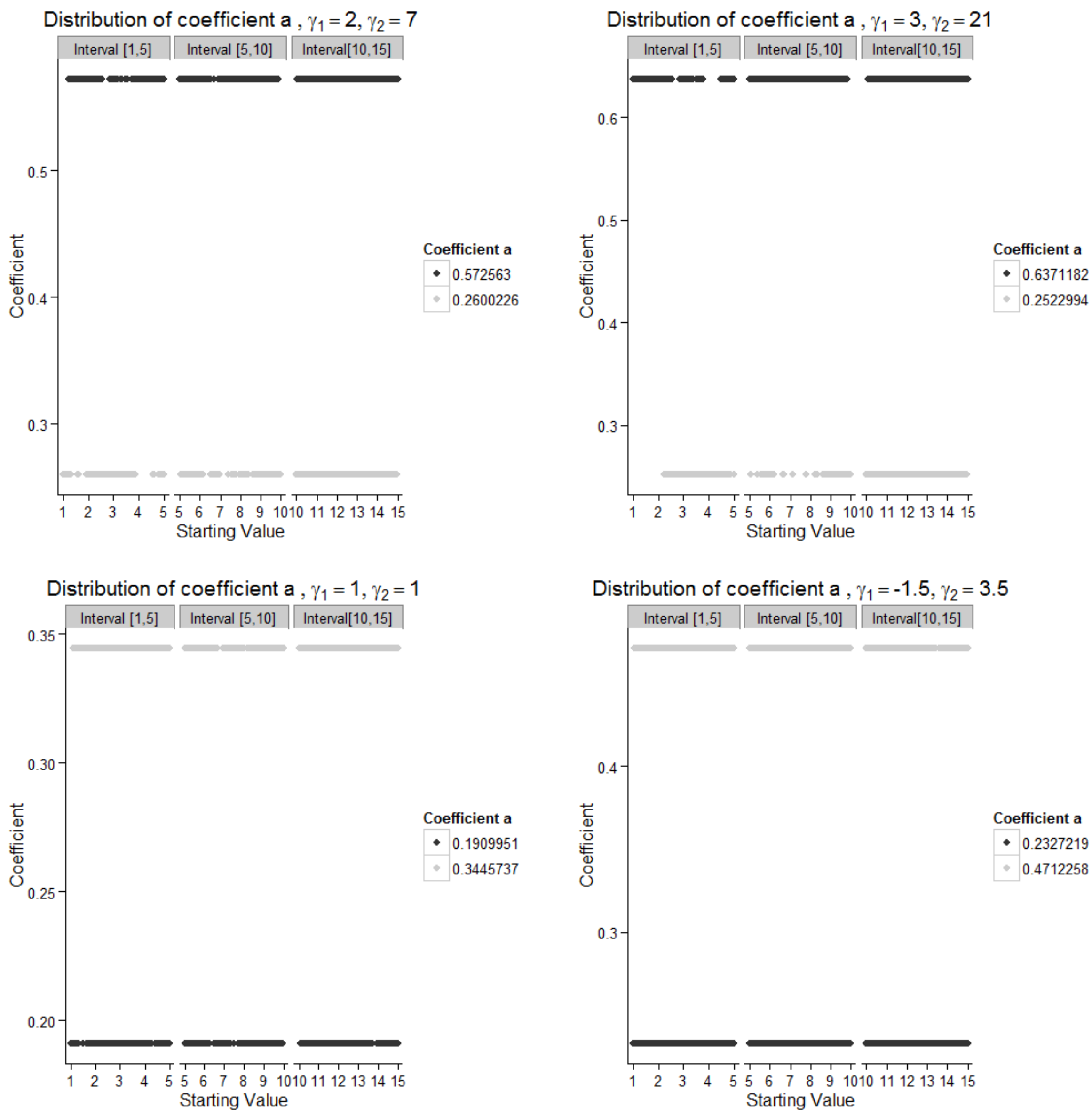
Table 4.2 shows the two different sets of polynomial coefficients for the skewness/kurtosis combinations found in the Table 4.1 articles. As it was described in the previous section, only the two qualitatively distinct polynomial solutions (i.e., those with different *a* and *c* coefficients) will be analyzed.

**Table 4.2** Two distinct sets of polynomial solutions for the skewness/kurtosis values.

| | | Solution 1 of polynomial coefficients | | | |
|---|---|---|---|---|---|
| Skewness | Kurtosis | $a_1$ | $b_1$ | $c_1$ | $d_1$ |
| 2 | 7 | 0.260 | -0.761 | -0.260 | -0.053 |
| 3 | 21 | 0.252 | -0.418 | -0.252 | -0.147 |
| 1 | 1 | 0.344 | 1.216 | -0.344 | -0.136 |
| -1.5 | 3.5 | 0.232 | -0.886 | -0.232 | -0.018 |
| | | Solution 2 of polynomial coefficients | | | |
| Skewness | Kurtosis | $a_2$ | $b_2$ | $c_2$ | $d_2$ |
| 2 | 7 | 0.572 | 0.848 | -0.572 | -0.108 |
| 3 | 21 | 0.637 | 0.681 | -0.637 | 0.148 |
| 1 | 1 | 0.190 | 1.017 | -0.190 | -0.018 |
| -1.5 | 3.5 | 0.471 | -1.049 | -0.471 | 0.122 |

In the parameter space defined on the interval [1, 20], 38000 solutions to the system of polynomial equations were found for each combination of skewness and kurtosis values. The Solution 1 set of coefficients was found 43.14% of the time for the skewness/kurtosis combination ($\gamma_1 = 2, \gamma_2 = 7$), 30.75% for the combination ($\gamma_1 = 3, \gamma_2 = 21$), 71.61% for the values ($\gamma_1 = 1, \gamma_2 = 1$) and 66.42% for ($\gamma_1 = -1.5, \gamma_2 = 3.5$). Figure 4.2 below highlights the fact that each solution (identified solely by coefficient *a*) is not evenly distributed in the parameter space. For purposes of identifying changes in the frequency of solutions, the scale of the horizontal axis was subdivided in intervals [1,5], [5,10] and [10,15]. In the [15,20] subsection no discernible pattern was found so that subinterval was not plotted. When the values of the 3$^{rd}$ and 4$^{th}$ order moment are similar (as presented in the bottom two panels) each solution appears more or less with the same frequency, regardless of the starting value. When the skewness and kurtosis are very different, there are certain values of the parameter space that tended to favour certain solutions. The clearest case of this can be seen in the ($\gamma_1 = 3, \gamma_2 = 21$) combination where Solution 1 is considerably less frequent, particularly in the [5,10] subinterval. For parameter space values less than 2, Solution 1 rarely appears, but for the values between 4 and 5, Solution 2 shows a clear gap where only coefficients for Solution 1 are found.

**Figure 4.2** Distribution of coefficient *a* across different starting values on the interval [1,20]




A similar (albeit less pronounced) pattern is also found for the set ($\gamma_1 = 2, \gamma_2 = 7$) although, in this case, it seems that at the lower end of the parameter space, Solution 1 appears more frequently. Towards the higher values of the parameter space the solution values appear to be more or less equally likely.

Table 4.3 further characterizes the type of solution depending on whether or not it results on a monotonic or non-monotonic transformation from $Z$ to $Y$.

**Table 4.3** Description of solutions by type of transformation (monotonic vs non-monotonic), range of possible final correlations, theoretical range of correlations as per Demritas and Hedeker (2011) and range coverage.

| | | Type | $r_{Y_1Y_2}$ range | Theoretical Range | Coverage |
|---|---|---|---|---|---|
| $\gamma_1 = 2,$ | Solution 1 | Monotonic | (-0.244, 0.514) | (-0.727, 1) | 43.97% |
| $\gamma_2 = 7$ | Solution 2 | Non-Monotonic | (-0.029, 1) | (-0.727, 1) | 59.59% |
| $\gamma_1 = 3,$ | Solution 1 | Monotonic | (-0.745, 1) | (-0.745, 1) | 100% |
| $\gamma_2 = 21$ | Solution 2 | Monotonic | (-0.592, 1) | (-0.745, 1) | 91.23% |
| $\gamma_1 = 1,$ | Solution 1 | Non-Monotonic | (-0.525, 1) | (-0.861, 1) | 81.91% |
| $\gamma_2 = 1$ | Solution 2 | Non-Monotonic | (-0.854, 1) | (-0.861, 1) | 99.58% |
| $\gamma_1 = -1.5,$ | Solution 1 | Monotonic | (-0.783, 1) | (-0.784, 1) | 99.93% |
| $\gamma_2 = 3.5$ | Solution 2 | Non-monotonic | (-0.111, 1) | (-0.784, 1) | 62.30% |

For each combination of skewness and kurtosis, every possible type of transformation was found, with some combinations being either strictly monotonic or strictly non-monotonic. There does not seem to be a relationship between the type of transformation and the width of the correlation ranges, although instances of more extreme moments, $(\gamma_1 = 3, \gamma_2 = 21)$ and $(\gamma_1 = 2, \gamma_2 = 7)$, were associated with the more restricted possible ranges. The plausible range for the final correlation $r_{Y_1Y_2}$ spanned more than 50% of the theoretical range most of the time (with one exception for the $\gamma_1 = 2, \gamma_2 = 7$ combination) and it appears that the boundaries were most pronounced towards the negative end of the correlation ranges as none of them were able to reach the lower bound of -1.

To understand the potential differences in sample estimates of skewness and kurtosis contingent on the choice of polynomial coefficients, Table 4.4 shows the results of a simulation study using the 3$^{rd}$-order polynomial transform where the only difference in the data-generation process is the type of coefficients employed. Although data under both conditions resulted in downward-biased estimates of skewness and kurtosis, it appears that there was always a solution in which both the bias and the variability were lower when compared to the other one. In the cases where the skewness/kurtosis combination resulted in both monotonic and non-monotonic transformations, non-monotonic transformations (Solution 2 for the $(\gamma_1 = 2, \gamma_2 = 7)$ and $(\gamma_1 = -1.5, \gamma_2 = 3.5)$ conditions) were always considerably more downward biased and with much higher standard deviations when compared with their monotonic counterparts. It is of particular interest to notice that for the skewness/kurtosis combination of 3/21 (where both coefficients result in a monotonic transformation) the sample skewness tends to be overestimated, on average, as opposed to underestimated. For the cases where skewness and kurtosis were smaller, $(\gamma_1 = 1, \gamma_2 = 1)$ and $(\gamma_1 = -1.5, \gamma_2 = 3.5)$, the

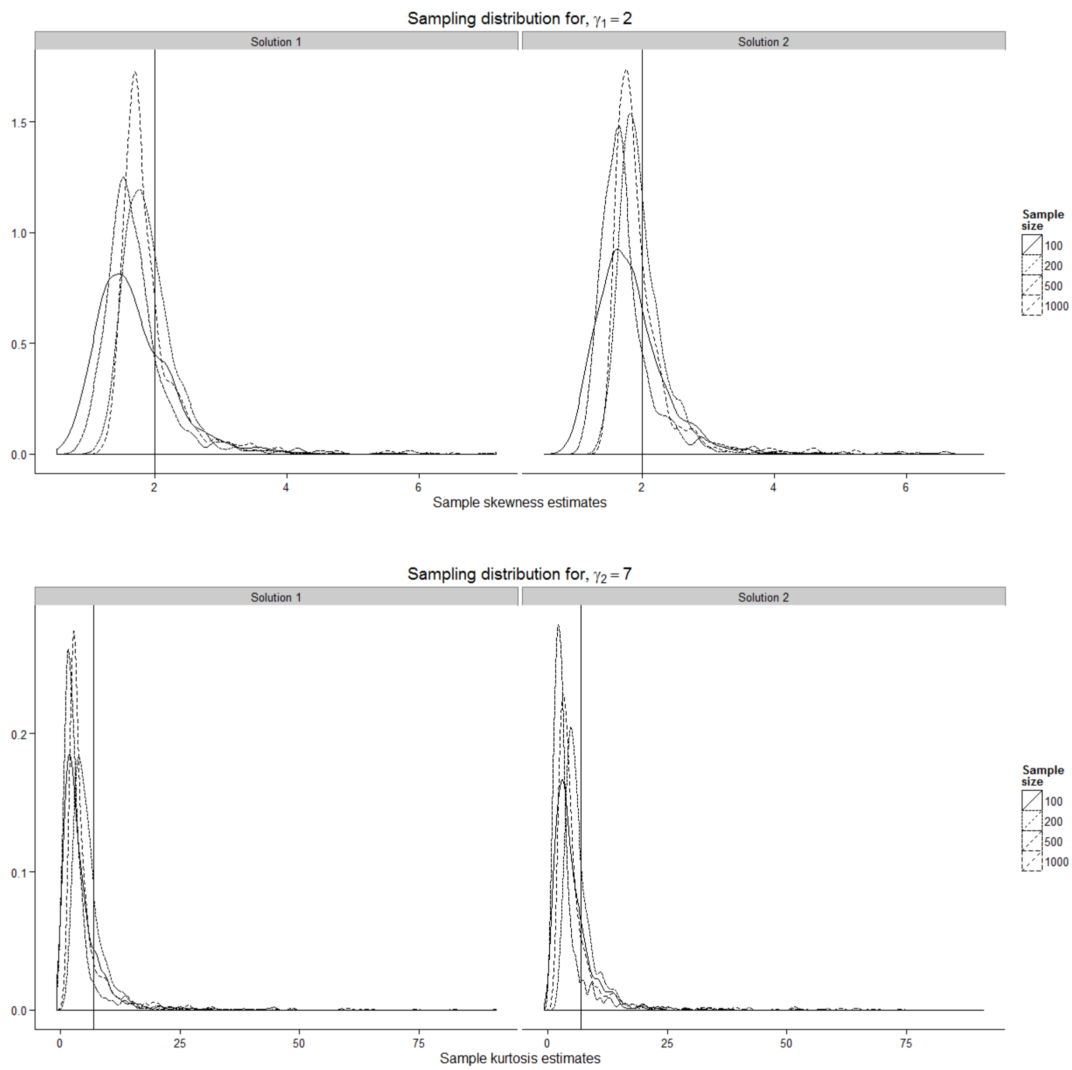
**Table 4.4** Mean and standard deviation (SD) of sample estimates of skewness and kurtosis at different sample sizes (n)

