University of Illinois at Chicago
School of Public Health
Division of Epidemiology and Biostatistics

Technical report#:2007-001
February 2007

Title: Multiple imputation under power polynomials

Authors: Hakan Demirtas and **Donald Hedeker**

Affiliation(s): University of Illinois at Chicago, Division of Epidemiology and Biostatistics.

Multiple imputation under power polynomials

Hakan Demirtas and Donald Hedeker*

February 07, 2007

Abstract

Although the normality assumption has been regarded as a mathematical convenience for inferential purposes due to its nice distributional properties, there has been a growing interest regarding generalized classes of distributions that span a much broader spectrum in terms of symmetry and peakedness behavior. In this respect, Fleishman's power polynomial method seems to have been gaining popularity in statistical theory and practice due because of its flexibility and ease of execution. In this article, we conduct multiple imputation for univariate continuous data under Fleishman polynomials to explore the extent to which this procedure works properly. We also make comparisons with normal imputation models via widely accepted accuracy and precision measures using simulated data that exhibit different distributional features as characterized by competing specifications of the third and fourth moments. Finally, we discuss generalizations to the multivariate case. Multiple imputation under power polynomials that cover most of the feasible area in the skewness-elongation plane appears to have substantial potential of capturing real missing-data trends.

Key Words: Multiple imputation; Normality; Symmetry; Skewness; Kurtosis

1 Introduction

The normality assumption is categorically one of the most extensively used and studied phenomena in statistics. Although real data rarely conform with normality, it has been regarded as a convenient assumption in model formation, estimation and testing due to its well-understood distributional properties. Despite its popularity, it only represents a single point in the skewness-elongation plane; and general classes of continuous distributions that span a broader spectrum in terms of symmetry and peakedness behavior (Genton [1])

*Hakan Demirtas (e-mail:demirtas@uic.edu) is an Assistant Professor, and Donald Hedeker (hedeker@uic.edu) is a Professor of Biostatistics, Division of Epidemiology and Biostatistics (MC923), University of Illinois at Chicago, 1603 West Taylor Street, Chicago, IL, 60612.

have received increased interest among statisticians. In this article, we describe multiple imputation (MI) under Fleishman's power polynomials ([2]) that can accommodate a wider range of distributional features. The salient aspects of the power method are elaborated in Section 2.

MI is a stochastic simulation technique that involves filling-in missing data with $m > 1$ plausible values through a predictive distribution ([3, 4]). Once m versions of the completed data sets are obtained, one can proceed with analyzing them with standard complete-data methods, and consolidating the results into a single inferential summary. As a result, with MI, uncertainty due to missing data is formally taken into account in the modeling process. Other key advantages of MI are reviewed by Rubin [3, 5] and Schafer [6, 7]. Methods and illustrative applications include [3-30]. For an extensive bibliography, see Rubin [5] and for a software review see Horton and Lipsitz [28]. The fundamental step in parametric MI is filling in the missing data by drawing from the conditional distribution of the missing data given the observed data. This usually entails positing a model for the data and using it to derive this conditional distribution. For continuous data, multivariate normality among the variables has been perceived as a natural assumption since the conditional distribution of the missing data given the observed data is then also multivariate normal. Recently, extending the practice of MI from normality to more general classes of densities has begun to receive attention (Liu [29], He and Raghunathan [30]).

Considering the restrictive nature of the normality assumption, employing a distributional setup that spans a wider range of symmetry-peakedness behavior in the imputation process may provide a reasonable way to handle non-Gaussian continuous data. In this regard, the power method can be thought as a sensible alternative because of the ability of accommodating a variety of distributional shapes depending on the choice of parameter values. Here, we explore the relative advantages of conducting imputation inferences under this more flexible approach via a limited simulation experiment that includes some univariate data generation mechanisms that may be encountered in practice. The rationale is to assess the feasibility of this technique as a possible impetus for extensions to the multivariate case, and to gauge its generalizability potential for creating imputations under a multivariate extension of the power method. Given that imputation under non-normal densities is a recently emerging notion, which has potential in many research areas, it is important to evaluate its performance in terms of commonly accepted bias and precision measures.

The organization of the rest of this paper is as follows. In Section 2, we describe essential aspects of Fleishman's power polynomials, and discuss possible augmentations that have appeared in the literature. In Section 3, we present our simulation design and give a simple algorithm to create multiply imputed data sets under power polynomials, followed by the examination of relative improvements over Gaussian imputation on incomplete data sets that exhibit different distributional characteristics. Subsequently, we explore the behavior of efficiency and accuracy measures to determine the extent to which the power procedure works properly. Section 4 includes concluding remarks, discussion and future directions in the sense of generalizing the proposed approach to the multivariate case. The technical details concerning the estimation procedure for obtaining the underlying parameters, and *R/Splus* code that implements this approach, are given in the Appendix.

