PSYCHOMETRIKA—VOL. 77, NO. 4, 803–826
OCTOBER 2012
DOI: 10.1007/S11336-012-9282-4

ROBUST STRUCTURAL EQUATION MODELING WITH MISSING DATA AND AUXILIARY VARIABLES

KE-HAI YUAN AND ZHIYONG ZHANG

UNIVERSITY OF NOTRE DAME

The paper develops a two-stage robust procedure for structural equation modeling (SEM) and an R package `rsem` to facilitate the use of the procedure by applied researchers. In the first stage, M-estimates of the saturated mean vector and covariance matrix of all variables are obtained. Those corresponding to the substantive variables are then fitted to the structural model in the second stage. A sandwich-type covariance matrix is used to obtain consistent standard errors (SE) of the structural parameter estimates. Rescaled, adjusted as well as corrected and F-statistics are proposed for overall model evaluation. Using R and EQS, the R package `rsem` combines the two stages and generates all the test statistics and consistent SEs. Following the robust analysis, multiple model fit indices and standardized solutions are provided in the corresponding output of EQS. An example with open/closed book examination data illustrates the proper use of the package. The method is further applied to the analysis of a data set from the National Longitudinal Survey of Youth 1997 cohort, and results show that the developed procedure not only gives a better endorsement of the substantive models but also yields estimates with uniformly smaller standard errors than the normal-distribution-based maximum likelihood.

Key words: auxiliary variables, estimating equation, missing at random, R package `rsem`, sandwich-type covariance matrix.

1. Introduction

Being capable of modeling latent variables and measurement errors simultaneously, structural equation modeling (SEM) has become one of the most popular statistical methods in social and behavioral research, where missing data are common, especially when data are collected longitudinally. Many procedures have been developed for modeling missing data in SEM, most of which are normal-distribution-based maximum likelihood (NML; see, e.g., Enders & Bandalos, 2001; Raykov, 2005). There are also a few developments accounting for the effect of nonnormality on test statistics and standard errors (SEs) associated with NML (Arminger & Sobel, 1990; Yuan & Bentler, 2000). However, NML[1] estimates (NMLE) can be very inefficient or even biased when data have heavy tails or are contaminated. Yuan & Bentler (2001) and Yuan, Marshall and Bentler (2002) provided outlines of modeling missing data in SEM and factor analysis using maximum likelihood (ML) based on the multivariate t-distribution. But their developments are limited to a rescaled statistic and they did not provide the details of implementing the procedures. Lee & Xia (2006) developed a missing data procedure for nonlinear SEM in which latent variables as well as measurement and prediction errors are symmetrically distributed. Model estimation and inference are through Monte Carlo and the Bayesian information criterion (BIC). Because both model structure and estimation method affect the value of BIC, Lee and Xia (2006, pp. 581–582) noted that the method should be used with caution. In this paper, we develop a two-stage procedure for robust SEM with missing data, where robust M-estimators of the saturated mean vector and covariance matrix are obtained in the first stage and are then fitted by

[1] Without missing values, NML is uniquely defined. With missing values, there are direct NML and 2-stage NML (see Yuan & Bentler, 2000). Unless explicitly mentioned, our discussion equally applies to both/either of them.

The research was supported by Grants DA00017 and DA01070 from the National Institute on Drug Abuse.

Requests for reprints should be sent to Ke-Hai Yuan, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556, USA. E-mail: kyuan@nd.edu

the structural model in the second stage. Model evaluation is done by using well-established test statistics or fit indices for complete data. Furthermore, we develop an R package `rsem` for the two-stage robust procedure so that applied researchers can use it when analyzing substantive data. We will also show how this procedure works by analyzing a data set from the National Longitudinal Survey of Youth 1997 (NLSY97) cohort with latent growth curve models.

Missing data can occur for many reasons. The process by which data become incomplete was called a missing data mechanism by Rubin (1976). Missing completely at random (MCAR) is a process in which missingness of data is independent of both the observed and the missing values. Missing at random (MAR) is a process in which missingness is independent of the missing values given the observed data. When missingness depends on the missing values themselves given the observed data, the process is missing not at random (MNAR). When ignoring the MAR or MCAR mechanism, ML estimates (MLEs) are still consistent. When missing values are MNAR, one has to correctly model the missing data process in order to get consistent parameter estimates in general. However, the MAR or MNAR mechanism depends on whether variables accounting for missingness are observed and included in the estimation. Auxiliary variables are those that are not directly involved in the structural model but have the potential to account for missingness in the substantive variables (Enders, 2010, pp. 127–163). Our procedure aims for SEM with missing data that are MAR after including potential auxiliary variables. In particular, auxiliary variables can be easily included in the first-stage robust M-estimation. Parallel to the procedures in Yuan & Lu (2008) and Savalei & Bentler (2009), only estimates of means and covariances corresponding to the substantive variables are selected and fitted by the structural model in the second-stage analysis.

With complete data, we can use existing procedures to check the distributional properties of the sample before choosing a parametric model (e.g., D'Agostino, Belanger & D'Agostino, 1990). With missing data, especially when missing values are MAR, the observed data can be skewed and possess excess kurtosis even when the underlying population is normally distributed. Similarly, when the population is non-normally distributed, the observed data may easily pass a test for normality due to MAR missing data mechanism (see, e.g., Yuan, Lambert & Fouladi, 2004b). Thus, we have to rely on the robust properties of the selected method in data analysis with missing values. In the context of complete data it has been shown that NMLEs suffer from severe biases when outliers or data contamination exists (see, e.g., Zu & Yuan, 2010). We do not expect the biases to disappear when a sample also contains missing values. In addition to biases, efficiency is also a key consideration in choosing a proper statistical method. The efficiency of NMLEs goes to zero as the kurtosis of the population increases. Compared to NML, a robust procedure typically yields much less biased estimates when outliers or data contamination exists. Robust estimates are also a lot more efficient with practical data typically having heavy tails (Zhong & Yuan, 2011).

The difference between NML and a robust procedure is in how each observation is treated in the estimation process. In NML, all observations are treated equally. In a robust procedure, each case gets a weight according to its distance from the center of the majority of the data. Cases far away from the center get smaller weights. Many weight functions can be used for such a purpose. In our implementation, we use the Huber-type weight function because it tends to yield more efficient parameter estimates than other weight functions for SEM with real complete data (Yuan, Bentler & Chan, 2004a). The tuning parameter in the Huber-type weight function is also explicitly related to the percentage of contaminated data or outliers that one would like to control. With robust estimates of means and variances–covariances, we also need an estimate of their asymptotic covariance matrix. This covariance matrix is a key element to obtain consistent standard errors (SEs) and reliable test statistics for overall model evaluation. The size of this matrix can be very large, and it is already a challenge for its estimate to be positive definite even without any missing data. Another consideration behind choosing the Huber-type weight function is that it does not assign zero weights to cases. When many cases get zero weights, it is very likely that the resulting estimate of the asymptotic covariance matrix is close to singular, then SEs and test statistics following from using such a covariance matrix become unreliable.

Robust estimates of means and covariances with missing values are studied by Little (1988) and Liu (1997) using a multivariate t-distribution. Cheng & Victoria-Feser (2002) provide an algorithm for obtaining minimum covariance determinant (MCD) estimates. Yuan (2011) extends M-estimators of means and covariances to samples containing missing values using estimating equations, and showed that these equations can be solved by an expectation robust (ER) algorithm. These robust estimates are parallel to the sample means and covariances, and provide the building blocks for robust SEM. However, it is technically a lot more involved to utilize these building blocks for robust SEM than the development of NML using the sample means and covariances. Existing development in this direction is Yuan & Bentler (2001) and Yuan et al. (2002), where a rescaled statistic is proposed for overall model evaluation, using robust estimates of means and covariances based on a multivariate t-distribution. However, the exact or asymptotic distribution of the rescaled statistic is unknown in general and other alternatives are available (Bentler & Yuan, 1999). As mentioned earlier, Huber-type weights tend to yield more efficient estimates with real complete data. To our knowledge, with missing data, there does not exist any development for using the M-estimates of means and covariance matrix based on Huber-type weight functions. The methodological contribution of this paper is to develop a robust SEM procedure with missing data using the Huber-type M-estimates of means and covariances. In particular, in addition to the rescaled statistic, we propose using an adjusted statistic, a corrected residual-based statistic, and a related F-statistic for overall model evaluation. These statistics have been shown to have either theoretical or practical advantages over the rescaled statistic in other contexts of SEM (Yuan & Bentler, 2010; Bentler & Yuan, 1999), and some of them have been implemented in software (e.g., EQS, Mplus) with the NML methodology. The novelty of the development is to use them in the context of robust SEM with missing data. Because the development is very technical, applied researchers will not be able to use the method if we just present the results with examples. Another contribution of the paper is to develop an easy-to-use R package `rsem` that implements the two-stage robust procedure. In particular, for any missing data with or without auxiliary variables, the package `rsem` can generate the standard EQS output (Bentler, 2008) that contains sound statistics for overall model evaluation, consistent SEs for structural parameter estimates, multiple fit-indices and standardized solutions.

Section 2 describes an expectation robust algorithm for obtaining M-estimators of means and covariances as well as a formula for evaluating their asymptotic covariance matrix. Section 3 contains the development of the second-stage analysis using the normal-distribution-based discrepancy function and the associated adjusted, rescaled, and residual-based statistics. Section 4 introduces the R package `rsem` and illustrates its use with EQS 6.1 by the test score data of Mardia, Kent and Bibby (1979). Section 5 presents the analysis of the NLSY97 data by comparing the results from the robust method against those from 2-stage NML. Section 6 concludes the paper with discussions.

2. M-estimates of the Saturated Mean Vector and Covariance Matrix

This section presents an ER algorithm as given in Yuan (2011). We also provide an asymptotic formula for estimating the covariance matrix of the robust M-estimates. In particular, we assume auxiliary variables are available.