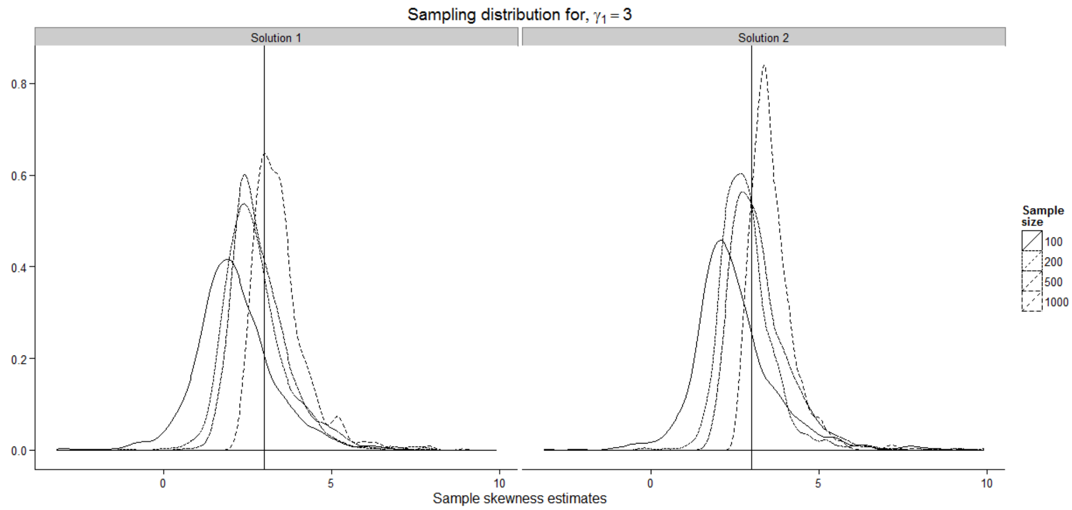
| | | n=100 | | n=200 | | n=500 | | n=1000 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| $\gamma_1 = 2$, $\gamma_2 = 7$ | Solution1 | 1.671 | 0.618 | 1.839 | 0.557 | 1.929 | 0.424 | 1.953 | 0.330 |
| | | 4.126 | 4.278 | 5.221 | 4.855 | 6.103 | 4.009 | 6.541 | 3.674 |
| | Solution 2 | 1.710 | 0.603 | 1.789 | 0.604 | 1.883 | 0.502 | 1.934 | 0.481 |
| | | 3.204 | 4.878 | 4.202 | 6.291 | 5.141 | 6.225 | 5.909 | 7.241 |
| $\gamma_1 = 3$, $\gamma_2 = 21$ | Solution 1 | 2.138 | 1.224 | 2.457 | 1.219 | 2.687 | 1.029 | 2.829 | 0.849 |
| | | 9.268 | 8.065 | 12.585 | 11.766 | 15.749 | 14.250 | 17.789 | 12.607 |
| | Solution 2 | 2.950 | 0.933 | 3.223 | 0.944 | 3.471 | 0.857 | 3.584 | 0.753 |
| | | 11.152 | 8.970 | 14.182 | 11.350 | 17.623 | 13.484 | 19.420 | 14.262 |
| $\gamma_1 = 1$, $\gamma_2 = 1$ | Solution 1 | 0.901 | 0.337 | 0.942 | 0.352 | 0.962 | 0.362 | 0.983 | 0.261 |
| | | 0.072 | 2.325 | 0.346 | 3.023 | 0.504 | 4.579 | 0.716 | 3.285 |
| | Solution 2 | 0.791 | 0.278 | 0.964 | 0.113 | 0.982 | 0.140 | 0.989 | 0.096 |
| | | 0.322 | 2.182 | 0.752 | 2.066 | 0.916 | 0.649 | 0.977 | 0.457 |
| $\gamma_1 = -1.5$, $\gamma_2 = 3.5$ | Solution 1 | -1.337 | 0.453 | -1.408 | 0.377 | -1.457 | 0.271 | -1.493 | 0.199 |
| | | 2.307 | 2.703 | 2.785 | 2.948 | 3.158 | 2.069 | 3.325 | 1.725 |
| | Solution 2 | -1.321 | 0.483 | -1.403 | 0.552 | -1.430 | 0.480 | -1.451 | 0.338 |
| | | 1.634 | 3.794 | 2.270 | 5.851 | 2.545 | 4.882 | 2.847 | 4.421 |

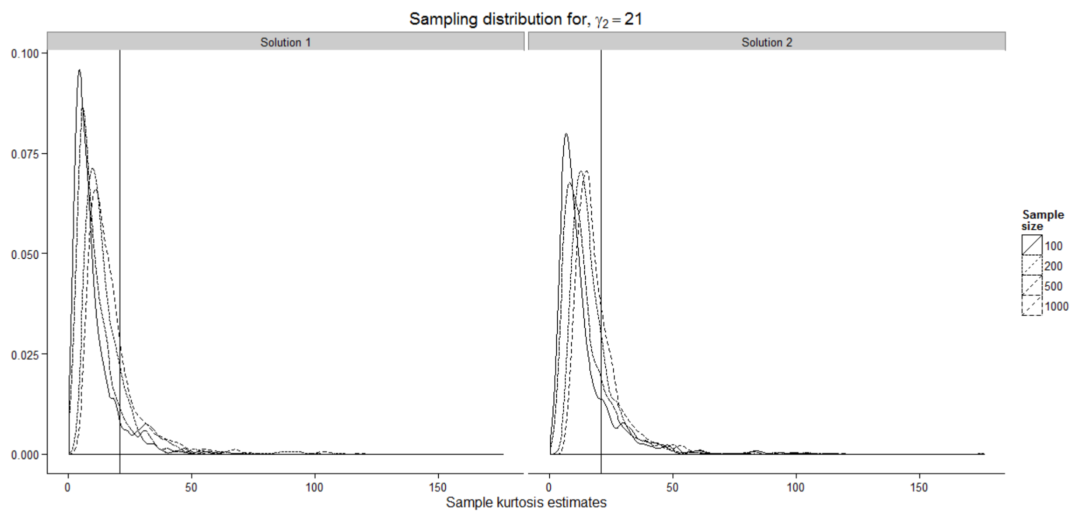
average sample estimates converged much faster to their population-level values than in the cases where these values were larger.

Figure 4.3 shows the empirical density plots for the most extreme combinations of skewness and kurtosis values at the different sample sizes. In each panel it is possible to observe that the type of solution does exert a certain influence on the overall behaviour of the sample estimates of these higher-

**Figure 4.3** Empirical density plot of for population values of $(\gamma_1 = 3, \gamma_2 = 21)$ and $(\gamma_1 = 2, \gamma_2 = 7)$ at sample sizes 100, 200, 500 and 100.






order moments. In each case, the moments are underestimated (with the exception of $\gamma_1 = 3$ in Solution 2) because the peak of each distribution lies to the left of the vertical line denoting the population value, evidencing that the regions of highest density are concentrated towards the lower end of the distribution. At the higher end, however, the tails of the distributions seem to behave differently depending on which solution one focuses. For the case of $\gamma_2 = 7$, the monotonic Solution 1 seems to smooth itself much more evenly than the non-monotonic Solution 2, where there is some variability of the estimates as the curve slopes down. Solution 1, however, shows some pockets of density very far out in the x-axis (with two recognizable sections beyond the mark of 75). Solution 2 does not show said pockets, albeit it does seem to have a tendency to generate higher values of kurtosis up to around the mark of 75 on the x-axis. For $\gamma_2 = 21$ the curves show substantial overlap in Solution 2, regardless of sample size. Solution 1, on the other hand, shows a distinct high peak at n=100, far from 21, with the subsequent curves progressively approaching the population value. In the case of $\gamma_2 = 7$ the pattern is not as clear, although it appears that Solution 1 curves are closer to the population parameter. The curves for skewness, $\gamma_1 = 3$, are mostly bell-shaped with the n=1000 for Solution 1 centered in its original population value. The curves for $\gamma_1 = 2$ do show a certain longer tail with irregularities as they spread out.

In order to gain better insight into the different nature of the multivariate distributions that each type of solution can generate, Table 4.5 documents the mean Mardia's kurtosis from the simulation study and its z-score transformation. Each solution generated data with different types of multivariate structures, with certain solutions (mostly the non-monotonic Solution 2) generating multivariate kurtosis estimates that are around 10 points lower in the standard normal metric than their counterpart. The value of Mardia's kurtosis is only asymptotically unbiased, so it is not a surprise to see that it increases as sample size increases as well, pointing towards the fact that for these values of skewness/kurtosis, the corresponding population value of multivariate kurtosis is probably much higher. It is also possible to see that as sample size grows larger, the differences between solutions become smaller. For the most extreme combination of values $(\gamma_1 = 3, \gamma_2 = 21)$, however, this difference reaches its highest at the largest sample size but for the lower combinations $(\gamma_1 = -1.5, \gamma_2 = 3.5)$ and$(\gamma_1 = 1, \gamma_2 = 1)$ the difference at a sample size of 1000 is barely over 5 points in the standard normal metric. Figure 4.4 shows the average Frobenius distance for each type of polynomial solution at the different sample sizes used in the simulation study. Overall, it is possible to see clear distinctions in the way the $\Gamma$matrix is calculated for each case. It is difficult to discern a pattern depending on the type

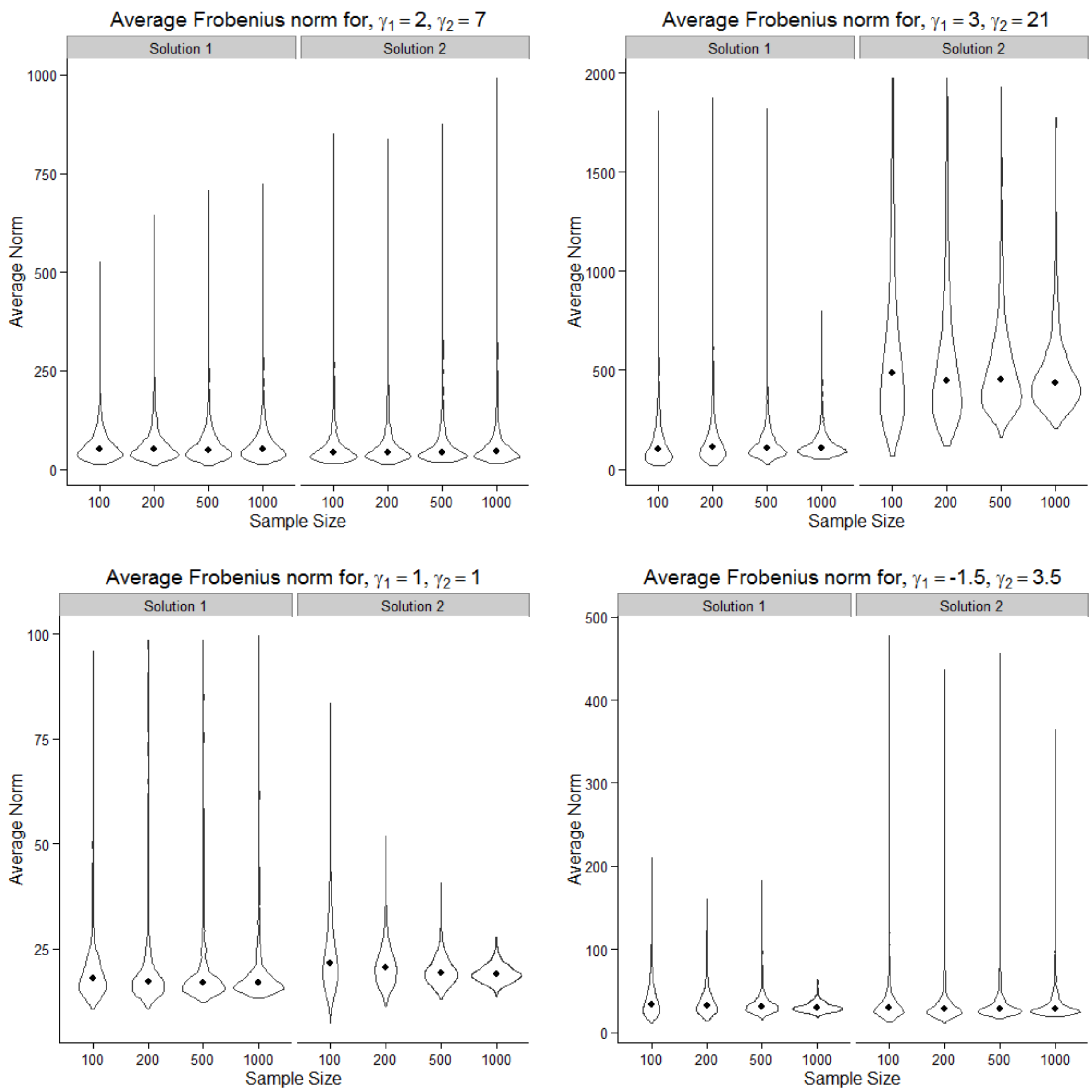
**Table 4.5.** Mean Mardia's kurtosis (Mku) and its corresponding average z-score at sample sizes 100, 200, 500 and 1000.