2 Fleishman polynomials

Fleishman ([2]) argued that real-life distributions of variables are typically characterized by their first four moments. He presented a moment-matching procedure that simulates non-normal distributions often used in Monte Carlo studies. It is based on the polynomial transformation, $Y = a + bZ + cZ^2 + dZ^3$, where Z follows a standard normal distribution, and Y is standardized (zero mean and unit variance). The distribution of Y depends on the constants a, b, c and d, whose values were tabulated for selected values of skewness ($r_1 = E[Y^3]$) and kurtosis ($r_2 = E[Y^4]-3$). This procedure of expressing any given variable by the sum of linear combinations of powers of a standard normal variate is capable of covering a wide area in the skewness-elongation plane whose bounds are given by the general expression $r_2 \geq r_1^2 - 2$. [1] High-order moment (the third and fourth moments) boundaries of the power method were given in the original paper ([2]) through an inequality; however, they were not entirely correct. Subsequently, Headrick and Sawilowsky ([31]) computed empirical lower bounds of kurtosis for a given value of skewness.

Assuming that $E[Y] = 0$, and $E[Y^2] = 1$, by utilizing the first 12 moments of the

[1] It is trivial to prove this by Cauchy-Schwarz inequality. Furthermore, one can show that equality condition is impossible to reach, but this is immaterial for the purposes of this work.

standard normal distribution, the following set of equations can be derived after simple but tedious algebra:

$a = -c$

$b^2 + 6bd + 2c^2 + 15d^2 - 1 = 0$

$2c(b^2 + 24bd + 105d^2 + 2) - r_1 = 0$

$24[bd + c^2(1 + b^2 + 28bd) + d^2(12 + 48bd + 141c^2 + 225d^2] - r_2 = 0$

Solving these equations can be accomplished by the Newton-Raphson method, or any other plausible root-finding or non-linear optimization routine. (The specifics of the Newton-Raphson algorithm for this particular setting and an implementation in the *R/Splus* software are given in the Appendix.)

Fleishman's method has been extended in several ways in the literature. One extension utilizes the fifth-order polynomials in the spirit of controlling for higher-order moments ([33]). The other one is in regard to a multivariate version of the power method. This extension is extremely interesting due to its potential for creating multiply imputed data sets in more realistic multivariate settings ([32, 34]). The generalizability to the multivariate case makes the polynomial method more compelling in the sense that it presents an advantage over other general distributions (Burr family [35], Johnson family [36], Pearson family [37], Schmeiser-Deutch system [38]) whose multivariate versions are either non-existent or very formidable to specify due to mathematical and/or computational difficulties. The scope of this paper is limited to the univariate case. For this reason, we only briefly mention in Section 4, the operational logic of the multivariate power approach and discuss its connection to the Bayesian multiple imputation technique under the assumption of multivariate normality. A convenience of adapting Fleishman's method to MI is that it allows one to take advantage of well-developed MI methods. In other words, employing suitable transformations makes it possible for practitioners to use existing MI software ([8, 39, 40, 41]). It should be noted that the power approach has been criticized by some authors ([42]) on the grounds that the exact distribution was unknown and thus lacked probability density and cumulative distribution functions (pdf and cdf, respectively). However, Headrick and Kowalchuk ([43]) recently derived the power method's pdf and cdf in general form.

After reviewing the fundamentals of the power approach, we describe a simulation study in the next section. For this, we describe how one creates multiply imputed data sets under power polynomials with competing moment structures.

3 Simulation design and imputation algorithm

Complete data generation: Complete data were generated using the generalized lambda density (GLD) which is a class of distributions used for parameter estimation, fitting distributions to data, or in simulation studies that primarily involve univariate data generation ([44, 45, 46]). The univariate GLD is attractive because its pdf and inverse distribution function are known and its associated algorithm for data generation can be implemented with relative ease. Simulated values, x, can be drawn by the inverse cdf method $x = \lambda_1 + [p^{\lambda_3} - (1-p)^{\lambda_4}]/\lambda_2$, where λ_1 and λ_2 are location and scale parameters, respectively, λ_3 and λ_4 are shape parameters, and $p \sim U([0,1]$. The right hand side of the equation is the inverse cdf of the GLD. Although the cdf does not exist in closed form, this is not a problem in practice since the same is true for the normal distribution. Ramberg and Schmeiser ([45]) showed that the k^{th} moment ($\lambda_1 = 0$) of the GLD, when it exists, is given by $E(X^k) = \lambda_2^{-k} \sum_{i=0}^{k} \binom{k}{i}(-1)^i \beta(\lambda_3(k-i)+1, \lambda_4 i + 1)$, where β denotes the beta function. Ramberg et al. ([46]) gave details of estimation and model-fitting procedures for this four-parameter probability distribution. It involves solving a set of equations that are formed through the first four moments that are capable of accommodating a wide variety of curve shapes. The specification of any feasible moment structure (again, not every combination is possible) that translates to corresponding values of λ's, enables us to generate random numbers via the inverse cdf method. The reason we use the GLD is that it covers about the same area in the skewness-elongation plane as the power method does. Another advantage of the use of the GLD is that it allows us to make a genuine assessment as to how Fleishman's method performs in the imputation context because of the fact that the GLD and the power method represent radically different distributional forms. The nine skewness-kurtosis pairs that were used in the simulated examples are given in Table 1. The number of observations, n, in the complete data set was chosen to be 100, 500, and 1000.