Let $\mathbf{y}$ represent a population of p random variables with $E(\mathbf{y}) = \boldsymbol{\mu}$ and $\mathrm{Cov}(\mathbf{y}) = \boldsymbol{\Sigma}$. We are interested in modeling $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by a structural model. A sample $\mathbf{y}_i$, $i = 1, 2, \ldots, n$, from $\mathbf{y}$

with missing values is available. In addition to the substantive variables in $\mathbf{y}$, there also exists a vector $\mathbf{u}$ of $q-p$ auxiliary variables with the associated sample realization $\mathbf{u}_i$, $i=1,2,\ldots,n$. Let $\mathbf{x}=(\mathbf{y}',\mathbf{u}')'$ with $E(\mathbf{x})=\boldsymbol{\nu}$ and $\mathrm{Cov}(\mathbf{x})=\mathbf{V}$. Due to missing values, the vector $\mathbf{x}_i=(\mathbf{y}_i',\mathbf{u}_i')'$ only contains q_i marginal observations of $\mathbf{x}$. Also, we do not know the distribution of $\mathbf{x}$, and the observations in $\mathbf{x}_i$ may contain outliers. In such a case, robust estimates of $\boldsymbol{\nu}$ and $\mathbf{V}$ are preferred to NMLEs. Let $\boldsymbol{\nu}_i$ and $\mathbf{V}_i$ be the mean vector and covariance matrix corresponding to the observed values in $\mathbf{x}_i$. Then the Mahalanobis distance between $\mathbf{x}_i$ and $\boldsymbol{\nu}_i$ is given by

$$d_i^2=d^2(\mathbf{x}_i,\boldsymbol{\nu}_i,\mathbf{V}_i)=(\mathbf{x}_i-\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(\mathbf{x}_i-\boldsymbol{\nu}_i).$$

Let $w_{i1}(d_i)$, $w_{i2}(d_i)$ and $w_{i3}(d_i)$ nonincreasing scalar functions of d_i. The estimating equations defining robust M-estimators are given by

$$\sum_{i=1}^{n} w_{i1}(d_i)\frac{\partial \boldsymbol{\nu}_i'}{\partial \boldsymbol{\nu}}\mathbf{V}_i^{-1}(\mathbf{x}_i-\boldsymbol{\nu}_i)=\mathbf{0} \tag{1}$$

and

$$\sum_{i=1}^{n}\frac{\partial\,\mathrm{vec}'(\mathbf{V}_i)}{\partial \boldsymbol{v}}\mathbf{W}_i\,\mathrm{vec}\big[w_{i2}(d_i)(\mathbf{x}_i-\boldsymbol{\nu}_i)(\mathbf{x}_i-\boldsymbol{\nu}_i)'-w_{i3}(d_i)\mathbf{V}_i\big]=\mathbf{0}, \tag{2}$$

where $\mathbf{W}_i=0.5(\mathbf{V}_i^{-1}\otimes\mathbf{V}_i^{-1})$ with $\otimes$ being the notation for the Kronecker product (Schott, 2005, p. 283), $\mathrm{vec}(\cdot)$ is the operator that transforms a matrix into a vector by stacking the columns of the matrix, and $\boldsymbol{v}=\mathrm{vech}(\mathbf{V})$ is the vector of stacking the columns in the lower-triangular part of $\mathbf{V}$. Notice that the subscript i in w_{i1}, w_{i2}, and w_{i3} is to adjust for varying number of observations in $\mathbf{x}_i$. When $w_{i1}(d_i)=w_{i2}(d_i)=w_{i3}(d_i)=1$, Equations (1) and (2) define NMLEs of $\boldsymbol{\nu}$ and $\mathbf{V}$ with missing data. When $w_{i1}(d_i)=w_{i2}(d_i)=(m+q_i)/(m+d_i^2)$ and $w_{i3}(d_i)=1$, Equations (1) and (2) define the MLEs of $\boldsymbol{\nu}$ and $\mathbf{V}$ based on the multivariate t-distribution with m degrees of freedom. When the three weight functions are chosen according to Lopuhaä (1989) or Rocke (1996), Equations (1) and (2) define S-estimators of $\boldsymbol{\nu}$ and $\mathbf{V}$ for samples with missing data. Let $0<\varphi<1$ and ρ_i be the $(1-\varphi)$-quantile corresponding to χ_{q_i}, the chi-distribution with q_i degrees of freedom. Huber-type weight functions with missing data are given by

$$w_{i1}(d_i)=\begin{cases}1 & \text{if } d_i\le\rho_i,\\ \rho_i/d_i & \text{if } d_i>\rho_i,\end{cases} \tag{3}$$

$w_{i2}(d_i)=[w_{i1}(d_i)]^2/\kappa_i$ and $w_{i3}(d_i)=1$, where κ_i is a constant defined by $E[\chi_{q_i}^2 w_{i1}^2(\chi_{q_i}^2)/\kappa_i]=q_i$ that aims to yield a consistent estimate of $\mathbf{V}$ when $\mathbf{x}\sim N(\boldsymbol{\nu},\mathbf{V})$. In using `rsem`, one only needs to specify φ, and the values of ρ_i and κ_i are functions of φ that will be automatically obtained by the package when evaluating each weight.

Equations (1) and (2) can be easily solved by an ER algorithm that consists of an (expectation) E-step and an (robust) R-step. Let $\mathbf{x}_{ic}=(\mathbf{x}_i',\mathbf{x}_{im}')'$ denote the complete data, where $\mathbf{x}_{im}$ is the vector containing the $q-q_i$ missing values. Of course, with real data the positions of missing values are not always at the end. We can perform permutations on each missing pattern so that all the missing variables are at the end before the start of each E-step, and put the expected values (including conditional variances–covariances) back to their original positions at the end of the E-step. Let $\boldsymbol{\nu}^{(j)}$ and $\mathbf{V}^{(j)}$ be the current values of $\boldsymbol{\nu}$ and $\mathbf{V}$, and $\boldsymbol{\nu}_i^{(j)}$ and $\mathbf{V}_i^{(j)}$ correspond to those of the observed $\mathbf{x}_i$ within a given missing data pattern. When $q_i<q$, we have

$$\boldsymbol{\nu}^{(j)}=\begin{pmatrix}\boldsymbol{\nu}_i^{(j)}\\ \boldsymbol{\nu}_{im}^{(j)}\end{pmatrix}\quad\text{and}\quad \mathbf{V}^{(j)}=\begin{pmatrix}\mathbf{V}_i^{(j)} & \mathbf{V}_{iom}^{(j)}\\ \mathbf{V}_{imo}^{(j)} & \mathbf{V}_{imm}^{(j)}\end{pmatrix},$$

where $\boldsymbol{\nu}_{im}^{(j)}$ corresponds to the means of $\mathbf{x}_{im}$, and $\mathbf{V}_{imm}^{(j)}$ and $\mathbf{V}_{imo}^{(j)}$ correspond to the covariances of

$\mathbf{x}_{im}$ with itself and $\mathbf{x}_i$, respectively. Let $d_i^{(j)} = d(\mathbf{x}_i, \boldsymbol{\nu}_i^{(j)}, \mathbf{V}_i^{(j)})$. The E-step of the ER algorithm obtains the weights $w_{i1}^{(j)} = w_{i1}(d_i^{(j)})$, $w_{i2}^{(j)} = w_{i2}(d_i^{(j)})$, $w_{i3}^{(j)} = w_{i3}(d_i^{(j)})$, the conditional means

$$\hat{\mathbf{x}}_{ic}^{(j)} = E_j(\mathbf{x}_{ic}|\mathbf{x}_i) = \begin{pmatrix} \mathbf{x}_i \\ \hat{\mathbf{x}}_{im}^{(j)} \end{pmatrix}, \tag{4}$$

and the conditional covariance matrix

$$\mathbf{C}_i^{(j)} = \mathrm{Cov}_j(\mathbf{x}_{ic}|\mathbf{x}_i) = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{imm}^{(j)} \end{pmatrix}, \tag{5}$$

where

$$\hat{\mathbf{x}}_{im}^{(j)} = \boldsymbol{\nu}_{im}^{(j)} + \mathbf{V}_{imo}^{(j)}\big(\mathbf{V}_i^{(j)}\big)^{-1}\big(\mathbf{x}_i - \boldsymbol{\nu}_i^{(j)}\big) \quad \text{and} \quad \mathbf{C}_{imm}^{(j)} = \mathbf{V}_{imm}^{(j)} - \mathbf{V}_{imo}^{(j)}\big(\mathbf{V}_i^{(j)}\big)^{-1}\mathbf{V}_{iom}^{(j)}.$$

The robust step is given by

$$\boldsymbol{\nu}^{(j+1)} = \frac{\sum_{i=1}^n w_{i1}^{(j)}\hat{\mathbf{x}}_{ic}^{(j)}}{\sum_{i=1}^n w_{i1}^{(j)}}, \tag{6}$$

$$\mathbf{V}^{(j+1)} = \frac{\sum_{i=1}^n [w_{i2}^{(j)}(\hat{\mathbf{x}}_{ic}^{(j)} - \boldsymbol{\nu}^{(j+1)})(\hat{\mathbf{x}}_{ic}^{(j)} - \boldsymbol{\nu}^{(j+1)})' + w_{i3}^{(j)}\mathbf{C}_i^{(j)}]}{\sum_{i=1}^n w_{i3}^{(j)}}. \tag{7}$$

The steps in (4) to (7) are repeated until convergence yields a solution to (1) and (2). For the Huber-type weight function in (3), the algorithm is implemented in the R package `rsem` to be introduced in Section 4.

Let $\hat{\boldsymbol{\nu}}$ and $\hat{\mathbf{V}}$ be the solution to (1) and (2). They play the role of sample means and covariance matrix in the second-stage analysis when estimating the structural parameters. We still need a consistent estimator of the covariance matrix of $\hat{\boldsymbol{\nu}}$ and $\hat{\boldsymbol{v}} = \mathrm{vech}(\hat{\mathbf{V}})$ to get consistent SEs of the structural parameter estimates and reliable statistics for overall model evaluation. We obtain such an estimator using a sandwich-type covariance matrix. Let $\boldsymbol{\alpha} = (\boldsymbol{\nu}', \boldsymbol{v}')'$ and

$$\mathbf{g}(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{i=1}^n \mathbf{g}_i(\boldsymbol{\alpha}),$$

where $\mathbf{g}_i(\boldsymbol{\alpha}) = (\mathbf{g}_{i1}'(\boldsymbol{\alpha}), \mathbf{g}_{i2}'(\boldsymbol{\alpha}))'$ with

$$\mathbf{g}_{i1}(\boldsymbol{\alpha}) = w_{i1}(d_i)\frac{\partial \boldsymbol{\nu}_i'}{\partial \boldsymbol{\nu}}\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i)$$

and

$$\mathbf{g}_{i2}(\boldsymbol{\alpha}) = \frac{\partial\, \mathrm{vec}'(\mathbf{V}_i)}{\partial \boldsymbol{v}}\mathbf{W}_i\, \mathrm{vec}\big[w_{i2}(d_i)(\mathbf{x}_i - \boldsymbol{\nu}_i)(\mathbf{x}_i - \boldsymbol{\nu}_i)' - w_{i3}(d_i)\mathbf{V}_i\big].$$

Under standard regularity conditions (Yuan & Jennrich, 1998), the estimators $\hat{\boldsymbol{\nu}}$ and $\hat{\mathbf{V}}$ are consistent and asymptotically normally distributed as described by

$$\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Upsilon}), \tag{8a}$$

where $\xrightarrow{\mathcal{L}}$ is the notation for convergence in distribution, $\boldsymbol{\alpha}$ satisfies $E[\mathbf{g}(\boldsymbol{\alpha})] = \mathbf{0}$ and $\boldsymbol{\Upsilon}$ can be consistently estimated by

$$\hat{\boldsymbol{\Upsilon}} = \left[\frac{1}{n}\sum_{i=1}^n \frac{\partial \mathbf{g}_i(\hat{\boldsymbol{\alpha}})}{\partial \hat{\boldsymbol{\alpha}}'}\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^n \mathbf{g}_i(\hat{\boldsymbol{\alpha}})\mathbf{g}_i'(\hat{\boldsymbol{\alpha}})\right]\left[\frac{1}{n}\sum_{i=1}^n \frac{\partial \mathbf{g}_i'(\hat{\boldsymbol{\alpha}})}{\partial \hat{\boldsymbol{\alpha}}}\right]^{-1}. \tag{8b}$$

The formulas for evaluating $\partial \mathbf{g}_i(\boldsymbol{\alpha})/\boldsymbol{\alpha}'$ with Huber-type weights are given in Appendix A, and coded in the R package `rsem`.

