INVESTIGATING THE PERFORMANCE OF

PROCEDURES FOR CORRELATIONS

UNDER NONNORMALITY

By

Miriam Kraatz

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

December, 2011

Nashville, Tennessee

Approved:

Professor James H. Steiger

Professor David A. Cole

Professor Andrew J. Tomarken

Professor David Lubinski

To my mother, who brought me into this world,

and who always kept fighting for the family.

ACKNOWLEDGEMENTS

I feel deeply indebted to my advisor, Dr. James H. Steiger, who enabled me to pursue this degree and very patiently helped me along the way during the past seven years. I feel that have not only found an academic advisor, but a friend. I am also grateful to the other members of my Dissertation Committee, Dr. David A. Cole, Dr. David Lubinski, and Dr. Andrew J. Tomarken, who have provided me with helpful advice during my time at Vanderbilt and who have been supportive of my career goals. Without the funding provided by the graduate school of Vanderbilt University and the Department of Psychology and Human Development, this dissertation would not have been possible.

Graduate school, in a foreign country no less, would have been infinitely harder without the support of David, who is the best friend one can imagine and like family to me. I would like to thank all my friends for bearing with me during this process: Peng Lai Kung Fu, Rocketown and the bboys and bgirls in Nashville, my colleagues at AdvanceMed, Dschin-u, Keri, Keith, and Georges.

Table of Contents

vi

List of Tables

List of Figures

ix

x

General Introduction

Assume you are a psychometrician at a large university on tenure track, and have developed an inferential technique for correlations. You want to ensure that the technique performs well under a wide range of circumstances, including nonnormal parent distributions, and therefore you choose a popular method for simulating nonnormal bivariate data by varying skewness and kurtosis. You run the Monte Carlo simulations, and things do not turn out quite as well as you had hoped — in some conditions of skewness and kurtosis, the performance of your method is poor. Nevertheless, you send your paper to *Psychological Methods* for review. The result — rejection.

Two years later, completely independently, a fellow researcher develops the *exact same inferential technique*. Just as you did, he examines the performance of the technique under a wide range of skewness and kurtosis conditions. However, there is one important difference between his work and yours: He uses a different simulation technique. His study finds reasonably good performance in the conditions where you found poor performance. He sends his paper to *Psychological Methods*, gets published, and, in the course of several years, receives many citations for the new correlational technique which gains great popularity. How could this happen?

Multivariate Data Analysis. Multivariate data analysis is a cornerstone in modern social sciences research. A fundamental part of such multivariate data analysis is the analysis of association between two or more variables, and many procedures exist that test hypotheses about or provide point and interval estimates for single estimates of association, coefficients of multiple determination, and patterns consisting of several estimates of association. Many

advanced statistical techniques such as Path Analysis, Structural Equations Modeling (SEM), and Factor Analysis are based on the calculation of some form of correlation. The broad involvement of correlational procedures in all kinds of scientific research makes it paramount that such procedures perform as expected and that their performance in a wide variety of (if not *all possible*) circumstances is investigated.

The traditional hypothesis tests and confidence intervals for the best known measure of association, Pearson's product moment correlation *r*, are based on the assumption of normal parent distributions, i.e., it is assumed that the variables for which the correlation is calculated have a bi–/multivariate normal distribution. If we calculate sample correlations for bivariate normal data, the Fisher *Z* transform can be used to transform the sample correlation into a variable that is close to normally distributed, which in turn can be used to create quite accurate test statistics and confidence intervals. While numerous specialized procedures and coefficients exist that cater more towards data that deviate from the ideal of normal parents, many researchers will still use the procedures based on the Pearson correlation coefficient, hoping that they will retain most of their validity even under nonnormality of parent distributions. Reasons for adhering to analyses based on normal theory may include: a) Lack of knowledge of the specialized procedures, or, indeed, of the robustness issue itself; b) The specialized procedures have not been implemented in data analysis software; c) Higher level analysis techniques such as factor analysis and SEM, were originally developed using the traditional coefficients; d) The advantages and improvements of the special procedure over the standard Pearson correlation coefficient are not fully established, or at least not salient to the researcher.

Nonnormality and Correlation. In his oft–cited 1989 study, Micceri demonstrates that we can expect the majority of univariate real data sets for achievement and attitude measurements to *not* have a normal distribution. An equivalent study for bi– or multivariate data seems to await completion, but even lacking concrete evidence, it is safe to assume that bivariate (and multivariate) nonnormality is at least as common as univariate nonnormality, for one simple reason: For the joint distribution of variables *X* and *Y* to be bivariate normal, the following condition needs to be true: Every linear combination $W = aX + bY$ with at least one *a* or *b* not equal to zero needs to be normally distributed as well. Hence, both the marginal distributions of *X* and *Y* (set either *a* or *b* equal to zero) *and* all the conditional distributions have to be normal. From this, we can expect that there are far more nonnormal than normal bi– and multivariate distributions in real data, as Micceri's study seems to indicate that the marginals alone will probably be nonnormal. Even if the marginal distributions are close to normal, their joint bivariate distribution may be distinctly nonnormal (see, e.g., Joliffe & Hope (1996) for a real bivariate data set with normal marginals but nonnormal joint distribution).

The ensuing dilemma is obvious: The Pearson correlation is employed almost universally. Inferences and confidence intervals based on correlations rely on the assumption of bivariate normal parent distributions. Yet bi- or multivariate normal data may be a rare exception in actual statistical practice. This discrepancy between the assumption of normally distributed data and nonnormal reality is also true for other data analysis procedures, but will often be mitigated by the central limit theorem, as is the case for the well-known *t*–test for a single mean. To be able to use the advantages of the central limit theorem, we need two things:

1) Increasing sample size and

2) Knowledge of the asymptotic variance σ^2 of the parameter in question, estimated by s^2.

A as sample size increases, the shape of sampling distribution of the mean can be approximated by a normal distribution, and its standard deviation can be approximated by $S/\sqrt{n}$.

What happens if we calculate correlations for samples drawn from a bivariate nonnormal parent distribution? The *shape* of the distribution of these sample correlations will be approximated by a normal distribution as sample size increases, but finding or approximating the standard deviation of the distribution is not as straightforward as for the sample mean. Further, the rate at which the approximation to the normal shape takes place is not well known. Several approaches have been taken to remedy these difficulties, including measures of association for nonnormal data other than the correlation, and some will argue that these are better at capturing the true nature of such data or that normalizing transformations of the univariate parent variables are the best option. However, such approaches shall not be elaborated on in my dissertation, and I will account for this later on. If we want to continue to use the product-moment correlation as well as the many analysis procedures based on normal theory, contrasted with reality which often is far from normal, we need to subject these procedures to strict tests regarding their performance under nonnormality. To do so, simulation studies have been the main staple, requiring the generation of bi–/multivariate nonnormal distributions.

My dissertation is divided into two parts. Part I explores available techniques for multivariate nonnormal data generation. The investigation of multivariate nonnormal distributions in psychometric literature seems to be not as well developed as it is in mathematical statistics. To bridge this gap and make the relevance more obvious to applied psychometricians, I will concentrate on bivariate nonnormality, specifically providing an in depth analysis of the most popular simulation technique to this date as well as two very recently published

4

competitors. Part I will conclude with a discussion of the implications of differences between simulated nonnormal distributions.

Part II examines the evaluation of confidence interval performance itself, and provides a new perspective on understanding and interpreting approximate CIs. Three conceptually different approaches to constructing a confidence interval around r will be applied to nonnormal distributions simulated from one of the methods discussed in Part I. The evaluation of these confidence intervals will reach a depth that has not been offered in previous Monte Carlo research.

PART I

Pitfalls of 'Blind' Monte Carlo Simulations

This dissertation begins with a key point that, in a sense, is an old story. In fact, many of us first received this point as undergraduates, and now routinely teach it in our introductory statistics classes. And yet, as psychological research has demonstrated repeatedly, even experienced senior scientists sometimes fail to identify all potential applications of a scientific concept when presented in a new context. As we shall show, that is unfortunately the case here.

This old story was told in 1973[1] by Francis John Anscombe as the well–known "Anscombe's Quartet." Anscombe presented four bivariate data sets with the same means, standard deviations, and correlation, but very different bivariate shapes as a whole; his original four graphs are reproduced in Figure 1. Using a total of six variables, X_1, X_2, Y_1, Y_2, Y_3, and Y_4, the graph in the upper left hand corner is constructed from X_1 and Y_1, graph 2 from X_1 and Y_2, graph 3 from X_1 and Y_3, and graph 4 from X_2 and Y_4. Both X variables have a mean of 9, and a variance of 10, and all four Y variables have a mean of 7.5, and a variance of 3.75. All four regression lines have the same equation: $Y = 3 + .5X$. The correctness of these descriptive statistics is undisputed, but it is quite striking how the linear model that is implied by the regression line is only appropriate for the first data set in graph 1. The second bivariate data set exhibits a strong nonlinear relationship, the third includes an outlier with formidable influence on

---

[1] Interestingly, Anscombe published his quartet right around the same time that both the investigation/definition of multivariate parameters and the simulation of multivariate distributions as well as the examination of r's sensitivity to multivariate nonnormality were under way.

the regression line and the fourth one may be interpreted as a data set with one dichotomous variable for which the model is just simply inadequate.
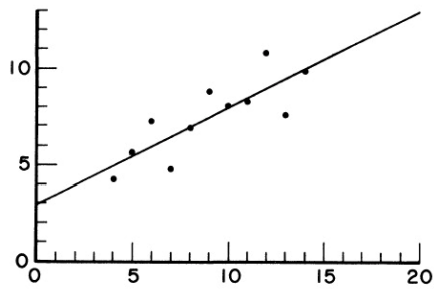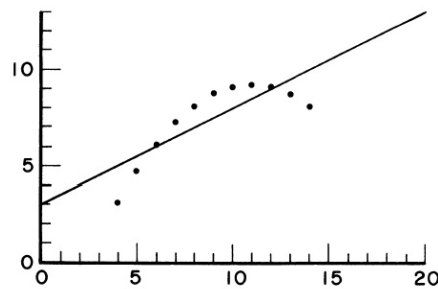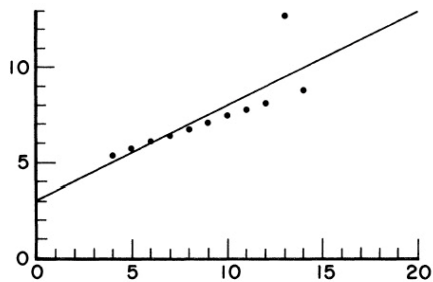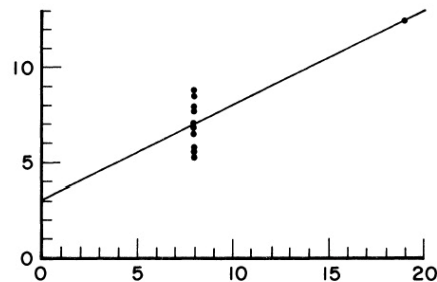
Figure 1: Anscombe's Quartet

The lesson conveyed by Anscombe's quartet is twofold: (1) While descriptive statistics may be numerically correct and can be reported without objection, taking the next step into inferential techniques such as hypothesis tests, model fitting, and prediction, requires a deeper analysis; (2) Data sets and distributions are not exhaustively described by their moments – unless you include all of them — and can in fact be radically different even if some of the moments between them agree. Of course, one might hope that the lesson of Anscombe's data and graphs would have been grasped and generalized to new situations by the researchers who were exposed to them. As we shall see below, this hope has not been justified.

7

Multivariate Nonnormal Data Generation

In the general introduction, we have already identified that shape and variance of the sampling distribution of the Pearson correlation coefficient and performance of hypothesis tests and confidence intervals for correlations under nonnormal parents must be investigated. Part I of my dissertation will explore different ways of simulating bi–/multivariate nonnormality. I will show that popular ways of describing bivariate nonnormal distributions may not be useful in capturing their shape and demonstrate that disregarded differences in bivariate nonnormal shape can have a strong impact on the sampling distribution of *r*.

Multivariate nonnormal data generation is not new and first attempts date back to the earlier years of the past century (see, e.g., Pearson, 1923, 1929; Baker, 1930). Pearson (1929), perhaps mindful that many real-world data sets are often mixtures of several distinct subpopulations, used mixtures of bivariate normal distributions to construct bivariate nonnormal distributions. Baker (1930), inspired by the skewed character of an actual data set, constructed a bivariate nonnormal distribution using two skewed marginal distributions.

As research and – most importantly – computing power progressed, various strategies for simulating bi–/multivariate nonnormal data were suggested. One prominent and rather straightforward approach is to subject a standard normal random variable *Z* to a nonlinear transformation, thus generating a nonnormal variable *Y*. To obtain multivariate samples, several standard normal random variables are first correlated (often the method described in Kaiser & Dickmann, 1962, is used) and then transformed. Such techniques have also been called "Transform and Calculate" (TC) methods (Ruscio & Kaczetow, 2008).

Three rather well known transformation methods are (1) polynomial transforms (see, e.g. Fan, Felsövalyi, Sivo & Keenan, 2003; Fleishman, 1978; Headrick, 2002; Headrick &

Sawilowsky, 1999; Vale & Maurelli, 1983) (2) the *g*-and-*h* distribution (Field & Genton, 2006; Hoaglin & Peters, 1979; Hoaglin, 1985; Kowalchuk & Headrick, 2010; Tukey, 1977), and, in a broader sense, (3) multivariate skew distributions (Azzalini & Dalla Valle, 1996; Arnold & Beaver, 1999). A good general introduction to "TC" techniques and issues they might face can be found in Li & Hammond (1975), while Ruscio & Kazcetow (2008) contrast them with iterative procedures (see next paragraph).

Other methods of simulating nonnormal multivariate data comprise mixture distributions (see, e.g., Kowalski, 1972; Edgell & Noon, 1984), a multivariate non–normal distribution based on the generalized lambda distribution (Headrick & Mugdadi, 2006), and others (Parrish, 1990). Finally, iterative procedures have been discussed or suggested by Hoyland, Kaut, & Wallace (2003), Lurie & Goldberg (1998), Ruscio & Kaczetow (2008), and Yang (2008). To briefly summarize one of the iterative techniques, consider Ruscio & Kaczetow's method and assume a multivariate nonnormal sample with k variables and a correlation matrix $\mathbf{P}$ is desired. As a first step, k univariate nonnormal samples are created by either bootstrapping from a real data set or sampling from some desired population (the user of the method will have to provide a function or quantiles for such sampling). Then, by means of factor analysis, the variables are iteratively rotated until they have a sample correlation matrix $\mathbf{R}$ which deviates from $\mathbf{P}$ only by sampling error.

During the 1990's, the proliferation of new methods to simulate multivariate nonnormal data gained momentum, as evidenced by a strong increase in related publications, mainly in mathematical statistics journals. A short list of methods and their number of citations (given are ISI web of knowledge citations with *Google scholar* citations in italics and parentheses) includes:

Table 1: Popularity of Multivariate Nonnormal Simulation Methods

3rd order polynomial method	Fleishman (1978): 173 *(263)*
	Vale & Maurelli (1983): 104 *(134)*
	Headrick & Sawilowsky (1999): 24 *(38)*
5th order polynomial method	Headrick (2002): 11, 7 of which are himself *(16)*
Multivariate skew-normal distribution	Azzalini & Dalla Valle (1996): 230 *(425)*
(Multivariate) *g*-and-*h* distribution	Hoaglin (1985): *(157)*
	Field & Genton (2006): 5 *(14)*
	Kowalchuk & Headrick (2010): 2 *(3)*
Iterative multivariate nonnormal distribution	Ruscio & Kaczetow (2008): 9 *(11)*
Using the generalized lambda distribution	Headrick & Mugdadi (2006): 6 *(10)*
Based on an exponential distribution	Arnold & Strauss (1991): 38 *(57)*
Based on the Pearson distribution system	Nagahara (2004): 11 *(17)*

Only some of these methods have been adopted into psychometrics. For example, a sizeable amount of research in purely statistical journals has focused on the (multivariate) skew(-normal) distribution (Azzalini & Dalla Valle, 1996), yet only 7 out of 230 citations on ISI are in a journal related to psychometrics. The most popular method in psychometrics to this day remains Vale & Maurelli's multivariate extension of the 3rd order polynomial transform by Fleishman. The purpose of Part I of my dissertation is to create a broader understanding of the potential issues of the 3rd order polynomial and *Z* score transformation methods in general and provide stepping stones for further evaluation of other already existing and forthcoming simulation methods. I will focus on transformation techniques which provide control over both marginal characteristics of *and* the correlation matrix between the nonnormal univariate variables. Such simulation methods include the 3rd order polynomial method and its multivariate

10

extension (Fleishman, 1978; Vale & Maurelli, 1983; see also Headrick & Sawilowsky, 1999), the 5th order polynomial method (Headrick, 2002), and the *g*-and-*h* distribution (Field & Genton, 2006; Kowalchuk & Headrick, 2010). The next section provides an in–depth look at the 3rd order polynomial transform and its multivariate extension, while the 5th order polynomial transform and the *g*-and-*h* distribution will be discussed afterwards.

The 3rd Order Polynomial Transform

Transformations for all methods discussed in my dissertation can be represented as a function applied to a standard normal random variable *Z* to obtain a nonnormal random variable *Y*:

$$Y = f(Z) \tag{1}$$

As mentioned before, the most popular transformation method in psychometrics is the 3rd order polynomial transform, also called the 3rd order polynomial method or power method, first suggested as a univariate version by Fleishman (1978), and later extended to the multivariate case by Vale & Maurelli (1983), which will be discussed below. The 3rd order power method uses a third-degree polynomial such that

$$Y = a_0 + a_1 Z + a_2 Z^2 + a_3 Z^3 \tag{2}$$

The polynomial coefficients are selected to produce desired values of skewness and kurtosis for *Y*, while mean and variance are fixed at 0 and 1.

Equation (2) uses four coefficients (a_0, a_1, a_2, and a_3) to create the nonnormal random variable (RV) *Y*. The moments about zero for a standard normal variable *Z* are well known, and the moments of *Y* are thus an easily calculated function of these coefficients. (See Appendix for details.) Skewness and kurtosis of *Y* are, in turn, a function of *Y*'s first four moments. Ultimately,

we arrive at a system of four nonlinear equations in terms of a_0 through a_3 which can then be used to determine the values of a_0 through a_3 that yield desired values of skewness and kurtosis: The mean of Y is

$$\mu_Y = E(Y) = a_0 + a_2 = 0 \tag{3}$$

This in turn means that $a_0 = -a_2$. For the variance, we have

$$
\begin{aligned}
\sigma_Y^2 &= E(Y^2) - (E(Y))^2 \\
&= E(Y^2) - 0 \\
&= a_1^2 + 6a_1 a_3 + 2a_2^2 + 15a_3^2
\end{aligned}
\tag{4}
$$

Skewness is defined as

$$
\begin{aligned}
\gamma_{1_Y} &= \left(E(Y^3) - 3E(Y^2)E(Y) + 2(E(Y))^3 \right) / \left(E(Y^2) - E(Y)^2 \right)^{3/2} \\
&= \left(E(Y^3) \right) / (1)^{3/2} \\
&= 2a_2 \left(3a_1^2 + 4a_2^2 + 36a_1 a_3 + 135a_3^2 \right)
\end{aligned}
\tag{5}
$$

because $E(Y) = 0$ and $E(Y^2) = 1$. Finally, the kurtosis of Y is defined as

$$
\begin{aligned}
\gamma_{2_Y} &= E(Y^4) - 4E(Y^3)E(Y) - 3\left(E(Y^2) \right)^2 + 12E(Y^2)(E(Y))^2 \\
&\quad + 6(E(Y))^4 / \left(E(Y^2) - E(Y)^2 \right)^2 \\
&= E(Y^4) - 3 \\
&= 3a_1^4 + 60a_1^2 a_2^2 + 60a_2^4 + 60a_1^3 a_3 + 936a_1 a_2^2 a_3 + 630a_1^2 a_3^2 \\
&\quad + 4500a_2^2 a_3^2 + 3780b a_3^3 + 10395a_3^4 - 3
\end{aligned}
\tag{6}
$$

Note that this defines skewness and kurtosis so that a normal distribution has $\gamma_1 = 0$ and $\gamma_2 = 0$.

If we substitute $a_0 = -a_2$, we are left with three unknowns (a_1, a_2, and a_3) in three equations (Equation (4), (5), and (6)). (Actually, since the variance has been set equal to 1, we

12

could solve for a_2, thereby having only two equations and two unknowns left, but the equations will become long and unsightly and the result will be the same.) To construct a random variable with desired skewness and kurtosis, we can now numerically solve the following set of nonlinear equations for the three coefficients a_1, a_2, and a_3 and later set $a_0 = -a_2$:

$$
\begin{aligned}
variance &= a_1{}^2 + 6a_1a_3 + 2a_2{}^2 + 15a_3{}^2 = 1 \\
skewness &= 2a_2\left(3a_1{}^2 + 4a_2{}^2 + 36a_1a_3 + 135a_3{}^2\right) \\
kurtosis &= 3a_1{}^4 + 60a_1{}^2a_2{}^2 + 60a_2{}^4 + 60a_1{}^3a_3 + 936a_1a_2{}^2a_3 + \\
&\quad 630a_1{}^2a_3{}^2 + 4500a_2{}^2a_3{}^2 + 3780a_1a_3{}^3 + 10395a_3{}^4 - 3
\end{aligned}
\tag{7}
$$

In his 1978 article, Fleishman provides a table of coefficients for a number of different skewness-kurtosis combinations. *Figure 2* is an excerpt of Table 1, pp. 524 from Fleishman (1978) (note that Fleishman chose $a_0 = A$, $a_1 = B$, $a_2 = C$, and $a_3 = D$ as names for the coefficients).

524   PSYCHOMETRIKA

TABLE 1

Power Method Weights

Skew	Kurto-sis	B	C	D
1.75	3.75	0.92966052480111	0.39949667453766	−0.03646699281275
1.50	3.75	0.86588620352314	0.22102762101262	0.02722069915809
1.50	3.50	0.88690855456083	0.23272187792846	0.01875401444244
1.50	3.25	0.91023877496903	0.24780864411835	0.00869952997029
1.50	3.00	0.93620992090360	0.26831868225542	−0.00368190099903
1.50	2.75	0.96443747224458	0.29807621191230	−0.01963521430303
1.50	2.50	0.99209856718687	0.34526935903177	−0.04181526211241
1.25	3.75	0.81888156132542	0.16064255561731	0.04916517172492
1.25	3.50	0.83472669039047	0.16546665419634	0.04385221308384
1.25	3.25	0.85174062710067	0.17101073821620	0.03803066692496
1.25	3.00	0.87016387886005	0.17749222807992	0.03157509494526
1.25	2.75	0.89031839050274	0.18523508277808	0.02430713561023
1.25	2.50	0.91264314105424	0.19474622768576	0.01596248199126

Figure 2: Fleishman's Transformation Coefficients

Equivalently to $a_0 = -a_2$, we have $A = -C$, therefore a separate column for A is not provided. As an example, to achieve a nonnormal random variable with a mean of 0, a variance of 1, a skewness of 1.5, and a kurtosis of 3.75, we need to construct Y as

$$Y_1 = -.22103 + .86589 \times Z + .22103 \times Z^2 + .02722 \times Z^3 \qquad (8)$$

Visual Representation of 3rd Order Polynomial Transformations and Resulting Distributions

To better understand the nature of these transformations, examine Figure 3, which graphs the transformation in Equation (8):
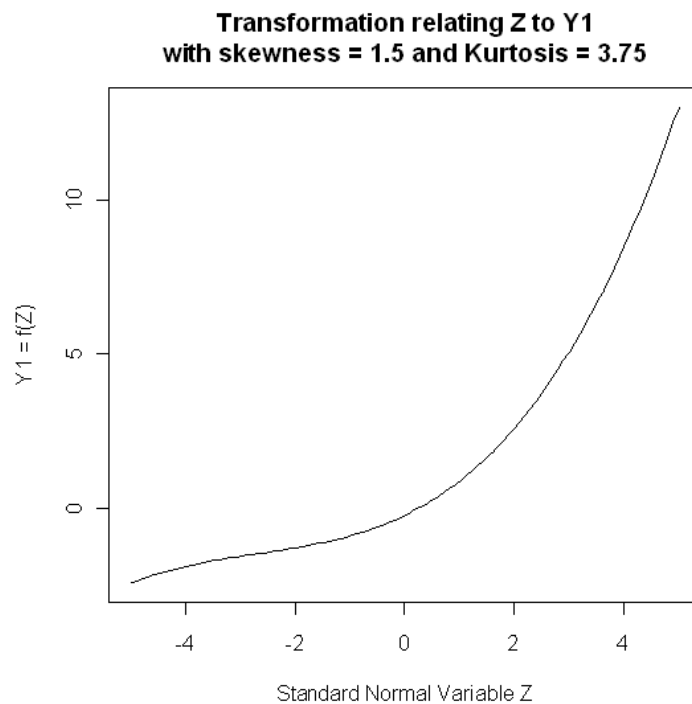
Figure 3: Example 3rd Order Polynomial Transformation for Y_1

The resulting empirical distribution (with $n = 1,000,000$), next to its parent standard normal distribution looks like this:
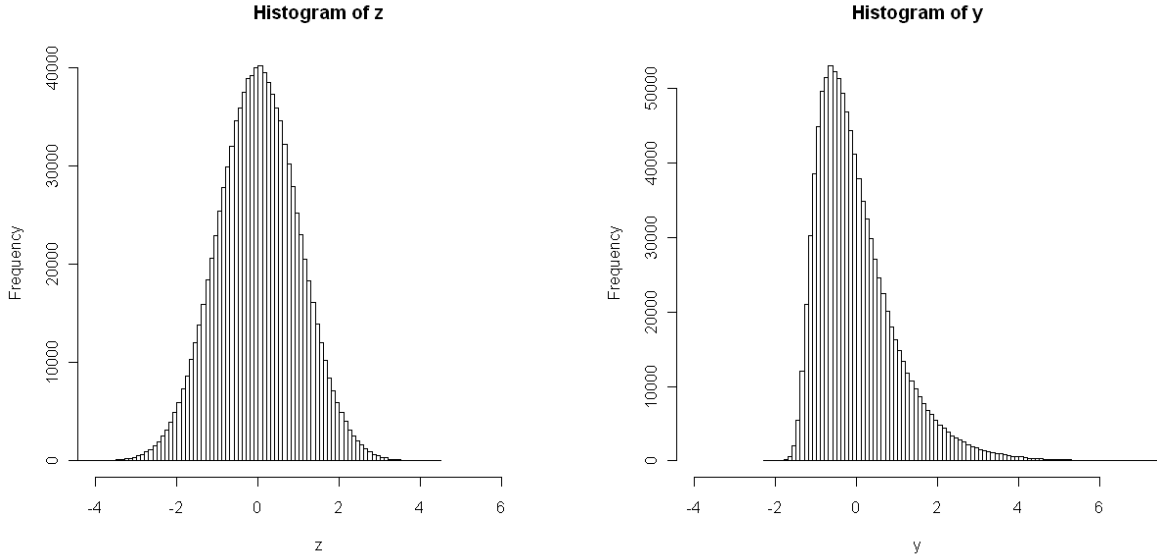
Figure 4: Parent and Final Distribution Histogram for Y_1

It is relatively easy to see how the function in Figure 3 relates the two distributions in Figure 4 to each other, as negative values for Z are matched with values between around –2 and 0 for Y while especially large positive values of Z translate into a long tail on the right hand side of Y's distribution.

So far, the univariate power method seems rather straightforward:

Step 1: Choose desired skewness and kurtosis

Step 2: Solve the set of nonlinear equations in (7) to find coefficients a_0, a_1, a_2, and a_3

Step 3: Use coefficients a_0, a_1, a_2, and a_3 to construct Y as in Equation (2).

Unfortunately, not all of the steps are quite as simple as they may seem. Specifically, the solutions to the set of nonlinear equations in step 2 are not necessarily *unique*. For example, for the combination $\gamma_1 = 1.5$, $\gamma_2 = 3.75$, aside from $a_0 = -.22103$, $a_1 = .86589$, $a_2 = .22103$, and $a_3 = .02722$, the coefficients $a_0 = -.472154$, $a_1 = 1.05881$, $a_2 = .472154$, and $a_3 = -.127813$,

15

present a solution to the set of equations in (7) as well. Using these coefficients in a second transformation to create Y_2, we have

$$Y_2 = -.472154 + 1.05881 \times Z + .472154 \times Z^2 - .127813 \times Z^3 \qquad (9)$$

The graph of the transformation in Equation (9) is pictured in Figure 5:
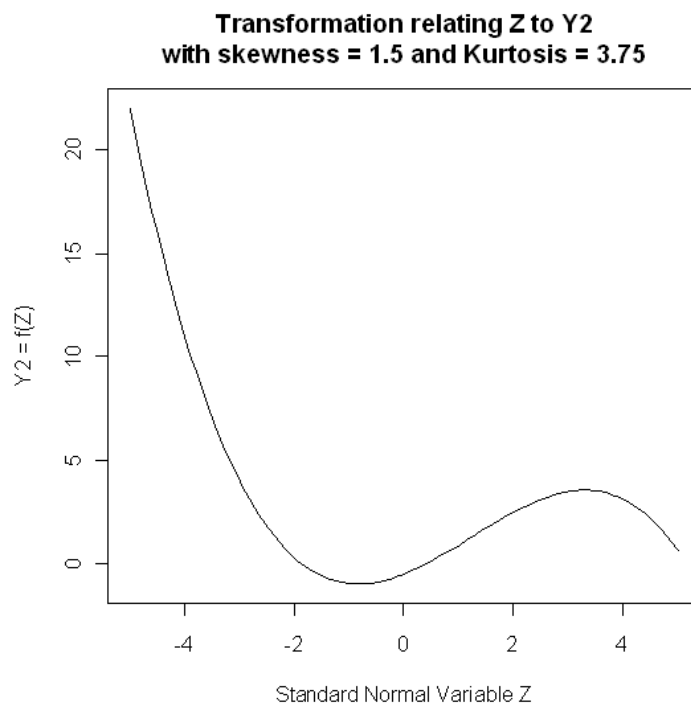
Figure 5: Example 3ʳᵈ Order Polynomial Transformation for Y_2

Figure 6 shows empirical pdfs of the normal RV Z and the transformed Y_2 next to each other:

16
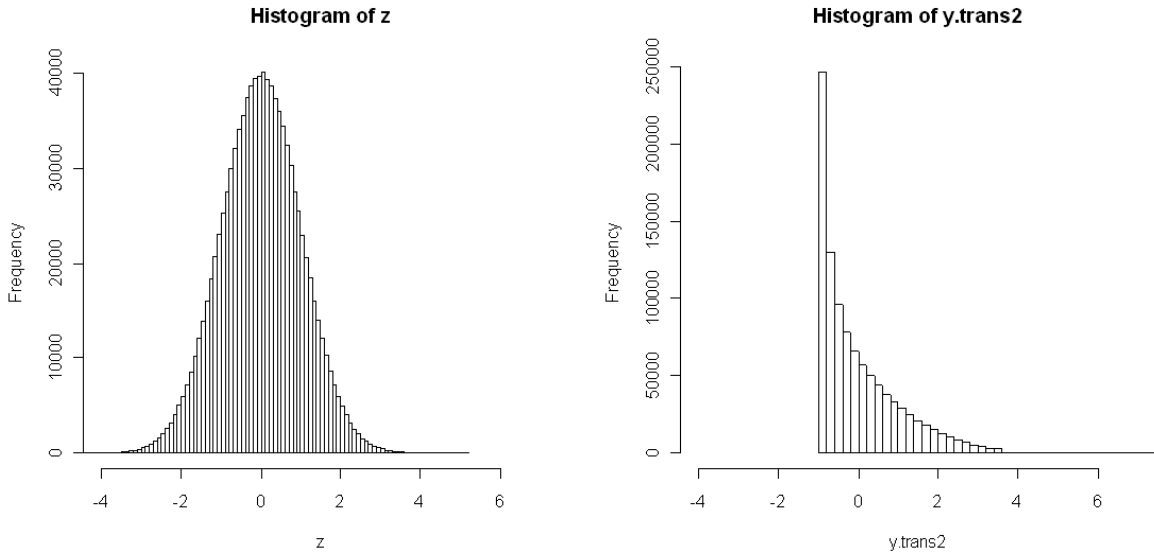
Figure 6: Parent and Final Distribution Histogram for Y_2

The distribution of Y_2, resulting from the second transformation, has quite a different shape compared to the distribution of Y_1! It is easy to see how the shape of Y_2's histogram results from the transformation in graphed Figure 5: Around $Z = -.5$, the transformation has a local minimum, relating almost all points on Z between –2 and 0 to a value on Y_2 close to about –1. This results in the mode we see at the low end of the distribution of Y_2. Values of Z between 0 and about 3 are almost linearly related to values on Y that lie between –1 and about 3.5, resulting in the relatively smooth right side tail of Y.

To establish the accurateness of these transformations, we can insert both sets of transformation coefficients into Equations (3), (4), (5), and (6). As a second way to assess the method's performance, we can calculate first four sample moments which are defined as:

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} \qquad (10)$$

17

$$S^2 = \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}{N-1} \qquad (11)$$

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^3 / N}{\left(\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 / N\right)^{3/2}} \qquad (12)$$

$$\hat{\gamma}_2 = \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^4 / N}{\left(\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 / N\right)^{2}} - 3 \qquad (13)$$

for a very large sample for each Y_1 and Y_2. The fifth and sixth sample moment, which will become relevant as we discuss the 5th order polynomial transform later on, are defined as:

$$\hat{\gamma}_3 = \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^5 / N}{\left(\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 / N\right)^{5/2}} - 10 \times \hat{\gamma}_1 \qquad (14)$$

$$\hat{\gamma}_4 = \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^6 / N}{\left(\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 / N\right)^{3}} - 15 \times \hat{\gamma}_2 - 10 \times \hat{\gamma}_1^2 - 15 \qquad (15)$$

$\bar{Y}$, S_Y^2, $\hat{\gamma}_1$, and $\hat{\gamma}_2$ for Y_1 and Y_2 turn out to be essentially the same (unless otherwise stated, empirical moments or correlations are based on sample sizes of 5,000,000):

Table 2: Empirical First Four Moments for the Transformed Variables Y_1 and Y_2

	Mean	Variance	Skewness	Kurtosis
Y_1	–0.000055	0.999947	1.498564	3.723273
Y_2	–0.000359	0.998610	1.490766	3.598177

As a matter of fact, the two solutions presented above are not the only ones. The set of equations

$$1 = a_1{}^2 + 6a_1a_3 + 2a_2{}^2 + 15a_3{}^2$$

$$1.5 = 2a_2\left(3a_1{}^2 + 4a_2{}^2 + 36a_1a_3 + 135a_3{}^2\right)$$

$$3.75 = 3a_1{}^4 + 60a_1{}^2a_2{}^2 + 60a_2{}^4 + 60a_1{}^3a_3 + 936a_1a_2{}^2a_3 +$$

$$630a_1{}^2a_3{}^2 + 4500a_2{}^2a_3{}^2 + 3780a_1a_3{}^3 + 10395a_3{}^4 - 3$$

$$(16)$$

has a total of four distinct real (and a substantial number of imaginary) solutions:

set1: $a_0 = -0.221028$, $a_1 = 0.865886$, $a_2 = 0.221028$, $a_3 = 0.0272207$

set2: $a_0 = -0.221028$, $a_1 = -0.865886$, $a_2 = 0.221028$, $a_3 = -0.0272207$

set3: $a_0 = -0.472154$, $a_1 = -1.05881$, $a_2 = 0.472154$, $a_3 = 0.127813$

set4: $a_0 = -0.472154$, $a_1 = 1.05881$, $a_2 = 0.472154$, $a_3 = -0.127813$

$$(17)$$

You may notice that the transformation based on the second set of coefficients is a mirror image about the y-axis of the transformation based on the first, with the sign of both a_1 and a_3 reversed. The same is true for sets 3 and 4. Closer inspection of the equations in (7) shows that if $a_1{}^*$, $a_2{}^*$, and $a_3{}^*$ present some particular solution, $-a_1{}^*$, $a_2{}^*$, and $-a_3{}^*$ must be a solution as well, since all products involving a_1 and a_3 have an even number of coefficients. The pdf of $Y = -a_2 + a_1Z + a_2Z^2 + a_3Z^3$ will be equal to the pdf of $Y = -a_2 - a_1Z + a_2Z^2 - a_3Z^3$, since the pdfs of Z and Z^3 are symmetric (see also Property 4.2 in Headrick & Kowalchuk, 2007). However, as we have already seen, which of the two *distinctly different* solutions (e.g. set 1 or set 4) is chosen can make a substantial difference. Finally, it seems that whenever *one* real solution to the set of equations in (7) exists, there will always be four real solutions for any given skewness-kurtosis combination, two of which will be distinctly different.

   A Second Example. Solving for $\gamma_1 = 0$ and $\gamma_2 = 25$ proves to be an even more striking example of how the resulting distributions can differ, depending on the choice of transformation coefficients. For $\gamma_1 = 0$ and $\gamma_2 = 25$, the four possible solutions are:

$$a_0 = 0, \ a_1 = -1.56668, \ a_2 = 0, \ a_3 = 0.348173$$
$$a_0 = 0, \ a_1 = 1.56668, \ a_2 = 0, \ a_3 = -0.348173$$
$$a_0 = 0, \ a_1 = -0.255283, \ a_2 = 0, \ a_3 = -0.203755 \qquad (18)$$
$$a_0 = 0, \ a_1 = 0.255283, \ a_2 = 0, \ a_3 = 0.203755$$

Choosing the first and the last set of coefficients, the resulting transformations look like this:
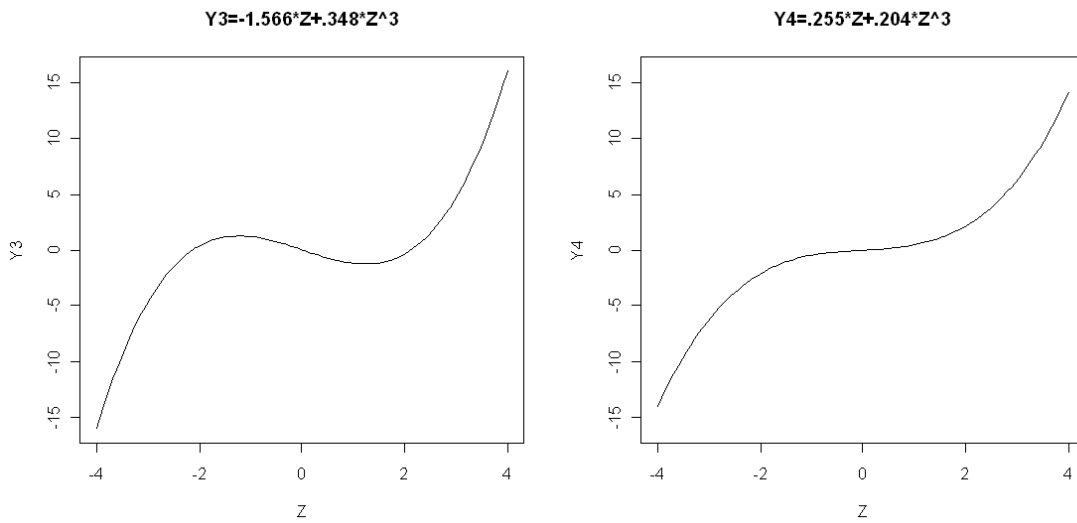
Figure 7: Two Distinctly Different Transformations for $\gamma_1 = 0$ and $\gamma_2 = 25$

These don't look *too* different, do they? What about the resulting distributions?

$$Y_3 = -1.56668 \times Z + .348173 \times Z^3 \qquad\qquad Y_4 = .255283 \times Z + .203755 \times Z^3$$

Nonnormal Distribution with g1 = 0, g2 = 25 **Nonnormal Distribution with g1 = 0, g2 = 25**
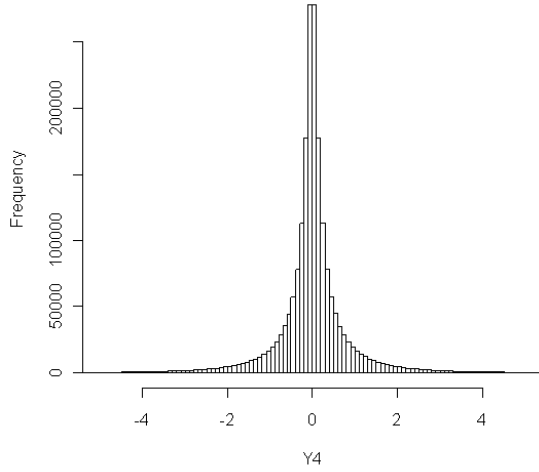
Figure 8: Univariate Distributions Resulting from Transformations in Figure 7

The two distributions look strikingly different, yet, they have essentially the same first four sample moments:

Table 3: Empirical First Four Moments for the Transformed Variables with $\gamma_1 = 0$ and $\gamma_2 = 25$

	Mean	Variance	Skewness	Kurtosis
Y_3	−.000199	0.999167	−.029793	24.133729
Y_4	−.000790	1.003323	.005143	25.608518

The question that presents itself now is which set of coefficients should be chosen when we want to construct a nonnormal univariate random variable with $\mu = 0$, $\sigma^2 = 1$, $\gamma_1 = 0$, and $\gamma_2 = 25$. If we compare all four sets of coefficients, we may notice that only the last set with $a_0 = 0$, $a_1 = 0.255283$, $a_2 = 0$, and $a_3 = 0.203755$ leads to a strictly increasing transformation function. Its mirror image, the third set of coefficients in Equation (18), will create a

21

monotonically decreasing transformation function, and the other two will lead to non-monotonic transformations. To be able to decide which one should be preferred, we need to investigate whether there will always be two monotonic and two non-monotonic transformations and what the properties of the resulting distributions are. This is done in the upcoming sections.

Limitations of the 3rd Order Polynomial Transform

Several limitations of the power method have been identified. The first of these limitations is a restriction of the range of possible skewness-kurtosis combinations (see Fleishman, 1978, whose equation (21) suffers from typographical errors; Headrick & Sawilowsky, 1999; Headrick, 2004; Tadikamalla 1980). Headrick proved in 2004 (see also Devroye, 1986) that for *any* univariate distribution, the range of valid skewness–kurtosis combinations is limited by the equation

$$\gamma_2 \geq \gamma_1^{\,2} - 2 \tag{19}$$

In other words, for a given value of skewness (γ_1), kurtosis (γ_2) can only take on values that are equal to or greater than $\text{skewness}^2 - 2$. The resulting range of valid skewness-kurtosis combinations is the area above the black line in the skewness-kurtosis plane in Figure 9.

The range of available kurtosis values for any given skewness value is further restricted when Fleishman's power method is employed. For any real-valued solution to the set of equations in (7) to exist, kurtosis must satisfy constraints that can be approximated by the following:

$$\gamma_2 > 1.588\gamma_1^{\,2} - 1.139 \tag{20}$$

Further, kurtosis also has an upper limit (see Chen & Tung, 2003): No real solutions to the set of equations in (7) exist if kurtosis exceeds approximately 101.38, and this value of allowable

22

kurtosis will be even lower for $\gamma_1 \neq 0$. The range of available skewness-kurtosis combinations for Fleishman's 3rd order polynomial is bounded by the red continuous line in Figure 9. Outside of that range, the set of equations in (7) does not have a real-valued solution.

In addition to the constraint on skewness-kurtosis combinations marked by the red line in Figure 9, only an even smaller subset of skewness-kurtosis combinations will have a set of coefficients that leads to a transformation function with $f'(Z) > 0 \; \forall Z \in \mathbb{R}$ (Chen & Tung, 2003; Headrick & Kowalchuk, 2007). Requiring $f'(Z) > 0$ is a slightly more stringent condition than *strictly increasing*, as strictly increasing allows for $f'(Z) = 0$ at individual points, so called saddle points. For example, the function $f(x) = x^3$ has a saddle point with $f'(x) = 3x^2 = 0$ at $x = 0$, but is still considered to be strictly increasing. In requiring that $f'(Z) > 0$ across the whole range of Z, we are following Headrick & Kowalchuk's *Definition 3.2*. We can find the subset of transformations with $f'(Z) > 0$ by investigating the derivative of the function in Equation (2), which is

$$f'(Z) = a_1 + 2a_2 Z + 3a_3 Z^2 \tag{21}$$

We will have

$$f'(Z) = a_1 + 2a_2 Z + 3a_3 Z^2 > 0 \tag{22}$$

if

$$\begin{aligned} &\text{a) } a_3 > 0 \text{ and} \\ &\text{b) } a_2{}^2 - 3a_1 a_3 < 0 \end{aligned} \tag{23}$$

(see also Chen & Tung, 2003).

23

Proof: Assume $Z = 0$, then Equation (22) will be satisfied only if $a_1 > 0$. Further, since the last

term for the derivative, $3a_3Z^2$, will exceed all other terms in absolute value for large enough Z,

we also need $a_3 > 0$. Any real solution to $3a_3Z^2 + 2a_2Z + a_1 = 0$ must be avoided, too. Solutions

are:

$$Z = \frac{-a_2 \pm \sqrt{a_2^2 - 3a_1a_3}}{3a_3}$$

Hence, there will only be a real-valued solution if $a_2^2 - 3a_1a_3 \geq 0$. We do *not* want a solution and

must therefore require $a_2^2 - 3a_1a_3 < 0$. A heavily modified version of this proof is also provided

in Headrick & Kowalchuk (2007). The range of skewness-kurtosis combinations that can be

constructed with a transformation that satisfies $f'(Z) > 0$ is enclosed by the blue line in Figure

9 (see also Headrick & Kowalchuk, 2007). It has been common in the existing literature to refer

to transformations which satisfy $f'(Z) > 0$ as being "monotonic" or "strictly increasing," which

is a misnomer, as these terms do include functions with saddle points (which have $f'(Z) = 0$).

However, from here on, we will adopt this terminology for the sake of consistency.

Range of possible Skewness-Kurtosis Combinations for Monotonic Fleishman Transforms, all Fleishman Transforms, and all Distributions
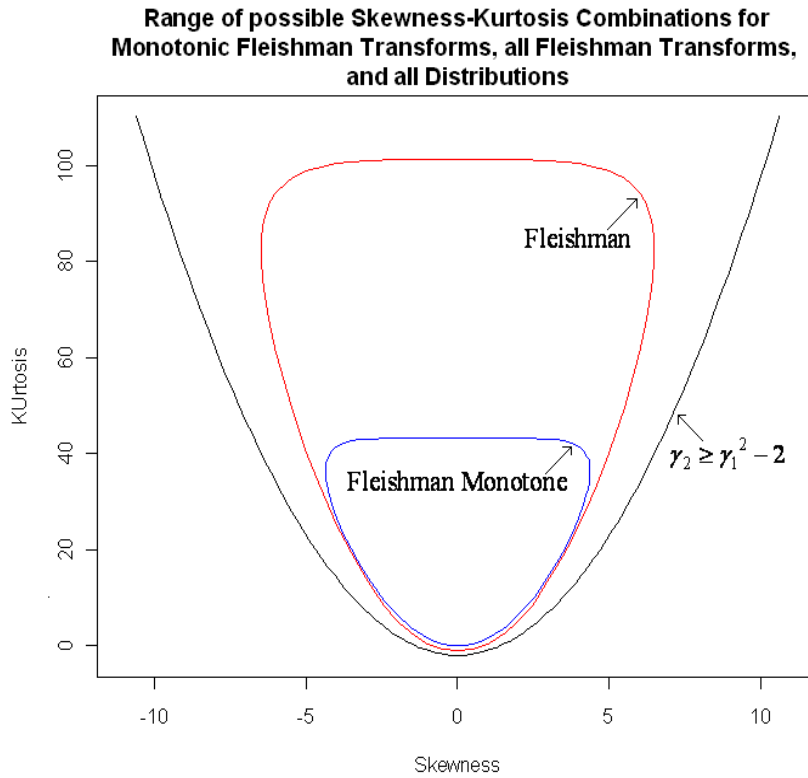
Figure 9: Univariate limitations of the range of skewness and kurtosis. The black line is the limitation due to the requirement $\gamma_2 \geq \gamma_1^2 - 2$, the red line is the limitation for any real-value solution to the set of equations in (7) to exist. The blue line bounds the set of skewness-kurtosis combinations for which a monotonic Fleishman transformation can be found.

Properties of monotonic vs. non-monotonic transformations based on the power method have been discussed by several authors (Chen & Tung, 2003; Headrick, 2004; Headrick & Kowalchuk, 2007). Headrick & Kowalchuk (2007) compare Fleishman's 3rd order polynomial method with Headrick's 5th order polynomial method (see below) and, among other things, derive the pdf and cdf as well as several properties regarding the pdf and cdf for a subset of 3rd and 5th order polynomial variables. They claim that some variables *Y* obtained through 3rd and 5th polynomial transforms will have a *valid* pdf and cdf while others will not. They argue that a nonnormal variable *Y* may result from a 3rd or 5th order polynomial transformation with a cdf that

25

is *not* a one-to-one relationship between values of Y and $P(Y \leq y)$ for $y \in \mathbb{R}$. The following section elaborates on their argument and reflects on its usefulness for the applied psychometrician.

The Issue of the "Valid pdf". For this section, the notation used in Headrick & Kowalchuk (2007) will be adopted. The main difference is that the authors use $p(z)$ to designate the transformation function instead of $f(Z)$. Further, they shift the indices for the transformation coefficients by one unit so that our a_0 is the same as their a_1, our a_1 is their a_2, and so on. They then define the family of sets of distributions obtainable through polynomial transforms of order r such that

$$Y = p(z) = \sum_{i=1}^{r} a_i z^{i-1} \tag{24}$$

with r being positive and even: $r \in \{2, 4, 6, ...\}$. Equation (2) is a special case of Equation (24), letting $r = 4$:

$$Y = p(z) = \sum_{i=1}^{4} a_i z^{i-1} = a_1 + a_2 z + a_3 z^2 + a_4 z^3 \tag{25}$$

We begin with (the relevant parts of) Definition 3.1 on page 230 in Headrick and Kowalchuk (2007):

"Let Z be a random variable that has a standard normal distribution with pdf $f_Z(z)$, cdf $F_Z(z)$ and t^{th} moment μ_t. Let $z = (x, y)$ be the auxiliary variable that maps the parametric curves of $f_Z(z)$ and $F_Z(z)$ as

$$f : z \to \mathbb{R}^2 := f_Z(z) = f_Z(x, y) = f_Z(z, f(z)) \quad \text{''}$$
$$F : z \to \mathbb{R}^2 := F_Z(z) = F_Z(x, y) = F_Z(z, F(z)) \tag{26}$$

The definition quoted in Equation (26) describes $f_Z(z) = f_Z(z, f_Z(z))$ as a map of the values

of Z, $Z \sim N(0,1)$, to the pdf of Z.

Next, we consider a slightly modified version of the first part of Definition 3.2 on their page 231.

In particular, we will keep to the special case of $r = 4$:

> "Let $r = 4$, then A_4 is the set of distributions stemming from the transformation
> $p : p(z) \to A_4$ (i.e. the set of distributions that can be created by the Fleishman transform
> from Equation (2)). The transformation is expressed as $p(z) = \sum_{i=1}^{4} a_i z^{i-1}$.
>
> (i) p is said to be a strictly increasing polynomial function in z with (a) degree 3, (b)
> one and only one real root, (c) derivative $p'(z) > 0$, and (d) constant coefficients
> $a_i \in \mathbb{R}$, $c_{r-1} \neq 0$, and where $0 < a_1 + 3a_3 \leq 1$.
>
> (ii)   … (not important for our purposes here)"

Finally, we have proposition 3.1 on page 232:

> "If the compositions $f \circ p$ and $F \circ p$, based on Definitions 3.1 and 3.2, map the
> parametric curves of $f_{p(Z)}(p(z))$ and $F_{p(Z)}(p(z))$, where $p(z) = p(x, y)$ as
>
> $$f \circ p : p(z) \to \mathbb{R}^2 := f_{p(Z)}(p(z)) = f_{p(Z)}(p(x, y)) = f_{p(Z)}\left(p(z), \frac{f_Z(z)}{p'(z)} \right)$$
> $$F \circ p : p(z) \to \mathbb{R}^2 := F_{p(Z)}(p(z)) = F_{p(Z)}(p(x, y)) = F_{p(Z)}(p(z), F_Z(z)) \tag{27}$$
>
> then $f_{p(Z)}(p(z), f_Z(z)/p'(z))$ and $F_{p(Z)}(p(z), F_Z(z))$ are the pdf and cdf of the
> power method transformation $p(Z)$."

This means that, equivalently to how the maps between a standard normal variable Z and its pdf

and cdf were described in Definition 3.1, $f_{p(Z)}(p(z), f_Z(z)/p'(z))$ and $F_{p(Z)}(p(z), F_Z(z))$

map the nonnormal variable Y to its pdf and cdf, respectively. Headrick & Kowalchuk provide a

proof on page 232.

Combining definition 3.1, definition 3.2, and proposition 3.1 essentially states that *if* the derivative of the transformation function is never equal to zero, you can draw the pdf of *Y* by plotting the transformed values $p(z)$ against the ratio of the normal pdf and the derivative of the transformation, $f_Z(z)/p'(z)$. Using this technique, I was able to replicate the "invalid" pdf in Panel A of Headrick & Kowalchuk's Figure 1 (created for a 5th order polynomial RV) in Figure 10. The coefficients for the figures in Panel A are: $a_0 \approx 0$, $a_1 \approx 1.36$, $a_2 \approx 0$, $a_3 \approx -.148$, $a_4 \approx 0$, $a_5 \approx .00126$. The pdf is created by plotting the *Y* variable $Y = 1.36 \times Z - .148 \times Z^3 + .00126 \times Z^5$ on the x-axis against $\varphi(Z)/(1.36 - .444 \times Z^2 + .0063 \times Z^4)$ on the y-axis, where $\varphi(Z)$ is the pdf of the standard normal distribution and the denominator is the first derivative of the transformation function. The cdf is created by plotting the *Y* variable $Y = 1.36 \times Z - .148 \times Z^3 + .00126 \times Z^5$ on the x–axis against $\Phi(Z)$, where $\Phi(Z)$ is the cdf of the standard normal distribution.

28
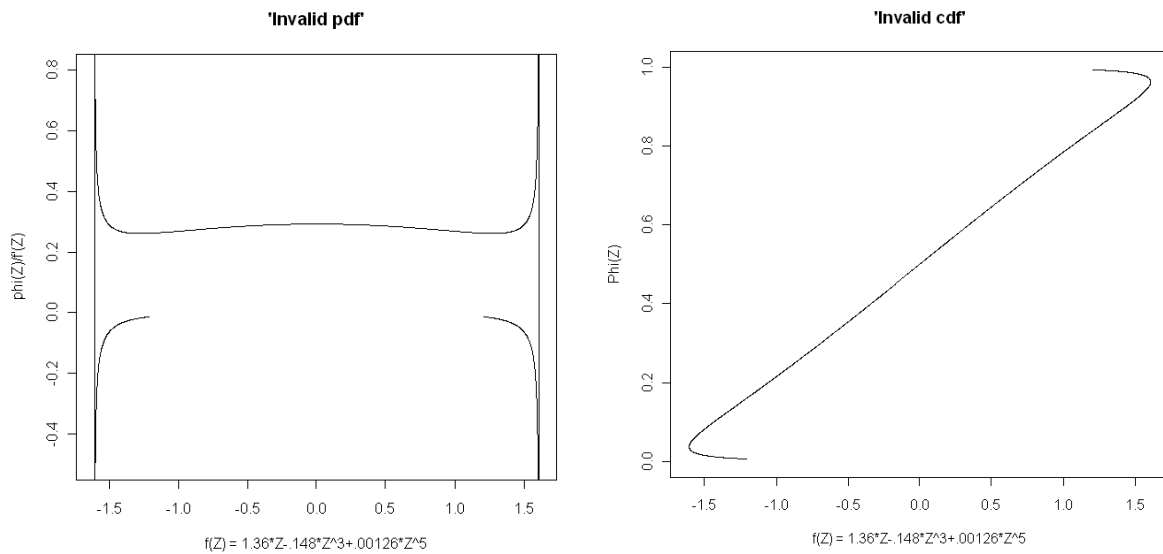
Figure 10 (a) & (b): "Invalid" pdf and cdf as shown in Headrick & Kowalchuk (2007), figure 1

Headrick & Kowalchuk have derived an analytic form of the pdf and cdf for variables *Y* created from monotonic polynomial transformations. With pdf and cdf available in analytical form, percentiles and mode, among other things, of a power method distribution can be found. However, they choose a slightly unfortunate terminology: They declare variables with monotonic transformations to have a "valid" pdf/cdf. Variables for which Equation (27) cannot be used to find the pdf or cdf are said to have an "invalid" pdf/cdf. This choice of terminology is unfortunate because it is safe to assume that variables produced with non-monotonic transformations will still have a pdf and cdf, only that Equation (27) cannot be used to find them. The pdf and cdf for variables that cannot use Equation (27) certainly exist, they are just *unknown*.

More importantly, Headrick & Kowalchuk have not identified why non-monotonic transformations whose pdf/cdf are not known need to be avoided when nonnormal random data generation is desired. Other than the necessity of a monotonic transformation in order to simulate

bi- or multivariate nonnormal data with a specified rank correlation (Headrick, 2004), the

practical importance of monotonic transformations remains unclear. It is also quite easy to find

an example of a power method distribution that does not have a "valid" pdf as defined by

Headrick & Kowalchuk (2007), yet whose pdf is well-known to even most undergraduate

students: Consider the nonnormal variable with $\gamma_1 = \sqrt{8}$ and $\gamma_2 = 12$. The sets of coefficients

that satisfy the set of equations in (7) for these values of skewness and kurtosis are:

$$
\begin{bmatrix} -.520676, .614598, .520676, .0200724 \end{bmatrix}
$$
$$
\begin{bmatrix} -.520676, -.614598, .520676, -.0200724 \end{bmatrix} \qquad (28)
$$
$$
\begin{bmatrix} -.707107, 0, .707107, 0 \end{bmatrix}
$$

None of these are monotonic:

Transformations for g1 = sqrt(8), g2 = 12
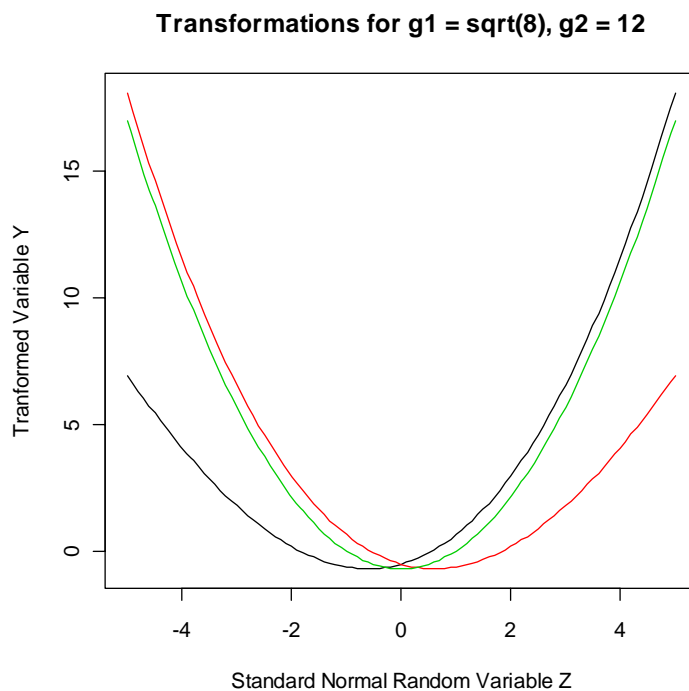
Figure 11: Non-monotonic Transformations to obtain a Chi-square Random Variable With One Degree of Freedom

Still, you may realize that the last transform is nothing other than a transformation of the form

$Y = a + bZ^2$ (where $Z \sim N(0,1)$). In other words, Y is a rescaled chi-square variable with one degree of freedom, and thus, it's pdf and cdf are certainly well known, even if they cannot be derived via Equation (27). I will demonstrate in the following sections that the choice between a monotonic and a non-monotonic transformation can still be an issue of very high practical importance. First, however, we need to introduce the multivariate extension of Fleishman's method.

The Multivariate Extension of the 3rd Order Polynomial Transform

Five years after Fleishman's publication of the univariate 3rd order polynomial transform, Vale & Maurelli (1983) extended the method to the multivariate case (Note that more recently, Headrick & Sawilowsky (1999) have suggested an alternate extension to Fleishman's power transformation). The multivariate extension of the 3rd order power method provides control over marginal skewness and kurtosis of all variables as well as over their correlation matrix. This ability to (relatively freely) manipulate marginal distribution parameters and the correlation matrix has been seen as an advantage over previously available techniques, and the method rose to considerable popularity, especially for use in Monte Carlo research in the social sciences.

A common and very tractable way of creating correlated variables is by post–multiplying a matrix of scores (in column vector format) with the Cholesky decomposition of the desired correlation matrix. Introducing some additional notation, let $\mathbf{z}$ be a $p \times 1$ random vector having a multivariate normal distribution with mean vector $\mathbf{0}$ and with covariance matrix $\mathbf{I}$. Let $\mathbf{L}$ be a Cholesky factor of $\mathbf{\Sigma}$, a $p \times p$ positive definite covariance matrix. That is, $\mathbf{L}$ is a unique lower-triangular matrix such that $\mathbf{LL}' = \mathbf{\Sigma}$. Then

$$\mathbf{z}^* = \mathbf{Lz} \qquad (29)$$

will have a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

Consequently, if the columns of an $n \times p$ sample data matrix $\mathbf{Z}$ represent n observations

from a $MVN(\mathbf{0}, \mathbf{I})$ distribution, then

$$\mathbf{Z}^* = \mathbf{ZL}^{'} \qquad (30)$$

will represent n observations from a $MVN(\mathbf{0}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\Sigma} = \mathbf{LL}^{'}$.

Sample Score Example. To illustrate this technique, I limit the following example to the

bivariate case. We want to create standard normal random scores $\mathbf{z}_1^*$ and $\mathbf{z}_2^*$ drawn from a

$MVN(\mathbf{0}, \boldsymbol{\Sigma})$ distribution, where the off-diagonal element of $\boldsymbol{\Sigma}$ is $\rho_{\mathbf{z}_1 \mathbf{z}_2} = .5$ (from here on, ρ_Z

will be used as a short for $\rho_{\mathbf{z}_1 \mathbf{z}_2}$). To achieve this, let $\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 \end{bmatrix}$ be an $n \times 2$ matrix of standard

normal random scores from a $MVN(\mathbf{0}, \mathbf{I})$ distribution. We now only need to post-multiply $\mathbf{Z}$ by

$\mathbf{L}^{'}$, where

$$\mathbf{L}^{'} = \text{chol} \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix} \approx \begin{bmatrix} 1 & .5 \\ 0 & .866 \end{bmatrix}, \qquad (31)$$

the transpose of Cholesky decomposition of the desired correlation matrix:

$$\begin{bmatrix} \mathbf{z}_1^* & \mathbf{z}_2^* \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 \end{bmatrix} \times \mathbf{L}^{'} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 \end{bmatrix} \times \begin{bmatrix} 1 & .5 \\ 0 & .866 \end{bmatrix}. \qquad (32)$$

$\begin{bmatrix} \mathbf{z}_1^* & \mathbf{z}_2^* \end{bmatrix}$ is now an $n \times 2$ matrix of *correlated* standard normal random scores; i.e., the

individual variables $\mathbf{z}_1^*$ and $\mathbf{z}_2^*$ still have a univariate standard normal distribution, but the

bivariate distribution exhibits a linear relationship between the two variables.

How can we create a set of *nonnormal* random scores $\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 \end{bmatrix}$ drawn from a bivariate nonnormal distribution with desired skewnesses and kurtoses and known correlation $\rho_{Y_1 Y_2}$, short ρ_Y, equal to .5? Can we, for example, just apply the same transformation of post-multiplying a vector $\begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 \end{bmatrix}$ with the Cholesky decomposition of the desired correlation matrix as in Equation (32)?

Although the above method is straightforward for creating multivariate normal variables with desired covariance matrix, it does *not* generalize directly to a method for creating multivariate nonnormal variates with desired covariance structure and specified skewness and kurtosis. To see why, consider the two possible approaches using the above method.

1. *Transform and Combine*. Using the power transform, create uncorrelated variates with desired skewness and kurtosis in $\mathbf{Y}$, then compute the linear combination $\mathbf{Y}^* = \mathbf{Y}\mathbf{L}'$. *Problem*. This approach will generate variates with the desired covariance matrix, but the process of linear recombination during the matrix multiplication will alter the skewness and kurtosis of all but the first variate.

2. *Combine and Transform*. Create multivariate normal variates with covariance matrix $\mathbf{\Sigma}$ as described above, then power transform the variates nonlinearly to have desired skewness and kurtosis. *Problem*. This approach will generate variables with the correct skewness and kurtosis, but the process of transforming nonlinearly will alter the correlations between the Y variables, so $\mathbf{\Sigma}_Y$ will no longer have the desired values.

To enable control over both the marginal parameters *and* the correlation matrix, a trick has to be employed: Assume we choose to first correlate normally distributed random variables as discussed above and then apply non-normalizing transforms. The influence of the non-

normalizing transformations on the correlations can be calculated and taken into account. The *final correlation* ρ_Y between the nonnormal variables can be expressed as a function of the *intermediate correlation* ρ_Z between the normal variables and the non-normalizing transforms (see Li & Hammond, 1975):

$$\rho_Y = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(Z_1) g(Z_2) f_{12} dZ_1 dZ_2 \tag{33}$$

where $f(Z_1)$ is the non-normalizing transform for the first variable, $g(Z_2)$ is the non-normalizing transform for the second variable, and

$$f_{12} = \frac{1}{2\pi\sqrt{1-2\rho^2{}_Z}} \exp\left(-\frac{Z_1{}^2 - 2\rho_Z Z_1 Z_2 + Z_2{}^2}{2(1-\rho^2{}_Z)}\right) \tag{34}$$

is the standard normal bivariate density.

A popular remedy therefore is to find an *intermediate correlation matrix* which to apply to standard normal random variables as in Equation (32) and then, in a second step, to subject the correlated standard normal random variables to the non-normalizing transformation. In this process the intermediate correlation matrix is chosen so that the final correlation matrix – the correlation matrix between the nonnormal random variables, after applying the non-normalizing transform – is as desired (Li & Hammond, 1975). This technique has been termed *TC*, transform and calculate, by Ruscio & Kaczetow (2008) and is the principle Vale & Maurelli used to extend the 3rd order polynomial transform to the multivariate case.

The first step in the Vale & Maurelli (1983) method is to find the Fleishman coefficients that are needed to create the nonnormal random variables with desired skewnesses and kurtoses. Secondly, one needs to find the required intermediate correlation matrix. As a special case of

Equation (33), the correlation ρ_Y between two nonnormal Fleishman variables Y_1 and Y_2 can be

expressed as (see equation (11) in Vale & Maurelli, 1983):

$$\rho_Y = \rho_Z \left(a_{11}a_{12} + 3a_{11}a_{32} + 3a_{31}a_{12} + 9a_{31}a_{32} \right) + 2a_{21}a_{22}\rho^2_Z + 6a_{31}a_{32}\rho^3_Z \qquad (35)$$

where ρ_Z is the intermediate correlation between two standard normal random variables, a_{11},

a_{21}, and a_{31} are the transformation coefficients for Y_1, and a_{12}, a_{22}, and a_{32} are the

transformation coefficients for Y_2. ρ_Z can therefore be found by numerically solving Equation

(35). Once ρ_Z is found, $\mathbf{\Sigma}^*$ is created as

$$\mathbf{\Sigma}^* = \mathrm{chol} \begin{bmatrix} 1 & \rho_Z \\ \rho_Z & 1 \end{bmatrix} \qquad (36)$$

In the next step, two standard normal random variables are correlated by using the intermediate

correlation matrix from Equation (36) in the process in Equation (32). Finally, the 3$^{\mathrm{rd}}$ order

polynomial transformations as in Equation (2) are applied individually to the now correlated

standard normal variables, using transformation coefficients a_{01}, a_{11}, a_{21}, a_{31}, a_{02}, a_{12}, a_{22},

and a_{32}.

An Example. Assume we want to create two nonnormal random variables Y_1 and Y_2 with

correlation ρ_Y. Each variable will have its own marginal skewness and kurtosis, which we

denote by γ_{11} and γ_{21} for the skewness and kurtosis of Y_1 and γ_{12} and γ_{22} for skewness and

kurtosis of Y_2. To demonstrate, we choose a condition used in Harwell & Serlin (1989) with

skewnesses $\gamma_{11} = \gamma_{12} = 0$, kurtoses $\gamma_{21} = \gamma_{22} = 25$, and final correlation $\rho_Y = .3$. The skewness-

kurtosis combination $\gamma_1 = 0$, $\gamma_2 = 25$ can be produced by two distinctly different sets of

coefficients for the 3$^{\mathrm{rd}}$ order power method, one resulting in a monotonic, the other in a non-

monotonic transformation. We have already plotted the two versions of the univariate distribution for this parameter combination (see Figure 8). What do the bivariate versions look like?

Assume we first choose the set of coefficients for the monotonic transformation for both Y_1 and Y_2: $a_{01} = a_{02} = 0$, $a_{11} = a_{12} = .2552828$, $a_{21} = a_{22} = 0$, $a_{31} = a_{32} = .2037548$. Inserting these coefficients into Equation (35), the relationship between final and intermediate correlation is now:

$$.3 = \rho_Y = 0.750904\rho_Z + 0.249096\rho^3{}_Z \tag{37}$$

We numerically solve for ρ_Z and obtain $\rho_Z = .381150$. Now we postmultiply two independent standard normal random variables by the Cholesky decomposition of a correlation matrix with off-diagonal element $\rho_Z = .381150$ and then apply the non-normalizing transform. The resulting bivariate nonnormal distribution with correlation $\rho_Y = .3$ is shown in Figure 12(a).

If both marginal variables are non-normalized with the non-monotonic transformation which has coefficients $a_0 = 0$, $a_1 = -1.566682$, $a_2 = 0$, and $a_3 = .3481727$, the intermediate correlation will be different, and have the value $\rho_Z = .579947$. The resulting distribution is shown in Figure 12(b).

Both marginals created with monotonic transform
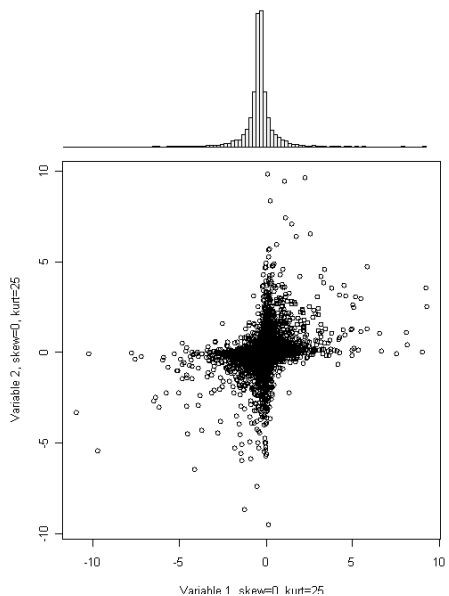
Both marginals created with non-monotonic transform
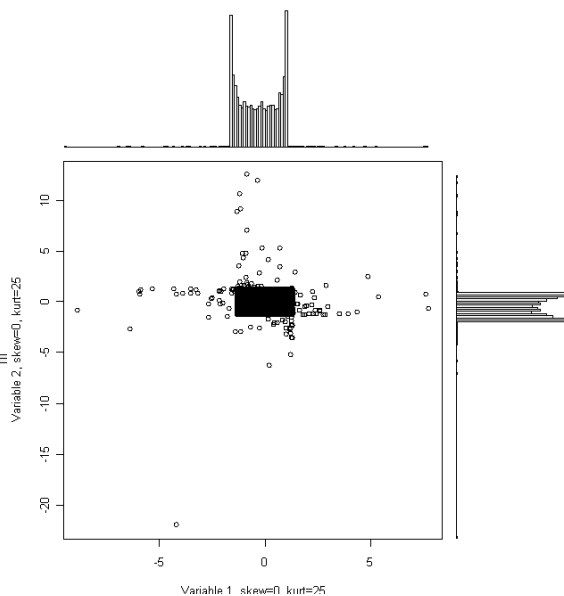
Figure 12 (a) & (b): V&M distributions for $\gamma_{11} = \gamma_{12} = 0$, $\gamma_{21} = \gamma_{22} = 25$, $\rho_Y = .3$, and $N = 5,000$.

Table 4: Sample Skewnesses, Kurtoses, and Correlations for Distributions in Figure 12

	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}_Y$
Both Marginals Transformed with Monotonic Transform	–0.0170	–0.0114	24.9536	25.1213	0.3002
Both Marginals Transformed with Non-monotonic Transform	0.0103	–0.0409	24.7616	23.3222	0.3006

These two distributions have essentially the same marginal means, variances, skewnesses, and kurtoses, as well as the same correlation, but very different shapes.

Identical Marginals vs. Non-Identical Marginals. At this point, I introduce new terminology that will help classify distributions for the remainder of my dissertation. In the

example above, Figure 12(a) & (b), both marginal distributions had the same skewnesses and

kurtoses, namely $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$. Further, because the marginal distributions are

constructed utilizing the same method and the same transformation coefficients, we can say that

the marginals are *identical*. Consequently, a distribution with such properties shall be called a

Distribution with Identical Marginals or *IM Distribution*. A distribution with $\gamma_{11} \neq \gamma_{12}$ or

$\gamma_{21} \neq \gamma_{22}$ or both will be called a *Distribution with Non-Identical Marginals* or *nIM Distribution*.

As we will see throughout the remainder of this dissertation, IM distributions and nIM

distributions differ in several ways, and that IM distributions have been simulated much more

often than nIM distributions.

Limitations of Vale & Maurelli's Multivariate Extension to the 3rd Order Polynomial Transform

Since Vale & Maurelli's method is directly based on applying the Fleishman 3rd order

transform to normally distributed (albeit correlated) random variables, limitations that have been

found for Fleishman's method also apply to their multivariate extension of the 3rd order

polynomial transform. Most prominently, each variable will be subjected to the limitations in the

range of skewness-kurtosis combinations discussed earlier and graphed in Figure 9. However,

multivariate distributions created by the Vale & Maurelli method, even if they are only bivariate,

are more complex by several orders of magnitude than any univariate distribution. This begins

with the 2nd order bivariate moment, represented by the correlation.

The Relationship Between Intermediate and Final Correlation. Li & Hammond (1975)

discuss for the general case that not every desired final correlation matrix has a valid

corresponding intermediate correlation matrix. Whether a valid intermediate correlation matrix

exists depends on the shape of the marginal distributions, i.e., the marginal skewness-kurtosis combinations. This in turn means that for certain choices of marginal skewness-kurtosis combinations for two variables, the final correlation ρ_Y cannot exceed some maximum permissible value $\rho_Y^* < 1$ (Carroll, 1961, provides a relatively interesting discussion). Also, the minimum value ρ_Y^{**} of the final correlation might exceed -1: $-1 < \rho_Y^{**}$. As we shall see, this limitation depends *both* on the selected values for skewnesses and kurtoses of the two variables *and* the subsequent choice of the particular set of transformation coefficients for Equation (2).

To explicate why this is so, assume we would like to simulate a bivariate nIM distribution with one normal and one fairly nonnormal marginal and desired final correlation ρ_Y. We have $\gamma_{11} = 0$ and $\gamma_{21} = 0$ for the normal distribution and choose $\gamma_{12} = 1$ and $\gamma_{22} = 15$ for the nonnormal distribution. The four sets of coefficients for the normal distribution are:

$$
\begin{aligned}
\text{set1} &= \begin{bmatrix} a_{01} = 0 & a_{11} = 1.49435 & a_{21} = 0 & a_{31} = -.214504 \end{bmatrix} \\
\text{set2} &= \begin{bmatrix} a_{01} = 0 & a_{11} = -1.49435 & a_{21} = 0 & a_{31} = .214504 \end{bmatrix} \\
\text{set3} &= \begin{bmatrix} a_{01} = 0 & a_{11} = 1 & a_{21} = 0 & a_{31} = 0 \end{bmatrix} \\
\text{set4} &= \begin{bmatrix} a_{01} = 0 & a_{11} = -1 & a_{21} = 0 & a_{31} = 0 \end{bmatrix}
\end{aligned}
\tag{38}
$$

and for the nonnormal distribution we have

$$
\begin{aligned}
\text{set1} &= \begin{bmatrix} a_{02} = -.170095 & a_{12} = 1.534711 & a_{22} = .170095 & a_{32} = -.306848 \end{bmatrix} \\
\text{set2} &= \begin{bmatrix} a_{02} = -.170095 & a_{12} = -1.534711 & a_{22} = .170095 & a_{32} = .306848 \end{bmatrix} \\
\text{set3} &= \begin{bmatrix} a_{02} = -.077327 & a_{12} = .445537 & a_{22} = .077327 & a_{32} = .157014 \end{bmatrix} \\
\text{set4} &= \begin{bmatrix} a_{02} = -.077327 & a_{12} = -.445537 & a_{22} = .077327 & a_{32} = -.157014 \end{bmatrix}
\end{aligned}
\tag{39}
$$

The choice of coefficients for the normal distribution is simple: $a_{11} = 1$ and $a_{01} = a_{21} = a_{31} = 0$; for the nonnormal distribution, let's just arbitrarily pick the very first set, $a_{02} = -.170095$,

$a_{12} = 1.534711$, $a_{22} = .170095$, and $a_{32} = -.306848$. We can now substitute the coefficients into

Equation (35), resulting in

$$\begin{aligned}
\rho_Y &= \rho_Z \left(a_{11}a_{12} + 3a_{11}a_{32} + 3a_{31}a_{12} + 9a_{31}a_{32} \right) + 2a_{21}a_{22}\rho^2_Z + 6a_{31}a_{32}\rho^3_Z \\
&= \rho_Z \left(a_{12} + 3a_{32} \right) \\
&= \rho_Z \left(1.534711 - 0.920544 \right) \\
&= 0.614167\rho_Z
\end{aligned} \tag{40}$$

Figure 13 depicts the relationship between final and intermediate correlation:

skew1=0, skew2=1, kurt1=0, kurt2=15
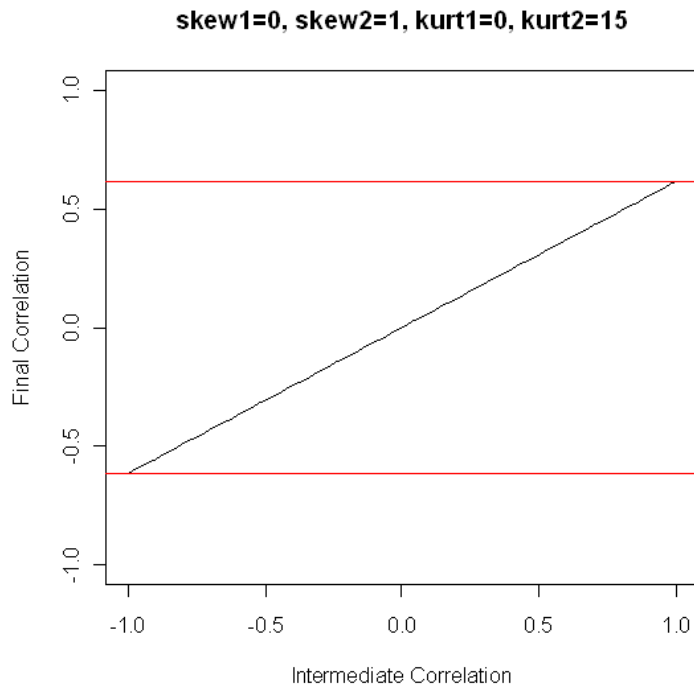
Figure 13: Relationship between Intermediate and Final Correlation for a Normal and a Nonnormal variable.