Table 1 goes here

Missingness mechanism: Missing values were assumed to be missing completely at random (MCAR). The nonresponse rate was chosen to be 25%, 50%, and 75%.

Imputation algorithm: We assume two imputation models for comparison purposes for each of the incomplete data sets generated. The first one is the normal model, where we create imputations following the standard approach of using a Bayesian predictive model of the missing data given the observed data ([6]). For the other (power polynomials), instead of adopting a Bayesian approach, we account for the parameter uncertainty by obtaining nonparametric bootstrap samples that anchor the subsequent estimation procedure for the parameters of the power expression, as was done by He and Raghunathan [30]. Denoting the data $Y = (Y_{obs}, Y_{mis}) = (y_1, y_2, ..., y_{n_1}, y_{n_1+1}, ..., y_n)^T$, of which the first n_1 elements are observed ($Y_{obs} = (y_1, y_2, ..., y_{n_1})^T$), and the remaining $n - n_1$ elements are missing ($Y_{mis} = y_{n_1+1}, ..., y_n)^T$), the imputation algorithm is as follows:

1. Center and scale the data so that mean is zero and variance is one. This standardization is needed for the subsequent estimation of the four power polynomial coefficients. Let the transformed data be Y^*_{obs}.

2. Draw a nonparametric bootstrap sample of size n_1 from Y^*_{obs}.

3. Estimate the model parameters (a, b, c, d) using the Newton-Raphson algorithm given in the Appendix.[2]

4. Simulate independent variates from these distributions for every missing data point in Y^*_{mis}.

5. Back transform the filled-in data and the transformed observed data to the original scale.

6. Repeat steps 2-5 independently $m = 10$ times.

Parameters of interest: We compared the relative performances of normal and Fleishman imputations on five quantiles (5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th}) that are known to be sensitive

[2]There is nothing magical about the Newton-Raphson method, one can use any other plausible root-finding algorithm.

to model misspecification.

Evaluation criteria: The simulation experiment was repeated $N = 500$ times for each of the $9 \times 3 \times 3 = 81$ scenarios (combinations of complete data distributions, data set sizes, and missingness rates, respectively). Obviously, N could have been chosen to be larger; however, with too many replications, the bias could turn out to be significant when it is actually not. Evaluation is conducted based on three quantities: a) *Standardized bias (SB)* is the relative magnitude of the raw bias to the overall uncertainty in the system. If the parameter of interest is θ, the standardized bias is $100 \times \frac{E(\hat{\theta})-\theta}{SE(\hat{\theta})}$, where SE stands for standard error. If the standardized bias exceeds 50% in a positive or negative direction, then the bias begins to have a noticeable adverse impact on efficiency, coverage and error rates (Demirtas [22]). b) *Coverage rate (CR)* is the percentage of times that the true parameter value is covered in the confidence interval. If a procedure is working well, the actual coverage should be close to the nominal rate (i.e. Type I error rates are properly controlled). We regard the performance of the interval procedure to be poor if its coverage drops below 90% (Collins et al. [12]). c) *Root-mean-square error (RMSE)* is an integrated measure of bias and variance. It is considered to be arguably the best criterion for evaluating $\hat{\theta}$ in terms of combined accuracy and precision. $RMSE(\hat{\theta})$ is defined as $\sqrt{E_{\theta}[(\hat{\theta}-\theta)^2]}$. Under this specification, SB is a pure accuracy measure, CR and $RMSE$ are the hybrid measures of accuracy and precision.

3.1 Results

Since the results across different sample sizes and nonresponse rates yielded little or no discernible differences, we present the results for $n = 1000$ and 50% nonresponse rate due to space limitations. In Tables 2, 3, and 4, we tabulate the average estimate (AE), SB, $RMSE$, and CR for the five quantiles under consideration across 500 simulation replicates for both the normal imputation model and the proposed power imputation. In these tables, the bias and coverage quantities that do not fall into the acceptable ranges ($>$ 50% for SB, and $<$ 90.0 for CR) are denoted with bold characters. The number of significant digits varies depending on the measures. A close examination of Tables 2-4 reveals that Fleishman's method outperforms the normal imputation method in terms of all three evaluation criteria to varying degrees across all scenarios except when both skewness and kurtosis are equal to zero. In this case, which corresponds to normal underlying data, the performances are comparable. Fleishman's method yields remarkable results with no exceptions as indicated by negligible biases and high coverage rates. In other words, the performance of the power method turns out to be superior *compared to* the MI normal model in terms of commonly accepted accuracy and precision quantities in an overwhelming majority of cases, with decent properties *in absolute terms*. This gives us hope in regard to the more general multivariate case which we briefly discuss in the next section.