3. Estimation and Inference with the Structural Model

The development in the previous section allows us to obtain $\hat{\boldsymbol{\nu}}$, $\hat{\mathbf{V}}$, and $\hat{\boldsymbol{\Upsilon}}$. Our interest is in modeling the mean vector and covariance matrix of $\mathbf{y}$. Let $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$, and $\hat{\boldsymbol{\Gamma}}$ be the parts of $\hat{\boldsymbol{\nu}}$, $\hat{\mathbf{V}}$, and $\hat{\boldsymbol{\Upsilon}}$ corresponding to the variables in $\mathbf{y}$, respectively; and $\boldsymbol{\beta} = (\boldsymbol{\mu}', \text{vech}'(\boldsymbol{\Sigma}))'$. It follows from (8a), (8b) that

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Gamma}), \tag{9}$$

where $\boldsymbol{\Gamma}$ is consistently estimated by $\hat{\boldsymbol{\Gamma}}$. With (9), the theory of robust SEM for samples containing missing values is the same as for SEM with complete data from an unknown population distribution. In particular, we can fit $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ by any structural model and use $\hat{\boldsymbol{\Gamma}}$ to obtain consistent SEs and test statistics or fit indices for overall model evaluation. Suppose $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are the structural models that satisfy $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ for certain $\boldsymbol{\theta}$. We choose estimating $\boldsymbol{\theta}$ by minimizing

$$F_{ML}(\boldsymbol{\theta}) = \big[\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}(\boldsymbol{\theta})\big]'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\big[\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}(\boldsymbol{\theta})\big] + \text{tr}\big[\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\big] - \log\big|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\big| - p \tag{10}$$

because minimizing $F_{ML}(\boldsymbol{\theta})$ for parameter estimates is the default procedure in essentially all SEM programs. Unless the sample size is huge, the resulting parameter estimates of minimizing (10) are also more efficient than the generalized least squares (GLS) estimates in which $\hat{\boldsymbol{\Gamma}}^{-1}$ is used as a weight matrix (see, e.g., Yuan & Bentler, 1997), although the GLS estimators are asymptotically more efficient. Let $\hat{\boldsymbol{\theta}}$ be the parameter estimates of minimizing (10). In the following, we will provide the formulas for obtaining consistent SEs of $\hat{\boldsymbol{\theta}}$ and test statistics for overall model evaluation. The output of the `rsem` package is based on these formulas. Let $\mathbf{D}_p$ be the duplication matrix such that $\mathbf{D}_p\,\text{vech}(\boldsymbol{\Sigma}) = \text{vec}(\boldsymbol{\Sigma})$ (Schott, 2005, p. 313), $\dot{\boldsymbol{\beta}} = \partial\boldsymbol{\beta}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and

$$\mathbf{W}_\beta = \begin{pmatrix} \boldsymbol{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & 0.5\mathbf{D}_p'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathbf{D}_p \end{pmatrix}.$$

It follows from (9) that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{\Omega}), \tag{11}$$

where

$$\boldsymbol{\Omega} = \big(\dot{\boldsymbol{\beta}}'\mathbf{W}_\beta\dot{\boldsymbol{\beta}}\big)^{-1}\big(\dot{\boldsymbol{\beta}}'\mathbf{W}_\beta\boldsymbol{\Gamma}\mathbf{W}_\beta\dot{\boldsymbol{\beta}}\big)\big(\dot{\boldsymbol{\beta}}'\mathbf{W}_\beta\dot{\boldsymbol{\beta}}\big)^{-1}$$

is consistently estimated when replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\Gamma}$ by $\hat{\boldsymbol{\Gamma}}$. Let $\hat{\boldsymbol{\Omega}} = (\hat{\omega}_{jk})$ be the resulting estimate of $\boldsymbol{\Omega}$. A consistent SE for the jth element of $\hat{\boldsymbol{\theta}}$ is given by $\hat{\omega}_{jj}^{1/2}/\sqrt{n}$.

Let $T_{ML} = nF_{ML}(\hat{\boldsymbol{\theta}})$ and k be the number of free parameters in $\boldsymbol{\theta}$. Although referring T_{ML} to the nominal chi-square distribution will most likely yield more reliable inference than the same procedure following NML, we do not recommend such a practice. Better theoretically justified test statistics are a rescaled statistic and an adjusted statistic derived from T_{ML}, a corrected residual-based asymptotically distribution free (ADF) statistic and a related F-statistic. Let $p^* = p(p+1)/2$, then $df = p^* + p - k$ is the nominal degrees of freedom. Let

$$\mathbf{U} = \mathbf{W}_\beta - \mathbf{W}_\beta\dot{\boldsymbol{\beta}}\big(\dot{\boldsymbol{\beta}}'\mathbf{W}_\beta\dot{\boldsymbol{\beta}}\big)^{-1}\dot{\boldsymbol{\beta}}'\mathbf{W}_\beta$$

and $\hat{m} = df/\,\text{tr}(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}})$. The rescaled statistic is given by

$$T_{RML} = \hat{m}T_{ML},$$

which asymptotically follows a distribution with mean equal to df (Satorra & Bentler, 1994). Let

$$\hat{m}_1 = \mathrm{tr}(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}}) / \mathrm{tr}\big[(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}})^2\big], \qquad \hat{m}_2 = \big[\mathrm{tr}(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}})\big]^2 / \mathrm{tr}\big[(\hat{\mathbf{U}}\hat{\boldsymbol{\Gamma}})^2\big].$$

The adjusted statistic

$$T_{AML} = \hat{m}_1 T_{ML}$$

asymptotically follows a distribution with mean and variance equal to that of $\chi^2_{m_2}$, where $m_2 = [\mathrm{tr}(\mathbf{U}\boldsymbol{\Gamma})]^2 / \mathrm{tr}[(\mathbf{U}\boldsymbol{\Gamma})^2]$. In practice, we refer T_{RML} to χ^2_{df} or T_{AML} to $\chi^2_{\hat{m}_2}$ for model inference. Although the exact distribution of neither T_{RML} nor T_{AML} is known even asymptotically, these chi-square distributions have been shown to provide good approximations both empirically (Hu, Bentler, & Kano, 1992) and asymptotically (Yuan & Bentler, 2010).

Let

$$\mathbf{Q} = \boldsymbol{\Gamma}^{-1} - \boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\beta}}\big(\dot{\boldsymbol{\beta}}'\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\beta}}\big)^{-1}\dot{\boldsymbol{\beta}}'\boldsymbol{\Gamma}^{-1}$$

and $\mathbf{r} = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}(\hat{\boldsymbol{\theta}})$. Then $T_{RADF} = n\mathbf{r}'\hat{\mathbf{Q}}\mathbf{r}$ is just the residual-based ADF statistic (Browne, 1984) applied to the setting of robust procedures with missing data, and T_{RADF} asymptotically follows χ^2_{df}. In particular, a corrected version of it,

$$T_{CRADF} = T_{RADF} / \big(1 + \mathbf{r}'\hat{\mathbf{Q}}\mathbf{r}\big)$$

also asymptotically follows χ^2_{df} and has been shown to work well when modeling the sample covariance matrices with complete data (Bentler & Yuan, 1999; Yuan & Bentler, 1998). Referring the F-statistic

$$T_{RF} = (n - df)T_{RADF} / \big[(n-1)df\big]$$

to an F-distribution with df and $n - df$ degrees of freedom has also been shown to work well with complete data at smaller sample sizes (Bentler & Yuan, 1999). Both $T_{CRADF} \sim \chi^2_{df}$ and $T_{RF} \sim F_{df,(n-df)}$ are asymptotically exact.

The details of the derivation or justification for the result in (11), as well as for the four test statistics, are essentially the same as for their counterparts in the context of SEM with complete data. We will not provide the details here. We would like to note that these four statistics are currently available in EQS (Bentler, 2008) for complete data or NML-based analysis with missing data. But they are not available in any software with a truly robust method. This motivated us to develop the statistical package to be introduced next.

4. R package `rsem` for Robust Estimation and Structural Models

This section introduces the R package `rsem` that generates the estimates $\hat{\boldsymbol{\nu}}$ and $\hat{\mathbf{V}}$ using the ER algorithm in (4) to (7) with the Huber-type weights in (3). The sandwich-type covariance matrix $\hat{\boldsymbol{\Upsilon}}$ in (8b) is also evaluated by the package. The vector $\hat{\boldsymbol{\mu}}$ and matrices $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Gamma}}$ are then fed into EQS for the second-stage analysis automatically by the package. We choose EQS because it outputs all the four test statistics described in the previous section, and it has the capability of talking with R since version[2] 6.1 for Windows (build 97). The output also contains multiple fit indices, standardized solutions, Lagrange multiplier, and Wald tests, which are widely used by applied researchers and are well documented in Bentler (2008). The use of the package is

[2]The R package for robust SEM does not work with earlier versions of EQS that do not have the capability of talking with R (Mair, Wu, & Bentler, 2010).

illustrated through a real data set, and missing values are created so that they are MAR when an auxiliary variable is included.

Table 1.2.1 of Mardia et al. (1979) contains test scores of $n = 88$ students on five subjects. The five subjects are: Mechanics, Vectors, Algebra, Analysis, and Statistics. The first two subjects were tested with closed-book exams and the last three were tested with open-book exams. Let $\mathbf{y}$ be the vector[3] of Mechanics, Vectors, Analysis, and Statistics. Yuan & Lu (2008) found that the sample means and covariances of these four variables are well explained by the two-factor model

$$\mathbf{y} = \mathbf{\Lambda f} + \mathbf{e}, \tag{12}$$

where

$$\mathbf{\Lambda} = \begin{pmatrix} 1.0 & \lambda_{21} & 0 & 0 \\ 0 & 0 & 1.0 & \lambda_{42} \end{pmatrix}'$$

is the factor loading matrix. Let $\boldsymbol{\tau} = E(\mathbf{f}) = (\tau_1, \tau_2)'$ be the vector of factor means, $\mathbf{\Phi} = (\phi_{jk}) = \text{Cov}(\mathbf{f})$ be the factor covariance matrix, and $\mathbf{\Psi} = \text{diag}(\psi_{11}, \psi_{22}, \psi_{33}, \psi_{44})$ be a diagonal matrix for the variances of unique factors or measurement errors. Then the mean and covariance structures of $\mathbf{y}$ are

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{\Lambda}\boldsymbol{\tau} \quad \text{and} \quad \mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi}. \tag{13}$$

There are 11 parameters in the model with

$$\boldsymbol{\theta} = (\tau_1, \tau_2, \lambda_{21}, \lambda_{42}, \phi_{11}, \phi_{21}, \phi_{22}, \psi_{11}, \psi_{22}, \psi_{33}, \psi_{44})'.$$

The normal-distribution-based likelihood ratio statistic is $T_{ML} = 3.259$, with an associated p-value $= 0.353$ when referred to χ^2_3.

