Residual Plots to Identify Outliers in Structural Equation Modeling

Laura A. Hildreth* Frederick O. Lorenz†

Abstract

Residual plots are routinely used in regression analysis to evaluate underlying model assumptions and identify potential outliers. Similar graphical tools are considerably more difficult to construct under the framework of structural equation modeling (SEM) as the use of latent variables complicates the construction of such residual-based diagnostics. The purpose of this paper is to introduce a method to construct residual plots under the SEM framework. First we present a class of residual estimators that are weighted linear functions of the observed variables. We then propose a method to construct residual plots under the SEM framework analogous to "residuals versus fitted values plots" in regression analysis. The utility of these plots to identify potential outliers is demonstrated by implementing our method using Mardia's exam data. This example illustrates that the choice of residual estimator affects the residual plots and provides insight into which residual estimator is the "best." Compared to previous diagnostics to identify outliers in SEM, we develop a graphical diagnostic that not only provides a more consistent method of identifying outliers but also identifies why an observation is outlying. Ultimately, this work provides the foundation for the use of residual plots and future development of residual analysis in SEM.

Key Words: structural equation modeling, residual plots, outliers, residual analysis

1. Introduction

Structural equation modeling (SEM) is a popular statistical methodology frequently used in the social and behavioral sciences that is becoming increasingly popular in the biological sciences and business. This popularity is due to the flexibility of SEM framework which allows for the modeling of complex systems of equations, the use of latent (unobserved) variables, and variables measured with error. While several methods exist to obtain parameter estimates, maximum likelihood estimation is most commonly used, making the validity of SEM analyses sensitive to the presence outliers. Consequently, the identification of potential outliers via model diagnosis is crucial in SEM. In many statistical methodologies, such as regression analysis, potential outliers are identified using graphical tools (e.g. scatterplots, QQ plots, and residual plots) or numerical diagnostics (e.g. Cook's distance, leverage, and standardized residuals) both of which frequently use residuals. Due to the complexity of structural equation models and the use of latent variables, few analogous tools and diagnostics have been developed for SEM as complications arise when calculating residuals. (Bollen and Arminger, 1991). The purpose of this paper is to introduce a method to construct residual plots under the SEM framework to aid in the identification of outliers. The plots we propose extend the standard "residuals versus fitted values" residual plots frequently used in regression analysis to the SEM framework. These plots allow researchers to visually assess the presence of outliers and identify in what portion(s) of the model observations are outlying.

As explained by Yuan and Hayashi (2010), the use of model diagnostics is crucial in any statistical analysis but is even more so in SEM due to the increased complexity of these models. While model diagnostics for SEM are not as extensively developed as in regression analysis, there have been several developments. The most basic of these diagnostics are distance measures including Bollen's $\boldsymbol{A}$ (Bollen, 1989; Mullen,

*Montana State University, Bozeman, MT

†Iowa State University, Ames, IA

Milne, and Doney, 1995) which is similar to the hat matrix in regression analysis and Mahalanobis distance (Campbell, 1980; Mullen et al, 1995; Yuan, Fung, and Reise, 2004). Several other diagnostics have been developed to assess local influence on the likelihood function (Coffman and Millsap, 2006; Ferrando, 2007, 2010; Lee and Lu, 2003; Rensvold and Cheung, 1999) and the observed covariance matrix (Mullen et al, 1995; Poon and Wong, 2004). A similar class of diagnostics adapt Cook's (1986) perturbation scheme to identify observations that are most influential on global fit indices such as the χ^2 statistic and the goodness of fit statistic (Cadigan, 1995; Lee, Lu, and Song, 2006; Lee and Tang, 2004; Lee and Wang, 1996; Poon, Wang, and Lee, 1999).

The use of residuals in model diagnosis was originally introduced by Bollen and Armminger (1991) who developed a method to calculate residuals. The utility of these residuals to identify outliers was demonstrated through the use of stem-and-leaf plots and QQ plots. Though the development and use of these residuals has been limited, their theoretical properties have been further investigated by Raykov and Penev (2001) and Hildreth (2013) and several numerical and graphical diagnostics have been introduced. Sánchez, Houseman, and Ryan (2009) developed several test statistics by extending the Kolmogorov-Smirnov and Cramér-von Mises tests to the SEM framework to assess the normality and linearity assumptions. Yuan and Hayashi (2010) introduced two scatterplots that use the residuals developed by Bollen and Arminger (1991) to aid in the identification of outliers which are available in the R package `semdiag` (Yuan and Zhang, 2012). The first plot displays the Mahalanobis distance of the residuals against the Mahalanobis distance of the estimated values of the latent variables and the second is a QQ plot of the Mahalanobis distance against quantiles of the χ^2 distribution.