Table 2, 3 and 4 go here

4 Discussion

We enthusiastically believe that an intimate connection between random number generation and MI can be established (see Demirtas and Hedeker [26] for an example). In this paper we have adapted methods developed in the random number generation literature to the context of MI. A major imputation principle is not to distort the marginal distributions and associations between observed and imputed variables; and random number generation is conducted via specified distributional properties. Observed data trends can be assumed to be applicable to the whole data set, and missing portions can be filled in with numbers that belong to the same distributional mechanism which, in a sense, is the operational logic for random number generation. This assertion clearly assumes ignorable nonresponse (once we have taken into account what we have observed, there remains no dependence on what we have not observed), where missingness fully depends on the observed quantities in the system in the sense of Rubin ([47]). Although this may be considered a limitation, it might serve as a milestone for nonignorable extensions.

In our view, the promising results in the univariate case substantiate the natural continuation of this imputation method under the multivariate case. We now describe its main characteristics. As articulated in Vale and Maurelli ([32]), one can compute estimated power coefficients marginally for each variable and correlations among them. Then, by a principle components factorization or another factorization method, an identity that involves powers of the correlation between pairs of standard normal variables and between pairs of the original variables could be obtained. Solving this third-order equation for each pair of standard normal variables, along with marginal coefficients, yields the set of parameters of a multivariate normal distribution. After the generation of the multivariate normal data matrix, it is back-transformed to the original scale. While this approach was suggested in the context of random number generation, it can easily be implemented for incomplete data problems in the following way. After computing the marginal coefficients and correlations for the observed part of the data, a conversion to the multivariate normal case can be done in a straightforward way. Subsequently, one may resort to well-developed Bayesian imputation techniques ([6]) under the normal model. Once multiply imputed data sets are created, a back transformation would result in completed data sets that preserve the original features of the data.

There are a few limitations that need to be addressed. First, while we recognize that real incomplete data often include many variables, our focus was on univariate data. We view this as a potential building block for more realistic situations. The behavior of the third and fourth moments typically requires more modeling flexibility in terms of the area covered in the symmetry-elongation plane as well as the association among variables. This work serves as an initial feasibility study for assessing the generalizability potential to the multivariate settings. On a related note, although the power approach is capable of picking some data trends that are unlikely to be captured by a normal model, it does not cover the entire symmetry-elongation plane. Nevertheless, considering the relative gains presented in this paper, it provides an indication that the multivariate version can lead to further improvements. Furthermore, the assumed missingness mechanism (MCAR) is generally too simplistic for real-life applications. However, our purpose was not to conduct a sensitivity analysis with respect to the mechanism that leads to the observed data. Rather, the current paper was motivated by how tenably MI inferences can be conducted with power polynomials. Finally, our simulation setup needed to be in manageable limits and does not span every imaginable scenario that may arise in practice. However, in our opinion, it is sufficiently comprehensive to demonstrate the superiority of the Fleishman imputation model in most cases.

The assumption of multivariate normality along with the Bayesian paradigm has often been regarded as a statistically defensible way of creating multiply imputed data sets for continuous data. While it is a convenient assumption and it has been shown to work well in some settings (e.g., with a large number of subjects), it is constructive to move the practice of MI to other distributions that cover a broader range of the third and fourth moments. In an attempt to go beyond the realm of normality to adequately model distributional properties that are not accommodated by a Gaussian model, there has been a growing interest in non-normal distributions. This work was motivated by the premise that the MI framework may be amenable to Fleishman's method. As mentioned before, forming Bayesian predictive distributions under the power approach is not a formidable task in a complex multivariate setting by applying a suitable transformation to normal variates, which makes the power method a potentially fruitful future research area.

5 Appendix

Here, an *R/Splus* routine is included to find the parameters of Fleishman polynomials. In this particular setting, the input data argument should be the observed data. Once the parameters are estimated, one can easily generate the complete data by imputing under these polynomials. Note that the parameters are estimated under the assumption that the mean is 0, and the standard deviation is 1; the resulting data set should be back-transformed to the original scale by multiplying every data point by the standard deviation and adding the observed data mean. Since $a = -c$, it comes down to solving the following equations:

$$g = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} b^2 + 6bd + 2c^2 + 15d^2 - 1 \\ 2c(b^2 + 24bd + 105d^2 + 2) - r_1 \\ 24[bd + c^2(1 + b^2 + 28bd) + d^2(12 + 48bd + 141c^2 + 225d^2] - r_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The first derivative matrix,

$$H = \begin{bmatrix} g_1'(b) & g_1'(c) & g_1'(d) \\ g_2'(b) & g_2'(c) & g_2'(d) \\ g_3'(b) & g_3'(c) & g_3'(d) \end{bmatrix}$$