The *intermediate* correlation ρ_Z will assume values between –1 and 1, but inserting its whole

range $[-1, \quad 1]$ into Equation (40) will only produce corresponding values of the final correlation

ρ_Y between $-.614167$ and $.614167$. For this choice of marginal skewnesses and kurtoses and this set of coefficients, it is *impossible* to create a final correlation of, say, .8 between Y_1 and Y_2.

Is this true for all sets of coefficients in equations (38) and (39)? Had we chosen the same coefficients for the normal distribution but $a_{02} = -.0773268$, $a_{12} = .4455373$, $a_{22} = .0773268$, and $a_{32} = .1570144$ for the nonnormal distribution, the relationship between ρ_Y and ρ_Z would have been

$$\begin{aligned} \rho_Y &= \rho_Z \left(a_{12} + 3a_{32} \right) \\ &= \rho_Z \left(.4455373 + .4710432 \right) \\ &= .9165805 \rho_Z \end{aligned} \qquad (41)$$

The picture of the relationship between intermediate and final correlation is now this:
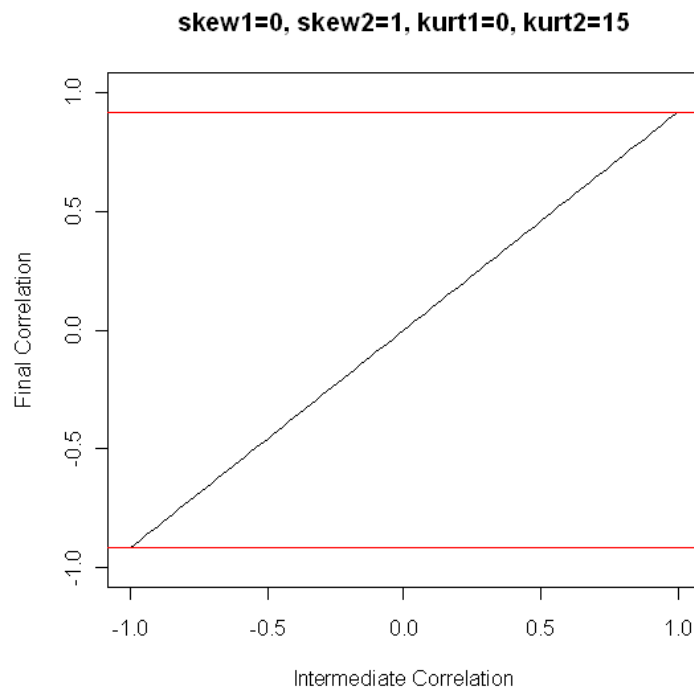
Figure 14: Relationship between Intermediate and Final Correlation for a Normal and a Nonnormal variable.

41

A much larger range of final correlations is available. By choosing $\rho_Z = .8728093$, we can produce a final correlation of $\rho_Y = .8$. However, the entire range of $[-1, \ 1]$ for ρ_Y still cannot be simulated. Obviously, the range of final correlations available to us depends both on the skewness-kurtosis combinations and the choice of transformation coefficients. To gain a better overview of the ranges of final correlations typically available, I examine three additional bivariate distributions, two of which have been popular choices in Monte Carlo studies (see Table 28).

Example 1. The first of these distributions has $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$. The transformation coefficients for this distribution have already been provided in (18). We choose

$$
\begin{aligned}
\text{Set1} &= \begin{bmatrix} 0 & 0.255283 & 0 & 0.203755 \end{bmatrix} \\
\text{Set2} &= \begin{bmatrix} 0 & 1.566682 & 0 & -0.348173 \end{bmatrix}
\end{aligned}
\tag{42}
$$

for the monotonic (Set1) and non-monotonic (Set2) transformation. This results in three distinctly different ranges of available final correlations (three because the combinations Set1–Set2 and Set2–Set1 will be the same). We get

Table 5: Range of Final Correlations, Example 1

	Min	Max
Set1 – Set1 ($monotonic_1$ & $monotonic_1$)	–1.000000	1.000000
Set1 – Set2 ($monotonic_1$ & $non-monotonic_2$)	–0.179564	0. 179564
Set2 – Set2 ($non-monotonic_2$ & $non-monotonic_2$)	–1.000000	1.000000

Both the monotonic/monotonic and the non-monotonic/non-monotonic combinations yield a range of $\begin{bmatrix} -1, & 1 \end{bmatrix}$ for ρ_Y, but the monotonic/non-monotonic combination does not; see Figure 15[2].
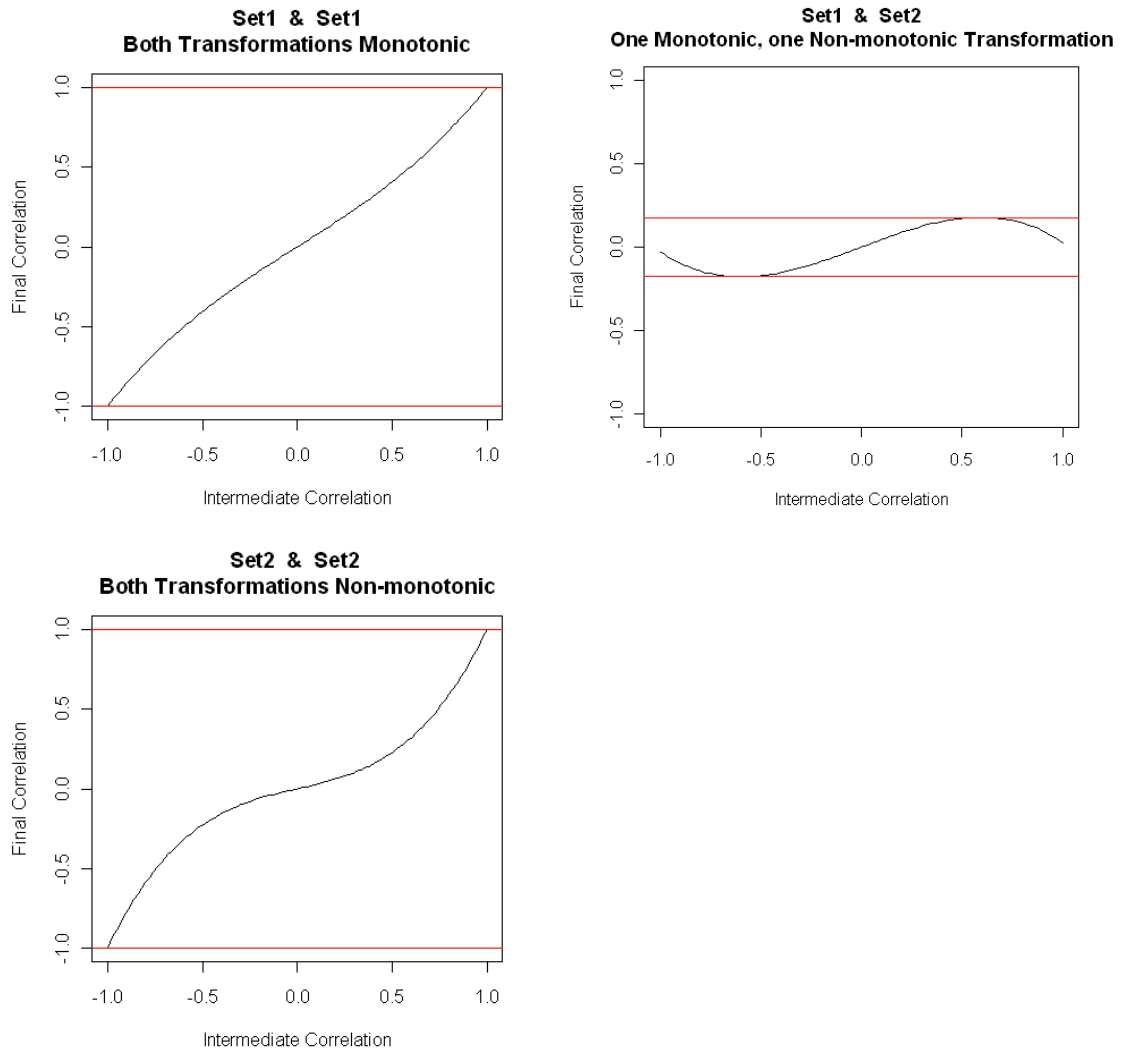
Figure 15: Relationship Between Intermediate and Final Correlation Depending on Choice of Transformation Coefficients

---

[2] Note that when more than two graphs are included in one figure, these will be numerated (a), (b), (c), etc. The order will always be

```
(a)    (b)
(c)    (d)
(e)    (f)
```

Example 2. The second distribution has $\gamma_{11} = \gamma_{12} = 2$ and $\gamma_{21} = \gamma_{22} = 6$. There are two distinctly different sets of coefficients, both resulting in a *non-monotonic* transformation (no monotonic transformation is available for $\gamma_1 = 2$ and $\gamma_2 = 6$):

$$
\begin{aligned}
\text{Set1} &= \begin{bmatrix} -0.313749 & 0.826324 & 0.313749 & 0.022707 \end{bmatrix} \\
\text{Set2} &= \begin{bmatrix} -0.569495 & -0.815741 & 0.569495 & 0.087793 \end{bmatrix}
\end{aligned}
\tag{43}
$$

The three unique ranges for the final correlation are:

Table 6: Range of Final Correlations, Example 2

	Min	Max
Set1 – Set1 ($non-monotonic_1$ & $non-monotonic_1$)	–.606249	1
Set1 – Set2 ($non-monotonic_1$ & $non-monotonic_2$)	–.167004	.839453
Set2 – Set2 ($non-monotonic_2$ & $non-monotonic_2$)	–.036503	.999999
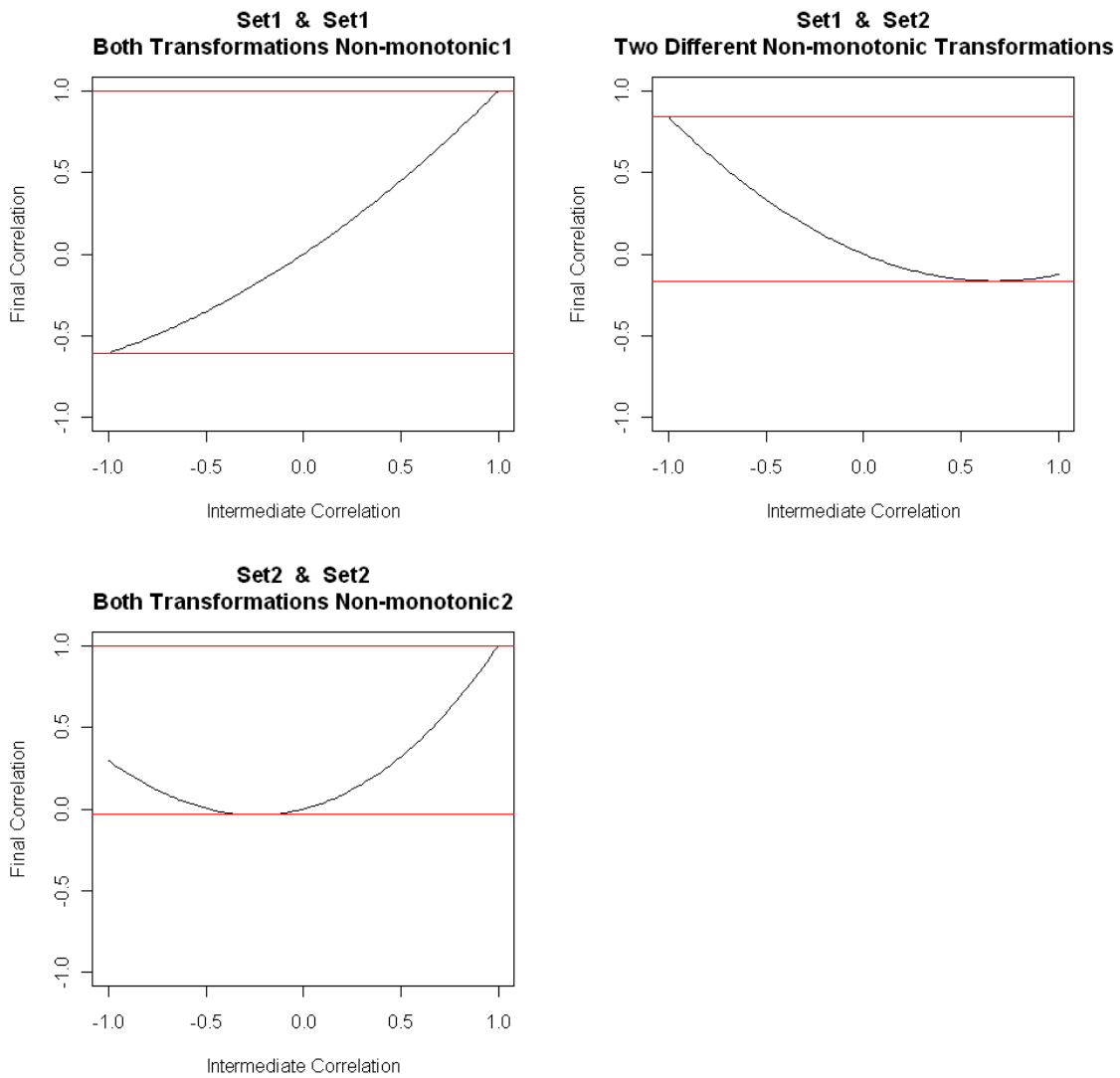
Figure 16: Relationship Between Intermediate and Final Correlation Depending on Choice of Transformation Coefficients

Even though both sets of coefficients lead to a non-monotonic transformation, the range of final correlations available is still dependent on which set of coefficient is chosen. Notice also that when combining Set1 or Set2 with Set2, the relationship between ρ_Z and ρ_Y is not one-to-one!

45

For example, if we chose Set2 for both marginals and want $\rho_Y = .2$, both $\rho_Z = -.877$ and

$\rho_Z = .365$ are a solution!

Example 3. The third example is an nIM distribution, with $\gamma_{11} = 1$ and $\gamma_{21} = 20$ for the

first variable and $\gamma_{12} = 2$ and $\gamma_{22} = 40$ for the second variable. The two distinctly different sets

of coefficients for the first variable are:

$$\begin{aligned} \text{Set1} &= \begin{bmatrix} -0.070726 & 0.348061 & 0.070726 & 0.180925 \end{bmatrix} \\ \text{Set2} &= \begin{bmatrix} -0.144537 & 1.544216 & 0.144537 & -0.325927 \end{bmatrix} \end{aligned} \tag{44}$$

The transform from Set1 is monotonic and the transform from Set2 is non-monotonic. For the

second variable, we have:

$$\begin{aligned} \text{Set3} &= \begin{bmatrix} -0.117429 & 0.061593 & 0.117429 & 0.242096 \end{bmatrix} \\ \text{Set4} &= \begin{bmatrix} -0.205819 & -1.465874 & 0.205819 & 0.354146 \end{bmatrix} \end{aligned} \tag{45}$$

Again, the first transform is monotonic and the second is non-monotonic. We can combine both

Set1 and Set2 with each Set3 and Set4, resulting in four ranges of possible final correlations:

Table 7: Range of Final Correlations, Example 3

	Min	Max
Set1 – Set3 ($monotonic_1$ & $monotonic_3$)	–.9480722	.9812932
Set1 – Set4 ($monotonic_1$ & $non-monotonic_4$)	–.1250746	.1432437
Set2 – Set3 ($non-monotonic_2$ & $monotonic_3$)	–.1565529	.1779114
Set2 – Set4 ($non-monotonic_2$ & $non-monotonic_4$)	–.8615769	.9805707

The largest possible range for ρ_Y is reached when the two monotonic transforms are coupled.

Coupling the two non-monotonic transforms results in a fairly large range for ρ_Y as well, but a

monotonic transformation coupled with a non-monotonic transformation results in a very small

range of possible final correlations. Graphing the range of available final correlations for the two

monotonic transformations (Set1 with Set3), we have:
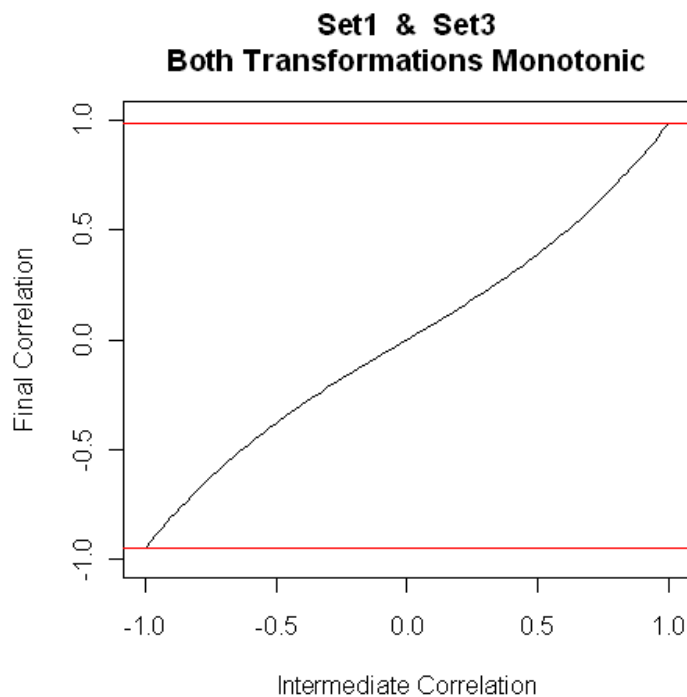
Figure 17: Relationship Between Intermediate and Final Correlation

For now, it largely remains an open question how marginal skewnesses and kurtoses as

well as transformation types (monotonic vs. non-monotonic) interact to produce a range of

available final correlations. It may be that two sets of coefficients that are relatively similar to

each other will yield a wider range for ρ_Y than two sets of coefficients that are fairly different

47

from each other. Even if we do not have a final answer on how to choose from different transformations available, it is incontrovertible that choice of transformation matters.

Mirror Image Transformations and the Relationship Between Intermediate and Final Correlation. Out of the four solutions to the set of equations in (7), only two are distinctly different, while the others are mirror images of first two. Whether we choose a set of coefficients or its mirror image has no effect on the range of final correlations. Consider the very first example from Figure 13 and Figure 14, for which we used the third set of coefficients from Equation (38) and the first set of coefficients from Equation (39). If we had used the second set of coefficients for the nonnormal variable from Equation (39), which is the mirror image of the first,

$$
\begin{aligned}
\text{set1} &= \left[-.170095, 1.53471, .170095, -.306848\right] \\
\text{set2} &= \left[-.170095, -1.53471, .170095, .306848\right]
\end{aligned}
\tag{46}
$$

while the set of coefficients for the normal variable remained the same ($a_0 = 0$, $a_1 = 1$, $a_2 = 0$, and $a_3 = 0$), Equation (40) would change to

$$
\begin{aligned}
\rho_Y &= \rho_Z \left(a_{11}a_{12} + 3a_{11}a_{32} + 3a_{31}a_{12} + 9a_{31}a_{32}\right) + 2a_{21}a_{22}\rho^2{}_Z + 6a_{31}a_{32}\rho^3{}_Z \\
&= \rho_Z \left(a_{12} + 3a_{32}\right) \\
&= \rho_Z \left(-1.53471 + .920545\right) \\
&= -.614166\rho_Z
\end{aligned}
\tag{47}
$$

The sign of the multiplicative factor in front of ρ_Z is reversed, but its absolute value is still the same. Since ρ_Z ranges from –1 to 1, this means that the range of ρ_Y is not changing either. The only aspect changing is that now $\rho_Z = -1$ maximizes ρ_Y and $\rho_Z = 1$ minimizes ρ_Y.

This is true for the general case. Consider Equation (35), the formula for ρ_Y, and slightly rearrange to

$$\rho_Y = g(\rho_Z) = a\rho_Z^3 + b\rho_Z^2 + c\rho_Z \tag{48}$$

with $a = 6a_{31}a_{32}$, $b = 2a_{21}a_{22}$, and $c = a_{11}a_{12} + 3a_{11}a_{32} + 3a_{31}a_{12} + 9a_{31}a_{32}$, with maximum

$$\rho_Y^* = a\left(\rho_Z^*\right)^3 + b\left(\rho_Z^*\right)^2 + c\rho_Z^*$$

and minimum

$$\rho_Y^{**} = a\left(\rho_Z^{**}\right)^3 + b\left(\rho_Z^{**}\right)^2 + c\rho_Z^{**}$$

where ρ_Z^* and ρ_Z^{**} are the values of the intermediate correlation that maximize and minimize ρ_Y, respectively. The coefficient of the quadratic term, b, will never change because the sign of both a_{21} and a_{22} stays the same. If the sign for the second and fourth coefficient for *both* sets are reversed (i.e. we change to the mirror image transformation for both variables), the changed signs will cancel out for both a and c. The only interesting case is a sign reversal for one of the variables, e.g. $a_{11}^* = -a_{11}$ and $a_{31}^* = -a_{31}$ thereby reversing the sign of a and b. Now the new equation for ρ_Y will be:

$$\rho_Y = h(\rho_Z) = -a\rho_Z^3 + b\rho_Z^2 - c\rho_Z \tag{49}$$

Notice that

$$
\begin{aligned}
h(-\rho_Z) &= -a(-\rho_Z)^3 + b(-\rho_Z)^2 - c(-\rho_Z) \\
&= a\rho_Z^3 + b\rho_Z^2 + c\rho_Z \\
&= g(\rho_Z)
\end{aligned}
\tag{50}
$$

Therefore, $h(\rho_Z) = -a\rho^3_Z + b\rho^2_Z - c\rho_Z$ is maximized by $-\rho_Z^*$ and minimized by $-\rho_Z^{**}$,

resulting in the same maximum ρ_Y^* and minimum ρ_Y^{**} as $g(\rho_Z)$. We can conclude that the

range for the final correlation ρ_Y is only influenced by the marginal skewness and kurtosis as

well as the choice between distinctly different sets of transformation coefficients for each

variable.

Odd–shaped Distributions

There is another and potentially even more important property of bi- and multivariate

distributions beyond marginal skewnesses, kurtoses, and correlation that has not been discussed

yet: Shape. Under "shape", we summarize *all* additional moments for bivariate or multivariate

distributions. We have to make ourselves aware that even just bivariate distributions are

substantially more complex than univariate ones. For example, bivariate distributions have a total

of two 1st order moments (the two means), three 2nd order moments (two variances and one

covariance) and *four* 3rd order moments:

- $\rho_{111} = \gamma_{11}$, skewness of variable 1

- ρ_{112}

- ρ_{122}

- $\rho_{222} = \gamma_{12}$, skewness of variable 2

where

$$\rho_{ijk} = \frac{E\left[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)\right]}{\sqrt{\sigma_{ii}\sigma_{jj}\sigma_{kk}}} \tag{51}$$

and

$$\sigma_{ii} = E\left[\left(X_i - \mu_i\right)^2\right] \tag{52}$$

Note that any permutations of the bivariate 3rd order moments are irrelevant; ρ_{112} is the same as ρ_{121} etc.

Likewise, there are five 4th order moments:

- $\rho_{1111} = \gamma_{21} + 3$

- ρ_{1112}

- ρ_{1122}

- ρ_{1222}

- $\rho_{2222} = \gamma_{22} + 3$

Equivalently to Equation (51),

$$\rho_{ijkh} = \frac{E\left[\left(X_i - \mu_i\right)\left(X_j - \mu_j\right)\left(X_k - \mu_k\right)\left(X_h - \mu_h\right)\right]}{\sqrt{\sigma_{ii}\sigma_{jj}\sigma_{kk}\sigma_{hh}}} \tag{53}$$

two of the 3rd order moments and three of the 4th order moments of a bivariate distribution are not controlled by the 3rd order power method. For higher moments, this figure will increase even more, e.g. for 5th order moments, we will have two marginal moments and four "mixed" moments.

For any bivariate distribution, there are

$$\frac{M^2 + 3M}{2} \tag{54}$$

moments up to M^{th} order. For example, if we are interested in the first $M = 4$ moments, we have $\sum_{i=1}^{4}(i+1) = (2+3+4+5) = 14$ moments in total. The V&M method provides control over

51

nine of these moments: Univariate means, variances, skewnesses, kurtoses, and the bivariate covariance, while leaving the other five moments uncontrolled. Of course, any other moments of even higher order are altogether uncontrolled as well.

What does this mean in practice? Figure 12 already shows that the shape of a bivariate V&M distribution depends not only on marginal skewnesses and kurtoses and the correlation, but also on the choice of coefficients used for the transformation. Below we will see another example of the striking impact the choice of a set of transformation coefficients can have. We create a bivariate nIM distribution with $\gamma_{11} = 2.5$ and $\gamma_{21} = 11.5$ for the first variable, $\gamma_{12} = 1.4$ and $\gamma_{22} = 5.6$ for the second variable, and $\rho_Y = .46$. For the first variable, the distinctly different coefficient sets are:

$$
\begin{aligned}
\text{Set1} &= \begin{bmatrix} -0.280057 & 0.640671 & 0.280057 & 0.084618 \end{bmatrix} & \text{(monotonic)} \\
\text{Set2} &= \begin{bmatrix} -0.646612 & -0.619375 & 0.646612 & 0.097626 \end{bmatrix} & \text{(non-monotonic)}
\end{aligned}
\tag{55}
$$

For the second variable, we have

$$
\begin{aligned}
\text{Set3} &= \begin{bmatrix} -0.159417 & 0.740037 & 0.159417 & 0.072615 \end{bmatrix} & \text{(monotonic)} \\
\text{Set4} &= \begin{bmatrix} -0.428141 & 1.198410 & 0.428141 & -0.177011 \end{bmatrix} & \text{(non-monotonic)}
\end{aligned}
\tag{56}
$$
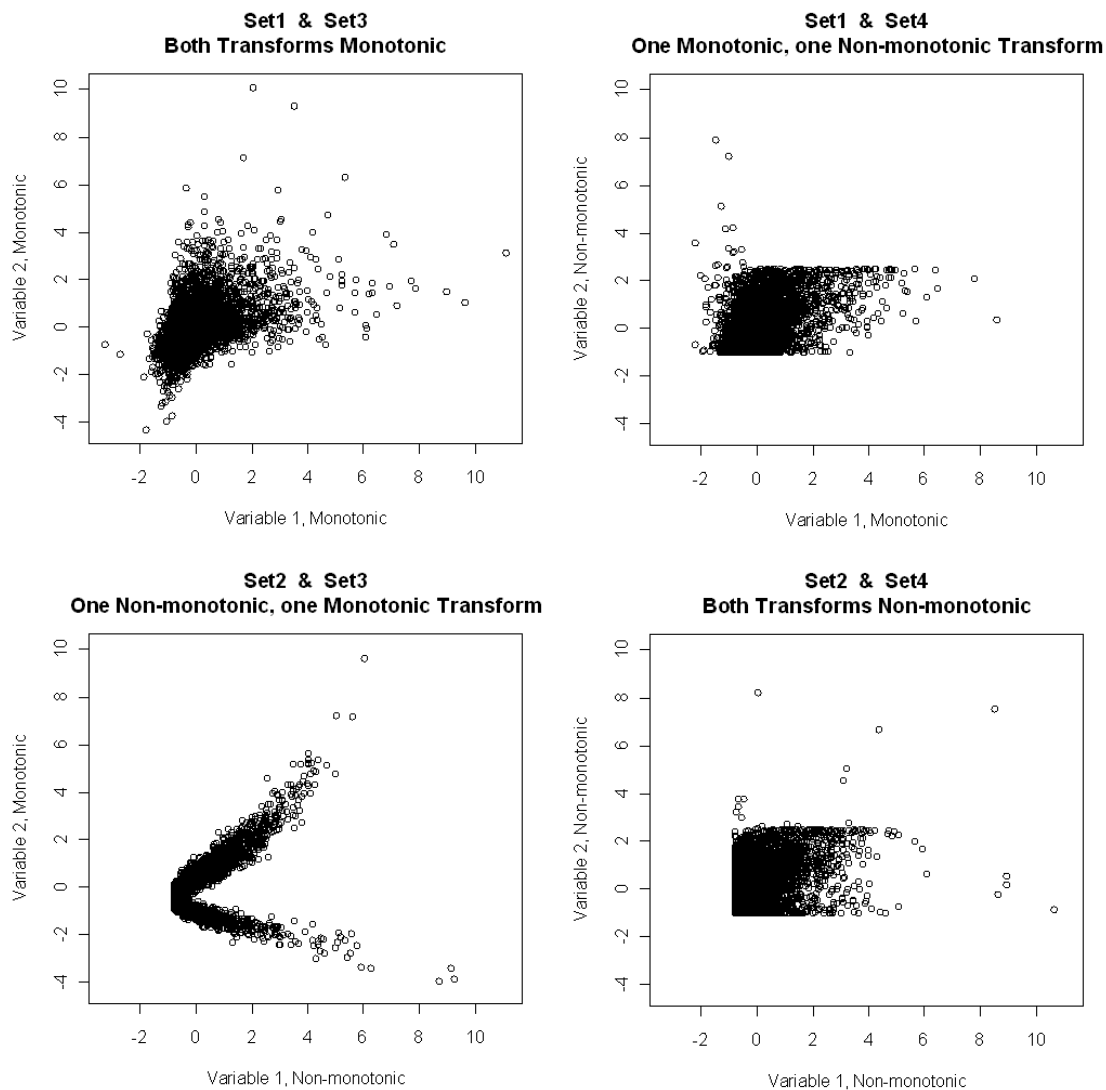
Figure 18 (a) through (d): Different Shapes Depending on Choice of Transformation Coefficients

Table 8: Skewnesses, Kurtoses, and Correlations for Distributions in Figure 18

Nominal Value	2.5	1.4	11.5	5.6	.46
Sample Estimate	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}_Y$
Figure 18(a), set 1 – set 3	2.4819	1.3901	11.2333	5.5354	.4597
Figure 18(b), set 1 – set 4	2.4979	1.4170	11.4382	5.9682	.4602
Figure 18(c), set 2 – set 3	2.4866	1.4064	11.2432	5.6918	.4612
Figure 18(d), set 2 – set 4	2.4978	1.3921	11.3700	5.4030	.4597

The shapes in Figure 18 are strikingly different from one another, despite all having the same

marginal skewnesses and kurtoses and the same correlation *and* all being produced with what has

been portrayed in the literature as a single unique method for simulating nonnormal distributions.

Notice that the only distribution with a completely "well-behaved" shape is the one constructed

from two monotonic transforms in Figure 18(a). The distribution created from the two non-

monotonic transforms (Figure 18(d)) comes quite close to the box-shaped form we have already

observed for the distribution in Figure 12(b). The distribution in Figure 18(c) almost looks like a

mixture of two relatively normal distributions with high correlations in opposite directions.

Interestingly, $\rho_Y = .46$ is very close to the maximum of available final correlations for the set of

parameters that creates the distribution in Figure 18(c).

First Conclusions on the 3rd Order Polynomial Transform. At this point, no

comprehensive summary of the findings on the univariate and multivariate 3rd order power

method, presented in the previous sections, exists. Univariate and bivariate shape as well as

range of final correlation available are aspects of V&M distributions that have not undergone any

formal examination. However, the above observations on the power method have a more general

implication for other bi- or multivariate distributions as well: Marginal skewnesses and kurtoses

for nonnormal variables Y_1 and Y_2 and the correlation between them are far from sufficient to

characterize a bivariate nonnormal distribution. Often, researchers employing the Fleishman or

the V&M method to simulate nonnormal distributions report no more than marginal skewnesses

and kurtoses, and it is very likely that few are aware of the drastic differences that can occur

between distributions even if these parameters have been fixed. The end of Part I of my

dissertation I will demonstrate that such differences in shape can have quite direct and drastic

consequences for studies in which nonnormal distributions are simulated. In other words, moments which are not controlled may have a major impact on inferences drawn from a Monte Carlo study. However, before we discuss the practical consequences of the results presented above, I introde two additional ways of simulating multivariate nonnormal scores, beginning with Headrick's (2002) extension of the 3rd order to a 5th order polynomial method.

The 5th Order Polynomial Transform

In 2002, Headrick suggested expanding Fleishman's 3rd order polynomial transform to a 5th order polynomial transform by simply adding two terms to Fleishman's transformation from Equation (2):

$$Y_H = a_0 + a_1 Z + a_2 Z^2 + a_3 Z^3 + a_4 Z^4 + a_5 Z^5 \tag{57}$$

Subsequently, he equated not only mean, variance, skewness, and kurtosis, but also the 5th and 6th moment (designated as γ_3 and γ_4 from here on) of the nonnormal variable Y_H with functions of the six coefficients a_0 through a_5, similar to the set of equations in (7) for the 3rd order power method:

$$
\begin{aligned}
\mu &= f_1(a_0, a_1, a_2, a_3, a_4, a_5) = 0 \\
\sigma^2 &= f_2(a_0, a_1, a_2, a_3, a_4, a_5) = 1 \\
\gamma_1 &= f_3(a_0, a_1, a_2, a_3, a_4, a_5) \\
\gamma_2 &= f_4(a_0, a_1, a_2, a_3, a_4, a_5) \\
\gamma_3 &= f_5(a_0, a_1, a_2, a_3, a_4, a_5) \\
\gamma_4 &= f_6(a_0, a_1, a_2, a_3, a_4, a_5)
\end{aligned}
\tag{58}
$$

This allows for control over the first six instead of the first four univariate moments. Detailed equations for the moments of Y_H as well as the final correlation ρ_Y are not included here, as they cover several pages. To get an impression of their size and character, the reader may

refer to Headrick (2002). To illustrate, consider that the largest of these functions,

$\gamma_4 = f_6\left(a_0, a_1, a_2, a_3, a_4, a_5\right)$, has about 112 terms. Some of the longer of these 112 terms are as

large as $27 \times 977816385000 c_4{}^2 c_5{}^4$. I will discuss possible implications of such large terms in the

following sections.

Headrick & Kowalchuk (2007) discuss additional properties of the 5th order power

method such as the pdf and cdf for subsets of the 3rd order and 5th order transform, an issue

which has already been examined in the section on the "valid pdf and cdf" above. Headrick,

Sheng, & Hodis (2007) provide *Mathematica* code to simulate desired distributions using the 5th

order polynomial transform. Headrick argues strongly that a need for control over additional

moments exists as this will improve the approximation of known univariate nonnormal

distributions such as the exponential or the uniform distribution (Headrick, 2002; Headrick,

2004).

An Example. We demonstrate the 5th order polynomial transform with two sets of

transformation coefficients from Table 1 in Headrick (2002) which are used to approximate

known nonnormal distributions. The first of these distributions is the logistic distribution (with

$\gamma_1 = 0$, $\gamma_2 = 6/5$, $\gamma_3 = 0$, and $\gamma_4 = 48/7$) and coefficients

$$\text{Set1} = \begin{bmatrix} 0.0000 & 0.8795 & 0.00005 & 0.0408 & 0.0000 & -0.0004 \end{bmatrix} \qquad (59)$$

The second is the Weibull distribution with parameters $\alpha = 6$ and $\beta = 10$ (with $\gamma_1 = -.373262$,

$\gamma_2 = .035455$, $\gamma_3 = .447065$, and $\gamma_4 = -1.022066$) and coefficients (Set2*b* is discussed below):

$$\text{Set2}a = \begin{bmatrix} 0.0655 & 0.9692 & -0.0652 & 0.0278 & -0.0001 & -0.0039 \end{bmatrix} \qquad (60)$$

The respective transformations and their resulting empirical distributions (along with the
theoretical densities of the logistic and the Weibull distribution plotted in red) are shown in
Figure 19:
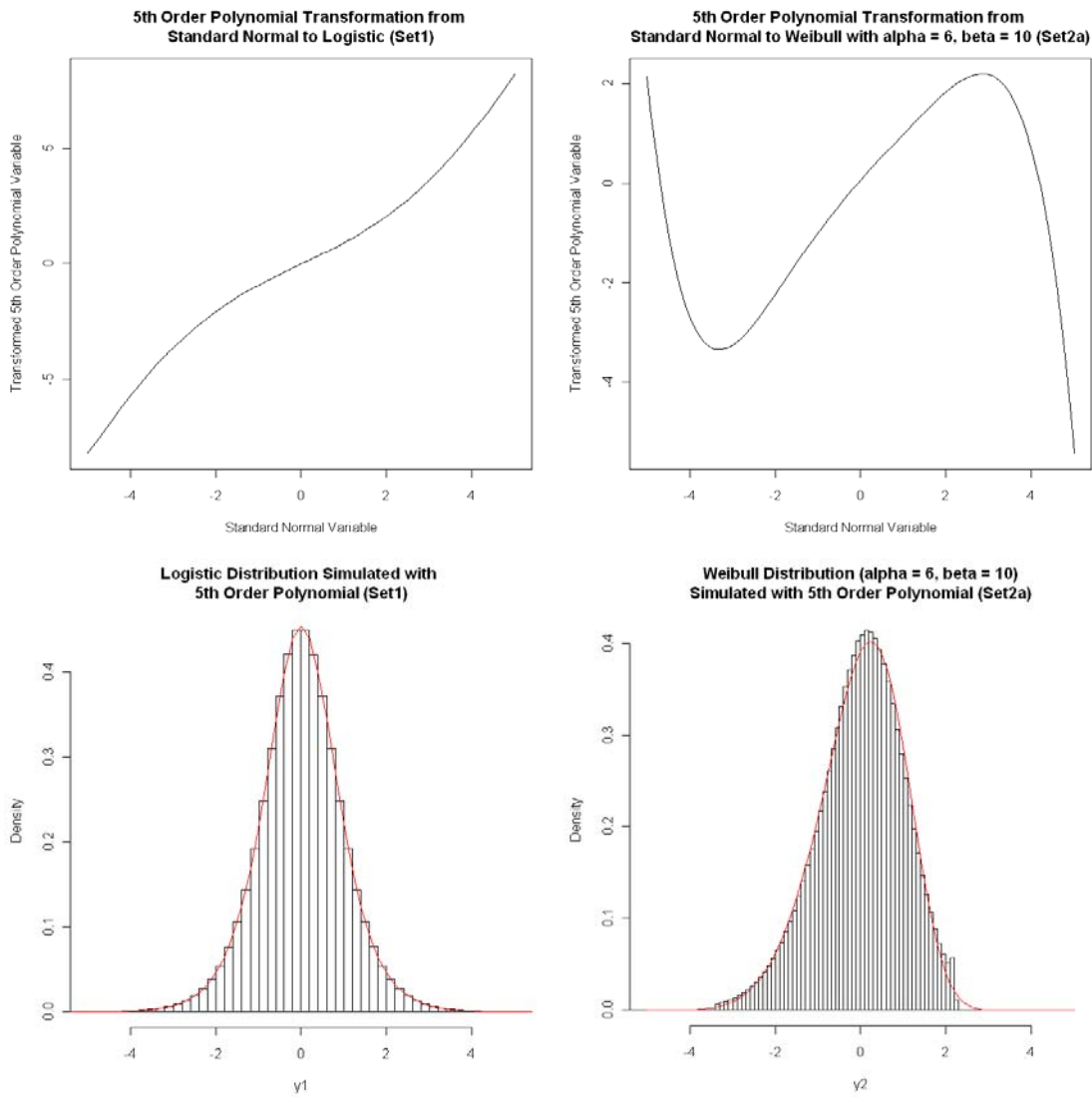
Figure 19: 5th Order Polynomial Transforms and Resulting Empirical Distributions

You may notice that the fit between theoretical and empirical distribution (as created by the 5th order power method) for the Weibull distribution is suboptimal. We will examine this issue in greater depth in the next section.

Limitations of the 5th Order Polynomial Transform

    Number of Solutions for Set of Coefficients. The 5th order polynomial transformation (or power method) is not free from the problems we identified for the 3rd order power method. Just as the 3rd order power method, the 5th order power method has multiple real-valued solutions to the set of equations that solves for the coefficients a_0 through a_5 (the *possibility* of multiple solutions is also briefly mentioned in a subordinate clause in Headrick & Kowalchuk, 2007, page 244). When solving the system of equations in (7) for the 3rd order power method, we saw that a total of four solutions, two of them distinctly different and sometimes leading to distributions with very different shapes, were found by *Mathematica* and R. It turns out that the 5th order polynomial can have even more distinctly different solutions to the set of equations referred to in (58).

    As an example, consider a scaled χ_1^2 distribution with first six moments $\mu = 0$, $\sigma = 1$, $\gamma_1 = \sqrt{8}$, $\gamma_2 = 12$, $\gamma_3 = 48\sqrt{2}$, and $\gamma_4 = 480$. Both *Mathematica* and R encountered difficulties when attempting to find *all* solutions for the six coefficients a_0 through a_5. Running for approximately 24 hours, *Mathematica* found six real-valued sets of coefficients (and in addition, a very large amount of sets of coefficients with at least some coefficients imaginary):

$$\text{Set1} = \begin{bmatrix} -0.6689 & 0.3194 & 0.6671 & -0.0007 & 0.0006 & -0.0001 \end{bmatrix}$$
$$\text{Set2} = \begin{bmatrix} -0.6689 & -0.3194 & 0.6671 & 0.0007 & 0.0006 & 0.0001 \end{bmatrix}$$
$$\text{Set3} = \begin{bmatrix} -0.3977 & 0.6211 & 0.4169 & 0.0684 & -0.0064 & 0.0000 \end{bmatrix}$$
$$\text{Set4} = \begin{bmatrix} -0.3977 & -0.6211 & 0.4169 & -0.0684 & -0.0064 & 0.0000 \end{bmatrix} \quad (61)$$
$$\text{Set5} = \begin{bmatrix} -0.7071 & -0.0325 & 0.7071 & -0.0335 & 0.0000 & -0.0001 \end{bmatrix}$$
$$\text{Set6} = \begin{bmatrix} -0.7071 & -0.0097 & 0.7071 & -0.0105 & 0.0000 & -0.0005 \end{bmatrix}$$

When either used to simulate data (see Equation (57)) or when substituted into equations for the first six moments of a 5th order polynomial variable, only sets 1, 2, and 6 produced distributions with values of $\hat{\gamma}_1$, $\hat{\gamma}_2$, $\hat{\gamma}_3$, and $\hat{\gamma}_4$ reasonably close to their nominal values. To investigate this further, I manipulated the existing root finding routine *nleqslv* in R to find as many different sets of coefficients as possible. R produced almost identical six sets of coefficients, all performing as expected (either by inserting into formulas for the first six moments or by simulating data):

$$\text{Set1} = \begin{bmatrix} -0.3977 & 0.6211 & 0.4169 & 0.0684 & -0.0064 & 0.00004 \end{bmatrix}$$
$$\text{Set2} = \begin{bmatrix} -0.3977 & -0.6211 & 0.4169 & -0.0684 & -0.0064 & 0.00004 \end{bmatrix}$$
$$\text{Set3} = \begin{bmatrix} -0.6689 & -0.3194 & 0.6671 & 0.0007 & 0.0006 & 0.00005 \end{bmatrix}$$
$$\text{Set4} = \begin{bmatrix} -0.6689 & 0.3194 & 0.6671 & -0.0007 & 0.0006 & -0.00005 \end{bmatrix} \quad (62)$$
$$\text{Set5} = \begin{bmatrix} -0.7071 & 0.0039 & 0.7071 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}$$
$$\text{Set6} = \begin{bmatrix} -0.7071 & 0.0335 & 0.7067 & 0.0001 & 0.0000 & 0.0000 \end{bmatrix}$$

Notice that Set1 and Set2 and Set3 and Set4 are, just as for the 3rd order polynomial, mirror images of each other. Further, Set5 and Set6 look quite similar, with a_0 close to $\sqrt{.5}$, a_2 close to $-\sqrt{.5}$, and all other coefficients close to 0. All of these solutions may very well be just an approximation to the exact solution $a_0 = -\sqrt{.5}$, $a_1 = 0$, $a_2 = \sqrt{.5}$, and $a_3 = a_4 = a_5 = 0$, which is the same transformation to a scaled χ_1^2 variable we already saw for the 3rd order power method. Both R and *Mathematica* had difficulties solving numerically for these solutions. Possibly, by increasing the complexity of the set of equations in (58), which relates the first six moments of

59

the *Y* variable to the six coefficients a_0 through a_5, and introducing numerically very large

terms, we may have met some limits of numerical root finding. This is somewhat contrary to the

claim by Headrick and co-authors in several papers that the 5th order polynomial is

computationally efficient and "simple."

Figure 20 plots the transformations for the sets of coefficients in Equation (62). All of

these transformations are non-monotonic – including transformation 1 and 2 (as can be verified

by plotting a smaller section on the interval $[-3, 0]$).
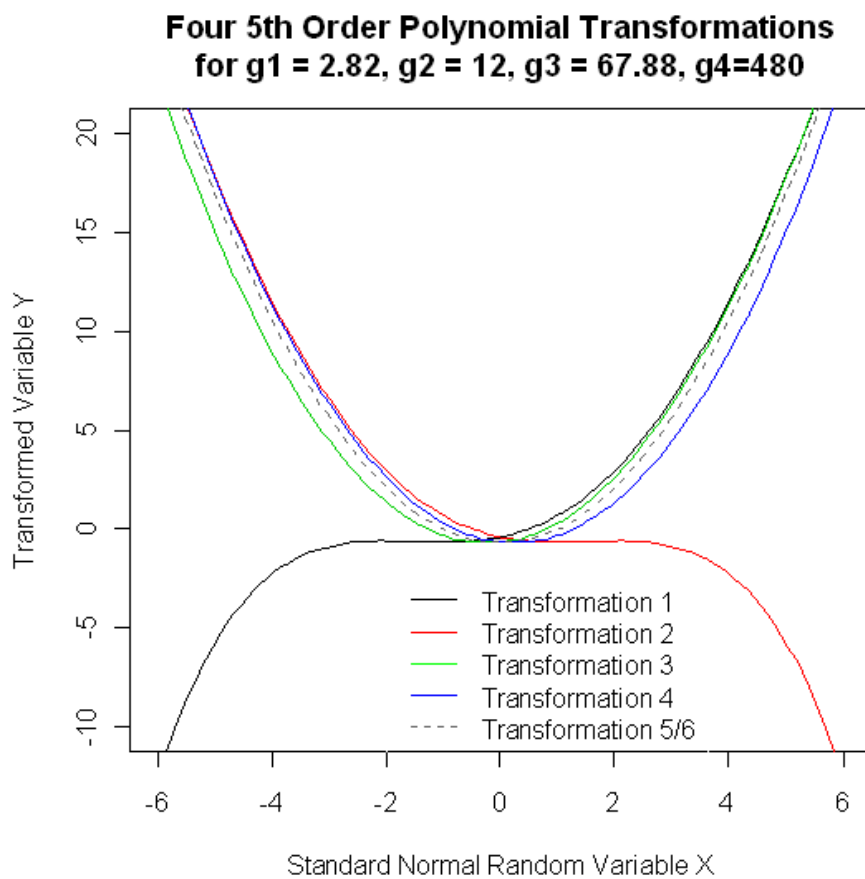
Figure 20: 5th Order Polynomial Transformations

In Figure 20, transformations 3, 4, and 5/6 are quite similar to each other, and it may seem as if the differences between these transforms should not be particularly important. However, the behavior of multivariate distributions with as many as $k = 10$ or more variables created with the 5th order polynomial method might become unpredictable, just as behavior has been shown to be somewhat unpredictable for the 3rd order polynomial transform and two dimensions. Again, requiring researchers to specify the coefficients used to create nonnormal distributions may be necessary to facilitate the replication of results through reviewers as well as consumers of the published research. For example, Beasley, DeShea, Toothaker, Mendoza, Bard, and Rodgers (2007) make use of the 5th order polynomial transformation to create an approximation to the χ^2_1 random variable. It is unknown which set of parameters was used and if a different set of parameters would – possibly – have resulted in different performance of the confidence intervals studied. It is not also clear whether Headrick et al.'s (2007) *Mathematica* program always chooses monotonic transformations when available.

When solving for the sets of transformation coefficients for a range of additional distributions, I found between two (for most of the distributions in Headrick's (2002) table 1) and four (for the Pareto distribution) distinctly different sets. Whether these were all available sets of coefficients or whether the number of coefficient sets will vary even more for other skewness-kurtosis combinations is, as of yet, uncertain, due to the computational difficulties I and possibly Headrick and others have encountered. Headrick & Kowalchuk (2007) do not discuss this issue and Headrick, Sheng, & Hodis (2007) only warn potential users of their *Mathematica* program to not alter starting values for the root finding routine that solves for the six coefficients.

It is also unclear whether Headrick is aware of the total number of existing solutions and their properties in all cases. As an example, consider Headrick's (2002) treatment of the

61

approximation to the Weibull distribution with $\alpha = 6$ and $\beta = 10$ as examined in Figure 19. The

set of coefficients he suggests is shown in Equation (60) and reproduced below:

$$\text{Set2}a = \begin{bmatrix} 0.065524 & 0.969217 & -0.065172 & 0.027783 & -0.000117 & -0.003879 \end{bmatrix} (63)$$

However, there is at least one more set of distinctly different transformation coefficients that

presents a solution to the set of equations in (58) for the this distribution:

$$\text{Set2}b = \begin{bmatrix} 0.072274 & 1.00608 & -0.080604 & -0.001916 & 0.002776 & -0.000307 \end{bmatrix} (64)$$

The transformation based on the set of coefficients from Equation (64) and the resulting

distribution look as follows:
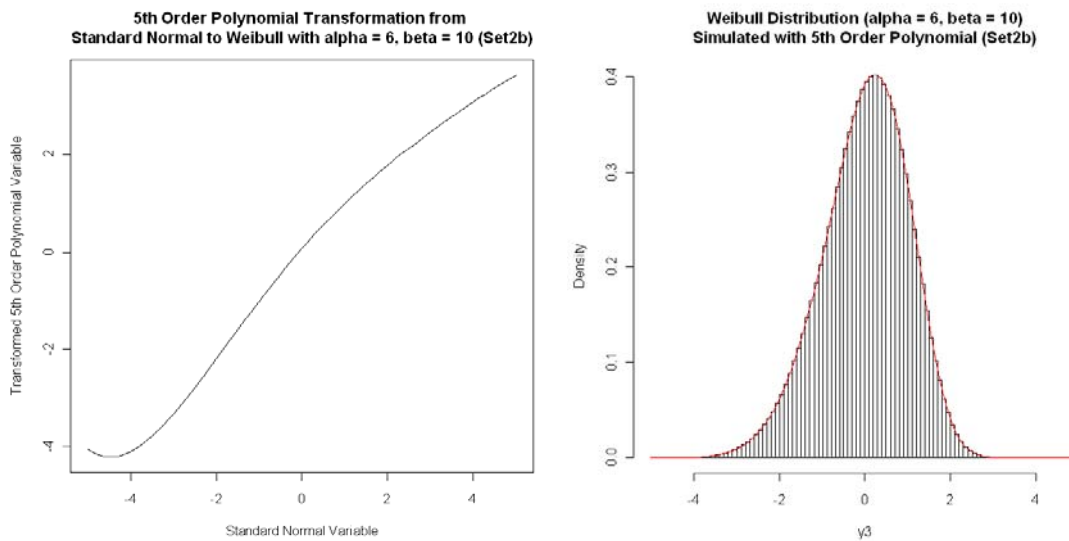
Figure 21: Transformation and Empirical Approximation of the Weibull Distribution Based on the Set of Coefficients from Equation (64)

The empirical distribution in Figure 21 fits the theoretical density function much better than the

empirical distribution in Figure 19.

I further investigated the performance of transformations from both sets of coefficients by

using Equation (57) to simulate 1,000 samples of size $N = 10,000$ for both Set2a and Set2b,

calculating $\hat{\gamma}_4$ for each of the samples and plotting the empirical sampling distribution of the

$\hat{\gamma}_4$ s in Figure 22:
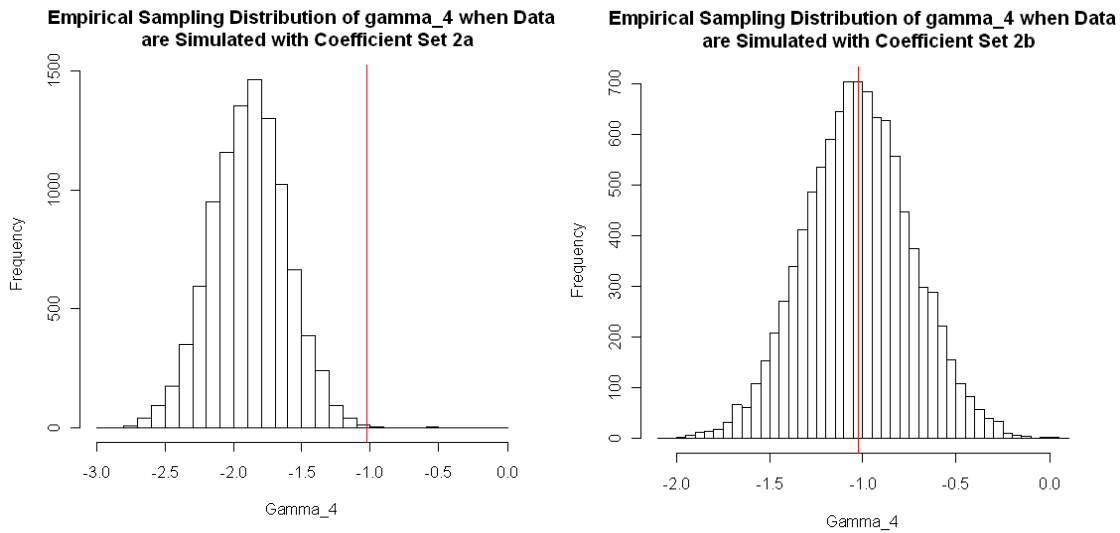
Figure 22(a) & (b): Sampling Distributions for $\hat{\gamma}_4$ Under the Two Sets of Transformation
Coefficients from Equations (63) and (64).

The bulk of the sampling distribution of γ_4 in Figure 22(a) lies below its nominal value of

$\gamma_4 = -1.022066$, while the sampling distribution in Figure 22(b) is more "on point." The 30 most

extreme values were excluded before the sampling distribution of γ_4 in Figure 22(a) was plotted.

The largest of these values was approximately $\hat{\gamma}_4 = 2008$, which does pull the *mean*, but not the

entire distribution significantly upwards.

Monotonicity of Transforms. Some of the transformations for the 3rd order polynomial

method were monotonically increasing, while others were not monotonic, and sometimes a

particular skewness-kurtosis combination could be simulated by both a monotonic and a non-

monotonic transform, in which case the monotonic transform seemed to generally have better performance and should be preferred. Both monotonic and non-monotonic transforms also exist for the 5th order polynomial method. Headrick & Kowalchuk (2007, properties 4.5 and 4.6) present constraints on the coefficients a_0 through a_5 which will guarantee monotonicity of 5th order polynomial transformations and which are similar to the constraints in Equation (23) on a_0 through a_3 for the 3rd order polynomial method. Property 4.5 states that a 5th order polynomial resulting in a symmetric distribution, i.e. $Y = a_1 Z + a_3 Z^3 + a_5 Z^5$, with $a_0 = a_2 = a_4 = 0$, will be monotonic if all numbers z with

$$z = \pm \sqrt{\frac{\pm\sqrt{9a_3{}^2 - 20a_1 a_5} - 3a_3}{10a_5}} \quad \text{or} \quad z = \mp \sqrt{\frac{\pm\sqrt{9a_3{}^2 - 20a_1 a_5} - 3a_3}{10a_5}} \tag{65}$$

have non-zero imaginary parts. Equivalently, property 4.6 states that a 5th order polynomial transform that results in an asymmetric distribution with will be monotonic if all numbers z with

$$z = \pm \frac{1}{2}\sqrt{S_4} \pm \frac{1}{2}\sqrt{S_6 \pm S_5} - \frac{a_4}{5a_5} \quad \text{or} \quad z = \mp \frac{1}{2}\sqrt{S_4} \mp \frac{1}{2}\sqrt{S_6 \pm S_5} - \frac{a_4}{5a_5} \tag{66}$$

have non-zero imaginary parts. Expressions for S_4, S_5, and S_6 can be found in the appendix of Headrick & Kowalchuk (2007) and are rather lengthy formulae of a_0 through a_5. These equations are not very practical. At best, they present a way to program a test for whether a particular transform is monotonically increasing. A graph similar to Figure 9, but for the 5th order power method is needed.

Range of Skewness-Kurtosis. Limitations to the range of skewness-kurtosis combinations available to the 3rd order polynomial method are plotted in Figure 9 and a similar plot for the 5th order polynomial method would be desirable. The 5th order power method's control over

additional moments seems to increase the proportion of the skewness-kurtosis plane $\gamma_2 \geq \gamma_1 - 2$

that can be simulated. In his table 3, Headrick (2004) provides a small set of skewness-kurtosis

combinations on the borders of the available range for the 5th order transform. These are plotted

and connected by the green line in Figure 23 and contrasted with the ranges for 3rd order

monotonic and all 3rd order transforms.
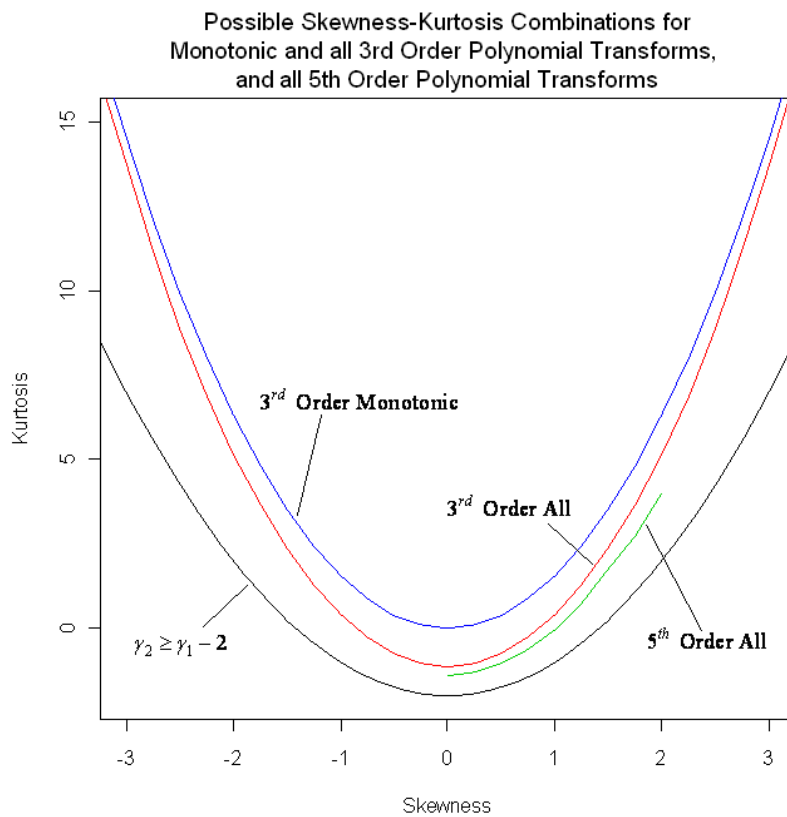
Figure 23: Ranges of Possible Skewness-Kurtosis Combinations for 3rd and 5th Order Polynomial Transforms

Apparently, Equation (57) allows one to solve for a larger range of skewness-kurtosis

combinations than Equation (2) because the larger number of variables and coefficients provides

more flexibility. On the other hand, this makes it more difficult to find the boundaries of the

available region, since the fifth and sixth moment are not part of the graph, but can be varied as well – this may be why Headrick (2004) does not provide more values. Notice, however, that the 5th order power method *can* create nonnormal variables with $\gamma_2 < 0$ via a monotonically increasing transform. An example is given in Headrick & Kowalchuk (2007). It should be a priority for the investigation of the 5th order polynomial transform to complete Figure 23 for both monotonic and *all* 5th order transformations. This will allow users and psychometricians alike to evaluate how much more of the skewness-kurtosis range is covered compared to the 3rd order polynomial transform and whether advantages gained through this increase outweighs disadvantages from the additional complexity of the 5th order polynomial transform. However, the long runtimes to solve for sets of coefficients make this more difficult since empirical exploration is the only known way to construct a figure for the 5th order transform similar to Figure 9 for the 3rd order transform.

The Multivariate 5th Order Polynomial Transform

The extension of the 5th order polynomial transform to the multivariate case can be carried out exactly like the multivariate extension of the 3rd order polynomial transform. After finding the transformation coefficients a_0 through a_5 for each individual variable, the final correlation ρ_Y is a function of the transformation coefficients of the individual variables and the intermediate correlation ρ_Z:

$$\rho_Y = f\left(\rho_Z, \mathbf{a_1}, \mathbf{a_2}\right) \tag{67}$$

where $\mathbf{a_1} = [a_0, a_1, a_2, a_3, a_4, a_5]$ for Y_1 and $\mathbf{a_2} = [a_0, a_1, a_2, a_3, a_4, a_5]$ for Y_2. The fully expressed equation is given in Headrick (2002) and not reproduced here due to its length. ρ_Z is then found

66

as the root of the function in Equation (67). Next, a set of standard normal variables is correlated with ρ_Z, with the same procedure of post-multiplying by the Cholesky decomposition of the intermediate correlation matrix as for the 3rd order power method. As a last step, the transformations to nonnormality are carried out on the individual variables.

Range of Final Correlations. We have seen that for the 3rd order polynomial transform the range of available final correlations can be smaller than $\begin{bmatrix} -1, & 1 \end{bmatrix}$ and that this range depends not only on the marginal moments but also the set of coefficients chosen for the univariate transformations. Assessment of this aspect for the 5th order power method is decidedly more difficult as it takes at least several hours for *Mathematica* to end its solution search process. We also cannot be certain that all solutions have been found. To nevertheless provide an example, I will again revert to the sets of coefficients found for the approximation to the χ_1^2 distribution. The following subsets of coefficients from Equation (62) are distinctly different (all are non-monotonic):

$$\begin{aligned}
\text{Set1} &= \begin{bmatrix} -0.3977 & 0.6211 & 0.4169 & 0.0684 & -0.0064 & .00004 \end{bmatrix} \\
\text{Set2} &= \begin{bmatrix} -0.6689 & -0.3194 & 0.6671 & 0.0007 & 0.0006 & 0.00005 \end{bmatrix} \quad (68) \\
\text{Set3} &= \begin{bmatrix} -0.7071 & 0.0000 & 0.7071 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}
\end{aligned}$$

resulting in six unique combinations. The approximate ranges for final correlations are displayed in Table 9:

Table 9: Range of Final Correlations for 5th Order Power Method Approximation to Bivariate χ_1^2 Distribution

	Min	Max
Set1 – Set1 ($non-monotonic_1$ & $non-monotonic_1$)	−.424858	1
Set1 – Set2 ($non-monotonic_1$ & $non-monotonic_2$)	−.033750	.769071
Set1 – Set3 ($non-monotonic_1$ & $non-monotonic_3$)	0	.535341
Set2 – Set2 ($non-monotonic_2$ & $non-monotonic_2$)	−.002792	1
Set2 – Set3 ($non-monotonic_2$ & $non-monotonic_3$)	0	.948543
Set3 – Set3 ($non-monotonic_3$ & $non-monotonic_3$)	0	1

Figure 24 plots the relationship between ρ_Z and ρ_Y for the first (Set1 – Set1) and the last

combination of transformation coefficient sets (Set3 – Set3):
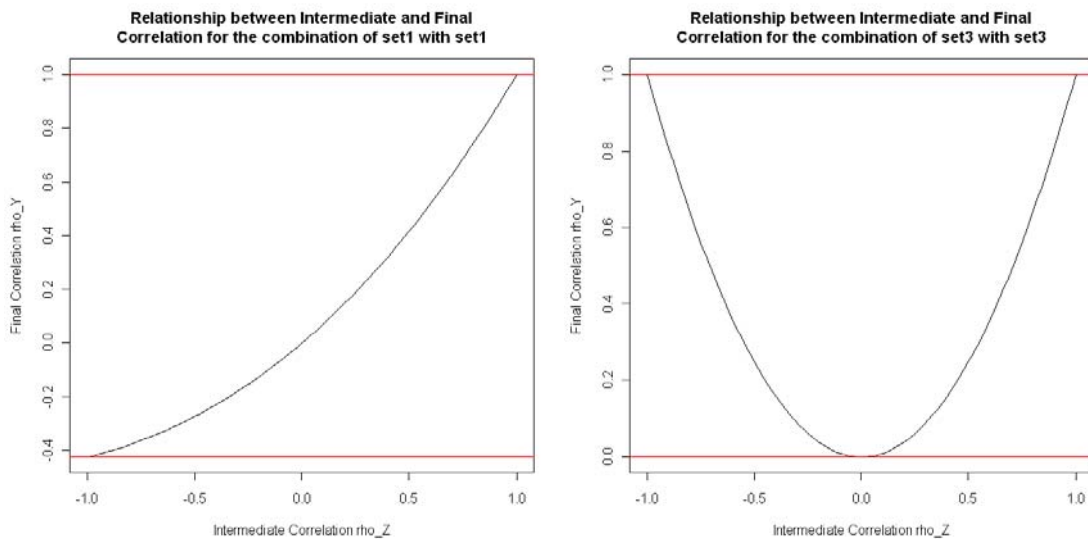
Figure 24(a) and (b): Relationship Between Intermediate and Final Correlation for 5th Order Power Method Example.

Surprisingly, the largest range of final correlations available is *not* for the most "natural"

transformation via the coefficients from Set3, which is of the form $Y = a + bZ^2$ and has been

recommended by Headrick & Kowalchuk (2007)! The largest range of final correlations is

achieved when both variables are transformed via Set1.

We can also compare these values to the ranges of final correlations available with the 3rd

order polynomial method. Sets of coefficients to approximate a χ_1^2 distribution are:

$$
\begin{aligned}
\text{Set1} &= \begin{bmatrix} -0.5207 & 0.6146 & 0.5207 & 0.0201 \end{bmatrix} \\
\text{Set2} &= \begin{bmatrix} -\sqrt{.5} & 0 & \sqrt{.5} & 0 \end{bmatrix}
\end{aligned}
\tag{69}
$$

Both sets of coefficients are non-monotonic. Resulting ranges for final correlations are:

Table 10: Range of Final Correlations for 3rd Order Power Method Approximation to Bivariate χ_1^2 Distribution

	Min	Max
Set1 – Set1 ($non-monotonic_1$ & $non-monotonic_1$)	−.095792	1
Set1 – Set2 ($non-monotonic_1$ & $non-monotonic_2$)	0	.736347
Set2 – Set2 ($non-monotonic_2$ & $non-monotonic_2$)	0	1

At least for this particular example it seems that the 5th order polynomial offers a wider range of

final correlations and therefore a wider range of bivariate distributions to be simulated – **if** the

right set of coefficients is chosen.

Odd-Shaped Distributions under the 5th Order Polynomial Transform. The 5th order

power method can, just like the 3rd order power method, result in odd-shaped distributions.

Figure 25 graphs an odd-shaped 5th order power method nIM distribution. Neither of the

variables has extreme values for skewness or kurtosis.
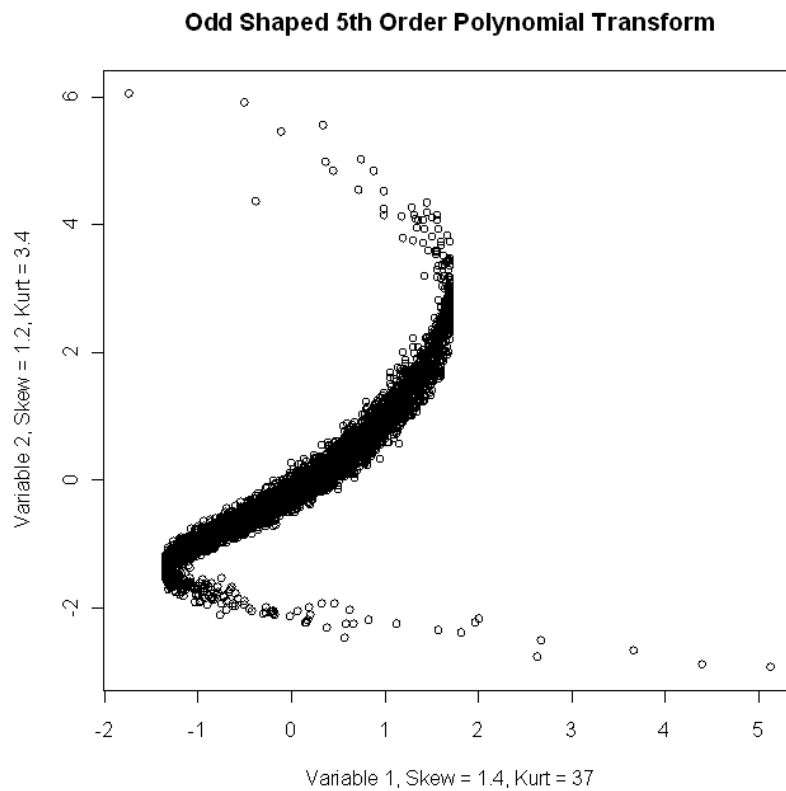
Figure 25: An Odd-Shaped Distribution under the 5th Order Polynomial Transformation

The *g*-and-*h* Distribution

Polynomial transforms are not the only transformation method for obtaining nonnormal

random variables. Another technique that allows control over univariate skewness and kurtosis of

the resulting nonnormal random variable as well as the covariance matrix for the multivariate

extension is the *g*-and-*h* distribution, initially suggested by Tukey (1977). The *g*-and-*h* distribution

has received a fair amount of attention in the literature. Its univariate version or slight variations

thereof have been described by several authors such as Headrick, Kowalchuk, & Sheng (2008),

Hoaglin (1985), Hoaglin & Peters (1979), and Martinez & Iglewicz (1984). Multivariate

extensions (see section on the multivariate *g*-and-*h* distribution below) have been elaborated on

by Field & Genton (2006) and Kowalchuk & Headrick (2010). The *g*-and-*h* distribution has been

utilized to simulate nonnormal distributions in Monte Carlo research (e.g., Keselman, Lix, &

Kowalchuk, 1998; Wilcox, 1994; Algina, Keselman, & Penfield, 2006) or to model extreme

events, e.g. relating to the stock market (e.g., Badrinath & Chatterjee, 1988; Badrinath &

Chatterjee 1991). See Headrick et al. (2008) for a more comprehensive list of applications.

The general form of the *g*-and-*h* transformation is

$$Y = \frac{\exp(gZ)-1}{g} \exp\left(\frac{hZ^2}{2}\right) \tag{70}$$

where *Z* is a standardized normal random variable. Example plots of *g*-and-*h* transformations and

histograms of the resulting distributions can be found below in Figure 26 and Figure 27.

Nonnormal distributions can also be created by setting one of the coefficients to zero,

thus reducing equation (70) to

$$Y = \frac{1}{g}\left(e^{gZ}-1\right) \tag{71}$$

if *h* is set to zero or

$$Y = Z\exp\left(\frac{hZ^2}{2}\right) \tag{72}$$

if *g* is set equal to zero. Note that if *g* is to be set equal to zero, Equation (72) *has* to be used for

transformations, since $g = 0$ would lead to an indeterminate term in Equation (70). Distributions

created with Equation (72) will always be symmetric. Mean, variance, skewness and kurtosis for

a univariate *g*-and-*h* variable are expressed as functions of *g* and *h* in the appendix, due to their

length. Moments for distributions that use only *g* or only *h*, i.e. that use transformations (71) or (72), can be found in Hoaglin (1985).

   Finding Coefficients g and h. As for the polynomial transforms, a set of nonlinear equations is solved numerically to find the transformation coefficients *g* and *h*: Equations (130) and (131) are set equal to the desired values for skewness and kurtosis and then solved for *g* and *h*. Notice that due to the term $\sqrt{1-4h}$ in Equation (131), we can only solve for $h < .25$ using this technique. However, it is certainly possible to enter values for *h* larger than .25 into Equation (70) or (72), and values of $h \geq .25$ have been employed in Monte Carlo studies (e.g. Wilcox, 1994).

   While some authors (e.g. Field & Genton, 2006; Headrick et al., 2008; Wilcox, 1995) seem to suggest that *g* controls the skewness of *Y* independently of *h*, while *h* alone controls kurtosis, both parameters interact to determine γ_1 and γ_2, as can be easily seen from Equations (130) and (131). This lack of clarity may stem from the conceptualization of "excess kurtosis" as discussed in Hoaglin (1985). Hoaglin treats tail elongation "naturally" resulting from increased skewness separately from additional tail heaviness not associated with skewness (see page 486 and page 504 in Hoaglin, 1985). However, this is not to be confused with the straightforward concepts of skewness and kurtosis as $E\left(X^3\right)$ and $E\left(X^4\right)-3$.

Limitations of the Univariate g–and–h Distribution

   In previous sections, we have encountered a range of issues with the 3rd (and 5th) order power method:

   (1) The power methods do not yield unique solutions when solving numerically for the transformation coefficients: The 3rd order power method yields 2 distinct solutions, while

the exact number of distinct solutions for the 5th order power method is not known. Further, not all power method transformations are monotonically increasing.

(2) The range of skewness and kurtosis combinations available with the power method is limited. Limitations for monotonic transforms vs. all transforms of the power method are shown in Figure 9 and Figure 23.

(3) The range of final correlations available between two power method variables Y_1 and Y_2 depends on the values of marginal skewnesses and kurtoses as well as the specific sets of coefficients chosen for the transformations. This range of available final correlations is not always -1 to 1, but can be severely limited.

We now investigate whether these issues apply to the *g-and-h* distribution as well.

(1) Multiple Solutions to the Set of Equations for the g-and-h Distribution & Monotonicity. When solving Equations (130) and (131) for the coefficients *g* and *h*, both *Mathematica*'s FindRoot routine and R's `nleqslv` function found a total of two real-valued solutions, one of which had comparatively large values for *g* and *h*. Both sets of solutions yield first four moments as desired when reentered into Equations (128), (129), (130) and (131).