The residual plots we propose build on this previous work but differ markedly in several crucial ways. First, the residual plots we propose are able to identify outliers at the equation level; that is, the plots are able to identify in what equation(s) of the model an observation may be outlying. This differs from many previous diagnostics which identify outliers at the model level such that the diagnostics identify an observation as potentially outlying but cannot determine in what regard the observation is outlying. While this may lead to the construction of a relatively large number of plots, it provides additional information that could not otherwise be obtained. Second, we extend tools that are already familiar to SEM practitioners to the SEM framework. The plots we propose can be interpreted analogously to residual plots in regression analysis which allows SEM practitioners to use the skills they already possess and extend them to the SEM framework.

The identification of outliers is a crucial part of model diagnosis, especially in SEM due to the complexity of the models. In the next section, we introduce the standard SEM notation and present the residuals introduced by Bollen and Arminger (1991). We then describe our method to construct residual plots and assess the utility of these plots by analyzing a data set frequently studied in the literature. From this analysis we are able to show the viability of these plots and recommend future research in this area in our conclusion and discussion.

2. Notation for Structural Equation Modeling and Its Associated Residual Estimators and Residuals

We use the notation from Bollen (1989) to introduce the standard structural equation model and the residual estimators of Bollen and Arminger (1991). Assume that there are $i = 1, \ldots, N$ independent individuals such that the equations representing a structural equation model are at the individual level. To avoid confusion arising from multiple subscripts, we omit the subscript i from our notation as done by Bollen (1989). A structural equation model consists of two systems of equations: the *latent variable model*, representing the relationships among the latent variables, and the *measurement model*, representing the relationships

between the latent and observed variables. The latent variable model describes the relationships among the m *endogenous* and n *exogenous* latent variables denoted as $\boldsymbol{\eta}_{m\times 1}$ and $\boldsymbol{\xi}_{n\times 1}$, respectively. The endogenous latent variables are described using p observed variables $\boldsymbol{x}_{p\times 1}$, while the exogenous latent variables are represented through q observed variables $\boldsymbol{y}_{q\times 1}$ (for notational ease we indicate the dimension of a vector or matrix the first time it is defined).

Without loss of generality, we assume that all variables (latent and observed) are deviated from their respective means leading to the following model:

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (1)$$

$$\boldsymbol{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon} \quad (2)$$

$$\boldsymbol{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}. \quad (3)$$

In equation (1), $\boldsymbol{\zeta}_{m\times 1}$ represents the *latent errors in equations* associated with $\boldsymbol{\eta}$, the matrix $\boldsymbol{B}_{m\times m}$ contains the coefficients relating the endogenous latent variables to one another such that $(\boldsymbol{I}-\boldsymbol{B})$ is non-singular and $\text{diag}(\boldsymbol{B}) = \boldsymbol{0}$, and $\boldsymbol{\Gamma}_{m\times n}$ contains the coefficients linking the exogenous latent variables $\boldsymbol{\xi}$ to the endogenous latent variables $\boldsymbol{\eta}$. In equation (2), $\boldsymbol{\epsilon}_{p\times 1}$ contains the *measurement errors of* $\boldsymbol{y}$ and $\boldsymbol{\Lambda}_{y,\,p\times m}$ represents the matrix of coefficients relating the endogenous latent variables $\boldsymbol{\eta}$ to its items $\boldsymbol{y}$. The *measurement errors of* $\boldsymbol{x}$ are denoted by $\boldsymbol{\delta}_{q\times 1}$ and the matrix $\boldsymbol{\Lambda}_{x,\,q\times n}$ in equation (3) consists of coefficients relating the exogenous latent variables $\boldsymbol{\xi}$ to its items $\boldsymbol{x}$. Regarding the error terms $\boldsymbol{\delta}$, $\boldsymbol{\epsilon}$, and $\boldsymbol{\zeta}$ we assume:

A1. $\text{E}(\boldsymbol{\delta}) = \boldsymbol{0}$, $\text{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$, and $\text{E}(\boldsymbol{\zeta}) = \boldsymbol{0}$.