, where

$g_1'(b) = 2b + 6d$, $g_1'(c) = 4c$, $g_1'(d) = 6b + 30d$

$g_2'(b) = 2c(2b + 24d)$, $g_2'(c) = 2(b^2 + 24bd + 105d^2 + 2)$, $g_2'(d) = 2c(24b + 210d)$

$g_3^{'}(b) = 24(d + 2bc^2 + 28c^2d + 48d^3)$ $g_3^{'}(c) = 24(2c + 2b^2c + 56bcd + 282cd^2)$
$g_3^{'}(d) = 24(b + 28bc^2 + 24d + 144bd^2 + 282c^2d + 900d^3)$

Updating equations in Newton-Raphson are

$$\begin{bmatrix} b^{(t+1)} \\ c^{(t+1)} \\ d^{(t+1)} \end{bmatrix} = \begin{bmatrix} b^{(t)} \\ c^{(t)} \\ d^{(t)} \end{bmatrix} - H^{-1}g$$

```
################################################################
# This R code finds four estimated coefficients in equation
# Y=a+bX+cX^2+dX^3 as appeared on Fleishman (1978)
# data = Data
# skewkurt = specified skewness and kurtosis
#            This is a list with two components.
# n.iter = maximum number of iterations
# Notation:
# r1 = Skewness    r2 = Kurtosis
# Restriction: Kurtosis > Skewness^2 - 2
# If skewness and kurtosis do not satisfy the general condition
# above, the function power.method() will give a warning message.
# Note that the power method does not cover the entire
# symmetry-elongation plane, and if the skewness-kurtosis
# specification does not fall in the feasible area, results will
# be unobtainable or unreliable. See Headrick and Sawilowsky (2000)
# for lower bounds of kurtosis for a given value of skewness.
##################################################################
# In what follows, notation on Pages 10-11 should be taken as a basis.
# The AUXILIARY FUNCTION that finds the derivative matrix
g.1st.der<-function(b0,c0,d0){
      b<-b0    # initial value of b
      c<-c0    # initial value of c
      d<-d0    # initial value of d
       g1.b<-2*b+6*d
       g1.c<-4*c
       g1.d<-6*b+30*d
       g2.b<-2*c*(2*b+24*d)
       g2.c<-2*(b^2+24*b*d+105*d^2+2)
       g2.d<-2*c*(24*b+210*d)
       g3.b<-24*(d+2*b*c^2+28*c^2*d+48*d^3)
       g3.c<-24*(2*c+2*b^2*c+56*b*c*d+282*c*d^2)
       g3.d<-24*(b+28*b*c^2+24*d+144*b*d^2+282*c^2*d+900*d^3)
H<-matrix(0,3,3)
H[1,1]<-g1.b; H[1,2]<-g1.c; H[1,3]<-g1.d
H[2,1]<-g2.b; H[2,2]<-g2.c; H[2,3]<-g2.d
H[3,1]<-g3.b; H[3,2]<-g3.c; H[3,3]<-g3.d
return(H)}
```

```
#################################################################
# The MAIN FUNCTION that finds the estimated coefficients
# using the Newton-Raphson algorithm. The data set or the
# skewness/kurtosis can be supplied as an argument.
power.method<-function(data,skewkurt,n.iter=500){
if(missing(skewkurt)){
y<-(data-mean(data))/sd(data)   # Centering and scaling
r1<-mean(y^3) ; r2<-mean(y^4)-3}
if(!missing(skewkurt)){
r1<-skewkurt$skew
r2<-skewkurt$kurtosis}
# Restriction
if (r2<=r1^2-2) #Unfeasible skewness-elongation plane
cat("WARNING: Computed a,b,c,d may not be correct.\n")
cat("skewness =",r1,"\n")
cat("kurtosis =",r2,"\n")
cat("Estimated four constants from power method are\n")
## NEWTON-RAPHSON ALGORITHM
      i<-0
      coeff<-matrix(0,n.iter,3)
      b<-1.001 ; c<-0.001 ; d<-0.001   # initial value of b, c and d
      g1<-b^2+6*b*d+2*c^2+15*d^2-1
      g2<-2*c*(b^2+24*b*d+105*d^2+2)-r1
      g3<-24*(b*d+c^2*(1+b^2+28*b*d)+d^2*(12+48*b*d+141*c^2+225*d^2))-r2
      coeff[1,]<-c(b,c,d)
      g<-c(g1,g2,g3)
      tolerance<-10^(-15)
      diff<-c(0,0,0)
          while(i<= n.iter && sum(diff)<3){
          i<-i+1
# Use Singular value decomposition to obtain pseudo-inverse
  s<-svd(g.1st.der(b,c,d))
  coeff[i+1,]<-coeff[i,]-t(s$v%*%diag(1/s$d)%*% t(s$u)%*%g)
# Update estimates
          b<-coeff[i+1,1]; c<-coeff[i+1,2]; d<-coeff[i+1,3]
# Compute three equations using updated estimates
      g1<-b^2+6*b*d+2*c^2+15*d^2-1
      g2<-2*c*(b^2+24*b*d+105*d^2+2)-r1
      g3<-24*(b*d+c^2*(1+b^2+28*b*d)+d^2*(12+48*b*d+141*c^2+225*d^2))-r2
      g<-c(g1,g2,g3)
# Check the convergence
            diff[abs(g1)<tolerance]<-1
            diff[abs(g2)<tolerance]<-1
            diff[abs(g3)<tolerance]<-1}
      if (i<=n.iter) cat("a =",-c,"b =", b,"c =", c,"d =", d,"\n")
      else cat ("Failed to converge in", n.iter,"iterations\n")}
###################################################################
```

References

[1] Genton MG. (Ed) *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman and Hall/CRC, 2004.