| | | n = 100 | | n = 200 | | n = 500 | | n = 1000 | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mku | z | Mku | z | Mku | z | Mku | z |
| $\gamma_1 = 2,$ $\gamma_2 = 7$ | Solution 1 | 144.62 | 16.21 | 163.38 | 32.35 | 180.15 | 64.48 | 188.09 | 100.11 |
| | Solution 2 | 119.85 | 7.4 | 138.01 | 19.6 | 166.42 | 53.56 | 181.66 | 92.88 |
| $\gamma_1 = 3,$ $\gamma_2 = 21$ | Solution 1 | 175.79 | 27.28 | 221.08 | 61.34 | 279.43 | 143.21 | 315.28 | 243.02 |
| | Solution 2 | 203.79 | 37.24 | 247.79 | 74.77 | 292.39 | 153.66 | 332.44 | 262.31 |
| $\gamma_1 = 1,$ $\gamma_2 = 1$ | Solution 1 | 106.05 | 2.5 | 111.58 | 6.32 | 117.15 | 14.42 | 120.67 | 24.35 |
| | Solution 2 | 107.11 | 2.88 | 111.51 | 6.29 | 114.41 | 12.25 | 115.28 | 18.3 |
| $\gamma_1 = -1.5,$ $\gamma_2 = 3.5$ | Solution 1 | 132.49 | 16.83 | 135.49 | 18.33 | 149.88 | 40.42 | 151.22 | 58.66 |
| | Solution 2 | 114.91 | 5.65 | 121.72 | 11.41 | 142.75 | 34.76 | 146.18 | 53.01 |

of solution, though. For example, for the combination $(\gamma_1 = 2, \gamma_2 = 7)$, Solution 2 (non-monotonic) showed a distinct tendency to generate more extreme values, resulting in higher variability of the estimates. However, in the case of $(\gamma_1 = 1, \gamma_2 = 1)$, Solution 1 (monotonic) was the one showing much longer tails and extreme values, even at the largest sample of 1000. At larger sample sizes (n=500 and n=1000), there is a larger density closer to the median of the estimates (marked with a dot) and shorter tails, although this also differed by solution type. Even though the median values across replications in each sample size are somewhat comparable, a clear exception is the most extreme combination of skewness/kurtosis values, $(\gamma_1 = 3, \gamma_2 = 21)$, which shows much fatter tails for Solution 2, implying higher values and higher variability. Solution 2 in this case also shows a much higher median Frobenius distance (around a value of 500) when compared to Solution 1 (closer to 100), suggesting that $\Gamma$ matrix for both types of solutions may be approximated differently. Overall, larger values of $\gamma_1, \gamma_2$ were associated with much higher sample estimates that took longer to converge to their population value than when $\gamma_1, \gamma_2$ are small. At n=1000, for instance, the norm for cases $(\gamma_1 = 1, \gamma_2 = 1)$ and $(\gamma_1 = -1.5, \gamma_2 = 3.5)$, already show plots with high-density regions closer to the centre of the distribution as opposed to the extremes. This was not constant across solution types though, as Solution

1 for the $(\gamma_1 = 1, \gamma_2 = 1)$ combination and Solution 2 for the $(\gamma_1 = -1.5, \gamma_2 = 3.5)$ combination still show much longer tails.

**Figure 4.4** Violin plots for the mean Frobenius norm at sample sizes 100, 200, 500 and 1000. Median value marked at the dot.




### 4.3.2 Study 2

Table 4.6 reproduces sections of Table 1 on page 22 of the Curran et al. (1996) study. It contrasts the average chi-square value and empirical rejection rates of the original authors to those obtained by

Solution 1 and Solution 2 of the 3rd-order polynomials. In both instances of skewness/kurtosis combinations, the mean chi-square values across conditions are relatively comparable with the one obtained by the original authors, although Solution 2, in both cases, resulted in

**Table 4.6** Mean chi-square values and empirical rejection rates from the Curran, West and Finch (1996) (CWF 1996) article obtained for both polynomial solutions at sample sizes (N) for normal theory maximum likelihood (ML), asymptotic distribution free (ADF) and Satorra-Bentler correction (SB).

| $\gamma_1 = 2, \gamma_2 = 7$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| n | Method | Expected | CWF (1996) | Solution 1 | Solution 2 | Nominal rejection | CWF (1996) | Solution 1 | Solution 2 |
| 100 | ML | 24 | 29.35 | 29.86 | 29.238 | 5% | 20% | 21.20% | 17.80% |
| | SB | 24 | 26.06 | 26.54 | 26.518 | 5% | 8.50% | 8.70% | 7.10% |
| | ADF | 24 | 38.04 | 31.007 | 30.531 | 5% | 49% | 47.80% | 23.10% |
| 200 | ML | 24 | 30.15 | 29.213 | 29.29 | 5% | 25% | 20.30% | 19.70% |
| | SB | 24 | 25.44 | 24.913 | 25.254 | 5% | 8% | 7.30% | 6.01% |
| | ADF | 24 | 29.27 | 27.279 | 27.35 | 5% | 19% | 14.30% | 9.40% |
| 500 | ML | 24 | 31.26 | 29.939 | 29.75 | 5% | 24% | 22.30% | 22.30% |
| | SB | 24 | 25.44 | 24.453 | 24.552 | 5% | 6.90% | 6.50% | 6.30% |
| | ADF | 24 | 26.42 | 25.716 | 25.515 | 5% | 6.70% | 6.30% | 7.72% |
| 1000 | ML | 24 | 30.78 | 30.133 | 30.235 | 5% | 24% | 21.40% | 23.30% |
| | SB | 24 | 24.77 | 24.406 | 24.461 | 5% | 7.50% | 6.40% | 6.00% |
| | ADF | 24 | 25.36 | 25.08 | 25.031 | 5% | 7.50% | 6.20% | 5.80% |

| $\gamma_1 = 3, \gamma_2 = 21$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| n | Method | Expected | CWF (1996) | Solution 1 | Solution 2 | Nominal rejection | CWF (1996) | Solution 1 | Solution 2 |
| 100 | ML | 24 | 33.54 | 31.46 | 27.747 | 5% | 30.00% | 34.10% | 20.30% |
| | SB | 24 | 27.26 | 24.41 | 25.644 | 5% | 13.00% | 10.40% | 6.79% |
| | ADF | 24 | 44.82 | 45.56 | 37.761 | 5% | 68.00% | 61.24% | 52.82% |
| 200 | ML | 24 | 34.4 | 33.98 | 29.949 | 5% | 36.00% | 35.60% | 28.30% |
| | SB | 24 | 25.8 | 25.12 | 20.463 | 5% | 6.50% | 7.00% | 6.46% |
| | ADF | 24 | 31.29 | 30.66 | 25.284 | 5% | 25.00% | 22.10% | 15.50% |
| 500 | ML | 24 | 35.55 | 32.95 | 31.414 | 5% | 40.00% | 42.50% | 32.40% |
| | SB | 24 | 24.85 | 24.78 | 24.799 | 5% | 8.50% | 7.80% | 6.70% |
| | ADF | 24 | 26.83 | 27.01 | 25.182 | 5% | 8.50% | 7.20% | 5.40% |
| 1000 | ML | 24 | 37.4 | 37.6 | 33.457 | 5% | 48.00% | 43.20% | 36.00% |
| | SB | 24 | 25.01 | 24.96 | 24.525 | 5% | 7.00% | 5.80% | 5.20% |
| | ADF | 24 | 25.47 | 25.11 | 24.929 | 5% | 7.20% | 6.20% | 5.10% |

somewhat lower values when compared to those shown in Solution 1. The differences between solutions were more pronounced, though, when looking at the empirical rejection rates. In both instances, Solution 1 was much closer to the original results than Solution 2, with Solution 2 being lower (sometimes by around 10 percent points) uniformly across the different types of chi-square estimates. For the ($\gamma_1 = 2, \gamma_2 = 7$) condition, the overall pattern of results is similar to the one reported by Curran et al. (1996) where the Satorra-Bentler correction outperforms both ML and ADF estimation (with the exception of the largest sample size). For the ($\gamma_1 = 3, \gamma_2 = 21$) condition, however, this pattern is not consistently present, with the ADF estimator approaching its nominal rejection rate of 5% much faster than the Satorra-Bentler correction. In both cases, it appears that estimating the model using data generated under Solution 2 resulted in a faster convergence to the chi-square distribution than under Solution 1. Overall, however, the differential effect that the type of solution had on the rejection rates disappears as sample size increases, as it is possible to see both results looking progressively similar at larger and larger samples.

### 4.3 Discussion

Being aware of the properties of data-generation algorithms is central to the design and implementation of Monte Carlo simulation studies. The issue of multiple solutions to the Fleishman (1978) polynomial equations is an example of these properties, but is also one that has received little attention in the literature in spite of the popularity of the method. With the exception of the work of Kraatz (2011) and Luo (2011), not much research exists regarding the potential impact that these solutions may have had in previously-published results, even though this algorithm continues to be routinely used. For instance, Google Scholar reports 18 and 31 citations of the Vale-Maurelli (1983) and Fleishman (1978) articles respectively in 2016, with more than half of those citations coming from articles using both methods in robustness-type Monte Carlo simulations.

Several important conclusions from the solution multiplicity issue can be highlighted in order to better inform the design and practice of simulation research. The first one is that, for each skewness/kurtosis combination studied, there does not seem to be a clear pattern of which set of polynomial roots is more frequent than the other when solving the system of equations. Although it appears that, for lower values of $\gamma_1, \gamma_2$, each solution is more or less equally likely (i.e., about 50% of the time Solution 1 was obtained and the other half of the time Solution 2 was obtained), for higher values one of the solutions is more frequent than the other. It would be beneficial for researchers to be aware that, if they are using low or moderate values of population skewness and kurtosis, there is some uncertainty regarding the type of solution on which their optimizer might settle. A second feature to point out is that there exists a differential effect that the type of solution had on the quality of the data that were generated. When both monotonic and non-monotonic transformations were available, solutions that implied a monotonic transformation resulted in overall less biased and less variable estimates of skewness and kurtosis. This apparent "good behaviour" of monotonic solutions was found previously in Kraatz (2011) and, although not referenced explicitly, the framework developed by Headrick and Kowalchuk (2007) labels these transformations as having a "valid" probability density function. Although it seems like polynomial solutions that imply monotonic transformations should be preferred, it is important to remind the reader that, even when no monotonic solution existed (or when both solutions were monotonic), one solution still showed less bias and variability, so recommending the choice of monotonic solutions may only work in the condition that both types are available.

Regarding the properties of the multivariate transformation (Vale & Maurelli, 1983), several new insights were found in this article that, to the authors' knowledge, are not yet present in the published literature. The first one is the fact that, even though quantitative researchers may indeed be limited in their ultimate choice of final correlation $r_{Y_1Y_2}$, the theoretical range that $r_{Y_1Y_2}$ can actually span is sometimes well approximated by the interplay of $r_{Z_1Z_2}$ and the polynomial coefficients. Demirtas and Hedeker (2011) lament the fact that, even though the Fréchet–Hoeffding bounds on the correlation have been known for more than 50 years, their use has not permeated the daily practice of applied statistics. Perhaps this situation has prompted a criticism of the available correlation range of the Vale-Maurelli method (and other polynomial-based transformations) that is not necessarily warranted. The Demirtas-Hedeker process is relatively straightforward to implement so it would be of benefit for applied researchers to check what range of correlations they can realistically choose from when studying non-normal distributions before selecting a final correlation $r_{Y_1Y_2}$. It also appears that, in most cases, the positive range of the Pearson correlation [0,+1] is well-approximated by the combinations of skewness/kurtosis being selected so this may not pose a great limitation as long as the skewness and kurtosis values point in the same direction (i.e., variables being positively skewed or negatively skewed simultaneously or both platykurtic or leptokurtic). These bounds would probably be much narrower if variables have their higher-order moments going in opposite directions.

Another new finding that was hinted by Astivia and Zumbo (2015) is that, even though the different solutions may asymptotically generate the same univariate skewness and kurtosis values, they do not necessarily generate the same multivariate structure at all sample sizes. Table 4.5 and Figure 4.4 point towards this fact. Of peculiar interest is the ( $\gamma_1 = 3, \gamma_2 = 21$) case where the median value of the Fobrenius distance metric is substantially different between Solution 1 (about 100) and Solution 2 (about 500). Two potential explanations can be offered regarding this result. The first one is that Solution 2 is over-approximating the distance between $\Gamma$and $\hat{\Gamma}$ and much larger samples are needed for it to converge. Although this is a plausible explanation, it would not elucidate why the median points at different sample sizes are clustered together (and far away from those of Solution 1). A second explanation, however, could be that the actual population values of $\Gamma$ are different for Solution 1 and Solution 2, which could also explain the disparate empirical rejection rates found in Study 2. Unfortunately, the values of $\Gamma$ had to be approximated via simulation because, to the authors' knowledge, there does not exist a data-generation algorithm yet that allows for full control of this matrix.

Although Study 2 was merely intended as an example, it does pose the question of which other types of studies exist where the recommendations and conclusions partly depend on value of the polynomial coefficients chosen to generate the data. An easy recommendation that would help clarify this issue (and which was originally advocated by Steiger (2014)) is to simply request that researchers report which polynomial coefficients were used in their simulation studies. Non-normality is, by all means, an elusive concept because there are many ways in which a distribution can be non-normal. Consider the 4$^{th}$-order moments present in the calculation of the $\Gamma$ matrix, $\sigma_{ijkl} = E[(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l)]$. The Vale-Maurelli (1983) procedure only allows the researcher control the case where $i = j = k = l$ per variable (i.e. elements found in the diagonal of $\Gamma$). All the other higher-order moments are free to vary and there is no way to know which type of non-normality will arise under which conditions.