To provide a few examples, I choose three relatively common skewness-kurtosis combinations (see Table 28): $\gamma_1 = 0$ and $\gamma_2 = -1$, $\gamma_1 = 1.25$ and $\gamma_2 = 3.75$, and last but not least, $\gamma_1 = 0$ and $\gamma_2 = 25$. The resulting sets of coefficients are:

Table 11: Two Sets of g-and-h Coefficients for Each Skewness-Kurtosis Combination

Skewness and Kurtosis	Transformation A		Transformation B	
1) $\gamma_1 = 0$, $\gamma_2 = -1$	$g = 0$	$h = -0.1610$	$g = 0$	$h = -5.1774$
2) $\gamma_1 = 1.25$, $\gamma_2 = 3.75$	$g = 0.3226$	$h = 0.0377$	$g = 3.5915$	$h = -67.9942$
3) $\gamma_1 = 0$, $\gamma_2 = 25$	$g = 0$	$h = 0.1930$	$g = 0$	$h = -1392.03$

You will notice that the sets of coefficients under the heading "Transformation A" are relatively close to zero, while the other set of coefficients is comparatively large in absolute value. What do these transformations look like when plotted as a function that relates standard normal scores to the nonnormal Y scores? In Figure 26, the first column of transformations (A) is plotted in black, and the transformations from the second column (B) are plotted in red. Transformations in the right column of Figure 26 display fairly extreme changes in direction close to a value of 0 on the Z scale, while being essentially horizontal everywhere else. Transformations in the left column of Figure 26 are considerably smoother compared to their counterparts in the right hand column. The transformation in Figure 26(d) relates the values of $-4 \leq Z \leq 4$ to values of Y between approximately $-.05$ and $.1$. However, note that the range of values on the Y variable is of only secondary importance, since this can always be adjusted by rescaling. The resulting univariate distributions are plotted in Figure 27.
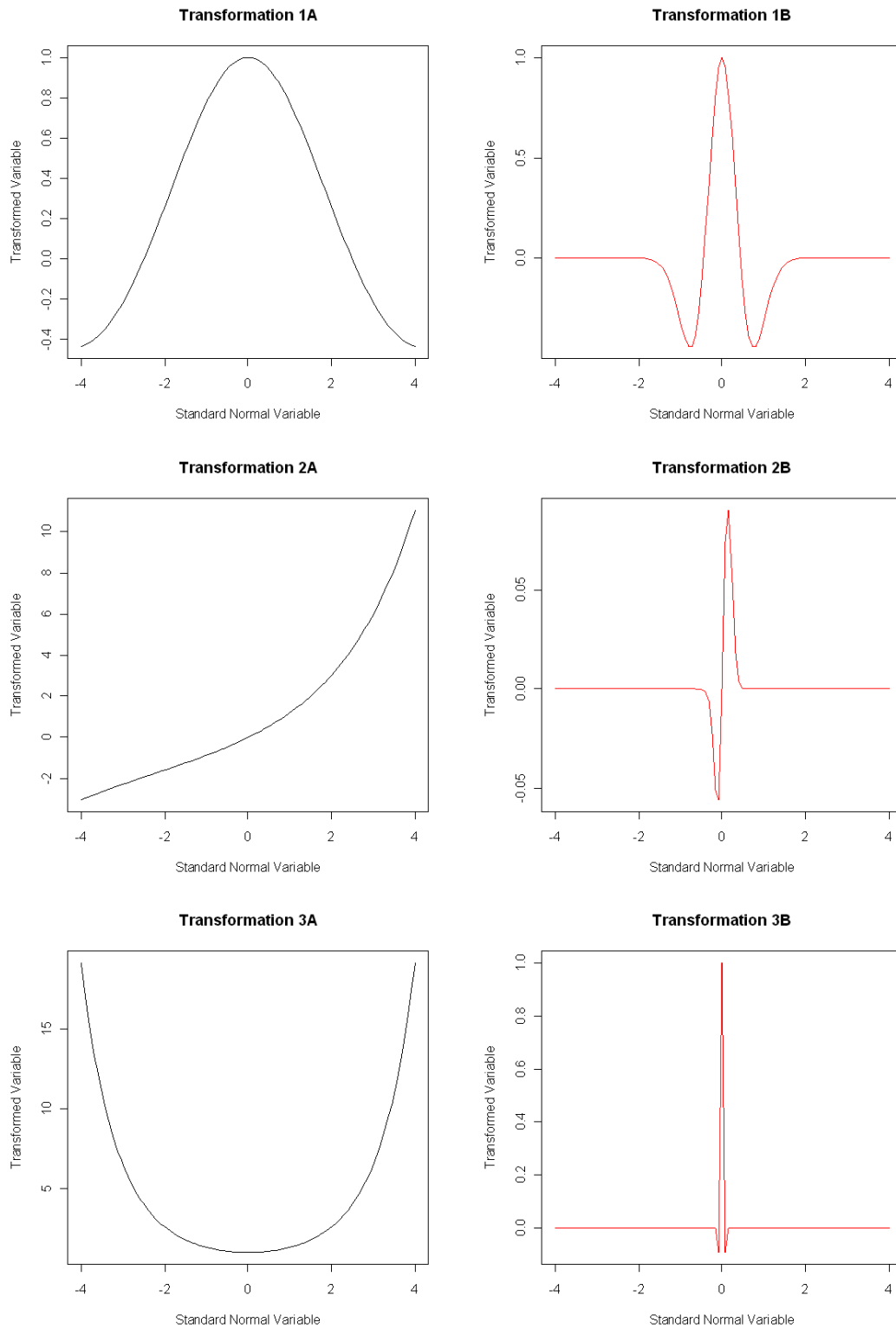
Figure 26(a) through (f): g-and-h Transformations corresponding to Table 11. Transformation 1 in black, transformation 2 in red.
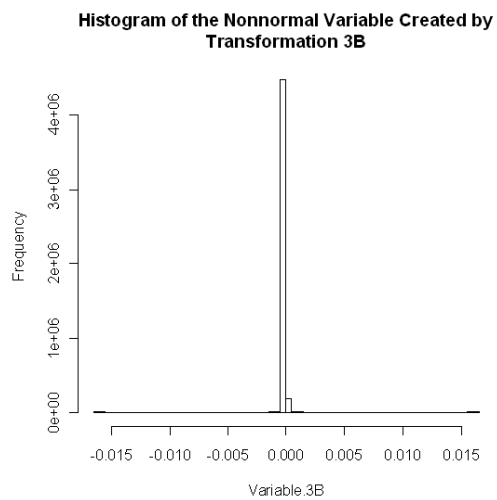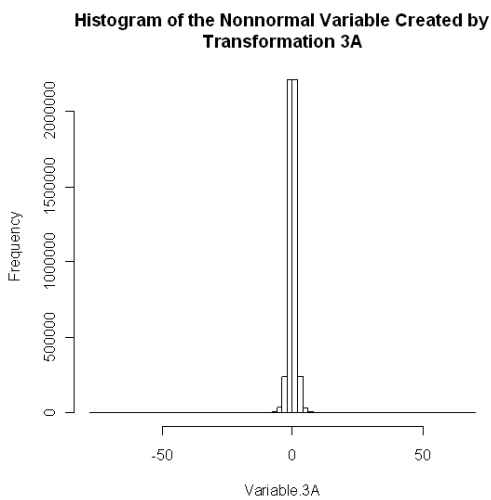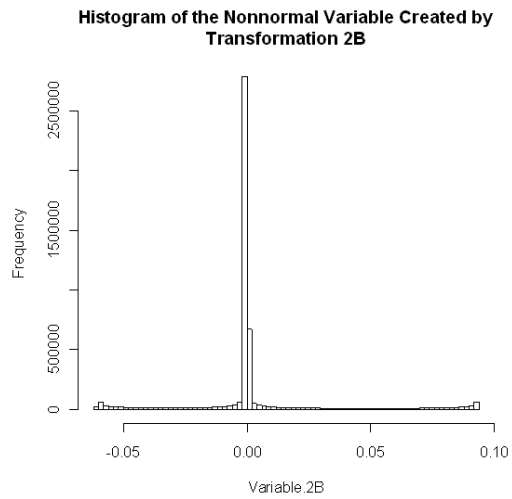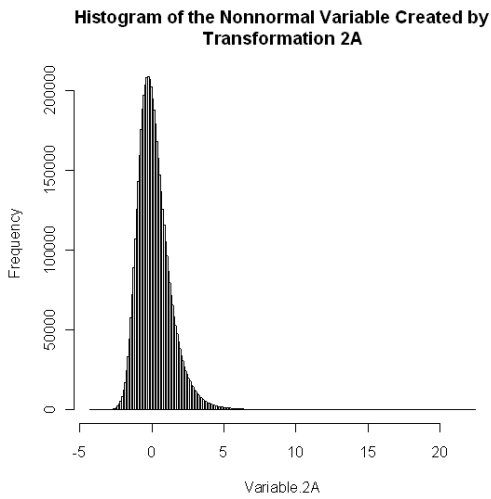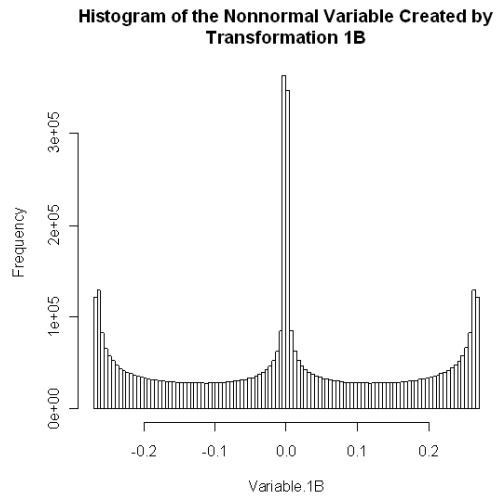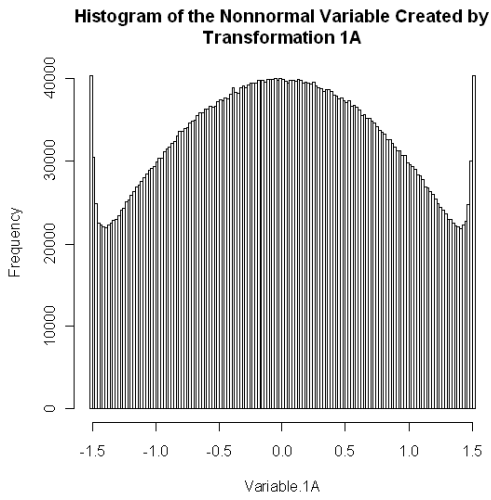
Figure 27: Histograms related to transformations in Figure 26, based on 5,000,000 replications.

Table 12: Empirical Skewnesses and Kurtoses for Distributions in Figure 27

	Nominal	$\hat{\gamma}_1$	$\hat{\gamma}_2$
Variable 1A	$\gamma_1 = 0$, $\gamma_2 = -1$	0.000367	–1.000040
Variable 1B		0.000326	–1.000222
Variable 2A	$\gamma_1 = 1.25$, $\gamma_2 = 3.75$	1.244497	3.682793
Variable 2B		1.250411	3.755162
Variable 3A	$\gamma_1 = 0$, $\gamma_2 = 25$	–0.040349	18.795985
Variable 3B		–0.011476	24.955543

Overall, the large sample estimates for $\hat{\gamma}_1$ and $\hat{\gamma}_2$ conform quite well with their nominal

counterparts, with the exception of kurtosis for Transformation 3A. It also seems that for the vast

majority of skewness-kurtosis combinations, the transformation with the larger coefficients will

lead to quite odd-shaped distributions, hence making the choice between coefficient sets

straightforward.

Monotonic and Non-monotonic g-and-h Transformations. In many cases, properties of

simulated power method distributions such as shape, distributional fit, and range of final

correlations are related to the monotonicity of the transformation function. Hence, it seems

necessary to examine monotonicity and its consequences for the *g*-and-*h* transform as well. Only

one of the six transformations in Figure 26 is monotonic. An examination of the derivatives for

Equation (70),

(71), and (72) can provide a first overview of when we can expect a monotonic *g*-and-*h*

transformation to be available. To assess whether a *g*-and-*h* transformation is monotonic,

consider the three possible cases (1) $g \neq 0, h = 0$, (2) $g = 0, h \neq 0$, and (3) $g \neq 0, h \neq 0$,

corresponding to Equations (71), (72), and (70).

For case (1), the derivative is

$$f'(Z) = e^{gZ} \tag{73}$$

This is always positive, hence, the transformation from Equation (71) with $g \neq 0, h = 0$ is always strictly increasing. For case (2), the derivative is

$$f'(Z) = e^{hZ^2}\left(1 + hZ^2\right) \tag{74}$$

The term e^{hZ^2} will always be greater than zero, therefore potential roots of Equation (74) must be roots of $1 + hZ^2$. For $Z = 0$, no root exists. For $Z \neq 0$, we must require $h \geq 0$ for the transformation to be monotonic; if $h < 0$, $Z = \sqrt{1/-h}$ will be a root (see Hoaglin, 1979).

Finally, case (3) is decidedly trickier. The derivative of the transformation in Equation (70) is

$$f(Z)' = e^{hZ^2/2}\left(e^{gZ} - \frac{hZ}{g} + \frac{hZ}{g}e^{gZ}\right) \tag{75}$$

We can certainly find values for *g* and *h* for which this derivative will always be strictly increasing, as, e.g., in transformation 2A: With $g = .322587$ and $h = .0377476$, we have

$$f(Z)' \approx e^{.019Z^2}\left(e^{.323Z} - .117 \times Z\left(1 - e^{.323Z}\right)\right) \tag{76}$$

When $Z = 0$, this reduces to $f(Z)' = 1$. When $Z < 0$, $0 < e^{.323Z} < 1$, therefore

$$\begin{aligned}
f(Z)' &\approx e^{.019Z^2}\left(e^{.323Z} - .117 \times Z\left(1 - e^{.323Z}\right)\right) \\
&\geq e^{.019Z^2}\left(e^{.323Z} + .117 \times Z\left(-1 + 1\right)\right) \\
&= e^{.019Z^2}e^{.323Z} \\
&> 0
\end{aligned} \tag{77}$$

Finally, when $Z > 0$, $1 < e^{.323Z}$, and

$$f(Z)' \approx e^{.019Z^2}\left(e^{.323Z} - .117 \times Z\left(1 - e^{.323Z}\right)\right)$$

$$\geq e^{.019Z^2}\left(e^{.323Z} + .117 \times Z\left(-1+1\right)\right) \tag{78}$$

$$= e^{.019Z^2}e^{.323Z}$$

$$> 0$$

Equivalently, it is easy to show that choices for $g \neq 0$ and $h \neq 0$ exist for which the transformation is non-monotonic. Let $g = -2.6604$ and $h = -.2$, then we have $f(Z)' = 0$ at $Z = 1$ and $f(.5)' \approx .231$, i.e. positive, and $f(1.5)' \approx -0.074$, i.e. negative. No further systematical evidence or analysis regarding the monotonicity of the *g-and-h* transform exists.

Headrick et al. (2008) (and based thereupon Kowalchuk & Headrick, 2010) suggest that transformations in Equation (70) will always be monotonic when $g \neq 0$ and $h > 0$, but do not provide a proof. In fact, Headrick et. al. (2008) and Kowalchuk & Headrick (2010) favor and almost exclusively focus on the subset of the *g-and-h* distributions with $g \neq 0$ and $h > 0$ and derive an analytical form of the pdf and cdf for this subset in a fashion quite similar to the one described for the power method in section on the "valid pdf/cdf" above. Kowalchuk & Headrick only present examples with positive kurtosis, and it is possible that when $h > 0$, only *g-and-h* variables with positive kurtosis can be created. However, they do not provide a compelling explanation of the need for a monotonically increasing transformation aside from procuring a way to derive a pdf and cdf. Hoaglin (1985) mentions that monotonicity is not obtained when $h < 0$ for the transformation in Equation (72) but makes no mention regarding monotonicity for the general *g-and-h* transformation from Equation (70).

(2) Range of Available Skewness-Kurtosis Combinations. For both power methods (3rd and 5th order), the portion of skewness-kurtosis combinations that can be simulated is limited,

especially if a monotonic transformation is desired (see Figure 23). We will now investigate

whether the *g*-and-*h* distribution suffers from similar limitations. Assessing the range of skewness

and kurtosis available to univariate *g*-and-*h* distributions turned out to be a somewhat challenging

task. As with the 3rd order power method, this range was assessed by attempting to numerically

solve Equations (130) and (131) for *g* and *h* for a variety of skewness-kurtosis combinations. If

the root finding procedure obtained a solution, that skewness-kurtosis combination was recorded

as available, if no solution was found, it was recorded as unavailable.

However, whether Equations (130) and (131) could be solved for *g* and *h* frequently

depended on specifics of numerical root finding. The availability of solutions was cross-validated

by using both R and *Mathematica*, which occasionally disagreed on whether the equations could

be solved for a given skewness-kurtosis combination. Furthermore, within each program,

solvability also depended on the various settings of the root finding algorithm (FindRoot for

Mathematica and `nleqslv` for R). Figure 28 provides insight into one of the problems that

were encountered. Plotted are (a) Total range of the skewness-kurtosis plane (black line), (b)

Portion of that plane that can be solved for by the 3rd order power method with monotonic and all

transformations (green and red lines), (c) Portion of the skewness-kurtosis plane that could be

solved for the *g*-and-*h* distribution in *Mathematica* (blue line), and (d) Portion of this plane that

could be solved for the *g*-and-*h* distribution in R (black dots). Across the three panels, number of

iterations for the numerical root finding `nleqslv` in R was varied from 100 to 500:

| 100 iterations | 300 iterations | 500 iterations |
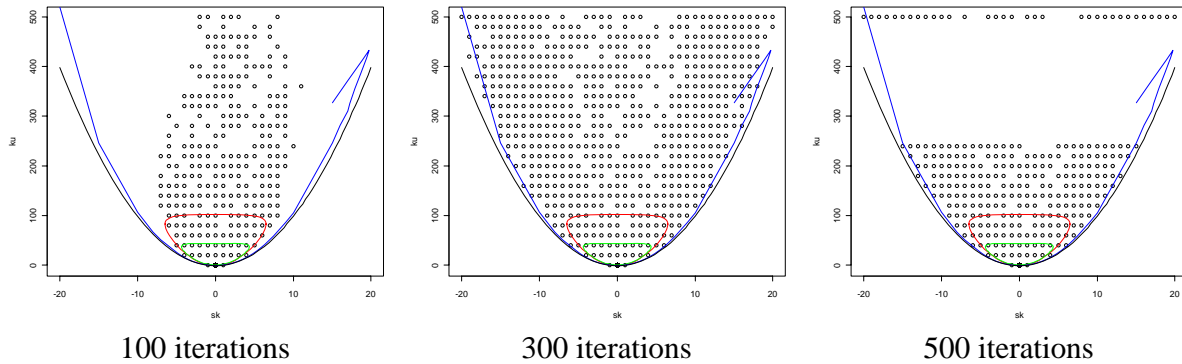
Figure 28 (a) – (c): Finding Solutions for the g-and-h Distribution Depends on Number of Iterations

As the maximum number of iterations increases from 100 to 300, R finds more solutions to Equations (130) and (131). However, increasing the number of iterations even more, R's root finding algorithm `nleqslv` encounters a singular Jacobian more often, and the number of solutions found decreases again. This could *possibly* be remedied by varying the starting values for the numerical iteration. It should further be noted that the maximum number of iterations (with 150 as the default value) is only one manipulable parameter of the `nleqslv` routine for numerical root finding. Others include, but are not limited to,

(a) Step length of x values (for f(x)). For example, if `xtol=1e-6`, the algorithm is stopped if the step for the next iteration is smaller than 1e-6 for all x-values;

(b) Function value tolerance. Convergence is declared when the largest absolute function value is smaller than `ftol`, e.g., 1e-6;

(c) Backtracking tolerance;

(d) An option to obtain detailed report of progress of iteration, which seems to be fairly comprehensive and informative;

(e) The initial trust region size.

81

The range of skewness-kurtosis combinations for which solutions for *g* and *h* were found with *Mathemtica* (as shown by the blue line in Figure 28) differed slightly from the range found with R. The indent in the upper right corner in the blue line is most probably an artifact occurring due to the settings in *Mathematica*'s FindRoot routine.

During my attempts to write my own R routine that would solve for *g* and *h*, I was occasionally able to increase the number of solutions found by manipulating the step length for *x* values (`xtol`), the function value tolerance (`ftol`), and the maximum number of iterations (`maxit`), as well as the starting values for the iteration. However, this is a time consuming process that requires good general knowledge of numerical optimization methods and how they relate to the function on which the root finding is performed. These technical requirements probably exceed the level of the many users of nonnormal random number generation. Because of the complicated interaction of settings in the root finding procedure with the desired skewness-kurtosis combination, very careful programming would be necessary to create a routine that finds *g* and *h whenever they exist*.

To get an overview of the total range of the skewness-kurtosis plane that can be simulated by the *g*-and-*h* distribution, I assume that if a solution for some specific combination of γ_1 and γ_2 was found in any one of the different root finding settings, this solution was only not found during other searches because of characteristics of the root finding process. Figure 29 then plots the range of skewness-kurtosis combinations available to the *g*-and-*h* distribution (stars are combinations for which R found a solution, the blue line spans the range of combinations for which *Mathematica* found a solution): Almost the entire plane spanned by $\gamma_2 \geq \gamma_1 - 2$ (see the black line) is available. Figure 29 also contains the range for the V&M method (red and green lines). Inspection of Figure 28 and Figure 29 shows that the *g*-and-*h* distribution can be used to

simulate practically any skewness-kurtosis combination possible. It provides a much wider range of distributions that can be simulated than the power method, especially distributions with large values for γ_2. The *g*-and-*h* distribution, compared to the V&M method, also offers simulation of a slightly extended range of platykurtic distributions, which can be useful for simulating floor and ceiling effects. To verify the validity of some of the solutions found, I both reentered the resulting *g* and *h* parameters into Equations (130) and (131) and used them to simulate sample distributions with large *N*, calculating the sample skewness and kurtosis for those, always with reassuring results.

It is unclear what the range of available skewness-kurtosis combinations for *monotonically increasing g*-and-*h* transformations is. If Headrick et al. (2008) are correct, and transformations for *g*-and-*h* distributions are only increasing when $g \neq 0$ and $h > 0$, it may be possible that no *g*-and-*h* distribution with $\gamma_2 < 0$ can be simulated via a monotonic transformation.

**Range of Skewness & Kurtosis
that can be Simulated by the g-and-h Distribution**
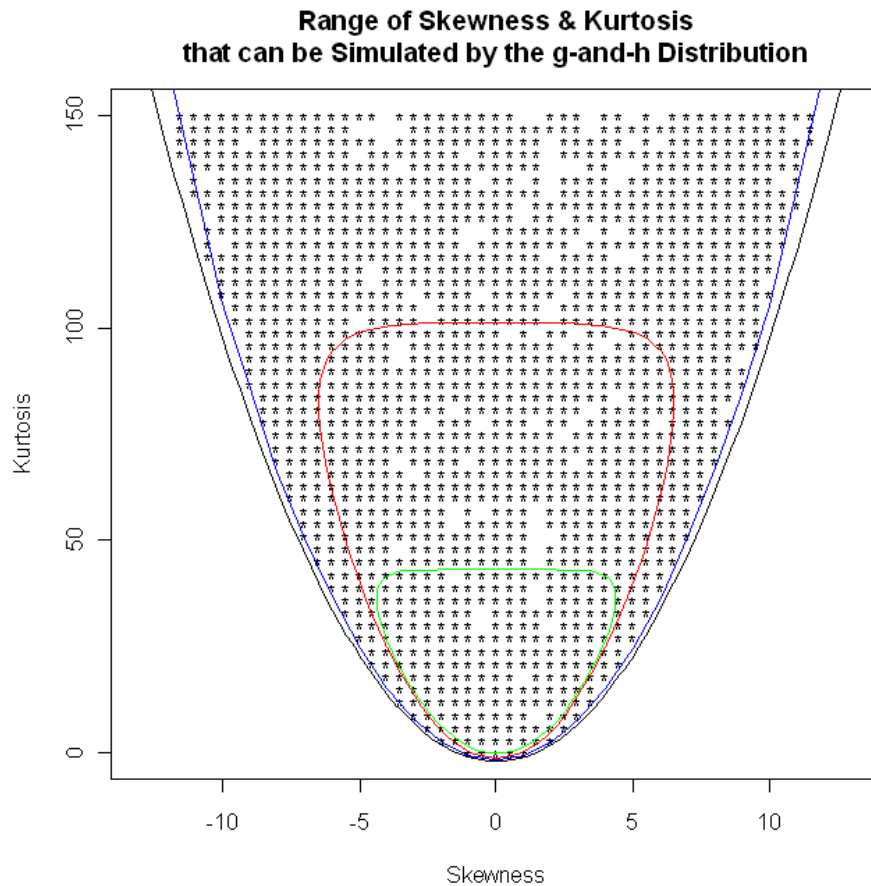
Figure 29: Range of Univariate Skewness and Kurtosis available for the g-and-h Distribution, Compared with 3rd Order Power Method

The Multivariate g-and-h Distribution

There have been two extensions of the univariate *g*-and-*h* distributions to the multivariate case, one by Field & Genton (2006), which is based on quantile fitting as described in great detail in Hoaglin (1985, see sections below), and one by Kowalchuk & Headrick (2010), which is very similar in procedure to the multivariate power method extensions. I will focus on the latter due to its comparability with the power methods. The extension is carried out as follows: First, the univariate coefficients g_i and h_i for the transformation to nonnormal marginal distributions are determined as described in the previous section. The second step then involves

84

solving for the entries of an intermediate correlation (matrix) individually, which will depend on the transformation coefficients for their respective variables and the desired final correlation. Just as for the power method, each final correlation ρ_Y between two *g*-and-*h* variables Y_1 and Y_2 can be expressed as a function of the transformation coefficients g_1, h_1, g_2, and h_2 and the intermediate correlation ρ_Z between the two standard normal random variables:

$$\rho_Y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\dfrac{\left(e^{g_1 z_1}-1\right)e^{h_1 \frac{z_1^2}{2}}}{g_1}-m_1}{s_1} \cdot \frac{\dfrac{\left(e^{g_2 z_2}-1\right)e^{h_2 \frac{z_2^2}{2}}}{g_2}-m_2}{s_2} \cdot \frac{\exp\left(-\dfrac{z_1^2 - 2\rho_Z z_1 z_2 + z_2^2}{2\left(1-\rho_Z^2\right)}\right)}{2\pi\sqrt{1-\rho_Z^2}} \, dz_1 dz_2 \quad (79)$$

Notice that Equation (79) is a direct adaptation of Equation (33). Once all entries of the intermediate correlation matrix have been found, a set of independent standard normal random scores is multiplied by the Cholesky factor of the intermediate correlation matrix, just as for the power method. Finally, the now correlated standard normal random scores are individually transformed with Equation (70), using the coefficients *g* and *h* found in the first step.

Special Issues of the Multivariate g-and-h Distribution

In addition to the issues discussed for the univariate case (non-uniqueness of solutions and especially the dependency on numerical root finding performance), the multivariate extension of the *g*-and-*h* distribution has an additional difficulty to master: The bivariate integral in Equation (79) cannot be solved analytically for ρ_Y. In practice, instead of having a (relatively) simple function that relates ρ_Y to ρ_Z, such as Equation (35) for the 3rd order power method, one must carry out a numerical integration of Equation (79), choosing integration limits for Z_1 and Z_2 that are large enough to yield results that are practically identical to what integration limits of

infinity would have produced. Kowalchuk & Headrick (2010) use $Z_1 = Z_2 = \begin{bmatrix} -8, & 8 \end{bmatrix}$. To this date, with the exception of one partial example of *Mathematica* code in Kowalchuk & Headrick (2010), no software implementation of the multivariate *g*-and-*h* distribution has been published (and we might soon see why). Therefore, I wrote my own R and *Mathematica* code to simulate multivariate *g*-and-*h* distributions. I encountered a number of difficulties. In R, these were:

- The most recent routine for multivariate numerical integration I could find was the *adapt* package, which was updated for Windows last in 2007 and is not available in the standard repositories anymore. A quick internet search for R package *adapt_1.0-4.zip* will find the necessary zip file which can then be installed locally (tar.gz files for LINUX distributions are available as well).

- The success of the numerical integral also depended on several adjustable parameters such as (a) `eps` – the "desired accuracy for the relative error" and (b) `minpts` – the "minimum number of function evaluations." If the minimum number of function evaluations was too small, e.g. 1,000, the results of the numerical integration often showed erratic behavior, such as discontinuities. Collaboration with mathematicians specializing in numerical optimization could possibly lead to a stable, yet efficient routine for finding ρ_Y.

- Once are successfully calculating ρ_Y from the transformation coefficients and ρ_Z, we still need to reverse the process and find the intermediate correlation ρ_Z based on the transformation coefficients and a desired final correlation ρ_Y. I wrote a simple bisection algorithm to accomplish this task, but finding ρ_Y often took considerable amounts of time.

Issues in *Mathematica*: To double-check results obtained in R, I also implemented the multivariate *g*-and-*h* distribution in *Mathematica*. Despite occasional error messages regarding slow convergence, *Mathematica* generally executed the numerical integration in a reliable fashion. As in R, I employed my own bisection algorithm to find the intermediate correlation ρ_Z needed for a desired ρ_Y. Generally, finding ρ_Z in *Mathematica* could be achieved in less computation time than in R, a circumstance that might be remediable with more careful programming. To summarize, we can say that the numerical optimization routines required for implementation of the *g*-and-*h* distribution are more challenging than routines needed for the implementation of the power methods. This may be a reason for the lack of a published program that will simulate multivariate *g*-and-*h* distributions with desired marginal skewnesses and kurtoses and correlation matrix.

Range of Final Correlations for the g-and-h Distribution. Another issue we encountered when investigating the multivariate extension of the power method was an occasional restriction of the *range* of final correlations available. Depending on marginal skewnesses and kurtoses as well as the choice of transformation coefficients, the range of available final correlations could be perfect, as, e.g., in the first and third panel in Figure 15, or severely limited, as in the second panel of Figure 15. To be able to compare relationships between intermediate and final correlations for the *g*-and-*h* distribution with the same relationships for the 3rd and 5th order power method, I chose a few skewness-kurtosis combinations which are displayed in Table 13 below. It includes 1) One bivariate IM distribution with $\gamma_1 = 0$ and $\gamma_2 = 25$ for both variables, a distribution that has been examined in great detail for the 3rd order polynomial, 2) One bivariate IM distribution with $\gamma_1 = \sqrt{8}$ and $\gamma_2 = 12$ for both variables, a skewness-kurtosis combination

that has been the focus of many of Headrick et al's investigations of the 5th order polynomial distribution, 3) One bivariate IM distribution with $\gamma_1 = 1.25$, $\gamma_2 = 3.75$ for both variables, a setting that has been quite popular with Monte Carlo studies employing the simulation of nonnormal distributions (see Table 28 below), and 4) One bivariate nIM distribution with a different set of marginal skewness and kurtosis for each variable, as examined for the 3rd order polynomial. In previous sections, each *g-and-h* distribution has two distinctly different sets of coefficients, leading to transformation A and B, listed in separate columns of Table 13:

Table 13: Sets of g-and-h Coefficients

Skewness and Kurtosis Values	Transformation A	Transformation B
1) $\gamma_{11} = \gamma_{12} = 0$, $\gamma_{21} = \gamma_{22} = 25$	$g = 0$, $h = 0.1930$	$g = 0$, $h = -1392.03$
2) $\gamma_{11} = \gamma_{12} = \sqrt{8}$, $\gamma_{21} = \gamma_{22} = 12$	$g = 1.1511$, $h = -0.1910$	$g = 11.4037$, $h = -263.784$
3) $\gamma_{11} = \gamma_{12} = 1.25$, $\gamma_{21} = \gamma_{22} = 3.75$	$g = 0.3226$, $h = 0.0377$	$g = 3.59147$, $h = -67.9942$
4i) $\gamma_{11} = 1$, $\gamma_{21} = 20$, 4ii) $\gamma_{12} = 2$, $\gamma_{22} = 40$	i) $g = 0.1017$, $h = 0.1760$ ii) $g = 0.1935$, $h = 0.1757$	i) $g = 4.2963$, $h = -909.117$ ii) $g = 11.4292$, $h = -3061.36$

For the first three bivariate distributions, there are three ways to uniquely combine distinctly different sets of transformation coefficients, while for the last distribution there are four ways to do so. Combined, there are 13 unique ranges of available final correlations for these four distributions, listed in Table 14. These ranges of available correlations and their corresponding graphs were created by executing the numerical integration for Equation (79) for individual values of ρ_Z, ranging from –.999 to .999.

Table 14: Range of Final Correlations

Skew & Kurt	Transformation	Max	Min
$\gamma_{11} = \gamma_{12} = 0$ $\gamma_{21} = \gamma_{22} = 25$	A & A	−.9882	.9882
	A & B	−.0377	.0377
	B & B	−.2361	.2361
$\gamma_{11} = \gamma_{12} = \sqrt{8}$ $\gamma_{21} = \gamma_{22} = 12$	A & A	−.4346	.9605
	A & B	−.0630	.0024
	B & B	−.5469	.7311
$\gamma_{11} = \gamma_{12} = 1.25$ $\gamma_{21} = \gamma_{22} = 3.75$	A & A	−.8642	.9679
	A & B	−.0831	.0484
	B & B	−.7838	.8555
$\gamma_{11} = 1$, $\gamma_{21} = 20$ $\gamma_{12} = 2$, $\gamma_{22} = 40$	A & A	−.9280	.9496
	A & B	−.0040	.0020
	B & A	−.0089	.0061
	B & B	−.2019	.2116

Available final correlations for the first skewness-kurtosis combination ($\gamma_{11} = \gamma_{12} = 0$ and

$\gamma_{21} = \gamma_{22} = 25$) and all transformation combinations are plotted in Figure 30:
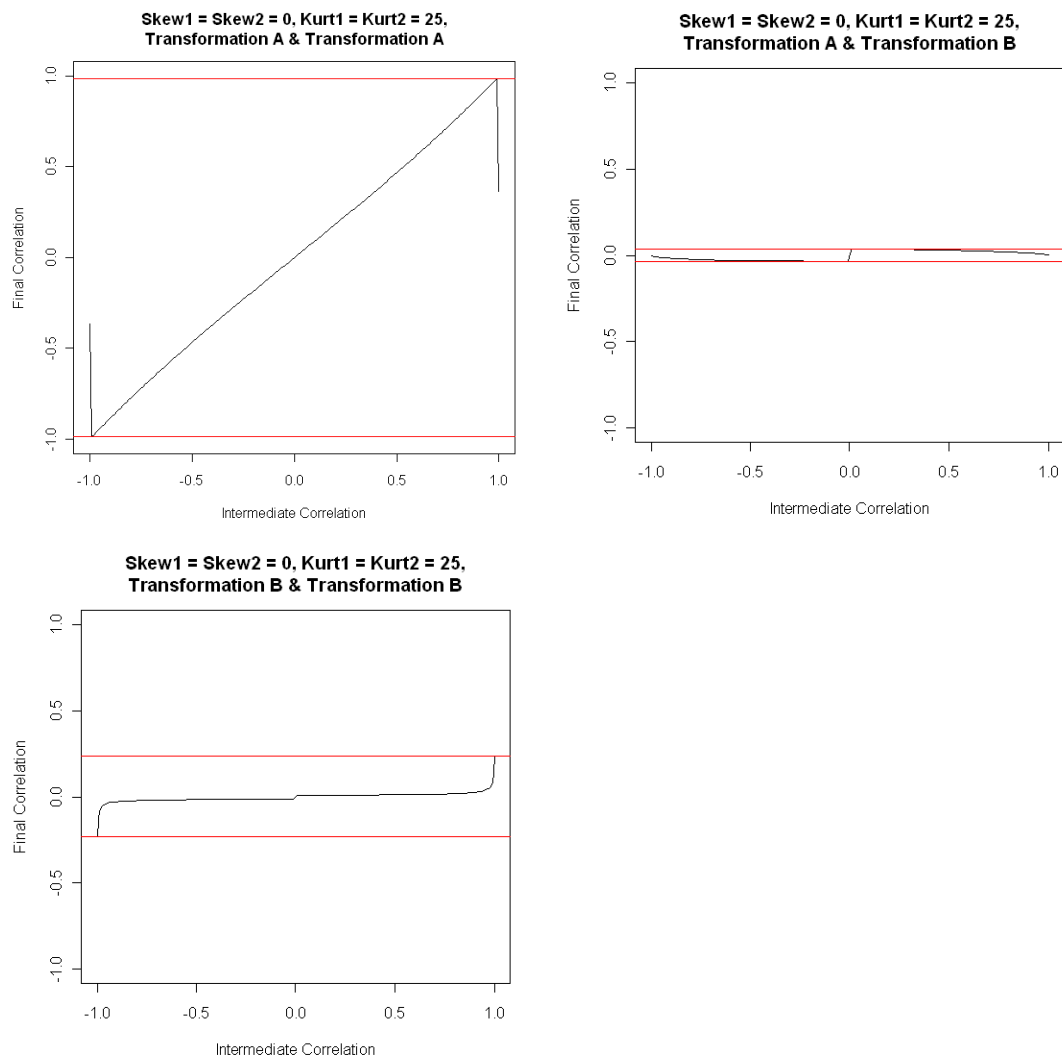
Figure 30 (a) – (c): Range of Final Correlations for Different Sets of g-and-h Transformations

Only bivariate distributions created with Transformation A for both variables offer a large, almost perfect range of final correlations. From Figure 30(b) and (c), we can see that Transformation B limits the range of available final correlations (the range in Figure 30(c) could be slightly extended as the range of the intermediate correlation was not fully exploited, but will never reach as far as the range for Transformation A in Figure 30(a)). For bivariate distributions created with at least one set of coefficients from the Transformation B column, some slight

numerical instability could be observed. Together with the odd shape already observed for univariate distributions, it is safe to disregard any distribution created with a Transformation B. however, notice that the nonlinearity in even Figure 30(a), which onvolves only Transformation A, could be problematic when trying to numerically solve for ρ_Z. Figure 31 shows the largest possible range of final correlations for *g*-and-*h* distributions with $\gamma_{11} = \gamma_{12} = \sqrt{8}$ and $\gamma_{21} = \gamma_{22} = 12$ and $\gamma_{11} = 1$, $\gamma_{12} = 2$, $\gamma_{21} = 20$, and $\gamma_{22} = 40$, both created with their respective Transformation A. Compare these to Figure 17 for the Vale & Maurelli method and Figure 24 for the 5th order polynomial method.
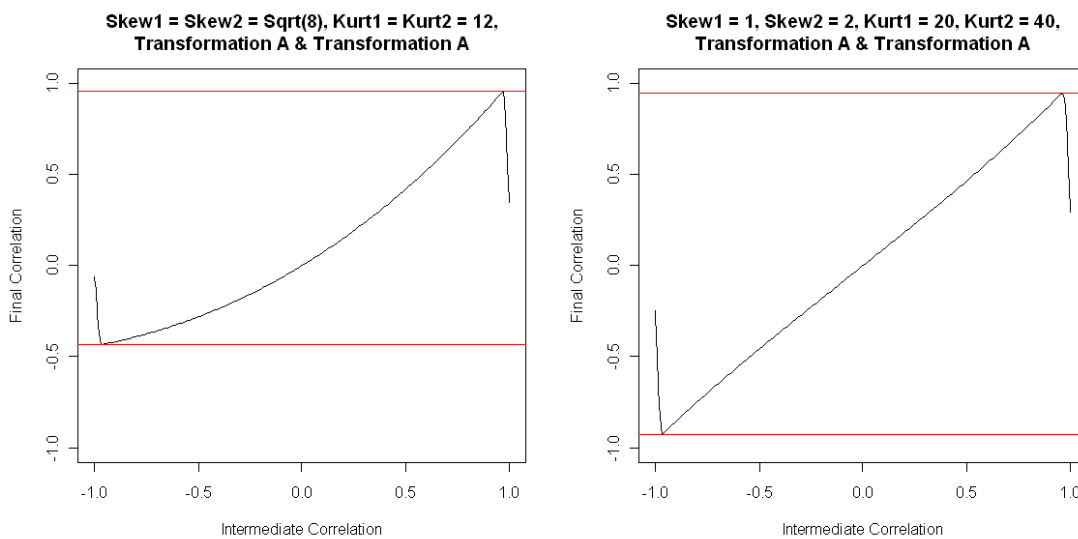
Figure 31: Ranges of Final Correlations Available for Various Bivariate g-and-h Distributions

How do the ranges of final correlations available for the *g*-and-*h* distribution compare to the ranges for the V&M method?

Table 15: Range of Final Correlations – V&M and g-and-h Comparison

	V&M		g-and-h
	Best	Worst	
$\gamma_{11}=0$, $\gamma_{21}=25$ $\gamma_{12}=0$, $\gamma_{22}=25$	$[-1,\ \ 1]$	$[-.1796,\ \ .1796]$	$[-.9882,\ \ .9882]$
$\gamma_{11}=\sqrt{8}$, $\gamma_{21}=12$ $\gamma_{12}=\sqrt{8}$, $\gamma_{22}=12$	$[-.0958,\ \ 1]$	$[0,\ \ .7364]$	$[-.4346,\ \ .9605]$
$\gamma_{11}=1$, $\gamma_{21}=20$ $\gamma_{12}=2$, $\gamma_{22}=40$	$[-.9481\ \ .9813]$	$[-.1251,\ \ .1432]$	$[-.9280\ \ .9496]$

The *g-and-h* distribution seems to sometimes span a range of final correlations somewhat below the range of the 3rd order power method, but it also offers a significantly wider range for other skewness-kurtosis combinations, such as for $\gamma_1 = \sqrt{8}$, $\gamma_2 = 12$.

The g-and-h Transformation and Odd–shaped Distributions. We have seen that distributions created with the 3rd or 5th order power method can occasionally be severely odd–shaped (see Figure 18(c) or Figure 25). Does the *g-and-h* distribution produce odd–shaped distributions as well? Among others, we have plotted bivariate distributions for the 3rd order power method with the following parameter choices:

   a) $\gamma_{11}=\gamma_{12}=0$, $\gamma_{21}=\gamma_{22}=25$, $\rho=.3$

   b) $\gamma_{11}=2.5$, $\gamma_{12}=1.4$, $\gamma_{21}=11.5$, $\gamma_{22}=5.6$, $\rho=.46$

Transformation coefficients for the corresponding *g-and-h* distributions with the same parameters can be found in Table 13. Distributions created with Transformation A are plotted in the left column, distributions created with Transformation B are plotted in the right column of Figure 32:
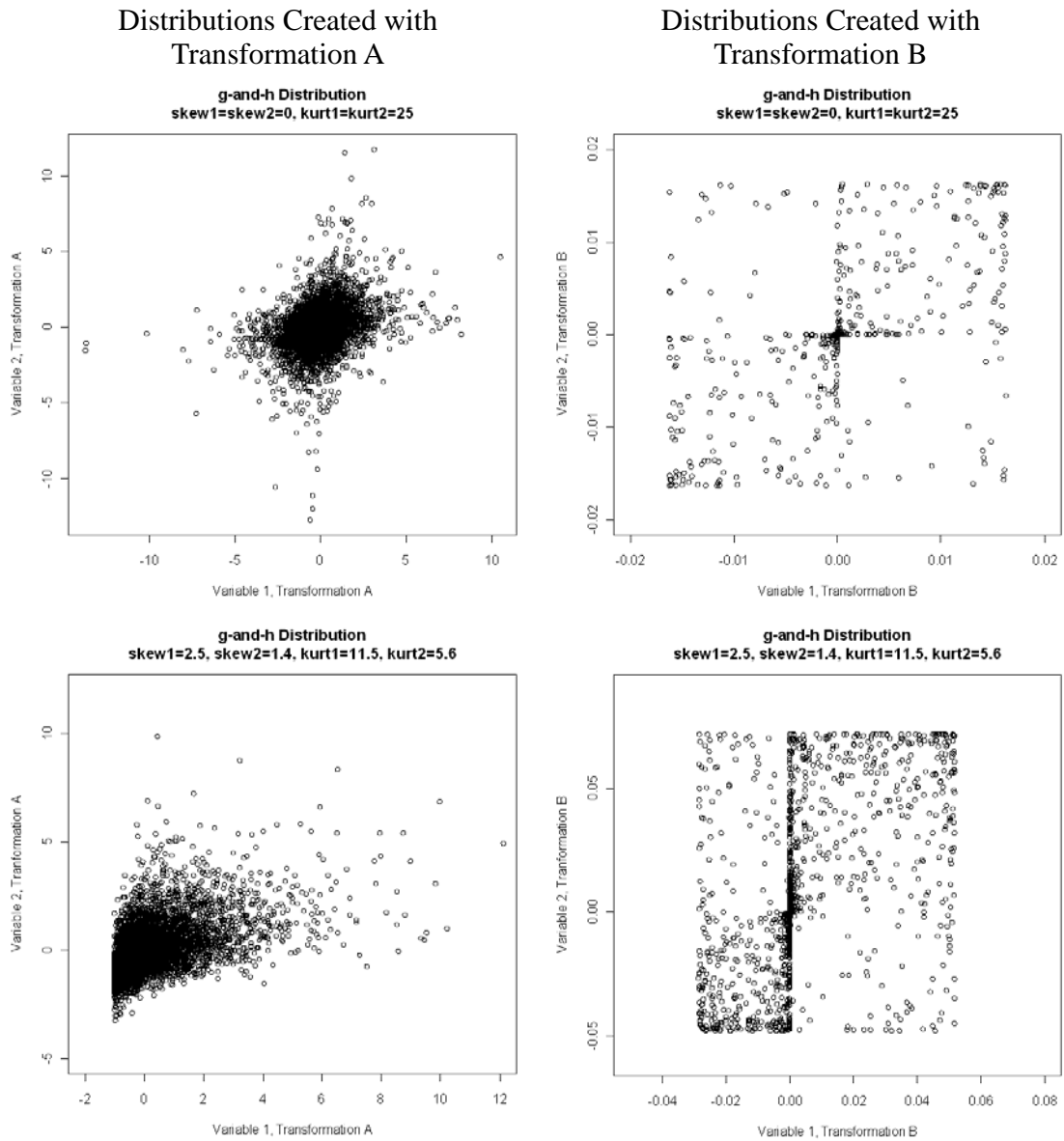
Figure 32(a) through (d): Various g-and-h distributions

Distributions created from Transformation B are unrepresentative of bivariate distributions typically observed in practice. At the same time, the shapes of the *g*-and-*h* distributions created from Transformation A seem relatively reasonable. Perhaps even more importantly, they are relatively similar to the well–behaved V&M distributions from Figure 12(b) and Figure 18(d).

For any following analyses, distributions created from Transformation B will be dropped from consideration, and only distributions created from Transformation A will be discussed.

But even if we only use distributions created from Transformation A, care is needed. Consider a bivariate distribution with $\gamma_{11} = 0$, $\gamma_{21} = -0.5$ and $\gamma_{12} = 0$, $\gamma_{22} = -1.5$. The two sets of coefficients are: $g_1 = 0$, $h_1 = -0.054383$, and $g_2 = 0$, $h_2 = -0.595744$. Further, we want $\rho_Y = .85$, e.g. for some robustness tests on reliability coefficients. The resulting *g*-and-*h* distribution will look like this:
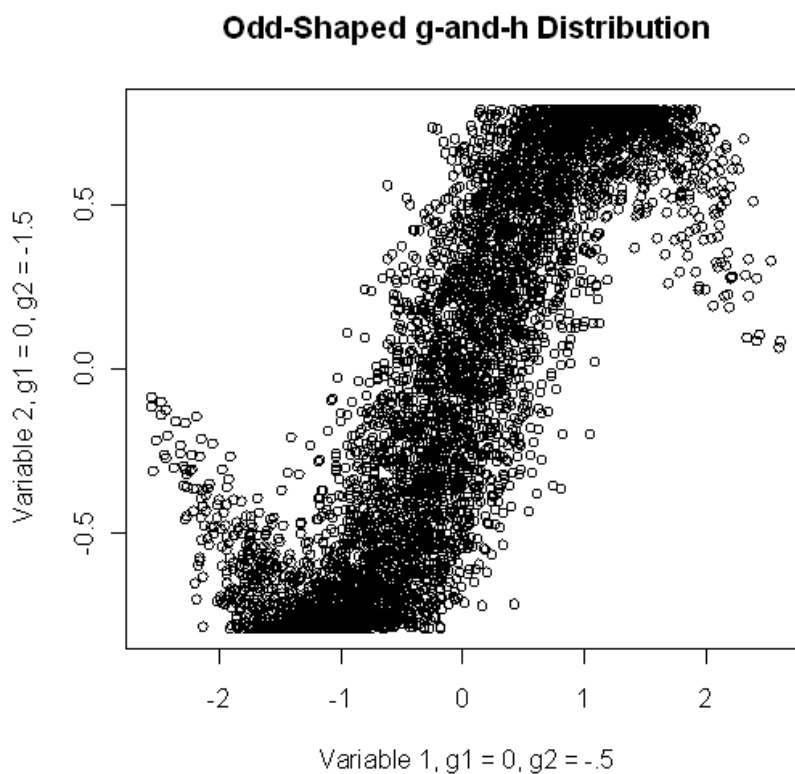
Figure 33: Odd-shaped g-and-h Distribution

94

Summary of the g-and-h Distribution

The *g*-and-*h* distribution was developed by Tukey around 1977 to provide a relatively

simple transform that would allow simulation of a wide range of distributions (e.g., covering a

large part of the skewness-kurtosis plane). While initially the main concern for the application of

the *g*-and-*h* distribution was to fit quantiles of nonnormal distributions, moment fitting is just as

feasible. For this study, we are focusing on moment fitting, as this has been a very popular

technique for creating nonnormal data with any desired skewness-kurtosis combination, using

the third and fourth moment of a nonnormal distribution as the only characteristics to control.

The *g*-and-*h* distribution has two solutions to the set of equations that relate the

transformation coefficients *g* and *h* to skewness and kurtosis. However, one of these coefficient

sets (labeled Transformation B in the sections above) will usually lead to transformations that

reproduce the third and fourth moment as requested, but with an overall rather odd-shaped

distribution and small range of available final correlations. Hence, it is safe to disregard

coefficients for Transformation B, leaving only one set of coefficients for Transformation A. It

seems that the range of correlations that can be simulated with the *g*-and-*h* distribution for a

given choice of marginal skewnesses and kurtoses tends to be larger than what is available for

the 3rd order polynomial method. Even multivariate *g*-and-*h* distributions from Transformation A

can be odd-shaped.

Despite some advantages over the power method, numerical optimization and integration

for the *g*-and-*h* uni- and multivariate distributions is decidedly more challenging than what is

required for the 3rd order power method. For the applied psychometrician the sensitivity of the *g*-

and-*h* distribution's results to settings of numerical procedures might very well present an

obstacle. As of today, no problem-free software solution to simulate *g*-and-*h* distributions is available.

Empirical Evaluation of Simulation Methods

When we draw a sample from a nonnormal distribution, the sample values for skewness and kurtosis, $\hat{\gamma}_1$ and $\hat{\gamma}_2$, will not have their nominal value but be subjected to sampling variability. Little is known about the sampling distributions of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ and whether these sampling distributions will be the same when different simulation methods are employed. For example, if we repeatedly draw samples of size $N = 20$ from a distribution with $\gamma_1 = 0$ and $\gamma_2 = 25$ using the 3rd order polynomial transform, what will the average sample skewness and kurtosis, $\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$, be? How much will they vary around their mean? If we draw a large number of samples, we can use $\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$ to estimate the expected value of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ at a given sample size, while $SD(\hat{\gamma}_1)$ and $SD(\hat{\gamma}_2)$ will estimate the standard error of $\hat{\gamma}_1$ and $\hat{\gamma}_2$.

Expected Value and Sampling Variability of $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\rho}$ for the 3rd Order Polynomial Transform

Attempts to evaluate the expected values of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ at given sample sizes for distributions created by the power method have been made. For example, Sharma, Durvasula & Dillon (1989) write that "…kurtosis tended, on average, to be underestimated in conditions characterized by smaller sample sizes." However, the authors provide no further explanations, demonstrations, or simulation results.

Harwell & Serlin (1989) report sample skewness and kurtosis (table 3, page 360) as well as average correlations (table 2, page 359) for samples with $N = 20$, $N = 40$, and $N = 100$, based on 4,000 replications each, from distributions with

I. $\gamma_1 = 0, \gamma_2 = 0$ (normal),

II. $\gamma_1 = 0, \gamma_2 = -1.12$ (symmetric and platykurtic),

III. $\gamma_1 = 0, \gamma_2 = 3$ (symmetric and leptokurtic),

IV. $\gamma_1 = 0, \gamma_2 = 25$ (symmetric and extremely leptokurtic), and

V. $\gamma_1 = 2, \gamma_2 = 6$ (heavily skewed).

Results in their table 3 seem to imply that for all four nonnormal distributions, $\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$ and based on all described sample sizes, showed very good congruence with the nominal values of γ_1 and γ_2, respectively. For example, for distribution IV with $\gamma_1 = 0, \gamma_2 = 25$, they report average sample skewness and kurtosis $\bar{\hat{\gamma}}_1 = -.141$ and $\bar{\hat{\gamma}}_2 = 22.80$ for $N = 20$. This seems to be quite satisfactory performance, and for all other distributions and sample sizes, $\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$ are even closer to their nominal values.

For average correlations, Harwell & Serlin observe a slight failure to reach nominal levels for both $\rho_Y = .3$ and $\rho_Y = .7$. For example, for distribution IV and $N = 100$, they report $\bar{r} = .62$ when $\rho_Y = .7$ is desired. Generally, average sample correlations seem to lie somewhat below their nominal level for all nonnormal distributions. Harwell & Serlin further describe (page 359) that they conducted additional simulations with larger sample sizes (up to $N = 400$) which showed increasing correspondence between sample correlation and nominal value.

However, there are a few problems with Harwell & Serlin's assessments. Several

publications (see Johnson & Lowe, 1979; Dalen, 1987; Kirby, 1974) contain derivations of

maximum possible sample skewness and kurtosis for a given sample size. Kirby (1974) provides

the most stringent upper boundary for sample skewness for any sample of size N:

$$\hat{\gamma}_1 \leq \frac{N-2}{\sqrt{N-1}} \qquad (80)$$

While Johnson & Lowe (1979) find an upper bound to sample kurtosis of $N-3$, Dalen (1987)

improves on this by deriving

$$\hat{\gamma}_2 \leq \frac{N^2 - 3N + 3}{N-1} - 3 = \frac{N^2 - 6(N-1)}{N-1} \qquad (81)$$

Therefore, we have the following boundaries for sample skewnesses and kurtoses for sample

sizes employed in Harwell & Serlin's study:

Table 16: Boundaries for Sample Skewness & Kurtosis

	$\hat{\gamma}_1$		$\hat{\gamma}_2$	
$N = 20$	–4.1295	4.1295	–2.0000	15.0526
$N = 40$	–6.0849	6.0849	–2.0000	35.0256
$N = 100$	–9.8494	9.8494	–2.0000	95.0101

This clearly implies that it would be impossible for Harwell & Serlin to observe an *individual*

sample kurtosis of $\hat{\gamma}_2 = 22.80$, let alone an average sample kurtosis of $\bar{\hat{\gamma}}_2 = 22.80$ for samples of

size $N = 20$. Their table 3 must contain simulation results for something other than $\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$

for $N = 20$. One possibility is that they – perhaps accidentally – treated their simulated data as

one large sample of 80,000 data points, calculating $\hat{\gamma}_1$ and $\hat{\gamma}_2$ for this one large sample. In that

case, Harwell & Serlin's table 3 would not show $\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$ based on 4,000 samples of sizes

$N = 20$, $N = 40$, and $N = 100$, but $\hat{\gamma}_1$ and $\hat{\gamma}_2$ for one large sample with $N = 80,000$,

$N = 160,000$, and $N = 400,000$, respectively. Following this assumption, I was able to

reproduce their table 3, and results from my simulations are displayed in Table 17. Remembering

that per skewness-kurtosis combination, there are two distinctly different sets of transformation

coefficients, I report results obtained with both sets of coefficients; within each cell, results from

the monotonic transformation (where available) are reported first and results from the non-

monotonic transformation second, in italics. When both transformations are non-monotonic, the

one with coefficients closer to [0   1   0   0] are reported first. Harwell & Serlin's results are

included on the last line of each row in red.

Table 17: Empirical Skewnesses and Kurtoses

N	$\gamma_1 = 0, \gamma_2 = -1.12$		$\gamma_1 = 0, \gamma_2 = 3$		$\gamma_1 = 0, \gamma_2 = 25$		$\gamma_1 = 2, \gamma_2 = 6$	
	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
80,000	−0.001	−1.126	0.041	3.032	0.072	22.321	2.017	6.099
	−0.006	*−1.218*	*0.015*	*3.651*	*0.002*	*13.696*	*2.004*	*6.060*
	0.001	−1.16	−0.007	3.06	−0.141	22.80	2.02	6.20
160,000	0.001	−1.128	0.021	2.954	−0.115	23.721	2.005	6.041
	0.005	*−1.132*	*−0.092*	*2.905*	*−0.156*	*31.482*	*1.979*	*5.561*
	−0.006	−1.15	−0.008	2.89	0.009	27.24	2.00	6.10
400,000	−0.001	−1.104	−0.032	2.982	−0.076	24.320	1.979	5.796
	0.003	*−1.148*	*−0.030*	*4.299*	*−0.125*	*22.391*	*1.999*	*6.044*
	0.002	−1.16	−0.006	2.98	−0.002	24.48	2.00	5.99
Average	0.000	−1.119	0.010	2.989	−0.040	23.454	2.000	5.979
	0.001	*−1.166*	*−0.036*	*3.618*	*−0.093*	*22.523*	*1.994*	*5.905*
	−0.001	−1.16	−0.007	2.98	−0.018	24.84	2.01	6.10

Sample skewness and kurtosis based on $N = 80,000$ or more show good agreement with Harwell & Serlin's results in table 3 and are not restricted by Equations (80) or (81).

I also attempted to replicate their results on average correlations by simulating samples of sizes 20, 40, and 100 and calculating average sample correlations $\overline{r}_{20}$, $\overline{r}_{40}$, and $\overline{r}_{100}$. Table 18 contains results based on 4,000 replications, as in Harwell & Serlin (1989; their results are included in red for comparison) as well as based on 100,000 replications, a figure more easily obtained with the increased computing power available today. As in Table 17, results for both distinctly different transformations are presented, with the results for the monotonic transformation first, if available.

Table 18: Sample Skewnesses and Kurtoses

12	Popula-tion Value	Normal	$\gamma_1 = 0,$ $\gamma_2 = -1.12$	$\gamma_1 = 0,$ $\gamma_2 = 3$	$\gamma_1 = 0,$ $\gamma_2 = 25$	$\gamma_1 = 2,$ $\gamma_2 = 6$
20	$\rho = .3$	.296 / .29	.295 / .295 / .26	.296 / .302 / .29	.324 / .327 / .25	.293 / .288 / .26
	$\rho = .7$	.689 / .69	.691 / .692 / .66	.694 / .706 / .68	.708 / .706 / .63	.692 / .676 / .65
40	$\rho = .3$	.294 / .30	.300 / .298 / .27	.301 / .309 / .29	.318 / .253 / .24	.305 / .294 / .26
	$\rho = .7$	.695 / .70	.698 / .700 / .66	.698 / .705 / .69	.711 / .708 / .63	.697 / .691 / .66
100	$\rho = .3$	.301 / .30	.299 / .299 / .27	.299 / .303 / .29	.311 / .313 / .24	.302 / .298 / .26
	$\rho = .7$	.699 / .70	.698 / .699 / .66	.698 / .703 / .69	.704 / .705 / .62	.700 / .698 / .66

Number of replications: 4,000; values obtained by Harwell & Serlin in red

N	Popula-tion Value	Normal	$\gamma_1 = 0,$ $\gamma_2 = -1.12$	$\gamma_1 = 0,$ $\gamma_2 = 3$	$\gamma_1 = 0,$ $\gamma_2 = 25$	$\gamma_1 = 2,$ $\gamma_2 = 6$
20	$\rho = .3$	.293	.295 / .297	.296 / .310	.324 / .333	.298 / .289
	$\rho = .7$	.690	.693 / .695	.693 / .704	.714 / .709	.691 / .679
40	$\rho = .3$	.297	.299 / .299	.299 / .309	.317 / .324	.301 / .296
	$\rho = .7$	.695	.696 / .697	.698 / .704	.709 / .706	.696 / .690
100	$\rho = .3$	.299	.299 / .300	.300 / .306	.309 / .314	.300 / .299
	$\rho = .7$	.698	.699 / .699	.699 / .703	.705 / .703	.699 / .697

Number of replications: 100,000

Neither for monotonic nor non-monotonic transformations, for 4,000 or 100,000 replications, was I able to observe the deviations from nominal levels of the magnitude reported in Harwell & Serlin (1989). Only for distribution V, with $\gamma_1 = 2$ and $\gamma_2 = 6$ does $\bar{r}$ fall slightly below its nominal value.

Expected Value and Sampling Variability of $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\rho}$ for the g–and–h Distribution

Estimates of $E(\hat{\gamma}_1)$ and $E(\hat{\gamma}_2)$ for samples of size $N = 10$, $N = 30$, $N = 50$, and $N = 100$ for the *g-and-h* distribution are presented by Headrick (2002, table 3, page 700) and Kowalchuk & Headrick (2010, table 6, page 70). Kowalchuk & Headrick calculate $\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$, based on 50,000 replications, for the following distributions:

I. $\gamma_1 = 3$, $\gamma_2 = 20$
II. $\gamma_1 = 1.5$, $\gamma_2 = 10$
III. $\gamma_1 = .75$, $\gamma_2 = 5$
IV. $\gamma_1 = .25$, $\gamma_2 = 1$

They do not report sampling variability of $\hat{\gamma}_1$ and $\hat{\gamma}_2$. According to their table 6, $\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$ fall fairly close to their nominal values, the largest deviation being observed for $\bar{\hat{\gamma}}_2 = 17.816$, when $\gamma_2 = 20$ (distribution I). However, remembering the derivations presented in Equation (80) and (81), we can construct a table similar to Table 16 for the sample sizes selected in Kowalchuk & Headrick's study:

Table 19: Boundaries for Sample Skewness and Kurtosis

	$\hat{\gamma}_1$		$\hat{\gamma}_2$	
$N = 10$	−2.6667	2.6667	−2.0000	5.1111
$N = 30$	−5.1995	5.1995	−2.0000	25.0345
$N = 50$	−6.8571	6.8571	−2.0000	45.0204
$N = 100$	−9.8494	9.8494	−2.0000	95.0101

For a sample size of $N = 10$, they cannot have observed $\overline{\hat{\gamma}}_1 = 2.958$ and $\overline{\hat{\gamma}}_2 = 17.816$ as reported in table 6 for distribution I. It may be that Kowalchuk & Headrick's table 6 is the result of something similar to what seems to have happened in Harwell & Serlin (1989): They may have combined *all* simulated data points and calculated $\hat{\gamma}_1$ and $\hat{\gamma}_2$ for one large sample instead of calculating $\overline{\hat{\gamma}}_1$ and $\overline{\hat{\gamma}}_2$ by averaging sample skewnesses and kurtoses across 50,000 samples. Headrick's (2002) Table 3 also suffers from an inconsistency between a supposedly observed average sample kurtosis of $\overline{\hat{\gamma}}_2 = 5.973$ for a sample size of $N = 10$ and a maximum possible sample kurtosis of 5.1111 at such sample size.

Simulation of Expected Value and Variability of Sample Skewness and Kurtosis

The smaller the sample, the narrower are the bounds on the possible values for $\hat{\gamma}_1$ and $\hat{\gamma}_2$. As sample size increases, skewness and kurtosis have a larger range to vary within, which is due to the fact that more scores are available, and therefore, individual scores can stand out further from the rest of the scores, thereby creating extreme skewness or kurtosis.

In light of the previous section, it seems very likely that estimates for $E(\hat{\gamma}_1)$ and $E(\hat{\gamma}_2)$ for small to moderate samples simulated with the 3rd order power method or the *g-and-h* distribution have not been reported correctly in the literature up to this point. Therefore, I present

new simulation results for eight bivariate IM distributions with popular skewness-kurtosis combinations:

I. $\gamma_1 = 0$, $\gamma_2 = 25$ (in Harwell & Serlin)

II. $\gamma_1 = 0$, $\gamma_2 = 3$ (in Harwell & Serlin)

III. $\gamma_1 = 1$, $\gamma_2 = 1$

IV. $\gamma_1 = 1.75$, $\gamma_2 = 3.75$

V. $\gamma_1 = 2$, $\gamma_2 = 6$ (in Harwell & Serlin)

VI. $\gamma_1 = 3$, $\gamma_2 = 21$ (similar to $\gamma_1 = 3$, $\gamma_2 = 20$ in Kowalchuk & Headrick)

VII. $\gamma_1 = -1.25$ $\gamma_2 = 3.75$

VIII. $\gamma_1 = 2$, $\gamma_2 = 40$

Samples for each skewness-kurtosis combination can be simulated in three ways: Using one of the two distinctly different Fleishman transformations or using the g-and-h distribution. To estimate $E(\hat{\gamma}_1)$ and $E(\hat{\gamma}_2)$, 100,000 samples of sizes $N_1 = 40$ and $N_2 = 100$ were drawn from the three different nonnormal distributions (two distinctly different 3rd order power method distributions and one g-and-h distribution) and the sample skewnesses and kurtoses calculated. Table 20 through Table 27 contain mean values, standard deviations, skewnesses and kurtoses for the empirical sampling distributions of both $\hat{\gamma}_1$ and $\hat{\gamma}_2$. For each skewness-kurtosis combination, cells with relatively good performance are colored in different shades of green: If the estimate for sample skewness or kurtosis is within $\pm 10\%$ of the respective population value (e.g., with $\gamma_1 = 2$, any observed value for $\overline{\hat{\gamma}}_1$ between 1.8 and 2.2), darker green is used, if it is within $\pm 50\%$ of the population value (e.g. $\overline{\hat{\gamma}}_1 \in [1.0, 3.0]$ when $\gamma_1 = 2$), a lighter green is used. These guidelines were chosen as opposed to confidence intervals of, e.g., ± 2 standard error estimates around the observed value to safeguard against "excessive power." With 100,000 replications, such confidence intervals would be quite narrow and less informative than the rules chosen above. For the special case of $\gamma_1 = 0$ or $\gamma_2 = 0$, the intervals $[-.05, .05]$ and $[-.25, .25]$ are

103

chosen. The worst departure from either asymptotic skewness or kurtosis is colored in a reddish

shade.

Table 20: First Four Moments of Sample Skewness and Kurtosis when $\gamma_1 = 0$ and $\gamma_2 = 25$

Skewness and Kurtosis values; Simulation method & Coefficients				$\overline{\hat{\gamma}}$	S_γ	Skewness of $\hat{\gamma}$	Kurtosis of $\hat{\gamma}$
$\gamma_1 = 0$, $\gamma_2 = 25$	mV&M: $a_1 = 0.2553$, $a_2 = 0$, $a_3 = 0.2038$	$N = 40$	γ_1	0.006	1.876	0.000	−0.111
			γ_2	6.594	5.317	1.602	2.770
		$N = 100$	γ_1	−0.001	1.996	−0.019	0.828
			γ_2	11.257	8.934	2.350	7.626
	nmV&M: $a_1 = -1.5667$, $a_2 = 0$, $a_3 = 0.3482$	$N = 40$	γ_1	0.005	1.003	−0.047	8.531
			γ_2	0.396	4.432	3.555	13.930
		$N = 100$	γ_1	−0.005	1.383	−0.041	8.114
			γ_2	2.957	9.195	3.571	15.677
	g-and-h: $g = 0$, $h = 0.1930$	$N = 40$	γ_1	0.002	1.144	−0.016	2.484
			γ_2	2.609	3.696	2.869	11.048
		$N = 100$	γ_1	0.005	1.254	0.006	6.097
			γ_2	4.829	6.671	4.293	26.901

Table 21: First Four Moments of Sample Skewness and Kurtosis when $\gamma_1 = 0$ and $\gamma_2 = 3$

Skewness and Kurtosis values; Simulation method & Coefficients			$\overline{\hat{\gamma}}$	S_γ	Skewness of $\hat{\gamma}$	Kurtosis of $\hat{\gamma}$
$\gamma_1 = 0$, $\gamma_2 = 3$	mV&M $a_1 = 0.7824$, $a_2 = 0$, $a_3 = 0.0679$	$N = 40$	γ_1	0.001	0.778	−0.002	1.839
			γ_2	1.382	2.064	2.748	12.187
		$N = 100$	γ_1	−0.005	0.650	0.007	3.087
			γ_2	2.108	2.346	3.678	25.063
	nmV&M $a_1 = 1.5504$, $a_2 = 0$, $a_3 = -0.2594$	$N = 40$	γ_1	0.001	0.510	0.105	25.361
			γ_2	−0.872	2.049	7.240	63.488
		$N = 100$	γ_1	0.001	0.623	−0.128	34.584
			γ_2	−0.354	4.036	7.566	75.678
	g-and-h $g = 0$, $h = 0.1089$	$N = 40$	γ_1	−0.003	0.737	−0.016	2.983
			γ_2	1.131	2.058	3.378	18.965
		$N = 100$	γ_1	−0.003	0.644	−0.025	6.137
			γ_2	1.838	2.598	5.110	48.329
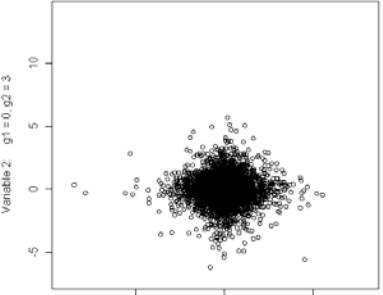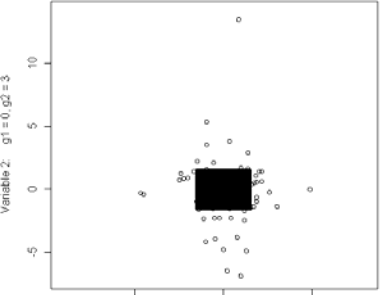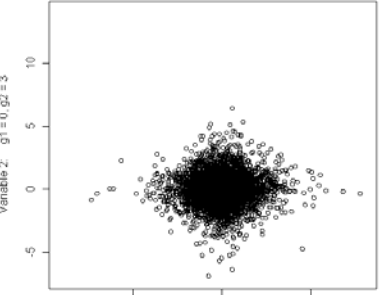
Table 22: First Four Moments of Sample Skewness and Kurtosis when $\gamma_1 = 1$ and $\gamma_2 = 1$

Skewness and Kurtosis values; Simulation method & Coefficients				$\overline{\hat{\gamma}}$	S_γ	Skewness of $\hat{\gamma}$	Kurtosis of $\hat{\gamma}$
$\gamma_1 = 1$, $\gamma_2 = 1$	nmV&M $a_1 = -1.0175$, $a_2 = 0.1910$, $a_3 = 0.0186$	$N = 40$	γ_1	0.871	0.156	0.713	1.209
			γ_2	0.462	1.903	2.181	8.102
		$N = 100$	γ_1	0.946	0.078	0.671	1.159
			γ_2	0.758	1.341	2.040	9.310
	nmV&M $a_1 = 1.2164$, $a_2 = 0.3446$, $a_3 = -0.1366$	$N = 40$	γ_1	0.882	0.359	2.202	13.700
			γ_2	0.008	1.412	6.774	74.981
		$N = 100$	γ_1	0.930	0.343	5.318	50.942
			γ_2	0.236	2.278	10.456	156.993
	g-and-h $g = 0.4168$, $h = -0.0746$	$N = 40$	γ_1	0.874	0.392	0.642	1.013
			γ_2	0.499	1.354	2.092	7.656
		$N = 100$	γ_1	0.949	0.272	0.511	0.595
			γ_2	0.783	1.081	1.633	4.949
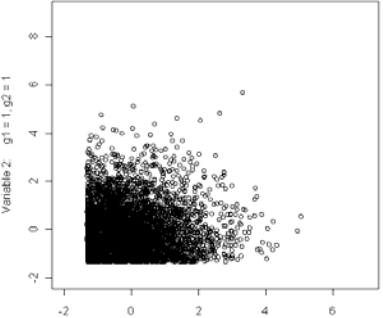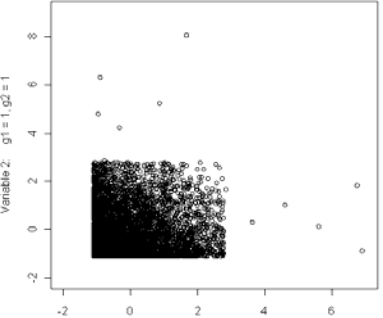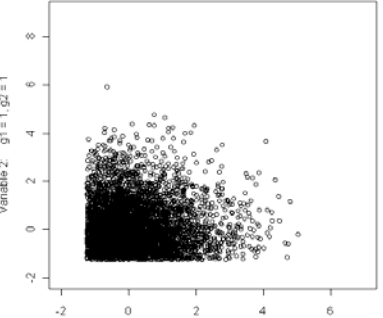
Table 23: First Four Moments of Sample Skewness and Kurtosis when $\gamma_1 = 1.75$ and $\gamma_2 = 3.75$

Skewness and Kurtosis values; Simulation method & Coefficients				$\overline{\hat{\gamma}}$	S_{γ}	Skewness of $\hat{\gamma}$	Kurtosis of $\hat{\gamma}$
$\gamma_1 = 1.75$, $\gamma_2 = 3.75$	nmV&M $a_1 = -0.9208$, $a_2 = 0.4868$, $a_3 = 0.0725$	$N = 40$	γ_1	1.506	0.253	0.952	1.979
			γ_2	2.117	6.028	2.130	7.860
		$N = 100$	γ_1	1.638	0.160	1.592	8.075
			γ_2	2.880	6.377	4.124	40.329
	nmV&M $a_1 = 0.9297$, $a_2 = 0.3995$, $a_3 = -0.0365$	$N = 40$	γ_1	1.494	0.517	0.901	1.461
			γ_2	2.144	2.518	1.953	5.864
		$N = 100$	γ_1	1.637	0.402	0.951	1.971
			γ_2	2.977	2.417	2.136	8.490
	g-and-h $g = 0.8339$, $h = -0.1680$	$N = 40$	γ_1	1.492	0.516	0.805	1.170
			γ_2	2.208	2.495	1.836	5.204
		$N = 100$	γ_1	1.643	0.395	0.739	1.036
			γ_2	3.067	2.311	1.717	5.129
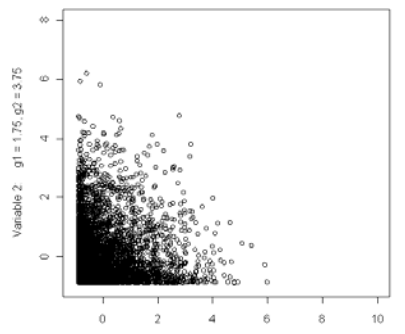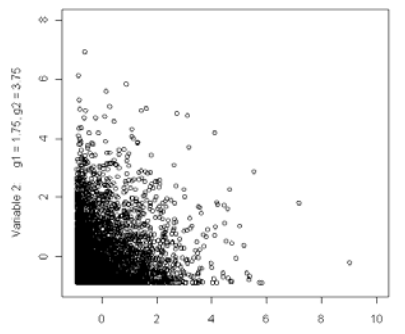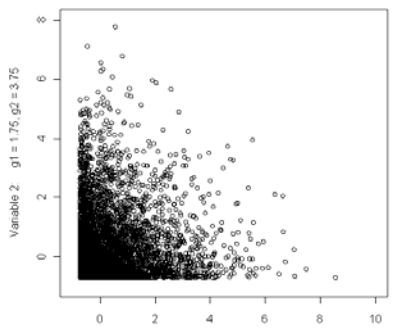
Table 24: First Four Moments of Sample Skewness and Kurtosis when $\gamma_1 = 2$ and $\gamma_2 = 6$

Skewness and Kurtosis values; Simulation method & Coefficients				$\overline{\hat{\gamma}}$	S_γ	Skewness of $\hat{\gamma}$	Kurtosis of $\hat{\gamma}$
$\gamma_1 = 2$, $\gamma_2 = 6$	nmV&M $a_1 = -0.8157$, $a_2 = 0.5695$, $a_3 = 0.0878$	$N = 40$	γ_1	1.621	0.317	1.215	2.998
			γ_2	2.610	8.859	2.422	9.516
		$N = 100$	γ_1	1.794	0.270	2.398	12.046
			γ_2	3.759	15.035	4.679	36.923
	nmV&M $a_1 = 0.8263$, $a_2 = 0.3137$, $a_3 = 0.0227$	$N = 40$	γ_1	1.575	0.615	1.056	1.781
			γ_2	2.726	3.203	2.066	6.048
		$N = 100$	γ_1	1.784	0.543	1.385	3.793
			γ_2	4.109	3.769	2.752	13.329
	g-and-h $g = 0.7594$, $h = -0.0979$	$N = 40$	γ_1	1.569	0.622	1.019	1.627
			γ_2	2.764	3.219	2.022	5.691
		$N = 100$	γ_1	1.789	0.545	1.269	2.978
			γ_2	4.204	3.762	2.521	10.340
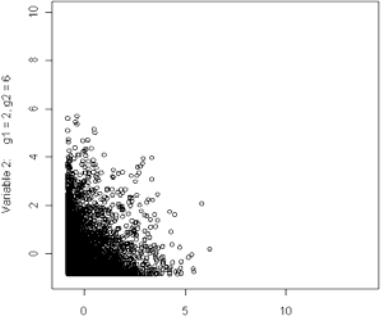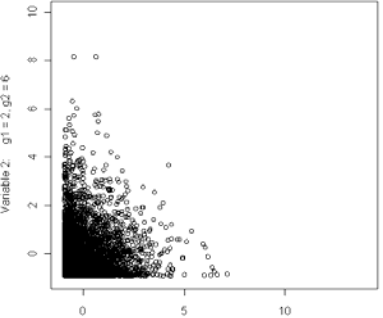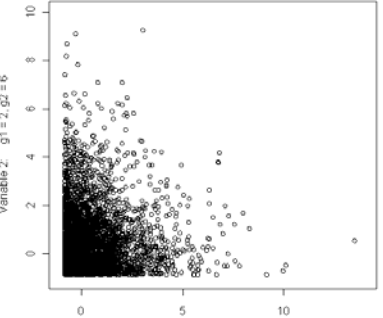
Table 25: First Four Moments of Sample Skewness and Kurtosis when $\gamma_1 = 3$ and $\gamma_2 = 21$

Skewness and Kurtosis values; Simulation method & Coefficients				$\overline{\hat{\gamma}}$	S_γ	Skewness of $\hat{\gamma}$	Kurtosis of $\hat{\gamma}$
$\gamma_1 = 3$, $\gamma_2 = 21$	mV&M $a_1 = 0.4186$, $a_2 = 0.2523$, $a_3 = 0.1476$	$N = 40$	γ_1	1.742	1.194	−0.272	1.713
			γ_2	5.774	5.178	1.668	3.106
		$N = 100$	γ_1	2.216	1.211	0.322	2.517
			γ_2	9.891	8.389	2.393	7.924
	nmV&M $a_1 = 0.6816$, $a_2 = 0.6371$, $a_3 = 0.1487$	$N = 40$	γ_1	1.713	0.797	1.649	3.225
			γ_2	3.361	4.740	2.448	6.840
		$N = 100$	γ_1	2.088	1.006	2.016	5.086
			γ_2	6.569	8.817	2.918	10.865
	g-and-h $g = 0.6542$, $h = 0.0242$	$N = 40$	γ_1	1.728	0.848	1.156	1.695
			γ_2	3.993	4.734	2.051	5.129
		$N = 100$	γ_1	2.149	0.950	1.810	5.118
			γ_2	7.299	7.999	3.037	13.316
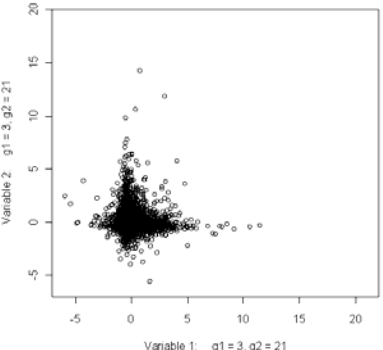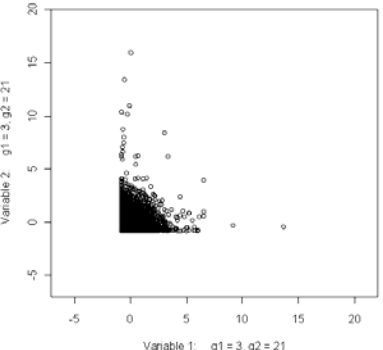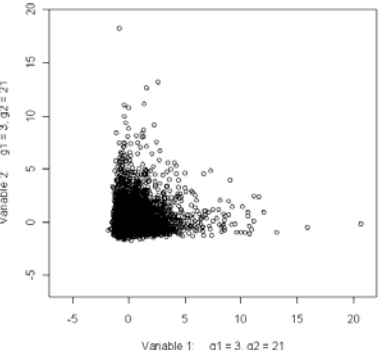
Table 26: First Four Moments of Sample Skewness and Kurtosis when $\gamma_1 = -1.25$ and $\gamma_2 = 3.75$

Skewness and Kurtosis values; Simulation method & Coefficients				$\overline{\hat{\gamma}}$	S_γ	Skewness of $\hat{\gamma}$	Kurtosis of $\hat{\gamma}$
$\gamma_1 = -1.25$ $\gamma_2 = 3.75$	mV&M $a_1 = 0.8189$, $a_2 = -0.1606$, $a_3 = 0.0492$	$N = 40$	γ_1	−0.923	0.646	−0.818	1.906
			γ_2	1.569	2.462	2.654	10.636
		$N = 100$	γ_1	−1.087	0.550	−1.295	4.214
			γ_2	2.503	2.976	3.589	23.032
	nmV&M $a_1 = 1.2091$, $a_2 = -0.4059$, $a_3 = -0.1663$	$N = 40$	γ_1	−0.952	0.468	−2.956	15.452
			γ_2	0.269	2.232	5.750	43.784
		$N = 100$	γ_1	−1.039	0.538	−4.367	27.622
			γ_2	0.874	4.018	6.660	59.670
	g-and-h $g = -0.3226$, $h = 0.0377$	$N = 40$	γ_1	−0.899	0.617	−1.187	2.775
			γ_2	1.320	2.466	2.947	13.062
		$N = 100$	γ_1	−1.066	0.548	−1.809	6.981
			γ_2	2.250	3.201	4.228	31.156

111
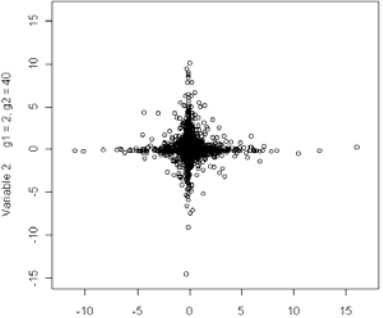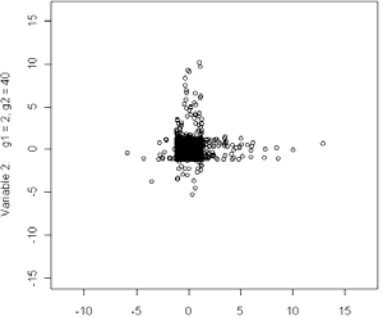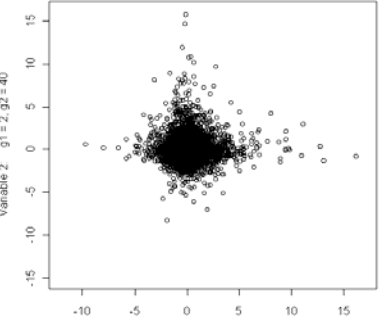
Table 27: First Four Moments of Sample Skewness and Kurtosis when $\gamma_1 = 2$ and $\gamma_2 = 40$

Skewness and Kurtosis values; Simulation method & Coefficients				$\overline{\hat{\gamma}}$	S_γ	Skewness of $\hat{\gamma}$	Kurtosis of $\hat{\gamma}$
$\gamma_1 = 2$, $\gamma_2 = 40$	mV&M $a_1 = 0.0616$, $a_2 = 0.1174$, $a_3 = 0.2421$	$N = 40$	γ_1	0.932	2.179	−0.419	−0.223
			γ_2	9.032	6.251	1.269	1.316
		$N = 100$	γ_1	1.239	2.381	−0.339	0.569
			γ_2	15.776	11.230	1.958	4.865
	nmV&M $a_1 = -1.4659$, $a_2 = 0.2058$, $a_3 = 0.3541$	$N = 40$	γ_1	0.517	1.169	0.770	5.143
			γ_2	1.299	5.420	2.779	8.026
		$N = 100$	γ_1	0.795	1.634	0.696	4.734
			γ_2	5.254	11.367	2.799	9.180
	g-and-h $g = 0.1935$, $h = 0.1757$	$N = 40$	γ_1	0.719	1.082	0.556	2.312
			γ_2	2.803	4.032	2.776	9.958
		$N = 100$	γ_1	1.003	1.183	1.328	5.443
			γ_2	5.316	7.454	3.919	21.180

112

Looking at Table 20 through Table 27 it becomes clear very quickly that it is rather an exception than the rule to find $\bar{\hat{\gamma}}_1 \approx \gamma_1$ or $\bar{\hat{\gamma}}_2 \approx \gamma_2$. In the vast majority of cases, $\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$ underestimate γ_1 and γ_2 in absolute value, often by a massive amount. As an example, compare $\gamma_1 = 2$ and $\gamma_2 = 40$ to $\bar{\hat{\gamma}}_1 = .7194$ and $\bar{\hat{\gamma}}_2 = 2.8028$ for $N = 40$ for the *g*-and-*h* distribution. It should be obvious that the results presented in Table 20 through Table 27 deviate gravely from Harwell & Serlin's (1989) table 3 or Kowalchuk & Headrick's (2010) table 6.