A2. $\text{Var}(\boldsymbol{\delta}) = \mathbf{c}_{\boldsymbol{\delta}}$, $\text{Var}(\boldsymbol{\epsilon}) = \mathbf{c}_{\boldsymbol{\epsilon}}$, and $\text{Var}(\boldsymbol{\zeta}) = \mathbf{c}_{\boldsymbol{\zeta}}$ where $\mathbf{c}_{\boldsymbol{\delta}}, \mathbf{c}_{\boldsymbol{\epsilon}}$ and $\mathbf{c}_{\boldsymbol{\zeta}}$ are vectors of constants such that each error term is homoskedastic.

A3. $\boldsymbol{\delta}$, $\boldsymbol{\epsilon}$, and $\boldsymbol{\zeta}$ are uncorrelated across individuals within an equation.

A4. $\text{Cov}(\boldsymbol{\xi}, \boldsymbol{\zeta}) = \boldsymbol{0}$, implying the latent errors in equations are uncorrelated with the exogenous latent variables.

A5. $\text{Cov}(\boldsymbol{\eta}, \boldsymbol{\delta}) = \boldsymbol{0}$, $\text{Cov}(\boldsymbol{\eta}, \boldsymbol{\epsilon}) = \boldsymbol{0}$, $\text{Cov}(\boldsymbol{\xi}, \boldsymbol{\delta}) = \boldsymbol{0}$, $\text{Cov}(\boldsymbol{\xi}, \boldsymbol{\epsilon}) = \boldsymbol{0}$, and $\text{Cov}(\boldsymbol{\delta}, \boldsymbol{\epsilon}) = \boldsymbol{0}$, such that the measurement errors are uncorrelated with both the exogenous and the endogenous latent variables and each other.

Associated with the latent variable model in equation (1) are covariance matrices of the latent errors in equation and the exogenous latent variables, $\boldsymbol{\Psi}_{m\times m} = \text{E}(\boldsymbol{\zeta}\boldsymbol{\zeta}^\top)$ and $\boldsymbol{\Phi}_{n\times n} = \text{E}(\boldsymbol{\xi}\boldsymbol{\xi}^\top)$, respectively. The covariance matrix of the endogenous latent variables is defined in terms of model parameters such that

$$\boldsymbol{\Sigma}_{\eta\eta,\,m\times m} = \text{E}(\boldsymbol{\eta}\boldsymbol{\eta}^\top) = (\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}^\top + \boldsymbol{\Psi})(\boldsymbol{I}-\boldsymbol{B})^{-\top}.$$

The measurement model in equations (2) and (3) relates the observed variables $\boldsymbol{y}$ and $\boldsymbol{x}$ to the endogenous and exogenous latent variables $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, respectively. Associated with this model are covariance matrices $\boldsymbol{\Theta}_{\delta,\,q\times q} = \text{E}(\boldsymbol{\delta}\boldsymbol{\delta}^\top)$ and $\boldsymbol{\Theta}_{\epsilon,\,p\times p} = \text{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top)$. The covariance matrices of $\boldsymbol{y}$ and $\boldsymbol{x}$ are expressed as functions of the other model parameters where $\boldsymbol{\Sigma}_{yy\,p\times p} = \text{E}(\boldsymbol{y}\boldsymbol{y}^\top) = \boldsymbol{\Lambda}_y\boldsymbol{\Sigma}_{\eta\eta}\boldsymbol{\Lambda}_y^\top$ and $\boldsymbol{\Sigma}_{xx\,q\times q} = \text{E}(\boldsymbol{x}\boldsymbol{x}^\top) = \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x^\top$.