[2] Fleishman AI. A method for simulating non-normal distributions. *Psychometrika* 1978; **43**, 521–532.

[3] Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley Classic Library, 2004.

[4] Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Second Edition. New York: Wiley, 2002.

[5] Rubin DB. Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association* 1996; **91**:473-520.

[6] Schafer JL. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall, 1997.

[7] Schafer JL. Multiple imputation: a primer. *Statistical Methods in Medical Research* 1999; **8**, 3–15.

[8] Schafer JL. *NORM: Multiple Imputation of Incomplete Multivariate Data Under a Normal Model, Software Library for S-PLUS*. University Park, PA: The Pennsylvania State University, Department of Statistics, 1999.

[9] Schafer JL, Graham, JW. Missing data: our view of the state of the art. *Psychological Methods* 2002; **7**: 147–177.

[10] Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Statistical Methods in Medical Research* 1999; **8**: 17–36.

[11] Schafer JL, Schenker N. Inference with imputed conditional means. *Journal of American Statistical Association* 2000; **95**: 144–154.

[12] Collins LM, Schafer JL, Kam CH. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* 2001; **6**: 330–351.

[13] van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* 1999; **18**: 681-694.

[14] Raghunathan TE, Lepkowski JM, van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* 2001, **27**, 85–95.

[15] Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research* 1998; **33**, 545–571.

[16] Lavori PW, Dawson R, Shera, D. A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in Medicine* 1995; **14**: 1913-1925.

[17] Allison, PD. Multiple imputation for missing data: A cautionary tale. *Sociological Methods and Research* 2000; **28**: 301–309.

[18] Schafer JL, Yucel RM. Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics* 2002; **11**: 437-457.

[19] Liu M, Taylor JM, Belin TR. Multiple imputation and posterior simulation for multivariate missing data in longitudinal studies. *Biometrics* 2000; **56**: 1157-63.

[20] Belin TR, Hu MY, Young AS, Grusky O. Using multiple imputation to incorporate cases with missing items in a mental health services study. *Health Services and Outcome Research Methodology* 2000; **1**: 7–22.

[21] Demirtas H, Schafer JL. On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine* 2003; **22**: 2553-2575.

[22] Demirtas, H. Simulation-driven inferences for multiply imputed longitudinal datasets. *Statistica Neerlandica* 2004; **58**: 466-482.

[23] Demirtas, H. Modeling incomplete longitudinal data. *Journal of Modern Applied Statistical Methods* 2004, **2**: 305–321.

[24] Demirtas H. Bayesian analysis of hierarchical pattern-mixture models for clinical trials data with attrition and comparisons to commonly used ad-hoc and model-based approaches. *Journal of Biopharmaceutical Statistics* 2005; **25**: 383-402.

[25] Demirtas H. Multiple imputation under Bayesianly smoothed pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine* 2005; **24**: 2345-2363.

[26] Demirtas H, Hedeker, D. Gaussianization-based quasi-imputation and expansion strategies for incomplete correlated binary responses. *Statistics in Medicine* (in press).

[27] Demirtas H, Freels SA, Yucel RM. Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: A simulation assessment. *Journal of Statistical Computation and Simulation* (in press).

[28] Horton JH, Lipsitz, SR. Multiple imputation in practice: Comparison of software packages for regression models with missing variables. *The American Statistician* 2001; **55**: 244–254.

[29] Liu C. Missing data imputation using the multivariate t distribution. *Journal of Multivariate Analysis* 1995; **53**, 139–158.

[30] He Y, Raghunathan, TE. Tukey's gh distribution for multiple imputation. *The American Statistician* 2006; **60**: 251–256.

[31] Headrick TC, Sawilowsky SS. Weighted simplex procedures for determining boundary points and constants for the univariate and multivariate power methods. *Journal of Educational and Behavioral Statistics* 2000; **25**, 417–436.

[32] Vale CD, Maurelli VA. Simulating multivariate nonnormal distributions. *Psychometrika* 1983; **48**, 465–471.

[33] Headrick TC. Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions. *Computational Statistics and Data Analysis* 2002; **40**, 685-711.

[34] Headrick TC, Sawilowsky SS. Simulating correlated multivariate nonnormal distributions. *Psychometrika* 1999; **64**, 25–35.