We use the variable Algebra to create missing data schemes, and therefore x_3 is an auxiliary variable. When data for $x_2 =$ Vectors and $x_5 =$ Statistics are removed corresponding to the smallest 31 scores of $x_3 =$ Algebra, and the variable Algebra is excluded from the analysis, the missing data mechanism is MNAR. The missing data mechanism is MAR when the five variables are considered simultaneously. The created data set is at http://rpackages.psychstat.org/examples/rsem/mardiamv25.dat, with −99 for missing values. The data set is also part of our R package and can be accessed through the function `data(mardiamv25)`.

To use the R package for the first time, it can be installed by issuing the following command

```
install.packages("rsem")
```

With the package installed, the robust SEM analysis can be conducted as illustrated below.

The R code in Lines 1 to 5 in Appendix B[4] illustrates a typical routine for using our R package. Specifically, `library(rsem)` loads the R package. The code `setwd("c:/rsemmv")` sets the working directory to the folder that contains the data file and the EQS model file (see Appendix C). Lines 3 and 4 use the R function `read.table` to read the raw data in the file `mardiamv25.dat` into R and save the data into an object called `mardiamv25`.[5] The argument `header=T` tells R that variable names are given in the data file and the argument `na.string="-99"` indicates that `-99` represents a missing datum in the data file. Line 5 uses the function `rsem` from the package `rsem` to conduct the robust analysis. The first argument

[3] When including Algebra, the means and covariances of the five variables cannot be well fitted by a two-factor model, as implied by a highly significant T_{ML}.

[4] The line numbers on the right margin of Appendix B are for the convenience of explaining the code, not part of R input. The same is true for the EQS input files in Appendices C to F.

[5] Any name can be used here and `mardiamv25` is used for convenience.

`mardiamv25` specifies the name of the data. The second argument `c(1,2,4,5)` is the vector to select the variables 1, 2, 4, and 5 to be further fitted by the structural model in (12) or (13), excluding the auxiliary variable $x_3 =$ Algebra. The third argument `"mcov.eqs"` is the name of the EQS input file. The content of `mcov.eqs` for estimating the model in (12) or (13) is given in Appendix C.[6] Readers are referred to Bentler (2008) for detailed instruction on specifying different models within EQS.

The default output from running the four lines of R code is given in Lines 9 to 49 of Appendix B. Lines 9 to 28 contain the information on estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ at the first stage. The basic information about the data set, including the sample size and the number of variables, is given in Lines 9 and 10. Line 13 lists the names of variables selected for the structural model. Lines 15 to 18 provide information on missing data patterns. Line 15 tells the number of total observed patterns in the original sample, 2 in this example. Each row from Lines 17 to 18 contains $q + 2$ numbers regarding the missing data information for a particular pattern. The first is the number of observed cases in the pattern, the second is the number of observed variables (q_i) in the pattern. The next q numbers are either 1 or 0 with 1 indicating that the data for the corresponding variables are observed and 0 indicating missing.

Line 21 gives the estimated mean vector $\hat{\boldsymbol{\mu}}$, and Lines 25 to 28 give the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ corresponding to the selected variables listed in Line 13. The $\hat{\boldsymbol{\Gamma}}$, a 14×14 matrix, is also calculated by the R package. Because its dimension is relatively large, the matrix[7] is saved in the file `weight.txt` and read into EQS directly rather than being part of the default output of our R package. Actually, $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\mu}}$ are also saved in the file `data.txt` and read by the EQS file in Appendix C to perform the second-stage analysis.

The output of EQS for the second-stage analysis is in the file `mcov.out` in the working directory. The four test statistics described in Section 3, T_{RML}, T_{AML}, T_{CRADF}, and T_{RF}, appear, respectively, as

```
SATORRA-BENTLER SCALED CHI-SQUARE = 1.3763 ON 3 DEGREES
OF FREEDOM PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC
IS 0.71109

MEAN- AND VARIANCE-ADJUSTED CHI-SQUARE = 1.220 ON 3 D.F.
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.74826

YUAN-BENTLER RESIDUAL-BASED TEST STATISTIC = 1.427
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS 0.69930

YUAN-BENTLER RESIDUAL-BASED F-STATISTIC = 0.472
DEGREES OF FREEDOM = 3, 85
PROBABILITY VALUE FOR THE F-STATISTIC IS 0.70228
```

These four statistics and p-values[8] are also part of the default output of our R package, as displayed in Lines 32 to 35 of Appendix B.

[6]The input file is also available at http://rpackages.psychstat.org/examples/rsem/mcov.eqs.

[7]Because EQS uses a different order from vech($\boldsymbol{\Sigma}$) when vectorizing the covariance matrix, the matrix in the file `weight.txt` is a permutation of $\hat{\boldsymbol{\Gamma}}$; it also has an extra row and column of zeros. To print the matrix in R console, use `ex1$sem`.

[8]For the adjusted statistic T_{AML}, EQS approximates the $\hat{m}_2$ using the nearest integer and obtains the p-value using the approximated degrees of freedom.

TABLE 1.
Test statistics and parameter estimates for model (13) with open–closed-book data.

(a) Statistics for overall model evaluation.

	MH(0.10)				2-stage NML			
	T_{RML}	T_{AML}	T_{CRADF}	T_{RF}	T_{RML}	T_{AML}	T_{CRADF}	T_{RF}
T	1.376	1.220	1.427	0.472	1.361	1.284	1.285	0.425
p-value	0.711	0.748	0.699	0.702	0.715	0.733	0.733	0.736

(b) Parameter estimates $\hat{\boldsymbol{\theta}}$, their SEs, and z-scores.

θ	MH(0.10)			2-stage NML		
	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z
τ_1	39.447	1.698	23.228	39.187	1.748	22.416
τ_2	47.192	1.524	30.969	46.660	1.585	29.435
λ_{21}	1.289	0.046	28.145	1.290	0.046	28.251
λ_{42}	0.876	0.024	35.850	0.882	0.026	33.681
ϕ_{11}	87.660	34.171	2.565	102.040	33.993	3.002
ϕ_{21}	78.812	23.546	3.347	81.852	23.668	3.458
ϕ_{22}	192.695	53.888	3.576	200.402	54.635	3.668
ψ_{11}	180.696	30.851	5.857	182.097	29.874	6.095
ψ_{22}	41.575	27.150	1.531	30.703	25.392	1.209
ψ_{33}	10.545	55.172	0.191	19.499	51.742	0.377
ψ_{44}	203.808	41.220	4.944	199.950	38.821	5.151

The file `mcov.out` also contains multiple fit indices, parameter estimates $\hat{\boldsymbol{\theta}}$ and their SEs, and standardized solutions. In particular, there are two SEs following each parameter estimate as shown below for λ_{21}.

```
VECTORS = V2 = 1.289*F1 + 1.000 E2
               0.050
              25.905@
              (0.046)
             (28.145@)
```

The one immediately below the parameter estimate is obtained from the normal-distribution-based information matrix by treating $\hat{\boldsymbol{\mu}}$ as a vector of sample means and $\hat{\boldsymbol{\Sigma}}$ as a sample covariance matrix, which should be ignored. The one based on (11) is within parentheses, which is consistent and should be used when inferring the significance of the estimate. EQS uses the sign `@` to indicate that the estimate is significant at 0.05 level. The parameter estimates and their SEs based on (11) are also part of the default output of our R package, as shown in lines 39 to 49 of Appendix B, where `(A,B)` denotes a path from `B` to `A`. For example, in Line 49, `(V4,F2)` represents the factor loading from `F2` to `V4`. The three numbers on the right side are the parameter estimate, its consistent SE based on (11), and the corresponding z-score.

Test statistics, parameter estimates, and their SEs are represented in Table 1, where results under MH(0.10) are obtained by the M-estimator with Huber-type weights at $\varphi = 0.10$. Parallel results using 2-stage NML[9] are also reported in Table 1 for comparison purpose, where the SEs and z-scores are also based on a sandwich-type covariance matrix (Yuan & Lu, 2008). Two-stage NML is chosen for comparison with MH(0.10) because it has advantages in including auxiliary

[9]The R code for the analysis is `rsem(mardiamv25, c(1,2,4,5), "mcov.eqs", varphi=0)`.

TABLE 2.
Test statistics and parameter estimates for model (13) with 5 cases of the open–closed-book data being contaminated.

(a) Statistics for overall model evaluation.

	MH(0.10)				2-stage NML			
	T_{RML}	T_{AML}	T_{CRADF}	T_{RF}	T_{RML}	T_{AML}	T_{CRADF}	T_{RF}
T	0.803	0.788	0.725	0.238	5.126	2.516	2.093	0.699
p-value	0.849	0.852	0.867	0.870	0.163	0.113	0.553	0.555

(b) Parameter estimates $\hat{\boldsymbol{\theta}}$, their SEs, and z-scores.

θ	MH(0.10)			2-stage NML		
	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z
τ_1	40.813	1.625	25.122	39.436	2.155	18.298
τ_2	49.397	1.547	31.925	52.918	2.900	18.251
λ_{21}	1.291	0.046	28.326	1.399	0.087	16.072
λ_{42}	0.899	0.028	32.088	1.019	0.073	14.005
ϕ_{11}	70.171	26.699	2.628	127.887	42.616	3.001
ϕ_{21}	68.254	20.463	3.336	226.830	104.116	2.179
ϕ_{22}	202.579	63.329	3.199	691.490	309.217	2.236
ψ_{11}	168.629	31.932	5.281	250.928	66.602	3.768
ψ_{22}	52.116	27.826	1.873	99.705	102.299	0.975
ψ_{33}	6.732	61.601	0.109	−62.183	122.096	−0.509
ψ_{44}	216.024	50.702	4.261	546.936	319.965	1.709

variables over direct NML (Savalei & Bentler, 2009; Yuan & Lu, 2008), and test statistics for 2-stage NML also perform better under varied conditions (Savalei & Falk, in press). The statistics under MH(0.10) and 2-stage NML in Table 1(a) are very comparable, suggesting that the model in (12) fits the data well. Most of the parameter estimates and SEs under MH(0.10) are also comparable to those under 2-stage NML in Table 1(b). This is because the sample is very close to being normally distributed. Actually, the normalized Mardia's (1970) multivariate kurtosis for the original open–closed-book data is 0.057, not statistically significant at all.