### 4.5 Conclusions and recommendations

There are several limitations in this article that will hopefully span future research on the properties of the 3$^{rd}$-order polynomial method and other data-generation algorithms. Aside from the fact that only a limited set of studies was used in these simulations, it is important to point out that cases where the skewness and kurtosis of the univariate distributions have different directions (e.g., a positively-skewed variable correlated with a negatively-skewed variable) were not studied. These are likely to change the Fréchet–Hoeffding bounds and will perhaps result in a much narrower range of values for $r_{Y_1Y_2}$. Another condition that warrants further study (particularly considering the results of Kraatz (2011)) is to explore the type of distributions that may arise when monotonic and non-monotonic solutions are used concurrently to generate data. All throughout these simulations, all variables were transformed uniformly using coefficients from either Solution 1 or Solution 2. It would be of interest to methodologists and psychometricians to see what happens when coefficients from Solution 1 and Solution 2 are used to generate data within the same simulation. One aspect of solution multiplicity that was not explored here (and, to the authors' knowledge has never been mentioned in the published literature) is the solution multiplicity issue of the intermediate correlation equation itself. Equation (4.9) is a $3^{rd}$-degree polynomial and, as such, it has 3 sets of solutions. Appendix F shows the mathematical derivation of when these solutions are real or imaginary. It would be of interest to know the effect that the intermediate correlation has for cases when all 3 solutions are real-valued numbers and all 3 solutions fall within the theoretical limits of the correlation coefficient.

There still is much to know and discover about these data-generation algorithms and the fact that they are being used in the published literature without a thorough understanding of their properties raises the question of whether or not Monte Carlo simulations are being operated from a "black box" perspective among quantitative social scientists, where little is known about what happens inside the "black box" and even less is known about how the "black box" influences the results and recommendations suggested in the literature. By pointing out the fact that even the most popular data-generation method still needs further exploration, the authors of this article hope other methodologists and statisticians will bring their efforts forward in order to help inform and design better simulation studies that will result in more robust and reliable recommendations to applied data analysts.

## Chapter 5: Conclusion

At the beginning of this dissertation, the claim was made that, within the social sciences, the process of Monte Carlo simulation can sometimes operate as if it were a “black box”, where the user has limited understanding of the actual process of the simulation itself and is mostly concerned with the hypotheses and statistical methods that she or he wishes to test. Throughout this document, three instances (among many that exist) were presented where the disconnection between the theory and practice of simulation either masked the proper Monte Carlo design or influenced potential recommendations related to data practice. In each instance, it was demonstrated that the ambiguity in the choice of population model and the unexpected properties of data-generating algorithms had a direct impact on the type and quality of synthetic data being used, on the final conclusions from the study and, potentially, on the decisions that applied researchers may make when evaluating a specific statistical method over another. Much like in real-life experiments, computer simulations are susceptible to design inconsistencies that may end up pushing the results of Monte Carlo studies in unexpected, unwanted directions.

With this issue in mind, the overall goal of this dissertation was successfully addressed both by highlighting how a “black box” understanding of simulation studies opens the possibility for researchers to accept unwarranted assumptions and by emphasizing the importance of becoming acquainted with the theory that underlies Monte Carlo simulations. The following sections summarize some of the novel findings contained within this dissertation and the implications that they have both on simulation studies and on overall statistical theory in general. They also point towards important limitations not only in the chapters presented here but also in the broader conceptualization of simulation studies and potential avenues of future research that could help inform practice.

### 5.1 Novel contributions

The novel contributions of this dissertation will be divided in two broad categories. The first category is a focuses on computer algorithms and technical details for Monte Carlo simulations. The second category deals with the advances in statistical theory and analytic results.

For the first set of novel contributions, it is important to note that comparing and evaluating computer algorithms and studying their impact on simulation study results are both relevant and relatively unexplored avenues of research. It is somewhat surprising to see how little is known about the data-generating mechanisms used regularly by social scientists, in spite of their popularity of use. With the exception of the work of Kraatz (2011) and Luo (2011), an extensive search both in Google Scholar and the ISI Web of Knowledge yielded no published articles tackling this issue in the social science literature. Even the work of these two authors in the area of data-generating methods is not the main focus of their research. It is merely a subsection within their doctoral dissertations dealing only with low-dimensional distributions. The understanding of the empirical properties of the Fleishman (1978) and Vale-Maurelli (1983) methods, such as their tendency to generate downward-biased estimates of skewness and kurtosis or their occasional propensity for outliers in the case of the $4^{th}$ order moment is, therefore, a significant contribution for applied researchers who may consider using these algorithms for their own studies. Investigating the Headrick (2002) method also provided novel insights into the role that the $5^{th}$ and $6^{th}$-order moments of the data have in obtaining more stable, consistent values of skewness and kurtosis, particularly at smaller samples. This information could potentially be of relevance to researchers who are looking to investigate the influence that non-normality and small sample sizes have concurrently on statistical methods.

To the author’s knowledge, chapter 3 of this dissertation is the first psychometric study comparing the impact that different types of algorithms can have when conducting a simulation focusing on the properties of a statistical method, such as the chi-square test of fit under non-normality in Structural Equation Modelling (SEM). This is one of the driving principles of this dissertation and poses the crucial question of what kind of distributions have the majority of social scientists been investigating for more than three decades, since the Fleishman (1978) article was first published. It has only been recently discovered, thanks to the work of Foldnes and Grønneberg (2015), that the $3^{rd}$-order polynomial transformation has more in common with the multivariate normal distribution than with the myriad of non-normal distributions that exist, so re-visiting published simulation work with other types of algorithms provides both new and necessary insights into the workings of the statistical methods used by researchers.

Of all of the novel findings related to the use of these data-generating methods, one of the most significant and, to a certain extent, worrisome is the multiplicity of solutions of the Fleishman (1978) method. Although the work of Kraatz (2011) foreshadows some of its impact, chapter 4 in this dissertation is the first account, to the authors' knowledge, of a systematic study aimed at investigating the empirical and analytical properties of this algorithm, focusing both on the univariate and multivariate characteristics of the data it can generate, and the potential impact it can have on simulation studies. Because of its popularity in the social sciences, any findings related to the $3^{rd}$-order polynomial method have the potential to alter a large number of the recommendations present in the published articles that have used it. Of particular interest (and a new finding in itself) is the issue that the multivariate structures that each type of solution can generate are not the same. This brings into question the ways in which social science researchers describe multivariate non-normality and the need to become better acquainted with this concept. As a curious final note on chapter 4, it is also important to point out that, to the authors' knowledge, no article has alluded to the fact that the intermediate correlation equation is also prone to multiple solutions and, up until this dissertation, there had been no previous investigation into its properties (shown in Appendix F).

Regarding the use of the Conover and Iman (1981) algorithm and the simulation of the rank correlation, one needs to pay attention to the fact that, within Monte Carlo studies in the social sciences, no previous study has explicitly employed a method where the generated data were rank-correlated. Although, in some cases, the results could generalize across correlation types (see Caruso & Cliff, 1997), the majority of published articles only used data where the Pearson correlation served as a population parameter. Acknowledging this mismatch between simulation design and research hypothesis is one of the most important contributions of this dissertation because it goes at the core of what any computer study attempts to do. Ensuring the concordance between the statistical model in the population and the algorithm used to generate the data from this model should be a top priority for any researcher and, if this concordance is not present, it is imperative to be cognizant of this fact and address the limitations that this may entail. For this chapter, the rank correlation presented itself as an amenable case study that, in its simplicity, helped remind quantitative social scientists that, even in straightforward cases, there can still be a disconnection between the population parameter implied by the researcher and what he or she actually simulates.

Regarding the advancement of formal statistical theory, the second main category of novel contributions of this work, chapter 2 presents findings that translate knowledge from the statistical and mathematical literature to a wider audience of methodologists and social scientists. At the crux of this paper is the population definition of the Spearman rank correlation in terms of copula distributions as shown in Schmid and Schmidt (2007), Nelsen (2010), and Joe (2014). The literature review for this chapter, however, showed that none of these authors (or subsequent ones that use this definition) shows that the sample rank correlation converges to Equation (2.5) as the sample size grows arbitrarily large. Appendix B made this connection explicit by showing that an algebraic re-arrangement of the sample Spearman $\hat{\rho}$ does converge in the limit to its copula population definition. Once this equality was obtained, Appendix C helped showcase the generality of the copula framework by re-deriving Pearson

(1914)'s identity of the graded correlation under the assumption of a bivariate Gaussian copula which, although not a new finding in itself, does help popularize this fact among social scientists that may be unaware of it.

Although chapter 3 is mostly focused on the actual computational results of comparing two data-generating algorithms, chapter 4 does provide some novel, theoretical insights into the nature of the 3rd-order polynomial transformation that had not been published before. First, although Kraatz (2011) mentions that there are four real solutions to the Fleishman system of polynomial equations; she presents no proof of this fact. Using Bézout's theorem, it was demonstrated that there are, at most, 24 potential solutions, four of which are real and the rest are either complex or with different multiplicities. In the process of deriving this, it was also made explicit why the symmetry in coefficients $\{b, d, -b, -d\}$ makes them members of different solution sets by re-formulating their expressions in terms of quadratic equations. To the best of the author's knowledge, up to this dissertation, no published article has acknowledged the multiplicity of solutions to the intermediate correlation equation or described the nature of its solutions, particularly the conditions for all three solutions to be real-valued. This series of theoretical insights help highlight the fact that not much is known about the properties of the 3rd-order power transformation and that, in spite of its popularity, there are many characteristics of this method that are still very much unknown.

### 5.2 Limitations and future directions

There exist several limitations within the studies presented in this dissertation that open the door for future developments and avenues of research. Since chapter 2's main purpose was to exemplify the mismatch between population models and simulation methods, its simulation study is small and limited in terms of the practical conclusions that can be derived from it. Having a full Monte Carlo study where the unidimensional marginals are modified systematically would provide novel insights into the properties of the Iman-Conover algorithm as well as the copula-based definition of Spearman's $\rho$. Another limitation in this study was that only one data-generating mechanism was used[2]. There are other computer algorithms capable of generating rank-correlated data, such as the one presented in Koran, Headrick, and Kuo (2015). This is an important research direction considering the theme of this dissertation because the literature review for this chapter yielded no empirical evaluations of the Conover and Iman (1981) method, in spite of its popularity of use across the natural and engineering sciences. The analytic and simulation work of this chapter is solely focused on the bivariate case, only hinting to multivariate extensions (i.e., the Spearman rank correlation matrix) towards its conclusion. Expanding the results from lower to higher-dimensional distributions could also be a potentially new direction that was not explored here.

Chapters 3 and 4 were mostly concerned with exploring polynomial-based transformation methods due to their popularity in the social sciences. Other methods to obtain correlated, non-normal data exist within the published literature; however, but none of them were explored in this dissertation. Gaining further understanding of these methods would help researchers determine which algorithms work better for which types of simulation designs. For instance, the study used to explore the differential effect of the data-generation process on published recommendations comes from the field of SEM. It would be of interest to see whether or not different ways of generating data also have an effect in different types of studies. In this specific case, it is a well-known fact that the kurtosis of the sample has a detrimental effect on the chi-square test statistic of model fit (Yuan, Bentler & Zhang, 2005). One of the findings of this dissertation was that the $3^{rd}$-order polynomial transform generates data that are prone to downward-biased kurtosis values, with the occasional exceedingly large one. Nevertheless, this method does generate relatively well-behaved skewness values at larger sample sizes. A potential new direction could be to document the effect of using other types of data-generating

[2] I would like to thank an anonymous reviewer of the published version of this chapter for this suggestion.

mechanisms on test statistics that may be influenced by the 3rd moment of the data, particularly location-based tests such as the *t*-test (Zimmerman & Zumbo, 1992). The use of the Headrick (2002) method also opens up questions that were unexplored in this dissertation regarding the role that the the 5th and 6th order moments play in describing data, sometimes referred to as “super” or “hyper” skewness and “super” or “hyper” kurtosis respectively (Yee, 1990). Any insight into how these moments influence datasets collected by social scientists would be a novel contribution given that there rarely is any mention of them.

Although several novel insights into the mathematical properties of the Fleishman system of equations were found in this dissertation, several questions still remain unresolved. To begin with, even though Fleishman (1978), Headrick (2004), Kraatz (2011), and Luo (2011) agree that the 3rd-order polynomial transform does not span the full $\gamma_2 \geq \gamma_1^2 - 2$ parabola from which the researcher can select skewness and kurtosis values, they all disagree on the bounds that these values can take. With the exception of Headrick (2004), the other three authors did not elaborate on the methodology they employed to find their respective upper and lower bounds so more research is needed in this area to see where the Fleishman (1978) method shows a good performance and where it becomes unstable. Chapter 4 also describes the distribution of solutions to the system of equations within a grid search of several different starting points, noting that certain solutions appear more frequently for certain intervals but offers no insights as for why this happens. This suggests that the solution space of the system is multimodal in certain directions so a more thorough description of it could help inform which starting values should be preferred if the researcher is looking to work with exclusively monotonic or non-monotonic solutions.