[35] Burr IW. Cumulative frequency functions. *Annals of Mathematical Statistics* 1942; **13**, 215–232.

[36] Johnson NL. Systems of frequency curves generated by methods of translation. *Biometrika* 1949; **36**, 149–176.

[37] Parrish RS. Generating random deviates from multivariate Pearson distributions. *Computational Statistics and Data Analysis* 1990; **9**, 283–295.

[38] Schmeiser BW, Deutch SJ. A versatile four parameter family of probability distributions suitable for simulation. *AIIE Transactions* 1977; **9**, 176–182.

[39] SAS/Stat User's Guide, Version 8.2, North Carolina, 2001.

[40] R Development Core Team. R: A language and environment for statistical computing, Version 2.3.0, Vienna, Austria, 2006.

[41] Schimert J, Schafer JL, Hesterberg T, Fraley C, Clarkson DB. *Analyzing Data with Missing Values in S-plus*. Seattle, WA: Data Analysis Products Division, Insightful Corp, 2001.

[42] Tadikamalla PR. On simulating non-normal distributions. *Psychometrika* 1980; **45**, 273–279.

[43] Headrick TC, Kowalchuk RK. The power method transformation: its probability density function, distribution function, and its further use for fitting data. *Journal of Statistical Computation and Simulation* 2007; **78**, n/a–n/a (in press).

[44] Ramberg JS, Schmeiser BW. An approximate method for generating symmetric random variables. *Communications of the ACM* 1972; **15**, 987–990.

[45] Ramberg JS, Schmeiser BW. An approximate method for generating asymmetric random variables. *Communications of the ACM* 1974; **17**, 78–82.

[46] Ramberg JS, Dudewicz EJ, Tadikamalla PR, Mykytka EF. A probability distribution and its uses in fitting data. *Technometrics* 1979; **21**, 201–214.

[47] Rubin DB. Inference and missing data. *Biometrika* 1976; **21**, 581–592.

Table 1: The skewness-kurtosis specifications with the prefixes meso, lepto, and platy stand for zero, positive, and negative kurtosis, respectively.

Skewness	Kurtosis	Property
0	0	symmetric-mesokurtic
0	3.75	symmetric-leptokurtic
0	-1	symmetric-platykurtic
0.5	0	right skewed-mesokurtic
-0.5	0	left skewed-mesokurtic
0.25	3	right skewed-leptokurtic
-0.25	3	left skewed-leptokurtic
0.75	-0.25	right skewed-platykurtic
-0.75	-0.25	left skewed-platykurtic

Table 2: The performance of imputation inferences under Fleishman's power polynomials and normal model for the five quantiles. AE, SB, $RMSE$, and CR stand for the average estimate, standardized bias, root-mean-square error, and coverage rate, respectively, for the three skewness-kurtosis specifications. The number of simulation replicates, N is 500; the length of the complete data vector, n is 1000; and missingness rate is 50%.

MI model	Skewness	Kurtosis	Quantile	AE	SB	RMSE	CR
			5	0.0499562	-0.70	0.0063	96.4
			25	0.2500195	0.15	0.0128	96.2
FLEISHMAN	0	0	50	0.5004404	2.96	0.0149	96.6
			75	0.7503166	2.39	0.0133	94.6
			95	0.9505163	8.80	0.0059	96.0
			5	0.0500160	0.27	0.0060	96.4
			25	0.2497439	-2.07	0.0123	96.8
NORMAL	0	0	50	0.5003455	2.47	0.0140	96.0
			75	0.7504248	3.34	0.0127	95.2
			95	0.9501423	2.51	0.0057	97.0
			5	0.0502699	4.12	0.0065	96.6
			25	0.2538158	28.84	0.0137	95.8
FLEISHMAN	0	3.75	50	0.4997814	-1.35	0.0162	96.2
			75	0.7458029	-32.50	0.0136	95.6
			95	0.9497558	-3.69	0.0066	96.8
			5	0.0537280	**58.21**	0.0074	94.6
			25	0.2742194	**203.09**	0.0270	**60.8**
NORMAL	0	3.75	50	0.4997814	-0.77	0.0132	97.2
			75	0.7253687	**-198.89**	0.0276	**61.0**
			95	0.9461702	**-57.60**	0.0077	93.6
			5	0.0498200	-2.76	0.0065	96.6
			25	0.2513454	9.88	0.0137	95.2
FLEISHMAN	0	-1	50	0.5006417	4.20	0.0153	96.2
			75	0.7505817	4.53	0.0128	96.8
			95	0.9509802	15.20	0.0065	95.4
			5	0.0509661	17.21	0.0057	98.4
			25	0.2323139	**-144.56**	0.0215	**74.2**
NORMAL	0	-1	50	0.5007650	5.34	0.0143	96.0
			75	0.7696790	**173.66**	0.0227	**68.8**
			95	0.0497664	-4.43	0.0053	98.2

Table 3: The performance of imputation inferences under Fleishman's power polynomials and normal model for the five quantiles. AE, SB, $RMSE$, and CR stand for the average estimate, standardized bias, root-mean-square error, and coverage rate, respectively, for the three skewness-kurtosis specifications. The number of simulation replicates, N is 500; the length of the complete data vector, n is 1000; and missingness rate is 50%.