Sample skewness generally comes closer to its nominal value than sample kurtosis, with sometimes considerable variability between the different simulation methods. Generally, when a monotonic 3rd order polynomial transform is available, it seems to produce nonnormal samples that have skewness and kurtosis closest to the desired nominal values. In these situations, the non-monotonic transform can clearly be disregarded. When no monotonic transformation is available, non-monotonic 3rd order polynomial transform and *g*-and-*h* distribution perform similarly, with the *g*-and-*h* distribution performing slightly better. The sampling distributions of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ often differ considerably, depending on whether one of the 3rd order polynomial transformations or the *g*-and-*h* distribution is used.

Notice that we cannot include the 5th order polynomial method in these tables. For the eight distributions with arbitrarily chosen skewness-kurtosis combination, we have no guidelines on how to choose values for the 5th and 6th moment. We could only evaluate expected value and variability of sample skewness and kurtosis for the distributions discussed in Headrick (2002).

Why This Matters

The differences in shape at equal marginal skewnesses and kurtoses, as well as equal correlation, are not only of theoretical interest. If we recall from the introduction, the main

motivation for simulating nonnormal distributions in many fields of applied statistics is to assess the performance of statistical procedures under nonnormality. The question then becomes: Keeping the selection of marginal skewnesses and kurtoses and correlation constant, will additional differences in shape have an effect on the performance of statistical procedures?

The 3rd Order Polynomial Transformation in Published Research

The 3rd order polynomial transform has been the most popular procedure for simulating nonnormal distributions in psychometrics, and has received a considerable amount of attention and referencing in statistics in general. Vale & Maurelli's paper has been cited over a hundred times, including at least 64 studies which employed the method to simulate bi- or multivariate nonnormal data in Monte Carlo analyses. The question arises whether odd-shaped distributions similar to the ones in Figure 12(b) or Figure 18(c) have been utilized in published research and whether they may have affected results in robustness studies. Table 28 and Figure 34 summarize the skewness-kurtosis combinations utilized in the most relevant of these about 64 studies. The center column lists the skewness-kurtosis combinations for which a monotonic solution exists (which does not necessarily means that the monotonic solution was used), while the right column lists skewness-kurtosis combinations for which no monotonic solution exists.

Table 28: List of Studies with Skewness-Kurtosis Combinations that Employed the 3rd Order Polynomial Method

Authors	Skew & Kurt Combinations where monotonic transform exists	Skew & Kurt Combinations where no monotonic transform exists
Rausch & Kelley (2009)	$\gamma_1 = 0$, $\gamma_2 = 3.75$ $\gamma_1 = 1.25$, $\gamma_2 = 3.75$ $\gamma_1 = -1.25$, $\gamma_2 = 3.75$	$\gamma_1 = 1.25$, $\gamma_2 = 1.5$ $\gamma_1 = -1.25$, $\gamma_2 = 1.5$
Lix & Fouladi (2007)	$\gamma_1 = 0$, $\gamma_2 = 3$	$\gamma_1 = 6.2$, $\gamma_2 = 110.9$
Vallejo et al. (2007)	$\gamma_1 = 3$, $\gamma_2 = 21$	$\gamma_1 = 1$, $\gamma_2 = .75$ $\gamma_1 = 1.75$, $\gamma_2 = 3.75$
Wang & Thompson (2007)		$\gamma_1 = 1$, $\gamma_2 = 1$ $\gamma_1 = -1.5$, $\gamma_2 = 3.5$
Flora & Curran (2004)	$\gamma_1 = .75$, $\gamma_2 = 1.75$ $\gamma_1 = .75$, $\gamma_2 = 3.75$ $\gamma_1 = 1.25$, $\gamma_2 = 3.75$	$\gamma_1 = 1.25$, $\gamma_2 = 1.75$
Hipp & Bollen (2003)		$\gamma_1 = 1.5$, $\gamma_2 = 3$
Lix, Keselman & Hinds (2005)	$\gamma_1 = 0$, $\gamma_2 = 3$	$\gamma_1 = 2$, $\gamma_2 = 6$ $\gamma_1 = 6.18$, $\gamma_2 = 110.93$
Hau & Marsh (2004)	$\gamma_1 = .5$, $\gamma_2 = .5$	$\gamma_1 = 1$, $\gamma_2 = 1.5$ $\gamma_1 = 1.5$, $\gamma_2 = 3.25$
Weathers, Sharma & Niedrich (2005)	$\gamma_1 = 0$, $\gamma_2 = 6$ $\gamma_1 = 1$, $\gamma_2 = 6$	$\gamma_1 = 1$, $\gamma_2 = 0$
Vallejo et al. (2004)		$\gamma_1 = 1.63$, $\gamma_2 = 4$
Fouladi & Yockey (2002)	$\gamma_1 = 0$, $\gamma_2 = 6$ $\gamma_1 = .75$, $\gamma_2 = 6$	$\gamma_1 = 0$, $\gamma_2 = -1$ $\gamma_1 = .75$, $\gamma_2 = 0$
Jedidi, Jagpal & Desarbo (1997)	$\gamma_1 = 0$, $\gamma_2 = 2.75$ $\gamma_1 = 1$, $\gamma_2 = 2.75$	$\gamma_1 = .75$, $\gamma_2 = 0$
Harwell & Serlin (1989)	$\gamma_1 = 0$, $\gamma_2 = 3$ $\gamma_1 = 0$, $\gamma_2 = 25$	$\gamma_1 = 0$, $\gamma_2 = -1.12$ $\gamma_1 = 2$, $\gamma_2 = 6$

Harwell (1991)	$\gamma_1 = 0$, $\gamma_2 = 3$	$\gamma_1 = 2$, $\gamma_2 = 6$ (with coefficients)
Berkovits et al. (2000)	$\gamma_1 = 3$, $\gamma_2 = 21$	$\gamma_1 = 1$, $\gamma_2 = .75$ $\gamma_1 = 1.75$, $\gamma_2 = 3.75$
Bauer & Curran (2003)	$\gamma_1 = 1.5$, $\gamma_2 = 6$	$\gamma_1 = 1$, $\gamma_2 = 1$
Enders & Bandalos (1999)	$\gamma_1 = 1.25$, $\gamma_2 = 3.5$ $\gamma_1 = 0$, $\gamma_2 = 3.75$	$\gamma_1 = 2.25$, $\gamma_2 = 7$
Ferrando & Lorenzo–Seva (1999)		$\gamma_1 = 1$, $\gamma_2 = 1$
Nevitt & Hancock (2000)	$\gamma_1 = 2$, $\gamma_2 = 7$ $\gamma_1 = 3$, $\gamma_2 = 21$	
Fouladi (2000)	$\gamma_1 = 0$, $\gamma_2 = 1$ $\gamma_1 = 0$, $\gamma_2 = 3$ $\gamma_1 = 0$, $\gamma_2 = 6$ $\gamma_1 = 0$, $\gamma_2 = 25$ $\gamma_1 = 1$, $\gamma_2 = 3$ $\gamma_1 = 1$, $\gamma_2 = 6$ $\gamma_1 = 1$, $\gamma_2 = 25$ $\gamma_1 = 2$, $\gamma_2 = 25$ $\gamma_1 = 3$, $\gamma_2 = 25$	$\gamma_1 = 0$, $\gamma_2 = -1$ $\gamma_1 = 1$, $\gamma_2 = 1$ $\gamma_1 = 2$, $\gamma_2 = 6$
Benson & Fleishman (1994)	$\gamma_1 = 1$, $\gamma_2 = 2$	$\gamma_1 = 2$, $\gamma_2 = 6$
Harwell & Serlin (1988)	$\gamma_1 = 0$, $\gamma_2 = 3$ $\gamma_1 = 0$, $\gamma_2 = 25$	$\gamma_1 = 2$, $\gamma_2 = 6$ (with coefficients)
Enders (2001)	$\gamma_1 = 1.25$, $\gamma_2 = 3.5$ $\gamma_1 = 3.25$, $\gamma_2 = 20$ $\gamma_1 = 0$, $\gamma_2 = 3.5$ $\gamma_1 = 0$, $\gamma_2 = 7$ $\gamma_1 = 0$, $\gamma_2 = 20$	$\gamma_1 = 2.25$, $\gamma_2 = 7$
Habib & Harwell (1989)	$\gamma_1 = 0$, $\gamma_2 = 3$ $\gamma_1 = 0$, $\gamma_2 = 20$ $\gamma_1 = 1$, $\gamma_2 = 3$	$\gamma_1 = 2$, $\gamma_2 = 6$ (with coefficients)

Olejnik & Algina (1984)	$\gamma_1 = 0$, $\gamma_2 = 1$ $\gamma_1 = 0$, $\gamma_2 = 3.75$ $\gamma_1 = .5$, $\gamma_2 = 3.75$ $\gamma_1 = 1$, $\gamma_2 = 3.75$	$\gamma_1 = 0$, $\gamma_2 = -1$ $\gamma_1 = .75$, $\gamma_2 = 0$ $\gamma_1 = 1.75$, $\gamma_2 = 3.75$
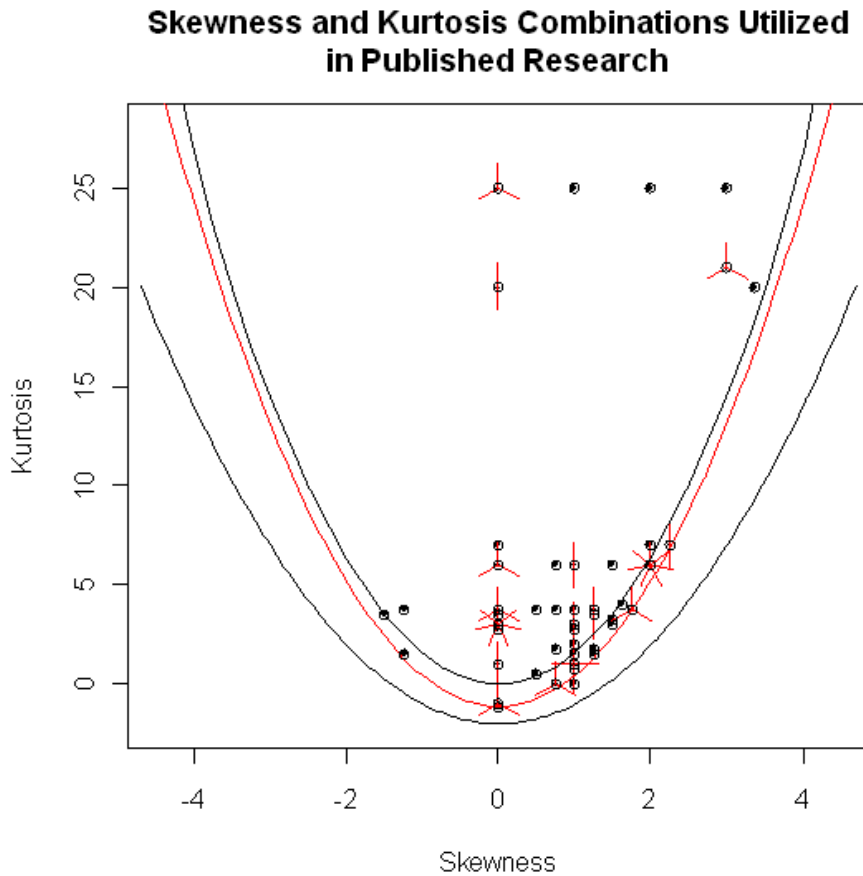
Figure 34: Sunflower Plot of Skewness–Kurtosis Combinations that have been Utilized in Published Research.

Particularly interesting are the studies by Lix & Fouladi (2007), Lix, Keselman & Hinds (2005), and Weathers et al. (2005): They use skewness-kurtosis combinations that, according to the skewness-kurtosis combination limits presented in Figure 9, do not have a real valued solution for the transformation coefficients. The combinations I am referring to are $\gamma_1 = 1, \gamma_2 = 0$,

117

$\gamma_1 = 6.18, \gamma_2 = 110.93$, and $\gamma_1 = 6.2, \gamma_2 = 110.9$ (the last two combinations are not included in Figure 34 due to scale considerations). Further, not a single study clearly mentioned including an nIM distribution, and only three studies (and one author) provide information on the transformation coefficients that were used for the simulation of nonnormal data.

Time constraints make it impossible to reconstruct the original simulations carried out in the studies listed in Table 28 and to then assess whether the choice of transformation coefficients would have an effect on the simulation results obtained in these studies. In lieu of testing the statistical procedures investigated in these studies, I will examine the performance of two confidence intervals for a single correlation for a small selection of nonnormal distributions, comparing the impact of different choices of transformation coefficients or even choosing the *g*-and-*h* distribution instead of the 3rd order power method. Results obtained in this study could provide a first indication regarding the severity of the influence of shape. The next section describes the traditional Fisher *Z* confidence interval as well as an asymptotically distribution-free confidence interval for ρ.

The Fisher Z Confidence Interval

The exact distribution of the sample correlation coefficient *r*, even when the underlying parent distribution is bivariate normal, is complex and therefore alternative procedures that allow hypothesis testing and confidence interval construction were developed early on. The Fisher *Z* transform for a single correlation is a well-known variance stabilizing and normalizing transform which allows hypothesis testing and confidence interval construction for *r* based on normal theory.

For the remainder of this study, define

$$z_\alpha = \Phi^{-1}(\alpha) \tag{82}$$

as the inverse standard normal cumulative distribution function, so that $z_{.05} = \Phi^{-1}(.05) = -1.645$ and $\Phi(1.96) = .975$. Further, let

$$\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1} \quad \text{and} \quad \tanh^{-1}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right) \tag{83}$$

be the hyperbolic tangent and inverse hyperbolic tangent of a variable x.

The Fisher Z confidence interval (see Fisher, 1915) for the product moment correlation r, calculated for a sample of size N, is constructed as follows:

$$\left[LL_z, UL_z\right] = \tanh\left(\tanh^{-1}(r) \pm z_{1-\alpha/2} \times \sqrt{1/(N-3)}\right) \tag{84}$$

When the parent distribution (the distribution of the variables for which the correlation is calculated) is bivariate normal, the performance of the Fisher Z confidence interval from Equation (100) is remarkably close to nominal. "Remarkably close to nominal" means that the empirical coverage rate of a $100(1-\alpha)\%$ confidence interval is quite close to $100(1-\alpha)\%$ and that both one-sided and two-sided hypothesis tests have empirical Type I error rates close to their nominal values. Note that the Fisher transform uses no information from the sample other than the value of the sample estimate $\hat{\rho}$ and the sample size, N. No information on skewness, kurtosis, or other moments is utilized in the construction of the confidence interval.

Fisher Z Confidence Interval Example. To illustrate the construction of all intervals utilized in Parts I and II of my dissertation, I introduce a very small sample of 5 data points: $(2,10)$, $(7,7)$, $(8,2)$, $(5,4)$, and $(1,9)$, with sample correlation $r = -.80965$:

Plot of Small Example
Correlation = -0.8096501

X	Y
2	10
7	7
8	2
5	4
1	9
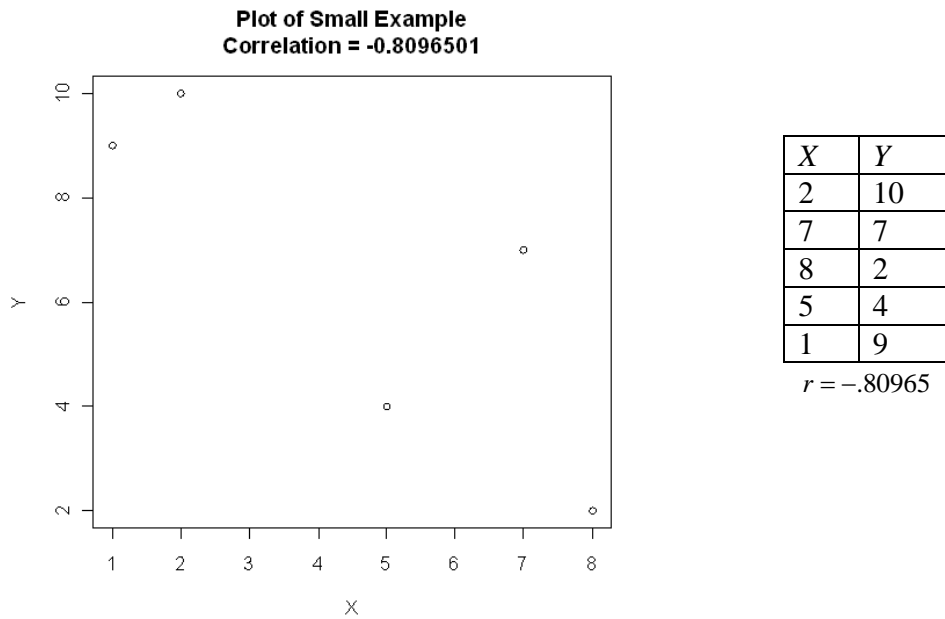
$r = -.80965$

Figure 35: Small Example for Demonstration of Confidence Intervals

The 95% Fisher Z confidence interval is constructed as

$$
\begin{aligned}
\left[LL_z, UL_z \right] &\approx \tanh\left(\tanh^{-1}\left(-.80965\right) \pm 1.96 \times \sqrt{1/2} \right) \\
&\approx \tanh\left(-1.126 \pm 1.386\right) \\
&= \left[-.987, .254\right]
\end{aligned}
\tag{85}
$$

This is a relatively wide confidence interval, but understandably so, since the sample size is only $N = 5$.

The Asymptotically Distribution-free Confidence Interval

In 1982, Steiger and Hakstian derived an asymptotically distribution free estimate of the covariance of two correlations (and therefore, as a special case, of the variance of a single correlation) based on second and fourth order moments. Equation (3.4) from their article is reproduced in Equation (86):

$$\gamma\left(r_{ij}, r_{kh}\right) = \rho_{ijkh} + \frac{1}{4} \rho_{ij}\rho_{kh}\left(\rho_{iikk} + \rho_{jjkk} + \rho_{iihh} + \rho_{jjhh}\right)$$
$$- \frac{1}{2}\rho_{ij}\left(\rho_{iikh} + \rho_{jjkh}\right) - \frac{1}{2}\rho_{kh}\left(\rho_{ijkk} + \rho_{ijhh}\right) \tag{86}$$

where $\gamma\left(r_{ij}, r_{kh}\right)$ is the asymptotic covariance between r_{ij} and r_{kh}. Steiger & Hakstian define:

$$\rho_{ijkh} = \frac{\sigma_{ijkh}}{\sqrt{\sigma_{ii}\sigma_{jj}\sigma_{kk}\sigma_{hh}}} \tag{87}$$

with

$$\sigma_{ijkh} = E\left[\left(X_i - \mu_i\right)\left(X_j - \mu_j\right)\left(X_k - \mu_k\right)\left(X_h - \mu_h\right)\right]. \tag{88}$$

Equation (87) is a general formula for fourth order moments of four variables. To simplify, we can assume that all variables are standardized, i.e. the variances are $\sigma_{ii} = \ldots = \sigma_{hh} = 1$, reducing Equation (87) to $\rho_{ijkh} = \sigma_{ijkh}$. The sample statistic corresponding to Equation (87) is

$$r_{ijkh} = \frac{s_{ijkh}}{\sqrt{s_{ii}s_{jj}s_{kk}s_{hh}}} \tag{89}$$

with

$$s_{ijkh} = \frac{1}{N}\sum_{m=1}^{N}\left(X_{mi} - \bar{X}_i\right)\left(X_{mj} - \bar{X}_j\right)\left(X_{mk} - \bar{X}_k\right)\left(X_{mh} - \bar{X}_h\right) \tag{90}$$

Note that if all coefficients are the same, i. e. $i = j = k = h$, Equation (89) estimates $\gamma_2 + 3$ as defined previously in Equation (13). To simplify the calculation of the fourth order moments in a sample, the scores will first be standardized, so that the denominator in Equation (89) does not need to be calculated and the sample estimate of the fourth order moments is

$$r_{ijkh} = \frac{1}{N}\sum_{m=1}^{N}\left(Z_{mi}\right)\left(Z_{mj}\right)\left(Z_{mk}\right)\left(Z_{mh}\right). \tag{91}$$

We then have the following sample estimate of the asymptotic covariance between two correlations r_{ij}, r_{kh} (compare with equation (5.1) in Steiger & Hakstian):

$$\hat{\gamma}_{ijkh} = r_{ijkh} + \frac{1}{4}r_{ij}r_{kh}\left(r_{iikk} + r_{jjkk} + r_{iihh} + r_{jjhh}\right) - \frac{1}{2}r_{ij}\left(r_{iikh} + r_{jjkh}\right) - \frac{1}{2}r_{kh}\left(r_{ijkk} + r_{ijhh}\right). \quad (92)$$

To construct a confidence interval for a single correlation, we can use Equation (92) and letting $i = k$ and $j = h$, obtain variance estimate:

$$\begin{aligned}
\hat{\gamma}_{ijij} &= r_{ijij} + \frac{1}{4}r_{ij}r_{ij}\left(r_{iiii} + r_{jjii} + r_{iijj} + r_{jjjj}\right) - \frac{1}{2}r_{ij}\left(r_{iiij} + r_{jjij}\right) - \frac{1}{2}r_{ij}\left(r_{ijii} + r_{ijjj}\right) \\
&= r_{iijj} + \frac{1}{4}r_{ij}^{2}\left(r_{iiii} + 2r_{iijj} + r_{jjjj}\right) - r_{ij}\left(r_{iiij} + r_{ijjj}\right)
\end{aligned} \quad (93)$$

This variance estimate takes into account all fourth order moments between two variables. The variance estimate in Equation (93) can now be used in any formula for a confidence interval around a single correlation. For example, a common approximate confidence interval around a single correlation is

$$r \pm z_{1-\alpha/2}\sqrt{\frac{\left(1-r^{2}\right)^{2}}{N-2}} \quad (94)$$

using $\sqrt{\left(1-r^{2}\right)^{2}/\left(N-2\right)}$ as a standard error estimate. This can now be substituted with the estimate from Equation (93), yielding

$$\left[LL_{ADF}, UL_{ADF}\right] = r \pm z_{1-\alpha/2}\sqrt{\frac{r_{iijj} + \frac{1}{4}r_{ij}^{2}\left(r_{iiii} + 2r_{iijj} + r_{jjjj}\right) - r_{ij}\left(r_{iiij} + r_{ijjj}\right)}{N-2}} \quad (95)$$

Using Equation (93) as a variance estimate receives further support from Yuan, Bentler, & Zhang (2005), who write:

> *"Results in Yuan and Bentler (1999a, 2000a, 2002a) imply that the asymptotic distributions of the covariance parameter estimates, the commonly used sample correlation coefficients, and sample reliability coefficients depend on only the joint fourth–order moments or kurtoses of the variables."*

Asymptotically Distribution Free Example. For the small example data set from above, we need to first standardize the scores and then calculate the correlation, squared correlation, and all fourth order moments between the two variables:

$$r_{12} = -.8096501$$
$$r_{12}{}^2 = .6555334$$
$$r_{1111} = 1.099827$$
$$r_{1112} = -.916156$$
$$r_{1122} = .9563612$$
$$r_{1222} = -1.034218$$
$$r_{2222} = 1.217355$$

Entering these values into Equation (95) provides us with the 95% confidence interval:

$$\left[LL_{ADF}, UL_{ADF}\right] = -.810 \pm$$

$$1.96\sqrt{\frac{.956 + \dfrac{1}{4}.656\left(1.010 + 2 \times .956 + 1.217\right) + .810\left(-.916 - 1.034\right)}{5 - 2}} \tag{96}$$

$$\approx -.810 \pm 1.96\sqrt{\frac{0.055}{3}}$$

$$\approx \left[-1.076, -.544\right]$$

The resulting CI includes values outside the parameter space for correlations, which may indicate that the asymptotically distribution free confidence interval may not work well for small sample sizes. Also, rounding to three decimal places introduced significant rounding error. A more exact value for the confidence interval is $[-1.110, -.509]$.

Evaluating Confidence Interval Performance: Coverage Rate & Balance Index

To evaluate and compare the performance of these two confidence intervals, I will assess the rate at which they cover the true parameter ρ, the so-called *coverage rate*. In addition to coverage rate, I will employ coverage balance as another indicator of performance. A confidence interval has (good) coverage balance if it misses ρ equally often above and below, while coverage imbalance occurs when a confidence interval does not miss ρ equally often above and below. I have found the ratio of the number of times the CI misses the true parameter above and below, called *balance index*, to be a measure that improves and quickens the assessment of coverage balance:

$$\text{Balance Index} = \frac{\text{Miss Rate Below}}{\text{Miss Rate Above}} \qquad (97)$$

This index will hover near 1 when a confidence interval exhibits good coverage balance, and will go toward 0 or ∞ as coverage balance worsens. For example, a value of 0.5 stands for a CI that missed ρ twice as often by lying above it and a value of 2 stands for a CI that missed twice as often by lying below. For a 95% confidence interval with nominal coverage rate, a balance index of 0.5 would mean that it missed the true ρ by lying 1.66% below it and 3.33% above it, a balance index of 2 would mean that it lies 3.33% below ρ and 1.66% above it. Coverage balance was already called to attention in DiCiccio & Efron (1996), who emphasize that they construct 90% confidence intervals which have noncoverage probabilities of 5% in each tail, not just 10% overall noncoverage (see also Efron, 2003).

Despite the importance of coverage balance being pointed out by Efron, most studies on the performance of bootstrap confidence intervals do not include this information, which is, in fact, not any more difficult to obtain than coverage rate. Earlier studies such as Rasmussen

(1988, 1989), Strube (1988), Efron (1988), Wilcox (1991), and Sievers (1996) do also not provide information on coverage balance. Neither do Lee & Rodgers (1998) or Beasley et al., who provide Type I error and power rates for hypothesis tests conducted with the BCa and univariate sampling bootstrap confidence intervals, but little to no information on coverage rate and balance.

Method

Four IM distributions (distribution 1 – 4) and four nIM distributions (distribution 5 – 8) were simulated. The four IM distributions were chosen from the pool of Vale & Maurelli distributions previously used in published research that simulates nonnormality (Table 28). To ensure that the present study is as representative as possible of previous research, the distributions from Table 28 were plotted and then visually grouped into four categories of distributions with

1. A cut-off effect on one side of a variable. Since all distributions in this group have identical marginals, this means that there will be a cut-off effect on one side of each variable. These distributions have a triangular shape and resemble a bivariate chi-square distribution with low degrees of freedom. (distribution 3a, b, & c)

2. A cut-off effect on both sides of a variable (and, because of symmetry, both variables), leading to a rather box-shaped bivariate distribution. (distribution 1b, 2b, & 4b)

3. A star-shaped appearance with long tails on both ends of both variables. (distribution 2a)

4. A relatively regular or fanned shape. (distribution 1a & c, 2c, 4a & c)

The distributions listed in Table 29 represent each category at least once, as indicated in parentheses in the list of categories. In addition to the two distinct Vale & Maurelli IM distributions that exist for a given skewness-kurtosis combination, the corresponding *g*-and-*h*

125

distribution was simulated as well, and all four IM distributions were created with underlying

population correlations $\rho_Y = -.2$, $\rho_Y = 0$, $\rho_Y = .4$, and $\rho_Y = .8$.

Table 29: Four IM Distributions Used to Demonstrate Effects of Shape on Simulation Results

Description of Distribution	Transformation Coefficients
$\gamma_{11} = \gamma_{12} = 0$, $\gamma_{21} = \gamma_{22} = 2.75$, $\rho_Y = -.2, .0, .4, .8$ (as utilized by Jedidi, Jagpal, & Desarbo with $\rho = .8$)	$\text{v\&m}_{1a} = \begin{bmatrix} 0 & 0.7948 & 0 & 0.0643 \end{bmatrix}$ $\text{v\&m}_{1b} = \begin{bmatrix} 0 & 1.5479 & 0 & -0.2569 \end{bmatrix}$ $\text{g\&h}_{1c} = \begin{bmatrix} 0 & 0.1046 \end{bmatrix}$
$\gamma_{11} = \gamma_{12} = 0$, $\gamma_{21} = \gamma_{22} = 25$, $\rho_Y = -.2, .0, .4, .8$ (as used by Harwell & Serlin, 1989, with $\rho = .3$ and $\rho = .7$)	$\text{v\&m}_{2a} = \begin{bmatrix} 0 & 0.2553 & 0 & 0.2038 \end{bmatrix}$ $\text{v\&m}_{2b} = \begin{bmatrix} 0 & 1.5667 & 0 & -0.3482 \end{bmatrix}$ $\text{g\&h}_{2c} = \begin{bmatrix} 0 & 0.1930 \end{bmatrix}$
$\gamma_{11} = \gamma_{12} = 1.25$, $\gamma_{21} = \gamma_{22} = 1.5$, $\rho_Y = -.2, .0, .4, .8$ (as used by Rausch & Kelley, 2009, with $\rho = .1$ and $\rho = .5$)	$\text{v\&m}_{3a} = \begin{bmatrix} -0.2823 & -1.0373 & 0.2823 & 0.0421 \end{bmatrix}$ $\text{v\&m}_{3b} = \begin{bmatrix} -0.3915 & -1.1058 & 0.3915 & 0.1044 \end{bmatrix}$ $\text{g\&h}_{3c} = \begin{bmatrix} 0.6074 & -0.1381 \end{bmatrix}$
$\gamma_{11} = \gamma_{12} = -1.25$, $\gamma_{21} = \gamma_{22} = 3.75$, $\rho_Y = -.2, .0, .4, .8$ (as used by Rausch & Kelley, 2009, with $\rho = .1$ and $\rho = .5$)	$\text{v\&m}_{4a} = \begin{bmatrix} 0.1606 & 0.8189 & -0.1606 & 0.0492 \end{bmatrix}$ $\text{v\&m}_{4b} = \begin{bmatrix} 0.4059 & 1.2091 & -0.4059 & -0.1663 \end{bmatrix}$ $\text{g\&h}_{4c} = \begin{bmatrix} 0.3226 & 0.0377 \end{bmatrix}$

Very few simulation studies conducted in the past utilized nIM distributions to test the

robustness of statistical procedures to nonnormality. However, it seems unrealistic to expect real

multivariate data to always have identical marginals, about as unrealistic as expecting real data to

always be normally distributed. With potentially all four marginal parameters γ_{11} through γ_{22}

different from each other, there is an immense range of possible nIM bivariate distributions that

can be simulated. Lacking systematic information on the behavior of real world nIM

distributions, I randomly sampled and combined values for γ_{11}, γ_{12}, γ_{21}, and γ_{22} from a weighted list of values obtained from Table 28. The values for skewness were sampled from the list [–1.5, –1.25, 0, 0.5, 0.75, 1, 1.25, 1.5, 1.63, 1.75, 2, 2.25, 3, 3.25], with weights according to the frequency with which they were used in the studies listed in Table 28. The values for kurtosis were sampled from the list [–1.12, –1, 0, 0.5, 0.75, 1, 1.5, 1.75, 2, 2.75, 3, 3.25, 3.5, 3.75, 4, 6, 7, 20, 21, 25]. Once two values for each skewness and kurtosis were chosen, I plotted the resulting bivariate nIM distribution with $\rho_Y = .5$ and $N = 5000$. Without providing a full list of graphs of these nIM distributions, I eventually settled on distributions with shapes that repeatedly occurred and that also resulted in noteworthy variations in confidence interval performance. However, these selections do not claim to be a comprehensive summary of any possible situation one may encounter when using the Vale & Maurelli method. Table 30 below contains information on simulation parameters (all nIM distributions were simulated with $\rho_Y = 0$ and $\rho_Y = .5$) and transformation coefficients.

Table 30: Four nIM Distributions Used to Demonstrate Effects of Shape on Simulation Results

Description of Distribution	Transformation Coefficients
$\gamma_{11} = 2$, $\gamma_{12} = 0$, $\gamma_{21} = 6$, $\gamma_{22} = 3.75$ $\rho_Y = .0, .5$	$\text{v\&m}_{5a} = \begin{bmatrix} -0.3137 & 0.8263 & 0.3137 & 0.0227 \end{bmatrix}$ $\text{v\&m}_{5b} = \begin{bmatrix} -0.5695 & -0.8157 & 0.5695 & 0.0878 \end{bmatrix}$ $\text{v\&m}_{5c} = \begin{bmatrix} 0 & 0.7480 & 0 & 0.0779 \end{bmatrix}$ $\text{v\&m}_{5d} = \begin{bmatrix} 0 & -1.5568 & 0 & 0.2662 \end{bmatrix}$ $\text{g\&h}_{5e} = \begin{bmatrix} 0.7594 & -0.0979 \end{bmatrix}$ $\text{g\&h}_{5f} = \begin{bmatrix} 0 & 0.1199 \end{bmatrix}$
$\gamma_{11} = 2$, $\gamma_{12} = 1$ $\gamma_{21} = 25$, $\gamma_{22} = 2.75$ $\rho_Y = .0, .5$	$\text{v\&m}_{6a} = \begin{bmatrix} -0.1379 & 0.2872 & 0.1379 & 0.1914 \end{bmatrix}$ $\text{v\&m}_{6b} = \begin{bmatrix} -0.2966 & -1.4301 & 0.2966 & 0.3059 \end{bmatrix}$ $\text{v\&m}_{6c} = \begin{bmatrix} -0.1321 & 0.8483 & 0.1321 & 0.0428 \end{bmatrix}$ $\text{v\&m}_{6d} = \begin{bmatrix} -0.3363 & -1.3060 & 0.3363 & 0.1831 \end{bmatrix}$ $\text{g\&h}_{6e} = \begin{bmatrix} 0.2396 & 0.1515 \end{bmatrix}$ $\text{g\&h}_{6f} = \begin{bmatrix} 0.2558 & 0.0446 \end{bmatrix}$
$\gamma_{11} = -1.25$, $\gamma_{12} = 1.25$ $\gamma_{21} = 3.75$, $\gamma_{22} = 3.5$ $\rho_Y = .0, .5$	$\text{v\&m}_{7a} = \begin{bmatrix} 0.1606 & 0.8189 & -0.1606 & 0.0492 \end{bmatrix}$ $\text{v\&m}_{7b} = \begin{bmatrix} 0.4059 & -1.2091 & -0.4059 & 0.1663 \end{bmatrix}$ $\text{v\&m}_{7c} = \begin{bmatrix} -0.1655 & 0.8347 & 0.1655 & 0.0439 \end{bmatrix}$ $\text{v\&m}_{7d} = \begin{bmatrix} -0.4086 & -1.1981 & 0.4086 & 0.1614 \end{bmatrix}$ $\text{g\&h}_{7e} = \begin{bmatrix} -0.3226 & 0.0377 \end{bmatrix}$ $\text{g\&h}_{7f} = \begin{bmatrix} 0.3372 & 0.0283 \end{bmatrix}$
$\gamma_{11} = 3$, $\gamma_{12} = 1.63$ $\gamma_{21} = 21$, $\gamma_{22} = 4$ $\rho_Y = .0, .5$	$\text{v\&m}_{8a} = \begin{bmatrix} -0.2523 & 0.4186 & 0.2523 & 0.1476 \end{bmatrix}$ $\text{v\&m}_{8b} = \begin{bmatrix} -0.6371 & -0.6816 & 0.6371 & 0.1487 \end{bmatrix}$ $\text{v\&m}_{8c} = \begin{bmatrix} -0.2581 & 0.8798 & 0.2581 & 0.0167 \end{bmatrix}$ $\text{v\&m}_{8d} = \begin{bmatrix} -0.4979 & -0.9894 & 0.4979 & 0.1112 \end{bmatrix}$ $\text{g\&h}_{8e} = \begin{bmatrix} 0.6542 & 0.0242 \end{bmatrix}$ $\text{g\&h}_{8f} = \begin{bmatrix} 0.5942 & -0.0633 \end{bmatrix}$

Confirmation of Population Moments. Table 31 through Table 38 contain values for

observed skewnesses, kurtoses, and correlations based on 2,000,000 replications each to compare

population moments with moments for very large samples. Notice that for distribution 7d, $\rho = .5$ cannot be simulated and that even at these very large samples, the population parameter values for skewness and kurtosis are not always reached. In particular, the (*g*-and-*h*) distribution 2c shows values of sample kurtoses perceivably lower than the nominal population kurtosis of 25.

Table 31: Large Sample Moments For IM Distribution 1

	$\gamma_{11} = 0$	$\gamma_{12} = 0$	$\gamma_{21} = 2.75$	$\gamma_{22} = 2.75$	$\rho = -.2, .0, .4, .8$
	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}$
V&M1	−0.0047	0.0087	2.7806	2.7903	−.1993
V&M2	0.0065	0.0049	3.0537	2.9053	−.1986
g&h	−0.0047	−0.0010	2.6288	2.7922	−.1988
V&M1	0.0048	−0.0008	2.7217	2.7111	0.0007
V&M2	−0.0033	−0.0167	2.7914	2.5818	0.0013
g&h	−0.0139	−0.0106	2.6370	2.7114	−0.0007
V&M1	0.0098	0.0057	2.7999	2.7967	.4008
V&M2	−0.0110	−0.0421	2.2067	2.8203	.4005
g&h	0.0018	0.0074	2.7760	2.8362	.4005
V&M1	0.0015	0.0016	2.7818	2.8067	.7997
V&M2	0.0112	−0.0363	3.3650	2.1665	.7996
g&h	−0.0010	−0.0023	2.7222	2.7762	.7997

Table 32: Large Sample Moments for IM Distribution 2

	$\gamma_{11} = 0$	$\gamma_{12} = 0$	$\gamma_{21} = 25$	$\gamma_{22} = 25$	$\rho = -.2, .0, .4, .8$
	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}$
V&M1	–0.0730	0.0285	25.1358	25.8449	–.1996
V&M2	0.0440	–0.0616	22.1144	25.8241	–.1991
g&h	–0.0650	0.0719	18.4026	14.1446	–.2011
V&M1	–0.0300	0.0777	23.8444	24.4979	.0005
V&M2	–0.0459	–0.0017	22.1743	21.7601	–.0003
g&h	0.1008	0.0146	17.5920	19.0393	–.0003
V&M1	–0.0028	–0.0009	25.1513	24.2231	.4012
V&M2	0.1053	0.0876	22.8050	21.8017	.3971
g&h	–0.1601	0.0032	34.1461	14.4994	.3999
V&M1	0.0274	0.0422	24.6004	23.9087	.7989
V&M2	0.1345	0.0352	26.9823	25.7209	.7993
g&h	–0.0629	–0.0009	19.1303	18.1340	.8003

Table 33: Large Sample Statistics for IM Distribution 3

	$\gamma_{11} = 1.25$	$\gamma_{12} = 1.25$	$\gamma_{21} = 1.5$	$\gamma_{22} = 1.5$	$\rho = -.38, -.2, .0, .4, .8$
	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}$
V&M1	1.2482	1.2488	1.4873	1.4952	–.3807
V&M2	1.2510	1.2485	1.5374	1.5136	–.3807
g&h	1.2464	1.2528	1.4842	1.5123	–.3799
V&M1	1.2495	1.2499	1.4983	1.4934	–.1997
V&M2	1.2462	1.2457	1.4641	1.4405	–.2007
g&h	1.2508	1.2500	1.5008	1.4980	–.1998
V&M1	1.2521	1.2479	1.5096	1.4901	–.0013
V&M2	1.2555	1.2490	1.5527	1.4877	–.0000
g&h	1.2491	1.2510	1.4986	1.5005	–.0007
V&M1	1.2518	1.2517	1.5066	1.5102	.4008
V&M2	1.2536	1.2570	1.5835	1.6138	.3994
g&h	1.2486	1.2502	1.4967	1.5007	.3999
V&M1	1.2544	1.2522	1.5217	1.5041	.8001
V&M2	1.2465	1.2486	1.4696	1.5429	.7999
g&h	1.2515	1.2529	1.5045	1.5094	.8004

Table 34: Large Sample Statistics for IM Distribution 4

	$\gamma_{11} = -1.25$	$\gamma_{12} = -1.25$	$\gamma_{21} = 3.75$	$\gamma_{22} = 3.75$	$\rho = -.2, .0, .4, .8$
	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}$
V&M1	−1.2555	−1.2495	3.7804	3.7602	−.2005
V&M2	−1.2658	−1.2444	4.2134	3.6442	−.1996
g&h	−1.2499	−1.2485	3.6960	3.7487	−.2003
V&M1	−1.2452	−1.2492	3.7048	3.7131	.0001
V&M2	−1.2395	−1.2549	3.5667	3.9131	−.0001
g&h	−1.2465	−1.2492	3.6651	3.7604	−.0000
V&M1	−1.2405	−1.2463	3.6589	3.7375	.4009
V&M2	−1.2532	−1.2460	3.7413	3.5296	.3996
g&h	−1.2490	−1.2460	3.6582	3.7544	.4114
V&M1	−1.2477	−1.2465	3.7499	3.6973	.8000
V&M2	−1.2431	−1.2338	3.5888	3.4300	.7995
g&h	−1.2576	−1.2699	3.8551	4.0037	.8001

Table 35: Large Sample Statistics for nIM Distribution 5

	$\gamma_{11} = 2$	$\gamma_{12} = 0$	$\gamma_{21} = 6$	$\gamma_{22} = 3.75$	$\rho = .0, .5$
	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}$
V&M1	1.9961	0.0111	5.9308	3.7014	−.0009
V&M2	2.0024	0.0036	5.9988	3.2789	−.0003
V&M3	2.0010	0.0037	6.0928	3.7266	.0002
V&M4	1.9867	0.0159	5.8580	3.6795	.0004
g&h	1.9953	0.0013	5.9596	4.0308	.0002
V&M1	2.0069	0.0045	6.0639	3.6915	.4986
V&M2	2.0135	−0.0312	6.0959	3.4364	.4978
V&M3	1.9935	0.0070	5.9532	3.7557	.5011
V&M4	1.9940	−0.0046	5.8941	2.8272	.5029
g&h	2.0044	0.0053	6.0417	3.5633	.4999

Table 36: Large Sample Statistics for nIM Distribution 6

	$\gamma_{11}=2$	$\gamma_{12}=1$	$\gamma_{21}=25$	$\gamma_{22}=2.75$	$\rho=.0,.5$
	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}$
V&M1	2.0495	1.0041	24.1686	2.8208	.0005
V&M2	2.0579	1.0039	26.2677	2.7218	−.0002
V&M3	2.1493	0.9954	34.4119	2.7462	.0017
V&M4	1.8947	1.0225	22.1066	2.9400	−.0011
g&h	2.0636	0.9933	26.1164	2.7256	−.0008
V&M1	2.0221	0.9951	24.7389	2.7300	.5013
V&M2	2.0574	0.9986	25.6209	2.6501	.5019
V&M3	2.0265	0.9997	26.7434	2.7531	.5005
V&M4	1.9816	1.0119	24.2023	2.8830	.4991
g&h	2.0576	0.9984	22.9781	2.7267	.5002

Table 37: Large Sample Statistics for nIM Distribution 7

	$\gamma_{11}=-1.25$	$\gamma_{12}=1.25$	$\gamma_{21}=3.75$	$\gamma_{22}=3.5$	$\rho=.0,.5$
	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}$
V&M1	−1.2502	1.2470	3.7441	3.4973	.0002
V&M2	−1.2555	1.2592	3.8081	3.9149	.0000
V&M3	−1.2391	1.2515	3.3933	3.5177	−.0017
V&M4	−1.2557	1.2321	3.8082	3.0608	.0002
g&h	−1.2512	1.2515	3.7866	3.4871	.0011
V&M1	−1.2541	1.2475	3.7844	3.4764	.5000
V&M2	−1.2524	1.2385	3.7343	3.2267	.4989
V&M3	−1.2624	1.2537	3.9677	3.5328	.4987
V&M4	NA	NA	NA	NA	NA
g&h	−1.2535	1.2483	3.7173	3.5512	.5009

Table 38: Large Sample Statistics for nIM Distribution 8

	$\gamma_{11} = 3$	$\gamma_{12} = 1.63$	$\gamma_{21} = 21$	$\gamma_{22} = 4$	$\rho = .0, .5$
	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}$
V&M1	3.0314	1.6264	21.5846	3.9782	.0002
V&M2	3.0283	1.6275	21.7421	4.0528	.0009
V&M3	3.0523	1.6288	22.2828	3.9889	.0014
V&M4	3.0037	1.6386	21.5028	4.1285	.0006
g&h	3.0318	1.6333	22.2933	4.0485	.0008
V&M1	3.0136	1.6327	20.8490	4.0095	.4999
V&M2	3.0386	1.6321	21.7794	4.0122	.5012
V&M3	3.0470	1.6276	24.2088	3.9747	.4993
V&M4	3.0138	1.6437	21.4155	4.3788	.4998
g&h	2.9622	1.6287	19.5347	4.0059	.5006

Using a sample size of $N = 50$ and 100,000 replications per condition, coverage rate and balance index were collected for the Fisher *Z* (Equation (84)) and the asymptotically distribution-free confidence interval (Equation (95)). These performance indices will be contrasted with marginal skewnesses and kurtoses as well as overall shape of the distributions in the upcoming result section.

Results

Table 39 through

Table *46* contain coverage rate and balance for both the Fisher Z CI and the asymptotically distribution-free CI for all eight distributions. In addition, average sample skewnesses and kurtoses and a small graph of the distributions at each simulated population correlation are provided for each distribution. This allows for a comparison of confidence

interval performance with average sample statistics and shape of distribution. Within each table, the population correlation changes horizontally, while type of distribution (V&M 1 and 2; g-and-h) changes vertically. Average sample statistics are recorded in the right most column, while coverage performance is reported in the lower section of the table: Each cell on coverage performancehas two columns, the first of which contains coverage rate information, and the second coverage balance. Rows within each cell represent the three (for IM distributions) or five (for nIM distributions) distribution types. Coverage performance horizontally across cells follows the changing correlations, while the two rows of results represent the two different confidence intervals.

Use Table 39 as an example: In the right most column, we can see that the first V&M distribution had average sample values $\hat{\gamma}_{11} = 0.000$, $\hat{\gamma}_{12} = 0.001$, $\hat{\gamma}_{21} = 1.458$ and $\hat{\gamma}_{22} = 1.465$. The second Vale & Maurelli distribution has $\hat{\gamma}_{11} = 0.000$, $\hat{\gamma}_{12} = 0.001$, $\hat{\gamma}_{21} = -0.789$ and $\hat{\gamma}_{22} = -0.789$. From the section 'Confidence Interval Performance,' we find that the Fisher Z CI covers the true correlation $\rho = -.2$ 94.9% of the time when samples are taken from V&M distribution 1a, 93.8% of the time when taken from V&M distribution 1b, and 94.7% of the time when taken from the g-and-h distribution. When $\rho = .8$, these values change to 93.5%, 87.6%, and 93.9%. The corresponding values for the balance index are 1.91, 1.73, and 1.86. For comparison, the coverage rate and balance index for the asymptotically distribution-free CI are 90.0%, 86.5%, and 90.5% as well as 3.25, 14.98, and 3.46.

Several aspects are of interest here: a) Is performance of a particular confidence interval procedure the same across distributions with equivalent marginal skewnesses and kurtoses as well as correlations, but different shapes? The answer to this question can be found in the confidence interval performance section for every distribution utilized. Focusing only on the

134

Fisher Z CI, we can observe sizeable differences in performance for distribution 2, $\rho_Y = .8$, for

distribution 3, $\rho_Y = -.38$, and all nIM distributions (5 through 8), when $\rho_Y = .5$. For example,

coverage rate varies from .375 to .928 and coverage balance varies from 1.999 to 3.756 for

distribution 6, $\rho_Y = .5$. For the asymptotically distribution-free confidence interval, equally

drastic performance deviations can be found; e.g., for distribution 6, coverage rate varies from

.525 to .891 and coverage balance varies from 1.687 to 22.698. Cells with performance

differences between distributions of at least .05 for coverage rate or a factor of three for coverage

balance are colored in yellow.

b) When comparing two interval procedures with each other, could our decision which
confidence interval is better depend on the choice of distribution? In other words, with marginal
parameters and correlation held constant, could aspects of shape have an influence on which
confidence interval we choose over the other? If we consider

Table 40, with distribution 2, we find that at $\rho = .8$, the Fisher Z CI covers the true ρ for

the first V&M distribution 75.3% of the time, and tends to overestimate it, indicated by a balance

index of 2.57. The asymptotically distribution-free CI misses the true ρ for the same distribution

78.6% of the time and has a balance index of 2.37. Therefore, by both standards, the

asymptotically distribution-free CI would be chosen over the Fisher Z CI. For the second V&M

distribution however, the situation is a little different: The Fisher Z CI covers the true ρ 79.8%,

with a balance index of 1.94, while the asymptotic CI covers ρ 81.5%, with a balance index of

10.15. While the asymptotic CI still covers more often, the gross overestimation as indicated by

the high balance index and the negligible coverage rate difference would lead most researchers to

prefer the Fisher Z CI. A slightly weaker case can be made for the results for distribution 8,

$\rho = .5$. For the last V&M distribution, the indices are 82.4% and 1.86 for the Fisher Z CI and

84.6% and 2.33 for the asymptotically distribution-free CI, while for the *g*-and-*h* distribution,

these values are 88.2% with 2.38 and 86.5% with 1.66. The performance indices are almost reversed from one distribution to the other. Finally, if we were to focus on only one of the parameters, e.g., just coverage rate, which confidence interval seems preferable varies even more, depending on the nonnormal distribution used. For example, for distribution 5, $\rho = .5$, the Fisher Z CI would be preferred for V&M distributions 1, and 2, and the *g*-and-*h* distribution, while the asymptotic CI performs better under V&M distribution 3 and 4.

c) Last but not least we may ask ourselves what influence average sample skewness and kurtosis have on the performance of the confidence intervals. In some instances, differences in average sample skewness and kurtosis seem to coincide well with differences in performance for the two confidence intervals. For distribution 1, $\rho = .8$, the most salient deviations from nominal coverage (rate or balance) occur for the second V&M distribution. This V&M distribution also has an average sample kurtosis of about $\hat{\gamma}_2 = -.79$, a value that, interestingly enough, deviates from kurtosis for a normal distribution in the opposite direction of the desired population kurtosis, $\gamma_2 = 2.75$. However, for distribution 5, $\rho = .5$, the first and third V&M distribution have very similar average sample skewnesses and kurtoses, with values of 1.64, 0.00, 3.09, and 1.91 and 1.67, 0.00, 2.91, and 1.90, respectively. The coverage performance for both confidence intervals differs drastically between the two distributions, though. This is an example for which moments other than skewness and kurtosis greatly influence the performance of the two confidence intervals. The main differences between these distributions lie in their shapes, and therefore their additional, uncontrolled moments.

Table 39: Coverage Rate and Balance forDistribution 1 (see detailed description on page 133)
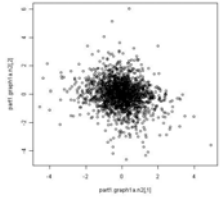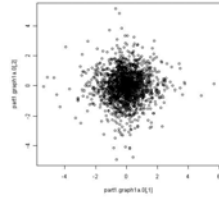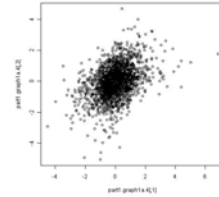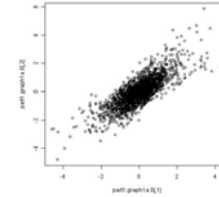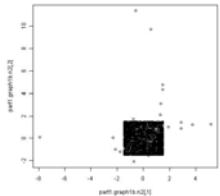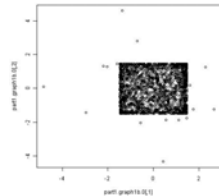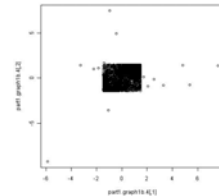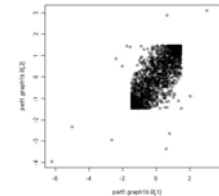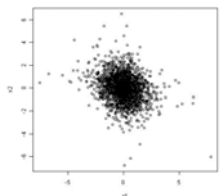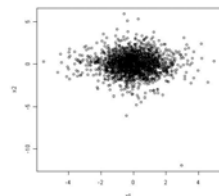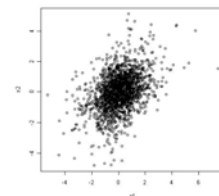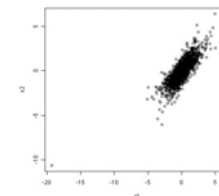
Distribution 1 a, b, & c: $\gamma_{11} = \gamma_{12} = 0$, $\gamma_{21} = \gamma_{22} = 2.75$

	$\rho_Y = -.2$	$\rho_Y = 0$	$\rho_Y = .4$	$\rho_Y = .8$	$\overline{\hat{\gamma}}_1$ and $\overline{\hat{\gamma}}_2$
V&M 1a					0.000 1.458 0.001 1.465
V&M 1b					0.000 −0.789 0.000 −0.789
G&h					0.000 1.244 −0.001 1.244

Confidence Interval Performance					
Fisher *Z* CI	0.9487 0.8136 0.9376 0.6673 0.9473 0.8447	0.9475 0.9778 0.9506 0.9956 0.9492 0.9996	0.9453 1.4769 0.9117 1.5428 0.9463 1.3719	0.9347 1.9118 0.8764 1.7287 0.9391 1.8610	
Asympt. CI	0.9074 0.7921 0.9170 0.5010 0.9075 0.7349	0.9068 0.9900 0.9268 1.0182 0.9093 1.0178	0.9061 1.7033 0.8987 3.9599 0.9079 1.7432	0.9003 3.2547 0.8650 14.9787 0.9045 3.4607	

Table 40: Coverage Rate and Balance for Distribution 2 (see detailed description on page 133)
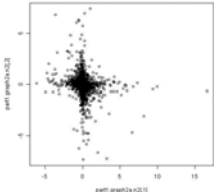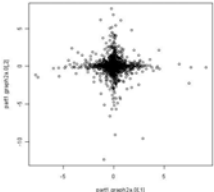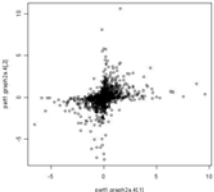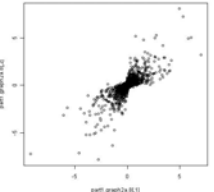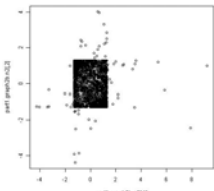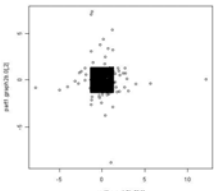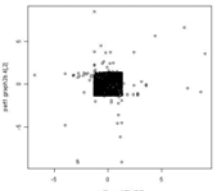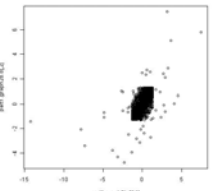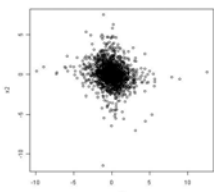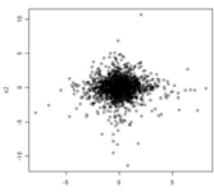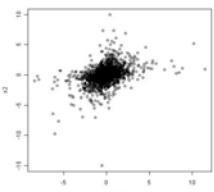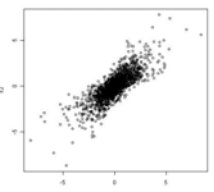
Distribution 2: $\gamma_{11} = \gamma_{12} = 0$, $\gamma_{21} = \gamma_{22} = 25$					
	$\rho_Y = -.2$	$\rho_Y = 0$	$\rho_Y = .4$	$\rho_Y = .8$	$\overline{\hat{\gamma}}_1$ and $\overline{\hat{\gamma}}_2$
V&M 2a					-0.001 7.690 -0.004 7.705
V&M 2b					0.002 0.877 -0.001 0.872
G&h					-0.002 3.107 0.001 3.106
Confidence Interval Performance					
Fisher Z CI	0.9230 0.2292 0.8940 0.3837 0.9452 0.5504	0.9413 0.9547 0.9520 1.0270 0.9467 0.9810	0.8770 5.3169 0.8441 1.9716 0.9338 2.6225	0.7527 2.5702 0.7976 1.9413 0.8957 3.8857	
Asympt. CI	0.8470 1.6653 0.8700 0.4134 0.8949 0.7682	0.8902 0.9720 0.9104 1.0043 0.8981 0.9889	0.8078 0.8197 0.8438 3.7371 0.8848 1.6439	0.7859 2.3708 0.8146 10.1540 0.8604 3.4647	

Table 41: Coverage Rate and Balance for Distribution 3 (see detailed description on page 133)
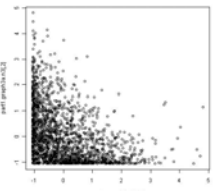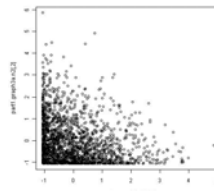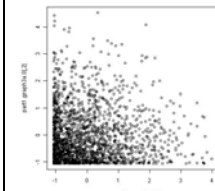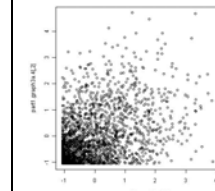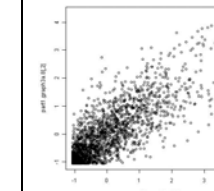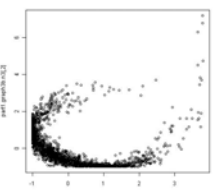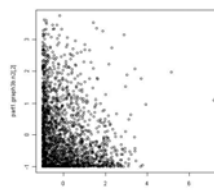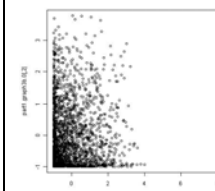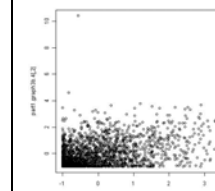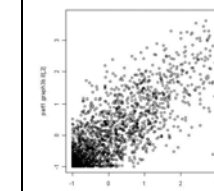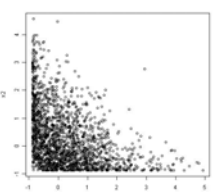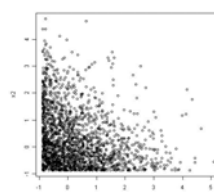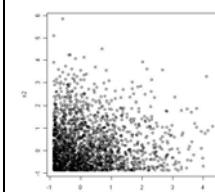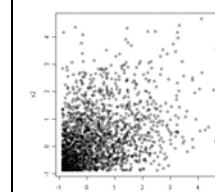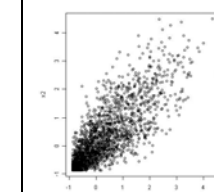
Distribution 3: $\gamma_{11} = \gamma_{12} = 1.25$, $\gamma_{21} = \gamma_{22} = 1.5$						
	$\rho_Y = -.38$	$\rho_Y = -.2$	$\rho_Y = 0$	$\rho_Y = .4$	$\rho_Y = .8$	$\overline{\hat{\gamma}}_1$ and $\overline{\hat{\gamma}}_2$
V&M 3a						1.139 0.992 1.139 0.989
V&M 3b						1.145 0.795 1.145 0.795
G&h						1.143 1.038 1.142 1.038
Confidence Interval Performance						
Fisher Z CI	0.9832 1.6609 0.7465 0.4677 0.9845 1.8391	0.9717 1.8769 0.9562 1.9058 0.9685 1.6829	0.9511 1.7440 0.9502 1.6116 0.9488 1.6620	0.9168 1.4257 0.9046 1.3546 0.9194 1.4890	0.8983 1.4605 0.8825 1.4425 0.8945 1.5744	
Asympt. CI	0.9161 0.1767 0.7396 0.0542 0.9205 0.1907	0.9152 0.3215 0.8956 0.2012 0.9104 0.3185	0.9062 0.5543 0.9055 0.5135 0.9049 0.5478	0.8992 1.4207 0.8959 1.5205 0.8998 1.3969	0.8968 4.1543 0.8896 5.6023 0.8919 4.1482	

Table 42: Coverage Rate and Balance for Distribution 4 (see detailed description on page 133)
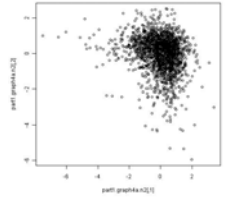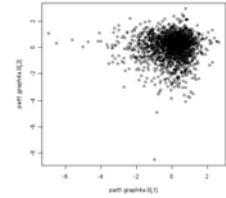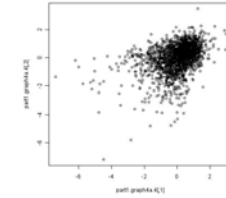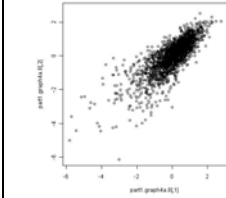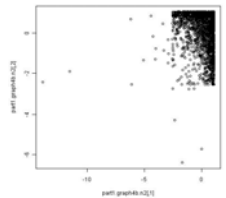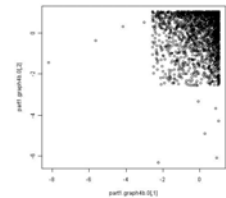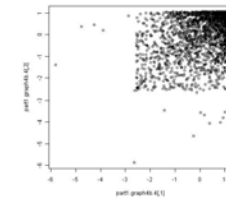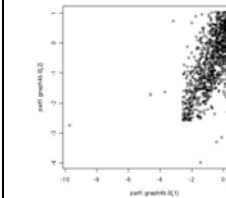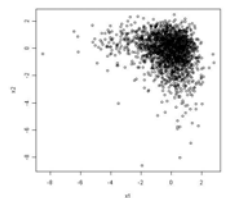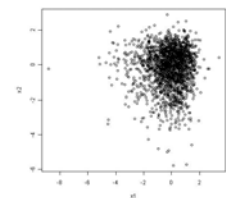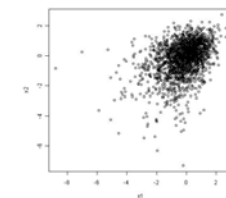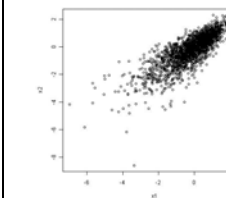
Distribution 4: $\gamma_{11} = \gamma_{12} = -1.25$, $\gamma_{21} = \gamma_{22} = 3.75$					
	$\rho_Y = -.2$	$\rho_Y = 0$	$\rho_Y = .4$	$\rho_Y = .8$	$\overline{\hat{\gamma}}_1$ and $\overline{\hat{\gamma}}_2$
V&M 2a					-0.970 1.815 -0.968 1.808
V&M 2b					-0.974 0.397 -0.975 0.397
G&h					-0.949 1.562 -0.950 1.567
Confidence Interval Performance					
Fisher Z CI	0.9582 1.1228 0.9036 1.0138 0.9590 1.1659	0.9486 1.3949 0.9498 1.4244 0.9497 1.3552	0.9282 1.7932 0.9024 1.5933 0.9305 1.7937	0.9033 1.9994 0.8683 1.6698 0.9109 2.0967	
Asympt. CI	0.9086 0.5632 0.8531 0.1597 0.9104 0.5349	0.9044 0.7427 0.9103 0.6077 0.9060 0.7474	0.8968 1.5252 0.8944 2.2496 0.8992 1.6714	0.8869 3.5634 0.8746 8.9571 0.8911 3.9491	

Table 43: Coverage Rate and Balance for Distribution 5 (see detailed description on page 133)
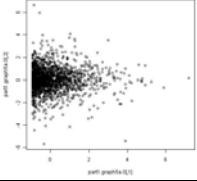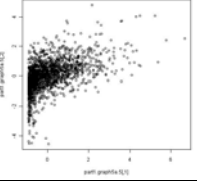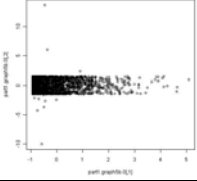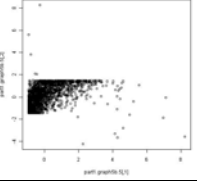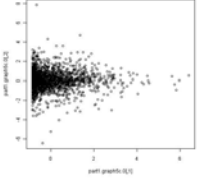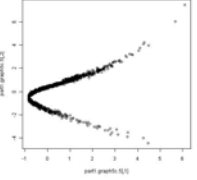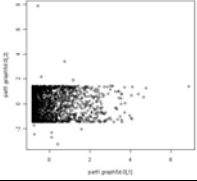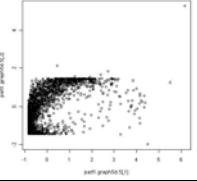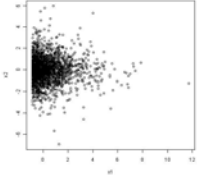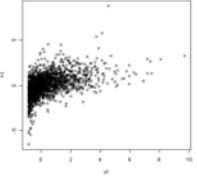
Distribution 5: $\gamma_{11}=2$, $\gamma_{12}=0$, $\gamma_{21}=6$, $\gamma_{22}=3.75$			
	$\rho_Y=.0$	$\rho_Y=.5$	$\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$
V&M1			1.635 3.092 0.002 1.908
V&M2			1.633 3.077 0.003 -0.709
V&M3			1.671 2.909 -0.001 1.904
V&M4			1.671 2.906 0.000 -0.705
G&h			1.633 3.149 0.000 1.521
Confidence Interval Performance			
Fisher Z CI	0.9488 1.0265 0.9524 1.0234 0.9483 0.9561 0.9510 1.0165 0.9484 0.9804	0.9430 2.5438 0.8045 1.9146 0.4536 2.1632 0.7788 1.7702 0.9452 2.5375	
Asymptotic CI	0.9010 1.0310 0.9057 1.0087 0.9030 0.9628 0.9058 0.9989 0.9028 0.9627	0.8919 1.6455 0.7011 21.7105 0.6889 9.8977 0.7515 16.0701 0.8959 1.9354	

Table 44: Coverage Rate and Balance for Distribution 6 (see detailed description on page 133)
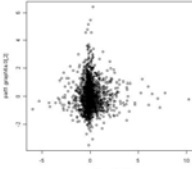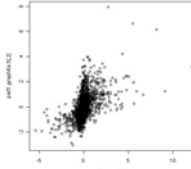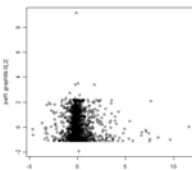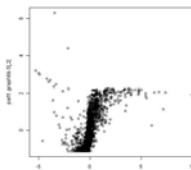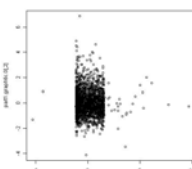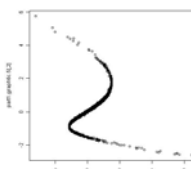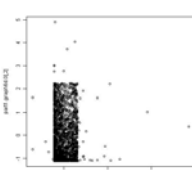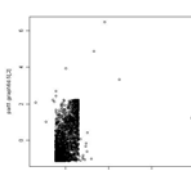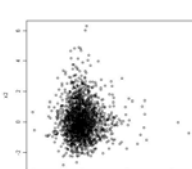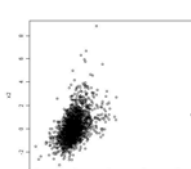
Distribution 6: $\gamma_{11}=2$, $\gamma_{12}=1$, $\gamma_{21}=25$, $\gamma_{22}=2.75$			
	$\rho_Y=.0$	$\rho_Y=.5$	$\overline{\hat{\gamma}}_1$ and $\overline{\hat{\gamma}}_2$
V&M1			1.094 7.577 0.786 1.375
V&M2			1.101 7.575 0.765 −0.137
V&M3			0.718 1.112 0.789 1.383
V&M4			0.721 1.137 0.765 −0.139
G&h			0.929 3.129 0.768 1.198

Confidence Interval Performance			
Fisher Z CI	0.9483 1.2901 0.9510 1.3864 0.9503 1.2180 0.9509 1.2408 0.9496 1.2462	0.9122 3.4431 0.5357 3.7559 0.3748 3.4137 0.8486 1.9992 0.9282 2.3054	
Asymptotic CI	0.8844 0.7971 0.8789 0.7527 0.9079 0.8402 0.9151 0.7877 0.9038 0.8529	0.8649 1.6865 0.5246 22.6984 0.5726 15.4896 0.8405 5.8355 0.8906 2.1984	

142

Table 45: Coverage Rate and Balance for Distribution 7 (see detailed description on page 133)
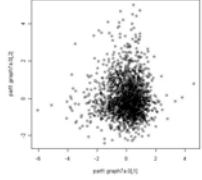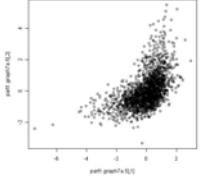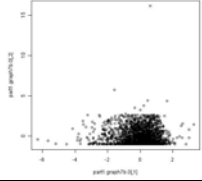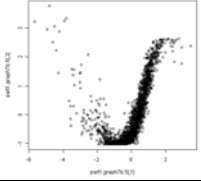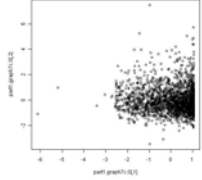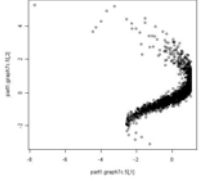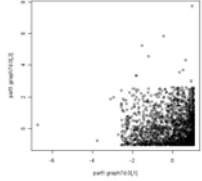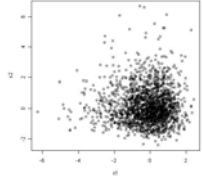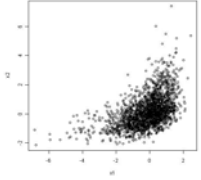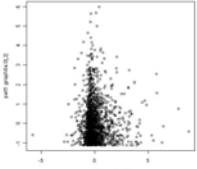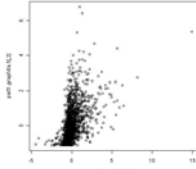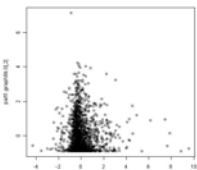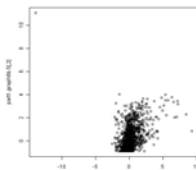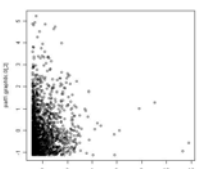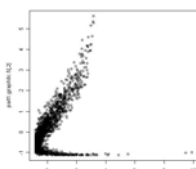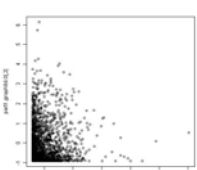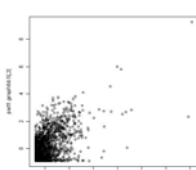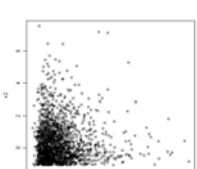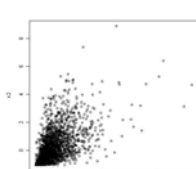
Distribution 7: $\gamma_{11}=-1.25$, $\gamma_{12}=1.25$, $\gamma_{21}=3.75$, $\gamma_{22}=3.5$			
	$\rho_Y=.0$	$\rho_Y=.5$	$\bar{\hat{\gamma}}_1$ and $\bar{\hat{\gamma}}_2$
V&M1			-0.969 1.812 0.985 1.722
V&M2			-0.969 1.813 0.990 0.421
V&M3			-0.972 1.382 0.980 1.704
V&M4			-0.973 0.386 0.991 0.420
G&h			-0.948 1.566 0.969 1.529
Confidence Interval Performance			
Fisher Z CI	0.9481 0.7186 0.9496 0.6890 0.9485 0.6615 0.9501 0.6986 0.9498 0.7101	0.9715 1.6208 0.4672 3.0776 0.4587 3.1389 0.9746 1.4858	
Asymptotic CI	0.9032 1.3456 0.9065 1.4779 0.9085 1.3980 0.9110 1.7031 0.9057 1.3097	0.9096 3.2154 0.6016 19.5376 0.5996 19.4918 0.9088 4.1462	

Table 46: Coverage Rate and Balance for Distribution 8 (see detailed description on page 133)

Distribution 8: $\gamma_{11}=3$, $\gamma_{12}=1.63$, $\gamma_{21}=21$, $\gamma_{22}=4$			
	$\rho_Y=.0$	$\rho_Y=.5$	$\overline{\hat{\gamma}}_1$ and $\overline{\hat{\gamma}}_2$
V&M1			1.859 6.722 1.357 2.125
V&M2			1.865 6.750 1.393 1.723
V&M3			1.802 4.007 1.359 2.133
V&M4			1.803 4.010 1.394 1.726
G&h			1.844 4.785 1.358 2.164
Confidence Interval Performance			
Fisher Z CI	0.9488 2.4244 0.9506 2.6776 0.9522 2.6883 0.9519 2.5476 0.9495 2.4865	0.8803 2.5633 0.7591 1.6855 0.5688 2.3916 0.8239 1.8626 0.8821 2.3791	
Asymptotic CI	0.8791 0.5043 0.8767 0.4310 0.8869 0.4493 0.8844 0.3860 0.8866 0.4965	0.8533 1.4400 0.8159 5.0433 0.7343 5.3154 0.8455 2.3335 0.8652 1.6611	

Other Nonnormal Distributions in Published Research.