The results in Table 1 suggest that the M-estimator with Huber-type weights generates results very close to those by 2-stage NML when data are close to normally distributed. However, practical data typically do not follow a normal distribution as close as the open–closed-book data. Actually, among all raw data that have been used in the SEM literature and are available to us, the distribution of the open–closed-book data is the closest to a normal distribution. In the created missing data set, only three variables are observed on each of the last five cases. Multiplying each of these 15 numbers by 5 created a contaminated data set.[10] Applying the same procedures that generated Table 1 to this new data set generates the results in Table 2, where the results under MH(0.10) and 2-stage NML are quite different. In particular, SEs under 2-stage NML in Table 2(b) are uniformly greater than those under MH(0.10). There is also a Heywood case under 2-stage NML. The statistics in Table 2(a) under 2-stage NML still endorse the model because they all account for non-normality by including fourth-order moments, but they are not as supportive as those under MH(0.10). In addition, with default starting values it took 391 iterations for EQS to obtain the estimates under 2-stage NML while it only took 12 iterations for EQS to obtain the estimates under MH(0.10).

The parameter estimates under MH(0.10) and 2-stage NML in Table 2 mostly differ in the estimates of factor variances–covariance (ϕs) and error variances (ψs). This is because model

[10]The data can be obtained at http://rpackages.psychstat.org/examples/rsem/mardiamv25_contaminated.dat.

identification is enforced by $\lambda_{11} = \lambda_{32} = 1$. If we set $\phi_{11} = \phi_{22} = 1$ for model identification, we will notice more difference in the 2-stage NML estimates of the factor loading parameters (λs) between using the original and contaminated data sets. The SEs under 2-stage NML, obtained using the sandwich-type covariance matrix, are often called robust SEs in the SEM literature. Comparing the results in Tables 1 and 2, we can observe that SEs under 2-stage NML change dramatically with data contamination. For example, the SE of $\hat{\phi}_{22}$ under 2-stage NML in Table 1 is 54.635 while that in Table 2 is 309.217. Thus, the "robust SEs" under 2-stage NML are not robust at all. In comparison, the changes in parameter estimates and SEs under MH(0.10) from Table 1 to Table 2 are much smaller.

The R code in Line 5 in Appendix B conducts the basic robust analysis. The function `rsem` will perform other robust analysis when supplied with different arguments. The full specification of this function is

```
rsem(dset, select, EQSmodel, moment = TRUE, varphi = 0.1,
     max.it = 1000, eqsdata = "data.txt",
     eqsweight = "weight.txt",
     EQSpgm = "C:/Progra~1/EQS61/WINEQS.EXE", serial="1234")
```

The first argument `dset` specifies the data to be used and this argument is required. The second argument `select` supplies the indices of variables that are used for analysis in the structural model. In the previous example, `select=c(1,2,4,5)`, meaning that the first, second, fourth, and fifth variables are selected. Not providing the argument `select` implies that all the variables in the data set will be used in the structural model or there is no auxiliary variable. The third argument `EQSmodel` provides the name of the EQS input file. In the previous example, `EQSmodel="mcov.eqs"`. If omitted, only the saturated mean vector and covariance matrix are estimated[11] and no structural model will be analyzed. The fourth argument is `moment` and its default value `TRUE` indicates that mean and covariance structure analysis will be conducted. Alternatively, if `moment=FALSE`, covariance structure analysis will be conducted without means. EQS code for covariance structure analysis with the open–closed-book data is provided in Appendix D. The fifth argument `varphi=0.1` specifies the Huber-type weight function according to (3) that gives the approximate proportion of cases to be down-weighted. The default value is 10 %. If `varphi=0`, 2-stage NML analysis is performed and no case is down-weighted. The sixth argument `max.it` defines the maximum number of iterations for the ER algorithm. The default is 1000 and if the number is exceeded, the user will be prompted to supply a greater number. The seventh argument `eqsdata` specifies the file name to save the estimates $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\mu}}$ from the ER algorithm and should be the same as the file name for the argument `data` in the EQS input file (e.g., Line 7 in Appendix C). The eighth argument `eqsweight` specifies the file name to save the sandwich-type covariance matrix $\hat{\boldsymbol{\Gamma}}$ from the ER algorithm and should be the same as the file name for the argument `weight` in the EQS input file (e.g., Line 5 in Appendix C). The next argument tells the path to the EQS program and it can be omitted typically. The last argument `serial` specifies the serial number of the EQS program (see Mair et al., 2010).

After running the function `rsem`, in addition to the default output discussed earlier, results from the analysis are also saved into the object `ex1`, according to Line 5 in Appendix B. For example, `ex1$misinfo` includes the missing data pattern information and sorted data according to missing data patterns; and `ex1$sem` provides the estimated mean vector, covariance matrix, and sandwich-type covariance matrix $\hat{\boldsymbol{\Gamma}}$. Other components of `ex1` can be viewed using the function `names(ex1)`.

[11] The SEs of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ according to (8a) and (8b) will be in the default output of R. The matrix $\hat{\boldsymbol{\Gamma}}$ according to the order of $\boldsymbol{\beta}$ in (8a) and (8b) will be saved into the object `ex1`, which is useful when SEM software other than EQS is used.

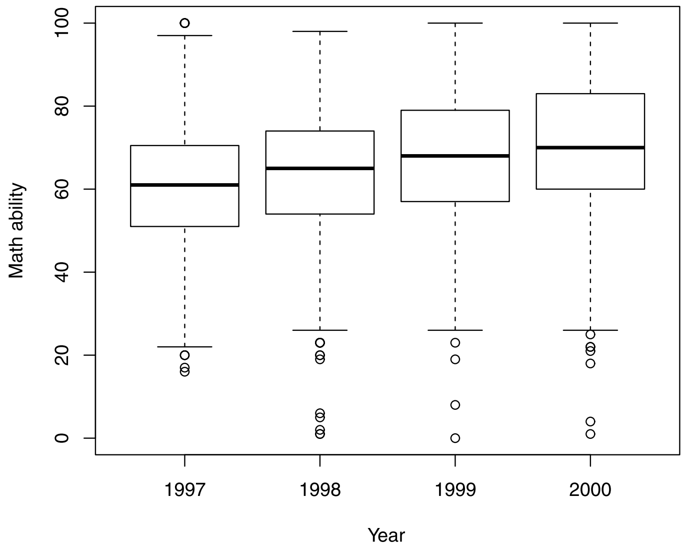

FIGURE 1.
Boxplots of Peabody Individual Achievement Test (PIAT) math data. Circles represent potential outliers that are more than 1.5 times interquantile range away from the first and third quantiles, respectively.

5. Robust Analysis of NLSY97 with Growth Curve Models

The NLSY97 consists of a nationally representative sample of approximately 9,000 youths who were 12 to 16 years old as of December 31, 1996. Many variables in the surveys were followed on an annual basis. The data set available to us consisted of yearly administration of the mathematics subtest of the Peabody Individual Achievement Test (PIAT) from 1997 to 2000 on $N = 399$ students. Information on family income, fathers' and mothers' years of education was also collected for this sample in 1997. We were interested in using this data set to investigate how mathematical ability grew over the 4-year period. Each of the seven variables contained missing values, only 126 cases (about one third) were completely observed, and there were a total of 44 observed data patterns. The data were also significantly non-normally distributed, with Mardia's measure of multivariate kurtosis = 20.376, and its normalized version = 21.020 (Yuan et al., 2004b). Figure 1 contains the boxplots of the 4 measures of mathematical ability, showing that each variable is skewed to the left due to outstanding cases.

More descriptive statistics, including the mean, the standard deviation (SD), the minimum (Min), the maximum (Max) and the percentage of complete (PC) data, for each variable are reported in Table 3. The average family income was about \$17,470 in 1997 and both parents had an average of about 12 years of education. Because the mathematical variables are longitudinal, we will use latent growth curve models to investigate the change of mathematical ability over the 4-year period. Both unconditional and conditional models will be studied (Preacher, Wichman, MacCallum, & Briggs, 2008).

5.1. *Unconditional Latent Growth Curve Model*

By using the unconditional latent growth curve model, we focus on the analysis of the growth rate of mathematical ability. The variables—family income, fathers' education and mothers' education—are included as auxiliary variables. Let $\mathbf{y}$ be the vector of mathematical abilities measured for the 4 years. The unconditional linear growth curve model can be written as

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \mathbf{e}, \tag{14}$$

TABLE 3.
Descriptive statistics of the NLSY97 sample ($N = 399$).

Variables	n	Mean	SD	Min	Max	PC
Math 1997	375	61.16	15.89	16	100	94
Math 1998	377	63.27	17.22	1	98	94
Math 1999	357	67.56	16.65	0	100	89
Math 2000	350	69.69	17.60	1	100	88
Family income ($1,000)	234	17.47	14.84	3	83	58
Fathers' education	275	12.24	2.86	3	20	69
Mothers' education	362	12.02	2.61	3	20	91

TABLE 4.
Unconditional latent growth curve analysis of mathematical ability.

(a) Statistics for overall model evaluation.

	MH(0.10)				2-stage NML			
	T_{RML}	T_{AML}	T_{CRADF}	T_{RF}	T_{RML}	T_{AML}	T_{CRADF}	T_{RF}
T	9.908	8.557	8.418	1.703	14.679	11.606	12.058	2.463
p	0.078	0.073	0.135	0.133	0.012	0.021	0.034	0.033

(b) Parameter estimates $\hat{\boldsymbol{\theta}}$, their SEs, and z-scores.

θ	MH(0.10)			2-stage NML		
	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z
τ_1	60.865	0.745	81.749	60.645	0.780	77.797
τ_2	3.177	0.226	14.038	3.100	0.258	12.005
ϕ_{11}	174.450	19.254	9.060	177.489	24.590	7.218
ϕ_{12}	−6.290	4.904	−1.283	−4.938	7.664	−0.644
ϕ_{22}	6.791	2.746	2.473	6.994	4.000	1.748
ψ_{11}	62.406	13.870	4.499	87.576	25.896	3.382
ψ_{22}	77.177	9.105	8.476	103.477	14.275	7.249
ψ_{33}	73.794	9.818	7.516	90.147	14.391	6.264
ψ_{44}	72.463	17.173	4.220	109.889	27.734	3.962

where

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix}'$$

and $\boldsymbol{\xi} = (\xi_1, \xi_2)'$ with ξ_1 being the individual initial level of mathematical ability in 1997 and ξ_2 being the growth rate from 1997 to 2000. Let $\boldsymbol{\tau} = E(\boldsymbol{\xi}) = (\tau_1, \tau_2)'$ with τ_1 representing the average initial level and τ_2 representing the average change rate; $\boldsymbol{\Phi} = (\phi_{jk}) = \text{Cov}(\boldsymbol{\xi})$ with ϕ_{11}, ϕ_{22}, and ϕ_{12} representing individual difference in the initial level and in the growth rate of mathematical ability and in their covariance, respectively; and $\boldsymbol{\Psi} = \text{Cov}(\mathbf{e}) = \text{diag}(\psi_{11}, \psi_{22}, \psi_{33}, \psi_{44})$ be a diagonal matrix containing the variances of unique factors or measurement errors. The mean and covariance structure of the $\mathbf{y}$ in (14) can be expressed as that in (13). The EQS input file for estimating the model (14) is given in Appendix E.