The multiplicity of solutions to the system of polynomial equations is not only an issue of the Fleishman (1978) method. The intermediate correlation equation derived by Vale and Maurelli (1983) also has multiple roots (three, to be more precise). This dissertation presents a proof of this fact and describes the solutions analytically, but it does not elaborate on the potential impact that different values of the intermediate correlation may have in simulation studies. Just as it was shown that monotonic and non-monotonic solutions to the 3rd order polynomial may have a differential effect on the data-generation process, it could be the case that different types of intermediate correlations can potentially describe different types of covariance structures. This issue also extends to the Headrick (2002) method itself, where there are even more solutions both for its intermediate correlation equation and for the polynomial coefficients it uses.

### 5.3 Concluding remarks

Although the primary goal of this dissertation was to describe some of the assumptions that social science researchers make when conducting simulation studies and documenting the impact that they may have, there is also a different, much more subtle dimension to this work that aims to open a dialogue regarding the epistemology of simulation studies and the type of knowledge they can generate. The types of conclusions that can be obtained from Monte Carlo investigations are defined from the early planning stages of the simulation and are continuously shaped by the researcher as the design of the study develops. What kind of conditions should be investigated? What types of models should be generated? How should results be reported? All these, and many more, are important questions that influence, in the short term, the kind of recommendations that can be offered to applied researchers and, in the long term, the overall thinking and practice of data analysis.

Take, for instance, the role that skewness and kurtosis have on characterizing the data. Since they are based on the normal distribution, these indices try to provide a descriptive measure of what data looks like if it were to move progressively from a classic “bell-shaped” curve to something that is not “bell-shaped”. But even the idea not being “bell-shaped” posits the question of the sheer number of ways in which datasets can be distributed and, consequently, the usefulness of skewness and kurtosis in describing data. A probability distribution with a mean of 0, variance of 1, skewness of 0, and a kurtosis of 3 is not immediately a normal distribution. To be more precise, Romano and Siegel (1986) have proven that, for any finite collection of moments of the normal distribution, it is possible to construct a discrete, non-normal distribution that matches them. Only when having the full set of infinite moments will *some* distributions be fully-defined, although this statement is not even true in general since there exist distributions (e.g., log-normal) that are not defined by their moments (Durrett, 2010). Other areas of research have taken note of this fact and are changing their conceptualization of how moments are to be interpreted with respect to the data (e.g., Westfall, 2014) or even considering the case of using more moments to provide a more accurate description of it (e.g., Yee, 1990).

Would questioning each and every type of simulation study bring forth a better understanding of statistical methods and more reliable simulation conclusions? Perhaps, but it would not be feasible given the time constraints to which most researchers are subject. What can be done, however, is to think critically about simulation studies when learning from them and to be fully aware of the methodology used in a Monte Carlo study when designing them. Being able to understand a simulation from a “glass box” perspective is crucial in order to evaluate its merits and, although it certainly complicates the process in most cases, it also, at the very least, sets the stage to have a clear, open understanding of what types of conclusions can be generalized from the results. A good practical benchmark to use regarding how far the questioning of a simulation design should go could potentially be the point at which switching designs or methodologies has no bearing on the conclusions obtained. Chapter 3 of this dissertation shows that, once the sample size of 1000 is reached, the type of data generated both by the Headrick (2002) and Vale-Maurelli (1983) methods and the practical recommendations they imply are somewhat similar. Therefore, for large-sample type simulations, using either method would probably make no difference and its choice would depend solely on the researcher’s preference.

Statistical theory, the design of the simulation, and its practical computer implementation are the three elements of a Monte Carlo study that researchers need to align in order to obtain reliable conclusions aimed at informing data practice. Whenever any of these three elements is not in concordance with the other two, the implicit assumption of their overlap is introduced in order to preserve the balance of the design and to obtain valid conclusions. Depending on the particular situation, sometimes this perfect overlap is not possible (in which case it needs to be acknowledged) and sometimes, even though it is possible, it can be missed by the researcher. The significance of this issue is crucial and cannot be emphasized enough in order to evaluate computer experiments with the same strict standards that we have for every day, real-life experiments. Those of us involved in the advancement of methodology and statistics need to set the example in terms of good design and data practice if we are to successfully navigate the changing zeitgeist taking place within the social sciences. The Crisis of Replicability and overall critiquing of the prevalent methods in science have brought into question many of the old research paradigms within psychology and its influence is beginning to reverberate across many other fields. It is my belief that this change in zeitgeist will eventually permeate the field of quantitative methods and we, as people of numbers, need to be ready to accept it and move on into whichever new endeavours the future holds for us.

## References

Astivia, O. L. O., & Zumbo, B. D. (2015). A Cautionary Note on the Use of the Vale and Maurelli Method to Generate Multivariate, Nonnormal Data for Simulation Purposes. *Educational and Psychological Measurement, 75*, 541-567. doi: 10.1177/0013164414548894

Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew–normal distribution. *Biometrika, 84*, 715- 726. doi: 10.1007/BF0260299

Bandalos, D. L., & Leite, W. (2013). The role of simulation in structural equation modeling. In G.R. Hancock and R. Mueller (Eds.), *A Second Course in Structural Equation Modeling*, *2nd Edition*. Greenwich, CT: Information Age Publishing.

Bentler, P. M. (1995). *EQS Structural Equations Program Manual*. Encino, CA: Multivariate Software.

Beisbart, C., & Norton, J. D. (2012). Why Monte Carlo simulations are inferences and not experiments. *International Studies in the Philosophy of Science, 26*, 403-422. doi: 10.1080/02698595.2012.748497

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods, 17*, 399–417. doi:10.1037/a0028087

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York, NY: John Wiley.

Borkowf, C. B. (2002). Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman's rank correlation. *Computational Statistics and Data Analysis, 39*, 271–286. doi:10.1016/S0167-9473(01)00081-0

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31*(2), 144-152. doi: 10.1111/j.2044-8317.1978.tb00581.x

Carsey, T. M., & Harden, J. J. (2013). *Monte Carlo Simulation and Resampling Methods for Social Science*. Thousand Oaks, CA: Sage Publications.

Caruso, J. C., & Cliff, N. (1997). Empirical size, coverage, and power of confidence intervals for Spearman's Rho. *Educational and Psychological Measurement*, *57*, 637-654. doi: 10.1177/0013164497057004009

Chen, X. & Tung, Y.K. (2003). Investigation of polynomial normal transform. *Structural Safety, 25,* 423-445. doi: 10.1016/S0167-4730(03)00019-5

Christine, P. D., & John, R. (2004). *Statistics Without Maths for Psychology Using SPSS for Windows*. Harlow, England: Pearson Prentice Hall.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician, 35*, 124-129. doi: 10.1080/00031305.1981.10479327

Crow, E. L., & Shimizu, K. (Eds.). (1988). *Lognormal distributions: Theory and applications* (Vol. 88). New York: M. Dekker.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16-29. doi: 10.1037/1082-989X.1.1.16

Dalen, J. (1987). Algebraic bounds on standardized sample moments. *Statistics & Probability Letters, 5,* 329–331. doi: 10.1067/715290005-8

David, F. N., & Mallows, C. L. (1961). The variance of Spearman's rho in normal samples. *Biometrika, 48,* 19-28. doi: 10.2307/2333126

Demirtas, H., & Hedeker, D. (2011). A practical way for computing approximate lower and upper correlation bounds. *The American Statistician, 65*, 104-109. doi: 10.1198/tast.2011.10090

du Boulay, B., O'Shea, T., & Monk, J. (1981). The black box inside the glass box: presenting computing concepts to novices. *International Journal of Man-Machine Studies, 14*, 237-249. doi: 10.1016/S0020-7373(81)80056-9

Durrett, R. (2010). *Probability: theory and examples*. Cambridge, England: Cambridge university press.

Elishakoff, I. (2003). Notes on philosophy of the Monte Carlo method. *International Applied Mechanics, 39*, 753-762. doi: 10.1023/A:1026236621486

Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods, 6*, 352-370. doi: 10.1037/1082-989X.6.4.352

Fan, X., Sivo, S., & Keenan, S. (2002). SAS for Monte Carlo studies: A guide for quantitative researchers. Cary, NC: SAS Institute.

Fieller, E. C., Hartley, H. O., & Pearson, E. S. (1957). Tests for rank correlation coefficients. *Biometrika, 44*, 470-481. doi: 10.2307/2332878

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal, 4*, 87-107. doi: 10.1080/10705519709540063

Fishman, G. (2013). *Monte Carlo: Concepts, Algorithms, and Applications*. New York, NY: Springer Science & Business Media.

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

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466-491. doi: 10.1037/1082-989X.9.4.466

Foldnes, N., & Grønneberg, S. (2015). How General is the Vale–Maurelli Simulation Approach? *Psychometrika*, *80*(4), 1066-1083. doi: 10.1007/s11336-014-9414-0

Foldnes, N., & Olsson, U. H. (2016). A Simple Simulation Technique for Nonnormal Data with Prespecified Skewness, Kurtosis, and Covariance Matrix. *Multivariate Behavioral Research*, *51*, 1-13. doi: 10.1080/00273171.2015.1133274

Headrick, T. C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate non-normal distributions. *Computational Statistics and Data Analysis, 40*, 685-711. doi: 10.1016/S0167-9473(02)00072-5

Headrick, T. C. (2004). Transformations for simulating multivariate distributions. *Journal of Modern Applied Statistical Methods, 3*(1), 65–71. Retrieved from http://digitalcommons.wayne.edu/jmasm/vol3/iss1/8 (November 19th, 2016)

Headrick, T. C. (2010). *Statistical Simulation: Power Method Polynomials and Other Transformations*. Boca Raton, Florida: Chapman & Hall/CRC.

Headrick, T. C., Aman, S. Y., & Beasley, T. M. (2008). Simulating Controlled Variate and Rank Correlations Based on the Power Method Transformation. *Communications in Statistics—Simulation and Computation, 37*, 602-616. doi: 10.1080/03610910701812394

Headrick, T. C. & Kowalchuk, R. K. (2007). The power method transformation: its probability density function, distribution function, and its further use for fitting data. *Journal of Statistical Computation and Simulation, 77*, 229–249. doi: 10.1080/10629360600605065

Headrick, T. C., & Mugdadi, A. (2006). On simulating multivariate non-normal distributions from the generalized lambda distribution. *Computational Statistics & Data Analysis, 50*, 3343-3353. doi: 10.1016/j.csda.2005.06.010

Headrick, T. C., & Pant, M. D. (2012). Simulating non-normal distributions with specified L-moments and L-correlations. *Statistica Neerlandica, 66*, 422-441. doi: 10.1111/j.1467-9574.2012.00523.x

Headrick, T. C., Sheng, Y., & Hodis, F. A. (2007). Numerical computing and graphics for the power method transformation using Mathematica. *Journal of Statistical Software*, *19,* 1-17. Retrieved from http://www.jstatsoft.org/v19/i03/paper(November 19th, 2016)

Headrick, T. C. & Sawilowsky, S.S. (1999). Simulating correlated multivariate non-normal distributions. *Psychometrika, 64*, 25-35. doi:10.1007/BF02294317

Hinton, P. R., McMurray, I., & Brownlow, C. (2014). *SPSS explained*. New York, NY: Routledge.

Hoeffding, W. (1948a). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics, 19*, 293–325. doi: 10.1007/978-1-4612-0919-5_20

Hoeffding, W. (1948b). A non-parametric test of independence. *The Annals of Mathematical Statistics, 19*, 546–557. doi: 10.1214/aoms/1177730150

Hoover, W.G. & Hoover, C.G. (2015). *Simulation and control of chaotic nonequilibrium systems*. New Jersey, NJ: World Scientific.

Hotelling, H., & Pabst, M. R. (1936). Rank correlation and tests of significance involving no assumption of normality. *The Annals of Mathematical Statistics, 7(1)*, 29-43. doi: 10.1214/aoms/1177732543

Howitt, D., & Cramer, D. (2005). *An introduction to statistics in psychology*. New York, NY:Pearson education.

Horn, J. L. (1965). A Rationale and Test For the Number of Factors in Factor Analysis. *Psychometrika, 30,* 179-185. doi: 10.1007/BF02289447

Joe, H. (2014). *Dependence modeling with copulas*. Boca Raton, FL: CRC Press.

Johansen, M. K., Savage, J., Fouquet, N., & Shanks, D. R. (2015). Salience not status: how category labels influence feature inference. *Cognitive Science, 39*, 1594-1621. doi: 10.1111/cogs.12206

Johnson, P.E. (2013). Monte Carlo Analysis in Academic Research. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods, Volume 1: Foundations*. Oxford, England: Oxford University Press.

Jones, O., Maillardet, R. & Robinson, A. (2009). *Introduction to scientific programming and simulation using R*. Boca Raton, FL: CRC Press.

Karlsson, C., Nellore, R., & Soderquist, K. (1998). Black box engineering: redefining the role of product specifications. *Journal of Product Innovation Management, 15*, 534-549. doi: 10.1111/1540-5885.1560534

Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods, 21*, 69-92. doi: 10.1037/a0040086

Koran, J., Headrick, T. C., & Kuo, T. C. (2015). Simulating univariate and multivariate nonnormal distributions through the method of percentiles. *Multivariate Behavioral Research, 50*, 216-232. doi: 10.1080/00273171.2014.963194

Kowalchuk, R. K. & Headrick, T. C. (2010). Simulating multivariate g-and-h distributions. *British Journal of Mathematical and Statistical Psychology, 63*, 63–74. doi: 10.1348/000711009X423067

Kraatz, M. (2011). *Investigating the Performance of Procedures for Correlations Under Nonnormality* (Unpublished doctoral dissertation). Vanderbilt University, Nashville, TN.