I conclude Part I examining some of the other nonnormal distributions utilized in previous simulation studies on the robustness of *r*. Many studies on the robustness of procedures for correlations have focused on nonnormal distributions with identical marginals and $\rho = 0$. Mixture distributions have been particularly popular, the most extreme case a mixture of two bivariate normal distributions, one with $\rho_1 = .9$, the other with $\rho_2 = -.9$, mixed at 50% each, as in Figure 36(a) below (Kowalski, 1972; Rasmussen, 1988 & 1989; Edgell & Noon, 1984). A hypothesis test or confidence interval for a *single* correlation is applied to such distribution and the performance examined, without taking into account that a single linear model may often be considered inadequate. The overall population correlation for the resulting distribution is $\rho = .5 \times \rho_1 + .5 \times \rho_2 = 0$, but generally it seems that such data should be analyzed with a model that distinguishes two groups. Something similar may be said for bivariate nonnormal distributions with a non-linear relationship between the two variables. A distribution with $Y_1 = Z$ (a standard normal variable) and $Y_2 = Z^2$ (the same standard norma variable squared), thus creating a perfect quadratic relationship between Y_1 and Y_2, has been used by Edgell & Noon (1984). Figure 36 plots these and several other distributions for which rather severe deviations from nominal Type I error rates or coverage rates for confidence intervals on a single correlation were observed.
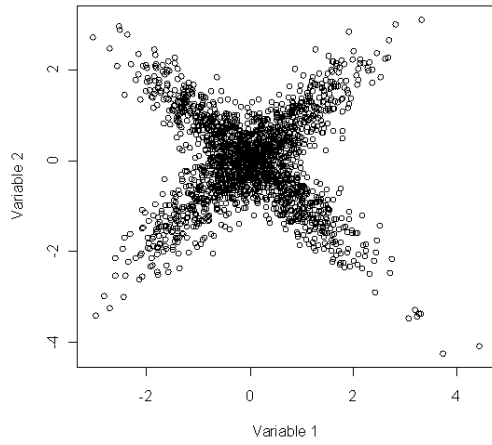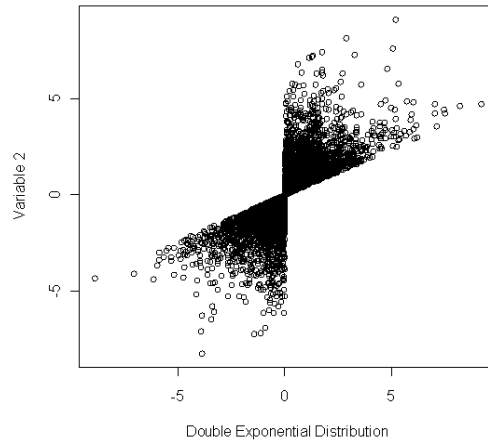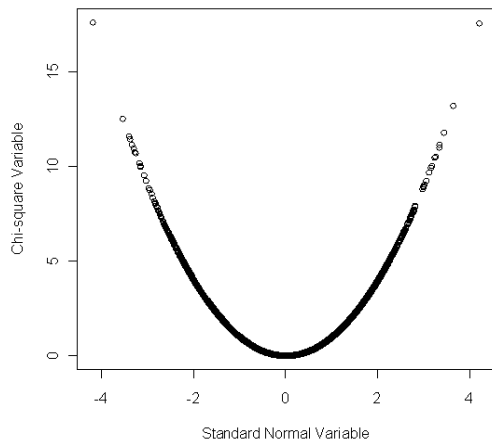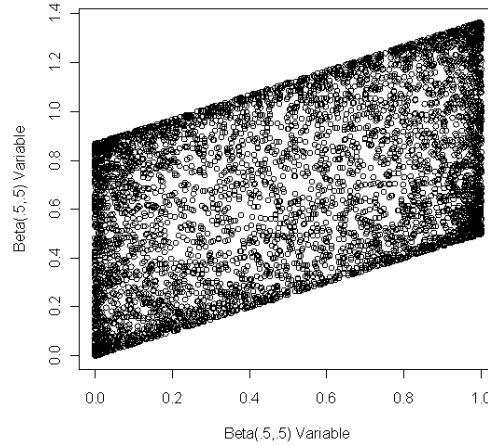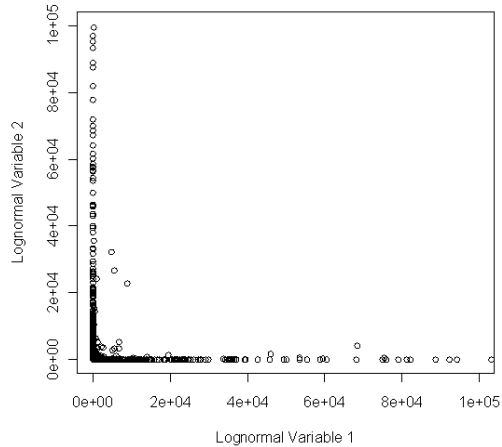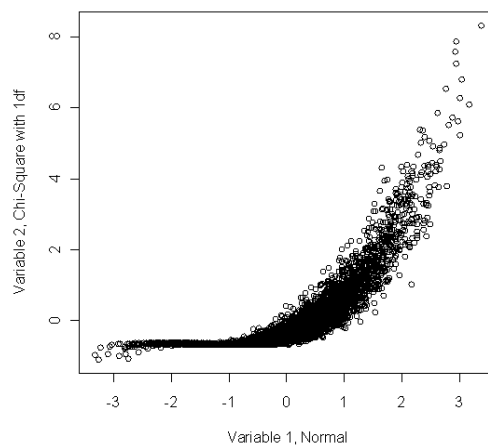
Figure 36(a) – (f): Nonnormal Distributions in Previous Research

146

Table 47 summarizes empirical estimates of asymptotic skewnesses, kurtoses, and correlations for the distributions plotted in Figure 36.

Table 47: Estimates for Expected Values of Skewnesses, Kurtoses, and Correlations for the Distributions in Figure 36, based on 2,000,000 Replications.

Distribution	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$	$\hat{\gamma}_{21}$	$\hat{\gamma}_{22}$	$\hat{\rho}$
(a) Kowalski (1972) & Rasmussen (1988 & 1989)	–0.002	–0.001	0.008	0.005	–.000
(b) Duncan & Layard (1973)	–0.001	–0.001	3.003	0.842	.749
(c) Edgell & Noon (1984)	–0.001	2.834	0.003	12.035	.001
(d) Sievers (1996)	0.001	0.000	–1.499	–0.937	.500
(e) Lee & Rodgers (1998)	1185.482	616.613	1502079.5	520139.0	.000
(f) Beasley et al (2007)	0.003	2.832	0.004	11.978	.800

It is safe to say that a reasonably sized sample (say with $N = 80$) drawn from any of these distributions will already forebode the extraordinary shape of its parent distribution, which should motivate any researcher to investigate further:
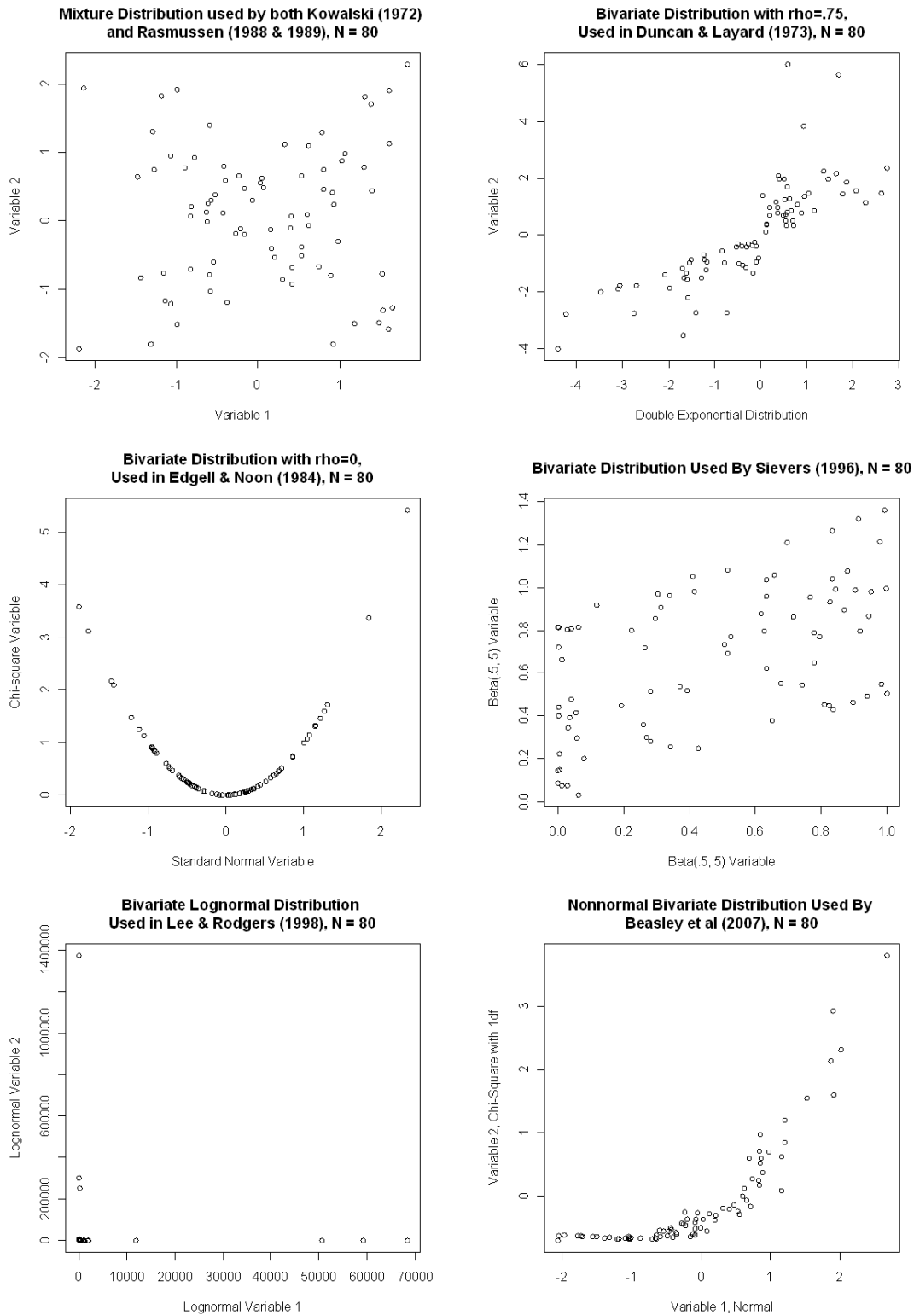
Figure 37: Samples of size $N = 80$ from the Distributions in Figure 36

Is it paramount to investigate whether a confidence interval for a single Pearson product-moment correlation is robust to the nonnormality of distribution such as in Figure 36(c)? If we require researchers to plot their data, is it likely that they will apply a confidence interval around r to a sample from such a distribution? And if not, is robustness of hypothesis tests and confidence intervals for a single correlation to the nonnormality of such a distribution of practical interest or *needed*?

The most extreme example in the writer's opinion is Figure 36(c), the case of $\rho = 0$, but total dependence between both variables with Variable 1 $\sim N(0,1)$ and Variable 2 $\sim \chi_1^2$. Edgell & Noon (1984) observed rejection rates of around .38 for sample sizes of $N = 5$ up to $N = 100$ when a Type I error of .05 was expected for this distribution (Table 3, page 581). Beasley et al (2007) employed a distribution with the same marginals, Variable 1 $\sim N(0,1)$, Variable 2 $\sim \chi_1^2$, and $\rho = 0$, but both variables truly independent of each other.For this distribution, they found empirical Type I error rates very close to their nominal value, even for sample sizes as small as $N = 5$. The distribution is plotted in Figure 38:

**Nonnormal Bivariate Distribution
Used By Beasley et al (2007)**
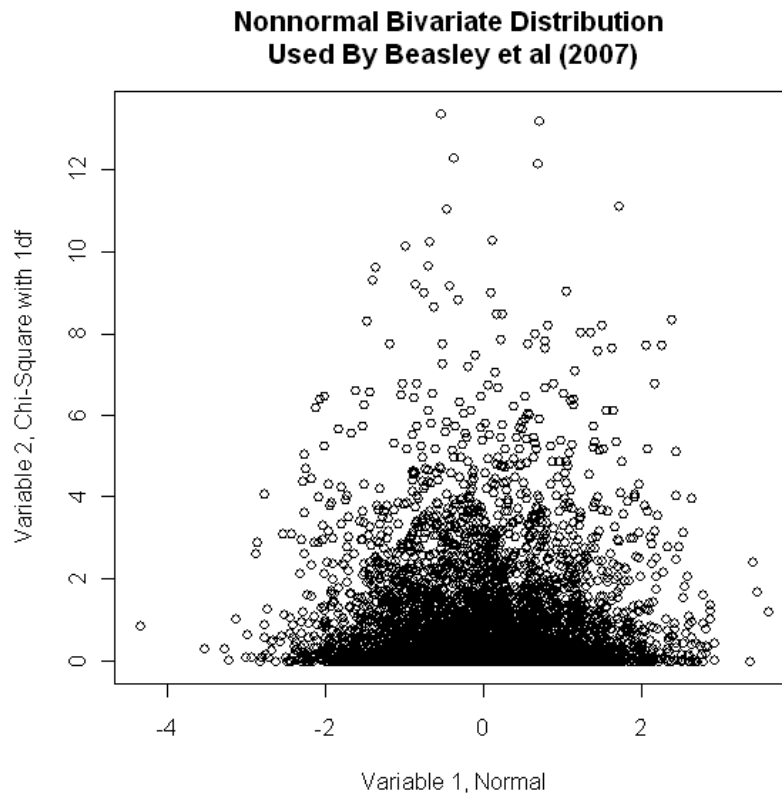
*Figure 38: Variable 1 $\sim N(0,1)$, Variable 2 $\sim \chi_1^2$ and true Independence.
Compare with Figure 36(c).*

The bivariate distributions in Figure 36(c) and Figure 38 have the same marginal

distributions, and the same population correlation of $\rho = 0$. Yet, robustness results of the Fisher

Z test or confidence interval to the nonnormality of these distributions are strikingly different

from each other. For Figure 36(c), Type I error rate is as high as .38 (nominal .05), while for

Figure 38, Type I error rate remains close to its nominal level. This observation is an extreme

instance of the observation already made above for some of the Vale & Maurelli distributions:

Something other than characteristics of the marginal distributions and the population correlation

has an influence on these robustness results. The shapes of Figure 36(c) and Figure 38 are

immensely different and generally we would say that a researcher will be more likely to test data

from a distribution such as in Figure 38 for a linear zero correlation than a data set that stems

150

from a distribution such as in Figure 36(c). As a corollary, we may say that robustness results for the distribution in Figure 38 are much more *interesting* than robustness results for the distribution in Figure 36(c).

The same will be true for other distributions. Knowledge of the first four marginal moments or even the entire marginal distribution and the population correlation will not suffice to judge whether a particular bivariate distribution is of great interest for a study on the robustness of *r*. If we simulate data using the 3rd order power method, we may either come up with the distribution in Figure 18(a) or the distribution in Figure 18(c), despite the same asymptotical marginal skewnesses and kurtoses and the same correlation. Differences in robustness results will not be trivial, and data drawn from a distribution such as the one in Figure 36(c) would most likely be analyzed by a researcher as a set of two subpopulations. Summarizing, we see that shape of a distribution is an important factor in the outcome of Monte Carlo studies designed to assess the robustness of statistical procedures has been neglected. Such neglect may, lead to failure to replicate results and contradicting evidence across Monte Carlo studies.

Discussion

Due to the convenience of and the familiarity with normal distribution theory, many statistical methods are still developed utilizing the assumption of an underlying normal distribution of the data, an assumption that is often severely violated for real data sets (Micceri, 1989). One hopes that subsequently the method in question can be shown to be robust to nonnormality, and robustness to nonnormality is traditionally demonstrated in Monte Carlo studies, utilizing simulated, artificial data sets. Simulation of nonnormal multivariate data and

151

subjection of multivariate procedures to tests of performance under nonnormality is becoming increasingly more feasible as computer power increases. Parallel to computing power, the number of methods available for generating nonnormal multivariate data has also rapidly increased. Because the development of multivariate simulation techniques is a rather new area of research, it is important to familiarize applied psychometricians and other methodologists with them, as they will eventually be the consumers of nonnormal data simulation techniques. At the same time, we need to provide applied researchers with tools that can help them understand the possibilities and limitations of the different methods for simulating multivariate nonnormal data.

Observations. Part I of my dissertation analyzes three nonnormal data simulation methods in depth: Multivariate extensions of the 3rd order and the 5th order polynomial transform and the multivariate *g*-and-*h* distribution. All three methods use a nonlinear transformation of a standard normal variable to create a nonnormal variable, and provide control over its first few moments. The 5th order polynomial method differs slightly from the other two methods: While the 3rd order polynomial method and the *g*-and-*h* distribution only provide control over the first four moments of the nonnormal variable, the 5th order polynomial method controls the first *six* moments. This additional control may seem to be an advantage and does provide improved fit to known distributions, as demonstrated in Headrick & Kowalchuk (2007). However, it is also accompanied by a fundamental problem: If we want to simulate nonnormal distributions with just any skewness-kurtosis combination, we will be hard-pressed to make a choice for the 5th and 6th moment. Very little is known about these moments and only when we strive to simulate distributions for which the 5th and 6th moments are known will we be able to make full use of the 5th order polynomial technique. Therefore, with respect to their properties and issues, only a

limited comparison between the 3rd order power and the *g*-and-*h* method and the 5th order power method is possible. While the 5th order polynomial method shares some of the concerns and cautions with the other methods, I concentrated on the 3rd order polynomial method and the *g*-and-*h* distribution.

This study provides particularly detailed insight into the limitations of the 3rd order polynomial method, which, to this date, has been the most popular simulation method in psychometrics. Its advantages and disadvantages compared to the 5th order power method and the multivariate *g*-and-*h* distribution are presented. Integrating and exceeding previous research, I discuss limitations and issues of the power method such as the range of skewness-kurtosis combinations available, monotonicity of transformations, range of final correlations, and odd-shaped distributions.

Because all three simulation methods (3rd and 5th order power method and *g*-and-*h* distribution) use a direct (nonlinear) transformation of a normal variable *Z* to create a nonnormal variable *Y*, it can be fruitful to investigate the relationships between properties of the transformation and properties of the resulting nonnormal distribution. The property that has received the greatest amount of attention so far is the monotonicity of the transformation (Headrick, 2002; Headrick, 2004; Headrick & Kowalchuk, 2007; Hoaglin, 1985; Kowalchuk & Headrick, 2010). Monotonicity of the transformation has been a focus of investigation for two reasons: 1) Hoaglin, who only discussed the *g*-and-*h* distribution, emphasizes the importance of monotonicity, as a monotonic transformation allows us to link quantiles of *Z* to the quantiles of *Y*. With a direct expression for the quantiles of *Y* – and he anticipates Headrick's work – we have access to the pdf of *Y* and can link its quantiles to any nonnormal shape we wish. 2) Headrick and others have placed emphasis on monotonicity of the transformation because their derivation

of an analytical form of the pdf of *Y* relies on $f'(Z) \neq 0$ for any *Z*, thereby requiring the derivation of the transformation to be strictly increasing (or decreasing).

Does this mean that we should only use monotonic transformations? When both a monotonic and a non-monotonic transformation are available for the 3rd order power method, we have seen that, aside from rank correlations and "valid" pdfs, monotonic transformations will also create better behaved distributions. But what if no monotonic transformation is available for a certain skewness-kurtosis combination? For example, no monotonic 3rd order polynomial transformation exists that can simulate a χ_1^2 variable. The criterion of monotonicity will also exclude the straightforward transformation $Y = a + bZ^2$. It remains to be decided whether that means that we should not simulate that particular distribution at all or whether we should find different criteria that help us choose the best non-monotonic transformation available.

The expected value and variability of sample skewness and kurtosis at a given sample size have received insufficient attention in the past, and some of the results that have been published seem to be substantially in error. These previous results create the impression that sample skewnesses and kurtoses will most often have values close to the nominal values specified by the researcher. However, the specified values are *population* values γ_1 and γ_2, and especially for smaller *N*, their sample equivalents $\hat{\gamma}_1$ and $\hat{\gamma}_2$ will differ substantially. Additonally, if we simulate the same values for γ_1 and γ_2 with different simulation techniques, e.g., 3rd order power method vs. *g*-and-*h* distribution, the sampling distributions for $\hat{\gamma}_1$ and $\hat{\gamma}_2$ at a given sample size can be gravely different as well. It seems that some transformations regulate skewness and kurtosis through outliers more than others. If a method heavily relies on extreme outliers to create a distribution with a large asymptotic skewness or kurtosis, it is likely that

smaller samples drawn from that distribution will not contain any of these extreme outliers, and therefore, the average sample skewness or kurtosis will be much smaller than their asymptotic values. If we test a method that is sensitive to skewness and kurtosis for its performance under nonnormality, choice of simulation method may have an influence on the results if expected value and variability of sample skewness and kurtosis for some given sample size differ greatly between simulation methods. An example could be the contrast between $\overline{\hat{\gamma}_1}$ and $\overline{\hat{\gamma}_2}$ for the monotonic vs. non-monotonic 3rd order power method distribution with $\gamma_1 = 2$ and $\gamma_2 = 40$ in Table 27.

The range of final correlations available between two nonnormal variables can be severely limited. Such limitation depends on marginal skewnesses and kurtoses, but also on the particular transformation to nonnormality for the variables. This issue was never broached before in the context of any specific simulation method, and only discussed theoretically in Li & Hammond (1975).

A few univariate moments such as marginal skewnesses and kurtoses are insufficient to describe multivariate nonnormal distributions. Aside from properties of sample statistics and ranges available for final correlations, some skewness-kurtosis combinations can lead to severely odd-shaped distributions. Without careful plotting and examination, these odd-shaped distributions may be used in simulation analyses unnoticed. Vale & Maurelli's method for simulating multivariate nonnormal distributions has been quite popular. Keeping marginal skewnesses and kurtoses as well as correlation constant, I compare coverage rate and balance of two different confidence intervals for a single correlation calculated from two or four differently shaped Vale & Maurelli distributions and the promising multivariate *g*-and-*h* distribution. Choice of distribution and thereby shape in some cases shows a strong influence on performance,

155

leading to very different results for coverage rate and balance. Whether odd-shaped distributions created by the Vale & Maurelli method, e.g., have been used in any of the studies listed in Table 28 is unknown. Finally, I demonstrate that some of the nonnormal distributions that have been utilized in Monte Carlo research on the robustness of correlational methods are of questionable relevance for the practical researcher, because the distributions in question deviate grossly from the simple linear model the product-moment correlation is intended for. Such deviation is obvious even for samples of relatively small size. Robustness to such less suitable distributions is of *less interest*, as researchers investigating real data will be less likely to apply a single correlation to samples drawn from such distributions, provided they plot their data. As this lack of suitability can generally not be detected through knowledge of marginal skewnesses and kurtoses alone, it is recommended that distributions to be used for robustness studies are plotted and visually inspected for their suitability.

Practical Consequences. Asymptotic marginal skewness and kurtosis are hardly sufficient to characterize univariate or multivariate nonnormal distributions. When multivariate nonnormal data are simulated in a Monte Carlo study, authors should be required to specify not only the method that was used for data simulation, but also transformation coefficients or other corresponding information that will enable any reader to exactly reproduce the results and avoid the unintentional use of odd-shaped distributions. Ideally, a routine for nonnormal data simulation should *automatically* provide the user with a representative graph of the distribution created with the program. With publication increasingly moving to online versions of journals (or pure online journals), space limitations are less severe and inclusion of figures, tables, and plots is much more viable. Whenever of potential interest, plotting nonnormal distributions utilized in

156

simulations should be encouraged, such as seen in Beasley et. al. (2007). Thus increased transparency will help in making research reproducible.

Alternative Approaches & Ideas. Alongside the three methods for nonnormal data generation discussed above, which all involve moment matching and subsequent transformation of a normal variable *Z*, other approaches for nonnormal data simulation have been suggested. For example, the *g*-and-*h* distribution was initially developed in the context of nonnormal distribution *quantile* fitting. Hoaglin & Peters (1979) describe several examples of how values for the coefficients *g* and *h* are determined by matching quantiles of an existing nonnormal distribution with quantiles of a *g*-and-*h* distribution. They further suggest expressing *g* and *h* each as a non-linear function of *Z*-values, thereby introducing even greater flexibility to the *g*-and-*h* distribution. Field & Genton (2006) extend the quantile fitting method to the multivariate case. The quantile method of fitting a *g*-and-*h* distribution to an existing nonnormal distribution is different from the moment fitting suggested by Headrick et al. (2008). Headrick et al (2008) and Kowalchuk & Headrick (2010) argue that the increased simplicity of generating nonnormal distributions by characterizing them by their (marginal) skewness and kurtosis is an improvement over the more involved quantile fitting. However, as we have seen, potential users of nonnormal data simulation procedures need to be made aware that such characterization can be inadequate to represent the full complexity of multivariate, or even of univariate distributions.

One of the most salient univariate examples are the two distributions in Figure 8, which have essentially the same first four moments, but radically different shapes. Hoaglin (1985, pp. 504) provides a very short discussion of quantile vs. moment fitting in which he hints at the "resistance and convenience" of working with quantiles. Hence, the advice of only relying on

marginal skewnesses and kurtoses cannot be taken without further precautions and simulated or fitted distributions must at least be plotted. As an advantage, the *g*-and-*h* moment matching procedure allows creation of nonnormal distributions with specified skewness and kurtosis without an already existing "blueprint" distribution.

Future Directions. Monotonicity vs. lack of monotonicity may not be a sufficient criterion to choose between transformations and we may need additional strategies. Headrick & Kowalchuk (2007) seem to make an attempt by discussing the correlation $\rho_{p(Z),Z}$ between normal scores and transformed scores (in their notation Z and $p(Z)$), requiring $\rho_{p(Z),Z}$ to be $\in (0,1]$, i.e. positive and as large as possible. However, I am not convinced that this is necessarily the best measure possible: Considering the example of a (scaled) χ_1^2 variable, which is simply constructed as $Y = a + bZ^2$, the correlation between the two variables is equal to zero, $\rho_{p(Z),Z} = \rho_{Y,Z} = 0$, but Y is a very good example of a well-behaved nonnormal variable. Since the standard normal variable Z has a distribution that is symmetric about 0, I suggest exploring the relationship between $Z \in (-\infty, 0]$ and Y and between $Z \in [0, \infty)$ and Y separately as a first step. Last but not least, some non-monotonic transformations seem to produce very good simulation results by some standard, see e.g. Figure 21(b). Additional examination of the properties of transformation functions and their connection with the intermediate correlation may be necessary if we want to identify the factors that lead to odd-shaped bi-/multivariate distributions. After showing that marginal skewness and kurtosis, the measures of nonnormality people are most familiar with, are insufficient for describing multivariate nonnormality, the question remains whether there are other measures of nonnormality that could provide us with

158

better information about the shape of bi- or even multivariate distributions. One such suggestion

has been made by Mardia (1970), who introduced a coefficient for multivariate skewness. No

attempt has been made in my study to verify whether Mardia's coefficient is able to detect

differences between distributions with the same marginal skewnesses and kurtoses but different

higher moments. As long as these questions remain unanswered and measures of multivariate

nonnormality that can make these important distinctions between distributions unidentified, we

must rely on careful visual inspection of simulated distributions when using them for Monte

Carlo studies. Needless to say that the same care has to be exercised when statistical models are

applied to real data.

The connection between moments of a distribution and the possible range of final

correlations requires closer examination. Carroll (1961, page 349) writes about distributional

shape and calculating correlation coefficients: "As the actual data depart from a fit to such a

model (bivariate normal or linear regression), the limits of the correlation coefficient may

contract…" As an example, he cites a bivariate dichotomous distribution: If the dichotomy is

asymmetrical between the two variables, the product-moment correlation (phi coefficient) will

not range between –1 and 1. He further writes: "But even when the distributions have more than

two class intervals, the possible range of the correlation coefficient is constricted to the extent

that the two marginal distributions are disparate, i.e., not of identical shape and skew." A general

framework for the shape of distributions may be worth investigating; for certain combinations of

marginal skewnesses and kurtoses, a particular shape may be predetermined. For example, when

γ_{11} and γ_{12} (the two skewnesses) for a bivariate distribution are different from each other, it is

possible that the relationship between the variables must be nonlinear. Halperin (1986; see also

Carroll, 1961) comments on page 4: "If the marginal distribution of X looks different from that of

159

variable *Y*, the variables cannot attain a perfect linear relationship, or a correlation of ± 1. … The problem is reduced if variables have relatively symmetric marginal distributions." If the overall shape of a bi- or multivariate distribution is partly predetermined once certain characteristics of the marginal distribution have been fixed, this will be of great interest for researchers who wish to simulate nonnormality. For example, Sterba (personal conversation) attempted to use the 3rd order polynomial method and the method developed by Ruscio & Kaczetow (2008) to simulate a bivariate nonnormal distribution with $Y_1 \sim N(0,1)$ and $Y_2 \sim \chi_3^2$. The bivariate relationship turned out to be nonlinear for both methods, an undesired outcome. Fortunately, the distributions were plotted, the nonlinear relationship discovered, and the simulation of nonnormality suspended. Researchers who plan to simulate multivariate nonnormality need to be made aware of such possible outcomes and limitations.

Distributions that are used to evaluate performance of a statistical procedure when assumptions are not met should reflect the violations of assumptions in the real world. Both the behavior of real-world distributions and the behavior of real-world *researchers* need to be taken into account. A recommendation for future research might be to compare nonnormality in existing real world data sets with nonnormal distributions that can be created with the 3rd order power method, the *g*-and-*h* distribution and other potential simulation techniques. It would be informative whether nonnormality in real world data sets leads to similar robustness or lack of robustness results as artificially simulated data sets. Real data may be lumpier, slightly non-linear, or suffer from cut-off effects as opposed to the distribution simulated by such methods as described in Part I of my dissertation. Interestingly, most studies exclusively simulate IM distributions – nonnormal distributions with identical marginals and therefore identical marginal skewnesses and kurtoses. However, it would be unrealistic to expect real data to always have

identical marginals, more often than not, marginal distributions of real data will have different skewness-kurtosis combinations. We may attempt to categorize bivariate nonnormal real data distributions, similar to the work done by Micceri (1989) for univariate distributions. Very large data sets like the one collected by Srivastava, John, Gosling, and Potter (2003) could be utilized for such an investigation. Subsequently, we may employ an appropriate selection from these bivariate nonnormal real data sets in Monte Carlo research instead of simulating bivariate nonnormal data blindly according to properties of marginal distributions. As storage and processing of large amounts of data become more feasible with increased computing power, it may be a valuable idea to consider bootstrapping from very large real nonnormal data sets to obtain smaller samples and conduct Monte Carlo studies. Ideas in that direction seem to be suggested with the iterative method by Ruscio & Kaczetow (2008), which can be used to emulate real data sets. Raju, Pappas, & Williams (1989) used a large set of real data to conduct an empirical Monte Carlo study on correlation and regression and Nanna & Sawilowsky (1998) also sampled from large real data sets to carry out a Monte Carlo study on the power of the t test vs. the Wilcoxon rank-sum test.

Lastly, multivariate binomial, Poisson, etc. distributions as discussed in Olkin (1994) or Sarabia & Gomez (2008) might be a better alternative to general nonnormal multivariate distributions if the process that creates these distributions can be used to closely model the event we are interested in. The most common example is the binomial distribution which mimics processes that are the sum of independent events with two possible outcomes, and constant probability of success, such as voting behavior in an election.

161

PART II

Robustness of Correlations Revisited

I began my dissertation by pointing out in the general introduction that many statistical procedures, starting with ANOVA, t-tests, and tests on correlations, as well as higher level techniques, which are often build on the former, make the assumption of normally distributed data. Because real data are more often than not nonnormally distributed (Micceri, 1989), we need to ensure that statistical procedures will still work as expected when they are applied to nonnormal data sets. If a study on robustness to nonnormality is executed and conclusions are to be drawn on whether the procedure can be safely applied to nonnormal data or whether one procedure is better than another, the method for simulating nonnormality needs to be chosen carefully. Part I of my dissertation introduced and examined three relatively popular or promising methods for simulating multivariate nonnormality, the 3rd order polynomial method, the 5th order polynomial method, and the multivariate g-and-h distribution. All three of these methods offer control over marginal skewness and kurtosis as well as the correlation matrix between the nonnormal variables. In addition, the 5th order polynomial method offers control over the 5th and 6th moment of the univariate nonnormal variable, but at the same time also *requires* these additional moments to be specified. Because little is known about these moments and generally researchers are only interested in varying skewness and kurtosis to create nonnormality, and because the g-and-h distribution suffers from several technical difficulties (see Part I), we will concentrate on the 3rd order power method only in Part II of my dissertation.

162

Part I has investigated numerous properties of the 3rd order power method. Most notably, even with same marginal skewnesses and kurtoses and the same correlation(s) between the nonnormal variables, distributions created from different sets of transformation coefficients may still have very different shapes. Further, properties such as expected value and variability of sample skewness and kurtosis for a given sample size may not be the same. part I also demonstrated that these newly discovered differences can lead to profound differences in results from Monte Carlo simulations. Part II is now going to test the robustness of several confidence intervals around a single correlation *r*, the sample estimate for ρ, when calculated for several nonnormal distributions. These nonnormal distributions will be characterized by their skewness-kurtosis combinations and simulated using the 3rd order power method. While Part I focused on the creation and selection of nonnormal data for simulation studies, Part II will take a close look at what to do with simulation results once they have been obtained. The evaluation of confidence interval performance for CIs around a single correlation takes on the role of the exemplary statistical procedure.

Robustness of the Sampling Distribution of r

The exact sampling distribution of Pearson's product moment correlation coefficient when calculated for an underlying bivariate normal parent distribution was derived in 1915 by Fisher. Early on, researchers have wondered whether and to what extent the sampling distribution of *r* is sensitive to violations of bivariate normality of the parent distribution. If such sensitivity could be demonstrated, the nominal performance of tests and confidence intervals for correlations that are based on the normality assumption would be in question as well.

163

Research on robustness to nonnormality commenced in the 1920's, and a good overview of early studies can be found in Kowalski (1972). Kowalski summarizes findings on the sensitivity of the sampling distribution of r to various deviations from bivariate normality. He cites both studies contending that the distribution of r is robust as well as studies that seem to have detected a sensitivity of r's distribution to deviations from normality. Haldane (1949) attempted to analytically derive the effects of marginal skewness vs. kurtosis of the parent distribution on r's sampling distribution. He concludes that marginal kurtosis had a much more substantial effect on r's sampling distribution than skewness. Norris & Hjelm (1961) hoped to settle the question of r's robustness to nonnormality by using larger numbers of replications, since previous studies, especially ones that found r's sampling distribution to be non-robust (Baker, 1930), used only a very small number of replications. They employed five different types of distributions: a) normal, b) rectangular, c) leptokurtic, d) slightly skewed, e) markedly skewed, each with both $\rho \approx 0$ and $\rho \approx .83$ and sample sizes of $N = 15$ (15,984 replications), $N = 30$ (7,992 replications), and $N = 90$ (2,664 replications). They observed surprisingly good performance for all sample sizes when $\rho \approx 0$, with only slightly low Type I error rate for the leptokurtic distribution. When $\rho \approx .83$, empirical Type I error rates deviated from the nominal value $\alpha = .05$, ranging from .0101 (rectangular distribution, $N = 90$) to .1562 (leptokurtic distribution, $N = 90$). Norris & Hjelm observe that deviations from nominal Type I error levels *increase* as sample size increases (table III, page 269). Kowalski (1972) concludes that the distribution of r is generally robust when $\rho = 0$, but for $\rho \neq 0$, especially with increasing kurtosis, the variance of r will deviate increasingly from its value under normality. He adds to existing studies by including results for a set of mixture distributions (distributions consisting of a mixture of two normal distributions with a different correlation for each) as discussed in Part I

164

that do lead to non-robustness of the distribution of *r* when overall $\rho = 0$. He also demonstrates

non-robustness for several other bivariate nonnormal distributions with $\rho \neq 0$. However, his

results are based on only 100 replications.

Newer studies on robustness of correlational procedures include work by Duncan &

Layard (1973), who compare coverage rate as well as Type I error and power of the Fisher *Z* CI

and significance test, Jackknife procedures, and a Box confidence interval and test. Zeller &

Levine (1974) seem to obtain results similar to Norris & Hjelm (1961): Overly conservative

Type I error rates for distributions with small kurtosis and increase of that effect as sample size

increases. Havlicek & Peterson (1977) and Edgell & Noon (1984) examined the effects of

nonnormality and type of scale on hypothesis tests of $\rho = 0$. Edgell & Noon (1984) included

additional conditions of bivariate nonnormality that had not been taken into account previously,

such as more extreme deviations from normality and bivariate distributions that had marginals

with different shapes. They generally found good agreement between nominal and empirical

Type I errors for $\alpha = .05$ and some inflation for $\alpha = .01$ test when $\rho = 0$ and the variables were

truly independent. Type I error tended to be inflated when bivariate distributions were

constructed from a mixture of two normal distribution with $\rho_1 \neq 0$ and $\rho_2 \neq 0$, but overall

$\rho = 0$.

What Am I Going to Add to the Robustness Literature?

As we have seen in the previous section, the robustness of the product-moment

correlation to nonnormality of the parent distributions has received much interest and resulted in

a number of simulation studies with mixed results. When two variables are truly independent, the

sampling distribution of ρ seems to be fairly robust to violations of nonnormality, and therefore

hypothesis tests and confidence intervals will be as well. When $\rho \neq 0$ or when $\rho = 0$ but the variables are statistically dependent, it seems to often be the case that the sampling distribution of ρ has an altered shape and statistical procedures are not robust. I believe that the perspective on robustness of CIs to nonnormality is incomplete in several respects. On the one hand, nonnormality is, in the vast majority of cases, described only in terms of marginal skewnesses and kurtoses. Traditionally, no further aspects of the underlying parent distribution such as shape are taken into account. We have already seen in Part I of this dissertation that shape of a distribution beyond marginal skewness and kurtosis can have a significant impact on the behavior of confidence intervals for a single correlation. On the other hand, I believe that the concept of confidence interval performance has not been explored in all its aspects. Often, the only measure employed in robustness studies is coverage rate, or, in some cases, Type I error rate, thereby treating the confidence interval solely as a substitute for hypothesis tests. Confidence intervals have a whole set of additional characteristics, however, and Part II of my dissertation shall be dedicated to exploring these additional properties for five different approaches to constructing an approximate confidence interval for ρ.

The Aspect of Sample Size. In addition to the odd shapes observed for some of the studies mentioned in Part I of my dissertation (see Figure 36), many robustness studies choose rather small sample sizes, which especially for tests and CIs on correlations lead to low precision. With a range of –1 to 1, the parameter space for ρ is relatively limited; hence, an informative confidence interval should part off a substantially smaller range to provide meaningful information. When $r = 0$, the Fisher Z confidence interval for a sample of size $N = 5$ is $\left[-.882, \quad .882 \right]$. Confidence intervals of this size have been routinely included in a wide range of

studies, both on the robustness of the traditional Fisher Z and of bootstrap CIs (see, e.g., Edgell & Noon, 1984; Lee & Rodgers, 1998; Rasmussen, 1988). For an observed sample correlation of $r = .8$ and $N = 5$, the Fisher Z CI is $[-.280, \quad .986]$, still not particularly interesting. Yet, this condition has been considered even in recent studies such as Beasley et al (2007). For which sample size does the width of the confidence interval shrink to a size that carries a substantial amount of information? Confidence intervals around correlations become narrower as either sample size or the absolute size of the correlation increase. For $r = .8$ and $N = 20$, the Fisher Z CI is $[.553, \quad .918]$, having a total width of just under .4; when $r = 0$ and $N = 20$, the Fisher Z CI is $[-.443, \quad .443]$, with a width of almost .9. When $N = 60$, the CI for $r = 0$ is $[-.254, \quad .254]$. I suggest choosing a sample size so that confidence interval width (as estimted by the Fisher I CI) does not exceed .5.

Confidence Intervals for Correlations

I will evaluate the performance of several different confidence intervals for a single correlation r. Confidence intervals will be constructed based on three main approaches: The traditional Fisher Z statistic, asymptotically distribution free theory, and the bootstrap approach. I will use the well-known Fisher Z confidence interval as an indicator of the severity of deviation from nominal performance, subsequently comparing its performance to the performance of an asymptotically distribution free confidence interval around r (derived from results presented in Steiger & Hakstian, 1982), an asymptotically distribution free Fisher Z CI, and two different approaches towards a bootstrap confidence interval around r. Nonnormality will be simulated using the 3rd order power method. A discussion of alternative approaches to measuring association under nonnormality is offered at the end of Part II.

The Fisher Z Confidence Interval and the Asymptotically Distribution-free Confidence Interval

These confidence intervals have already been described in Part I of this study. For convenience, remember the following definitions:

$$z_\alpha = \Phi^{-1}(\alpha) \tag{98}$$

is the inverse standard normal cumulative distribution function, so that $z_{.05} = \Phi^{-1}(.05) = -1.645$ and $\Phi(1.96) = .975$. Let

$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \quad \text{and} \quad \tanh^{-1}(x) = \frac{1}{2} \ln\left(\frac{1+x}{1-x}\right) \tag{99}$$

be the hyperbolic tangent and inverse hyperbolic tangent of a variable x.

The Fisher Z confidence interval (see Fisher, 1915) for the product moment correlation r is constructed as follows:

$$[LL_z, \quad UL_z] = \tanh\left(\tanh^{-1}(r) \pm z_{1-\alpha/2} \times \sqrt{1/(N-3)}\right) \tag{100}$$

To recall the construction of the asymptotically distribution-free confidence interval, define

$$r_{ijkh} = \frac{1}{N} \sum_{m=1}^{N} (Z_{mi})(Z_{mj})(Z_{mk})(Z_{mh}). \tag{101}$$

and

$$\begin{aligned}
\hat{\gamma}_{ijij} &= r_{ijij} + \frac{1}{4} r_{ij} r_{ij} \left(r_{iiii} + r_{jjii} + r_{iijj} + r_{jjjj}\right) - \frac{1}{2} r_{ij} \left(r_{iiij} + r_{jjij}\right) - \frac{1}{2} r_{ij} \left(r_{ijii} + r_{ijjj}\right) \\
&= r_{iijj} + \frac{1}{4} r_{ij}^2 \left(r_{iiii} + 2r_{iijj} + r_{jjjj}\right) - r_{ij} \left(r_{iiij} + r_{ijjj}\right)
\end{aligned} \tag{102}$$

The asymptotically distribution-free CI is then:

$$\left[LL_{ADF}, \quad UL_{ADF} \right] = r \pm z_{1-\alpha/2} \sqrt{\dfrac{r_{iijj} + \dfrac{1}{4} r_{ij}^{2} \left(r_{iiii} + 2r_{iijj} + r_{jjjj} \right) - r_{ij} \left(r_{iiij} + r_{ijjj} \right)}{N-2}} \quad (103)$$

Asymptotically Distribution-Free Fisher Z Confidence Interval. The variance estimator

from Equation (102) can also be used for an asymptotic Fisher Z confidence interval (see section

4 in Steiger & Hakstian, 1982). Instead of $\sqrt{1/(N-3)}$ as a standard error estimate, we can use

$\sqrt{\hat{\gamma}_{1212} / \left(\left(1 - r_{12}^{2} \right)^{2} (N-3) \right)}$, modifying Equation (100) to

$$\left[LL_{z^*}, UL_{z^*} \right] = \tanh \left(\tanh^{-1}(r) \pm z_{\alpha/2} \times \sqrt{\dfrac{\hat{\gamma}_{1212}}{\left(\left(1 - r_{12}^{2} \right)^{2} (N-3) \right)}} \right) \quad (104)$$

Applied to the small sample of 5 data points $(2,10)$, $(7,7)$, $(8,2)$, $(5,4)$, and $(1,9)$ with

sample correlation $r = -.80965$ from Part I (see Figure 35 for a graph of these data points), we

find:

$$r_{12} = -0.8096501$$
$$r_{12}^{2} = 0.6555334$$
$$r_{1111} = 1.099827$$
$$r_{1112} = -0.916156$$
$$r_{1122} = 0.9563612$$
$$r_{1222} = -1.034218$$
$$r_{2222} = 1.217355$$

So that

$$\hat{\gamma}_{1212} = r_{1122} + \frac{1}{4} r_{12}^{2} \left(r_{1111} + 2r_{1122} + r_{222} \right) - r_{12} \left(r_{1112} + r_{1222} \right)$$
$$\approx 0.956 + \frac{1}{4} 0.656 \left(1.010 + 2*0.956 + 1.217 \right) - -0.810 \left(-0.916 - 1.034 \right)$$
$$\approx 0.055$$

The interval is constructed as

$$\left[LL_{z^*}, UL_{z^*} \right] \approx \tanh\left(-1.127 \pm 1.96 \times \sqrt{\frac{.055}{2(1-.656)^2}} \right)$$

$$= \tanh\left[-2.072, -.182 \right] \tag{105}$$

$$\approx \left[-.969, -.180 \right]$$

Again, this example suffers from rounding and the final confidence interval limits in Equation (105) are displayed for consistency in the calculation "path" only. A more exact value for the asymptotic Fisher Z confidence interval is $[-.975, -.058]$.

Bootstrapping Procedures for Correlations

Bootstrapping is a computationally intensive estimation technique that was suggested by Efron around 1977/79 (Efron, 1979). It has inspired a substantial amount of research, and, as computing power advances, is becoming increasingly feasible for a variety of statistical procedures. Bootstrapping circumvents the necessity of making distributional assumptions (other than that the sample obtained is sufficiently representative), such as underlying population normality, by basing inferences solely on information carried by the sample at hand. From the observed sample of size N, a *sampling frame* is constructed. This sampling frame can be just simply the observed data set or some other set of points based on the originally observed data. N data points are drawn randomly and with replacement from the sampling frame into a *bootstrap sample* and, if the parameter of interest is a correlation, $\hat{\rho}^*$, the sample estimator of ρ, is calculated for this bootstrap sample. This process is repeated B times, creating an empirical sampling distribution for ρ out of the $\hat{\rho}^*$'s. The empirical sampling distribution is used to construct confidence intervals or execute hypothesis tests. Efron & Gong (1983) provide a clear

introduction to the bootstrap in general as well as a description of bootstrap confidence intervals for a single correlation; a less technical description can be found in Diaconis & Efron (1983) or Lunneborg (1985).

Several bootstrap hypothesis tests and confidence intervals for a single correlation r have been suggested, with probably the first description of bootstrap techniques for r presented in Efron (1982). Among these confidence intervals were a simple *percentile* bootstrap CI, an *adjusted* bootstrap CI, a *bias-corrected* bootstrap CI, and a *bias-corrected and accelerated* bootstrap CI (BCa CI). All of these confidence intervals use a *bivariate* sampling frame as described in the following section which illustrates both the simple percentile method and the BCa CI.

The Percentile Bootstrap Confidence Interval

The construction of the percentile bootstrap confidence interval around r is conceptually simple:

1. Collect a sample of N bivariate data points (pairs of X and Y values). This is also the sampling frame, called a *bivariate sampling frame*.

2. Randomly select, with replacement, N (bivariate) data points from the sampling frame and calculate the bootstrap sample correlation r^* for this bootstrap sample.

3. Repeat step 2. B times (B tends to range from 999 to 9,999, see Efron (1987) for bootstrap sample size calculations). The set of bootstrap sample correlations created by repeating steps 2 is the empirical sampling distribution of r, $\left\{ r_1^*, ..., r_B^* \right\}$.

4. A $100(1-\alpha)\%$ confidence interval is constructed by choosing the $100(\alpha/2)^{th}$ and the $100(1-\alpha/2)^{th}$ percentile of this empirical sampling distribution.

171

If we define $B(x)$ as the empirical cumulative distribution function of a bootstrap sample and if we have an *ordered* empirical sampling distribution of 100 bootstrap correlations, $\left\{ r_{(1)}^{*},...,r_{(100)}^{*} \right\}$, the value of the empirical cdf at the smallest bootstrap correlation is $B\left(r_{(1)}^{*} \right) = .01$. The percentile bootstrap confidence interval is then defined as

$$\left[LL_{p}, UL_{p} \right] = \left[B^{-1}\left(\alpha / 2 \right), B^{-1}\left(1 - \alpha / 2 \right) \right] \tag{106}$$

Bivariate Sampling Bootstrap Example. We use the same small sample as for the Fisher *Z* and the asymptotically distribution free confidence intervals. Since the sampling frame is nothing other than the original sample, no additional graph or table is presented. Next, we draw $B = 2000$ samples of size $N = 5$ from the sampling frame with replacement. Note that with such a small sampling frame, it can be expected that some of these samples will include all equal points and no correlation can be calculated; in that case, the correlation for the specific bootstrap sample is set equal to zero. This occurred six times. The first and last 100 values of the empirical sampling distribution thus created are displayed in Table 48:

Table 48: Bootstrap Correlations for Small Example, Bivariate Sampling Approach

−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
									…
−0.325	−0.325	−0.325	−.325	−0.099	−0.099	−0.099	−0.099	−0.099	−0.099
−0.099	−0.099	−0.099	−.099	−0.099	−0.099	−0.099	−0.059	−0.059	−0.059
−0.059	−0.059	−0.059	−.059	−0.059	−0.059	−0.059	−0.059	−0.059	−0.059
−0.059	−0.059	−0.059	−.059	−0.059	0.000	0.000	0.000	0.000	0.000
0.000	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034
0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	1.000	1.000
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

For a 95% confidence interval, we choose the middle of the interval between the 50th and 51st and the middle of the interval between the 1950th and 1951st of these correlations:

$$\frac{r^*_{50} + r^*_{51}}{2} \approx \frac{-(1.000 + 1.000)}{2} = -1.000 \quad \text{and}$$

$$\frac{r^*_{1950} + r^*_{1951}}{2} \approx \frac{(.034 + .034)}{2} = .034 \tag{107}$$

The 95% OI confidence interval for this (very small) example is:

$$\text{Percentile Bootstrap CI} = \left[LL_p, UL_p \right] = \left[-1.000, .034 \right] \tag{108}$$

Bias-corrected and Accelerated (BCa) Confidence Interval

To construct a bias-corrected and accelerated bootstrap CI, we calculate the bias-correction factor

$$\hat{z}_0 = \Phi^{-1}\left[\frac{\#\left(r^*(b) < r\right)}{B}\right] \tag{109}$$

where $\#\left(r^*(b) < r\right)/B$ is the proportion of bootstrapped correlations r^* that are smaller than the

observed sample value r and Φ^{-1} is the inverse of a standard normal cumulative distribution

function. For our small example with $r = -.80965$, $\#\left(r^*(b) < r\right)/B = 1140/2000 = .57$ and

therefore $\hat{z}_0 = \Phi^{-1}(.57) = 0.176$. Further, an acceleration index $\hat{\alpha}$ is computed as

$$\hat{\alpha} = \frac{\sum_{i=1}^{n}\left(r_{[\cdot]} - r_{[i]}\right)^3}{6\left(\sum_{i=1}^{n}\left(r_{[\cdot]} - r_{[i]}\right)^2\right)^{3/2}} \tag{110}$$

where $r_{[i]}$ is the correlation of the original data set with the ith value deleted and

$r_{[\cdot]} = \sum_{i=1}^{n} r_{[i]}/n$, i.e. the mean of the $r_{[i]}$. With $r_{[1]} = -.7446$, $r_{[2]} = -.9552$, $r_{[3]} = -.6634$,

$r_{[4]} = -.8534$, $r_{[5]} = -.7739$, and $r_{[\cdot]} = -.7981$, we calculate the acceleration factor as

$\hat{\alpha} = 0.0218$.

The bias-correction and the acceleration factor are subsequently used to calculate the quantiles

for the lower and upper end of the BCa confidence interval:

$$\left[LL_{BCa}, UL_{BCa}\right] = \left[r^*_{\alpha_1 * B}, r^*_{\alpha_2 * B}\right] \tag{111}$$

where

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{\alpha}\left(\hat{z}_0 + z_{\alpha/2}\right)}\right) \tag{112}$$

and

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{\alpha}\left(\hat{z}_0 + z_{1-\alpha/2}\right)}\right) \tag{113}$$

Entering the values we found for $\hat{z}_0$ and $\hat{\alpha}$ for this particular example into Equations (112) and (113), we find for a 95% confidence interval with $\alpha = .05$:

$$\alpha_1 = \Phi\left(.176 + \frac{.176 - 1.96}{1 - .0218(.176 - 1.96)}\right) = \Phi(-1.541) = .0617 \tag{114}$$

and

$$\alpha_2 = \Phi\left(.176 + \frac{.176 + 1.96}{1 - .0218(.176 + 1.96)}\right) = \Phi(2.4163) = .9922 \tag{115}$$

The corresponding correlations (that is, only the upper limit of the CI) are colored in green in Table 48 and the resulting confidence interval is

$$\left[LL_{BCa}, UL_{BCa}\right] = \left[r_{123}^*, r_{1984}^*\right] = \left[-1, 1\right] \tag{116}$$

This is not a very convincing confidence interval – it covers the entire parameter space of ρ. The BCa method seems to generally require a larger number B of bootstrap samples in the empirical sampling distribution than the percentile method to perform well (Efron, 1987).

After their introduction, the performance of bootstrap hypothesis tests and confidence intervals for r was evaluated in a number of Monte Carlo studies (Efron, 1988; Lunneborg, 1985; Mendoza, Hart & Powell, 1991; Rasmussen, 1987, 1988, and 1989; Strube, 1988). Early simulation studies, such as Lunneborg (1985), Rasmussen (1987), and Efron (1988) concerned themselves with examining whether bootstrap confidence intervals would perform as well as the Fisher Z confidence interval from Equation (100) when $\rho = 0$, often choosing an underlying bivariate normal distribution, i.e. complete independence. Conclusions were somewhat mixed, which may in part stem from methodological differences between studies. Rasmussen (1987)

observed inflated Type I error rates for the percentile bootstrap CI (Equation (106)) calculated for both a bivariate normal and a nonnormal (one variable normal, the other lognormal) distribution, each with $\rho = 0$, while the standard Fisher Z confidence interval performed almost perfectly. Strube (1988) expanded on Rasmussen's study by including the bias-corrected bootstrap CI, which seemed to work better than the percentile CI, and in fact performed close to nominal.

Deciding that the bootstrap should be tested under conditions for which the Fisher Z confidence interval has been shown to perform poorly, Rasmussen (1988) applied the percentile and the adjusted bootstrap (the latter is not discussed in my study) to bivariate mixture distributions with $\rho = 0$ such as used by Kowalski (1972), the most extreme of which has been plotted in Figure 36(a). In cases where the parametric approach is extremely liberal, the bootstrap CI demonstrated better performance, in other conditions performance was about the same or slightly better for the parametric CI. In another study, Rasmussen (1989) showed that the adjusted bootstrap performs better than the Fisher Z CI for the same set of mixture distributions for medium to large sample sizes. Additional studies on the performance of bootstrap confidence intervals for correlations were carried out by Mendoza, Hart & Powell (1991) and Wilcox (1991). Sievers (1996) compares (among others) a version of the asymptotically distribution free CI (similar to Equation (103)) and the percentile bootstrap CI at the 90% confidence level with $N = 19$, observing better performance for the bootstrap confidence interval. To construct bivariate nonnormal distributions, Sievers first creates two nonnormal univariate variables and then multiplies them with the Cholesky factor of the desired correlation matrix, as has been described for normal variables in the section on the 3rd order polynomial method. Notice that this

alters marginal moments for one of the variables. One of these distributions is pictured in Figure 36(d).

In 1998, Lee & Rodgers introduced the *univariate sampling* approach, originally designed to allow for "logically more sound" *nil* hypothesis testing, i.e. testing the null hypothesis $H_0 : \rho = 0$. Potentially motivated by considerations presented in Dolker, Halperin, & Divgi (1982), they argue that when a bivariate sampling bootstrap confidence interval is constructed around $r \neq 0$, the sampling frame reflects the relationship between the variables, and a sampling frame with independent variables should be preferred. While it is slightly unclear how Lee & Rodgers (1998) created their bivariate nonnormal distributions with $\rho \neq 0$ (I was unable to construct a bivariate distribution with one normal, one lognormal marginal distribution and $\rho = .8$), it appears that their univariate sampling approach provides better overall Type I error control than several bivariate sampling bootstrap CIs when testing $H_0 : \rho = 0$. They claim that univariate sampling will show superior performance for very small samples because of a larger sampling frame. Recently, Beasley, DeShea, Toothaker, Mendoza, Bard, and Rodgers (2007) generalized the univariate sampling approach to allow testing the null hypothesis $H_0 : \rho = a$, with $a \in (-1,1)$. The resulting confidence interval is described in the next section.

The OI Univariate Sampling Confidence Interval

The bootstrap confidence interval around a single correlation as derived by Beasley et al (2007) is constructed in the following steps:

1. Collect a sample of *N* bivariate data points (pairs of *X* and *Y* values, Figure 39(a)).
2. Construct a sampling frame.

a. The notable difference between univariate sampling bootstrap for a correlation and more traditional methods (named "bivariate sampling bootstrap" in Beasley et al. (2007)) lies in the construction of the sampling frame. The number of elements in the sampling frame is increased by combining all X values with all Y values, thereby creating a rectangular sampling frame with a correlation of 0, independently of what the observed correlation in the sample has been. See Figure 39(b).

b. Next, the variables in the sampling frame are standardized.

c. Finally, the up until now uncorrelated variables in the sampling frame are correlated by multiplying by the Cholesky decomposition of the desired correlation matrix (see Kaiser & Dickmann, 1962), with the *observed* correlation as off-diagonal element. This changes not only the correlation between the two variates in the sampling frame from zero to the observed correlation, but also the skewness and kurtosis of the second variate. See Figure 39(c).

3. Randomly select N data point pairs from this sampling frame with replacement, thus creating one bootstrap sample.

4. Calculate the sample correlation r^* for the bootstrap sample in 3; repeat B times.

5. The set of B bootstrap sample correlations $\left\{ r^*_1,...,r^*_B \right\}$ is the empirical sampling distribution of r. A $100(1-\alpha)\%$ OI confidence interval $\left[LL_{OI}, UL_{OI} \right]$ is constructed by choosing the $100(\alpha/2)^{th}$ and the $100(1-\alpha/2)^{th}$ percentile of this empirical sampling distribution.

6. Finally, an asymptotic adjustment is applied by widening the CI from step 5 by a factor of $\sqrt{(N+2)/(N-1)}$: Both the center and width of the confidence interval are found: $\text{center} = (UL_{OI} + LL_{OI})/2$, $\text{width} = (UL_{OI} - LL_{OI})/2$. The width is multiplied by the adjustment factor, creating the final margin of error: $\text{ME} = (UL_{OI} - LL_{OI})/2 \times \sqrt{(N+2)/(N-1)}$. Lastly, this margin of error is added to and subtracted from the center: $[LL_{OIa}, UL_{OIa}] = [\text{center} - \text{ME}, \text{center} + \text{ME}]$.

To exemplify the adjustment procedure, a confidence interval of $[-.100 \quad .100]$ around $r = 0$ obtained in step 5 with $N = 5$ will be extended to $[-.132 \quad .132]$, and a confidence interval of $[.100, \quad .135]$ around $r = .24$ will be extended to $[.060, \quad .390]$. Notice that the extension is symmetric about the center of the confidence interval, not necessarily about r.

OI Univariate Sampling Confidence Interval Example. We have the same sample of 5 data points as above: $(2,10)$, $(7,7)$, $(8,2)$, $(5,4)$, and $(1,9)$, with sample correlation $r = -.80965$ (Figure 39(a)). The (uncorrelated) sampling frame is constructed by pairing each X value with each Y value, creating a total of 25 points (see Figure 39(b)). Next, the two variables are standardized and correlated with the originally observed sample correlation $r = -.80965$ (Figure 39(c)).
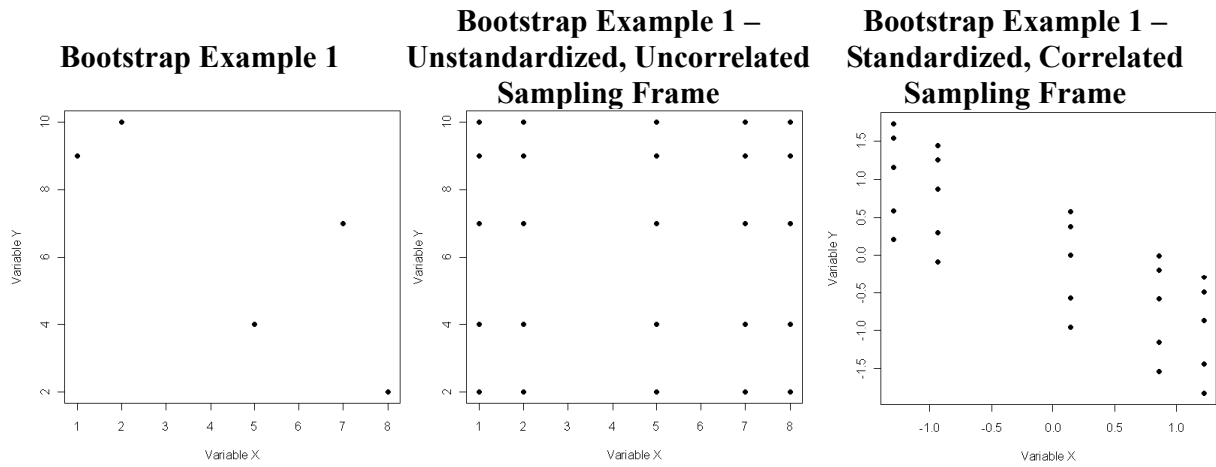
Figure 39(a), (b), and (c): Development of the sampling frame for the univariate sampling bootstrap

Now bootstrap samples of size $N = 5$ are randomly drawn from the sampling frame from Figure 39(c) and sample correlations are calculated for each sample. With $B = 2000$, the first and last 100 bootstrap correlations are listed (and ordered) in Table 49.

Table 49: Bootstrap Correlations for the Small Example, OI Univariate Sampling Approach

–1.000	–1.000	–1.000	–1.000	–1.000	–1.000	–0.999	–0.998	–0.998	–0.998
–0.998	–0.997	–0.997	–0.996	–0.996	–0.996	–0.996	–0.995	–0.995	–0.995
–0.995	–0.995	–0.995	–0.995	–0.995	–0.995	–0.994	–0.994	–0.994	–0.994
–0.994	–0.994	–0.994	–0.994	–0.994	–0.993	–0.993	–0.993	–0.993	–0.993
–0.993	–0.992	–0.992	–0.992	–0.992	–0.991	–0.991	–0.991	–0.991	–0.990
–0.990	–0.990	–0.990	–0.990	–0.990	–0.990	–0.990	–0.990	–0.989	–0.989
–0.989	–0.989	–0.988	–0.988	–0.988	–0.988	–0.988	–0.988	–0.987	–0.987
–0.987	–0.987	–0.987	–0.986	–0.986	–0.986	–0.986	–0.986	–0.986	–0.986
–0.986	–0.986	–0.986	–0.986	–0.986	–0.985	–0.985	–0.985	–0.985	–0.985
–0.985	–0.985	–0.985	–0.984	–0.984	–0.984	–0.984	–0.984	–0.984	–0.984
									…
–0.366	–0.359	–0.357	–0.349	–0.345	–0.344	–0.342	–0.331	–0.330	–0.325
–0.321	–0.313	–0.313	–0.311	–0.311	–0.310	–0.308	–0.300	–0.300	–0.299
–0.297	–0.297	–0.288	–0.287	–0.270	–0.260	–0.252	–0.248	–0.245	–0.238
–0.237	–0.231	–0.231	–0.230	–0.223	–0.220	–0.217	–0.216	–0.215	–0.214
–0.202	–0.194	–0.193	–0.180	–0.177	–0.164	–0.149	–0.142	–0.134	–0.132
–0.119	–0.115	–0.115	–0.109	–0.107	–0.076	–0.072	–0.069	–0.067	–0.063
–0.063	–0.056	–0.045	–0.045	–0.036	–0.022	–0.021	–0.012	–0.011	–0.009
–0.003	0.000	0.000	0.000	0.024	0.025	0.044	0.051	0.061	0.088
0.102	0.124	0.128	0.162	0.168	0.192	0.199	0.253	0.307	0.311
0.312	0.328	0.354	0.415	0.418	0.421	0.463	0.505	0.545	0.563

For a 95% OI confidence interval, we choose the middle of the interval between the 50th and 51st and the middle of the interval between the 1950th and 1951st of these correlations:

$$\frac{r^*_{50} + r^*_{51}}{2} \approx \frac{-(.990 + .990)}{2} = -.990 \quad \text{and} \quad \frac{r^*_{1950} + r^*_{1951}}{2} \approx \frac{-(.132 + .119)}{2} = -.126 \,(117)$$

As a last step, the confidence interval is widened by a factor of $\sqrt{(5+2)/(5-1)} \approx 1.323$: The center of the CI is –.558, and its width is .864. Hence, the new width will be $.864 \times 1.323 = 1.143$. Arranged symmetrically around the center of –.558, we find the adjusted 95% OI bootstrap confidence interval to be $[-1.129, .013]$.

Evaluating Confidence Interval Performance – The Exact CI

The main focus of Part II of my dissertation is the evaluation of confidence interval performance. Aside from Type I error rate, coverage rate is probably the most established index of CI performance, and both coverage rate and coverage balance as described in Part I of this study will be included. Coverage rate and balance will be contrasted with several new approaches to understanding CI performance, described in the remainder of this section. In my dissertation proposal, I included several predictions regarding the performance of the five approximate confidence intervals for *r*. In particular, I predicted the Fisher *Z* CI to depart from nominal performance considerably only when a linear, single population model was inadequate for the distribution at hand. I further stipulated that the distribution-free CIs would outperform the Fisher *Z* CI as sample size increased. Finally, I predicted that overall, the BCa CI would outperform the Univariate Sampling CI, especially for nonlinear relationships, since the sampling fram for the BCa method will increasingly resemble the true population, while the sampling frame for the Univariate CI will not. I also adopted the common prediction that confidence interval performance will improve as sample size increases. In particular, preliminary results had shown that coverage balance will increase, while coverage rate may in some cases decrease. The next section will provide us with an additional tool to investigate there predictions.

Additional Ways of Evaluating Performance: An Exact CI for ρ. Confidence intervals are recommended by methodologists because they provide an interval estimate for the parameter of interest, carrying information on the precision the parameter estimate. Thus, another important property of confidence intervals is their width. The narrower an interval, the more precise is the estimation, assuming equal confidence levels. Some authors have even gone further by

comparing different interval procedures applied to a handful of samples one-by-one. Rasmussen (1988) tabulates endpoints of 99% and 95% confidence intervals, while Efron (1988) provides average confidence limits as well as a one to one comparison for a smaller set of instances for parametric and bootstrap CIs. Sievers (1996) attempts a comparison of parametric and bootstrap confidence intervals with "exact" confidence limits. However, there seems to be a certain inconsistency in his findings: Considering his Table 1, we can see for $\rho = .5$ and distribution 3 (the bivariate exponential distribution), the percentile *t*-method has confidence limits $-.108$ and .824 and coverage rate 86.8%, while at the same time, the "exact" 90% confidence interval has limits .068 and .798, thus being considerably narrower, but supposedly covering the true ρ more often (Sievers does not provide the empirical coverage rate for his exact CI).

Taking a closer look at Sievers' description of an "exact" two-sided central 90% confidence interval, we find that he defines its confidence limits to be the 5th and 95th percentile of an empirical distribution obtained by repeatedly sampling correlations from the distribution of interest (page 385, second paragraph). This does *not* necessarily result in a confidence interval that will miss the true ρ a total of 10% of the time by lying below and above it 5% of the time each! To see why, consider the following: As ρ changes, the shape of the sampling distribution will change as well, including the size of the standard error. Generally, the standard error of ρ will decrease in size as ρ increases in absolute value. This effect is reflected in the simple approximation to the standard error as described, e.g., in Olkin & Finn (1995):

$$SE_r = \sqrt{\frac{\left(1-\rho^2\right)^2}{n}} \qquad (118)$$

It is easy to see that as ρ increases, SE_r from Equation (118) decreases. Hence, width of intervals changes over the range of the parameter space of ρ. Now, choose $\rho = .5$, based on a

bivariate normal distribution: The 5th and 95th percentile of the sampling distribution of *r* when calculated for samples of size $N = 19$ are approximately equal to .153 and .750. However, this does not mean that the 5th percentile of the sampling distribution around $\rho = .750$ is equal to approximately .5! When $\rho = .750$, and $N = 19$, the 5th percentile of the sampling distribution is approximately .529. Therefore, confidence intervals around sample correlations smaller than .750 would already not cover $\rho = .5$, and the 5th and 95th percentile of the sampling distribution prove to be inadequate confidence limits *if* coverage balance is desired. The remainder of this study focuses on comparisons between the approximate CIs introduced in previous sections and a CI with exact coverage. We define an exact two-sided $1 - \alpha$ confidence interval based on work in Steiger & Fouladi (1997), who write (page):

> "An *upper confidence limit* (or *upper confidence bound*) is a statistic that, over repeated samples of size *N*, exceeds an unknown parameter θ a certain proportion of the time. For example, function $B(X)$ is a $1 - \alpha$ upper confidence limit for θ if, …
> $$P(B(X) \geq \theta) = 1 - \alpha ."$$

An *exact* two-sided $100(1 - \alpha)\%$ confidence interval for ρ is then defined as a confidence interval that will lie entirely below and above the true ρ exactly $100(\alpha / 2)\%$ of the time. That is, if $LL_e(X)$ stands for the lower and $UL_e(X)$ stands for the upper limit of the exact CI,

$P(UL_e(X) \leq \rho) = \alpha / 2$ and $P(\rho \leq LL_e(X)) = \alpha / 2$. Such an interval can be constructed with the technique of *pivoting the cdf* as described in Steiger & Fouladi. The technique of pivoting the cdf utilizes the relationship between the population parameter (such as ρ) and the statistic used to obtain sample estimates (correspondingly, *r*) that is created when the cdf is known.

Creating an Exact CI Empirically. When we attempt to apply this procedure for creating

an exact CI to correlations drawn from a nonnormal distribution (potentially even with an

unknown cdf itself), where the cdf of *r* is infeasible or impossible to obtain, we can adapt by

using an empirical approximation to the cdf of *r*. An empirical cdf can be created if large

numbers of samples are drawn, the correlation for each sample is calculated, and the respective

quantiles of the resulting empirical sampling distribution recorded. For example, we may create

50,000 samples of size $N = 20$ from a Vale & Maurelli distribution with known $\gamma_{11} = \gamma_{12} = 0$,

$\gamma_{21} = \gamma_{22} = 25$, and $\rho_Y = .5$. For each sample, we calculate r_Y, resulting in 50,000 sample

correlations, and find the approximate 2.5[th] and 97.5[th] percentile of this empirical sampling

distribution to be .126 and .886. If such simulation of empirical cdfs is repeated for values across

the entire range of ρ_Y, with a sufficient amount of intervals, a functional relationship between

ρ_Y and r_Y can be created. To see how the confidence interval works, consider Figure 40 below,

which is similar to Figure 9.3 in Steiger & Fouladi (1997):

Relating True and Sample Correlation to Another Via the 2.5th and 97.5th Percentile of the Empirical Sampling Distribution
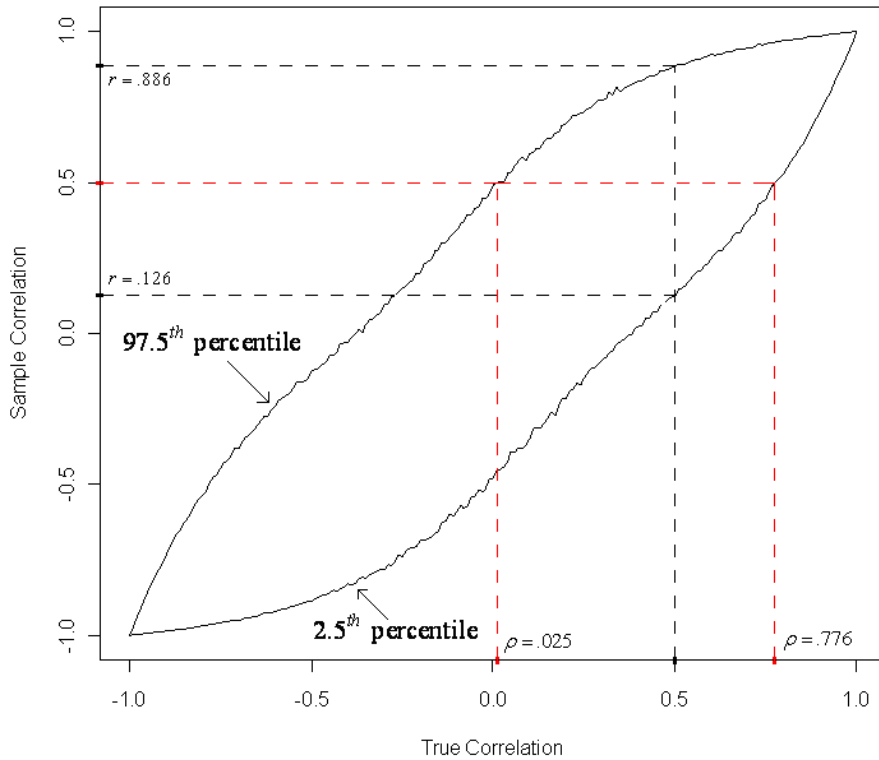
Figure 40: Construction of an Exact CI for ρ when Sampling from a 3rd Order Power Method Distribution

On the x-axis, the entire possible range of population correlations is plotted, while sample correlations are plotted on the y-axis. The two slightly zigzagged graphs are the 2.5th and 97.5th percentile of the empirical sampling distribution of *r* for respective values of ρ on the x-axis. Lack of continuity can be compensated by using a suitable smoothing technique. For example, for $\rho = .5$, we follow the dashed black line vertically to the lower graph and then over to the y-axis to find that the 2.5th percentile of the sampling distribution is approximately equal to .126. Equivalently, the 97.5th percentile is approximately equal to .886. Reversing this process and assuming that we found $r = .5$, following the red dashed line horizontally to the two slightly zigzagged lines and then dropping vertically onto the scale of the population correlation, we

obtain about .025 as a lower and .776 as an upper confidence limit for a 95% confidence interval. The novelty about the interval constructed from Figure 40 is that it is not for a correlation based on a bivariate normal parent distribution, but on a Vale & Maurelli distribution with $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$ (transformation coefficients for both variables: [0   0.2553   0   0.2038]) as in Figure 12(a) with sample size $N = 20$. The graph in Figure 40 was created with 10,000 samples per empirical distribution, dividing the range of ρ into 201 parts:

$\{-1.00, \ -.99, \ -.98, \ \ldots, \ .99, \ 1.00\}$. No smoothing or interpolation were applied.

This empirical way of constructing an exact CI for a single correlation drawn from a nonnormal distribution will be utilized to create an exact CI that the five approximate CIs can be compared to. For the present study, empirical sampling distributions were created for population correlations spaced apart in intervals of .001, i.e., an empirical sampling distribution was created for $\rho_Y = -1$, $\rho_Y = -.999$, $\rho_Y = -.998$ and so forth, up to $\rho_Y = .999$, $\rho_Y = 1$. Since random variation may lead to a situation in which the 2.5th quantile for $\rho_Y = -.999$ is larger than the 2.5th quantile for $\rho_Y = -.998$, smoothing using the lowess function in R with smoothing parameter $f = 0.01$ was applied. For this study, during the gathering of percentiles of the empirical sampling distribution, not only the 2.5th and the 97.5th percentile were recorded, but the 0.5th, 5th, 10th, 16th, 84th, 90th, 95th, and 99.5th percentiles as well. This allows us to construct exact CIs with 68%, 80%, 90%, 95%, and 99% coverage. Once the construction of the Exact CI is complete and a sample correlation obtained, the population correlation values for which this sample correlation would lie at the 97.5th percentile or the 2.5th percentile, respectively, are found as demonstrated in Figure 40 and make up the lower and upper limit for a 95% CI.

In the results section, I provide a small demonstration of the accuracy of coverage rate and balance for the exact CI. However, exact coverage is not the only thing this CI has to offer. In fact, it is actually quite simple to create a confidence interval with nominal coverage for a single correlation. For a 95% CI, let

$$LL = -1 \text{ and } UL = 1 \quad \text{with } p = .95$$
$$LL = UL = -1 \quad\quad\quad\quad \text{with } p = .025$$
$$LL = UL = 1 \quad\quad\quad\quad\;\; \text{with } p = .025$$

This interval will show perfect coverage in all conditions, but have very uninteresting behavior, unrelated to the properties of the obtained sample, in addition to being overly wide. Instead, the lower and upper confidence interval limits of the Exact CI as constructed above provide us with information on the probability of observing a sample correlation given a certain population correlation. For example, if $r = .5$ and $LL_E = .025$ and $UL_E = .776$, we know that

$P(r \geq .5 \mid \rho = .025) = .025$ and $P(r \leq .5 \mid \rho = .776) = .025$. These probabilities will be *larger* for any $.025 < \rho < .776$. Therefore, the Exact CI provides information on the sampling distribution produced at each level of ρ. Interestingly, some of the most popular instances of confidence intervals have this property as well. An example is the traditional confidence interval for a single mean, often the first confidence interval taught in statistics courses and certainly formative as students and researchers develop their expectations towards the statistical concept of confidence interval construction. The traditional CI for a single mean possesses these properties as well, because the standard error is conveniently assumed to be known, or, when it is not, mean and standard error are assumed to be independent of each other.

Utilizing the Exact Confidence Interval. Notice that the construction and use of the exact confidence interval turns out to be more complicated when the simulation technique and the

188

choice of marginal skewnesses and kurtoses for the nonnormal parent distribution does not allow the simulation of the entire range of $\begin{bmatrix} -1, & 1 \end{bmatrix}$ for ρ_Y. This poses a problem because the exact confidence interval is constructed by creating empirical sampling distributions of sample correlations calculated on samples drawn from the underlying nonnormal parent distribution for several levels of ρ. Assume we simulate an nIM V&M distribution with $\gamma_{11} = 2.5$, $\gamma_{12} = 2.0$, $\gamma_{21} = 10$, and $\gamma_{22} = 12$, with simulation coefficients $\begin{bmatrix} -0.6415 & 0.6487 & 0.6415 & -0.1057 \end{bmatrix}$ and $\begin{bmatrix} -0.5435 & 0.9802 & 0.5435 & -0.1553 \end{bmatrix}$. This choice of transformation coefficients cannot simulate population correlations outside the range of $\rho_Y = \begin{bmatrix} -.0106 & .9663 \end{bmatrix}$. Utilizing only the available range and creating sampling distributions for $\rho = -.010$ through $\rho = .966$ in steps of .001 and plotting the 2.5th and 97.5th percentile equivalent to Figure 40, we obtain:

Exact CI for a nDIM Vale & Maurelli Distribution with Skewnesses 2.5 and 2.0 and Kurtoses 10 and 12
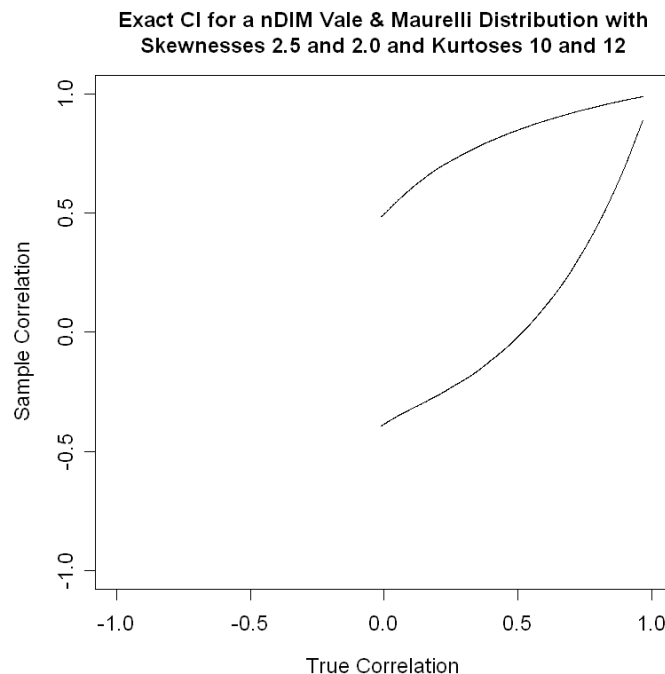
Figure 41: The Exact CI for an nIM Distribution with $\rho \in \begin{bmatrix} -.0106 & .9663 \end{bmatrix}$

189

The issue here is obvious: How would we construct the confidence limits of the exact confidence interval if we observed $r = -.3$, e.g.? There exists a straightforward answer: Since ρ_Y cannot lie outside the range $[-.0106 \quad .9663]$, and we have knowledge of this property of the particular V&M distribution at hand, we could restrict all values possible for confidence limits to this range, $[-.0106 \quad .9663]$. Therefore, the lower limit of the CI would be $-.0106$, and the upper limit probably something around .10. However, would it be fair to use the knowledge about the range of ρ_Y for the construction of the exact CI, but not for the construction of the other confidence intervals? Answering this question requires the consideration of several other issues, such as simulating nonnormal distributions that produce sample correlations that lie *outside* the range of ρ_Y. A thorough discussion of these questions is deferred to some other occasion. For the purposes of this study, we will limit ourselves to constructing the exact confidence interval for distributions for which the entire range of $[-1, \quad 1]$ for ρ_Y can be simulated.