MI model	Skewness	Kurtosis	Quantile	AE	SB	RMSE	CR
			5	0.0497666	-3.84	0.0061	96.4
			25	0.2498870	-0.86	0.0132	94.6
FLEISHMAN	0.5	0	50	0.4991600	-5.35	0.0157	94.6
			75	0.7491879	-5.90	0.0138	94.4
			95	0.9498987	-1.61	0.0063	95.0
			5	0.0598241	**180.99**	0.0112	**82.4**
			25	0.2385139	**-100.52**	0.0162	**89.6**
NORMAL	0.5	0	50	0.4808458	**-132.78**	0.0240	**78.2**
			75	0.7447645	-38.74	0.0145	93.4
			95	0.9566509	**114.21**	0.0088	**78.4**
			5	0.0499155	-1.35	0.0063	95.8
			25	0.2499144	-0.67	0.0127	96.8
FLEISHMAN	-0.5	0	50	0.4994420	-3.88	0.0144	97.2
			75	0.7496098	-2.98	0.0131	96.0
			95	0.9504249	6.48	0.0066	95.2
			5	0.0432604	**-115.66**	0.0089	**76.8**
			25	0.2545417	34.81	0.0138	92.8
NORMAL	-0.5	0	50	0.5183019	**132.77**	0.0229	**82.6**
			75	0.7616396	**101.83**	0.0163	**87.4**
			95	0.9402742	**-167.59**	0.0113	**82.0**
			5	0.0502682	3.83	0.0070	95.4
			25	0.2530172	22.50	0.0160	94.6
FLEISHMAN	0.25	3	50	0.5000550	0.34	0.0160	95.2
			75	0.7471826	-21.34	0.0135	95.0
			95	0.9502364	3.74	0.0063	96.2
			5	0.0561272	**90.97**	0.0091	90.2
			25	0.2679351	**144.67**	0.0218	**79.0**
NORMAL	0.25	3	50	0.4947805	-38.14	0.0146	95.4
			75	0.7272889	**-178.19**	0.0260	**64.8**
			95	0.9499097	-1.45	0.0062	96.6

Table 4: The performance of imputation inferences under Fleishman's power polynomials and normal model for the five quantiles. AE, SB, $RMSE$, and CR stand for the average estimate, standardized bias, root-mean-square error, and coverage rate, respectively, for the three skewness-kurtosis specifications. The number of simulation replicates, N is 500; the length of the complete data vector, n is 1000; and missingness rate is 50%.

MI model	Skewness	Kurtosis	Quantile	AE	SB	RMSE	CR
			5	0.0493868	-9.61	0.0064	96.8
			25	0.2522979	17.86	0.0130	97.4
FLEISHMAN	-0.25	3	50	0.4995691	-2.75	0.0157	95.8
			75	0.7468822	-24.13	0.0133	96.2
			95	0.9499511	-0.71	0.0068	95.8
			5	0.0497283	-4.20	0.0065	95.4
			25	0.2724619	**177.26**	0.0258	**66.4**
NORMAL	-0.25	3	50	0.5049756	35.37	0.0149	96.2
			75	0.7319635	**-144.39**	0.0219	**75.6**
			95	0.9442302	**-86.29**	0.0088	90.4
			5	0.0563114	40.97	0.0166	90.0
			25	0.2530365	22.27	0.0139	94.8
FLEISHMAN	0.75	-0.25	50	0.5007446	4.93	0.0151	97.2
			75	0.7493247	-5.16	0.0131	94.4
			95	0.9490807	-13.66	0.0068	95.2
			5	0.0841698	**632.25**	0.0346	**0.2**
			25	0.2224906	**-262.32**	0.0294	**41.0**
NORMAL	0.75	-0.25	50	0.4587111	**-292.79**	0.0436	**24.8**
			75	0.7496099	-2.80	0.0139	93.2
			95	0.9620676	**225.43**	0.0131	**37.8**
			5	0.0510310	15.63	0.0067	96.0
			25	0.2510702	8.05	0.0133	95.8
FLEISHMAN	-0.75	-0.25	50	0.4992120	-5.37	0.0147	96.4
			75	0.7466113	-25.92	0.0135	95.4
			95	0.9434810	-42.86	0.0165	90.2
			5	0.0380095	**-223.62**	0.0131	**39.6**
			25	0.2508764	6.41	0.0137	93.6
NORMAL	-0.75	-0.25	50	0.5415639	**305.81**	0.0437	**21.4**
			75	0.7774824	**266.93**	0.0293	**40.6**
			95	0.9160428	**-623.68**	0.0344	**0.0**