Kruschke, J. K. (2010). *Doing Bayesian data analysis: A tutorial with R and BUGS*. New York, NY: Elsevier.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association, 53*, 814-861. doi:10.2307/2281954

Lang, S. (2014). *Introduction to Algebraic Geometry*. Eastford, CT: Martino Fine Books.

Li, T. S. & Hammond, J. L. (1975). Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Trans. On Systems, Man, and Cybernetics, 5*, 557–561. doi: 10.1109/TSMC.1975.5408380

Lix, L.M. & Fouladi, R.T. (2007). Robust step-down tests for multivariate independent group designs. *British Journal of Mathematical & Statistical Psychology*, *60*, 245–265. doi: 10.1348/000711006X117853

Lix, L.M., Keselman, H.J. & Hinds, A.M. (2005) Robust tests for the multivariate Behrens-Fisher problem. *Computer Methods and Programs in Biomedicine, 77*, 129–139. doi: 10.1016/j.cmpb.2004.09.002

Luo, H. (2011). *Some Aspects on Confirmatory Factor Analysis of Ordinal Variables and Generating Non-normal Data*. (Unpublished doctoral dissertation). Uppsala University, Uppsala, Sweden.

Mair, P., Satorra, A., & Bentler, P. M. (2012). Generating nonnormal multivariate data using copulas: Applications to SEM. *Multivariate Behavioral Research*, *47*, 547-565. doi: 10.1080/00273171.2012.692629

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519-530.

Martínez-López, B., Ivorra, B., Ngom, D., Ramos, A. M., & Sánchez-Vizcaíno, J. M. (2012). A novel spatial and stochastic model to evaluate the within and between farm transmission of classical swine fever virus: II Validation of the model. *Veterinary Microbiology, 155*(1), 21-32. doi: 10.1016/j.vetmic.2011.08.008

Mattson, S. (1997). How to generate non-normal data for simulation of structural equation models. *Multivariate Behavioral Research*, *32*, 355-373. doi: 10.1207/s15327906mbr3204_3

Mildenhall, S. J. (2006). *Correlation and aggregate loss distributions with an emphasis on the Iman–Conover method*. Casualty Actuarial Society Forum (Winter), 103–204. Retrieved from http://www.casact.org/pubs/forum/06wforum/06w107.pdf (November 19th, 2016)

Mollin, R.A. (1947). Algebraic Number Theory. Boca Raton, FL: CRC press.

Monagan, M. B., Geddes, K. O., Heal, K. M., Labahn, G., Vorkoetter, S. M., Devitt, J. S. & Rickard, K. M. (2012). *Maple V Programming Guide: for Release 5*. Chicago, IL: Springer Science & Business Media

Moran, P. A. P. (1948). Rank correlation and product-moment correlation. *Biometrika*, *35*, 203-206. doi: 10.2307/2332641

Muthén, L.K. & Muthén, B.O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal, 4*, 599-620. doi: 10.1207/S15328007SEM0904_8

Nagahara, Y. (2004). A method of simulating multivariate nonnormal distributions by the Pearson distribution system and estimation. *Computational Statistics & Data Analysis, 47*(1), 1-29. doi: 10.1016/j.csda.2003.10.008

Nelsen, R. B. (2010). *An introduction to copulas*. New York, NY: Springer Science & Business Media.

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 8*, 353-377. doi: 10.1207/S15328007SEM0803_2

Pant, M. D., & Headrick, T. C. (2015). A Characterization of the Burr Type III and Type XII Distributions through the Method of Percentiles and the Spearman Correlation. *Communications in Statistics-Simulation and Computation, (just-accepted)*, 00-00. doi: 10.1080/03610918.2015.1048878

Pearson, K. (1914). On an extension of the method of correlation by grades or ranks. *Biometrika*, *10*, 416-418. doi: 10.2307/2331792

Pouillot, R. and Delignette-Muller, M.L. (2010), Evaluating variability and uncertainty in microbial quantitative risk assessment using two R packages. *International Journal of Food Microbiology, 142*.330-40. doi: 10.1016/j.ijfoodmicro.2010.07.011

R Core Team (2015). R: A language and environment for statistical computing. *R Foundation for Statistical Computing.* Vienna, Austria. URL http://www.R-project.org/ (November 19th, 2016)

Revelle, W. (2016) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.6.9 (November 19th, 2016)

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354-373. doi: doi.org/10.1037/a0029315

Romano, J. P., & Siegel, A. F. (1986). *Counterexamples in probability and statistics*. Boca Raton, FL: CRC Press.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36. doi: 10.18637/jss.v048.i02

Rubinstein, R. Y. & Kroese, D. P. (2007). *Simulation and the Monte Carlo Method* (2nd ed.). New York: John Wiley & Sons.

Rupinski, M. T., & Dunlap, W. P. (1996). Approximating Pearson product-moment correlations from Kendall's tau and Spearman's rho. *Educational and Psychological Measurement*, *56*, 419-429. doi: 10.1177/0013164496056003004

Ruscio, J. & Kaczetow, W. (2008). Simulating multivariate nonnormal data using an iterative algorithm. *Multivariate Behavioral Research, 43*, 335–381. doi: 10.1080/00273170802285693

Sapp, M. (2006). *Basic psychological measurement, research designs, and statistics without math*. Springfiled, IL:Charles C Thomas Publisher.

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science, 4*, 409-423. doi: 10.1214/ss/1177012413

Sawilowsky, S. S. & Fahoome, G.C. (2003). *Statistics via Monte Carlo Simulation with Fortran*. Rochester Hills, MI: JMASM.

Schmid, F., & Schmidt, R. (2007). Multivariate extensions of Spearman's rho and related statistics. *Statistics & Probability Letters*, *77*, 407-416. doi:10.1016/j.spl.2006.08.007

Schweizer, B., & Wolff, E. F.. (1981). On Nonparametric Measures of Dependence for Random Variables. *The Annals of Statistics*, *9*, 879–885. doi: 10.1214/aos/1176345528

Seco, G. V., Gras, J. A., & García, M. A. (2007). Comparative robustness of recent methods for analyzing multivariate repeated measures designs. *Educational and Psychological Measurement, 67*, 410-432. doi: 10.1177/0013164406294777

Skidmore, S. T., & Thompson, B. (2011). Choosing the Best Correction Formula for the Pearson $r^2$ Effect Size. *The Journal of Experimental Education, 79*, 257-278. doi: 10.1080/00220973.2010.48443

Sklar, A. (1959). Fonctions de Répartition à n Dimensions et Leurs Marges. *Publications de l'Institut de Statistique de L'Université de Paris, 8*, 229–231.

Spearman, C. (1906). 'Footrule' for measuring correlation. *British Journal of Psychology, 1904-1920, 2*(1), 89-108. doi: 10.1111/j.2044-8295.1906.tb00174.x

Steiger, J. H. (2014, October). *Still Crazy After All These Years: Complexity, Principles, and Practice in Multivariate Statistics*. Paper presented at the Society of Multivariate Experimental Psychology, Nashville, TN.

Szolnoki, A., & Perc, M. (2015). Conformity enhances network reciprocity in evolutionary social dilemmas. *Journal of The Royal Society Interface, 12*, 1214-1229. doi: 10.1098/rsif.2014.1299

Tadikamalla, P. R. (1980). On simulating non-normal distributions. *Psychometrika*, *45*, 273-279. doi:10.1007/BF02294081

Tong, X., & Bentler, P. M. (2013). Evaluation of a new mean scaled and moment adjusted test statistic for SEM. *Structural Equation Modeling: A Multidisciplinary Journal, 20*, 148-156. doi: 10.1080/10705511.2013.742403

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48*, 465-471. doi: 10.1007/BF02293687

Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Springer, New York.

Weathers, D., Sharma, S. & Niedrich, R.W. (2005). The impact of the number of scale points, dispositional factors, and the status quo decision heuristic on scale reliability and response accuracy. *Journal of Business Research, 58*, 1516–1524. doi: S0148296304001596

Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. RIP. *The American Statistician, 68*, 191-195. doi: 10.1080/00031305.2014.917055

Wilcox, R. R., & Tian, T. (2008). Comparing dependent correlations. *The Journal of General Psychology*, *135*, 105-112. doi: 10.3200/GENP.135.1.105-112

Yee, E. (1990). The shape of the probability density function of short-term concentration fluctuations of plumes in the atmospheric boundary layer. *Boundary-Layer Meteorology, 51*, 269-298. doi:10.1007/BF00122141

Yuan, K. H., Bentler, P. M., & Zhang, W. (2005). The effect of skewness and kurtosis on mean and covariance structure analysis the univariate case and its multivariate implication. *Sociological Methods & Research, 34*, 240-258. doi: 10.1177/0049124105280200

Zimmerman, D. W., & Zumbo, B. D. (1992). Parametric alternatives to the Student t test under violation of normality and homogeneity of variance. *Perceptual and Motor Skills, 74*, 835-844. doi: 10.2466/pms.1992.74.3.835

Zimmerman, D. W., Zumbo, B. D., & Williams, R. H. (2003). Bias in Estimation and Hypothesis Testing of Correlation. *Psicologica : Revista de Metodologica Psicologica Experimental, 24* , 133-158. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=647711 (November 19th, 2016)

## Appendix A: Bounds of correlation coefficient for lognormal-distributed marginals.

Without loss of generality, assume:

$$(Z_1, Z_2)' \sim \boldsymbol{\mathcal{N}}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right).$$

Let $X_1 = e^{Z_1}$ and $X_2 = e^{Z_2}$ so that $(X_1, X_2)'$ follows a standard log-normal distribution.

From the definition of the log-normal distribution's mean and variance and the population values specified above, it is possible to determine:

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = e^{\mu+\frac{\sigma^2}{2}} = e^{0+\frac{1^2}{2}} = e^{1/2}.$$

$$Var(X_1) = Var(X_2) = \left(e^{\sigma^2} - 1\right)e^{2\mu+\sigma^2} = (e^1 - 1)e^{2(0)+1} = (e-1)e = e^2 - e.$$

For the readers interested in knowing how the mean and variance of the log-normal distribution are derived, consult Crow and Shimizu (1988) for further details.

To obtain the expected value of the cross-product $\mathbb{E}(X_1X_2)$ we will rely on the moment-generating function (MGF) of the multivariate normal distribution:

$$M_X(\boldsymbol{t}) = e^{\boldsymbol{t}'\boldsymbol{\mu}+(1/2)\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}.$$

Since the MGF is expressed in terms of exponentiated random variables, it is a natural framework to obtain the moments of the log-normal distribution. By setting ***t=1***:

$$\mathbb{E}(X_1X_2) = \mathbb{E}(e^{Z_1}e^{Z_2}) = M_{Z_1Z_2}(\mathbf{1}) = \exp\left([1 \quad 1]\begin{bmatrix}0\\0\end{bmatrix} + \frac{1}{2}[1 \quad 1]\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\right) =$$
$$\exp\left(0 + \frac{1}{2}(2\rho + 2)\right) = e^0 e^{\rho+1} = e^{\rho+1}.$$

Now all the elements are in place to define the correlation between $X_1$and$X_2$ in the population:

$$Cor(X_1, X_2) = \frac{\mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)}{\sqrt{Var(X_1)}\sqrt{Var(X_2)}} = \frac{e^{\rho+1} - e^{1/2}e^{1/2}}{\sqrt{e^2-e}\sqrt{e^2-e}} = \frac{ee^{\rho} - e}{e^2 - e} = \frac{e(e^{\rho}-1)}{e(e-1)} = \frac{e^{\rho}-1}{e-1}$$

For $\rho = +1$:

$$Cor(X_1, X_2) = \frac{e-1}{e-1} = +1$$

For $\rho = -1$:

$$Cor(X_1, X_2) = \frac{e^{-1}-1}{e-1} = \frac{\frac{1-e}{e}}{e-1} = \frac{1-e}{e(e-1)} = \frac{-1(e-1)}{e(e-1)} = \frac{-1}{e} \approx -0.3678794$$

Therefore as $n \to \infty$ the Pearson correlation between two standard log-normal random variables can never go below $-1/e$.

A quick simulation in R yields:

```
n <- 10000000
z1 <- rnorm(n, 0, 1)
z2 <- rnorm(n, 0, 1)

x1 <- exp(z1)
x2 <- exp(z2)

cor(sort(x1, decreasing=T), sort(x2, decreasing=F), method="pearson")
[1] -0.3678432
```

## Appendix B: Convergence of the sample $\rho_s$ to its population definition

Consider the formula for the sample product-moment correlation coefficient.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Since both $x_i$and $y_i$ are ranked variables with no ties, they consist of natural numbers from 1 to *n*.

Therefore, the denominator can be expressed as:

$$\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{x})^2} = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sum_{i=1}^{n}(x_i - \bar{x})^2,$$

because the order of the ranks does not influence their mean value or their sum of squared differences.