Once the Exact CI around r of a sample drawn from a particular nonnormal distribution has been found, confidence limits of an approximate confidence interval can be directly compared to the limits of the Exact CI in several ways: Histograms of lower and upper CI limits for the Exact and the approximate CI can be plotted alongside each other. Another graphing approach presented in this study entails calculating difference scores between the lower and upper CI limits of the Exact CI and the approximate CI and plotting this bivariate set of difference scores in a scatterplot. Differences between interval limits can also be summarized numerically using a mean or median. These techniques will be demonstrated and applied in great detail in the results section below.

Widths of the confidence intervals will be recorded as well. Narrower confidence intervals provide more precise information on the parameter of interest. Should all other performance measures be approximately equal for two competing confidence interval procedures, the narrower confidence interval should be preferred. In general, however, width should be of secondary importance.

Method

Several concerns play a role in the assessment of confidence interval performance. First, of course, the selection of conditions of nonnormality is important. In this section, we will examine coverage rate and balance, width, and variability around the exact confidence interval for five approximate confidence intervals: 1) Fisher's Z confidence interval (Equation (100)), 2) the asymptotically distribution-free confidence interval (Equation (103)), 3) the asymptotically distribution-free Fisher Z confidence interval (Equation (105)), 4) the BCa bootstrap confidence interval (Equation (107)) and 5) the univariate sampling bootstrap confidence interval as suggested by Beasley et al. (2007, Equation (117)).In previous sections, we have already seen that properties of distributions that have been disregarded in previous literature can have a tremendous impact on simulation outcomes. We will be able to observe more of these effects in the results section of this part; however, they will not be discussed in detail here.

Conditions Simulated. I investigated a total of eight different nonnormal distributions. Due to its popularity and in order to limit the different factors that may influence results, the 3rd order power method was used to simulate nonnormality. In light of the discussion above on Exact CIs constructed for nIM distributions, we limit the construction of exact CIs to IM

191

distributions which are capable of simulating ρ_Y over the entire range of $[-1, \quad 1]$. There exists

no equation that allows for the determination of skewness-kurtosis combinations that produce the

entire range $\rho_Y = [-1, \quad 1]$, but trial and error manipulation showed that only symmetric IM Vale

& Maurelli distributions seem to be able to produce final correlations spanning the entire interval

$[-1, \quad 1]$. Hence, two distributions with $\gamma_{11} = \gamma_{12} = 0$ and small kurtosis, namely $\gamma_{21} = \gamma_{22} = 2.75$

and two distributions with $\gamma_{11} = \gamma_{12} = 0$ and large kurtosis, namely $\gamma_{21} = \gamma_{22} = 25$, were used to

compare confidence interval limits to the limits of the exact CI. The coefficient sets for these

distributions are as follows:

Table 50: Moments and Transformation Coefficients for IM Distributions Simulated

Distribu-tion	Moments	Transformation Coefficients
1a	$\gamma_{11} = \gamma_{12} = 0$ $\gamma_{21} = \gamma_{22} = 2.75$	$\begin{bmatrix} 0.0000 & 0.7948 & 0.0000 & 0.0643 \end{bmatrix}$ (monotonic)
1b	$\gamma_{11} = \gamma_{12} = 0$ $\gamma_{21} = \gamma_{22} = 2.75$	$\begin{bmatrix} 0.0000 & 1.5479 & 0.0000 & -0.2569 \end{bmatrix}$ (non-monotonic)
2a	$\gamma_{11} = \gamma_{12} = 0$ $\gamma_{21} = \gamma_{22} = 25$	$\begin{bmatrix} 0.0000 & 0.2553 & 0.0000 & 0.2038 \end{bmatrix}$ (monotonic)
2b	$\gamma_{11} = \gamma_{12} = 0$ $\gamma_{21} = \gamma_{22} = 25$	$\begin{bmatrix} 0.0000 & -1.5667 & 0.0000 & 0.3482 \end{bmatrix}$ (non-monotonic)

All four distributions were simulated with population correlations $\rho_Y = -.8$, $\rho_Y = -.4$, $\rho_Y = 0$,

$\rho_Y = .4$, and $\rho_Y = .8$. While results for distribution 1a and 1b are limited to samples of size

$N = 50$, simulations for distribution 2a and 2b span samples of size $N = 20$, $N = 50$, $N = 100$,

$N = 200$, and $N = 500$. A graph for each distribution and population correlation size based on

$N = 5000$ is provided in Figure 42 and Figure 43.
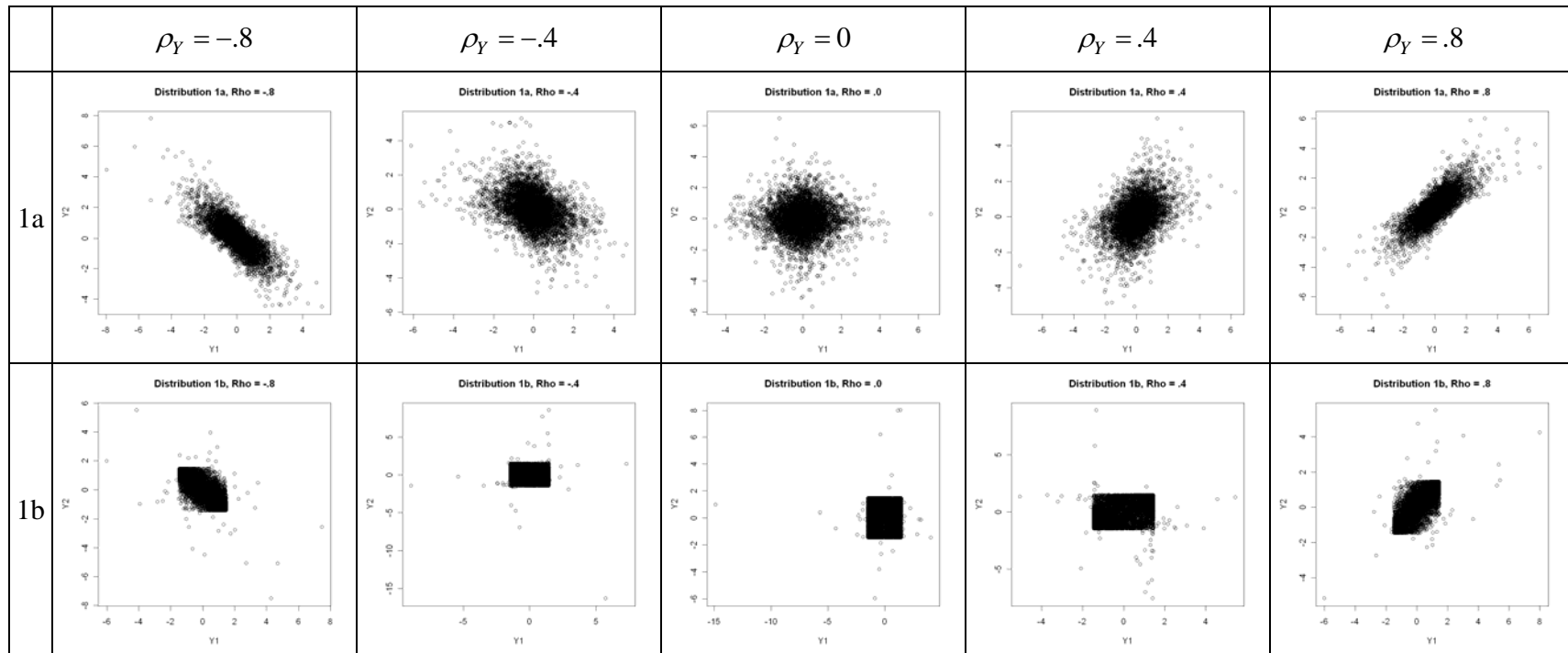
Figure 42: Plots of Distributions 1a and 1b, ρ = −.8 through .8, N = 5000

	$\rho_Y = -.8$	$\rho_Y = -.4$	$\rho_Y = 0$	$\rho_Y = .4$	$\rho_Y = .8$
2a					
2b					
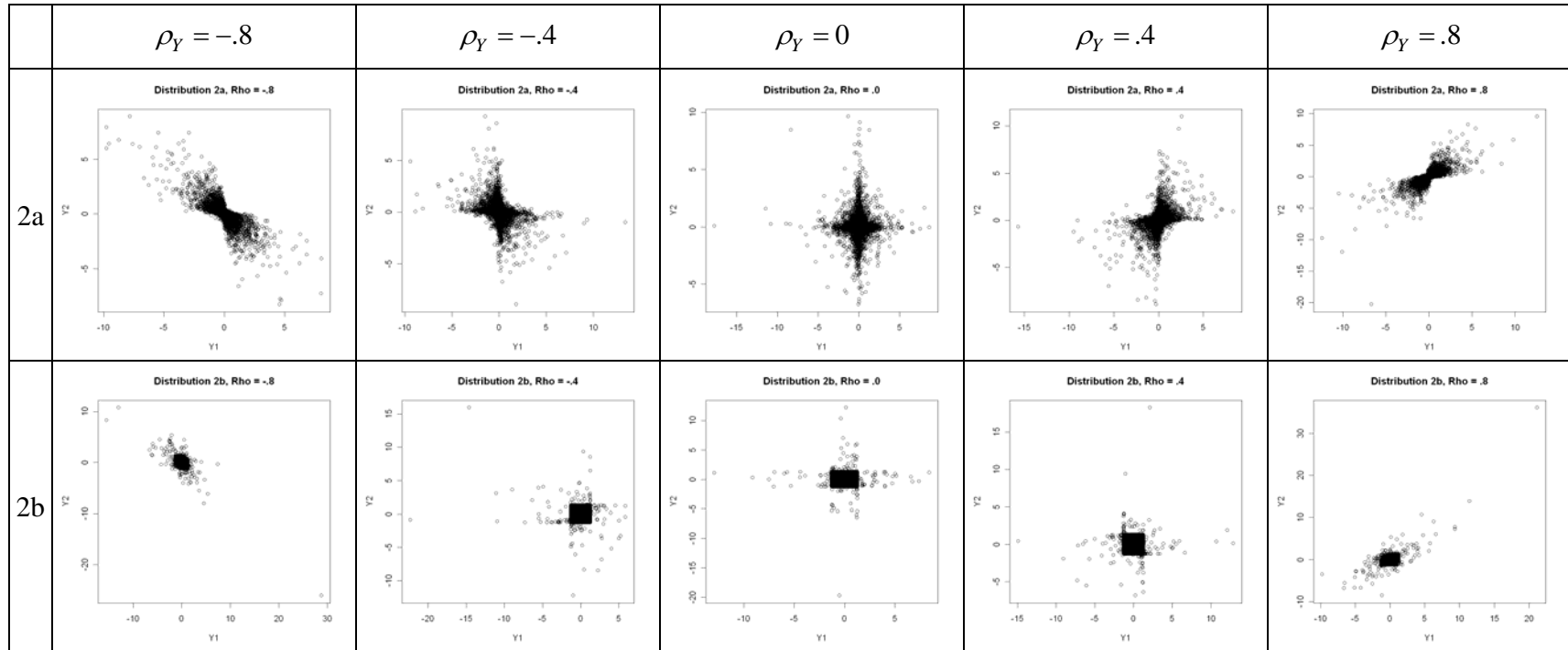
Figure 43: Plots of Distributions 2a and 2b, $\rho = -.8$ through .8, N = 5000

One more combination of marginal skewnesses and kurtoses was chosen, an nIM V&M

distribution with $\gamma_{11} = 2.5$, $\gamma_{12} = 2.0$, $\gamma_{21} = 10$, and $\gamma_{22} = 12$. Each variable (Y_1 and Y_2) has two

distinct sets of transformation coeffcients (Table 51), which combine into four differently shaped

distributions.

Table 51: Moments and Transformation Coefficients for nIM Distributions Simulated

Variable	Distribution	Moments	Transformation Coefficients
Y_1	1a	$\gamma_{11} = 2.5$ $\gamma_{21} = 10$	$\begin{bmatrix} -0.2694 & -0.6214 & 0.2694 & -0.0918 \end{bmatrix}$
	1b	$\gamma_{11} = 2.5$ $\gamma_{21} = 10$	$\begin{bmatrix} -0.6415 & 0.6487 & 0.6415 & -0.1057 \end{bmatrix}$
Y_2	2a	$\gamma_{11} = 2.0$ $\gamma_{21} = 12$	$\begin{bmatrix} -0.2034 & 0.6304 & 0.2034 & 0.0987 \end{bmatrix}$
	2b	$\gamma_{11} = 2.0$ $\gamma_{21} = 12$	$\begin{bmatrix} -0.5435 & 0.9802 & 0.5435 & -0.1553 \end{bmatrix}$

The ranges of the final correlations that can be simulated for these four distributions are less than

the maximum $\begin{bmatrix} -1, & 1 \end{bmatrix}$:

	ρ_Y	
	MIN	MAX
41a & 42a = 41a2a	−.7759	.9951
41a & 42b = 41a2b	−.1530	.6685
41b & 42a = 41b2a	−.0805	.5055
41b & 42b = 41b2b	−.0106	.9663

Population correlations and sample sizes for the nIM distributions are $\rho_Y = 0$ and $\rho_Y = .5$, and

$N = 20$ ($\rho_Y = .5$ only), $N = 50$, $N = 100$, $N = 200$, and $N = 500$, respectively. No exact CI

was constructed for this distribution, but confidence limits for all five approximate confidence

intervals were recorded, which could then be used to find coverage rate and balance and compare

CIs amongst each other. The distributions are plotted in Figure 44 below, based on $N = 5000$:

	$\rho_Y = .0$	$\rho_Y = .5$
41a2a	Distribution 41a2a, Rho = .0	Distribution 41a2a, Rho = .5
41a2b	Distribution 41a2b, Rho = .0	Distribution 41a2b, Rho = .5
41b2a	Distribution 41b2a, Rho = .0	Distribution 41b2a, Rho = .5
41b2b	Distribution 41b2b, Rho = .0	Distribution 41b2b, Rho = .5
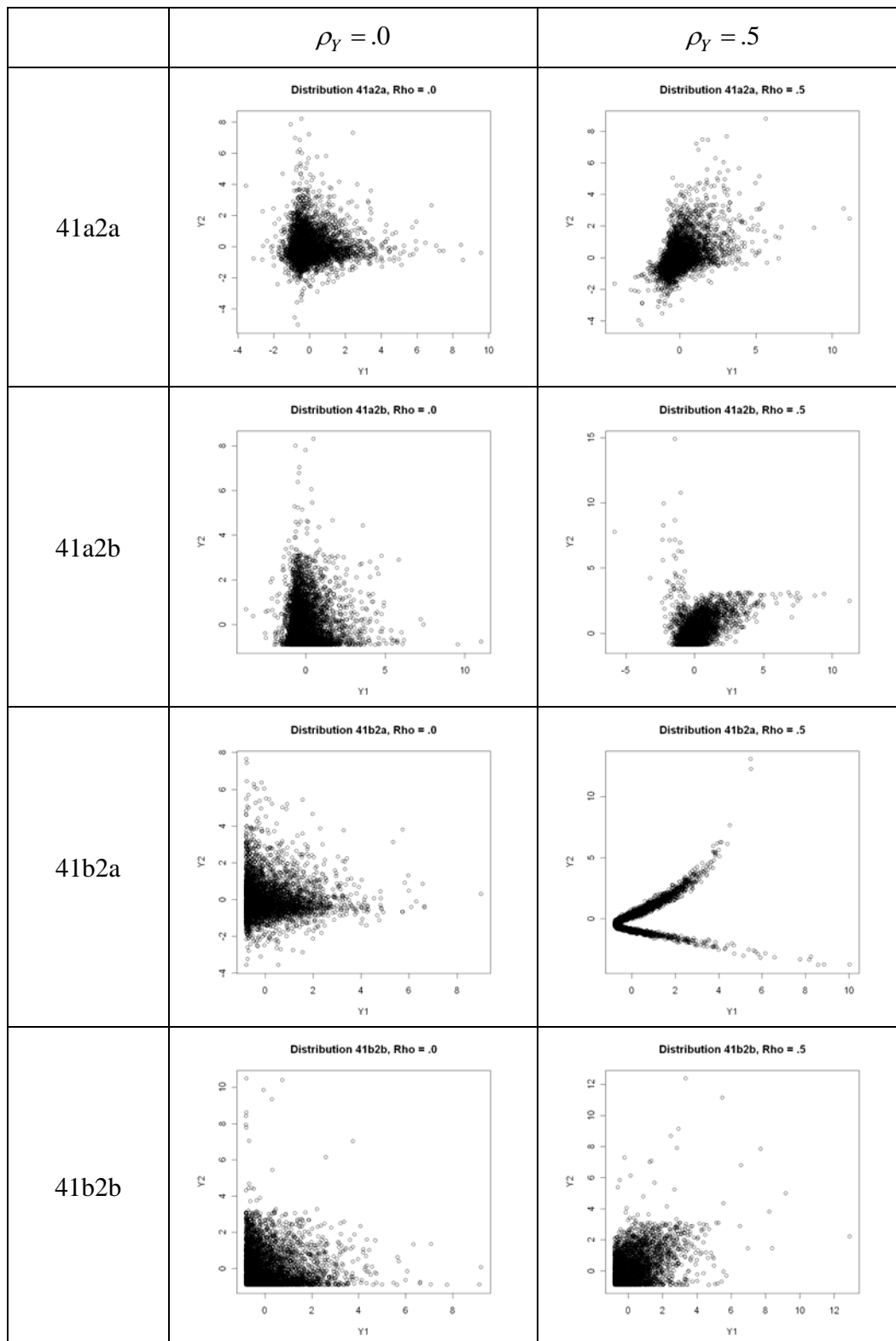
Figure 44: Plots of Distribution 41a2a through 41b2b $\rho = 0$ and .5, N = 5000

Table 52 below summarizes the conditions (sample size, size of ρ_Y, and confidence level) simulated:

Table 52: AllDistributions - Part II

Distribution	Correlation	Sample size	Type I Error Level
1a	$\rho = -.8$	$N = 50$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = -.4$	$N = 50$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = 0$	$N = 50$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = .4$	$N = 50$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = .8$	$N = 50$	$\alpha = .32, .2, .1, .05, .01$
1b	$\rho = -.8$	$N = 50$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = -.4$	$N = 50$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = 0$	$N = 50$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = .4$	$N = 50$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = .8$	$N = 50$	$\alpha = .32, .2, .1, .05, .01$
2a	$\rho = -.8$	$N = 20, 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = -.4$	$N = 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = 0$	$N = 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = .4$	$N = 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = .8$	$N = 20, 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
2b	$\rho = -.8$	$N = 20, 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = -.4$	$N = 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = 0$	$N = 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = .4$	$N = 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho = .8$	$N = 20, 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
41a2a	$\rho_Y = .0$	$N = 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho_Y = .5$	$N = 20, 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
41a2b	$\rho_Y = .0$	$N = 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho_Y = .5$	$N = 20, 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
41b2a	$\rho_Y = .0$	$N = 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho_Y = .5$	$N = 20, 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
41b2b	$\rho_Y = .0$	$N = 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$
	$\rho_Y = .5$	$N = 20, 50, 100, 200, 500$	$\alpha = .32, .2, .1, .05, .01$

Results

Performance of the Exact Confidence Interval

During the construction and execution of the exact CI, both a lowess procedure and some approximation are required and therefore a more extensive performance check for the Exact CI was done. I randomly chose two combinations of sample size and population correlation for each of the four distributions 1a, 1b, 2a, and 2b and generated coverage rate and balance based on 1,000,000 replications. The eight randomly selected distributions were:

- Distribution 1a: $N = 50$ with $\rho = -.8$ and $\rho = -.4$

- Distribution 1b: $N = 50$ with $\rho = .8$ and $\rho = 0$

- Distribution 2a: $N = 50$ with $\rho = 0$ and $N = 20$ with $\rho = -.8$

- Distribution 2b: $N = 500$ with $\rho = .8$ and $N = 200$ with $\rho = -.4$.

Due to the very large sample size ($N = 1,000,000$), a simple approximation can be used to create a 95% confidence interval for coverage rate, which is just a proportion:

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{119}$$

For coverage balance, the values were bundled into packages of 50,000, and the balance index for each package then treated like one sample value. Calculating mean and standard deviation for these 20 values, the interval was constructed as

$$m \pm t_{\alpha/2,19}s \tag{120}$$

where $t_{\alpha,19} = 1 - \alpha^{th}$ percentile of the *t* distribution with 19 degrees of freedom. A visual representation of these confidence intervals around coverage rate and balance is provided in Figure 45 below. The first cell shows 95% CIs for coverage rate of the exact CIs at the 68% confidence level and each horizontal line stands for the 95% CI for one of the distributions,

199

resulting in a total of eight CIs. Five of the CIs cover the nominal value of 68% (distributions

1a.1 with $\rho = -.8$ and $N = 50$, 1b.1 with $\rho = .8$ and $N = 50$, 2a.2 with $\rho = 0$ and $N = 50$, 2b.1

with $\rho = .8$ and $N = 500$, and 2b.2 with $\rho = -.4$ and $N = 200$), and CIs for three conditions do

not cover the nominal value. Equivalent to Figure 45(a), Figure 45(b) shows CIs for 80%

coverage rate and so on.
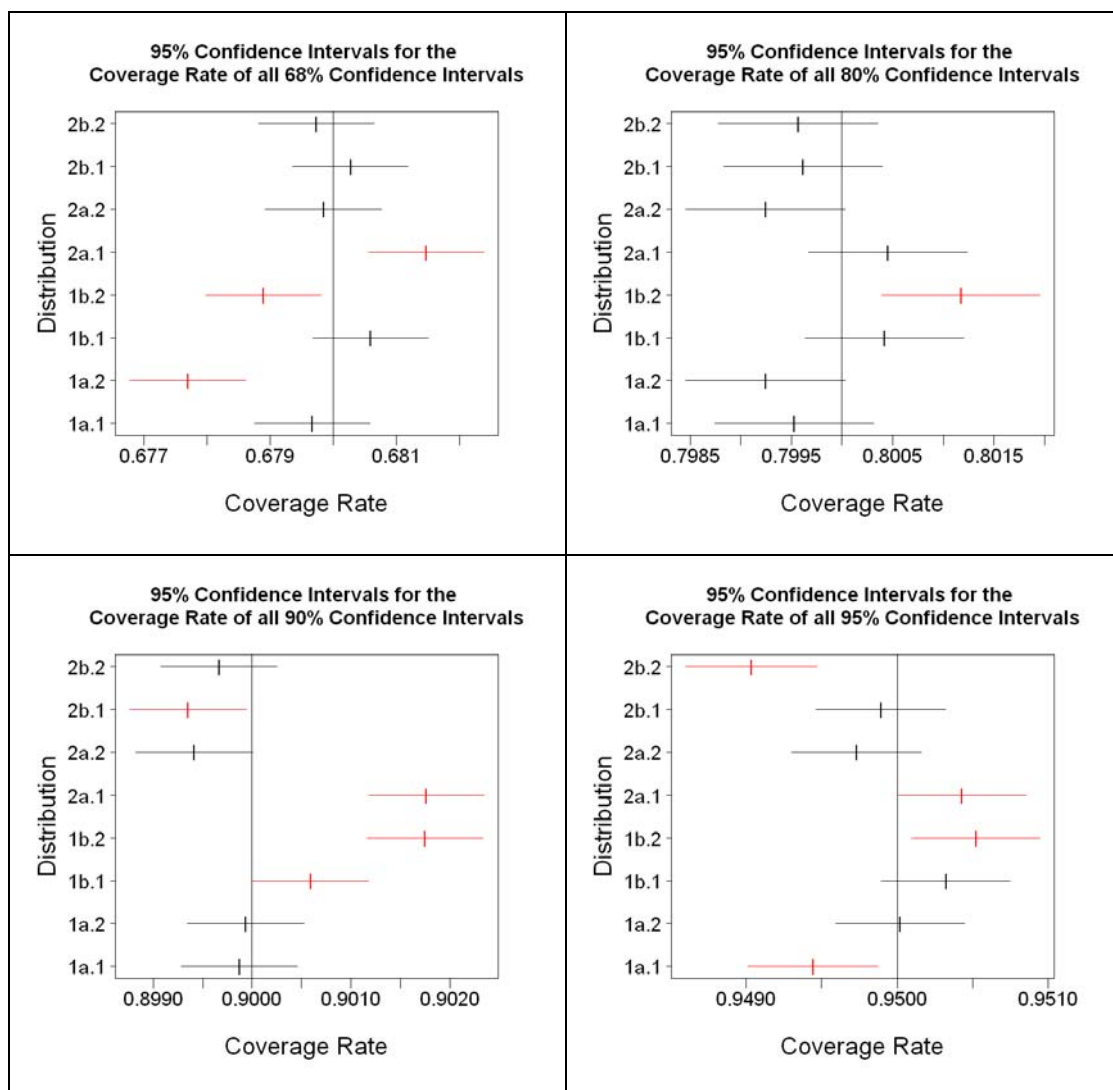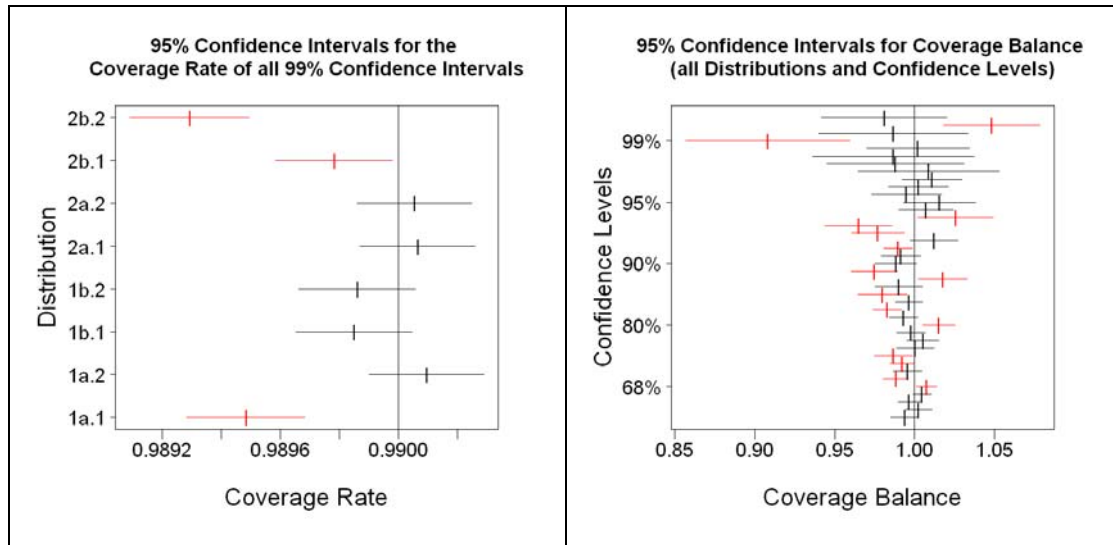
Figure 45(a) through (f): Confidence Intervals Summarizing the Performance of the Exact CI

15 out of the 40 coverage rate CIs and 15 out of the 40 balance index CIs (37.5% each) do not cover the nominal value. This rate is above the random non-coverage that would be expected for 95% confidence intervals. Most likely reasons are that smoothing and rounding applied in the construction and application of the CI do have some slight but noticeable effects. Notice that the deviations from nominal performance are very small and the confidence intervals are narrow, having high precision. In Figure 45(f), as confidence level increases, the width of the CIs increases as well, as the number of values going into the calculation of the balance index decreases, leading to more variability.

Evaluating Confidence Intervals – Distributions with Identical Marginals.

As promised, the presentation of study results will be more elaborate than what is common and focus on exploring several avenues of conceptualizing performance of confidence intervals. Not all simulated conditions will be discussed, and the greatest emphasis will be given to simulations for the IM distributions. All subsequent results are based on 50,000 replications.

201

Coverage Rate. The most common approach is to compare coverage rates (in lieu of Type I error performance). Oftentimes, a two-sided confidence interval is used as a substitute for a non-directional hypothesis test, and the information provided by coverage rate for the confidence interval is equivalent to the Type I error rate of the corresponding hypothesis test. Below are the coverage rates for all six intervals for distribution 2a, $N = 20$, $\rho = .8$:

Table 53: Coverage Rate for Distribution 2a, N = 20, ρ = .8

Exact CI	Fisher CI	Asymptotic CI	Asymptotic Fisher CI	BCa CI	Univariate Sampling CI
0.680	0.497	0.446	0.484	0.557	0.737
0.801	0.608	0.539	0.599	0.677	0.874
0.900	0.724	0.637	0.717	0.796	0.960
0.951	0.799	0.702	0.795	0.867	0.987
0.990	0.893	0.793	0.891	0.945	0.999

Empirical coverage rates for the exact CI are very close to the nominal coverage rates of $1-\alpha_1 = .68$, $1-\alpha_2 = .80$, $1-\alpha_3 = .90$, $1-\alpha_4 = .95$, and $1-\alpha_5 = .99$. For the approximate confidence intervals, one may decide that the order of performance is 2) Univariate Sampling CI, 3) BCa CI, 4) Fisher Z CI, 5) asymptotically distribution-free Fisher Z CI, and 6) asymptotically distribution-free CI. Notice that the Univariate Sampling CI has coverage rates above the nominal values, while all other approximate CIs have coverage rates below the nominal values.

Coverage Balance. More recently (e.g. in Zou, 2007, discussed in Efron, 2003, Newcombe 1998) researchers have begun to include miscoverage rates for the lower and upper end of the confidence interval as discussed in the section on the balance index in Part I of my dissertation. The balance index divides the rate at which a confidence interval lies above the true value by the rate at which a CI lies below the true value. Coverage rates in conjunction with balance index values are presented in Table 54:

Table 54: Coverage Rate and Balance for Distribution 2a, N = 20, ρ = .8

Exact CI		Fisher CI		Asymptotic CI	
0.680	1.019	0.497	2.077	0.446	1.975
0.801	1.028	0.608	2.542	0.539	2.272
0.900	1.059	0.724	3.553	0.637	2.728
0.951	1.068	0.799	5.256	0.702	3.204
0.990	1.051	0.893	15.539	0.793	4.327

Asymptotic Fisher CI		BCa CI		Univariate Sampling CI	
0.484	1.506	0.557	0.962	0.737	2.822
0.599	1.407	0.677	0.884	0.874	3.856
0.717	1.212	0.796	0.785	0.960	6.878
0.795	1.041	0.867	0.749	0.987	14.119
0.891	0.697	0.945	0.781	0.999	37.000

Again, the exact confidence interval's empirical coverage rate and balance values are very close to their nominal counterparts. For the other CIs, however, balance performance varies greatly between procedures and can be far removed from the nominal value of 1. In particular, the Univariate Sampling CI and tends to overestimate the true value of ρ; out of 38 times that the CI misses the true ρ at the 99% confidence level, it lies above ρ 37 times and only once below. While the Fisher *Z* CI tends to also overestimate ρ, the asymptotically distribution-free Fisher Z CI and the BCa CI perform rather well. With this additional information available, a researcher may now decide that the BCa CI outperforms the Univariate Sampling CI. Of course, coverage rate for the BCa CI is not great, and some researchers may be unwilling to accept such deviations, but the considerable coverage imbalance displayed by the Univariate Sampling CI should not be ignored. Notice that the asymptotically distribution-free Fisher Z CI has the second best coverage balance performance.

Width. Additional insight about the confidence intervals might be gained if we also include information on some average distance between the lower and upper confidence limits.

The table below includes the median difference between lower and upper confidence limits as a third column:

Table 55: Coverage Rate, Coverage Balance, and Width for Distribution 2a, N = 20, ρ = .8

Exact CI			Fisher CI			Asymptotic CI		
0.680	1.019	0.260	0.497	2.077	0.155	0.446	1.975	0.137
0.801	1.028	0.342	0.608	2.542	0.202	0.539	2.272	0.176
0.900	1.059	0.452	0.724	3.553	0.265	0.637	2.727	0.222
0.951	1.068	0.554	0.799	5.256	0.322	0.702	3.203	0.260
0.990	1.051	0.743	0.893	15.539	0.442	0.793	4.327	0.325
Asymptotic Fisher CI			BCa CI			Univariate Sampling CI		
0.484	1.506	0.145	0.557	0.962	0.179	0.737	2.822	0.247
0.599	1.407	0.190	0.677	0.884	0.238	0.874	3.856	0.336
0.717	1.212	0.250	0.796	0.785	0.312	0.960	6.878	0.459
0.795	1.041	0.307	0.867	0.749	0.374	0.987	14.119	0.576
0.891	0.697	0.428	0.945	0.781	0.477	0.999	37.000	0.806

Overall, the Univariate Sampling CI tends to be the widest, while the asymptotically distribution-free CI tends to be the most narrow. The Univariate Sampling CI is rather close in width to the exact CI, but for the 68% confidence level, the Univariate Sampling CI is narrower than the Exact CI, but has *greater* coverage. While this may seem appealing at first, the interval also tends to overestimate the true ρ quite often; more specifically, 19.4% of the time, while it underestimates the true parameter only 6.9% of the time. Nevertheless, it estimates the width of the exact CI best.

Plotting. With three indices of performance, ranking the confidence intervals becomes increasingly difficult, as there are interactions between confidence interval type and type of performance index. While the triad of coverage rate, coverage balance, and width is more than is presented in the vast majority of studies on confidence interval performance, these are only *point-value summaries* of an entire bivariate distribution of confidence interval limits. We now create a number of different plots to aid the evaluation of CI performance. For distribution 2a,

$N = 20$, $\rho = .8$, Figure 46(a) shows frequency plots for both the lower and upper confidence

limits of the Fisher Z CI, at a confidence level of $1 - \alpha_5 = .99$ in red. The two distributions of the

lower and upper confidence interval limit of the Exact CI in black and a green vertical line for

the true value of ρ are included for comparison. While the upper limits of both intervals lie

nicely on top of each other, suggesting near equivalence, the lower interval limits are radically

different from each other. One can also see that the lower limit of the Fisher Z CI lies above the

true ρ more often than the upper limit lies below. Figure 46(b) plots the lower and upper

confidence limits for the Univariate Sampling CI, at a confidence level of $1 - \alpha_5 = .99$ in red,

with the exact CI limits in black as before.
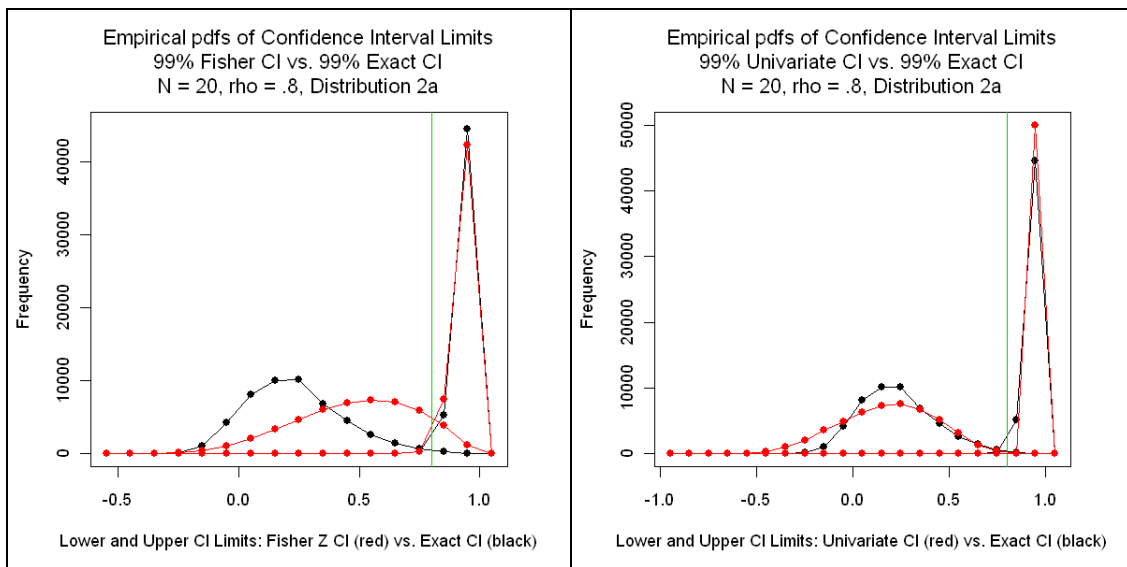
Figure 46(a) and (b): Histograms of Upper and Lower Confidence Limits for the Fisher Z, the Univariate Sampling, and the Exact CI.

The Univariate Sampling CI seems to perform better than the Fisher Z CI. The upper confidence

limit distributions lie fairly well on top of each other and the distributions for the lower

confidence limits overlap more than they do for the Fisher Z CI as well.

The graphs in Figure 46 are quite interesting, but there are additional ways in which we can graphically portray the behavior of confidence limits. The next type of graph plots values for both the lower and upper limit on the y-axis against the observed sample correlations on the x-axis, and Figure 47 demonstrates this type of plot for the Exact CI:
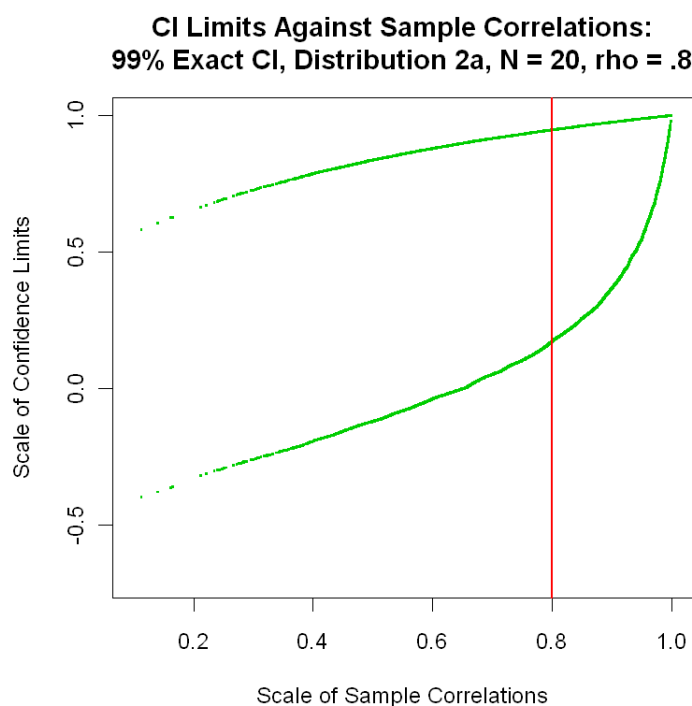
Figure 47: CI Limit-Sample Correlation Plot for the Fisher Z CI.

The red line demonstrates that for a sample correlation of $r = .8$, we would observe lower and upper CI limits of approximately .2 and .95. This graph is, in some ways, similar to Figure 40, however, now it is the sample correlation that is on the x-axis. We can use this type of graph to compare different confidence interval procedures side by side. Figure 48 shows the Exact CI and the Fisher Z CI in one graph:

206

99% Exact CI vs. 99% Fisher CI:
Distribution 2a, N = 20, rho = .8
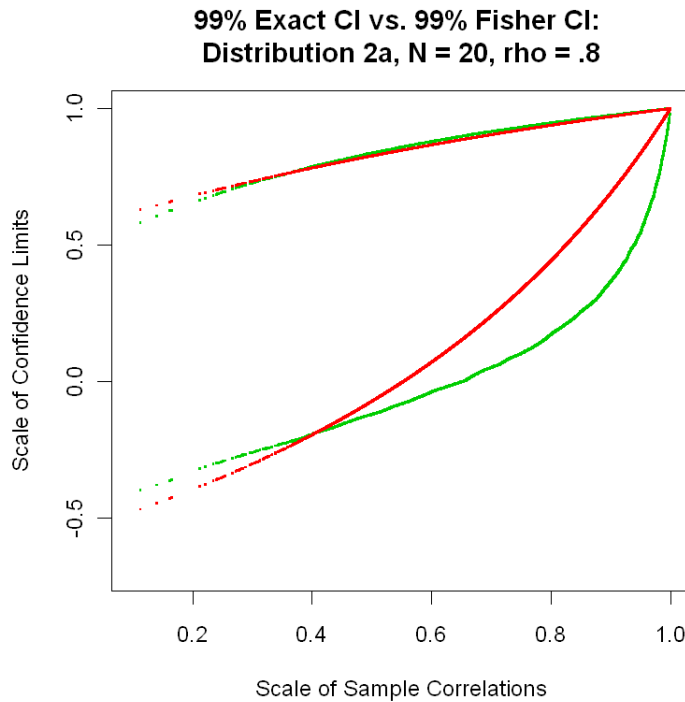
Figure 48: CI Limit-Sample Correlation Plot for the Fisher Z and the Exact CI.

This graph allows us to not only assess over and underestimation, but also where on the scale of

r these occur. It seems that the Fisher *Z* CI has its worst departure from the Exact CI right around

the value of the true ρ, i.e., 0.8. The upper limits lie almost perfectly on top of each other, while

the lower limit of the Fisher *Z* CI overestimates the lower limit of the Exact CI by a substantial

amount. Lastly, we can include one more confidence interval in the graph. Figure 49 plots the

Exact, the Fisher *Z*, and the Univariate Sampling CI side-by-side:

99% Exact, Fisher, and Univariate Sampling CI:
Distribution 2a, N = 20, rho = .8
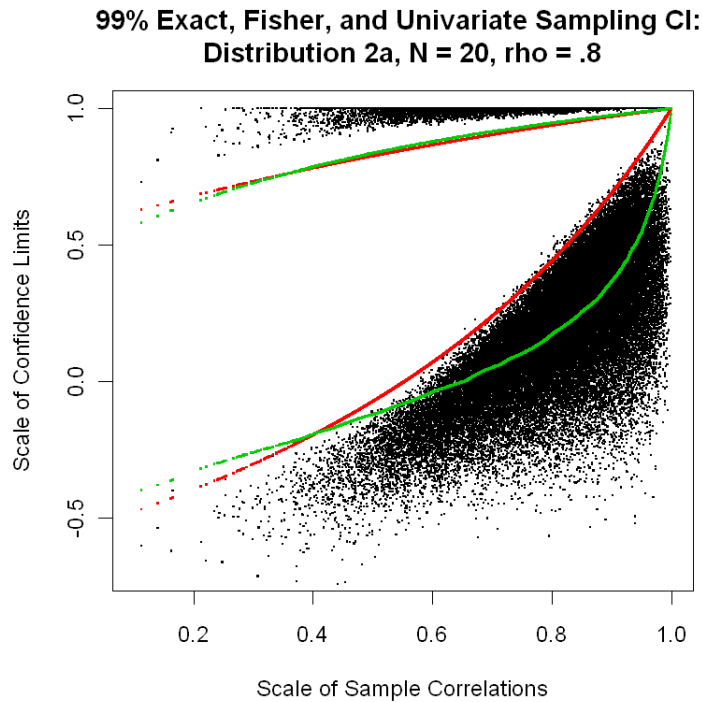
Figure 49: CI Limit-Sample Correlation Plot for the Fisher Z, the Univariate Sampling, and the Exact CI.

Several things can be read from Figure 49: The most striking observation will probably be that for any observed sample correlation, there will be exactly one value for the lower and upper limit of the Exact and the Fisher Z CI. The lower and upper limits of the Univariate Sampling CI, are subjected to *sampling variability*, that is, for a given sample r, the limits of the CI are not determined, but depend on other characteristics of the sample as well. The upper limit of the Univariate Sampling CI overestimates the upper limit of the Exact CI, and interestingly, the lower limit of the Fisher Z almost looks like an upper boundary to the lower limit of the Univariate Sampling CI.

As one last option for graphing confidence interval limits, I focus on difference scores between confidence interval limits for two different procedures. In particular, I am interested in the differences between confidence interval limits for the Exact vs. some approximate CI.

208

Assume, e.g., we had observed four sample correlations for some nonnormal distribution and created both the Fisher Z CIs and the Exact CIs around these correlations. The (fictional) observed values for the sample correlations, the Fisher Z CIs and the exact CIs are:

	Sample Correlations	Fisher Z CIs	Exact CIs
Sample 1	$r_1 = .67$	$CI_1 = \begin{bmatrix} .48 & .80 \end{bmatrix}$	$CI_1 = \begin{bmatrix} .38 & .72 \end{bmatrix}$
Sample 2	$r_2 = .84$	$CI_2 = \begin{bmatrix} .73 & .91 \end{bmatrix}$	$CI_2 = \begin{bmatrix} .75 & .86 \end{bmatrix}$
Sample 3	$r_3 = .81$	$CI_3 = \begin{bmatrix} .69 & .89 \end{bmatrix}$	$CI_3 = \begin{bmatrix} .73 & .96 \end{bmatrix}$
Sample 4	$r_4 = .53$	$CI_4 = \begin{bmatrix} .30 & .70 \end{bmatrix}$	$CI_4 = \begin{bmatrix} .25 & .74 \end{bmatrix}$

Notice how for the first sample, both the lower and the upper limit of the Fisher Z CI lie above the lower and the upper limit, respectively, of the exact CI. Similarly, for the third sample, both the lower and upper limit of the Fisher Z CI lie *below* the lower and upper limit of the exact CI. We would say that for the first sample, the Fisher Z CI is *overestimating* the exact CI and for the third sample, the Fisher Z CI is *underestimating* the exact CI. For the second sample, the Fisher Z CI fully contains the exact CI, making the Fisher Z CI *too wide*, and for the fourth sample, the exact CI fully contains the Fisher Z CI, which is therefore *too narrow*.

These relationships can be portrayed graphically. For each sample, calculate the differences between lower and upper confidence interval limits of the two interval procedures. For sample 1, the difference $LL_F - LL_E$ is $.48 - .38 = .10$, and the difference $UL_F - UL_E$ is $.80 - .72 = .08$. These two difference scores can be plotted as one point in a bivariate plane. Choosing the x-axis as the scale for differences between upper limits of confidence intervals and the y-axis as the scale for differences between lower limits of confidence intervals, the point portraying the difference scores for sample 1 is plotted in the upper right corner of Figure 50:

Four Categories of Behavior:
Overestimating, Too Wide, Underestimating, Too Narrow
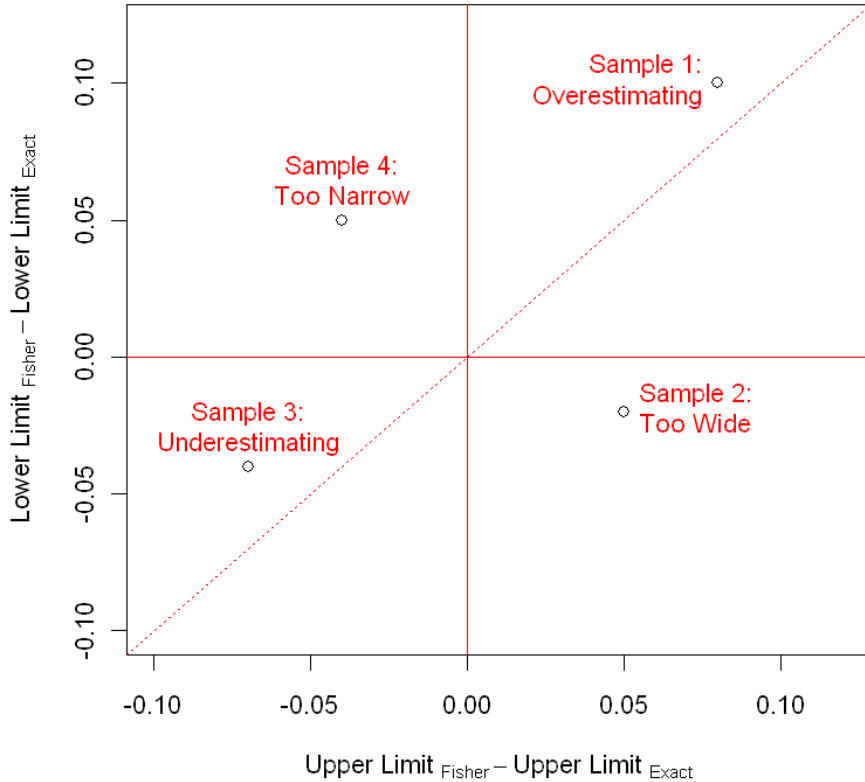
Figure 50: Four Categories of Bias.

Any point plotted in the first quadrant of Figure 50, with a positive difference between lower CI limits *and* a positive difference between upper CI limits, is a confidence interval that overestimates the exact CI. Moving on to the second sample, we have $LL_F - LL_E = .73 - .75 = -.02$ and $UL_F - UL_E = .91 - .86 = .05$; the interval is too wide and the coordinates of the differences are plotted in the second quadrant of Figure 50. For the third sample, $LL_F - LL_E = .69 - .73 = -.05$ and $UL_F - UL_E = .89 - .96 = -.07$, the Fisher Z CI is underestimating, and for the fourth sample, $LL_F - LL_E = .30 - .25 = .05$ and $UL_F - UL_E = .70 - .74 = -.04$, the Fisher Z CI is too narrow. Points for the third and fourth

sample are plotted in the third and fourth quadrant of Figure 50, respectively.

Assume we now had 100 sample correlations with their respective Fisher Z and Exact CIs and created the lower and upper CI limit difference scores as we have done for samples 1 through 4. Plotting these bivariate difference scores just as we did in Figure 50, the resulting scatter plot shows the general behavior of the Fisher Z CI relative to the Exact CI for the 100 samples:
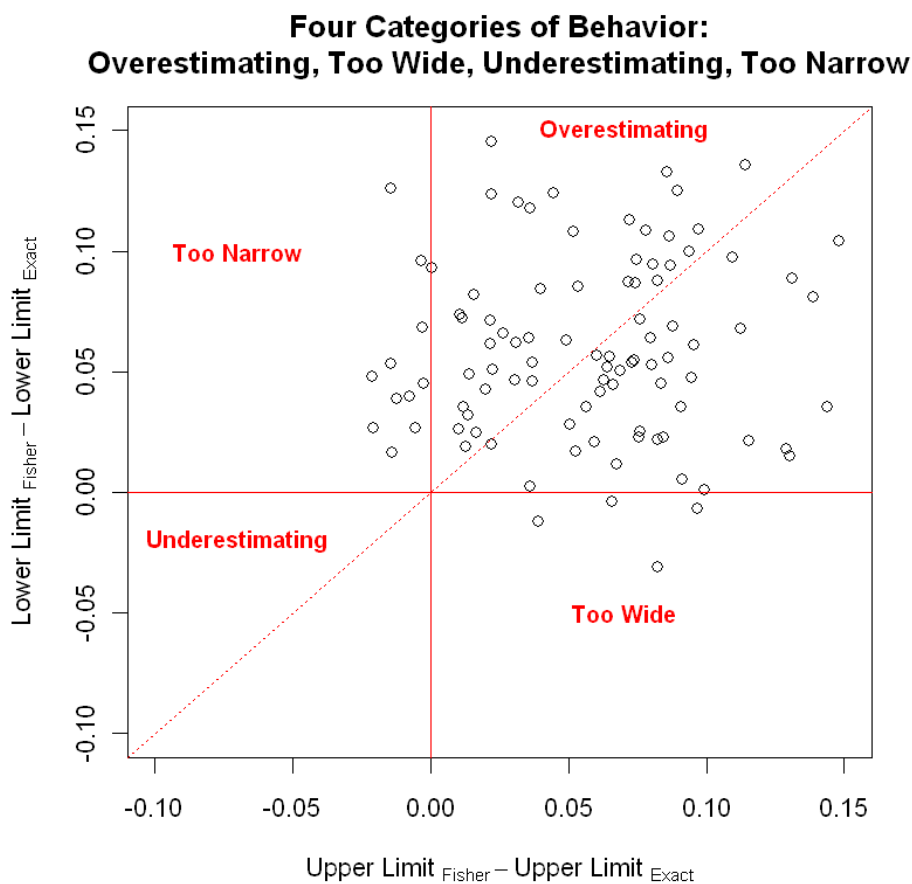
Figure 51: Example of a CI That Tends to Overestimate

Most of the difference score pairs lie in the upper right quadrant, and therefore it is the tendency of the Fisher Z CI to overestimate, that is, generally lie higher than, the exact CI.

We can now plot the CI limit difference scores for the two intervals we have compared so

far, the Fisher Z CI and the Univariate CI for distribution 2a, $N = 20$, $\rho = .8$, and a confidence

level of 99%. Most instances of the Fisher Z CI are too narrow compared to the Exact CI, mainly

due to a lower limit that is higher than the lower limit of the Exact CI (Figure 52):
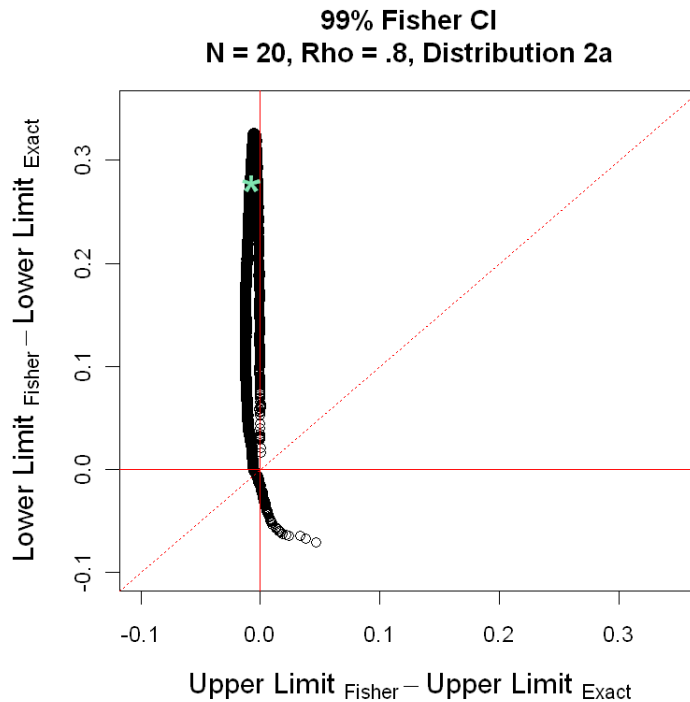
Figure 52: CI Limit Difference Score Plot for the Fisher Z CI

The center of the CI limit difference score distribution, defined as the median of the difference

scores for lower and upper CI limits and graphed as a bluish green star, shows little bias in

estimating the upper limit of the Exact CI, as it lies close to the y-axis. It does indicate

considerable bias in estimating the lower limit of the Exact CI, however, as it lies far above the

x-axis. The univariate sampling CI on the other hand is entirely concentrated on the right hand of

the graph, with practically all values in the first and second quadrant, either overestimating the

exact CI or being too wide (Figure 53).
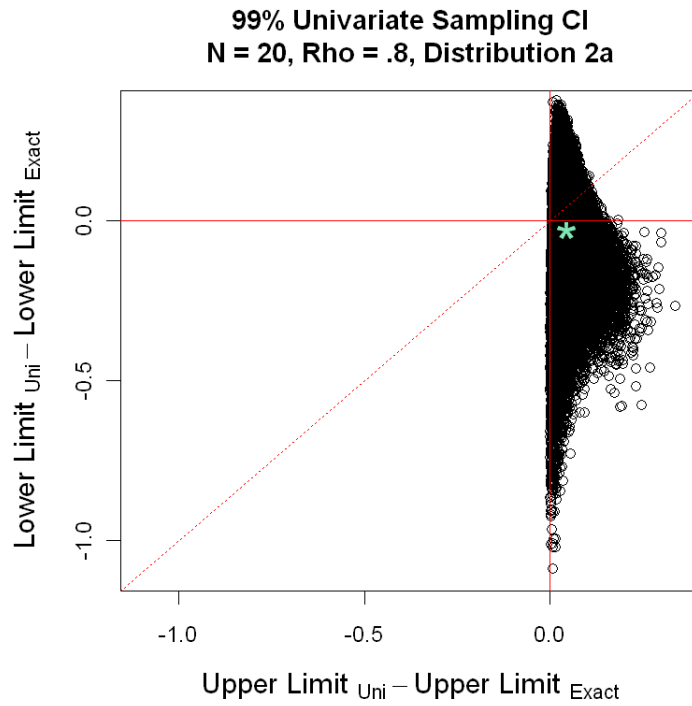
Figure 53: CI Limit Difference Score Plot for the Univariate Sampling CI

If we plot difference scores for both the Fisher *Z* and the Univariate Sampling confidence interval side by side in the same graph, we obtain the following picture:

Comparison 99% Fisher vs. Univariate Sampling CI
N = 20, Rho = .8, Distribution 2a
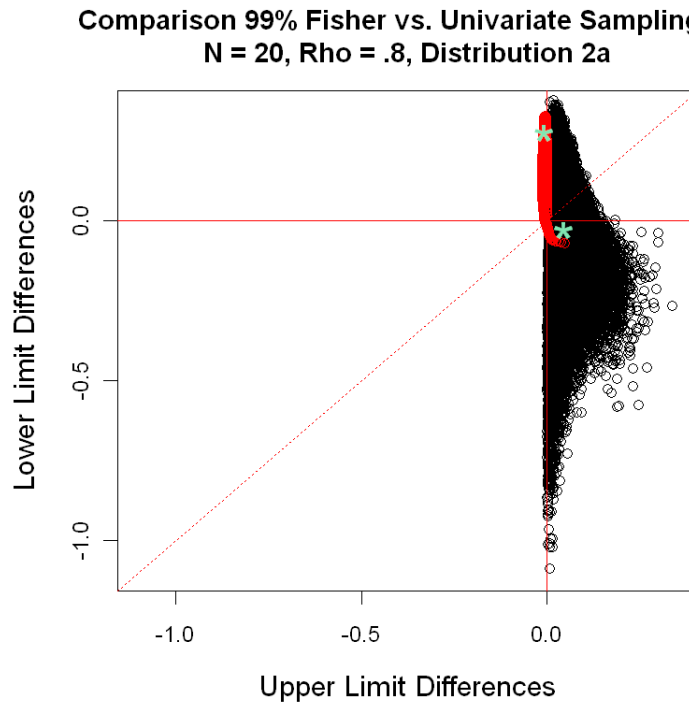
Figure 54: CI Limit Difference Score Plot for the Fisher Z and the Univariate Sampling CI

The black dots represent differences between the Univariate Sampling CI and the Exact CI, while the red dots represent differences between the Fisher Z CI and the Exact CI. Based on the centers of the two distributions, we can see that the Fisher Z CI shows more bias in the estimation of the lower limit of the Exact CI, while the Univariate Sampling CI shows more bias in estimating the upper limit. Deviation from the Exact CI seems to also vary much more for the Univariate Sampling CI than the Fisher Z CI, especially for the upper, but also for the lower confidence interval limit. These observations are in contrast with the univariate histograms of confidence interval limits in Figure 46(a) and (b) and also with the conclusions one might have drawn if only the coverage rates for both confidence intervals (89.2% for the Fisher Z CI and 99.9% for the Univariate Sampling CI) had been considered.

When using CI limit difference score plots, we have several choices for scaling the plot:

1) Choose a scale for just the individual plot at hand; 2) When graphing several conditions next to each other, choose the same scale for all plots; 3) Use the same range of scale for both x- and y-axis. Option 3 was used in Figure 52 and Figure 53, and helps identify which CI limit is affected by bias in the estimation of the Exact CI the most. Option 1 could be particularly useful when detailed insight in the functioning of a particular CI under specific conditions is desired or for the development of definitions of confidence interval bias and the like.

Number Summaries. As an aid to the interpretation of these graphs, I supplement some of the graphs with median and standard deviation of the confidence limit difference scores. As an example, for the 99% Fisher Z CI, distribution 2a, $N = 20$, and $\rho = .8$, median and SD for CI limit difference scores are

$MED\left(LL_F - LL_E\right) \approx 0.278$	$MED\left(UL_F - UL_E\right) \approx -0.008$
$SD\left(LL_F - LL_E\right) \approx 0.073$	$SD\left(UL_F - UL_E\right) \approx 0.003$

The formidable overestimation of the lower limit as well as the greater variability of the lower limit around the Exact CI limits is clearly recognizable. The distribution of the lower Fisher Z CI limit around the Exact CI is heavily skewed. The median difference between the lower limits is positive and the median difference between the upper limits is slightly negative, indicating that the Fisher Z CI is too narrow. For comparison, the same values for the Univariate Sampling CI are

$MED\left(LL_U - LL_E\right) \approx -0.029$	$MED\left(UL_U - UL_E\right) \approx 0.045$
$SD\left(LL_U - LL_E\right) \approx 0.178$	$SD\left(UL_U - UL_E\right) \approx 0.035$

The negative median for lower CI limit differences in combination with the positive median value for upper CI limit differences indicates that the Univariate Sampling CI is wider than the Exact CI in over 50% of the cases.

Additional Remarks. The graphs and numerical summaries introduced here, focusing on the comparison with an interval that, for several reasons, can be used as a form of gold standard, open up opportunities for investigating many different aspects of confidence interval performance. The additional insight that can be gained shall be demonstrated in the following sections. Rather than discussing every single distribution – sample size – correlation size and confidence level combination that has been simulated, the purpose of this results section is to present new and hopefully more informative ways of understanding confidence interval performance. The next section will utilize the newly introduced CI limit difference score plots to explore the performance of the five approximate confidence interval procedures. There are several conditions that have been varied: Type of confidence interval procedure, distribution (and therefore, population values of skewness and kurtosis as well as sample values and higher moments), size of correlation, confidence level, and sample size.

Varying The Type Of Confidence Interval and Nonnormal Distribution. The next four figures provide comparative histograms as well as CI limit difference scores plots for all five approximate confidence intervals and distributions 1a, 1b, 2a, and 2b, at $N = 50$, $\rho = .8$, at a confidence level of 99%. For the CI limit difference score plots, I combine option 2 and 3 for maximum comparability across conditions.
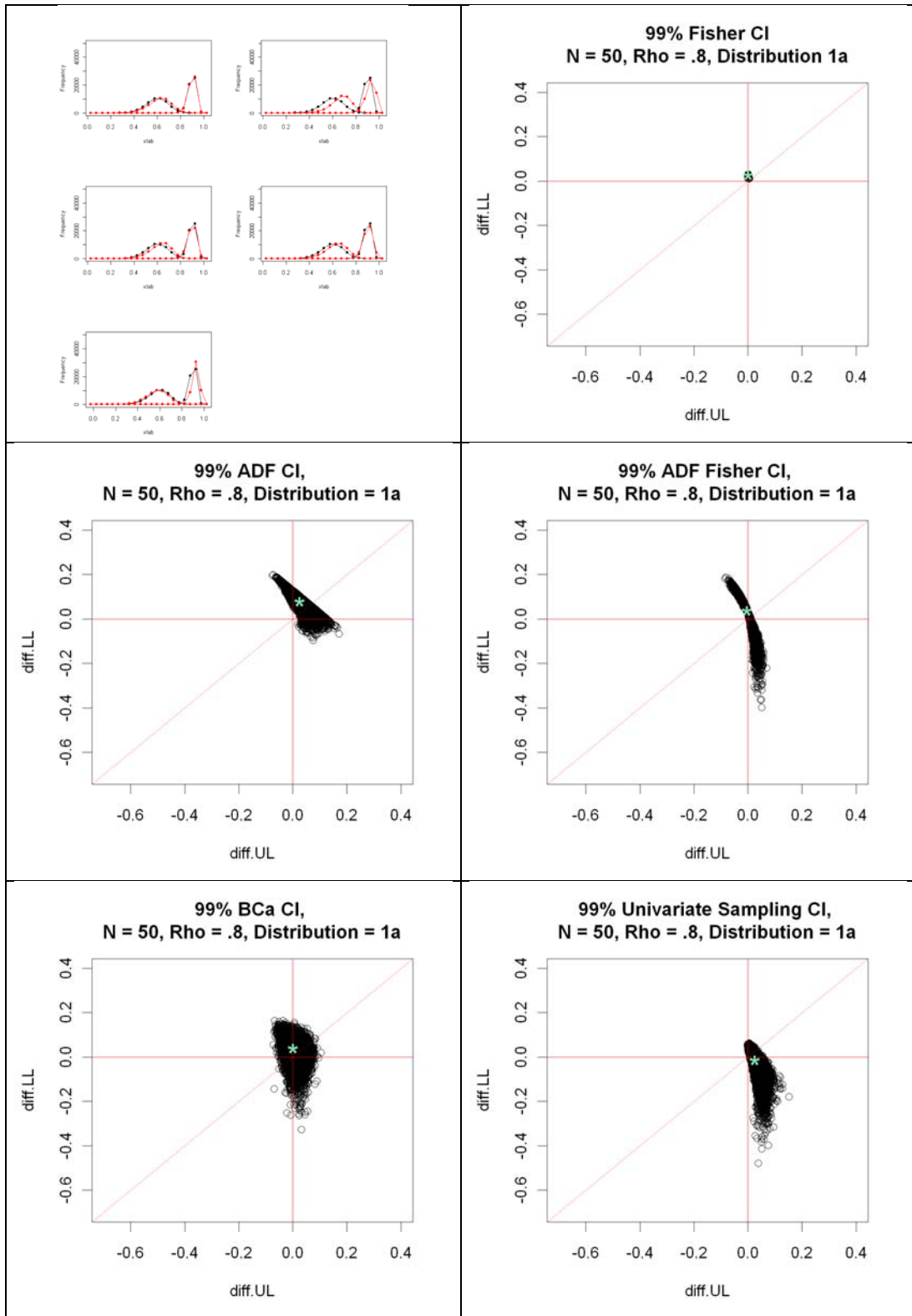
Figure 55: CI Limit Difference Score Plots for Distribution 1a, ρ = .8, 99% Confidence Level
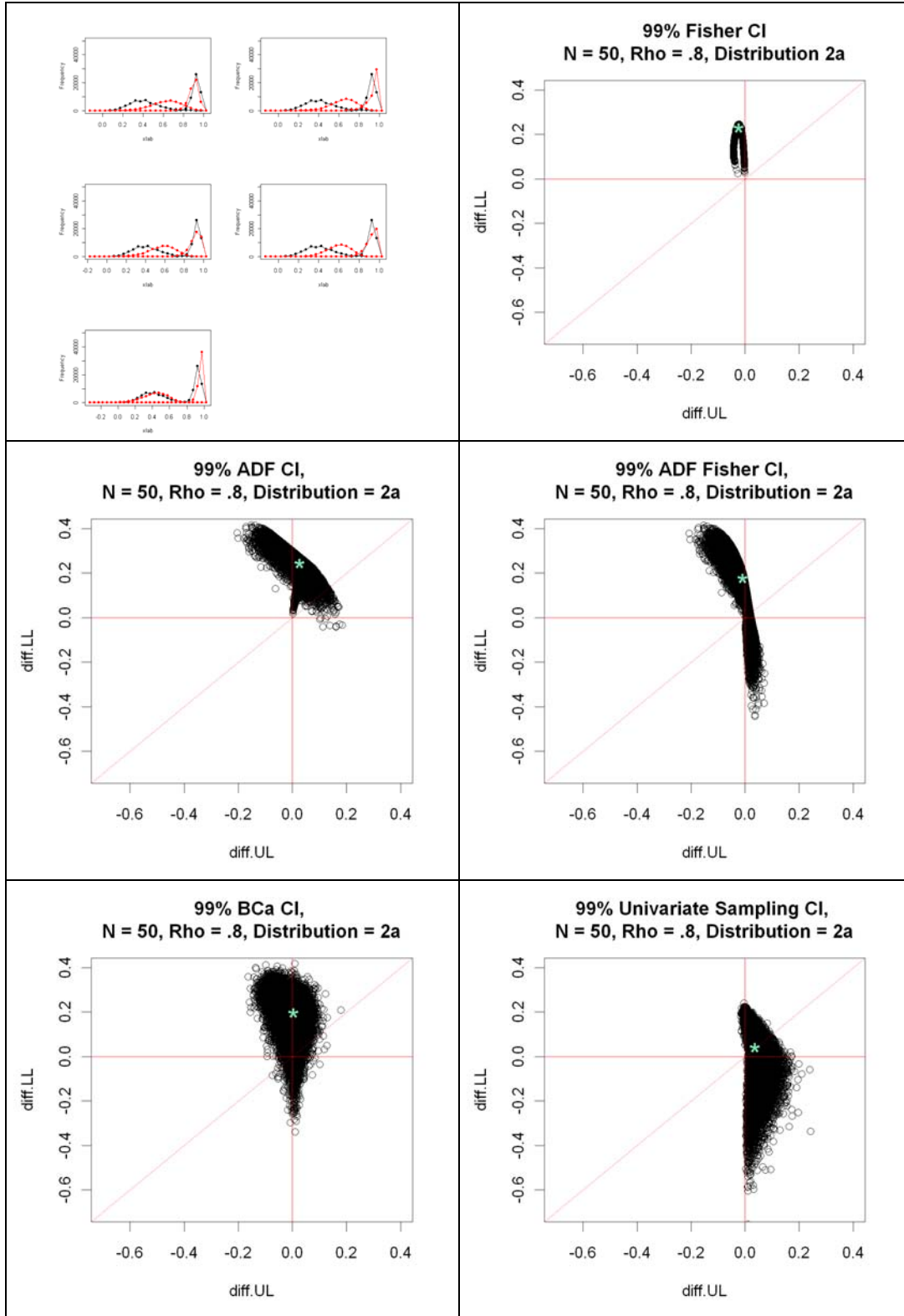
Figure 56: CI Limit Difference Score Plots for Distribution 2a, ρ = .8, 99% Confidence Level
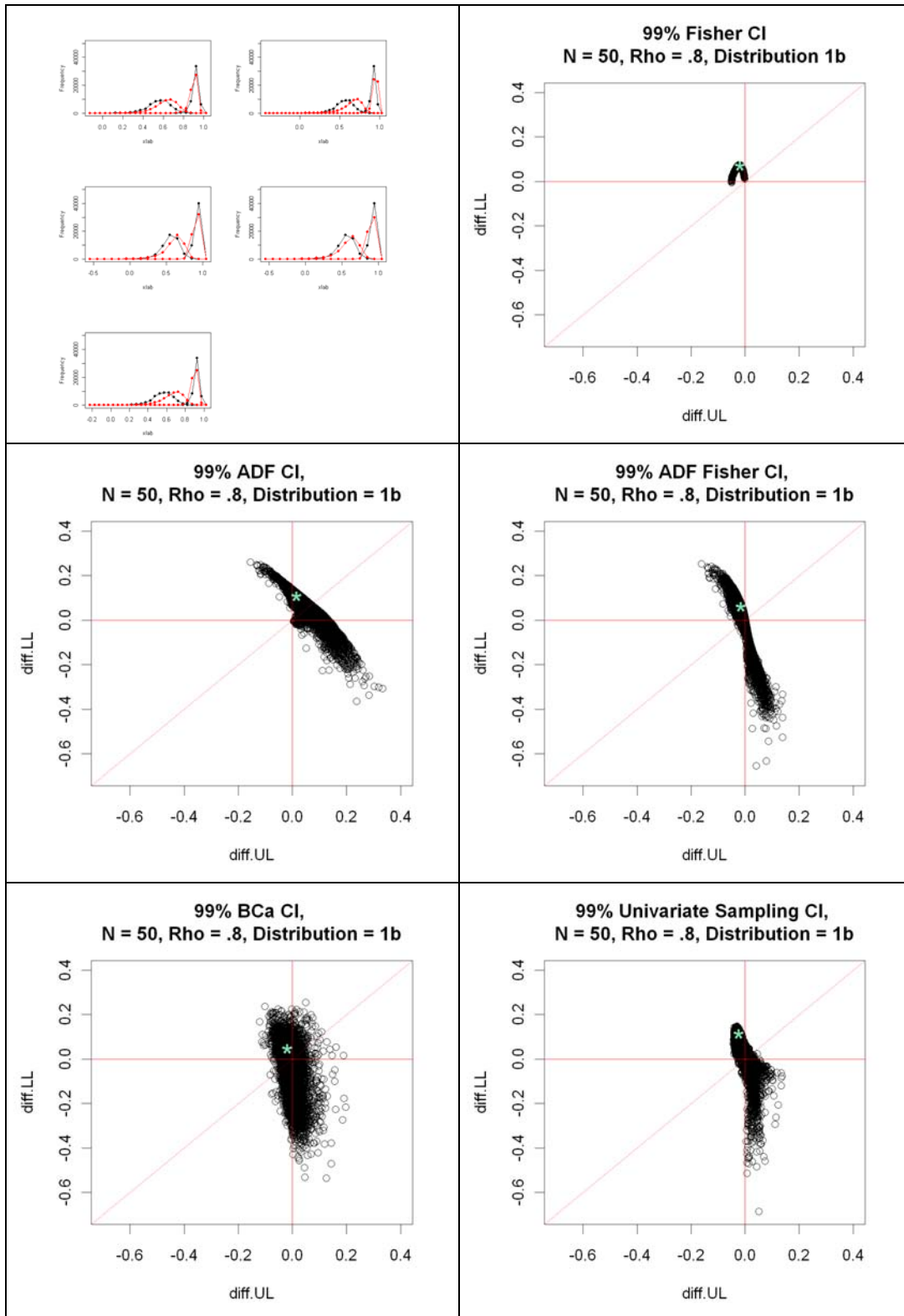
Figure 57: CI Limit Difference Score Plots for Distribution 1b, ρ = .8, 99% Confidence Level
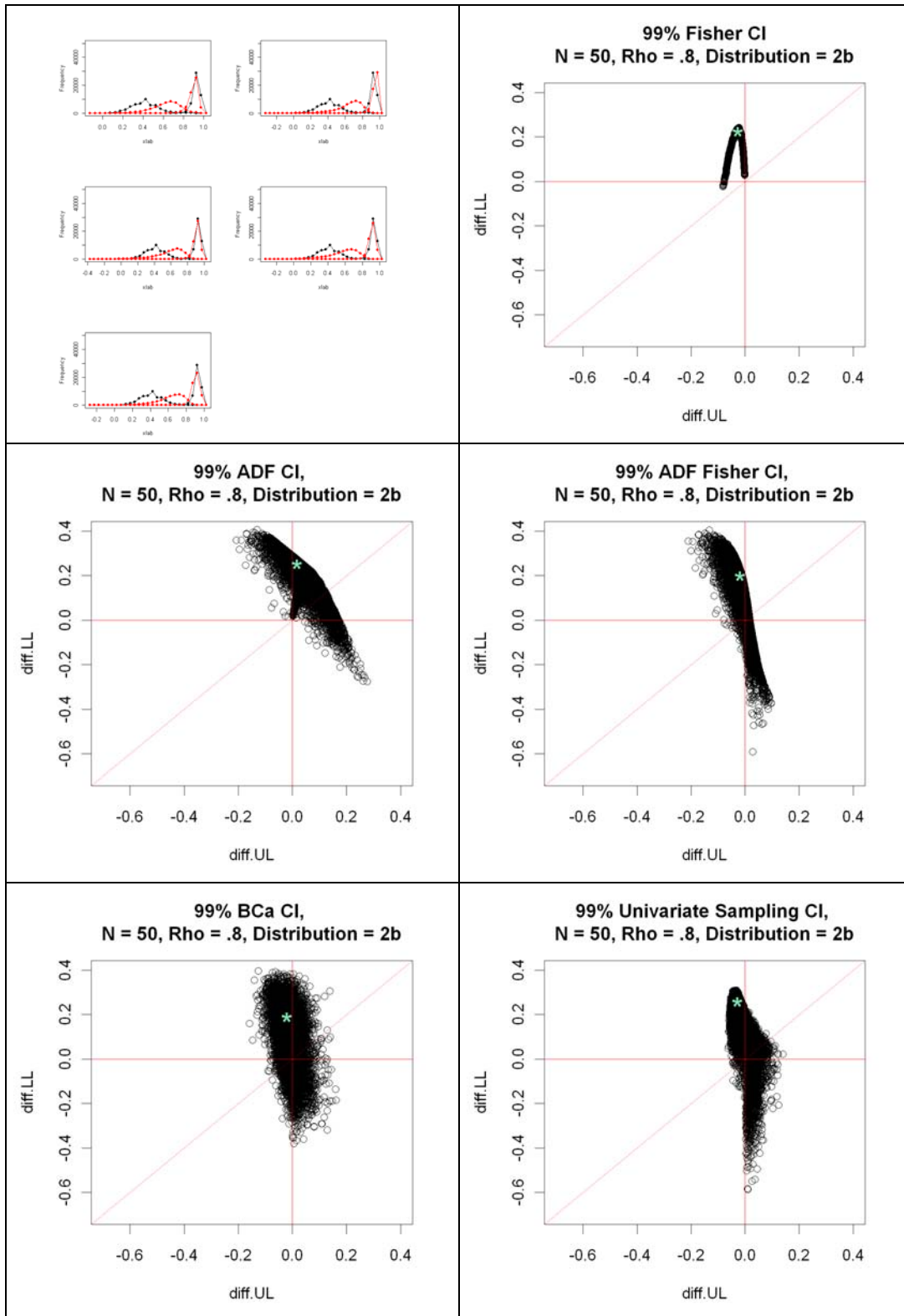
Figure 58: CI Limit Difference Score Plots for Distribution 2b, ρ = .8, 99% Confidence Level

Several things can be read from Figure 55 through Figure 58: Most plainly visible is perhaps that the Fisher *Z* confidence interval shows the least amount of variability around the Exact CI, thereby displaying the most consistent behavior. The other approximate confidence intervals show considerably more variation around the limits of the Exact CI, with more variability around the lower limit, as all the difference score distributions are somewhat stretched along the y-axis. The center of the distributions as well as their entire body usually lie above the x-axis, indicating that all CIs tend to overestimate the lower limit of the Exact CI; only the Univariate Sampling CIs center lies below the x-axis for distribution 1a. In many settings, the difference score distribution stretches somewhat along the too wide-too narrow dimension, while the asymptotically distribution-free CI tends to overestimate the exact CI more systematically. Notice also that for all CIs, the least variability around the Exact CI occurs for distribution 1a, and is increased with higher levels of marginal kurtosis (distribution 2a and 2b), but also for distribution 1b, which has the same marginal population kurtosis as distribution 1a, but a non-monotonic transformation function. The univariate CI displays the largest variety of patterns, and is not as consistent as the other intervals. Table 56 includes coverage rate, balance, width, and the difference score summaries:

Table 56: Coverage Rate, Coverage Balance, and Difference Score Summaries for Figure 55 through Figure 58.