Both 2-stage NML and MH(0.10) are used for the analysis of the unconditional model, and the results are reported in Table 4. All the four test statistics following 2-stage NML suggest that there is a significant difference between the model and the data. However, none of the statistics under MH(0.10) is statistically significant at the 0.05 level. Given the boxplots in Figure 1 and the highly significant multivariate kurtosis, we would trust the results following MH(0.10) more than those following 2-stage NML. Actually, estimates for the structural parameters under MH(0.10) are very comparable to those under 2-stage NML, while estimates for error variances under MH(0.10) are smaller. In particular, the SEs under MH(0.10) are uniformly smaller, implying that the robust estimates are more efficient. Due to being less efficient, $\hat{\phi}_{22}$ under 2-stage NML is not statistically significant at the 0.05 level.

According to the robust analysis, the average initial level of mathematical ability in 1997 is about 60.865 and the growth rate from 1997 to 2000 is about 3.177. Individuals are significantly different in both initial level and growth rate. Students with higher initial levels tend to have lower growth rates although $\hat{\phi}_{12}$ is not significant at the 0.05 level. Actually, the p-value associated with $\hat{\phi}_{12} = -6.29$ ($z = -1.283$) for a one-sided test is about 0.1.

5.2. *Conditional Latent Growth Curve Model*

With the conditional latent growth curve model, we evaluate how family income and parents' education are related to the initial level and growth rate of mathematical ability of children. The conditional model can be specified as

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \mathbf{e}, \tag{15}$$

$$\boldsymbol{\xi} = \boldsymbol{\tau} + \mathbf{B}\mathbf{v} + \boldsymbol{\zeta}, \tag{16}$$

where $\mathbf{y}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\xi}$, and $\mathbf{e}$ are the same as used in the unconditional model. The $\mathbf{v}$ is a vector of family income, fathers' education and mothers' education. The matrix

$$\mathbf{B} = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{pmatrix}$$

contains the regression coefficients of initial level and growth rate of math ability on family income, fathers' education and mothers' education, and $\boldsymbol{\tau}$ contains the intercepts. The vector $\boldsymbol{\zeta}$ contains the residuals of $\boldsymbol{\xi}$ after being predicted by $\mathbf{v}$. There is no auxiliary variable in the analysis. The EQS input file for estimating the model in (15) and (16) is given in Appendix F.

The results following the analyses by MH(0.10) and 2-stage NML for the conditional model are reported in Table 5. Although test statistics following 2-stage NML in Table 5(a) are not significant at the 0.05 level, those following MH(0.10) give stronger support for the substantive model in (15) and (16). Similar to those in Table 4, while the estimates for the structural parameters following the two methods are comparable, those for the error variances are clearly smaller under MH(0.10). The SEs for the parameter estimates following MH(0.10) in Table 5 are again uniformly smaller. The z-scores for estimates $\hat{\tau}_2$ and $\hat{\phi}_{22}$ under MH(0.10) imply that, after controlling the family background variables, the average growth rate is still statistically significant and individuals are significantly different in growth rate.

The z-scores for the six beta coefficients indicate that only the variable income positively predicts the initial level of mathematical ability after controlling for parents' education. Neither fathers' education nor mothers' education significantly predicts the initial level or the growth rate of children's math ability. Such a conclusion might also be drawn by comparing the estimates $\hat{\phi}_{11}$, $\hat{\phi}_{12}$, and $\hat{\phi}_{22}$ in Tables 4 and 5, and those in Table 5 are only slightly smaller.

6. Discussion and Conclusion

In social and behavioral sciences, data are typically non-normally distributed (Micceri, 1989). The aim of the paper is to develop a robust SEM procedure for real data like NLSY97, which have missing values and a significant multivariate kurtosis. According to Rubin (1976), MAR mechanism can be ignored if the analysis is proceeded by ML. However, both the missing data mechanism and the population distribution are typically unknown. The first stage of the

TABLE 5.
Conditional latent growth curve analysis of mathematical ability.

(a) Statistics for overall model evaluation.

	MH(0.10)				2-stage NML			
	T_{RML}	T_{AML}	T_{CRADF}	T_{RF}	T_{RML}	T_{AML}	T_{CRADF}	T_{RF}
T	13.735	11.884	13.332	1.223	16.089	12.975	15.242	1.405
p	0.248	0.293	0.272	0.270	0.138	0.164	0.172	0.168

(b) Parameter estimates $\hat{\boldsymbol{\theta}}$, their SEs, and z-scores.

θ	MH(0.10)			2-stage NML		
	$\hat{\theta}$	SE	z	$\hat{\theta}$	SE	z
τ_1	51.696	4.266	12.117	51.548	4.292	12.010
τ_2	3.717	1.462	2.542	2.866	1.851	1.548
β_{11}	0.231	0.058	3.999	0.247	0.058	4.225
β_{12}	−0.114	0.411	−0.277	−0.125	0.443	−0.283
β_{13}	0.552	0.410	1.345	0.539	0.430	1.254
β_{21}	0.001	0.018	0.079	0.010	0.021	0.461
β_{22}	0.150	0.135	1.112	0.163	0.149	1.094
β_{23}	−0.197	0.123	−1.598	−0.157	0.143	−1.102
ϕ_{11}	160.875	18.256	8.812	161.062	23.266	6.923
ϕ_{12}	−5.514	4.793	−1.150	−4.804	7.468	−0.643
ϕ_{22}	6.370	2.744	2.321	6.571	4.015	1.637
ψ_{11}	64.164	14.137	4.539	89.091	25.920	3.437
ψ_{22}	76.687	9.044	8.480	102.918	14.178	7.259
ψ_{33}	72.774	9.705	7.498	89.010	14.159	6.287
ψ_{44}	73.807	17.045	4.330	111.824	27.419	4.078

developed procedure allows us to easily include auxiliary variables so that it is more realistic to assume that the missing data mechanism is MAR. The tuning parameter φ allows us to choose different weighting schemes according to (3) so that the resulting estimating equations in (1) and (2) approximate those obtained when setting the score functions corresponding to the true unknown likelihood function to zero. Thus, the robust estimates are closer to the true population values of the parameters that generated the data than pseudo NMLEs.

We set the default value of φ at 0.10 in the `rsem` package, which implies that observations with $d_i > \rho_i = c_{0.10}$ will get weights smaller than 1.0 according to (3), where $c_{0.10}$ is the critical value corresponds to the 90 % quantile of a chi-distribution. If letting $\varphi = 0.20$, then additional cases with $d_i \in (c_{0.20}, c_{0.10}]$ will also be downweighted in the estimation process. Because, at $\varphi = 0.20$, cases with $d_i \in (c_{0.20}, c_{0.10}]$ have weights only slightly smaller than 1.0, parameter estimates corresponding to the two φs may only differ slightly. Outlying cases with extreme d_i will get weights close to zero whether $\varphi = 0.05, 0.10$, or 0.20. Empirical results in Zhong & Yuan (2011) for complete data indicate that using a robust method matters far more than choosing a particular φ. A greater φ gives the estimates more protection against data contamination while the estimates will be less efficient when data are truly from a normally distributed population without contamination. In our experience, $\varphi = 0.10$ keeps a good balance between efficiency and protection against anomalies. Researchers who like to pursue optimality are referred to Yuan et al. (2004a), where empirical efficiency of parameter estimates by bootstrap is used to select the tuning parameter φ.

Statistical theory for robust estimation has been developed primarily within the class of elliptical distributions (Huber, 1981; Hampel, Ronchetti, Rousseeuw & Stahl, 1986), mainly because Equations (1) and (2) are the score functions defining the MLEs of $\boldsymbol{\nu}$ and $\mathbf{V}$ when $w_{i1}(d_i)$, $w_{i2}(d_i)$

and $w_{i3}(d_i)$ are properly chosen. In practice, data contamination or outliers make a sample from a truly elliptical distribution skewed at the sample level. In such a situation, a robust procedure is definitely preferred. If the true distribution of $\mathbf{x}$ is skewed, then the M-estimators $\hat{\boldsymbol{\nu}}$ and $\hat{\mathbf{V}}$ may not converge to the population means and covariance matrix; and NML estimates may not converge to the population means and covariance matrix either when missing values are MAR. Even in situation where NMLEs are known to be consistent and there is no data contamination or outliers, the estimates of variance parameters by NML may contain biases that are larger than the parameter values themselves (Yuan, Wallentin & Bentler, in press). Monte Carlo studies and empirical results with real and simulated complete data in Zu & Yuan (2010) and Zhong & Yuan (2011) indicate that robust methods lead to less biased and more efficient parameter estimates than NML even when populations are skewed. Preliminary results reported in Tong, Zhang and Yuan (2011) indicate that the two-stage robust procedure developed in this paper also leads to less biased parameter estimates and more reliable test statistics than NML when samples contain missing values. Thus, we expect this procedure to yield more reliable analysis than NML in most practical situations. Of course, the performance of the robust method over NML does not mean that the former is the best. Actually, with typical unknown population distributions in practice, it is unlikely to find a method that yields unbiased and most efficient parameter estimates.

In the `rsem` package, we implemented the Huber-type M-estimators because they are very close to NMLEs for normally distributed population and have been shown to work well in practical data analysis (Yuan et al., 2004a; Zhong & Yuan, 2011; Zu & Yuan, 2010). However, the Huber-type M-estimator cannot handle the situation when the proportion of extreme values is greater than $1/(q+1)$, as measured by a property called the breakdown point. If there is a suspicion that a large proportion of extreme observations due to contamination exists, one may need to choose the S- or the MCD-estimators (Rocke, 1996; Cheng & Victoria-Feser, 2002). Both can have a breakdown point of approximately $1/2$, while the S-estimator is usually more efficient. Since S-estimators also satisfy Equations (1) and (2), the ER algorithm and the methodology development in Sections 2 and 3 also apply to S-estimators. However, these estimators tend to be less efficient than an M-estimator. Actually, many observations get weights of zero in an estimator with a high breakdown point. When the sample size is not large enough, an estimator with a high breakdown point may end up with a singular $\hat{\boldsymbol{\Gamma}}$, which does not permit us to get valid statistics at the second-stage analysis. In additional to breakdown point, the starting values for the ER algorithm may affect the robustness of the converged estimators. We set the starting values of $\boldsymbol{\nu}$ and $\mathbf{V}$ in the ER algorithm at respectively $\mathbf{0}$ and $\mathbf{I}$ in the R package `rsem`, which are obviously not affected by contaminated cases.

The development of the paper parallels that of 2-stage NML. Another approach is to directly estimate the structural parameters without explicitly estimating the saturated model. This can be done by embedding the structural model into a multivariate t-distribution or Equations (1) and (2) of the paper. With the existing evidence on advantages of 2-stage NML over direct NML (Savalei & Bentler, 2009; Savalei & Falk, in press), we do not expect that a direct robust approach will out-perform the 2-stage approach as developed in this paper.