Expanding the binomial one obtains:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - 2x_i\bar{x} + \bar{x}^2 = \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

And by using the identities of the sum of *n* and $n^2$ consecutive natural numbers, the above expression can be formulated as:

$$\sum_{i=1}^{n} x_i^2 - n\bar{x} = \frac{n(n+1)(2n+1)}{6} - n\left[\frac{(n+1)}{2}\right]^2 = n(n+1)\left[\frac{(2n+1)}{6} - \frac{(n+1)}{4}\right] =$$
$$n(n+1)\left[\frac{8n+4-6n-6}{24}\right] = n(n+1)\left[\frac{2(4n+2-3n-3)}{24}\right] = n(n+1)\left[\frac{n-1}{12}\right] = \frac{n(n^2-1)}{12}$$

For the numerator:

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - x_i\bar{y} - y_i\bar{x} + \bar{x}\bar{y} = \sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \bar{x}\bar{y} =$$
$$\sum_{i=1}^{n} x_i y_i - n\bar{y}\bar{x} - n\bar{x}\bar{y} + n\bar{x}\bar{y} = \sum_{i=1}^{n} x_i y_i - n\bar{y}\bar{x} = \sum_{i=1}^{n} x_i y_i - n\left[\frac{n+1}{2}\right]^2$$

And by letting $x_i y_i = r_{ix} r_{iy}$to define them as ranks, it is possible to obtain the same expression for the Spearman rank correlation as found in equation (2.4):

$$\widehat{\rho_s} = \frac{\sum_{i=1}^{n} r_{ix} r_{iy} - n[(n+1)/2]^2}{n(n^2-1)/12}.$$

In order to obtain the population value of this expression, one needs to take the limit → ∞ . Split the previous expression as:

$$\widehat{\rho_s} = \frac{\sum_{i=1}^{n} r_{ix} r_{iy}}{n(n^2-1)/12} - \frac{n[(n+1)/2]^2}{n(n^2-1)/12}$$

And take each limit independently. For the section of the previous expression to the right of the negative sign, one can do:

$$\frac{n[(n+1)/2]^2}{n(n^2-1)/12} = \frac{\frac{(n+1)^2}{4}}{\frac{(n^2-1)}{12}} = \frac{12(n+1)^2}{4(n^2-1)} = 3\frac{(n+1)^2}{(n+1)(n-1)} = 3\frac{n+1}{n-1}$$

The limit $n \to \infty$ is a classical application of L'Hôpital's rule which results in:

$$3 \lim_{n\to\infty} \frac{n+1}{n-1} = (3)(1) = 3$$

For the left section, invoke the definition of an integral as the limit of a Riemann sum:

$$\int f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i) \cdot (x_{i+1} - x_i)$$

Therefore, for the cumulative density functions $F_X(x), F_Y(y)$:

$$\int F_X(x)F_Y(y)dF(x,y) = \lim_{n\to\infty} \sum_{i=1}^{n} F_X(x_i)F_Y(y_i) \cdot (F(x_{i+1}, y_{i+1}) - F(x_i, y_i))$$

Without loss of generality, remember that, by definition, $F_X(x_i) = \mathbb{P}(X \leq x_i)$. Since $x_1, x_2, x_3, \ldots, x_i, \ldots x_n$ are randomly selected samples (i.e. each $x_i$ has equal probability of being selected), it is possible to see that:

$$\mathbb{P}(x_1, x_2, x_3, \ldots, x_{i-1}, x_{i+1}, \ldots x_n \leq x_i) = \frac{\#\ elements\ in\ list\ less\ than\ x_i}{total\ \#\ of\ elements\ in\ the\ list} = \frac{rank\ position\ of\ x_i}{n}$$

By the same logic and because $F_1(x_i)$and $F_2(y_i)$ are ordered to be strictly increasing, $F(x_{i+1}, y_{i+1})$ is the smallest value bigger than $F(x_i, y_i)$. Therefore, as $n \to \infty$ it is possible to see:

$$\lim_{n\to\infty} F_X(x_i) = \frac{r_{ix}}{n} \qquad \lim_{n\to\infty} F_Y(y_i) = \frac{r_{iy}}{n} \qquad \lim_{n\to\infty} F(x_{i+1}, y_{i+1}) - F(x_i, y_i) = \frac{1}{n}$$

with $r_{ix}$ and $r_{iy}$ being the ranking position of $x_i$ , $y_i$ as defined above. Finally joining all the limiting quantities one can see:

$$\int F_X(x)F_Y(y)dF(x,y) = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{r_{ix}}{n}\frac{r_{iy}}{n}\frac{1}{n} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} r_{ix} r_{iy}}{n^3}$$

Since for simple polynomial expression, like in the denominator, only the largest exponent defines the rate of convergence on the limit, one can trivially re-arrange the denominator such that:

$$\int F_X(x)F_Y(y)dF(x,y) = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} r_{ix}r_{iy}}{n^3} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} r_{ix}r_{iy}}{n^3 - n} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} r_{ix}r_{iy}}{n(n^2 - 1)}$$

Therefore, all the elements are present to obtain the population version of the Spearman rank correlation expressed as:

$$\rho_S = 12 \int F_X(x)F_Y(y)dF(x,y) - 3$$

To re-express $\rho_S$ using copulas, it suffices to remind the reader that that both $F_1(x)$ and $F_1(y)$ are uniformly distributed over the domain [0,1] and that, by Sklar's theorem, one can express any multidimensional joint distribution function in terms of its univariate marginals and respective copula function. Therefore:

$$\rho_S = 12 \int F_X(x)F_Y(y)dF(x,y) - 3 = 12 \int u_x u_y dC(u_x, u_y) - 3$$

As shown in Joe (2014), notice that $\int u_x u_y dC(u_x, u_y) = \mathbb{P}(u_x \le U_X, u_y \le U_Y)$ with $U_X$, $U_Y$ and $(U_X, U_Y)$ independent of one another. $U_X$ and $U_Y$ follow standard uniform distributions and the joint distribution $(U_X, U_Y)$ can be expressed as the copula $C(u_x, u_y)$. Taking the complement of the copula function (i.e. the survival function), it is possible to switch the integrand as:

$$\rho_S = 12 \int u_x u_y dC(u_x, u_y) - 3 = 12 \int \bar{C}(u_x, u_y) du_x du_y - 3$$

And using the identity $\bar{C}(u_x, u_y) = 1 - u_x - u_y + C(u_x, u_y)$ found in Nelsen (2010), the pervious integral can be evaluated to result in:

$$\rho_S = 12 \int C(u_x, u_y) du_x du_y - 3$$

## Appendix C: Generalization of the Pearson identity for graded correlation

Consider the population version of the Spearman rank correlation expressed as:

$$\rho_S = 12\int C(u_x, u_y, \rho)du_x du_y = 12\int F_X(z_x)F_Y(z_y)dF(z_x, z_y) - 3$$

Where *F* is the cumulative density function for the bivariate normal random variable $(Z_x, Z_y)$ with standard normal marginal distributions $F_X, F_Y$ and correlation coefficient $\rho$.

Notice that:

$$12\int F_X(z_x)F_Y(z_y)dF(z_x, z_y) - 3 = 12\int F_X(z_x)F_Y(z_y)f(z_x, z_y)dz_x dz_y - 3 = 12\mathbb{E}[F_X(Z_x)F_Y(Z_y)] - 3$$

Where $f(z_x, z_y)$ is the probability density function of the bivariate normal distribution and $F_X(Z_x), F_Y(Z_y)$ are uniformly distributed.

Consider the bivariate normal random variable $(Z'_x, Z'_y)$ with standard normal marginal distributions $Z'_x$ and $Z'_y$ but with components independent of one another and of $(Z_x, Z_y)$ so that $Z'_x \perp\!\!\!\perp Z'_y \perp\!\!\!\perp (Z_x, Z_y)$.

Define the bivariate distribution $(Z_x - Z'_x, Z_y - Z'_y)$. This distribution is in itself bivariate normal with:

$$\mathbb{E}[Z_x - Z'_x] = \mathbb{E}[Z_y - Z'_y] = 0 \qquad Var[Z_x - Z'_x] = Var[Z_y - Z'_y] = 2 \qquad Cor[(Z_x - Z'_x), (Z_y - Z'_y)] = \frac{\rho}{2}$$

Then it is possible to see:

$$\mathbb{P}\{(Z_x - Z'_x)(Z_y - Z'_y) > 0\} = \mathbb{P}\{(F_X(Z_x) - F_X(Z'_x))(F_Y(Z_y) - F_Y(Z'_y)) > 0\}$$
$$\mathbb{P}\{(Z_x - Z'_x)(Z_y - Z'_y) > 0\} = \mathbb{E}[\mathbb{P}\{(F_X(Z_x) - F_X(Z'_x))(F_Y(Z_y) - F_Y(Z'_y)) > 0|(F_X(Z_x), F_Y(Z_Y))\}]$$
$$\mathbb{P}\{(Z_x - Z'_x)(Z_y - Z'_y) > 0\} = \mathbb{E}\left[F_X(Z_x)F_Y(Z_y) + (1 - F_X(Z'_x))\left(1 - F_Y(Z'_y)\right)\right]$$
$$\mathbb{P}\{(Z_x - Z'_x)(Z_y - Z'_y) > 0\} = 2\mathbb{E}[F_X(Z_x)F_Y(Z_y)]$$

By the law of total probability and symmetry we also have:

$$\mathbb{P}\{(Z_x - Z'_x)(Z_y - Z'_y) > 0\} = \mathbb{P}\{Z_x - Z'_x, Z_y - Z'_y > 0 + Z_x - Z'_x, Z_y - Z'_y < 0\}$$
$$\mathbb{P}\{(Z_x - Z'_x)(Z_y - Z'_y) > 0\} = 2\mathbb{P}\{Z_x - Z'_x, Z_y - Z'_y > 0\}$$

And so this equality holds:

$$\mathbb{E}[F_X(Z_x)F_Y(Z_y)] = \mathbb{P}\{Z_x - Z'_x, Z_y - Z'_y > 0\}$$

For the case of the bivariate standard normal distribution (i.e. $(Z_1, Z_2)$ follow a bivariate normal distribution with standard normal marginals), the normal quadrant probability theorem states:

$$\mathbb{P}(z_1 \leq 0, z_2 \leq 0) = \mathbb{P}(z_1 \geq 0, z_2 \geq 0) = \frac{1}{4} + \frac{\sin^{-1}\rho}{2\pi}$$

Therefore one can make the equivalence:

$$\mathbb{E}\left[F_X(Z_x)F_Y(Z_y)\right] = \mathbb{P}\{Z_x - Z_x', Z_y - Z_y' > 0\} = \frac{1}{4} + \frac{\sin^{-1}\left(\frac{\rho}{2}\right)}{2\pi}$$

Substituting this in the population definition of the Spearman correlation with a Gaussian copula then results in:

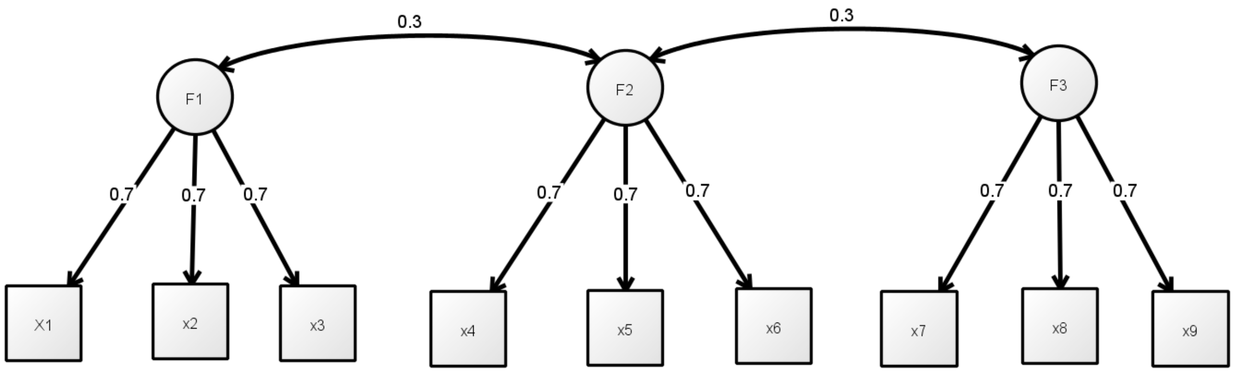
$$\rho_S = 12\int F_X(z_x)F_Y(z_y)dF(z_x, z_y) - 3 = 12\mathbb{E}\left[F_X(Z_x)F_Y(Z_y)\right] - 3 = 12\left[\frac{1}{4} + \frac{\sin^{-1}\left(\frac{\rho}{2}\right)}{2\pi}\right] - 3 = \frac{6}{\pi}\sin^{-1}\left(\frac{\rho}{2}\right)$$

which yields the desired identity:

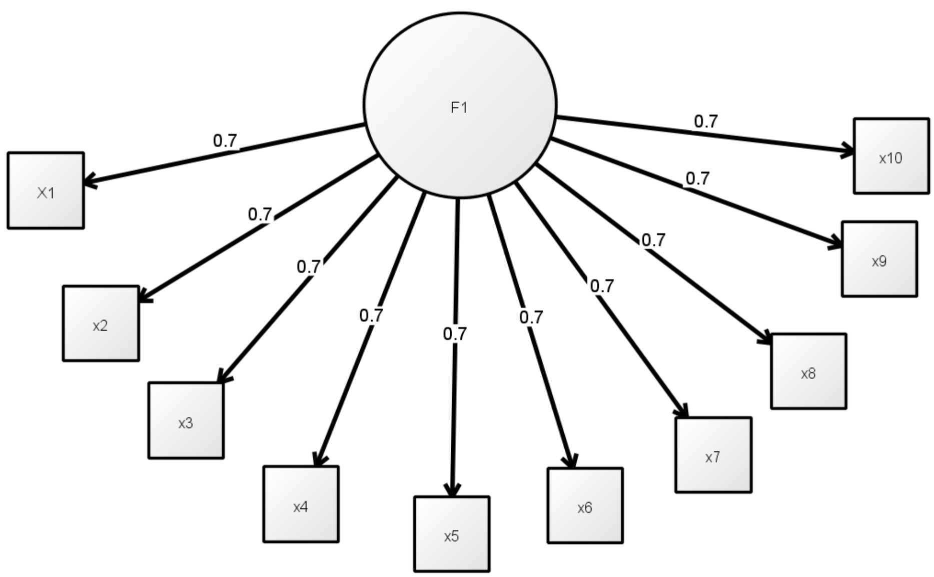
$$\rho_S = \frac{6}{\pi}\sin^{-1}\left(\frac{\rho}{2}\right)$$