Distri-bution	Fisher		Asymptotic		Asymptotic Fisher		BCa		Univariate Sampling	
Coverage Rate and Balance										
1a	0.982	2.956	0.959	6.597	0.974	0.576	0.975	0.865	0.997	2.190
2a	0.862	4.639	0.874	3.223	0.925	0.412	0.944	0.599	0.999	3.100
1b	0.953	1.502	0.925	35.194	0.967	3.464	0.970	3.165	0.906	8.363
2b	0.897	2.306	0.886	18.324	0.948	2.313	0.947	4.244	0.900	6.538
Difference Scores: Median (SD)										
1a	0.001 (0.001) 0.028 (0.002)		0.026 (0.025) 0.080 (0.031)		-0.003 (0.015) 0.039 (0.050)		0.002 (0.018) 0.040 (0.038)		0.025 (0.013) -0.013 (0.043)	
2a	-0.024 (0.008) 0.231 (0.025)		0.028 (0.040) 0.246 (0.059)		-0.009 (0.029) 0.177 (0.111)		0.004 (0.030) 0.198 (0.080)		0.037 (0.022) 0.042 (0.104)	
1b	-0.019 (0.006) 0.072 (0.007)		0.017 (0.032) 0.109 (0.039)		-0.015 (0.016) 0.062 (0.067)		-0.017 (0.015) 0.049 (0.069)		-0.024 (0.012) 0.115 (0.046)	
2b	-0.026 (0.011) 0.226 (0.027)		0.017 (0.037) 0.254 (0.063)		-0.017 (0.021) 0.201 (0.103)		-0.020 (0.021) 0.190 (0.090)		-0.027 (0.022) 0.258 (0.101)	

Particularly noticeable from Table 56 are the above nominal coverage rate for the Univariate Sampling CI and distribution 1a and 2a, and the much lower variability of CI limit difference scores for the Fisher Z CI (the two corresponding cells are shaded in light red). Surprisingly, the Fisher Z CI performs better for the more odd-shaped distributions 1b and 2b, while the opposite is true for all other CIs. Notice that the BCa CI has a coverage balance of less than 1 for distributions 1a and 2a, but a center for the difference scores that indicates overestimation – both the median for the lower CI limit difference scores and the upper CI limit difference scores are positive (shaded in light blue).

What if $\rho = 0$. The first few analyses have focused on performance when ρ is large. However, the mostly utilized value for ρ in past research on confidence intervals for single correlations has been $\rho = 0$. This choice is possibly rooted in the tradition of nil hypothesis

testing (Steiger, 2004). This section demonstrates that even when $\rho = 0$, coverage rate is not the only interesting property of a confidence interval procedure. Defaulting again to distribution 2a for comparability, we use $\rho = 0$, $N = 50$ and $N = 500$ to compare the approximate CIs across all confidence levels. As shown in Table 57, all CIs seem to have great coverage balance at both sample sizes. With respect to coverage rate, all CIs perform satisfactorily at $N = 500$, but the Fisher Z and the Univariate Sampling CI are clearly superior at $N = 50$ over the other CIs.

Table 57: Coverage Rate and Balance for Distribution 2a, $\rho = 0$

	Fisher Z CI		ADF CI		Fisher Z ADF CI		BCa CI		Univariate Sampling CI	
$N = 50$	0.728	1.013	0.566	0.995	0.577	0.997	0.627	0.986	0.711	1.015
	0.825	1.016	0.695	0.991	0.711	0.995	0.752	1.000	0.836	1.019
	0.902	1.040	0.817	1.008	0.836	1.006	0.859	0.993	0.929	1.049
	0.941	1.054	0.891	1.002	0.909	0.996	0.917	0.997	0.969	1.063
	0.979	1.098	0.963	0.951	0.974	0.923	0.973	1.029	0.996	0.713
$N = 500$	0.705	1.003	0.651	0.996	0.652	0.998	0.631	1.002	0.690	1.000
	0.817	0.998	0.777	0.998	0.779	1.001	0.748	1.002	0.811	0.991
	0.905	0.992	0.887	1.008	0.889	1.004	0.852	1.001	0.908	1.012
	0.949	1.015	0.944	0.994	0.946	0.990	0.912	0.997	0.956	1.026
	0.986	0.964	0.989	1.011	0.990	1.004	0.971	0.999	0.992	1.035

Figure 59 provides corresponding CI limit-sample r plots for the 99% confidence level along with coverage performance and the difference score summaries (within each cell, the upper CI limit is above the lower CI limit). As before, the Exact CI is plotted in green and the approximate CIs in red. Huge variability of confidence interval limits for the asymptotic, the asymptotic Fisher, and the BCa CI can be observed, for both $N = 50$ and $N = 500$. In terms of coverage rate and balance, all CIs would likely be regarded comparable at $N = 500$, but the Fisher Z and the Univariate Sampling CI are clearly superior over the other confidence intervals based on sampling variability of the CI limits *and* accuracy of estimating the Exact CI.

	Fisher CI	ADF CI	ADF Fisher CI	BCa CI	Univariate Sampling
N = 50					
N = 500					
Coverage	0.979 1.098	0.963 0.951	0.974 0.923	0.973 1.029	0.996 0.713
	0.986 0.964	0.989 1.011	0.990 1.004	0.971 0.999	0.992 1.035
Median (SD) of Diff Scores	0.0340 (0.022)	−0.069 (0.118)	−0.071 (0.107)	−0.027 (0.116)	0.091 (0.047)
	−0.0322 (0.022)	0.069 (0.118)	0.072 (0.107)	0.027 (0.117)	−0.090 (0.047)
	0.0021 (0.002)	−0.013 (0.032)	−0.014 (0.032)	−0.013 (0.042)	0.009 (0.007)
	−0.0020 (0.002)	0.013 (0.032)	0.014 (0.032)	0.013 (0.042)	−0.009 (0.007)

Figure 59: CI Limit-Sample Correlation Plot For All Five Approximate CIs, Distribution 2a, $\rho = 0$

Does Performance Improve as Sample Size isIincreased? For many statistical procedures, increasing the sample size will lead to a decrease in bias introduced because normal distribution theory has been applied to nonnormal parent distributions. A classic example is the confidence interval for a sample proportion, where the measurable outcome has a binomial distribution. Especially when the population proportion π is close to 0 or 1, the normal distribution theory CI

$$p \pm z_\alpha \sqrt{\frac{p(1-p)}{n}} \qquad (121)$$

with p as an estimate for π performs poorly with substantial bias. But as N increases, the CI from Equation (121) can profit from the Central Limit Theorem and its performance steadily improves. Does the performance of any of the approximate CIs for a single correlation also improve with increasing sample size? To answer this question, I again use distribution 2a, with $\rho = .8$, and compare the Fisher Z and the Univariate Sampling CI, increasing N from 20 to 500. Consider Figure 60 below, which shows CI limit difference score plots for both CIs (for this figure, only scaling option 2 but not option 3 was used for better visual detail). The range of difference scores for the Fisher Z CI changes from $[-0.02, \quad 0.01]$ for the upper and about $[-0.1, \quad 0.3]$ for the lower limit when $N = 20$ to $[-0.05, \quad -0.01]$ and $[0.04, \quad 0.09]$ when $N = 500$. For the Univariate CI, the same ranges are $[0.0, \quad -0.2]$ and $[-1.0, \quad 0.4]$ when $N = 20$, to $[-0.02, \quad 0.07]$ and $[-0.2, \quad 0.05]$ when $N = 500$. These seem to be fair improvements, particularly for the bias in estimating the lower CI limit. The difference score summaries for all five CIs in the second half of Table 58 seem to support this observation of decrease in estimation bias and variability not only for the Fisher Z and Univariate Sampling CI, but for the other intervals as well. But when we consider coverage rate and balance (first half of Table 58), we can make two especially peculiar observations: For the Fisher Z CI, despite the

decrease in bias, coverage rate worsens, and for the asymptotically distribution-free Fisher Z CI,

again despite decreasing bias, coverage balance deteriorates. Why is this so?
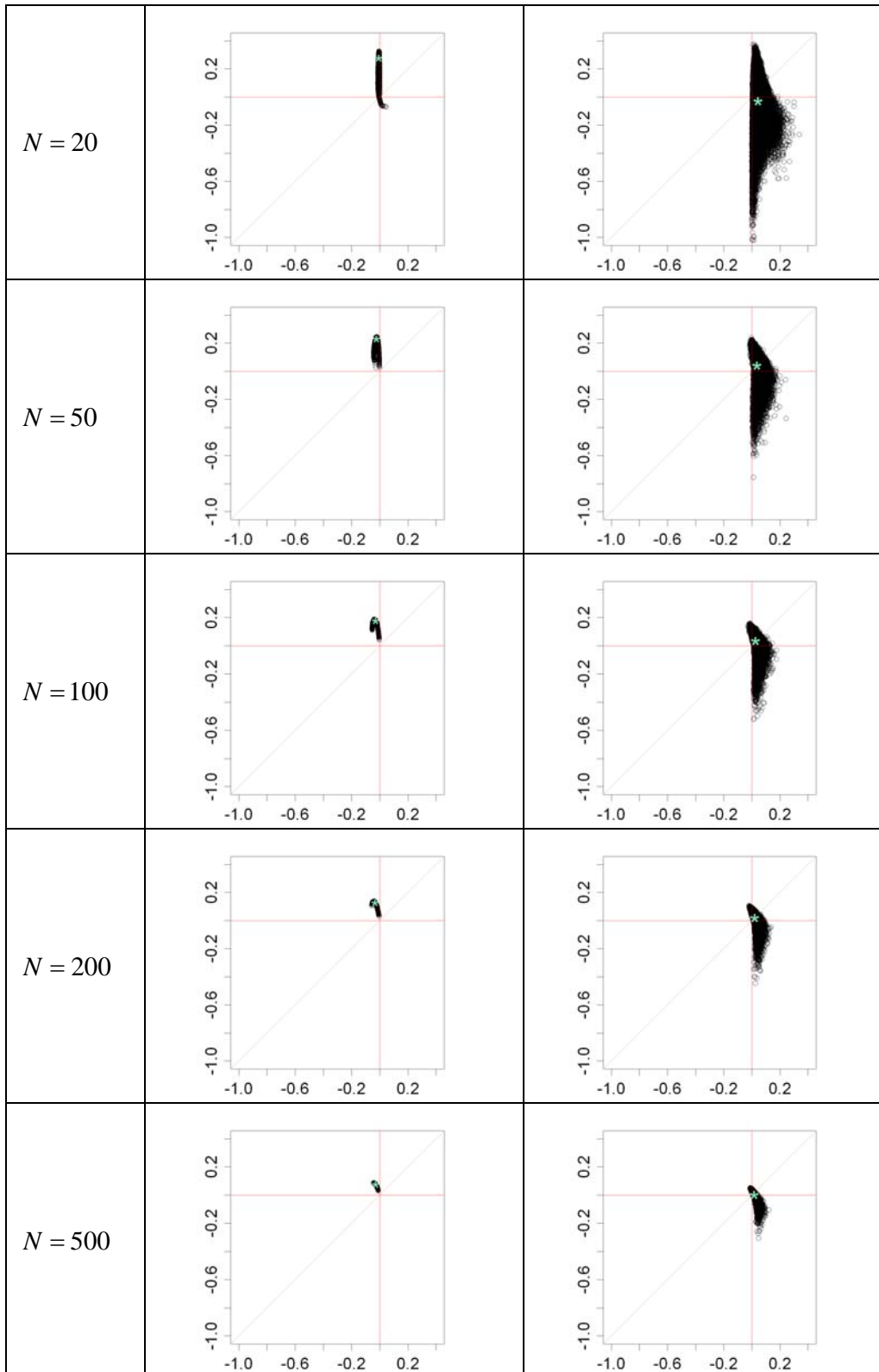
Figure 60: Absolute CI Limit Difference Score Plots for Increasing N; Dist. 2(a), $\rho = .8$

Table 58: Coverage Rate, Coverage Balance, and Difference Score Summaries for Figures Figure 55 through Figure 58.

N	Fisher		ADF		ADF Fisher		BCa		Univariate Sampling	
Coverage Rate and Balance										
20	0.892	15.538	0.793	4.327	0.891	0.697	0.945	0.781	0.999	37.000
50	0.862	4.639	0.874	3.223	0.925	0.412	0.944	0.599	0.999	3.100
100	0.833	2.723	0.919	2.637	0.945	0.364	0.946	0.677	0.999	0.762
200	0.813	1.937	0.950	2.101	0.961	0.330	0.953	0.673	0.999	0.395
500	0.786	1.497	0.971	1.831	0.975	0.293	0.961	0.749	0.999	0.462
Difference Scores: Median (SD)										
20	-0.008 (0.003) 0.278 (0.073)		0.024 (0.046) 0.393 (0.081)		-0.007 (0.039) 0.265 (0.156)		0.007 (0.033) 0.247 (0.157)		0.045 (0.035) -0.029 (0.178)	
50	-0.024 (0.008) 0.231 (0.025)		0.028 (0.040) 0.246 (0.059)		-0.009 (0.029) 0.177 (0.111)		0.004 (0.030) 0.198 (0.080)		0.037 (0.022) 0.042 (0.104)	
100	-0.030 (0.009) 0.178 (0.015)		0.022 (0.035) 0.161 (0.046)		-0.009 (0.023) 0.118 (0.082)		0.001 (0.027) 0.137 (0.054)		0.027 (0.018) 0.037 (0.072)	
200	-0.033 (0.007) 0.128 (0.011)		0.013 (0.029) 0.099 (0.034)		-0.007 (0.019) 0.073 (0.056)		-0.002 (0.023) 0.087 (0.036)		0.021 (0.015) 0.020 (0.049)	
500	-0.029 (0.004) 0.077 (0.006)		0.006 (0.020) 0.047 (0.022)		-0.005 (0.014) 0.034 (0.032)		-0.002 (0.018) 0.043 (0.020)		0.016 (0.012) 0.003 (0.029)	

The difference score summaries and the graphs in Figure 60 are on the *absolute* CI limit difference scores, but these absolute differences do not take confidence interval width into account. As sample size increases, confidence intervals tend to narrow; a deviation of size $|.1|$ for one of the limits means something different for a CI with overall width .6 when $N = 20$ as opposed to a CI with overall width of .15 when $N = 500$. Therefore, Figure 61 below revisits the plots from Figure 60, but this time, the CI limit difference scores are relativized by the median width of the exact confidence interval.
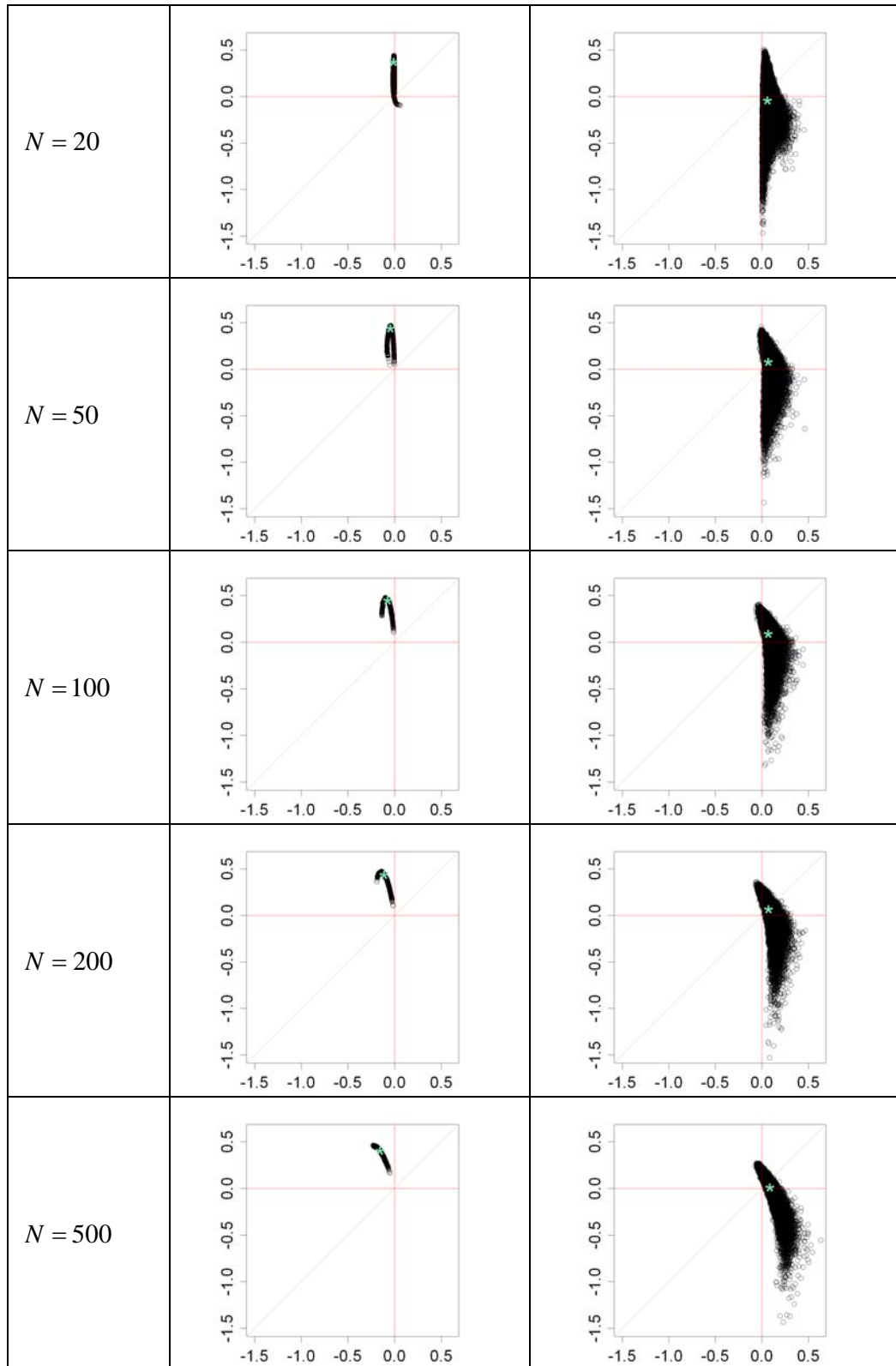
Figure 61: Relative CI Limit Difference Score Plots for Increasing N; Dist. 2(a), $\rho = .8$

The ranges of difference scores for the Fisher Z CI and $N = 20$ are now $[-0.02, \quad 0.01]$

and $[-0.05, \quad 0.45]$ and for $N = 500$, they are $[-0.14, \quad -0.08]$ and $[0.2, \quad 0.46]$. For the

Univariate CI, we have $[0.0, \quad 0.3]$ and $[-1.2, \quad 0.5]$ at $N = 20$ and $[-0.1, \quad 0.35]$ and

$[-0.9, \quad 0.3]$ at $N = 500$. Note that at $N = 500$, the Fisher Z CI is always too narrow, keeping

sort of equal distances from the upper and lower CI limit of the exact confidence interval, and

that the Univariate CI tends to be too wide. The scaled difference score summaries (second half

of Table 60) confirm this: All else remaining approximately the same, the estimation bias of the

upper CI limit for the Fisher Z CI increases with sample size, while variability increases for the

upper and decreases for the lower CI limit. For the Univariate Sampling CI, variability decreases

slightly. The largest improvements in estimation bias can probably be seen for the asymptotically

distribution-free CI and Fisher Z CI.

Table 59: Coverage Rate and Balance and Number Summaries – Various Sample Sizes

N	Fisher		Asymptotic		Asymptotic Fisher		BCa		Univariate Sampling	
Coverage Rate and Balance										
20	0.892	15.538	0.793	4.327	0.891	0.697	0.945	0.781	0.999	37.000
50	0.862	4.639	0.874	3.223	0.925	0.412	0.944	0.599	0.999	3.100
100	0.833	2.723	0.919	2.637	0.945	0.364	0.946	0.677	0.999	0.762
200	0.813	1.937	0.950	2.101	0.961	0.330	0.953	0.673	0.999	0.395
500	0.786	1.497	0.971	1.831	0.975	0.293	0.961	0.749	0.999	0.462
Difference Scores: Median (SD) – Relativized										
20	-0.010 (0.005) 0.374 (0.099)		0.032 (0.061) 0.530 (0.109)		-0.009 (0.052) 0.356 (0.210)		0.007 (0.045) 0.247 (0.211)		0.060 (0.048) -0.039 (0.239)	
50	-0.045 (0.016) 0.439 (0.047)		0.052 (0.075) 0.466 (0.112)		-0.017 (0.055) 0.337 (0.211)		0.008 (0.058) 0.375 (0.152)		0.070 (0.041) 0.080 (0.197)	
100	-0.075 (0.022) 0.449 (0.037)		0.055 (0.088) 0.405 (0.116)		-0.022 (0.059) 0.298 (0.207)		0.003 (0.069) 0.346 (0.135)		0.069 (0.045) 0.092 (0.182)	
200	-0.111 (0.023) 0.436 (0.037)		0.044 (0.098) 0.338 (0.117)		-0.025 (0.063) 0.249 (0.192)		-0.005 (0.080) 0.298 (0.121)		0.073 (0.052) 0.068 (0.167)	
500	-0.155 (0.021) 0.405 (0.031)		0.033 (0.105) 0.250 (0.114)		-0.025 (0.071) 0.181 (0.167)		-0.008 (0.097) 0.229 (0.108)		0.087 (0.065) 0.014 (0.151)	

Does Performance Vary With Confidence Level? Last but not least we may wonder whether the bias in estimating the limits of the Exact CI varies with confidence level. In Figure 62 I compare the Fisher Z to the BCa CI, for confidence levels 68%, 80%, 90%, 95%, and 99%, using a relativized CI limit difference score plot similar to Figure 61. The nonnormal distribution is again distribution 2a, with $N = 50$ and $\rho = .8$.

We can see that there is not much change in the performance of the Fisher Z CI, other than perhaps a slight increase in the bias of the lower CI limit (the center of the distribution moves up the y-axis a little). For the BCa CI, the overall variability decreases, i.e. relative to the width of the Exact CI, the CI behaves more consistently. There is a some increase in bias of lower CI limit estimation as well. When we consult Table 60, it is interesting to note that bias for

estimation of the lower limit increases for all CIs but the Univariate Sampling CI, which

maintains approximately equal performance across all confidence levels.

Table 60: Coverage Rate and Balance and Number Summaries – Various Confidence Levels

	Fisher		Asymptotic		Asymptotic Fisher		BCa		Univariate Sampling	
Coverage Rate and Balance										
68%	0.453	1.572	0.511	1.614	0.535	1.239	0.562	0.942	0.751	1.751
80%	0.558	1.741	0.615	1.787	0.652	1.124	0.680	0.879	0.876	1.765
90%	0.670	2.065	0.719	2.087	0.770	0.922	0.796	0.785	0.960	1.932
95%	0.752	2.561	0.787	2.373	0.843	0.746	0.867	0.708	0.988	1.966
99%	0.862	4.639	0.874	3.223	0.925	0.412	0.944	0.599	0.999	3.100
Difference Scores: Median (SD)										
68%	-0.068 (0.021) 0.376 (0.064)		0.018 (0.117) 0.337 (0.148)		-0.020 (0.093) 0.281 (0.181)		-0.032 (0.137) 0.222 (0.214)		0.113 (0.107) 0.108 (0.239)	
80%	-0.068 (0.022) 0.377 (0.059)		0.024 (0.114) 0.355 (0.143)		-0.025 (0.086) 0.285 (0.186)		-0.021 (0.118) 0.244 (0.203)		0.090 (0.088) 0.091 (0.233)	
90%	-0.063 (0.021) 0.397 (0.053)		0.035 (0.107) 0.385 (0.135)		-0.025 (0.076) 0.297 (0.194)		-0.009 (0.097) 0.276 (0.189)		0.088 (0.070) 0.083 (0.224)	
95%	-0.056 (0.019) 0.410 (0.049)		0.044 (0.097) 0.413 (0.127)		-0.022 (0.068) 0.313 (0.200)		-0.001 (0.082) 0.308 (0.176)		0.082 (0.058) 0.078 (0.215)	
99%	-0.045 (0.016) 0.439 (0.047)		0.052 (0.075) 0.466 (0.112)		-0.017 (0.055) 0.337 (0.211)		0.008 (0.058) 0.375 (0.152)		0.070 (0.041) 0.080 (0.197)	
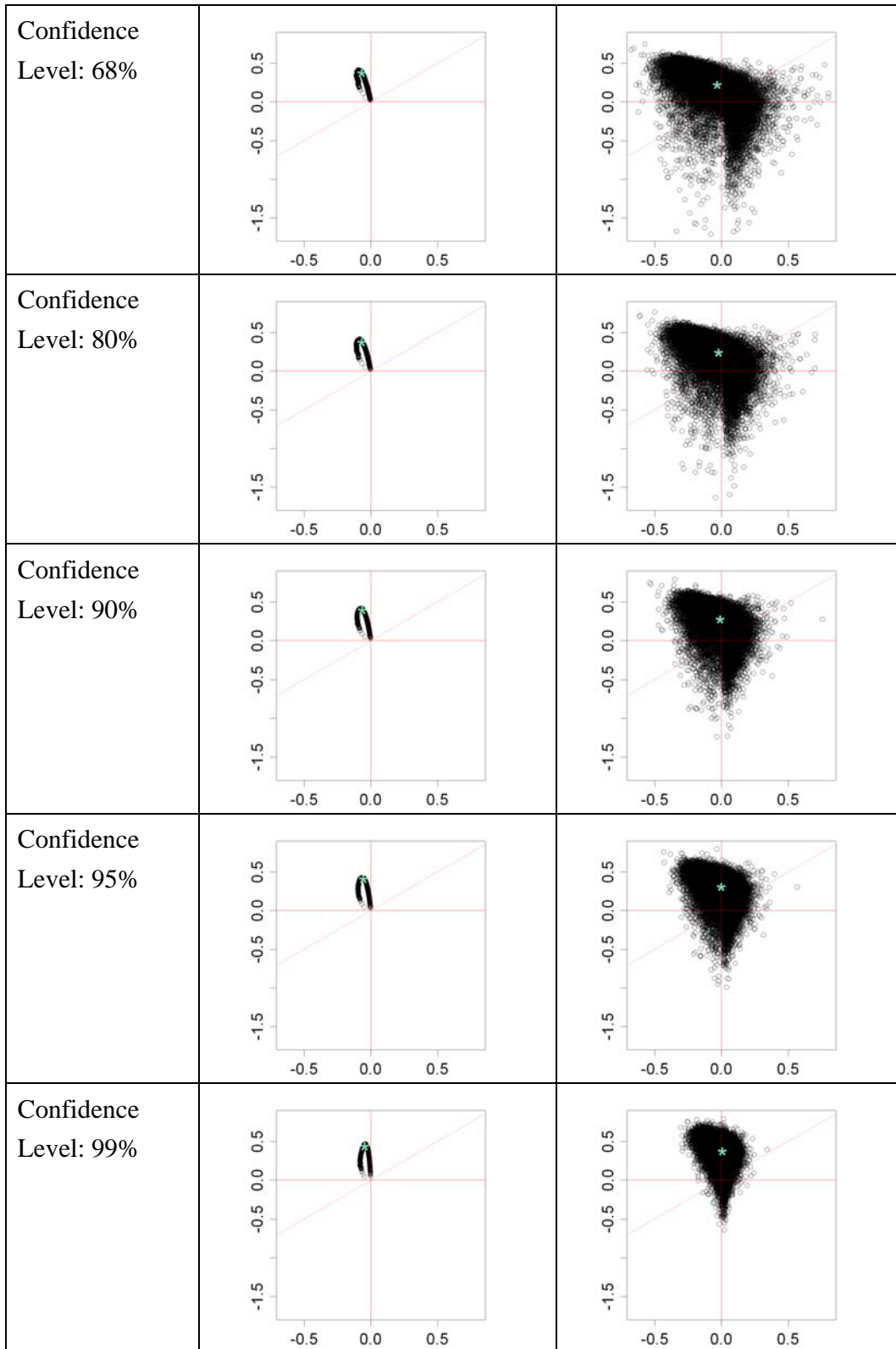
Figure 62: Relative CI Limit Difference Score Plots for Decreasing α ; Dist. 2(a), $\rho = .8$

Evaluating Confidence Intervals –Distributions with non-Identical Marginals.

Without the Exact CI available for nIM distributions, we cannot create CI limit difference scores and evaluate the performance of the approximate interval procedures in relation to the Exact CI. Instead, I choose one nIM distribution for a more detailed comparison of coverage rate and balance as well as a variation of plot type 1, which allows us to examine the variability of the confidence interval limits across different CI procedures. Observed values for coverage rate, coverage balance, and width for all nIM distributions can be found in Appendix C. Distribution 41b2a is the most aberrant distribution in terms of shape, suggesting two distinctly different subpopulations. One of my predictions on CI performance was that the BCa CI would outperform the Univariate Sampling CI in such a situation. Some interesting patterns can be observed in Table 61: Coverage rate worsens for the Fisher Z CI and the Univariate Sampling CI, albeit not dramatically, while their balance indices improve. The opposite is true for the two asymptotically distribution-free CIs, with significant worsening of coverage balance for the asymptotically distribution-free CI in particular. Only for the BCa CI do coverage rate and balance improve with increasing sample size.

Table 61: Coverage Rate and Balance For Five Approximate CIs, Distribution 41b2a, ρ = .5

N	Fisher		ADF		ADF Fisher		BCa		Univariate	
20	0.237	2.005	0.370	2.824	0.403	2.390	0.507	3.925	0.268	2.094
	0.303	2.112	0.441	3.420	0.499	2.691	0.606	5.284	0.360	2.196
	0.384	2.286	0.515	4.380	0.602	3.269	0.712	9.146	0.477	2.323
	0.450	2.459	0.566	5.444	0.677	4.042	0.776	3.664	0.584	2.479
	0.570	2.873	0.635	7.754	0.784	6.350	0.858	29.472	0.760	2.838
50	0.232	1.737	0.469	2.707	0.492	2.184	0.550	2.811	0.274	1.787
	0.297	1.805	0.558	3.477	0.597	2.480	0.656	3.670	0.354	1.860
	0.374	1.919	0.643	4.890	0.708	3.074	0.762	5.701	0.453	1.913
	0.439	2.011	0.699	6.937	0.781	3.906	0.828	8.771	0.536	1.959
	0.554	2.219	0.774	12.609	0.877	7.107	0.907	21.104	0.685	2.024
100	0.223	1.574	0.523	2.576	0.538	2.042	0.578	2.274	0.276	1.619
	0.286	1.625	0.626	3.527	0.655	2.398	0.693	2.901	0.355	1.662
	0.364	1.693	0.720	5.410	0.766	3.118	0.800	4.257	0.449	1.701
	0.425	1.748	0.776	8.083	0.835	4.014	0.862	6.040	0.529	1.733
	0.541	1.911	0.848	21.023	0.918	8.013	0.934	13.522	0.669	1.774
200	0.219	1.436	0.567	2.342	0.574	1.904	0.598	1.855	0.285	1.478
	0.281	1.468	0.674	3.231	0.691	2.228	0.715	2.213	0.362	1.498
	0.354	1.506	0.772	5.302	0.803	2.860	0.823	2.973	0.457	1.537
	0.418	1.560	0.830	8.597	0.870	3.828	0.886	3.902	0.534	1.557
	0.529	1.646	0.898	27.406	0.944	7.959	0.953	7.607	0.665	1.578
500	0.212	1.310	0.611	2.083	0.613	1.747	0.621	1.504	0.292	1.352
	0.271	1.333	0.724	2.844	0.732	2.084	0.740	1.732	0.372	1.368
	0.345	1.365	0.822	4.862	0.838	2.714	0.845	2.101	0.466	1.384
	0.406	1.369	0.878	8.513	0.899	3.859	0.906	2.771	0.546	1.414
	0.519	1.422	0.938	36.155	0.962	8.555	0.966	4.658	0.674	1.444

Figure 63 is a variation of the CI limit-sample *r* plot: The Fisher *Z* CI in green – a well performing interval if the parent distribution is bivariate normal – is contrasted with the other four confidence intervals in red, one CI in each cell. The true ρ is marked as a blue line; makes it easier to identify CIs that miss the true ρ above and below. At $N = 20$, this nonnormal distribution produces sample correlations across the entire interval $[-1, \ 1]$. The two CIs with the worst coverage rate, the Fisher and the Univariate Sampling CI, behave quite similarly, and they are much narrower than the other CIs, thereby not containing the true ρ. On the other hand,

the Fisher Z CI and Univariate Sampling CI have a considerably lower sampling variability for their lower and upper CI limits at any given value for the sample correlation.
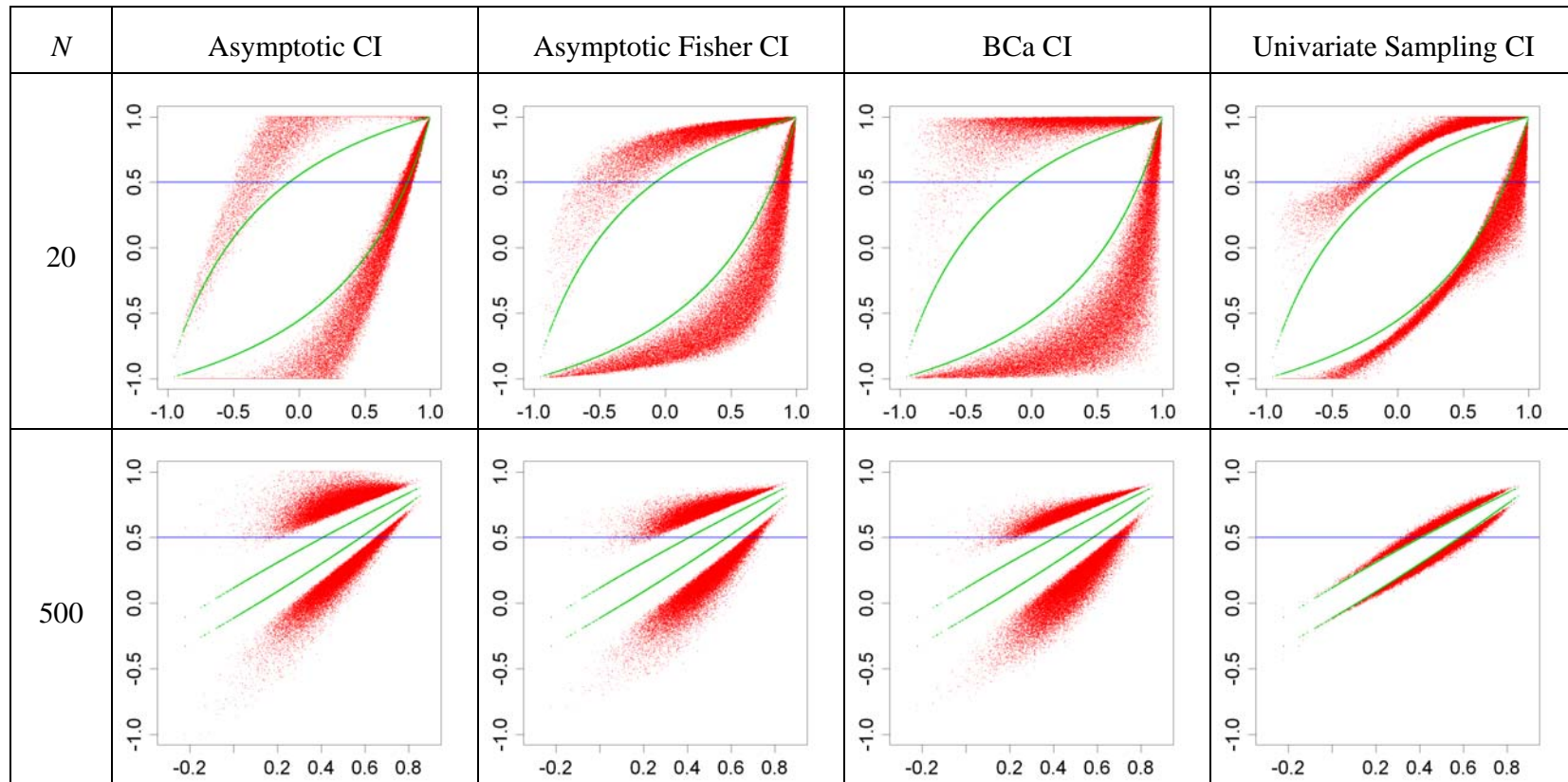
Figure 63: CI Limit-Sample Correlation Plots Comparing the Fisher Z CI to the Other Four Approximate CIs; Distribution 4(a), $\rho = .5$

Discussion

The main goal of Part II of my dissertation is to explore the performance of confidence intervals for the Pearson product-moment correlation under nonnormality of the underlying parent distribution. I compare three different approaches to constructing a confidence interval around r, resulting in five different intervals: a) The parametric Fisher Z CI, b) Two asymptotically distribution-free confidence intervals, and c) Two bootstrap confidence intervals. Confidence intervals for the simple correlation r were chosen because r is a well-known statistic with a large body of research behind it, but a considerably complicated sampling distribution. Research on its robustness has not lost its relevance and robust confidence intervals for a single correlation r or a simple pattern hypothesis for correlations are still currently being investigated (Beasley et al, 2007; Zou, 2007).

Previous Monte Carlo studies exploring the robustness of confidence intervals (such as the Fisher Z CI) and hypothesis tests for r have left somewhat of a rag rug of results. When based on assessment of simple Type I error rate or coverage rate, most studies indicate that the Fisher Z CI is robust when $\rho = 0$ and the variables are truly independent (Pearson, 1929; Strube, 1988), but find lack of robustness when $\rho = 0$, but the variables are dependent in some nonlinear fashion (Edgell & Noon, 1984). When $\rho \neq 0$, both the presence of robustness (Pearson, 1929) and lack of robustness (Beasley et al., 2007) have been demonstrated. Sievers (1996) concluded that the performance of an asymptotically distribution free CI was not superior to other intervals available such as the Fisher Z CI or bootstrap confidence intervals. When comparing BCa and univariate sampling bootstrap confidence intervals for r, Lee & Rodgers (1998) and Beasley et al. (2007) seem to have observed superior performance for the univariate sampling approach.

Some of this cloud of inconclusiveness may be precipitated as none of these studies evaluate all the information confidence intervals can provide. The overwhelming majority of Monte Carlo research on confidence intervals, especially in psychometrics, only presents Type I error rate or occasionally coverage rate, while ignoring such aspects as coverage balance and width, for example. Closeness to some form of exact confidence interval has – to my knowledge – not been investigated in a quantitative manner at all. On occasion, researchers (Efron, 1988, Rasmussen, 1987; Sievers, 1996) have conducted individual comparisons between bootstrap and other approximate confidence intervals for a handful of samples, but never have they included quantifying analyses.

Further, many studies on the robustness of confidence intervals around a single correlation comprise results for sample sizes that are too small to yield a confidence interval of information-bearing width. A CI that covers 80% of the entire parameter space, as is the case when sample sizes are as small as $N = 5$, is barely helpful at all, and most researchers will not calculate confidence intervals around r for such small samples.

As has been shown in Part I of my dissertation, some simulation studies contain at least one bi-/multivariate nonnormal distribution with an odd shape. The shape of these nonnormal distributions are often best explained as consisting of several subpopulations or demonstrating a blatantly nonlinear relationship. Under such circumstances, a single linear relationship is not a particularly meaningful model. Users need to be made aware that performance of seemingly all confidence interval procedures for a single correlation is severely hampered if the population being investigated does not warrant a single, linear model. Frequently, the combination of small sample size and odd-shaped distribution produces the most striking results, which in turn will make the greatest impression in the discussion of results and be remembered the most clearly.

Researchers should make an effort to utilize nonnormal distributions that are likely to occur in real data at sample sizes that are realistically used in correlational analyses.

A New Approach. Based on these observations from previous studies, I propose a more detailed approach of examining properties of confidence intervals, transcending simple Type I error rate and coverage rate. The properties that have been examined in this study, coverage rate and balance, width, closeness to the Exact CI, character and degree of bias when estimating the Exact CI, and variability of CI limits, are extremely useful since they touch on the *nature* of confidence intervals. If we were just interested in Type I error rate, in most cases we wouldn't need a confidence interval, as a hypothesis test will be readily available and often easier to calculate. But confidence intervals provide additional, crucial information such as precision with which a parameter has been estimated, and lower and upper confidence limits, whose behavior is often subject to certain expectations.

Coverage rate can tell us how likely it is that a confidence interval will cover the true parameter, and some may argue that it is the most important aspect of a confidence interval, but it provides only limited information on the behavior of the confidence interval *limits*. Yet, the behavior of the confidence interval limits themselves is highly important. Confidence intervals were developed to estimate uncertainty, to gauge the variability of the sampling distribution of the parameter of interest. Wide (relative to some standard, such as standard deviation in the sample or, as for the correlation, width of parameter space) confidence interval limits indicate low precision, narrow CI limits indicate high precision. Knowing that a two-sided CI will cover the true parameter 95% of the time without any information on coverage balance, the actual confidence interval limits provide nothing but conservative boundaries of lower or upper CI

240

bounds for a *one-sided* 95% CI, or, equivalently, a two-sided *90%* CI. Why is this so? If we do not know anything about coverage balance, a 95% CI may lie below and above the true parameter equally at a rate of 2.5%. It may also lie above the true parameter 1% of the time and below it 4% of the time. In the most extreme case, the entire 5% of miscoverage would occur because the CI systematically misses the true parameter on one side only. A fair few may protest now and object that the interval still covers the true value 95% of the time, so how can it be compared to a 90% CI? One may call this observation the "Confidence Interval Paradox."

If we add coverage balance to the evaluation of a CI procedure, we can now answer the question of whether the CI misses the true parameter evenly on either side. However, there will still be a few caveats. We do not know any additional properties of the CI limits themselves, and the little we know is limited to the confidence level we included, say 95% (note that for some of the CIs, the balance index switches from being greater than 1 at the 68% level to being less than 1 at the 99% level, see Table 68, Univariate Sampling CI, $N = 100$). As a remedy, we could cover a large amount of confidence levels in our analyses, as has been done in this study. Nevertheless, we would still fail to keep track of one other aspect: The amount by which the approximate CI misses the Exact CI limits. This level of depth of performance analysis is reached in the graphs and numerical summaries presented in this study.

Users of confidence intervals tend to expect certain behavior from the lower and upper limit of a two-sided confidence interval. Most commonly, people presume that both limits of a two-sided confidence interval provide a kind of boundary of parameter values. It is generally believed that parameter values more extreme than the values of the lower and upper confidence interval limits are unlikely to have produced the observed data. For the traditional confidence interval around a single mean, which is frequently the first CI taught in undergraduate statistics

241

classes, this expectation is fulfilled. It is also fulfilled for the Exact CI introduced in this part of my dissertation. The confidence interval limits of the Exact CI provide information on the probability with which a certain value for ρ would have produced the observed sample r. The CI limit vs. sample r scatterplots and CI limit difference score scatterplots can therefore be used to explore the relationship between approximate CI limits, sample correlation value, and exact CI limits in great detail. And when providing these graphs for all conditions in a simulation study is not feasible, the median and standard deviation of the CI limit difference scores can summarize these relationships. The median differences provide information on whether the confidence interval procedure of interest systematically over- or underestimates the Exact CI, or whether it is too wide or too narrow. The standard deviation of the CI limit difference scores brings us to the last aspect of confidence interval performance discussed in my dissertation: Variability of CI limits. Less variability means that the confidence interval displays more steady behavior, independently of whether it, e.g., grossly overestimates the Exact CI, or whether it tends to lie right on top of it. The Fisher Z CI shows very consistent behavior; and in fact, for a given sample correlation, confidence level, and sample size, the lower and upper confidence interval limits are determinate. The other CIs, however, react to additional properties of the sample, thereby introducing more variability in the values of lower and upper CI limits which translates into more uncertainty in the estimation of a lower and upper limit for the value of the true parameter. This quality is related to predictability in general: Assume we have two confidence intervals to choose from and investigate their coverage rate at the 90% confidence level. One confidence interval has a coverage rate of 90% in most simulation conditions, but in 10% of cases has a coverage rate of about 50%. The other confidence interval always has a coverage rate of 80%. We may prefer the second confidence interval due to its predictability, especially if we cannot be

sure how often the second type of data occurs in the real world or if we are unable to simulate all kinds of nonnormality.

Including width in evaluation results may be very informative for a researcher that needs to choose between different methods, and doing so could have potentially prevented some authors of Monte Carlo studies to include results for correlations calculated on samples of size $N = 5$. As most researchers will calculate correlations on samples of size $N = 50$ and up, I concentrated on simulation results obtained for sample sizes of $N = 50$ through $N = 500$.

Alternative Approaches to Dealing with Nonnormality

In addition to developing and finding statistical procedures for correlations that will perform well under both normality and nonnormality of the parent distribution, other approaches have been suggested. These include transformations of the marginals and robust measures of association.

Transformations

Transformation of the marginal distributions is a popular technique and statistics textbooks tend to devote a section or chapter to it. Often, a decidedly nonnormal distribution with strong skewness in the marginal distributions or a nonlinear relationship between variables is subjected to transformation of one or both marginals to create a bivariate shape that is much closer to bivariate normality. For example, a nonlinear, concave relationship between two variables X and Y, with X on the x-axis, can often be linearized by transforming X to $\sqrt{X}$ or $\ln(X)$. However, after the transformation, we are left with *transformed* variables, which may or may not retain their interpretability.

There is an extended literature discussing the advantages and disadvantages of such transformations for bivariate or multivariate data. Halperin (1986) offers an overview and discussion of the most common uses and issues of transformations of the marginals, recommending transformation when skewness is present. Kowalski (1972) presents findings according to which the sampling distribution of r for a bivariate exponential distribution whose marginals have been transformed is close to the sampling distribution of r under bivariate normality of the parent distribution. He offers further readings, citing research that found better performance of normal-based tests for independence after transformation of marginals to normal. Dunlap, Burke, & Greer (1995) discuss the effect of skew on product-moment correlations, observing an increase in the size of $|r|$ after skewness-reducing transformations. The inverse, a decrease in correlation after transforming bivariate normal variables to nonnormality has been stated and investigated by Lancaster (1957). Games (1983) offers additional discussion on pros and cons of transforming variables, albeit mainly in the context of ANOVA, cautioning against indiscriminate use of transformations.

Certainly, whether a nonlinear transformation of content-bearing variables is appropriate must be decided on a case to case basis and cannot be answered here. The present author sees value in not transforming variables but instead describing a nonlinear relationship as nonlinear, e.g. in the form $Y = f(X) = a_0 + a_1 X + a_2 X^2$. It seems that transforming variables to create a more linear relationship or to make methods that have been developed for linear, bivariate normal data applicable will disguise nonlinear relationships when that specific property of nonlinearity may provide valuable information. As one small example, one may consider some intervention that loses its effectiveness after a certain point because the intervention-result relationship asymptotes.

Robust Measures of Association

A few robust measures of association have been developed. Among them are robust estimators described in Devlin, Gnanadesikan, & Kettenring (1975), the winsorized correlation (Wilcox, 1993) and the percentage bend correlation (Wilcox, 1994). Robust estimators of correlation and regression seem to have gained relatively little popularity outside of statistical or psychometric research. A search of studies citing these three articles found almost no studies in psychology or other behavioral sciences making use of these techniques. Without being able to explore and examine all robust correlational methods available in great detail, this lack of acceptance may, in part, come from a lack of knowledge of these methods and their functionality. Further, some of these robust measures of association may lack the rich methodology that has been developed for parametric estimators, including confidence intervals.

In addition, I would like to draw attention to one issue I encountered while familiarizing myself with the percentage bend correlation (Wilcox, 1994): For two variables *X* and *Y*, the sample percentage bend correlation is calculated as

$$r_{pb} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2 \sum_{i=1}^{n} B_i^2}} \tag{122}$$

where $A_i = \Psi\left[\left(X_i - \hat{\phi}_{pbx}\right)/\hat{\omega}_x\right]$ and $B_i = \Psi\left[\left(Y_i - \hat{\phi}_{pby}\right)/\hat{\omega}_y\right]$. $\hat{\phi}_{pbx}$ and $\hat{\phi}_{pby}$ are percentage bend location estimators, that is, some robust estimator of location, and $\hat{\omega}_x$ and $\hat{\omega}_y$ are robust measures of scale (in his 1994 article, Wilcox uses the 90th percentile of the deviation scores from the median). Finally, the function $\Psi(x) = \max\left[-1, \quad \min(1, x)\right]$ is used to confine all A_i and B_i scores to not exceed the interval $\left[-1, \quad 1\right]$. Notice that this last step rearranges outliers to

245

not exceed some value based on the rescaled distribution and thath the assessment of the status "outlier" is on the *univariate* level.

Wilcox demonstrates the percentage bend correlation on a contaminated distribution, which is constructed as follows:

1. One sample of size *N* from a standard normally distributed variable *X* and one sample of size *N* from a univariate contaminated variable *C*, independent of *X* are constructed. The contaminated sample consists of 90% scores sampled from $N(0,1)$ and 10% scores sampled from $N(0,10)$.

2. A new variable *Q* is created from *X* and *C* with $\rho_{xc} = .8$ as follows:

$$Q = \rho_{xc}X + \sqrt{1 - \rho_{xc}^2}\,C \tag{123}$$

3. The bivariate distribution used to demonstrate the percentage bend correlation is the distribution that exists between *X* and *Q*, and is plotted in Figure 64:
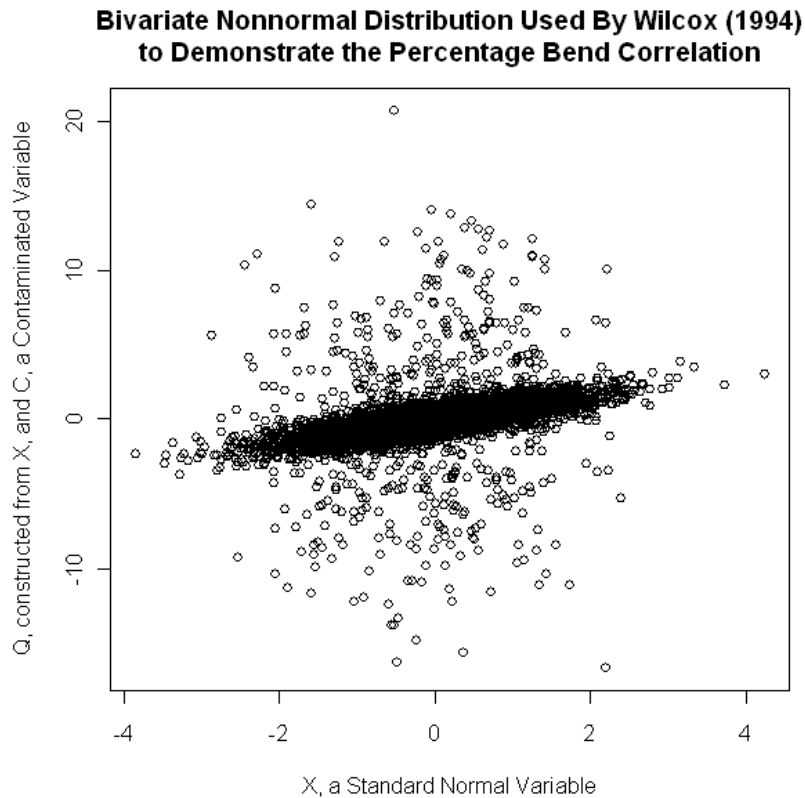
**Bivariate Nonnormal Distribution Used By Wilcox (1994)
to Demonstrate the Percentage Bend Correlation**

*Figure 64: Bivariate Nonnormal Distribution Used to Demonstrate the Percentage Bend
Correlation*

Wilcox argues that the part of the distribution we would be interested in is the dense cloud of

points that exhibits a strong relationship. While the traditional Pearson product-moment

correlation for the *entire* distribution in Figure 64 is only $\rho = .374$, the percentage bend

correlation is estimated at $r_{pb} = .634$, much closer to the underlying correlation of $\rho = .8$ for the

"interesting" part of the distribution. However, there are two questions that need to be answered:

(1) Which of these correlations is more meaningful, more representative of the distribution? (2)

How exactly does the percentage bend correlation work? The first question is strongly related to

the issue of whether robustness to some particular distribution is *desired*. The bivariate

distribution between the variables *X* and *Q* is a mixture distribution and before we use this

distribution to investigate the robustness of r, we need to ensure that a single correlation is a statistical procedure researchers would realistically apply. Do we want the measure of association calculated for a sample from the distribution in Figure 64 to turn out as estimating $\rho = .8$, that is, do we want to only measure the correlation for the majority of the points, and discard the small subpopulation that demonstrates very different behavior? Why then do we not attempt to model the two separate subpopulations? Definite answers to these questions must be found elsewhere.

To answer the second question, reconsider Equation (122): The percentage bend correlation is a regular product-moment correlation calculated on the rescaled and, most importantly, *rearranged* (Wilcox calls them trimmed, however they are not truly trimmed, since they are still used in the calculation of the percentage bend correlation) scores A_i and B_i. We can plot A_i and B_i to visualize the distribution for which r_{pb} is computed:

Altered Bivariate Nonnormal Distribution
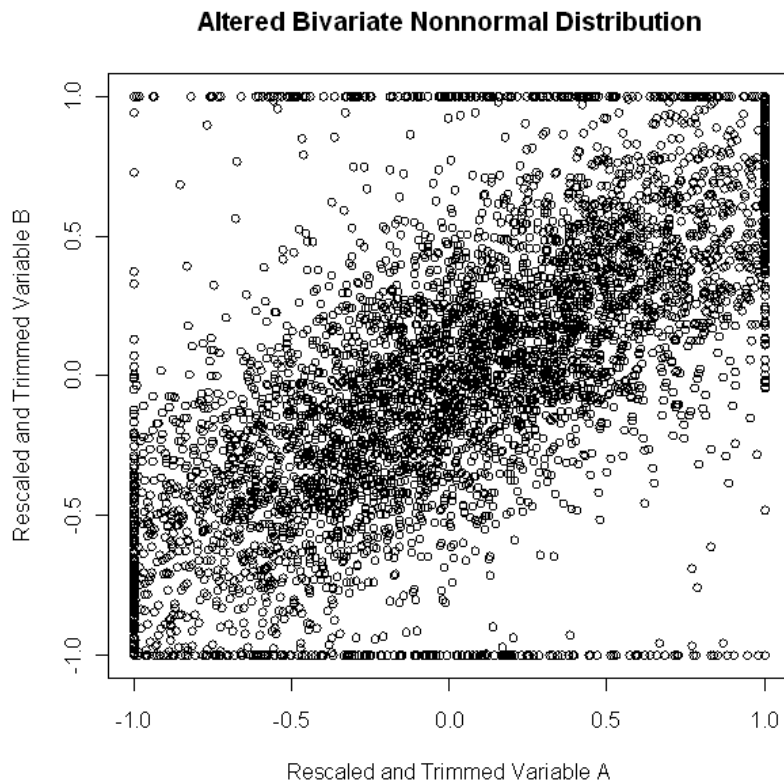
Figure 65: Bivariate Distribution Based on Which the Percentage Bend Correlation is Calculated

To our surprise, the distributional shape in Figure 65 is not particularly representative of what Wilcox considers to be the 'interesting' part of the distribution in Figure 64. While the percentage bend correlation yields a higher numerical value than the product-moment correlation in this case, it does not help uncover the nature of the distribution in Figure 64. In the end, a procedure that identifies bivariate rather than just univariate outliers may be more promising.

It does not lie within the scope of this study to provide a complete answer on the advantages and disadvantages of alternative approaches to tackling nonnormal data. My objective is merely to draw attention to possible consequences of attempting to automate the handling of nonnormal data, be it real or simulated. A plot should always be drawn, and Figure

249

65 demonstrates that the data the percentage bend correlation is eventually based on may be far from representative for the measured phenomenon. Several additional examples could be provided in which the percentage bend correlation procedure substantially distorts the distribution for which a measure of association is computed (some of these, e.g. Wilcox's own version of a multivariate *g*-and-*h* distribution with $\rho = .5$, appear directly in his 1994 article). Note that a bootstrap confidence interval for the percentage bend correlation is available (Wilcox, 2005).

Future Directions. Alongside with lower and upper confidence interval limits and sample correlations, sample skewnesses and kurtoses were recorded for each replication as well. I believe that including these values in additional analyses has potential, since especially the two asymptotic confidence intervals are built to react to sample skewnesses and kurtoses. These analyses have been set aside for future research due to sheer volume.

Some of the predictions I presented in the proposal of my dissertation were confirmed, while others were not. For example, the improvement of performance for the asymptotically distribution-free and bootstrap CIs with increased sample size was slower than predicted or did not occur at all. It would be interesting to investigate the behavior of approximate CIs for very large sample sizes and see whether the asymptotically distribution-free and bootstrap CIs eventually estimate the true CI very well.

Lastly, the discussion of over- and underestimation of confidence interval limits for an Exact CI, of being too wide or too narrow could possibly be refined. The perfect coverage displayed by the Exact CI is only a by-product of its origin, namely the construction by pivoting the (empirical) cdf. The relationship between lower and upper confidence limits and the

percentiles of an empirical sampling distribution given the population ρ make it a very strong

candidate for a gold standard that can be used to establish a pragmatic definition for confidence

interval bias.

GENERAL DISCUSSION

Both Part I and II of my dissertation wish to improve the interconnection between common practice in Monte Carlo research, which is traditionally used to investigate performance of statistical methods under a variety of conditions, and the actual application of such statistical methods. Part I examines three techniques for simulating multivariate nonnormal distributions and demonstrates their properties, including neglected or to this point unknown ones. The 3rd order power method, 5th order power method, and the g-and-h distribution all have more than one set of coefficients that can be used to simulate a given skewness-kurtosis combination. This has several implications: Even though population skewness and kurtosis will be the same from one set of coefficients to another, the expected values of sample skewness and kurtosis and the shape of the distribution will almost always be different. I have demonstrated this for both univariate and bivariate distributions, and it will be true for higher dimensions as well; therefore, marginal population skewnesses and kurtoses are hardly sufficient when describing a nonnormal distributions. What range of values for the final correlation ρ_Y between two nonnormal variables can be simulated depends on the values chosen for marginal skewnesses and kurtoses, the simulation method, and the set of transformation coefficients once a simulation method is chosen. Part I concludes by showing that the choice of transformation coefficients, holding everything else constant, can also impact simulation results in Monte Carlo studies on the robustness of some statistical procedure, using two confidence intervals for a single correlation to demonstrate these effects.

Part II chooses five approximate confidence intervals for a single correlation to closely examine CI performance under nonnormality, discussing several criteria, techniques, and

viewpoints. The traditional method of CI performance evaluation commonly boils down to just a simple comparison of coverage rates, thereby disregarding large amounts of information that can provide a more detailed picture of the confidence interval's behavior. I introduce an exact confidence interval built from empirical sampling distributions, which is closely related to the 'pivoting the cdf' technique described by Steiger & Fouladi (1997). This exact CI combines exact coverage with a first meaningful definition of a gold standard when no theoretically derived 'optimal' interval is available. Thereby, in addition to coverage rate, coverage balance, and width of the CIs, we can compare approximate CIs to the Exact CI, paving the way for a practical definition of the bias of a confidence interval. Two particularly insightful scatterplot techniques display the variability of CI limits for a given sample correlation or around the CI limits of the Exact CI. They can also provide information on the type of bias an approximate CI shows, such as over- and underestimation, or being too wide or too narrow compared to the Exact CI.

Implications and Future Research

The main message from Part I and II of my dissertation must be: Use Explorative Data Analysis (EDA) to reconnoiter your output! We should not rely on just a few numerical values to summarize the behavior of very complex objects such as nonnormal multivariate distributions or the behavior of confidence intervals. In quantitative research particularly, there is no Type I error inflation or other types of penalty for experimentally exploring results. Using EDA has helped us to discover that the traditional approach of characterizing multivariate nonnormal distributions by their marginal population skewnesses and kurtoses is insufficient. It has also uncovered the behavior of approximate confidence intervals in relation to the sample correlation and to an

253

Exact CI that may be worthy of being called a gold standard to compare other CIs to. We may hope that visual inspection of results will inspire better summaries for assessing confidence intervals and maybe even new strategies for developing robust confidence intervals.

Secondly, the intent of my dissertation is also to motivate an increase in the quality of Monte Carlo studies by asking what the most important aspects of simulation conditions and result interpretations are. I go into details of distributional shape, sample size, and extensiveness with which to interpret results. By keeping such questions in mind, the high hope is that we may help research become more efficient.

In view of the fast proliferation of techniques to simulate multivariate nonnormal data and the large amount of already existing methods, this study naturally has to be limited. My dissertation only addresses the effects of distributional shape on the robustness of confidence intervals for the Pearson product moment correlation. Whether differences in distributional shape not captured by differences in marginal population skewnesses and kurtoses have an equally profound effect on other statistical procedures such as ANOVA, e.g., cannot be simple implied and poses an interesting opportunity for future research.

Moments of The 3rd Order Polynomial Transformation

We can find the first four moments about zero by expanding and applying expected value theory to the expected values of Y, Y^2, Y^3, and Y^4.

Using $E(Z) = E(Z)^3 = E(Z)^5 = E(Z)^7 = E(Z)^9 = ... = 0$, as well as $E(Z^2) = 1$, $E(Z^4) = 3$, $E(Z^6) = 15$, $E(Z^8) = 105$, $E(Z^{10}) = 945$, $E(Z^{12}) = 10395$, $E(Z^{14}) = 135135$, we find that

$$
\begin{aligned}
E(Y) &= E\left(a_0 + a_1 Z + a_2 Z^2 + a_3 Z^3\right) \\
&= a_0 + a_1 E(Z) + a_2 E(Z^2) + a_3 E(Z^3) \\
&= a_0 + 0 + a_2 + 0 \\
&= a_0 + a_2
\end{aligned}
\tag{124}
$$

If we set the mean of Y to be zero ($E(Y) = 0$), we have $a_0 = -a_2$.

$$
\begin{aligned}
E(Y^2) &= E\left(\left(a_0 + a_1 Z + a_2 Z^2 + a_3 Z^3\right)^2\right) \\
&= a_0^2 + 2a_0 a_2 E(Z^2) + a_1^2 E(Z^2) + 2a_1 d E(Z^4) \\
&\quad + a_2^2 E(Z^4) + a_3^2 E(Z^6) \\
&= a_0^2 + 2a_0 a_2 + a_1^2 + 6a_1 a_3 + 3a_2^2 + 15a_3^2 \\
&= a_1^2 + 6a_1 a_3 + 2a_2^2 + 15a_3^2
\end{aligned}
\tag{125}
$$

$$E\left(Y^3\right) = E\left(\left(a_0 + a_1Z + a_2Z^2 + a_3Z^3\right)^3\right)$$

$$= E\left(\left(-a_2 + a_1Z + a_2Z^2 + a_3Z^3\right)^3\right)$$

$$= -a_2{}^3 - 3a_1{}^2a_2E\left(Z^2\right) + 3a_2{}^3E\left(Z^2\right) + 3a_1{}^2a_2E\left(Z^4\right)$$

$$-3a_2{}^3\left(Z^4\right) - 6a_1a_2a_3E\left(Z^4\right) + a_2{}^3E\left(Z^6\right) + \tag{126}$$

$$6a_1a_2a_3E\left(Z^6\right) - 3a_2a_3{}^2E\left(Z^6\right) + 3a_2a_3{}^2E\left(Z^8\right)$$

$$= -a_2{}^3 - 3a_1{}^2a_2 + 3a_2{}^3 + 9a_1{}^2a_2 - 9a_2{}^3 - 18a_1a_2a_3$$

$$+15a_2{}^3 + 90a_1a_2a_3 - 45a_2a_3{}^2 + 315a_2a_3{}^2$$

$$= 2a_2\left(3a_1{}^2 + 4a_2{}^2 + 36a_1a_3 + 135a_3{}^2\right)$$

$$E\left(Y^4\right) = E\left(\left(a_0 + a_1Z + a_2Z^2 + a_3Z^3\right)^4\right)$$

$$= E\left(\left(-a_2 + a_1Z + a_2Z^2 + a_3Z^3\right)^4\right)$$

$$= a_2{}^4 + 6a_1{}^2a_2{}^2E\left(Z^2\right) - 4a_2{}^4E\left(Z^2\right) + a_1{}^4E\left(Z^4\right) - 12a_1{}^2a_2{}^2E\left(Z^4\right)$$

$$+6a_2{}^4E\left(Z^4\right) + 12a_1a_2{}^2a_3E\left(Z^4\right) + 12a_2{}^3a_3E\left(Z^5\right) + 6a_1{}^2a_2{}^2E\left(Z^6\right)$$

$$-4a_2{}^4E\left(Z^6\right) + 4a_1{}^3a_3E\left(Z^6\right) - 24a_1a_2{}^2a_3E\left(Z^6\right) + 6a_2{}^2a_3{}^2E\left(Z^6\right)$$

$$+a_2{}^4E\left(Z^8\right) + 12a_1a_2{}^2a_3E\left(Z^8\right) + 6a_1{}^2a_3{}^2E\left(Z^8\right) - 12a_2{}^2a_3{}^2E\left(Z^8\right)$$

$$+6a_2{}^2a_3{}^2E\left(Z^{10}\right) + 4a_1a_3{}^3E\left(Z^{10}\right) + a_3{}^4E\left(Z^{12}\right)$$

$$= a_2{}^4 + 6a_1{}^2a_2{}^2 - 4a_2{}^4 + 3a_1{}^4 - 36a_1{}^2a_2{}^2 + 18a_2{}^4 + 36a_1a_2{}^2a_3 + 90a_1{}^2a_2{}^2$$

$$-60a_2{}^4 + 60a_1{}^3a_3 - 360a_1a_2{}^2a_3 + 90a_2{}^2a_3{}^2 + 105a_2{}^4 + 1260a_1a_2{}^2a_3$$

$$+630a_1{}^2a_3{}^2 - 1260a_2{}^2a_3{}^2 + 5670a_2{}^2a_3{}^2 + 3780a_1a_3{}^3 + 10395a_3{}^4$$

$$= 3b^4 + 60b^2c^2 + 60c^4 + 60b^3d + 936bc^2d + 630b^2d^2 + 4500c^2d^2$$

$$+3780bd^3 + 10395d^4 \tag{127}$$

Moments of the g-and-h Distribution

$$\mu_1\left(g,h\right) = \frac{e^{g^2(2-2h)^{-1}} - 1}{g\sqrt{1-h}} \tag{128}$$

256

$$\text{var}(g,h) = \dfrac{\dfrac{1-2e^{g^2(2-4h)^{-1}}+e^{2g^2(1-2h)^{-1}}}{\sqrt{1-2h}}+\dfrac{\left(e^{g^2(2-2h)^{-1}}-1\right)^2}{h-1}}{g^2} \tag{129}$$

$$\gamma_1(g,h) = \dfrac{3e^{\frac{g^2}{2-6h}}+e^{\frac{9g^2}{2-6h}}-3e^{\frac{2g^2}{1-3h}}-1}{\sqrt{1-3h}}-\dfrac{3(1-2e^{\frac{g^2}{2-4h}}+e^{\frac{2g^2}{1-2h}})(e^{\frac{g^2}{2-2h}}-1)}{\sqrt{1-2h}\sqrt{1-h}}$$

$$+\dfrac{2\left(e^{\frac{g^2}{2-2h}}-1\right)^3}{\sqrt{(1-h)^3}}/g^3\sqrt{(\dfrac{1-2e^{\frac{g^2}{2-4h}}+e^{\frac{2g^2}{1-2h}}}{\sqrt{1-2h}}+\dfrac{\left(e^{\frac{g^2}{2-2h}}-1\right)^2}{h-1}/g^2)^3} \tag{130}$$

$$\gamma_2(g,h) = \left(\exp\left(\frac{8g^2}{1-4h}\right)\right.$$

$$\frac{\left(1 + 6\exp\left(\frac{6g^2}{4h-1}\right) + \exp\left(\frac{8g^2}{4h-1}\right) - 4\exp\left(\frac{7g^2}{8h-2}\right) - 4\exp\left(\frac{15g^2}{8h-2}\right)\right)}{\sqrt{1-4h}}$$

$$-\frac{4\left(3\exp\left(\frac{g^2}{2-6h}\right) + \exp\left(\frac{9g^2}{2-6h}\right) - 3\exp\left(\frac{2g^2}{1-3h}\right) - 1\right) * \left(\exp\left(\frac{g^2}{2-2h}\right) - 1\right)}{\sqrt{1-3h}\sqrt{1-h}}$$

$$-\frac{6\left(\exp\left(\frac{g^2}{2-2h}\right) - 1\right)^4}{(h-1)^2}$$

$$-\frac{12\left(1 - 2\exp\left(\frac{g^2}{2-4h}\right) + \exp\left(\frac{2g^2}{1-2h}\right)\right)\left(\exp\left(\frac{g^2}{2-2h}\right) - 1\right)^2}{\sqrt{1-2h}\,(h-1)} +$$

$$\left.\frac{3\left(1 - 2\exp\left(\frac{g^2}{2-4h}\right) + \exp\left(\frac{2g^2}{1-2h}\right)\right)^2}{2h-1}\right) \tag{131}$$

$$\Bigg/ \left(\frac{1 - 2\exp\left(\frac{g^2}{2-4h}\right) + \exp\left(\frac{2g^2}{1-2h}\right)}{\sqrt{1-2h}} + \frac{\left(\exp\left(\frac{g^2}{2-2h}\right) - 1\right)^2}{h-1}\right)^2$$

APPENDIX B

Studies That Used V&M to Simulate Nonnormal Data, Listed in Table 28:

Bauer & Curran (2003). Distributional Assumptions of Growth Mixture Models: Implications for Overextraction of Latent Trajectory Classes. *Psychological Methods, 8*(3), 338–363.

Benson, J., Fleishman, J.A. (1994). The robustness of maximum likelihood and distribution free estimators to nonnormality in confirmatory factor analysis. *Quality & Quantity, 28*(2), 117–136. see page 122.

Berkovits I, Hancock GR, Nevitt J. (2000). Bootstrap resampling approaches for repeated measure designs: Relative robustness to sphericity and normality violations. *Educational and Psychological Measurement, 60*(6), 877–892.

Enders CK. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data, *Psychological Methods, 6*(4), 352–370.

Enders CK, Bandalos DL. (1999). The effects of heterogeneous item distributions on reliability. *Applied Measurement in Education, 12*(2), 133–150.

Ferrando PJ, Lorenzo-Seva U. (1999). Implementing a test of underlying normality for censored variables, *Multivariate Behavioral Research, 34*(4), 421–439.

Flora DB, Curran PJ. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466–491.

Fouladi, R.T. (2000). Performance of Modified Test Statistics in Covariance and Correlation Structure Analysis Under Conditions of Multivariate Nonnormality. *Structural Equation Modeling, 7*(3), 356–410.

Fouladi RT, Yockey RD. (2002). Type I error control of two-group multivariate tests on means under conditions of heterogeneous correlation structure and varied multivariate distributions. *Communications in Statistics Simulation and Computation, 31*(3), 375–400.

Habib, A.R., Harwell, M.R. (1989). An empirical study of the type I error rate and power for some selected normal theory and nonparametric tests of the independence of 2 sets of variables. *Communications in Statistics Simulation and Computation, 18* (2), 793–826

Harwell, M.R. (1991). Using randomization tests when errors are unequally correlated. *Computations Statistics & Data Analysis, 11*(1), 75–85. - see page 80, table 1.

Harwell, M. R. & Serlin, R. C. (1988). An experimental study of a proposed test of nonparametric analysis of covariance. Psychol. Bull. 104, 268–281.

Harwell, M. R. & Serlin, R.C. (1989). A nonparametric test statistic for the general linear model. J. Educational Statist. 14, 351–371.

Hau KT, Marsh HW. (2004). The use of item parcels in structural equation modeling: Non-normal data and small sample sizes. *British Journal of Mathematical & Statistical Psychology, 57*, 327–351.

Hipp JR, Bollen KA. (2003). Model fit in structural equation models with censored, ordinal, and dichotomous variables: Testing vanishing tetrads. Sociological Methodology, Vol. 33, 267–305.

Jedidi K, Jagpal HS, DESarbo WS. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16*(1), 39–59. See page 54, right column: combined skewness (0 and 1) with kurtosis (0 and 2.75). At least one combination not possible under constraints.

Lix LM, Keselman HJ, Hinds AM. (2005) Robust tests for the multivariate Behrens-Fisher problem. *Computer Methods and Programs in Biomedicine, 77*(2), 129–139.

Lix LM, Fouladi RT. (2007). Robust step-down tests for multivariate independent group designs. *British Journal of Mathematical & Statistical Psychology, 60*, 245–265.

Nevitt J, Hancock GR. (2000). Improving the root mean square error of approximation for nonnormal conditions in structural equation modeling. *Journal of Experimental Education, 68*(3), 251–268.

Olejnik, S.F., Algina, J., 1984. Parametric ANCOVA and the rank transform ANCOVA when the data are conditionally non-normal and heteroscedastic. J. Educational Statist. 9, 129–150.

Rausch JR, Kelley K. (2009). A comparison of linear and mixture models for discriminant analysis under nonnormality. *Behavior Research Methods, 41*(1), 85–98.

Vallejo G, Fernandez P, Herrero FJ & Conejo, N. M. (2004). Alternative procedures for testing fixed effects in repeated measures designs when assumptions are violated, *Psicothema, 16*(3), 498–508.

Vallejo, G., Gras, J. A. & Garcia, M. A. (2007). Comparative robustness of recent methods for analyzing multivariate repeated measures designs. *Educational and Psychological Measurement, 67*(3), 410–432.

Wang ZM, Thompson B. (2007). Is the Pearson r(2) biased, and if so, what is the best correction formula? *Journal of experimental Education, 75*(2), 109–125.

Weathers D, Sharma S, Niedrich RW. (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*(11), 1516–1524.

Tables below contain three columns in each cell: First column = coverage rate (nominal values .68, .80,. 90,. 95, and .99); second column = coverage balance (nominal value = 1); third column = mean width of the 50,000 sample confidence intervals in each condition. This is for comparison with the mean width of the exact confidence interval. I might change this to the median, since some of the confidence interval limit distributions are quite skewed.

Each cell has five rows, moving from the 68% confidence level (first row) to the 99% confidence level (fifth row).

The order of the columns is as follows:

1) Exact CI

2) Fisher Z CI

3) Asymptotically Distribution-free CI

4) Asymptotically Distribution-free Fisher Z CI

5) BCa CI

6) Univariate Sampling CI (Rodgers approach)

For distribution 1(a) and 1b, rows of the tables stand for different size of ρ, top row: $\rho = -.8$, bottom row: $\rho = .8$. For distributions 2a – 41b2b, rows of the tables progress with sample size. Top row: $N = 20$, second row $N = 50$, and so on, last row, $N = 500$.