Following the typical practice of SEM, this paper does not consider prior information on model parameters. When prior information is available, one may include the information using a Bayesian analysis. In particular, the Bayesian procedure developed in Lee & Xia (2008) allows the sample to contain missing values and is robust to data contamination.

Acknowledgements

We would like to thank Dr. Alberto Maydeu-Olivares and two reviewers for their very constructive comments on an earlier version of the paper.

Appendix A. Mathematical Details for Evaluating the Matrix $\hat{\boldsymbol{\Upsilon}}$

This appendix provides the development and formulas for evaluating the $\hat{\boldsymbol{\Upsilon}}$ in (8b) with Huber-type weight. The formulas are programmed in the R package introduced in Section 4.

With the Huber-type weight, $w_{i3}(d_i) = 1$. Then the estimating equations in (1) and (2) are derived from

$$g_{i1}(\boldsymbol{\alpha}) = w_{i1}(d\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i), \tag{A.1}$$

and

$$g_{i2}(\boldsymbol{\alpha}) = \frac{1}{2}\operatorname{tr}\left\{\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}\left[w_{i2}(\mathbf{x}_i - \boldsymbol{\nu}_i)(\mathbf{x}_i - \boldsymbol{\nu}_i)' - \mathbf{V}_i\right]\right\}, \tag{A.2}$$

where d is for differentials. It follows from (A.1) and (A.2) that

$$\begin{aligned} dg_{i1}(\boldsymbol{\alpha}) = & -w_{i1}(d\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\boldsymbol{\nu}_i) - w_{i1}(d\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i) \\ & + (dw_{i1})(d\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i) \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} dg_{i2}(\boldsymbol{\alpha}) = & -\operatorname{tr}\left\{\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}\left[w_{i2}(\mathbf{x}_i - \boldsymbol{\nu}_i)(\mathbf{x}_i - \boldsymbol{\nu}_i)'\right]\right\} \\ & + \frac{1}{2}\operatorname{tr}\left\{\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(d\mathbf{V}_i)\right\} \\ & - \frac{1}{2}\operatorname{tr}\left\{\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}w_{i2}\left[(d\boldsymbol{\nu}_i)(\mathbf{x}_i - \boldsymbol{\nu}_i)' + (\mathbf{x}_i - \boldsymbol{\nu}_i)(d\boldsymbol{\nu}_i)'\right]\right\} \\ & + \frac{1}{2}\operatorname{tr}\left\{\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}\left[(dw_{i2})(\mathbf{x}_i - \boldsymbol{\nu}_i)(\mathbf{x}_i - \boldsymbol{\nu}_i)'\right]\right\}. \end{aligned} \tag{A.4}$$

Noting that both w_{i1} and w_{i2} are function of $d_i = [(\mathbf{x}_i - \boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i)]^{1/2}$, we have

$$dw_{i1}(d_i) = \begin{cases} 0 & \text{if } d_i \le \rho_i, \\ \frac{\rho_i}{d_i^3}[(d\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i) + 0.5(\mathbf{x}_i - \boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i)] & \text{if } d_i > \rho_i \end{cases} \tag{A.5}$$

and

$$dw_{i2}(d_i) = \begin{cases} 0 & \text{if } d_i \le \rho_i, \\ \frac{\rho_i^2}{\kappa_i d_i^4}[2(d\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i) + (\mathbf{x}_i - \boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i)] & \text{if } d_i > \rho_i. \end{cases} \tag{A.6}$$

Thus, when $d_i > \rho_i$, we have

$$\begin{aligned} dg_{i1}(\boldsymbol{\alpha}) = & -w_{i1}(d\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\boldsymbol{\nu}_i) - w_{i1}(d\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i) \\ & + \frac{\rho_i}{d_i^3}(d\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i)(\mathbf{x}_i - \boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\boldsymbol{\nu}_i) \\ & + \frac{\rho_i}{2d_i^3}(\mathbf{x}_i - \boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(\mathbf{x}_i - \boldsymbol{\nu}_i)(\mathbf{x}_i - \boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\boldsymbol{\nu}_i) \end{aligned} \tag{A.7}$$

and

$$
\begin{aligned}
dg_{i2}(\boldsymbol{\alpha}) = &-\operatorname{tr}\left\{\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}\left[w_{i2}(\mathbf{x}_i-\boldsymbol{\nu}_i)(\mathbf{x}_i-\boldsymbol{\nu}_i)'\right]\right\} \\
&+\frac{1}{2}\operatorname{tr}\left\{\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(d\mathbf{V}_i)\right\} \\
&-\frac{1}{2}\operatorname{tr}\left\{\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}w_{i2}\left[(d\boldsymbol{\nu}_i)(\mathbf{x}_i-\boldsymbol{\nu}_i)'+(\mathbf{x}_i-\boldsymbol{\nu}_i)(d\boldsymbol{\nu}_i)'\right]\right\} \\
&+\frac{\rho_i^2}{\kappa_i d_i^4}\operatorname{tr}\left\{\mathbf{V}_i^{-1}(\mathbf{x}_i-\boldsymbol{\nu}_i)(\mathbf{x}_i-\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(\mathbf{x}_i-\boldsymbol{\nu}_i)(d\boldsymbol{\nu}_i)'\right\} \\
&+\frac{\rho_i^2}{2\kappa_i d_i^4}\operatorname{tr}\left\{\mathbf{V}_i^{-1}(\mathbf{x}_i-\boldsymbol{\nu}_i)(\mathbf{x}_i-\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\mathbf{V}_i)\mathbf{V}_i^{-1}(\mathbf{x}_i-\boldsymbol{\nu}_i)(\mathbf{x}_i-\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1}(d\mathbf{V}_i)\right\}.
\end{aligned} \tag{A.8}
$$

Notice that, for matrices **A**, **B**, **C**, and **D** of proper orders, there exists

$$
\operatorname{tr}(\mathbf{ABCD}) = \operatorname{vec}'(\mathbf{D})\left(\mathbf{A}\otimes\mathbf{C}'\right)\operatorname{vec}\left(\mathbf{B}'\right) = \operatorname{vec}'\left(\mathbf{D}'\right)\left(\mathbf{C}'\otimes\mathbf{A}\right)\operatorname{vec}(\mathbf{B}). \tag{A.9}
$$

Let

$$
\mathbf{b}_i = \mathbf{V}_i^{-1}(\mathbf{x}_i-\boldsymbol{\nu}_i), \qquad \mathbf{H}_i = \mathbf{V}_i^{-1}(\mathbf{x}_i-\boldsymbol{\nu}_i)(\mathbf{x}_i-\boldsymbol{\nu}_i)'\mathbf{V}_i^{-1},
$$

$$
\mathbf{E}_i = \frac{\partial\boldsymbol{\nu}_i}{\partial\boldsymbol{v}'}, \quad \text{and} \quad \mathbf{F}_i = \frac{\partial\operatorname{vec}(\mathbf{V}_i)}{\partial\boldsymbol{v}'}.
$$

Using (A.9), it follows from (A.3) to (A.8) that, when $d_i \le \rho_i$,

$$
\frac{\partial\mathbf{g}_{i1}(\boldsymbol{\alpha})}{\partial\boldsymbol{v}'} = -\mathbf{E}_i'\mathbf{V}_i^{-1}\mathbf{E}_i, \qquad \frac{\partial\mathbf{g}_{i1}(\boldsymbol{\alpha})}{\partial\boldsymbol{v}'} = -\mathbf{E}_i'\left(\mathbf{V}_i^{-1}\otimes\mathbf{b}_i'\right)\mathbf{F}_i,
$$

$$
\frac{\partial\mathbf{g}_{i2}(\boldsymbol{\alpha})}{\partial\boldsymbol{v}'} = -\frac{1}{\kappa_i}\mathbf{F}_i'\left(\mathbf{b}_i\otimes\mathbf{V}_i^{-1}\right)\mathbf{E}_i, \qquad \frac{\partial\mathbf{g}_{i2}(\boldsymbol{\alpha})}{\partial\boldsymbol{v}'} = -\mathbf{F}_i'\left[\frac{1}{\kappa_i}\left(\mathbf{H}_i\otimes\mathbf{V}_i^{-1}\right)-\mathbf{W}_i\right]\mathbf{F}_i;
$$

and when $d_i > \rho_i$,

$$
\begin{aligned}
\frac{\partial\mathbf{g}_{i1}(\boldsymbol{\alpha})}{\partial\boldsymbol{v}'} &= -\mathbf{E}_i'\left(w_{i1}\mathbf{V}_i^{-1}-\frac{\rho_i}{d_i^3}\mathbf{H}_i\right)\mathbf{E}_i \\
\frac{\partial\mathbf{g}_{i1}(\boldsymbol{\alpha})}{\partial\boldsymbol{v}'} &= -\mathbf{E}_i'\left[w_{i1}\left(\mathbf{V}_i^{-1}\otimes\mathbf{b}_i'\right)-\frac{\rho_i}{2d_i^3}\left(\mathbf{H}_i\otimes\mathbf{b}_i'\right)\right]\mathbf{F}_i, \\
\frac{\partial\mathbf{g}_{i2}(\boldsymbol{\alpha})}{\partial\boldsymbol{v}'} &= -\mathbf{F}_i'\left[w_{i2}\left(\mathbf{b}_i\otimes\mathbf{V}_i^{-1}\right)-\frac{\rho_i^2}{\kappa_i d_i^4}(\mathbf{H}_i\otimes\mathbf{b}_i)\right]\mathbf{E}_i, \\
\frac{\partial\mathbf{g}_{i2}(\boldsymbol{\alpha})}{\partial\boldsymbol{v}'} &= -\mathbf{F}_i'\left[w_{i2}\left(\mathbf{H}_i\otimes\mathbf{V}_i^{-1}\right)-\mathbf{W}_i-\frac{\rho_i^2}{2\kappa_i d_i^4}(\mathbf{H}_i\otimes\mathbf{H}_i)\right]\mathbf{F}_i.
\end{aligned}
$$

Appendix B. R Code for Robust SEM and Its Output

```
library(rsem)
setwd("c:/rsemmv")
mardiamv25<-read.table("mardiamv25.dat", header=T,
                       na.string="-99")
ex1<-rsem(mardiamv25, c(1,2,4,5), "mcov.eqs")

## Sample output from the above analysis.

Sample size n = 88
```

```
Total number of variables q= 5

The following 4 variables are selected for SEM models
Mechanics Vectors Analysis Statistics

There are 2 missing data patterns. They are
           n nvar Mechanics Vectors Algebra Analysis Statistics
Pattern 1 57    5         1       1       1        1          1
Pattern 2 31    3         1       0       1        1          0

Estimated means:
[1] 39.18057 50.90668 47.20384 41.05387

Estimated covariance matrix:
     Mechanics  Vectors Analysis Statistics
[1,]  289.3151 124.7135 105.4223   102.6819
[2,]  124.7135 182.4506  95.7065   108.5030
[3,]  105.4223  95.7065 202.1062   180.7338
[4,]  102.6819 108.5030 180.7338   373.3670

Test statistics:
            T       p
RML   1.37630 0.71110
AML   1.21980 0.74826
CRADF 1.42660 0.69930
RF    0.47245 0.70229

Parameter estimates:
            Parameter          SE          z
(E1,E1)   180.6959600 30.85084800  5.8570824
(E2,E2)    41.5744760 27.15028700  1.5312721
(E3,E3)    10.5459460 55.17236400  0.1911454
(E4,E4)   203.8073000 41.21980500  4.9444023
(D1,D1)    87.6599860 34.17127900  2.5653118
(D1,D2)    78.8119050 23.54609100  3.3471333
(D2,D2)   192.6941800 53.88799900  3.5758273
(F1,V999)  39.4471330  1.69824770 23.2281386
(F2,V999)  47.1918230  1.52386050 30.9685978
(V2,F1)     1.2892989  0.04580930 28.1449160
(V4,F2)     0.8755553  0.02442281 35.8498963
```