## Appendix D: Population factor models for chapter 3, study 1

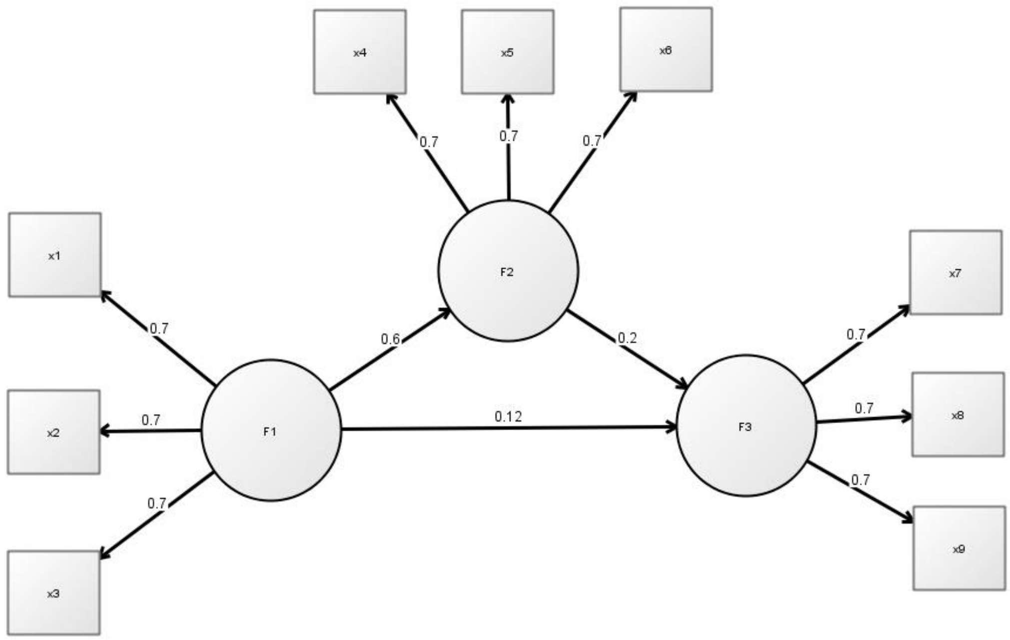
Three Factor Model for Curran, West & Finch (1996) with 9 indicators ("Model 1")



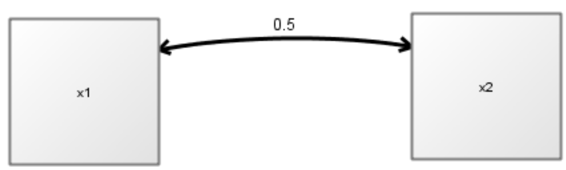

One Factor Base Model from Flora & Curran (2004) with 10 indicators.



Latent Mediation Model from Finch, West and MacKinnon (1997)




Bivariate correlation from Skidmore and Thompson (2011)



## Appendix E: Analysis of solutions for the Fleishman system of polynomial equations

The polynomial system defined by Fleishman (1978) is:

$$a = -c$$

$$b^2 + 6bd + 2c^2 + 15d^2 = 1$$

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

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

By Bézout's theorem of homogeneous polynomial equations, the maximum number of solutions for this system is 24. Readers interested in the details of this theorem please consult Lang (2014). What follows is a demonstration of the number of real-valued solutions to this system.

Take $b^2 + 6bd + 2c^2 + 15d^2 = 1$ and express it in terms of $c$. It then becomes apparent that:

$$a = -c \qquad\qquad c = \pm\sqrt{\frac{1 - (b^2 + 6bd + 15d^2)}{2}}$$

are solutions to the system in terms of $b$ and $d$. Coefficients $b$ and $d$ also need to satisfy the constraints:

$$b^2 + 6bd + 15d^2 \leq 1 \qquad\qquad c = \frac{\gamma_1}{2(b^2 + 24bd + 105d^2 + 2)}$$

By the symmetry of the inequality $b^2 + 6bd + 15d^2 \leq 1$ it is possible to see that if the set $\{b, d\}$ is a solution to the system, then $\{-b, -d\}$ is also a solution to the system. It now becomes apparent that whether one uses $\{b, d\}$ or $\{-b, -d\}$, coefficient $c$ has two possible real-valued solutions in terms its positive and the negative radical expression. And since coefficient $a$ is fully determined by the sign of $c$ it can only take on two values,$\{c, -c\}$. Therefore we have two solutions (with switching signs) for $\{b, d\}$ and two solutions for $\{a, c\}$ making it a total of 4 possible solution sets.

With this fact established, the task is now to explore possible bounds for the $(\gamma_1, \gamma_2)$ values to obtain real-valued solutions to the system. For simplicity we will not consider solutions of $a$ because its values are fully determined by any solution obtained for $c$.

Take $b^2 + 6bd + 2c^2 + 15d^2 = 1$ and solve for $b$ to obtain:

$$b = \pm\sqrt{1 - 2c^2 - 6d^2} - 3d$$

We will only now work with $b = \sqrt{1 - 2c^2 - 6d^2} - 3d$ while keeping in mind that, by symmetry, any algebraic solution to the positive radical of $b$ will also be a solution to the negative radical.

Substituting this expression for $b$ in the other two equations yields the new system:

$$4c^3 + 6c\left[-1 + 6d\left(-2d + \sqrt{1 - 2c^2 - 6d^2}\right)\right] + \gamma_1 = 0$$

$$-24\left[2c^4 + c^2\left(-2 - 60d^2 + 22d\sqrt{1 - 2c^2 - 6d^2}\right) + d\left(-9d - 81d^3 + \sqrt{1 - 2c^2 - 6d^2} + 48d^2\sqrt{1 - 2c^2 - 6d^2}\right)\right] = \gamma_2$$

which has two equations in with two unknowns.

Maple version 13 (Monagan *et.al.*, 2012) was used to solve the second equation containing $\gamma_2$ in order to derive closed form expressions for $c$ and $d$. In total, 12 possible solutions were found. given the following conditions. Notice that not all solutions are reproduced given the complexity both of the conditions and the high-degree polynomial expressions.

If $d = 0$:

$$c_1 = \pm\sqrt{\frac{6 - \sqrt{3}\sqrt{12 - \gamma_2}}{12}}, \quad c_2 = \pm\sqrt{\frac{6 + \sqrt{3}\sqrt{12 - \gamma_2}}{12}}$$

If $\pm\sqrt{366 + 11\sqrt{6}\sqrt{-1638 - 799\gamma_2}} \neq 0$ then:

$$d_1 = \pm\sqrt{\frac{366 + 11\sqrt{6}\sqrt{-1638 - 799\gamma_2}}{28764}}, \quad d_1 = \pm\sqrt{\frac{366 - 11\sqrt{6}\sqrt{-1638 - 799\gamma_2}}{28764}},$$

$$c_1 = \pm\sqrt{\frac{-351 + 4\sqrt{6}\sqrt{-1638 - 799\gamma_2}}{4794}}, \quad c_2 = \pm\sqrt{\frac{-351 - 4\sqrt{6}\sqrt{-1638 - 799\gamma_2}}{4794}}$$

Notice that in the second case, there truly are only 4 (not 8) different solutions because the conditions:

$$\sqrt{366 + 11\sqrt{6}\sqrt{-1638 - 799\gamma_2}} \neq 0 \text{ and } -\sqrt{366 + 11\sqrt{6}\sqrt{-1638 - 799\gamma_2}} \neq 0$$

are the same for purposes of obtaining the bounds of $(\gamma_1, \gamma_2)$ that make the system have real-valued roots.

Substituting either the positive or negative radical expressions for $\{c, d\}$ in the first equation of the new system gives the final conditions for $(\gamma_1, \gamma_2)$ to be real-valued as follows.

IF $\pm\sqrt{366 \pm 11\sqrt{6}\sqrt{-1638 - 799\gamma_2}} \neq 0$ AND:

$$\gamma_1 \leq \frac{A\left(-2880\sqrt{6} + 3\sqrt{5130 - 3\sqrt{6}\sqrt{-1638 - 799\gamma_2}}\sqrt{366 + 11\sqrt{6}\sqrt{-1638 - 799\gamma_2}} - 74\sqrt{-1638 - 799\gamma_2}\right)}{2397\sqrt{799}}$$

where $A = -\sqrt{-351 - 4\sqrt{6}\sqrt{-1638 - 799\gamma_2}}$

the original Felishman (1978) polynomial system will generate real-valued solutions.

At least another 7 possible conditions like this one can be derived, but some of them are high-degree polynomials that will have no closed-form expression. Still, this is the first account within the published literature where explicit bounds to the 3$^{rd}$-order polynomial method are reported and could be used to evaluate the performance of optimizing algorithms close to the solution space.

## APPENDIX F: Analysis of solutions for the intermediate correlation equation

Vale and Maurelli's (1983) intermediate correlation equation is defined as:

$$r_{Y_1Y_2} = r_{Z_1Z_2}(b_1b_2 + 3b_1d_2 + 3d_1b_2 + 9d_1d_2) + r_{Z_1Z_2}^2(2c_1c_2) + r_{Z_1Z_2}^3(6d_1d_2)$$

Let $z \coloneqq r_{Z_1Z_2},\ q \coloneqq -r_{Y_1Y_2},\ \delta \coloneqq (b_1 + 3d_1)(b_2 + 3d_2),\ \beta \coloneqq 2c_1c_2,\ \alpha \coloneqq 6d_1d_2$ to simplify it as:

$$\alpha z^3 + \beta z^2 + \delta z + q = 0$$

It becomes immediately apparent that this is a polynomial of the third degree. Notice that since $q$ is chosen by the researcher, it is always a pre-determined constant.

By relying on the Cardano method to find the roots of a cubic equation (readers interested in the details of this method can consult Mollin (1947)) it is possible to see that:

$$z_1 = S + T - \frac{\beta}{3\alpha}$$

$$z_2 = -\frac{S+T}{2} - \frac{\beta}{3\alpha} + \frac{i\sqrt{3}}{2}(S - T)$$

$$z_3 = -\frac{S+T}{2} - \frac{\beta}{3\alpha} - \frac{i\sqrt{3}}{2}(S - T)$$

where:

$$S = \sqrt[3]{R + \sqrt{Q^3 + R^2}}$$

$$T = \sqrt[3]{R - \sqrt{Q^3 + R^2}}$$

with:

$$Q = \frac{3\alpha\delta - \beta^2}{9\alpha^2}, \qquad R = \frac{9\alpha\beta\delta - 27\alpha^2 q - 2\beta^3}{54\alpha^3}$$

The expression $Q^3 + R^2$ is referred to as the discriminant and it describes the nature of the solutions as falling in one of three possible scenarios.

If $Q^3 + R^2 = 0$:

$$Q^3 + R^2 = 0 \Rightarrow S - T = 0$$

$$\Rightarrow \frac{i\sqrt{3}}{2}(S - T) = 0$$

which implies $z_1, z_2, z_3$ are all real-valued numbers, with $z_2 = z_3$.

If $Q^3 + R^2 > 0$:

$$Q^3 + R^2 > 0 \Rightarrow \sqrt{Q^3 + R^2} \in \mathbb{R}$$

$$\Rightarrow S \neq T, \qquad S, T \in \mathbb{R}$$

$$\Rightarrow \frac{i\sqrt{3}}{2}(S - T) \neq 0, \quad \frac{i\sqrt{3}}{2}(S - T) \in \mathbb{C}$$

which implies $z_1$is the only real-valued root, with $z_2, z_3$ being complex conjugates of one another.

If $Q^3 + R^2 < 0$:

$$Q^3 + R^2 < 0 \Rightarrow \sqrt{Q^3 + R^2} \in \mathbb{C}$$

$$\Rightarrow S \neq T, \qquad S, T \in \mathbb{C}$$

$$\Rightarrow \frac{i\sqrt{3}}{2}(S - T) \neq 0, \quad \frac{i\sqrt{3}}{2}(S - T) \in \mathbb{R}$$

which implies $z_1, z_2, z_3$ are all real-valued numbers and all three are solutions are different.

A back substitution to the original polynomial coefficients yields:

$$R = \frac{9\alpha\beta\delta - 27\alpha^2 q - 2\beta^3}{54\alpha^3} = \frac{54c_1c_2d_1d_2(b_1 + 3d_1)(b_2 + 3d_2) + 486d_1^2d_2^2r_{Y_1Y_2} - 8d_1^3d_2^3}{5832d_1^3d_2^3}$$

$$Q = \frac{3\alpha\delta - \beta^2}{9\alpha^2} = \frac{9d_1d_2(b_1 + 3d_1)(b_2 + 3d_2) - 2c_1^2c_2^2}{162d_1^2d_2^2}$$

Since $R$ in the discriminant is being squared, it is always positive. Therefore, the nature of the solution is almost exclusively defined by the numerator of $Q$. Expanding the numerator we obtain:

$$9(d_1d_2b_1b_2 + 3d_1^2b_2d_2 + 3d_2^2b_1d_1 + 9d_1^2d_2^2) - 2c_1^2c_2^2$$

which does not simplify further. The cases where this expression is less than 0 (primarily for large $c_1, c_2$) determine the cases where there could potentially be more than one intermediate correlation value.

## APPENDIX G: Population factor model for study 2, chapter 4

Three Factor Model for Curran, West & Finch (1996) with 9 indicators (“Model 1”)