Table 62: Coverage Rate, Coverage Balance, and Median Width for Cis on Distribution 1a

	Distribution 1a: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 2.75$, $N = 50$																	
ρ	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
−.8	.683	0.995	0.11	.660	0.771	0.10	.626	0.680	0.10	.639	0.909	0.10	.642	1.058	0.10	.753	0.711	0.12
	.802	0.989	0.14	.780	0.710	0.13	.744	0.579	0.12	.759	0.960	0.13	.762	1.074	0.13	.869	0.696	0.16
	.902	0.987	0.19	.884	0.623	0.17	.846	0.431	0.16	.865	1.124	0.16	.867	1.161	0.17	.949	0.682	0.21
	.951	0.977	0.22	.936	0.535	0.21	.904	0.314	0.19	.922	1.271	0.20	.923	1.184	0.20	.980	0.649	0.26
	.989	0.949	0.30	.983	0.349	0.28	.961	0.136	0.25	.975	1.600	0.26	.976	1.042	0.27	.997	0.473	0.35
−.4	.676	0.976	0.24	.675	0.849	0.24	.628	0.829	0.22	.639	0.953	0.22	.648	1.024	0.23	.708	0.817	0.26
	.798	0.983	0.31	.795	0.818	0.31	.743	0.779	0.28	.758	1.000	0.28	.768	1.044	0.29	.830	0.823	0.34
	.899	0.957	0.40	.894	0.742	0.40	.847	0.694	0.36	.862	1.075	0.36	.871	1.075	0.37	.923	0.838	0.44
	.948	0.964	0.47	.943	0.680	0.47	.903	0.607	0.43	.920	1.151	0.43	.924	1.067	0.44	.967	0.809	0.53
	.989	0.790	0.62	.987	0.448	0.61	.964	0.412	0.56	.976	1.318	0.56	.978	0.988	0.58	.995	0.779	0.70
0	.684	1.001	0.28	.687	1.001	0.29	.631	1.002	0.26	.641	1.003	0.26	.656	1.000	0.26	.692	1.003	0.29
	.803	1.023	0.36	.803	1.018	0.37	.748	1.018	0.33	.761	1.022	0.33	.774	1.019	0.34	.816	1.010	0.38
	.900	1.009	0.46	.899	0.984	0.47	.850	1.013	0.42	.865	1.019	0.42	.875	1.000	0.44	.914	0.989	0.49
	.951	1.008	0.54	.949	0.987	0.55	.906	1.005	0.50	.922	0.989	0.50	.927	0.986	0.52	.963	0.986	0.59
	.990	0.951	0.70	.989	0.937	0.71	.966	0.987	0.66	.976	0.957	0.65	.978	0.926	0.68	.995	0.877	0.77
.4	.676	1.018	0.24	.673	1.149	0.24	.622	1.169	0.22	.632	1.021	0.22	.643	0.955	0.23	.706	1.202	0.26
	.796	1.006	0.31	.792	1.209	0.31	.739	1.271	0.28	.753	0.984	0.28	.764	0.943	0.29	.828	1.202	0.34
	.899	0.995	0.40	.894	1.294	0.40	.847	1.428	0.36	.863	0.917	0.36	.871	0.922	0.37	.925	1.155	0.44
	.950	0.990	0.47	.945	1.448	0.47	.905	1.614	0.43	.921	0.805	0.43	.927	0.892	0.44	.968	1.145	0.53
	.990	1.168	0.62	.988	2.034	0.61	.965	2.093	0.56	.976	0.565	0.56	.979	0.801	0.58	.996	0.935	0.69
.8	.682	1.014	0.11	.656	1.313	0.10	.623	1.467	0.10	.636	1.100	0.10	.639	0.952	0.10	.750	1.416	0.12
	.803	1.017	0.14	.777	1.429	0.13	.742	1.728	0.12	.757	1.025	0.13	.759	0.923	0.13	.866	1.455	0.16
	.901	1.011	0.19	.880	1.614	0.17	.844	2.293	0.16	.863	0.916	0.16	.863	0.892	0.17	.948	1.480	0.21
	.950	1.029	0.22	.934	1.899	0.21	.901	3.252	0.19	.920	0.792	0.20	.921	0.876	0.20	.980	1.629	0.26
	.989	1.094	0.31	.982	2.956	0.28	.959	6.597	0.25	.974	0.576	0.26	.975	0.865	0.27	.997	2.190	0.35

Table 63: *Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 1b*

Distribution 1b: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 2.75$, $N = 50$

ρ	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
−.8	.680	1.009	0.13	.575	0.609	0.10	.602	0.379	0.10	.608	0.531	0.10	.624	0.611	0.11	.488	0.522	0.08
	.801	1.017	0.16	.694	0.588	0.13	.715	0.250	0.13	.730	0.471	0.13	.747	0.542	0.14	.606	0.447	0.10
	.899	1.013	0.21	.810	0.588	0.17	.813	0.126	0.17	.839	0.402	0.17	.851	0.468	0.18	.728	0.347	0.14
	.950	1.044	0.26	.878	0.577	0.20	.866	0.058	0.20	.904	0.339	0.21	.913	0.400	0.22	.808	0.260	0.16
	.990	1.081	0.36	.955	0.619	0.27	.926	0.027	0.26	.967	0.234	0.28	.971	0.258	0.29	.907	0.109	0.22
−.4	.676	1.007	0.26	.619	0.716	0.24	.623	0.581	0.24	.636	0.670	0.24	.647	0.729	0.24	.603	0.624	0.23
	.800	1.003	0.34	.740	0.685	0.31	.741	0.474	0.31	.757	0.610	0.31	.770	0.668	0.31	.728	0.545	0.30
	.899	0.991	0.43	.851	0.666	0.39	.843	0.348	0.40	.866	0.558	0.40	.878	0.592	0.40	.841	0.442	0.39
	.948	1.023	0.51	.912	0.637	0.47	.900	0.252	0.47	.923	0.496	0.47	.932	0.522	0.48	.904	0.360	0.47
	.990	1.004	0.67	.973	0.563	0.61	.957	0.122	0.62	.978	0.387	0.61	.982	0.392	0.63	.967	0.206	0.62
0	.677	0.995	0.28	.680	0.990	0.29	.657	0.993	0.28	.667	0.991	0.28	.678	0.984	0.28	.686	0.990	0.29
	.798	0.983	0.36	.799	0.983	0.37	.773	0.975	0.36	.787	0.964	0.36	.799	0.969	0.36	.808	0.986	0.38
	.899	0.963	0.47	.898	0.981	0.47	.873	0.979	0.46	.888	0.996	0.46	.900	0.992	0.46	.907	0.986	0.49
	.950	1.010	0.57	.950	0.996	0.55	.926	0.998	0.55	.941	0.991	0.54	.951	0.978	0.55	.956	1.000	0.58
	.989	1.255	0.76	.989	1.237	0.71	.976	1.065	0.72	.985	1.108	0.70	.990	1.115	0.71	.990	1.174	0.76
.4	.681	1.012	0.26	.621	1.412	0.24	.625	1.731	0.24	.637	1.494	0.24	.649	1.391	0.24	.606	1.613	0.23
	.801	0.988	0.34	.738	1.444	0.31	.740	2.115	0.31	.758	1.631	0.31	.771	1.509	0.31	.727	1.820	0.30
	.899	1.010	0.43	.850	1.481	0.39	.842	2.832	0.40	.865	1.785	0.40	.876	1.674	0.40	.839	2.224	0.39
	.950	1.001	0.51	.912	1.553	0.47	.899	4.018	0.47	.921	2.043	0.47	.931	1.908	0.48	.904	2.850	0.47
	.990	0.884	0.67	.974	1.849	0.61	.957	8.179	0.62	.977	2.666	0.61	.981	2.668	0.63	.966	4.821	0.63
.8	.680	0.977	0.13	.575	1.638	0.10	.602	2.586	0.10	.607	1.840	0.10	.624	1.611	0.11	.490	1.899	0.08
	.798	0.979	0.16	.694	1.673	0.13	.716	3.947	0.13	.729	2.090	0.13	.744	1.814	0.14	.606	2.180	0.10
	.900	0.967	0.22	.807	1.708	0.17	.811	7.615	0.17	.840	2.496	0.17	.851	2.127	0.18	.726	2.760	0.14
	.949	0.941	0.26	.876	1.659	0.20	.865	15.432	0.20	.902	2.863	0.21	.912	2.421	0.22	.807	3.656	0.16
	.989	1.023	0.36	.953	1.502	0.27	.925	35.194	0.26	.967	3.464	0.28	.970	3.165	0.29	.906	8.363	0.22

Table 64: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 2a, $\rho = -.8$

Distribution 2a: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = -.8$

N	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	.681	0.978	0.26	.498	0.483	0.16	.444	0.514	0.14	.482	0.671	0.14	.556	1.052	0.18	.738	0.355	0.25
	.800	1.000	0.35	.610	0.400	0.20	.540	0.445	0.18	.598	0.719	0.19	.677	1.154	0.24	.875	0.252	0.34
	.898	0.937	0.45	.728	0.291	0.27	.636	0.372	0.22	.716	0.815	0.25	.795	1.284	0.31	.960	0.141	0.46
	.950	0.914	0.55	.801	0.187	0.32	.703	0.312	0.26	.795	0.960	0.31	.867	1.317	0.37	.988	0.052	0.58
	.990	0.885	0.75	.894	0.057	0.44	.795	0.219	0.33	.893	1.437	0.43	.947	1.210	0.48	.999	0.000	0.81
50	.679	0.987	0.18	.446	0.627	0.10	.506	0.615	0.11	.531	0.803	0.12	.560	1.072	0.13	.752	0.564	0.18
	.799	0.957	0.24	.555	0.563	0.13	.613	0.555	0.15	.652	0.882	0.15	.681	1.145	0.17	.875	0.542	0.24
	.899	0.940	0.31	.672	0.470	0.17	.717	0.482	0.19	.769	1.062	0.20	.795	1.282	0.21	.958	0.521	0.32
	.949	0.950	0.38	.752	0.377	0.20	.785	0.415	0.22	.842	1.335	0.24	.865	1.399	0.25	.986	0.482	0.39
	.990	0.913	0.53	.859	0.219	0.27	.873	0.302	0.29	.924	2.206	0.33	.943	1.613	0.32	.999	0.464	0.52
100	.676	1.034	0.13	.412	0.728	0.07	.542	0.701	0.10	.560	0.890	0.10	.567	1.070	0.10	.758	0.728	0.14
	.799	1.031	0.18	.516	0.676	0.09	.656	0.644	0.12	.684	0.994	0.13	.688	1.138	0.13	.878	0.769	0.19
	.900	1.028	0.23	.630	0.610	0.12	.767	0.567	0.16	.802	1.255	0.16	.803	1.274	0.17	.959	0.827	0.24
	.951	1.079	0.28	.714	0.534	0.14	.836	0.499	0.19	.872	1.647	0.20	.871	1.407	0.20	.986	0.926	0.29
	.990	1.203	0.40	.836	0.371	0.19	.919	0.388	0.25	.945	2.847	0.26	.946	1.525	0.25	.999	0.889	0.39
200	.676	0.984	0.10	.392	0.811	0.05	.578	0.774	0.08	.590	0.960	0.08	.581	1.082	0.08	.767	0.825	0.11
	.796	0.994	0.13	.492	0.772	0.07	.696	0.717	0.10	.715	1.067	0.10	.701	1.111	0.10	.882	0.914	0.15
	.897	0.981	0.17	.603	0.707	0.08	.806	0.651	0.13	.828	1.354	0.13	.812	1.196	0.13	.959	1.042	0.19
	.948	0.961	0.21	.684	0.646	0.10	.873	0.586	0.15	.894	1.797	0.16	.878	1.306	0.16	.985	1.345	0.22
	.990	1.075	0.29	.807	0.520	0.13	.948	0.459	0.20	.960	3.122	0.21	.951	1.441	0.20	.999	3.333	0.29
500	.678	1.016	0.07	.371	0.880	0.03	.615	0.862	0.06	.623	1.027	0.06	.604	1.076	0.06	.784	0.952	0.08
	.798	1.007	0.09	.466	0.854	0.04	.737	0.849	0.07	.747	1.175	0.07	.722	1.126	0.07	.892	1.090	0.10
	.899	1.006	0.11	.575	0.815	0.05	.849	0.789	0.09	.857	1.457	0.09	.831	1.212	0.09	.962	1.277	0.13
	.949	1.017	0.14	.656	0.774	0.06	.909	0.728	0.11	.918	1.846	0.11	.893	1.230	0.11	.986	1.411	0.16
	.990	0.950	0.19	.786	0.667	0.08	.971	0.591	0.15	.972	3.173	0.15	.959	1.345	0.15	.999	2.000	0.20

Table 65: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 2a, $\rho = -.4$

Distribution 2a: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = -.4$																		
N	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	N/A			N/A			N/A			N/A			N/A			N/A		
50	.683	1.021	0.28	.583	0.680	0.24	.520	0.982	0.21	.540	1.113	0.21	.578	1.320	0.23	.707	0.630	0.30
	.802	1.028	0.36	.703	0.530	0.31	.631	1.035	0.27	.660	1.324	0.27	.698	1.487	0.29	.832	0.541	0.39
	.901	1.035	0.46	.815	0.329	0.40	.739	1.123	0.35	.774	1.813	0.35	.812	1.746	0.38	.932	0.407	0.51
	.950	0.994	0.55	.880	0.176	0.47	.810	1.206	0.41	.845	2.565	0.41	.880	1.999	0.45	.975	0.299	0.62
	.990	0.973	0.71	.941	0.037	0.61	.899	1.422	0.54	.921	5.278	0.54	.952	2.397	0.59	.998	0.304	0.83
100	.679	0.982	0.22	.550	0.760	0.17	.558	1.033	0.17	.572	1.146	0.17	.584	1.226	0.18	.700	0.692	0.22
	.798	0.976	0.28	.670	0.646	0.22	.673	1.118	0.22	.692	1.391	0.22	.703	1.327	0.23	.822	0.607	0.29
	.898	0.968	0.36	.784	0.474	0.28	.783	1.286	0.28	.806	1.978	0.28	.815	1.567	0.29	.922	0.514	0.38
	.948	1.021	0.43	.855	0.320	0.33	.851	1.516	0.33	.869	2.869	0.33	.880	1.805	0.35	.967	0.435	0.45
	.989	1.116	0.56	.933	0.114	0.43	.928	2.030	0.44	.938	6.416	0.44	.952	2.320	0.45	.997	0.237	0.60
200	.679	1.026	0.16	.525	0.861	0.12	.589	1.129	0.13	.598	1.248	0.13	.595	1.236	0.14	.703	0.781	0.17
	.800	1.029	0.21	.643	0.768	0.15	.709	1.266	0.17	.721	1.538	0.17	.712	1.343	0.18	.824	0.703	0.22
	.899	1.032	0.27	.762	0.619	0.20	.821	1.581	0.22	.834	2.305	0.22	.825	1.530	0.22	.922	0.617	0.28
	.950	1.005	0.32	.839	0.482	0.23	.885	1.935	0.26	.893	3.376	0.26	.889	1.747	0.27	.966	0.513	0.33
	.991	1.057	0.43	.929	0.221	0.31	.954	3.190	0.35	.956	9.282	0.34	.958	2.280	0.35	.997	0.315	0.44
500	.682	1.010	0.11	.500	0.901	0.08	.626	1.137	0.10	.631	1.224	0.10	.616	1.147	0.10	.708	0.831	0.11
	.803	1.001	0.14	.617	0.853	0.10	.748	1.285	0.12	.753	1.468	0.12	.736	1.206	0.12	.827	0.772	0.15
	.901	1.034	0.18	.738	0.743	0.12	.854	1.562	0.16	.860	2.064	0.16	.840	1.313	0.16	.920	0.669	0.19
	.952	1.015	0.22	.817	0.638	0.15	.913	2.086	0.19	.916	3.192	0.19	.902	1.494	0.19	.965	0.553	0.22
	.990	1.008	0.29	.917	0.390	0.19	.972	3.672	0.25	.972	9.150	0.25	.966	1.769	0.24	.995	0.384	0.29

Table 66: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 2a, $\rho = 0$

	Distribution 2a: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = 0$																	
N	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	N/A			N/A			N/A			N/A			N/A			N/A		
50	.683	1.013	0.25	.728	1.013	0.29	.566	0.995	0.20	.577	0.997	0.20	.627	0.986	0.22	.711	1.015	0.29
	.800	1.025	0.33	.825	1.016	0.37	.695	0.991	0.26	.711	0.995	0.26	.752	1.000	0.29	.836	1.019	0.38
	.901	1.015	0.42	.902	1.040	0.47	.817	1.008	0.33	.836	1.006	0.33	.859	0.993	0.38	.929	1.049	0.50
	.951	1.063	0.50	.941	1.054	0.55	.891	1.002	0.40	.909	0.996	0.39	.917	0.997	0.46	.969	1.063	0.62
	.991	0.946	0.65	.979	1.098	0.71	.963	0.951	0.52	.974	0.923	0.51	.973	1.029	0.63	.996	0.713	0.84
100	.680	1.021	0.18	.723	1.013	0.20	.598	1.015	0.15	.603	1.016	0.15	.615	1.010	0.16	.700	1.014	0.20
	.799	1.044	0.24	.826	1.045	0.26	.728	1.012	0.20	.736	1.004	0.20	.735	1.017	0.21	.824	1.041	0.26
	.900	1.072	0.30	.905	1.048	0.33	.852	1.020	0.25	.861	1.017	0.25	.846	1.040	0.27	.921	1.064	0.34
	.950	1.117	0.36	.944	1.113	0.39	.919	1.073	0.30	.927	1.075	0.30	.908	1.045	0.33	.963	1.110	0.42
	.990	1.225	0.48	.981	1.155	0.51	.978	1.081	0.39	.982	1.057	0.39	.968	1.103	0.44	.993	1.191	0.58
200	.681	1.023	0.13	.714	1.036	0.14	.619	1.013	0.11	.622	1.013	0.11	.613	1.005	0.12	.691	1.030	0.14
	.800	1.037	0.17	.819	1.026	0.18	.750	1.003	0.15	.754	1.005	0.15	.731	1.012	0.15	.813	1.025	0.18
	.899	1.005	0.22	.904	1.014	0.23	.870	0.998	0.19	.875	0.995	0.19	.843	1.009	0.20	.913	1.007	0.24
	.948	0.960	0.26	.945	1.011	0.28	.932	1.041	0.22	.936	1.046	0.22	.904	1.022	0.24	.958	0.961	0.29
	.990	0.930	0.35	.982	0.991	0.36	.985	1.027	0.30	.987	1.037	0.29	.966	0.985	0.32	.992	0.939	0.40
500	.688	0.998	0.09	.705	1.003	0.09	.651	0.996	0.08	.652	0.998	0.08	.631	1.002	0.08	.690	1.000	0.09
	.805	0.964	0.11	.817	0.998	0.12	.777	0.998	0.10	.779	1.001	0.10	.748	1.002	0.10	.811	0.991	0.11
	.902	0.968	0.14	.905	0.992	0.15	.887	1.008	0.13	.889	1.004	0.13	.852	1.001	0.13	.908	1.012	0.15
	.952	1.030	0.17	.949	1.015	0.18	.944	0.994	0.15	.946	0.990	0.15	.912	0.997	0.16	.956	1.026	0.18
	.990	0.996	0.23	.986	0.964	0.23	.989	1.011	0.20	.990	1.004	0.20	.971	0.999	0.21	.992	1.035	0.25

Table 67: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 2a, $\rho = .4$

Distribution 2a: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = .4$	

N	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	N/A			N/A			N/A			N/A			N/A			N/A		
50	.679	1.002	0.28	.582	1.505	0.24	.523	1.060	0.21	.542	0.936	0.21	.581	0.789	0.23	.705	1.616	0.30
	.799	0.991	0.36	.701	1.916	0.31	.630	0.995	0.27	.660	0.777	0.27	.700	0.699	0.30	.833	1.955	0.39
	.902	0.986	0.46	.814	3.172	0.40	.741	0.913	0.35	.777	0.560	0.35	.813	0.579	0.38	.932	2.556	0.52
	.950	1.014	0.55	.878	5.795	0.47	.811	0.823	0.41	.845	0.398	0.42	.879	0.494	0.46	.975	3.922	0.62
	.991	1.238	0.71	.943	35.909	0.61	.901	0.710	0.54	.922	0.204	0.54	.952	0.421	0.59	.998	4.474	0.83
100	.678	1.009	0.22	.551	1.295	0.17	.554	0.952	0.17	.567	0.854	0.17	.580	0.804	0.18	.699	1.431	0.22
	.798	1.003	0.28	.667	1.530	0.22	.670	0.868	0.22	.689	0.696	0.22	.701	0.718	0.23	.823	1.599	0.29
	.900	0.980	0.36	.784	2.106	0.28	.782	0.749	0.28	.804	0.477	0.28	.813	0.609	0.29	.925	1.895	0.38
	.950	0.993	0.43	.856	3.116	0.33	.851	0.634	0.33	.870	0.325	0.33	.881	0.527	0.35	.969	2.106	0.45
	.990	0.864	0.56	.936	9.304	0.43	.929	0.457	0.44	.938	0.130	0.43	.951	0.417	0.45	.997	3.735	0.60
200	.681	1.005	0.16	.525	1.197	0.12	.590	0.912	0.13	.599	0.833	0.13	.595	0.839	0.14	.701	1.315	0.17
	.798	1.032	0.21	.644	1.350	0.15	.710	0.818	0.17	.721	0.680	0.17	.714	0.777	0.18	.821	1.469	0.22
	.900	0.989	0.27	.760	1.657	0.20	.819	0.688	0.22	.830	0.486	0.22	.823	0.695	0.22	.921	1.672	0.28
	.951	0.977	0.32	.836	2.149	0.23	.883	0.569	0.26	.893	0.323	0.26	.888	0.621	0.27	.966	1.982	0.33
	.990	1.008	0.43	.927	4.653	0.31	.953	0.354	0.35	.955	0.119	0.34	.957	0.468	0.35	.996	3.109	0.44
500	.677	0.981	0.11	.499	1.086	0.08	.623	0.872	0.10	.626	0.818	0.10	.612	0.866	0.10	.702	1.185	0.11
	.797	0.990	0.14	.614	1.165	0.10	.742	0.777	0.12	.749	0.673	0.12	.730	0.815	0.12	.822	1.284	0.15
	.898	0.994	0.18	.734	1.317	0.12	.851	0.610	0.16	.857	0.465	0.16	.836	0.746	0.16	.919	1.480	0.19
	.949	0.987	0.22	.812	1.548	0.15	.912	0.451	0.19	.915	0.304	0.19	.900	0.645	0.19	.964	1.729	0.22
	.989	0.952	0.29	.913	2.441	0.19	.970	0.265	0.25	.969	0.113	0.25	.964	0.513	0.24	.995	2.306	0.29

Table 68: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 2a, $\rho = .8$

<div align="center">

Distribution 2a: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = .8$

</div>

N	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	.680	1.019	0.26	.497	2.077	0.16	.446	1.975	0.14	.484	1.506	0.15	.557	0.962	0.18	.737	2.822	0.25
	.801	1.028	0.34	.608	2.542	0.20	.539	2.272	0.18	.599	1.407	0.19	.677	0.884	0.24	.874	3.856	0.34
	.900	1.059	0.45	.724	3.553	0.27	.637	2.727	0.22	.717	1.212	0.25	.796	0.785	0.31	.960	6.878	0.46
	.950	1.068	0.55	.799	5.256	0.32	.702	3.203	0.26	.795	1.041	0.31	.867	0.749	0.37	.987	14.119	0.58
	.990	1.051	0.74	.892	15.538	0.44	.793	4.327	0.33	.891	0.697	0.43	.945	0.781	0.48	.999	37.000	0.81
50	.684	0.989	0.18	.453	1.572	0.10	.511	1.614	0.11	.535	1.239	0.12	.562	0.942	0.13	.751	1.751	0.18
	.799	0.988	0.24	.558	1.741	0.13	.615	1.787	0.15	.652	1.124	0.15	.680	0.879	0.17	.876	1.765	0.24
	.900	1.028	0.31	.670	2.065	0.17	.719	2.087	0.19	.770	0.922	0.20	.796	0.785	0.21	.960	1.932	0.32
	.950	1.067	0.38	.752	2.561	0.20	.787	2.373	0.22	.843	0.746	0.24	.867	0.708	0.25	.988	1.966	0.39
	.989	1.145	0.53	.862	4.639	0.27	.874	3.223	0.29	.925	0.412	0.33	.944	0.599	0.32	.999	3.100	0.52
100	.678	1.015	0.13	.414	1.380	0.07	.540	1.446	0.10	.558	1.132	0.10	.564	0.948	0.10	.758	1.427	0.14
	.796	1.013	0.18	.515	1.494	0.09	.653	1.562	0.12	.681	1.009	0.13	.684	0.886	0.13	.880	1.352	0.19
	.898	0.986	0.23	.628	1.676	0.12	.764	1.796	0.16	.801	0.796	0.16	.801	0.800	0.17	.958	1.211	0.24
	.950	1.036	0.28	.712	1.898	0.14	.836	2.038	0.19	.870	0.626	0.20	.870	0.721	0.20	.986	0.950	0.29
	.990	0.841	0.40	.833	2.723	0.19	.919	2.637	0.25	.945	0.364	0.26	.946	0.677	0.25	.999	0.762	0.39
200	.682	0.986	0.10	.395	1.221	0.05	.580	1.270	0.08	.593	1.015	0.08	.584	0.916	0.08	.773	1.181	0.11
	.801	0.992	0.13	.494	1.290	0.07	.702	1.370	0.10	.718	0.903	0.10	.703	0.867	0.10	.886	1.090	0.15
	.901	0.985	0.17	.607	1.390	0.08	.811	1.510	0.13	.832	0.724	0.13	.815	0.816	0.13	.960	0.936	0.19
	.951	0.985	0.21	.689	1.529	0.10	.878	1.729	0.15	.896	0.539	0.16	.880	0.745	0.16	.986	0.789	0.22
	.990	1.068	0.29	.813	1.937	0.13	.950	2.101	0.20	.961	0.330	0.21	.953	0.673	0.20	.999	0.395	0.29
500	.680	0.995	0.07	.375	1.158	0.03	.617	1.182	0.06	.625	0.992	0.06	.605	0.950	0.06	.785	1.057	0.08
	.801	0.986	0.09	.468	1.198	0.04	.739	1.220	0.07	.749	0.872	0.07	.724	0.909	0.07	.894	0.940	0.10
	.901	0.994	0.11	.576	1.249	0.05	.850	1.322	0.09	.860	0.703	0.09	.834	0.845	0.09	.963	0.784	0.13
	.950	1.005	0.14	.658	1.323	0.06	.911	1.366	0.11	.919	0.551	0.11	.895	0.806	0.11	.987	0.611	0.16
	.990	1.031	0.19	.786	1.497	0.08	.971	1.831	0.15	.975	0.293	0.15	.961	0.749	0.15	.999	0.462	0.20

Table 69: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 2b, $\rho = -.8$

	Distribution 2b: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = -.8$																	
N	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	.680	1.014	0.22	.551	0.477	0.15	.527	0.299	0.14	.544	0.453	0.15	.586	0.467	0.16	.498	0.366	0.13
	.801	0.994	0.30	.667	0.438	0.20	.628	0.192	0.18	.661	0.400	0.19	.706	0.389	0.21	.624	0.277	0.18
	.900	1.039	0.39	.780	0.400	0.26	.718	0.106	0.23	.773	0.341	0.25	.817	0.293	0.28	.748	0.151	0.24
	.950	1.042	0.48	.852	0.357	0.32	.770	0.065	0.27	.846	0.300	0.31	.882	0.216	0.34	.828	0.057	0.30
	.991	0.966	0.69	.938	0.238	0.44	.838	0.031	0.35	.929	0.251	0.43	.952	0.109	0.48	.914	0.007	0.43
50	.678	0.997	0.16	.488	0.600	0.10	.540	0.396	0.11	.547	0.539	0.11	.572	0.600	0.11	.474	0.547	0.09
	.798	0.999	0.20	.602	0.579	0.13	.655	0.270	0.14	.671	0.485	0.14	.694	0.528	0.15	.589	0.498	0.12
	.900	1.017	0.27	.718	0.551	0.16	.757	0.154	0.17	.791	0.425	0.18	.807	0.444	0.19	.714	0.426	0.15
	.950	1.002	0.34	.796	0.519	0.20	.814	0.098	0.21	.865	0.393	0.22	.875	0.363	0.23	.799	0.320	0.19
	.990	0.879	0.52	.897	0.420	0.26	.887	0.053	0.27	.949	0.389	0.30	.947	0.221	0.30	.900	0.145	0.25
100	.677	0.987	0.12	.449	0.705	0.07	.556	0.487	0.09	.559	0.619	0.09	.564	0.710	0.09	.491	0.689	0.07
	.797	0.986	0.16	.555	0.678	0.09	.672	0.352	0.11	.681	0.568	0.11	.684	0.653	0.12	.607	0.672	0.09
	.898	1.066	0.22	.669	0.657	0.12	.779	0.216	0.14	.805	0.530	0.15	.799	0.575	0.15	.727	0.642	0.12
	.950	1.067	0.27	.751	0.629	0.14	.843	0.139	0.17	.879	0.499	0.18	.867	0.499	0.18	.807	0.577	0.15
	.991	1.178	0.43	.859	0.561	0.19	.918	0.067	0.22	.958	0.515	0.24	.944	0.350	0.23	.904	0.405	0.19
200	.682	1.013	0.10	.417	0.812	0.05	.578	0.601	0.07	.581	0.744	0.07	.570	0.870	0.07	.523	0.847	0.06
	.801	1.020	0.13	.519	0.792	0.07	.701	0.460	0.09	.707	0.701	0.09	.689	0.822	0.10	.643	0.885	0.08
	.901	1.012	0.17	.629	0.766	0.08	.813	0.281	0.12	.828	0.646	0.12	.801	0.723	0.12	.762	0.939	0.10
	.951	0.976	0.22	.710	0.731	0.10	.876	0.179	0.14	.901	0.598	0.15	.870	0.638	0.15	.835	0.919	0.12
	.990	0.887	0.34	.827	0.642	0.13	.945	0.069	0.19	.970	0.584	0.19	.945	0.433	0.19	.924	0.772	0.16
500	.681	0.997	0.07	.374	0.906	0.03	.600	0.731	0.06	.603	0.869	0.06	.575	1.004	0.06	.576	1.040	0.05
	.800	0.963	0.09	.470	0.886	0.04	.730	0.605	0.07	.736	0.865	0.07	.697	0.961	0.07	.702	1.179	0.07
	.899	0.925	0.12	.576	0.861	0.05	.846	0.432	0.09	.857	0.839	0.09	.810	0.898	0.09	.817	1.406	0.08
	.950	0.938	0.15	.656	0.828	0.06	.910	0.276	0.11	.923	0.841	0.11	.876	0.800	0.11	.882	1.567	0.10
	.990	0.981	0.23	.778	0.722	0.08	.970	0.116	0.14	.980	0.738	0.14	.949	0.566	0.14	.952	1.685	0.13

Table 70: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 2b, $\rho = -.4$

	Distribution 2b: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = -.4$																	
N	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	N/A			N/A			N/A			N/A			N/A			N/A		
50	.675	0.961	0.31	.525	0.630	0.24	.554	0.515	0.25	.566	0.583	0.25	.592	0.612	0.25	.534	0.565	0.24
	.797	0.977	0.39	.643	0.596	0.30	.667	0.426	0.32	.689	0.534	0.32	.715	0.538	0.33	.655	0.504	0.31
	.900	0.951	0.49	.761	0.543	0.39	.774	0.324	0.41	.807	0.478	0.41	.830	0.439	0.42	.776	0.420	0.40
	.949	0.991	0.58	.842	0.482	0.46	.841	0.249	0.49	.877	0.457	0.49	.897	0.357	0.50	.854	0.345	0.48
	.990	1.098	0.71	.934	0.327	0.60	.918	0.160	0.64	.957	0.474	0.63	.964	0.209	0.65	.944	0.214	0.64
100	.685	0.996	0.24	.502	0.724	0.17	.575	0.588	0.19	.582	0.659	0.19	.594	0.721	0.20	.529	0.656	0.17
	.801	0.999	0.31	.615	0.685	0.22	.692	0.497	0.25	.705	0.604	0.25	.717	0.643	0.25	.647	0.606	0.23
	.902	0.960	0.39	.734	0.654	0.28	.803	0.372	0.32	.823	0.541	0.32	.829	0.543	0.32	.768	0.549	0.29
	.951	0.924	0.46	.813	0.589	0.33	.868	0.291	0.38	.892	0.506	0.38	.894	0.442	0.39	.848	0.478	0.35
	.990	0.959	0.59	.915	0.432	0.43	.941	0.174	0.50	.963	0.484	0.49	.961	0.265	0.50	.942	0.366	0.45
200	.679	0.987	0.19	.470	0.802	0.12	.588	0.677	0.15	.593	0.742	0.15	.589	0.838	0.15	.523	0.748	0.13
	.800	0.979	0.24	.580	0.781	0.15	.709	0.580	0.20	.718	0.687	0.20	.711	0.773	0.20	.640	0.708	0.17
	.901	1.012	0.31	.696	0.732	0.20	.821	0.460	0.25	.834	0.635	0.25	.823	0.668	0.25	.762	0.649	0.21
	.949	1.037	0.37	.778	0.672	0.23	.886	0.372	0.30	.902	0.600	0.30	.888	0.553	0.30	.844	0.600	0.26
	.990	1.139	0.49	.886	0.559	0.30	.957	0.217	0.39	.968	0.540	0.39	.956	0.371	0.39	.939	0.519	0.33
500	.679	0.983	0.13	.434	0.904	0.08	.609	0.811	0.11	.612	0.871	0.11	.594	0.983	0.11	.528	0.852	0.09
	.800	0.992	0.17	.542	0.882	0.10	.735	0.725	0.14	.740	0.836	0.14	.717	0.943	0.14	.648	0.811	0.12
	.900	1.035	0.22	.657	0.829	0.12	.848	0.625	0.18	.857	0.824	0.18	.827	0.851	0.18	.768	0.752	0.15
	.950	1.026	0.26	.737	0.781	0.15	.912	0.548	0.21	.920	0.812	0.21	.891	0.785	0.21	.848	0.709	0.18
	.989	0.918	0.36	.852	0.673	0.19	.973	0.353	0.28	.977	0.738	0.28	.958	0.542	0.28	.943	0.646	0.23

Table 71: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 2b, $\rho = 0$

Distribution 2b: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = 0$

N	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	N/A			N/A			N/A			N/A			N/A			N/A		
50	.681	0.996	0.29	.680	1.006	0.29	.636	1.011	0.27	.647	1.012	0.27	.671	1.011	0.28	.689	1.007	0.29
	.800	0.965	0.38	.801	0.993	0.37	.752	0.996	0.35	.766	1.004	0.35	.794	0.995	0.36	.815	0.980	0.38
	.900	0.966	0.50	.902	0.969	0.47	.853	0.989	0.45	.869	0.980	0.45	.899	0.970	0.46	.916	0.971	0.49
	.951	0.952	0.61	.953	0.952	0.55	.912	0.973	0.54	.926	0.973	0.53	.951	0.939	0.54	.962	0.947	0.59
	.990	0.917	0.81	.990	0.944	0.71	.969	0.971	0.71	.979	0.961	0.69	.992	1.029	0.70	.993	0.978	0.76
100	.683	0.986	0.21	.681	0.990	0.20	.648	0.983	0.19	.653	0.983	0.19	.670	0.985	0.20	.687	0.987	0.20
	.801	0.988	0.27	.802	0.989	0.26	.765	0.990	0.25	.771	0.994	0.25	.793	0.975	0.25	.811	0.981	0.26
	.900	0.986	0.36	.902	0.981	0.33	.866	0.991	0.32	.873	1.002	0.32	.896	0.993	0.32	.911	0.996	0.34
	.949	0.950	0.44	.950	0.948	0.39	.921	0.999	0.38	.928	0.990	0.38	.947	0.975	0.38	.958	0.970	0.40
	.990	1.080	0.60	.990	1.068	0.51	.974	0.977	0.50	.979	0.976	0.49	.989	1.073	0.50	.992	1.043	0.53
200	.688	1.000	0.15	.686	1.009	0.14	.659	1.009	0.14	.662	1.009	0.14	.676	1.010	0.14	.687	1.010	0.14
	.809	1.023	0.19	.807	1.026	0.18	.777	1.029	0.18	.781	1.028	0.18	.796	1.033	0.18	.810	1.031	0.18
	.905	1.064	0.25	.904	1.064	0.23	.877	1.031	0.23	.881	1.027	0.23	.897	1.038	0.23	.911	1.046	0.24
	.954	1.079	0.31	.954	1.069	0.28	.931	1.030	0.27	.935	1.031	0.27	.948	1.022	0.27	.959	1.040	0.28
	.991	0.975	0.42	.990	0.933	0.36	.980	1.058	0.35	.982	1.011	0.35	.989	0.967	0.35	.992	0.928	0.37
500	.682	0.999	0.09	.683	1.001	0.09	.665	1.004	0.09	.666	1.004	0.09	.674	1.001	0.09	.682	1.000	0.09
	.805	0.992	0.12	.806	0.989	0.12	.786	0.992	0.11	.787	0.991	0.11	.797	0.986	0.11	.807	0.994	0.12
	.903	0.995	0.16	.904	0.979	0.15	.888	0.982	0.14	.890	0.985	0.14	.898	0.976	0.14	.906	0.970	0.15
	.952	0.969	0.19	.953	0.951	0.18	.940	0.960	0.17	.941	0.956	0.17	.948	0.928	0.17	.956	0.933	0.18
	.990	0.776	0.26	.990	0.776	0.23	.984	0.826	0.22	.985	0.839	0.22	.989	0.852	0.22	.992	0.771	0.23

Table 72: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 2b, $\rho = .4$

| | Distribution 2b: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = .4$ | | | | | | | | | | | | | | | | | |
N	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	N/A			N/A			N/A			N/A			N/A			N/A		
50	.683	0.982	0.31	.530	1.566	0.24	.559	1.905	0.25	.573	1.668	0.25	.597	1.596	0.25	.538	1.747	0.24
	.804	0.989	0.39	.650	1.622	0.30	.676	2.308	0.32	.697	1.819	0.32	.723	1.804	0.33	.662	1.941	0.31
	.903	0.993	0.49	.770	1.786	0.39	.781	2.974	0.41	.812	1.977	0.41	.836	2.167	0.42	.784	2.276	0.40
	.951	1.012	0.58	.847	1.981	0.46	.846	3.721	0.49	.883	2.024	0.49	.902	2.630	0.50	.860	2.699	0.48
	.990	1.004	0.71	.937	3.054	0.60	.924	6.206	0.64	.959	2.189	0.63	.967	4.771	0.65	.947	4.308	0.64
100	.679	0.991	0.24	.500	1.366	0.17	.573	1.662	0.19	.581	1.496	0.19	.591	1.356	0.20	.527	1.489	0.17
	.802	0.980	0.31	.614	1.417	0.22	.690	1.989	0.25	.702	1.619	0.25	.712	1.519	0.25	.646	1.604	0.23
	.901	0.980	0.39	.732	1.533	0.28	.800	2.586	0.32	.820	1.806	0.32	.826	1.811	0.32	.765	1.778	0.29
	.951	0.968	0.46	.814	1.628	0.33	.866	3.351	0.38	.891	1.936	0.38	.893	2.205	0.39	.848	2.001	0.35
	.991	1.036	0.59	.913	2.208	0.43	.941	5.719	0.50	.964	2.044	0.49	.962	3.564	0.50	.941	2.668	0.46
200	.679	1.021	0.19	.465	1.259	0.12	.584	1.505	0.15	.589	1.372	0.15	.586	1.216	0.15	.519	1.363	0.13
	.798	1.019	0.24	.578	1.314	0.15	.705	1.735	0.20	.714	1.459	0.20	.707	1.315	0.20	.637	1.455	0.17
	.900	1.017	0.31	.697	1.383	0.20	.820	2.164	0.25	.833	1.572	0.25	.824	1.500	0.25	.759	1.560	0.21
	.950	1.000	0.37	.779	1.458	0.23	.888	2.790	0.30	.905	1.656	0.30	.891	1.792	0.30	.843	1.692	0.26
	.990	0.793	0.49	.888	1.825	0.30	.958	4.106	0.39	.969	1.747	0.39	.958	2.718	0.39	.940	2.001	0.33
500	.683	1.003	0.13	.442	1.090	0.08	.616	1.210	0.11	.619	1.122	0.11	.601	1.005	0.11	.534	1.171	0.09
	.803	1.018	0.17	.545	1.124	0.10	.738	1.349	0.14	.743	1.165	0.14	.718	1.051	0.14	.653	1.237	0.12
	.903	1.036	0.22	.657	1.180	0.12	.851	1.597	0.18	.857	1.194	0.18	.829	1.151	0.18	.773	1.327	0.15
	.952	1.035	0.26	.740	1.268	0.15	.915	1.966	0.21	.922	1.301	0.21	.893	1.324	0.21	.850	1.436	0.18
	.990	1.094	0.37	.855	1.567	0.19	.974	3.141	0.28	.978	1.431	0.28	.959	1.836	0.28	.945	1.685	0.23

Table 73: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 2b, $\rho = .8$

Distribution 2b: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = .8$

N	Exact CI			Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	.680	0.998	0.22	.547	2.112	0.15	.525	3.354	0.14	.542	2.232	0.15	.586	2.166	0.16	.498	2.727	0.13
	.800	1.019	0.30	.663	2.330	0.20	.627	5.263	0.18	.658	2.573	0.19	.704	2.654	0.21	.624	3.687	0.18
	.899	1.034	0.39	.780	2.584	0.26	.714	9.503	0.23	.774	3.006	0.25	.816	3.518	0.27	.748	6.685	0.24
	.950	1.012	0.48	.851	2.889	0.32	.766	15.632	0.27	.845	3.375	0.31	.881	4.670	0.34	.825	15.472	0.30
	.989	1.019	0.68	.935	4.060	0.44	.837	32.159	0.34	.929	3.855	0.42	.952	8.882	0.48	.913	119.833	0.43
50	.678	0.982	0.16	.488	1.640	0.10	.542	2.519	0.11	.549	1.855	0.11	.571	1.654	0.11	.473	1.822	0.09
	.800	0.994	0.20	.598	1.723	0.13	.654	3.629	0.14	.670	2.039	0.14	.691	1.869	0.15	.590	1.977	0.12
	.900	0.990	0.27	.717	1.838	0.16	.754	6.143	0.17	.789	2.333	0.18	.804	2.265	0.19	.717	2.357	0.15
	.949	0.990	0.34	.795	1.961	0.20	.814	9.430	0.21	.863	2.467	0.22	.872	2.642	0.23	.798	3.012	0.19
	.990	1.016	0.52	.897	2.306	0.26	.886	18.324	0.27	.948	2.313	0.30	.947	4.244	0.30	.900	6.538	0.25
100	.680	0.975	0.12	.450	1.411	0.07	.558	2.040	0.09	.560	1.587	0.09	.566	1.377	0.09	.491	1.454	0.07
	.798	0.952	0.16	.558	1.445	0.09	.677	2.860	0.11	.684	1.743	0.11	.689	1.517	0.12	.604	1.484	0.09
	.901	0.953	0.21	.673	1.508	0.12	.783	4.883	0.14	.807	1.982	0.15	.802	1.782	0.15	.725	1.546	0.12
	.951	0.963	0.27	.751	1.554	0.14	.845	8.072	0.17	.881	2.136	0.18	.872	2.080	0.18	.803	1.756	0.14
	.989	1.063	0.43	.864	1.839	0.19	.917	19.547	0.22	.958	2.332	0.24	.945	3.279	0.23	.905	2.692	0.19
200	.684	0.989	0.10	.416	1.234	0.05	.580	1.696	0.07	.581	1.368	0.07	.571	1.160	0.07	.524	1.183	0.06
	.804	1.022	0.13	.520	1.267	0.07	.703	2.218	0.09	.709	1.470	0.09	.690	1.258	0.10	.645	1.130	0.08
	.903	1.001	0.17	.630	1.339	0.08	.813	3.564	0.12	.833	1.576	0.12	.804	1.413	0.12	.764	1.087	0.10
	.951	0.985	0.21	.713	1.400	0.10	.877	5.458	0.14	.900	1.674	0.15	.871	1.627	0.15	.838	1.107	0.12
	.990	0.888	0.34	.828	1.608	0.13	.945	12.730	0.19	.969	1.665	0.19	.945	2.241	0.19	.925	1.290	0.16
500	.678	0.982	0.07	.372	1.093	0.03	.600	1.318	0.06	.603	1.108	0.06	.576	0.973	0.06	.578	0.943	0.05
	.799	0.971	0.09	.466	1.112	0.04	.728	1.575	0.07	.733	1.112	0.07	.695	0.993	0.07	.700	0.825	0.07
	.898	0.943	0.12	.574	1.153	0.05	.845	2.211	0.09	.853	1.127	0.09	.806	1.070	0.09	.812	0.686	0.08
	.949	0.917	0.15	.653	1.189	0.06	.908	3.081	0.11	.919	1.135	0.11	.873	1.149	0.11	.878	0.597	0.10
	.990	0.985	0.23	.777	1.332	0.08	.968	6.402	0.14	.978	1.116	0.14	.947	1.532	0.14	.950	0.545	0.13

Table 74: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 41a2a, $\rho = 0$

N	Exact CI	Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
		Distribution 41a2a: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = 0$														
20	N/A	N/A			N/A			N/A			N/A			N/A		
50	N/A	.696	1.002	0.29	.597	0.734	0.23	.607	0.722	0.23	.640	0.762	0.25	.697	1.003	0.29
		.814	1.173	0.37	.714	0.666	0.30	.728	0.647	0.30	.760	0.735	0.32	.831	1.366	0.38
		.904	1.543	0.47	.822	0.592	0.38	.839	0.545	0.38	.866	0.672	0.41	.924	2.258	0.49
		.948	2.177	0.55	.886	0.524	0.46	.901	0.465	0.45	.925	0.654	0.49	.964	3.907	0.59
		.985	4.818	0.71	.956	0.433	0.60	.965	0.355	0.59	.977	0.624	0.65	.993	12.185	0.78
100	N/A	.693	1.004	0.20	.628	0.743	0.17	.633	0.735	0.17	.645	0.818	0.18	.692	1.002	0.20
		.811	1.134	0.26	.745	0.667	0.22	.752	0.652	0.22	.764	0.771	0.23	.819	1.279	0.26
		.904	1.464	0.33	.852	0.548	0.28	.860	0.519	0.28	.868	0.716	0.30	.916	2.137	0.34
		.950	2.057	0.39	.912	0.457	0.34	.919	0.421	0.34	.924	0.675	0.35	.960	3.950	0.41
		.987	4.716	0.51	.970	0.311	0.44	.975	0.263	0.44	.977	0.621	0.46	.991	12.486	0.54
200	N/A	.688	1.016	0.14	.646	0.765	0.13	.648	0.763	0.13	.650	0.865	0.13	.685	1.013	0.14
		.807	1.126	0.18	.765	0.677	0.16	.769	0.669	0.16	.768	0.834	0.17	.810	1.228	0.18
		.903	1.346	0.23	.870	0.541	0.21	.874	0.525	0.21	.871	0.753	0.21	.910	1.796	0.24
		.949	1.735	0.28	.925	0.416	0.25	.928	0.404	0.25	.927	0.708	0.25	.956	3.119	0.28
		.988	3.607	0.36	.977	0.263	0.33	.980	0.215	0.32	.978	0.635	0.33	.990	9.938	0.38
500	N/A	.684	1.008	0.09	.662	0.802	0.08	.663	0.801	0.08	.658	0.925	0.08	.682	1.005	0.09
		.806	1.082	0.12	.783	0.705	0.11	.784	0.702	0.11	.778	0.894	0.11	.806	1.160	0.12
		.903	1.217	0.15	.886	0.557	0.14	.888	0.552	0.14	.881	0.842	0.14	.906	1.502	0.15
		.951	1.489	0.18	.939	0.426	0.16	.940	0.419	0.16	.934	0.782	0.17	.954	2.236	0.18
		.989	2.589	0.23	.984	0.218	0.22	.985	0.214	0.21	.983	0.697	0.22	.989	7.548	0.23

Table 75: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 41a2a, $\rho = .5$

		Distribution 41a2a: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = .5$															
N	Exact CI	Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	N/A	.603	1.539	0.35	.514	1.378	0.29	.543	1.138	0.29	.602	0.918	0.33	.690	1.690	0.42
		.722	1.867	0.45	.618	1.515	0.37	.662	1.071	0.38	.728	0.881	0.43	.824	2.040	0.56
		.831	2.659	0.58	.724	1.763	0.47	.778	0.967	0.49	.843	0.848	0.56	.931	2.751	0.73
		.894	3.858	0.68	.790	2.019	0.56	.849	0.866	0.58	.908	0.899	0.68	.974	4.437	0.88
		.958	12.361	0.88	.876	2.781	0.72	.931	0.672	0.76	.970	1.420	0.88	.997	15.100	1.16
50	N/A	.571	1.278	0.22	.563	1.149	0.21	.581	0.973	0.21	.600	0.858	0.22	.684	1.288	0.27
		.691	1.470	0.28	.678	1.211	0.27	.701	0.881	0.27	.719	0.815	0.29	.814	1.386	0.35
		.806	1.873	0.36	.788	1.311	0.35	.817	0.743	0.35	.833	0.791	0.37	.917	1.461	0.45
		.873	2.433	0.42	.855	1.404	0.42	.882	0.609	0.42	.897	0.731	0.44	.966	1.498	0.54
		.951	5.560	0.55	.932	1.639	0.54	.952	0.415	0.55	.963	0.743	0.57	.996	2.633	0.70
100	N/A	.554	1.164	0.15	.597	1.040	0.16	.608	0.900	0.16	.607	0.853	0.17	.686	1.143	0.20
		.672	1.273	0.19	.713	1.061	0.21	.727	0.811	0.21	.727	0.820	0.22	.811	1.188	0.25
		.788	1.528	0.25	.820	1.078	0.27	.836	0.658	0.27	.834	0.755	0.28	.911	1.209	0.32
		.860	1.884	0.30	.885	1.096	0.32	.900	0.522	0.32	.898	0.690	0.33	.962	1.209	0.39
		.944	3.663	0.39	.956	1.152	0.42	.964	0.316	0.42	.963	0.631	0.43	.995	1.159	0.50
200	N/A	.539	1.137	0.11	.621	1.014	0.13	.628	0.903	0.13	.619	0.902	0.13	.692	1.115	0.14
		.657	1.215	0.14	.742	0.988	0.16	.751	0.790	0.16	.739	0.852	0.16	.812	1.121	0.18
		.776	1.362	0.18	.851	0.956	0.21	.860	0.631	0.21	.846	0.803	0.21	.913	1.133	0.24
		.850	1.576	0.21	.911	0.959	0.25	.917	0.499	0.25	.907	0.756	0.25	.960	1.093	0.28
		.938	2.477	0.27	.971	0.888	0.32	.973	0.261	0.32	.968	0.654	0.32	.994	0.906	0.37
500	N/A	.527	1.060	0.07	.647	0.965	0.09	.650	0.888	0.09	.638	0.918	0.09	.699	1.055	0.09
		.644	1.110	0.09	.768	0.932	0.11	.772	0.790	0.11	.756	0.899	0.11	.818	1.045	0.12
		.763	1.222	0.11	.872	0.852	0.14	.875	0.632	0.14	.860	0.824	0.14	.912	1.035	0.16
		.840	1.340	0.13	.927	0.788	0.17	.930	0.487	0.17	.916	0.788	0.17	.960	1.014	0.19
		.931	1.789	0.17	.980	0.644	0.22	.979	0.258	0.22	.973	0.707	0.22	.993	0.867	0.24

Table 76: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 41a2b, $\rho = 0$

			Distribution 41a2b: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = 0$			
N	Exact CI	Fisher Z CI	ADF CI	ADF Fisher Z CI	BCa CI	Univariate Sampling CI
20	N/A	N/A	N/A	N/A	N/A	N/A
50	N/A	.688 1.032 0.29 .808 1.210 0.37 .904 1.702 0.47 .952 2.600 0.55 .988 8.413 0.71	.605 0.714 0.24 .719 0.631 0.31 .824 0.516 0.40 .883 0.437 0.48 .951 0.317 0.63	.615 0.702 0.24 .732 0.603 0.31 .838 0.474 0.40 .898 0.373 0.47 .962 0.245 0.61	.647 0.770 0.25 .767 0.699 0.33 .870 0.626 0.42 .927 0.543 0.51 .979 0.550 0.66	.702 1.063 0.29 .830 1.480 0.38 .926 2.737 0.49 .967 5.358 0.59 .994 43.143 0.77
100	N/A	.681 0.999 0.20 .804 1.146 0.26 .905 1.490 0.33 .951 2.029 0.39 .988 5.242 0.51	.627 0.717 0.18 .744 0.620 0.23 .851 0.490 0.30 .909 0.389 0.35 .966 0.271 0.46	.633 0.709 0.18 .751 0.608 0.23 .860 0.458 0.29 .916 0.359 0.35 .971 0.207 0.46	.649 0.813 0.18 .770 0.769 0.24 .877 0.682 0.30 .932 0.630 0.36 .980 0.544 0.48	.687 1.004 0.20 .817 1.309 0.26 .918 2.217 0.34 .961 4.166 0.40 .991 17.667 0.53
200	N/A	.683 0.994 0.14 .805 1.129 0.18 .902 1.412 0.23 .951 1.759 0.28 .989 3.516 0.36	.646 0.738 0.13 .766 0.656 0.17 .870 0.518 0.22 .925 0.400 0.26 .974 0.236 0.34	.649 0.735 0.13 .769 0.650 0.17 .874 0.502 0.21 .928 0.383 0.26 .976 0.212 0.33	.658 0.872 0.13 .777 0.839 0.17 .882 0.779 0.22 .935 0.732 0.26 .982 0.606 0.34	.685 0.997 0.14 .810 1.242 0.18 .909 1.880 0.24 .955 3.097 0.28 .990 11.725 0.37
500	N/A	.681 1.004 0.09 .803 1.093 0.12 .902 1.322 0.15 .949 1.579 0.18 .988 3.217 0.23	.661 0.803 0.09 .782 0.710 0.11 .884 0.599 0.14 .936 0.486 0.17 .982 0.308 0.22	.662 0.801 0.09 .783 0.707 0.11 .886 0.592 0.14 .938 0.480 0.17 .983 0.286 0.22	.664 0.934 0.09 .785 0.910 0.11 .886 0.887 0.14 .939 0.895 0.17 .984 0.845 0.22	.679 1.001 0.09 .803 1.157 0.12 .904 1.615 0.15 .951 2.347 0.18 .988 8.231 0.23

Table 77: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 41a2b, $\rho = .5$

Distribution 41a2b: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = .5$

N	Exact CI	Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	N/A	.420	2.177	0.31	.432	2.830	0.31	.465	2.370	0.32	.558	2.852	0.36	.512	2.052	0.38
		.531	2.241	0.40	.528	3.415	0.40	.580	2.536	0.42	.678	3.348	0.48	.655	1.986	0.51
		.657	2.276	0.51	.626	4.434	0.52	.704	2.848	0.54	.797	3.971	0.62	.797	1.772	0.67
		.747	2.210	0.61	.692	5.556	0.61	.791	2.970	0.65	.870	4.338	0.75	.885	1.500	0.82
		.876	1.949	0.80	.788	8.301	0.76	.900	3.481	0.85	.949	5.591	0.99	.969	1.092	1.08
50	N/A	.383	2.005	0.20	.491	3.009	0.25	.506	2.543	0.26	.554	2.884	0.27	.478	1.875	0.25
		.481	2.047	0.26	.594	4.025	0.33	.618	2.966	0.33	.668	3.690	0.35	.603	1.830	0.32
		.599	2.056	0.33	.695	5.955	0.42	.736	3.641	0.42	.779	5.094	0.45	.741	1.634	0.41
		.687	2.012	0.39	.762	8.508	0.50	.815	4.607	0.51	.848	6.500	0.54	.832	1.415	0.50
		.828	1.797	0.51	.850	18.992	0.65	.914	7.051	0.66	.933	9.334	0.71	.939	0.909	0.65
100	N/A	.358	1.798	0.14	.528	2.920	0.21	.535	2.489	0.21	.569	2.506	0.22	.461	1.751	0.18
		.455	1.841	0.18	.636	4.208	0.27	.648	3.104	0.27	.683	3.260	0.28	.580	1.736	0.23
		.567	1.862	0.24	.738	7.256	0.35	.764	4.368	0.35	.791	4.720	0.36	.711	1.633	0.30
		.655	1.829	0.28	.803	12.196	0.41	.836	6.272	0.42	.857	6.737	0.44	.804	1.493	0.36
		.794	1.676	0.37	.882	35.457	0.54	.923	12.242	0.54	.934	12.332	0.57	.918	1.040	0.47
200	N/A	.346	1.654	0.10	.563	2.774	0.17	.566	2.392	0.17	.589	2.102	0.18	.455	1.669	0.13
		.438	1.684	0.13	.675	4.007	0.22	.682	3.041	0.22	.705	2.643	0.23	.571	1.691	0.17
		.548	1.694	0.17	.776	7.378	0.28	.789	4.524	0.28	.809	3.731	0.29	.696	1.642	0.22
		.632	1.681	0.20	.835	13.778	0.33	.855	7.030	0.33	.872	5.467	0.35	.784	1.545	0.26
		.768	1.583	0.27	.907	55.890	0.44	.935	17.623	0.44	.944	10.305	0.45	.904	1.228	0.34
500	N/A	.332	1.439	0.07	.602	2.378	0.12	.600	2.099	0.12	.610	1.634	0.13	.455	1.491	0.09
		.422	1.450	0.09	.718	3.422	0.16	.720	2.734	0.16	.730	1.933	0.16	.566	1.524	0.12
		.526	1.473	0.11	.817	6.556	0.20	.824	4.293	0.20	.835	2.519	0.21	.689	1.547	0.15
		.608	1.460	0.13	.874	12.820	0.24	.886	6.860	0.24	.896	3.313	0.25	.772	1.542	0.18
		.743	1.408	0.17	.937	64.458	0.32	.952	19.517	0.32	.960	6.252	0.33	.889	1.413	0.23

Table 78: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 41b2a, $\rho = 0$

Distribution 41b2a: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = 0$																			

	Exact CI	Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
N	N/A	N/A			N/A			N/A			N/A			N/A		
20	N/A	.691	1.015	0.29	.600	0.716	0.24	.610	0.704	0.24	.642	0.755	0.25	.694	1.014	0.29
		.810	1.211	0.37	.715	0.638	0.30	.729	0.608	0.30	.760	0.689	0.32	.827	1.408	0.38
		.905	1.628	0.47	.819	0.528	0.39	.835	0.483	0.39	.867	0.615	0.41	.923	2.580	0.49
		.950	2.431	0.55	.883	0.450	0.46	.898	0.391	0.46	.924	0.573	0.50	.963	4.716	0.59
		.985	5.328	0.71	.953	0.356	0.61	.963	0.272	0.60	.976	0.556	0.65	.991	17.240	0.78
50	N/A	.690	0.991	0.20	.627	0.727	0.17	.633	0.719	0.17	.648	0.818	0.18	.689	0.988	0.20
		.809	1.150	0.26	.747	0.638	0.22	.753	0.623	0.22	.767	0.767	0.23	.816	1.298	0.26
		.904	1.513	0.33	.852	0.515	0.29	.860	0.485	0.29	.872	0.704	0.30	.915	2.271	0.34
		.952	2.092	0.39	.911	0.413	0.34	.917	0.378	0.34	.927	0.620	0.36	.960	4.258	0.41
		.988	4.941	0.51	.968	0.249	0.45	.973	0.190	0.45	.979	0.530	0.47	.990	17.346	0.54
100	N/A	.686	1.029	0.14	.647	0.768	0.13	.650	0.764	0.13	.653	0.888	0.13	.684	1.019	0.14
		.805	1.138	0.18	.768	0.659	0.17	.772	0.653	0.17	.772	0.845	0.17	.807	1.235	0.18
		.903	1.360	0.23	.871	0.512	0.21	.874	0.500	0.21	.877	0.750	0.22	.908	1.823	0.24
		.950	1.721	0.28	.926	0.404	0.25	.929	0.378	0.25	.931	0.695	0.26	.955	3.119	0.28
		.988	3.277	0.36	.976	0.222	0.33	.978	0.188	0.33	.979	0.578	0.34	.989	10.298	0.37
200	N/A	.684	1.021	0.09	.661	0.813	0.08	.662	0.813	0.08	.660	0.949	0.08	.681	1.013	0.09
		.803	1.086	0.12	.783	0.713	0.11	.784	0.712	0.11	.781	0.907	0.11	.803	1.161	0.12
		.901	1.301	0.15	.883	0.574	0.14	.884	0.568	0.14	.881	0.861	0.14	.904	1.596	0.15
		.949	1.627	0.18	.936	0.445	0.17	.937	0.437	0.17	.935	0.811	0.17	.952	2.472	0.18
		.989	3.118	0.23	.983	0.269	0.22	.984	0.264	0.22	.983	0.826	0.22	.989	7.304	0.23

Table 79: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 41b2a, $\rho = .5$

N	Exact CI	Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
		Distribution 41b2a: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = .5$													
20	N/A	.237	2.005	0.26	.370	2.824	0.44	.403	2.390	0.46	.507	3.925	0.54	.268	2.094	0.33
		.303	2.112	0.34	.441	3.420	0.55	.499	2.691	0.60	.606	5.284	0.70	.360	2.196	0.44
		.384	2.286	0.44	.515	4.380	0.68	.602	3.269	0.77	.712	9.146	0.93	.477	2.323	0.58
		.450	2.459	0.53	.566	5.444	0.76	.677	4.042	0.93	.776	3.664	1.12	.584	2.479	0.71
		.570	2.873	0.70	.635	7.754	0.90	.784	6.350	1.21	.858	29.472	1.44	.760	2.838	0.94
50	N/A	.232	1.737	0.19	.469	2.707	0.41	.492	2.184	0.42	.550	2.811	0.45	.274	1.787	0.24
		.297	1.805	0.24	.558	3.477	0.53	.597	2.480	0.54	.656	3.670	0.58	.354	1.860	0.31
		.374	1.919	0.31	.643	4.890	0.68	.708	3.074	0.69	.762	5.701	0.75	.453	1.913	0.40
		.439	2.011	0.37	.699	6.937	0.79	.781	3.906	0.82	.828	8.771	0.89	.536	1.959	0.48
		.554	2.219	0.49	.774	12.609	0.95	.877	7.107	1.06	.907	21.104	1.15	.685	2.024	0.62
100	N/A	.223	1.574	0.14	.523	2.576	0.35	.538	2.042	0.35	.578	2.274	0.37	.276	1.619	0.18
		.286	1.625	0.18	.626	3.527	0.45	.655	2.398	0.45	.693	2.901	0.47	.355	1.662	0.23
		.364	1.693	0.23	.720	5.410	0.58	.766	3.118	0.58	.800	4.257	0.61	.449	1.701	0.30
		.425	1.748	0.27	.776	8.083	0.68	.835	4.014	0.68	.862	6.040	0.72	.529	1.733	0.36
		.541	1.911	0.36	.848	21.023	0.87	.918	8.013	0.89	.934	13.522	0.93	.669	1.774	0.46
200	N/A	.219	1.436	0.10	.567	2.342	0.28	.574	1.904	0.28	.598	1.855	0.29	.285	1.478	0.14
		.281	1.468	0.13	.674	3.231	0.36	.691	2.228	0.36	.715	2.213	0.37	.362	1.498	0.17
		.354	1.506	0.17	.772	5.302	0.46	.803	2.860	0.46	.823	2.973	0.48	.457	1.537	0.22
		.418	1.560	0.20	.830	8.597	0.55	.870	3.828	0.55	.886	3.902	0.57	.534	1.557	0.27
		.529	1.646	0.26	.898	27.406	0.72	.944	7.959	0.72	.953	7.607	0.74	.665	1.578	0.35
500	N/A	.212	1.310	0.07	.611	2.083	0.20	.613	1.747	0.20	.621	1.504	0.21	.292	1.352	0.09
		.271	1.333	0.08	.724	2.844	0.26	.732	2.084	0.26	.740	1.732	0.26	.372	1.368	0.12
		.345	1.365	0.11	.822	4.862	0.33	.838	2.714	0.33	.845	2.101	0.34	.466	1.384	0.15
		.406	1.369	0.13	.878	8.513	0.40	.899	3.859	0.39	.906	2.771	0.40	.546	1.414	0.18
		.519	1.422	0.17	.938	36.155	0.52	.962	8.555	0.52	.966	4.658	0.53	.674	1.444	0.23

Table 80: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 41b2b, $\rho = 0$

N	Exact CI	Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	N/A	N/A			N/A			N/A			N/A			N/A		
50	N/A	.682	1.053	0.29	.608	0.714	0.25	.618	0.698	0.25	.649	0.789	0.26	.694	1.078	0.29
		.804	1.280	0.37	.721	0.621	0.32	.735	0.590	0.32	.768	0.712	0.33	.826	1.573	0.38
		.904	1.849	0.47	.824	0.485	0.41	.839	0.441	0.41	.872	0.606	0.43	.922	3.161	0.49
		.951	2.927	0.55	.883	0.404	0.49	.897	0.338	0.48	.928	0.508	0.51	.963	7.393	0.59
		.988	14.895	0.71	.949	0.263	0.64	.959	0.171	0.63	.978	0.385	0.67	.992	204.500	0.76
100	N/A	.687	0.999	0.20	.637	0.705	0.18	.642	0.697	0.18	.661	0.826	0.19	.693	1.006	0.20
		.808	1.169	0.26	.754	0.598	0.23	.761	0.583	0.23	.780	0.771	0.24	.820	1.337	0.26
		.905	1.574	0.33	.855	0.468	0.30	.862	0.438	0.30	.882	0.690	0.31	.916	2.347	0.34
		.953	2.282	0.39	.910	0.360	0.36	.917	0.324	0.36	.935	0.635	0.37	.959	4.538	0.40
		.988	6.658	0.51	.967	0.219	0.47	.971	0.165	0.46	.982	0.517	0.48	.990	35.357	0.53
200	N/A	.687	0.987	0.14	.653	0.732	0.13	.656	0.729	0.13	.666	0.873	0.13	.689	0.987	0.14
		.806	1.097	0.18	.773	0.629	0.17	.776	0.624	0.17	.787	0.833	0.17	.811	1.211	0.18
		.904	1.332	0.23	.871	0.498	0.22	.874	0.481	0.22	.886	0.787	0.22	.910	1.803	0.24
		.950	1.767	0.28	.924	0.396	0.26	.927	0.374	0.26	.937	0.727	0.26	.955	3.052	0.28
		.989	4.586	0.36	.974	0.209	0.34	.976	0.185	0.34	.983	0.620	0.35	.989	15.688	0.37
500	N/A	.685	0.997	0.09	.668	0.791	0.09	.669	0.789	0.09	.672	0.934	0.09	.685	0.992	0.09
		.804	1.056	0.12	.786	0.707	0.11	.787	0.703	0.11	.791	0.913	0.11	.807	1.125	0.12
		.903	1.248	0.15	.886	0.562	0.14	.887	0.555	0.14	.890	0.861	0.14	.905	1.535	0.15
		.951	1.538	0.18	.938	0.449	0.17	.939	0.443	0.17	.942	0.871	0.17	.952	2.225	0.18
		.991	2.762	0.23	.983	0.240	0.22	.984	0.235	0.22	.987	0.754	0.22	.990	6.132	0.23

Distribution 41b2b: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = 0$

Table 81: Coverage Rate, Coverage Balance, and Median Width for CIs on Distribution 41b2b, $\rho = .5$

	Distribution 41b2b: $\gamma_{11} = \gamma_{12} = 0$ and $\gamma_{21} = \gamma_{22} = 25$, $\rho = .5$																
N	Exact CI	Fisher Z CI			ADF CI			ADF Fisher Z CI			BCa CI			Univariate Sampling CI		
20	N/A	.542	1.203	0.35	.525	1.273	0.34	.555	1.021	0.35	.616	0.911	0.38	.622	1.175	0.42
		.661	1.281	0.46	.628	1.394	0.43	.668	0.959	0.45	.738	0.869	0.50	.765	1.257	0.55
		.783	1.455	0.58	.728	1.574	0.56	.779	0.845	0.57	.847	0.778	0.65	.889	1.434	0.72
		.857	1.688	0.69	.790	1.724	0.66	.846	0.747	0.68	.908	0.707	0.78	.954	1.889	0.87
		.946	2.817	0.89	.871	2.103	0.84	.923	0.557	0.88	.968	0.654	1.02	.994	6.000	1.14
50	N/A	.525	1.158	0.22	.585	1.256	0.25	.599	1.044	0.25	.626	0.990	0.26	.607	1.097	0.25
		.642	1.201	0.28	.697	1.363	0.32	.717	0.972	0.32	.745	0.969	0.33	.738	1.091	0.33
		.762	1.308	0.36	.800	1.570	0.41	.827	0.876	0.41	.853	0.926	0.43	.863	1.068	0.42
		.839	1.451	0.42	.862	1.794	0.48	.891	0.753	0.49	.914	0.887	0.51	.931	1.042	0.51
		.933	2.129	0.55	.934	2.210	0.63	.958	0.547	0.64	.970	0.807	0.66	.987	1.051	0.66
100	N/A	.511	1.099	0.15	.612	1.190	0.19	.620	1.012	0.19	.630	0.969	0.19	.606	1.037	0.18
		.628	1.129	0.19	.730	1.304	0.24	.743	0.956	0.24	.754	0.956	0.25	.733	1.011	0.23
		.747	1.214	0.25	.834	1.459	0.31	.849	0.875	0.31	.857	0.952	0.31	.850	0.966	0.30
		.824	1.325	0.30	.892	1.672	0.37	.909	0.765	0.37	.913	0.952	0.37	.918	0.913	0.36
		.919	1.801	0.39	.957	2.315	0.48	.970	0.545	0.48	.972	0.933	0.48	.981	0.891	0.47
200	N/A	.493	1.074	0.11	.630	1.121	0.14	.635	0.985	0.14	.632	0.948	0.14	.604	1.023	0.13
		.611	1.101	0.14	.750	1.203	0.18	.758	0.934	0.18	.753	0.947	0.18	.732	1.011	0.17
		.729	1.191	0.18	.857	1.321	0.23	.867	0.834	0.23	.861	0.954	0.23	.848	0.962	0.22
		.809	1.291	0.21	.916	1.460	0.28	.925	0.731	0.28	.918	0.947	0.28	.915	0.885	0.26
		.909	1.633	0.27	.973	1.930	0.36	.978	0.497	0.36	.974	1.013	0.36	.979	0.710	0.34
500	N/A	.475	1.028	0.07	.642	1.027	0.09	.645	0.939	0.09	.633	0.934	0.09	.607	0.996	0.09
		.588	1.057	0.09	.766	1.042	0.12	.769	0.870	0.12	.753	0.912	0.12	.731	0.980	0.11
		.706	1.116	0.11	.873	1.080	0.16	.877	0.760	0.16	.859	0.900	0.16	.848	0.907	0.14
		.787	1.193	0.13	.931	1.115	0.19	.935	0.654	0.19	.918	0.909	0.18	.912	0.818	0.17
		.892	1.425	0.17	.982	1.226	0.24	.983	0.419	0.24	.975	0.938	0.24	.976	0.641	0.22

REFERENCES

Algina, J., Keselman, H. J. & Penfield, R. D. (2006). Confidence interval coverage for Cohen's effect size statistic. *Educational and Psychological Measurement, 66*(6).

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician, 27*(1), 17–21.

Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew–normal distribution. *Biometrika, 84*(4), 715–726.

Arnold, B. C. & Beaver, R. J. (2000). The skew-Cauchy distribution. *Statistics & Probability Letters, 49*, 285–290.

Arnold, B. C. & Strauss, D. J. (1991). Bivariate distributions with conditionals in prescribed exponential families. *Journal of the Royal Statistical Society. Series B (Methodological), 53*(2), 365–375.

Badrinath, S. G. & Chatterjee, S. (1988). On measuring skewness and elongation in common stock return distributions: The case of the market index. *Journal of Business, 61*, 451–472.

Badrinath, S. G. & Chatterjee, S. (1991). A data-analytic look at skewness and elongation in common-stock return distributions. *Journal of Business & Economic Statistics, 9*, 223–233.

Baker, G. A. (1930). The significance of the product-moment coefficient correlation with special reference to the character of the marginal distribution. *Journal of the American Statistical Association, 25*(172), 387–396.

Beasley, W. H., DeShea, L., Toothaker, L. E. et al. (2007). Bootstrapping to test for nonzero population correlation coefficients using univariate sampling. *Psychological Methods, 12*(4), 414–433.

Carroll, J. B. (1961). The nature of the data, or how to choose a correlation coefficient*. *Psychometrika, 26*(4), 347–372.

Chen, X. & Tung, Y.–K. (2003). Investigation of polynomial normal transform. *Structural Safety, 25*, 423–445.

Dalen, J. (1987). Algebraic bounds on standardized sample moments. *Statistics & Probability Letters, 5*, 329–331.

Devlin, S. J., Gnanadesikan, R. & Kettenring, J. R. (1975). Robust estimation and outlier detection with correlation coefficients. *Biometrika, 62*(3), 531–545.

Devroye, L. (1986). *Non–Uniform Random Variate Generation*. Springer–Verlag, New York.

Diaconis, P. & Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American, 248*(5), 116–130.

DiCiccio, T. J. & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science, 11*(3), 189–212.

Dolker, M., Halperin, S. & Divgi, D. R. (1982). Problems with bootstrapping Pearson correlations in very small bivariate samples. *Psychometrika, 47*(4), 529–530.

Duncan, G. T. & Layard, M. W. J. (1973). A Monte-Carlo study of asymptotically robust tests for correlation coefficients. *Biometrika, 60*(3), 551–558.

Dunlap, W. P., Burke, M. J. & Greer, T. (1995). The effect of skew on the magnitude of product-moment correlations. *The Journal of General Psychology, 122*(4). 365–377.

Edgell, S. E. & Noon, S. M. (1984). Effect of violation of normality on the *t* test of the correlation coefficient. *Psychology Bulletin, 95*(3), 576–583.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*(1), 1–26.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association, 82,* 171-200.

Efron, B. (1988). Bootstrap confidence intervals: Good or bad? *Psychological Bulletin, 104*(2), 293–296.

Efron, B. (2003). Second thoughts on the bootstrap. *Statistical Science, 18*(2), 135–140.

Efron, B. & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician, 37*(1), 36–48.

Fan, X., Felsövalyi, A., Sivo, S. A. & Keenan, S. C. (2003). *SAS for Monte Carlo Studies: a guide for quantitative researchers*. SAS Institute, Inc., Cary, North Carolina.

Field, C. & Genton, M. G. (2006). The multivariate **g**-and-**h** distribution**.** *Technometrics, 48*(1).

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika, 10*(4), 507–521.

Fleishman, A. I. (1978). A method for simulating non–normal distributions. *Psychometrika 43*(4).

Games, P. A. (1983). Curvilinear transformations of the dependent variable. *Psychological Bulletin, 93*(2), 382–387.

Haldane, J. B. S. (1949). A Note on Non–Normal Correlation. *Biometrika, 36,* 467–468.

Halperin, S. (1986). Spurious correlations – causes and cures. *Psychoneuroendocrinology, 11*(1), 3–13.

Harwell, M. R. & Serlin, R. C. (1989). A nonparametric test statistic for the general linear model. *Journal of Educational Statistics, 14*(4), 351–371.

Havlicek, L. L. & Peterson, N. L. (1977). Effect of the violation of assumptions upon significance levels of the Pearson r. *Psychological Bulletin, 84*(2), 373–377.

Headrick, T. C. (2002). Fast fifth–order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics & Data Analysis 40*, 685–711.

Headrick, T. C. (2004). Transformations for simulating multivariate distributions. *Journal of Modern Applied Statistical Methods, 3*(1), 65–71.

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*(3), 229–249.

Headrick, T. C., Kowalchuk, R. K. & Sheng, Y. (2008). Parametric probability densities and distribution functions for Tukey *g*-and-*h* transformations and their use for fitting data. *Applied Mathematical Sciences, 2*(9), 449–462.

Headrick, T. C. & Mugdadi, A. (2006). On simulating multivariate non-normal distributions from the generalized lambda distribution. *Computational Statistics & Data Analysis, 50*, 3343–3353.

Headrick, T. C. & Sawilowsky, S. S. (1999). Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method. *Psychometrika, 64*(1), 25–35.

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*(3)

Hoaglin, D.C., Peters, S.C. (1979). Software for exploring distribution shape. *Proceedings of computer science and statistics*

Hoaglin, D. C. (1985). Summarizing shape numerically: The *g*-and-*h* distributions. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), Exploring data tables, trends, and shapes (pp. 461–513). New York: Wiley.

Hoyland, K., Kaut, M. & Wallace, S. W. (2003). A heuristic for moment-matching scenario generation. *Computational Optimization and Applications, 24*, 169–185.

Johnson, M. E. & Lowe, V. W. (1979). Bounds on the sample skewness and kurtosis. *Technometrics, 21*(3), 377–378.

Jolliffe, I. T. & Hope, P. B. (1996). Bounded Bivariate Distributions with Nearly Normal Marginals. *The American Statistician, 50*(1), 17–20

Kaiser, H. F. & Dickmann, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika, 27*(2).

Keselman, H. J., Lix, L. M. & Kowalchuk, R. K. (1998). Multiple comparison procedures for trimmed means. *Psychological Methods, 3*(1), 123–141.

Kirby, W. (1974). Algebraic boundedness of sample statistics. *Water Resources Research, 10*(2), 220–222.

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

Kowalski, C. J. (1972). On the effects of non–normality on the distribution of the sample product–moment correlation coefficient. *Journal of the Royal Statistical Society. Series C, 21*(1), 112.

Lancaster, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika, 44*(1/2), 289–292.

Lee, W-C. & Rodgers, J. L. (1998). Bootstrapping correlation coefficients using univariate and bivariate sampling. *Psychological Methods, 3*(1), 91–103.

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.

Lunneborg, C. F. (1985). Estimating the correlation coefficient: The bootstrap approach. *Psychological Bulletin, 98*(1), 209–215.

Lurie, P. M. & Goldberg, M. S. (1998). An approximate method for sampling correlated random variables from partially-specified distributions. *Management Science, 44*(2), 203–218.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519–530.

Martinez, J. & Iglewicz, B. (1984). Some properties of the Tukey g and h family of distributions. *Communications in Statistics – Theory and Methods, 13*(3), 353–369.

Mendoza, J. L., Hart, D. E. & Powell, A. (1991). A bootstrap confidence interval based on a correlation corrected for range restriction. *Multivariate Behavioral Research, 26*(2), 255–269.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156–166.

Nagahara, Y. (2004). A method of simulating multivariate nonnormal distributions by the Pearson distribution system and estimation. *Computational Statistics & Data Analysis, 47*, 1–29.

Nanna, M. J. & Sawilowski, S. S. (1998). Analysis of likert scale data in disability and medical rehabilitation research. *Psychological Methods, 3*(1), 55–67.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine, 17*, 857–872.

Norris, R. C. & Hjelm, H. F. (1961). Nonnormality and product moment correlation. *Journal of Experimental Education, 29*(3), 261–270.

Olkin, I. (1994). Multivariate Non–Normal Distributions and Models of Dependency. *Lecture Notes–Monograph Series, Vol. 24, Multivariate Analysis and Its Applications*, 37–53.

Olkin, I. & Finn, J. D. (1995). Correlations redux. *Psychological Bulletin, 118*(1), 155–164.

Parrish, R. S. (1990). Generating random deviates from multivariate Pearson distributions*. *Computational Statistics & Data Analysis, 9*, 283–295.

Pearson, K. (1923). Notes on skew frequency surfaces. *Biometrika, 15*(3/4), 222–230.

Pearson, E. S. (1929). Some notes on sampling tests with two variables. *Biometrika, 21*(1/4), 337–360.

Raju, N. S., Pappas, S. & Williams, C. P. (1989). An empirical Monte Carlo test of the accuracy of the correlation, covariance, and regression slope models for assessing validity generalization. *Journal of Applied Psychology, 74*(6), 901–911.

Rasmussen, J. L. (1987). Estimating correlation coefficients: Bootstrap and parametric approaches. *Psychological Bulletin, 101*(1), 136–139.

Rasmussen, J. L. (1988). "Bootstrap confidence intervals: Good or bad": Comments on Efron (1988) and Strube (1988) and further evaluation. *Psychological Bulletin, 104*(2), 297–299.

Rasmussen, J. L. (1989). Computer-intensive correlational analysis: Bootstrap and approximate randomization techniques. *British Journal of Mathematical and Statistical Psychology, 42*, 103–111.

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

Sarabia, J. M. & Gomez–Deniz, E. (2008). Construction of multivariate distributions: a review of some recent results. *SORT 32* (1), 3–36.

Sharma, S., Durvasula, S., Dillon, W. R. (1989). Some results on the behavior of alternate covariance structures estimation procedures in the presence of non-normal data. *Journal of Marketing Research, 26*(2), 214–221.

Sievers, W. (1996). Standard and bootstrap confidence intervals for the correlation coefficient. *British Journal of Mathematical and Statistical Psychology, 49*, 381–396.

Srivastava, S., John, O. P., Gosling, S. D. & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology, 84*(5), 1041–1053.

Steiger, J. H. (2004). Beyond the *F* test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9*(2), 164–182.

Steiger, J. H. & Hakstian, A. R. (1982). The asymptotic distribution of elements of a correlation matrix: Theory and application. *British Journal of Mathematical and Statistical Psychology, 35*, 208–215.

Steiger, J. H. & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.) *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum Associates.

Strube, M. J. (1988). Bootstrap type I error rates for the correlation coefficient: An examination of alternate procedures. *Psychological Bulletin, 104*(2), 290–292.

Tadikamalla, P. R. (1980). On simulating non–normal distributions. *Psychometrika, 45*(2). 273–279.

Tukey, J.W. (1977). *Modern techniques in data analysis*. NSF sponsored regional research conference at Southern Massachusetts University, North Dartmouth, MA.

Vale, C. D. & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48*(3).

Wilcox, R. R. (1991). Bootstrap inferences about the correlation and variances of paired data. *British Journal of Mathematical and Statistical Psychology, 44*, 379–382.

Wilcox, R. R. (1993). Some results on a winsorized correlation coefficient. *British Journal of Mathematical and Statistical Psychology, 46*, 339–349.

Wilcox, R. R. (1994). The percentage bend correlation coefficient. *Psychometrika, 59*(4), 601–616.

Wilcox, R. R. (1995). Simulation results on solutions to the multivariate Behrens-Fisher problem via trimmed means. *The Statistician, 44*(2), 213–225.

Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing* (2nd ed.). San Francisco, CA: Elsevier.

Yang, I.T. (2008). Distribution-free Monte Carlo simulation: Premise and refinement. *Journal of construction engineering and management, 134*(5), 352–360.

Zeller, R. A. & Levine, Z. H. (1974). The effects of violating the normality assumption underlying *r*. *Sociological Methods & Research, 2*(4), 511–519.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods, 12*(4), 399–413.