Appendix C. EQS Code for the Model in Equations (12) and (13)

```
/TITLE
EQS 6.1: Mean and covariance structure analysis.
The file name is mcov.eqs.
/SPECIFICATION
weight="weight.txt";
cases=88; variables=4; matrix=covariance;
analysis=moment; methods=ML, robust; data="data.txt";
/LABELS
V1=Mechanics; V2=Vectors; V3=Analysis; V4=Statistics;
/EQUATIONS
V1=  F1+E1;
V2= *F1+E2;
V3=  F2+E3;
V4= *F2+E4;
F1= *V999+D1;
```

```
F2= *V999+D2;
/VARIANCES
E1-E4= *;
D1=*;
D2=*;
/COVARIANCES
D1,D2= *;
/TECHNICAL
conv=0.0001;
itera=500;
/Means
/INEQUALITY
(E3, E3)>-100;
/OUTPUT
CODEBOOK;
DATA="mcov.ETS";
PARAMETER ESTIMATES;
STANDARD ERRORS;
LISTING;
/END
```

Appendix D. EQS Code for Confirmatory Factor Analysis with Four Variables

```
/TITLE
EQS 6.1: Covariance structure analysis. The file name is cov.eqs
/SPECIFICATION
weight="weight.txt";
cases=88; variables=4; matrix=covariance;
analysis=covariance; methods=ML, robust; data="data.txt";
/LABELS
V1=Mechanics; V2=Vectors; V3=Analysis; V4=Statistics;
/EQUATIONS
V1=  F1+E1;
V2= *F1+E2;
V3=  F2+E3;
V4= *F2+E4;
/VARIANCES
E1-E4= *;
F1=*;
F2=*;
/COVARIANCES
F1,F2= *;
/TECHNICAL
conv=0.0001;
itera=500;
/OUTPUT
CODEBOOK;
DATA="cov.ETS";
PARAMETER ESTIMATES;
STANDARD ERRORS;
LISTING;
/END
```

Appendix E. EQS Code for the Unconditional Latent Growth Curve Model in Equation (14)

```
/TITLE
Unconditional latent growth curve analysis.
```

```
The file name is nlsy4.eqs.
/SPECIFICATION
weight="weight.txt";
cases=399; variables=4; matrix=covariance;
analysis=moment; methods=ML, robust; data="data.txt";
/EQUATIONS
V1=  F1 + E1;
V2=  F1 + F2 + E2;
V3=  F1 + 2F2 + E3;
V4=  F1 + 3F2 + E4;
F1= *V999+D1;
F2= *V999+D2;
/VARIANCES
E1-E4= *;
D1=*;
D2=*;
/COVARIANCES
D1,D2= *;
/TECHNICAL
conv=0.0001;
itera=500;
/Means
/OUTPUT
CODEBOOK;
DATA="nlsy4.ETS";
PARAMETER ESTIMATES;
STANDARD ERRORS;
LISTING;
/END
```

Appendix F. EQS Code for the Conditional Latent Growth Curve Model in Equations (15) and (16)

```
/TITLE
Conditional latent growth curve analysis.
The file name is nlsy4p.eqs.
/SPECIFICATION
weight="weight.txt";
cases=399; variables=7; matrix=covariance;
analysis=moment; methods=ML, robust; data="data.txt";
/EQUATIONS
V1=  F1 + E1;
V2=  F1 + F2 + E2;
V3=  F1 + 2F2 + E3;
V4=  F1 + 3F2 + E4;
V5=*V999+E5;
V6=*V999+E6;
V7=*V999+E7;
F1= *V999+*V5+*V6+*V7+D1;
```

```
F2= *V999+*V5+*V6+*V7+D2;
/VARIANCES
E1-E7= *;
D1=158*;
D2=7*;
/COVARIANCES
D1,D2= *;
E5,E6=*;
E5,E7=*;
E6,E7=*;
/TECHNICAL
conv=0.0001;
itera=1000;
/Means
/OUTPUT
CODEBOOK;
DATA="nlsy4p.ETS";
PARAMETER ESTIMATES;
STANDARD ERRORS;
LISTING;
/END
```

References

Arminger, G., & Sobel, M.E. (1990). Pseudo-maximum likelihood estimation of mean and covariance structures with missing data. *Journal of the American Statistical Association*, *85*, 195–203.

Bentler, P.M. (2008). *EQS 6 structural equations program manual*. Encino: Multivariate Software.

Bentler, P.M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: test statistics. *Multivariate Behavioral Research*, *34*, 181–197.

Browne, M.W. (1984). Asymptotic distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical & Statistical Psychology*, *37*, 62–83.

Cheng, T.-C., & Victoria-Feser, M.-P. (2002). High-breakdown estimation of multivariate mean and covariance with missing observations. *British Journal of Mathematical & Statistical Psychology*, *55*, 317–335.

D'Agostino, R.B., Belanger, A., & D'Agostino, R.B. Jr. (1990). A suggestion for using powerful and informative tests of normality. *American Statistician*, *44*, 316–321.

Enders, C.K. (2010). *Applied missing data analysis*, New York: Guilford.

Enders, C.K., & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., & Stahel, W.A. (1986). *Robust statistics: the approach based on influence functions*. New York: Wiley.

Hu, L., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351–362.

Huber, P.J. (1981). *Robust statistics*. New York: Wiley.

Lee, S.Y., & Xia, Y.M. (2006). Maximum likelihood methods in treating outliers and symmetrically heavy-tailed distributions for nonlinear structural equation models with missing data. *Psychometrika*, *71*, 565–585.

Lee, S.Y., & Xia, Y.M. (2008). A robust Bayesian approach for structural equation models with missing data. *Psychometrika*, *73*, 343–364.

Little, R.J.A. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics*, *37*, 23–38.

Liu, C. (1997). ML estimation of the multivariate t distribution and the EM algorithm. *Journal of Multivariate Analysis*, *63*, 296–312.

Lopuhaä, H.P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariances. *Annals of Statistics*, *17*, 1662–1683.

Mair, P., Wu, E., & Bentler, P.M. (2010). EQS goes R: simulations for SEM using the package REQS. *Structural Equation Modeling*, *17*, 333–349.

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

Mardia, K.V., Kent, J.T., & Bibby, J.M. (1979). *Multivariate analysis*. New York: Academic Press.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166.

Preacher, K.J., Wichman, A.L., MacCallum, R.C., & Briggs, N.E. (2008). *Latent growth curve modeling*. Thousand Oaks: Sage.

Raykov, T. (2005). Analysis of longitudinal studies with missing data using covariance structure modeling with full-information maximum likelihood. *Structural Equation Modeling*, *12*, 493–505.

Rocke, D.M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *Annals of Statistics*, *24*, 1327–1345.

Rubin, D.B. (1976). Inference and missing data (with discussions). *Biometrika*, *63*, 581–592.

Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), *Latent variables analysis: applications for developmental research* (pp. 399–419). Newbury Park: Sage.

Savalei, V., & Bentler, P.M. (2009). A two-stage ML approach to missing data: theory and application to auxiliary variables. *Structural Equation Modeling*, *16*, 477–497.

Savalei, V., & Falk, C. (in press) Robust two-stage approach outperforms robust FIML with incomplete non-normal data. *Structural Equation Modeling*.

Schott, J. (2005). *Matrix analysis for statistics* (2nd ed.). New York: Wiley.

Tong, X., Zhang, Z., & Yuan, K.-H. (2011, October). *Evaluation of test statistics for robust structural equation modeling with non-normal missing data*. Paper presented at the graduate student pre-conference of the annual meeting of the society of multivariate experimental psychology, Norman, OK.

Yuan, K.-H. (2011). *Expectation-robust algorithm and estimating equation for means and covariances with missing data*. Manuscript under review.

Yuan, K.-H., & Bentler, P.M. (1997). Improving parameter tests in covariance structure analysis. *Computational Statistics & Data Analysis*, *26*, 177–198.

Yuan, K.-H., & Bentler, P.M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical & Statistical Psychology*, *51*, 289–309.

Yuan, K.-H., & Bentler, P.M. (2000). Three likelihood-based methods for mean and covariance structure analysis with non-normal missing data. *Sociological Methodology*, *30*, 167–202.

Yuan, K.-H., & Bentler, P.M. (2001). A unified approach to multigroup structural equation modeling with nonstandard samples. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling: new developments and techniques* (pp. 35–56). Mahwah: Lawrence Erlbaum Associates.

Yuan, K.-H., & Bentler, P.M. (2010). Two simple approximations to the distributions of quadratic forms. *British Journal of Mathematical & Statistical Psychology*, *63*, 273–291.

Yuan, K.-H., Bentler, P.M., & Chan, W. (2004a). Structural equation modeling with heavy tailed distributions. *Psychometrika*, *69*, 421–436.

Yuan, K.-H., & Jennrich, R.I. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis*, *65*, 245–260.

Yuan, K.-H., Lambert, P.L., & Fouladi, R.T. (2004b). Mardia's multivariate kurtosis with missing data. *Multivariate Behavioral Research*, *39*, 413–437.

Yuan, K.-H., & Lu, L. (2008). SEM with missing data and unknown population using two-stage ML: theory and its application. *Multivariate Behavioral Research*, *62*, 621–652.

Yuan, K.-H., Marshall, L.L., & Bentler, P.M. (2002). A unified approach to exploratory factor analysis with missing data, non-normal data, and in the presence of outliers. *Psychometrika*, *67*, 95–122.

Yuan, K.-H., Wallentin, F., & Bentler, P.M. (in press) ML versus MI for missing data with violation of distribution conditions. *Sociological Methods & Research*.

Zhong, X., & Yuan, K.-H. (2011). Bias and efficiency in structural equation modeling: maximum likelihood versus robust methods. *Multivariate Behavioral Research*, *46*, 229–265.

Zu, J., & Yuan, K.-H. (2010). Local influence and robust procedures for mediation analysis. *Multivariate Behavioral Research*, *45*, 1–44.

Manuscript Received: 5 OCT 2011
Final Version Received: 12 DEC 2011
Published Online Date: 22 AUG 2012