To simplify the notation of the measurement model and the resulting residual estimators, we follow the approach of Bollen and Arminger (1991) and let $\boldsymbol{z}_{(p+q)\times 1} = [\boldsymbol{y}^\top \quad \boldsymbol{x}^\top]^\top$, $\boldsymbol{L}_{(m+n)\times 1} = [\boldsymbol{\eta}^\top \quad \boldsymbol{\xi}^\top]^\top$, $\boldsymbol{\nu}_{(p+q)\times 1} = [\boldsymbol{\epsilon}^\top \quad \boldsymbol{\delta}^\top]^\top$, and $\boldsymbol{\Lambda}_{(p+q)\times(m+n)} = \text{diag}(\boldsymbol{\Lambda}_y, \boldsymbol{\Lambda}_x)$. The measurement model is rewritten as:

$$\boldsymbol{z} = \boldsymbol{\Lambda}\boldsymbol{L} + \boldsymbol{\nu}. \quad (4)$$

Associated with equation (4) are the covariance matrix of measurement errors $\boldsymbol{\Sigma}_{\nu\nu,\,(p+q)\times(p+q)} = \text{diag}(\boldsymbol{\Theta}_\epsilon, \boldsymbol{\Theta}_\delta)$, the covariance matrix of the latent variables

$$\boldsymbol{\Sigma}_{LL,\,(m+n)\times(m+n)} = \begin{bmatrix} \boldsymbol{\Sigma}_{\eta\eta,\,m\times m} & (\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi} \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}^\top(\boldsymbol{I}-\boldsymbol{B})^{-\top} & \boldsymbol{\Phi} \end{bmatrix},$$

and the covariance matrix of the observed variables

$$\boldsymbol{\Sigma}_{zz\,(p+q)\times(p+q)} = \text{E}(\boldsymbol{z}\boldsymbol{z}^\top) = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Lambda}_x^\top\boldsymbol{\Phi}\boldsymbol{\Gamma}^\top(\boldsymbol{I}-\boldsymbol{B})^{-\top}\boldsymbol{\Lambda}_y^\top \\ \boldsymbol{\Lambda}_y(\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x & \boldsymbol{\Sigma}_{xx} \end{bmatrix}.$$

Rearranging equation (4) yields the *measurement errors*:

$$\boldsymbol{\nu}(\boldsymbol{\theta}) = \boldsymbol{z} - \boldsymbol{\Lambda}\boldsymbol{L} \tag{5}$$

where $\boldsymbol{\theta}$ is a vector of the unique elements of $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Sigma}_{LL}$, and $\boldsymbol{\Sigma}_{\nu\nu}$ whose dimension depends on the model. Similarly, by rearranging equation (1) the *latent errors* are:

$$\boldsymbol{\zeta}(\boldsymbol{\theta}) \;\; = \;\; (\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\eta} - \boldsymbol{\Gamma}\boldsymbol{\xi} = [(\boldsymbol{I}-\boldsymbol{B})^{-1} \;\; -\boldsymbol{\Gamma}]\boldsymbol{L} = \boldsymbol{M}\boldsymbol{L} \tag{6}$$

where $\boldsymbol{M}_{m\times(m+n)} = [(\boldsymbol{I}-\boldsymbol{B})^{-1} \;\; -\boldsymbol{\Gamma}]$ and $\boldsymbol{\theta}$ is defined as above. In SEM, unlike other common statistical methods such as regression analysis, there is a *set* of $(p+q+m)$ residuals $\{\nu_1, \nu_2, \ldots, \nu_{(p+q)}, \zeta_1, \zeta_2, \ldots, \zeta_m\}$ associated with each individual.

The values of $\boldsymbol{L}$ in equations (5) and (6) are, by definition, unknown and must be estimated. These estimates, $\widehat{\boldsymbol{L}}$, are known as *factor scores* and frequently are constructed as linearly weighted functions of the observed variables in $\boldsymbol{z}$, such that $\widehat{\boldsymbol{L}} = \boldsymbol{W}_{(m+n)\times(p+q)}$ where $\boldsymbol{W}$ is referred to as the weight matrix (see DiStefano, Zhu, and Mindrila (2009) for a more thorough discussion of computing factor scores). Substituting the factor scores $\widehat{\boldsymbol{L}}$ for $\boldsymbol{L}$ in equation (5) we define the *measurement residuals*, denoted as $\widehat{\boldsymbol{\nu}}(\boldsymbol{\theta})$:

$$\widehat{\boldsymbol{\nu}}(\boldsymbol{\theta}) \;\; = \;\; \boldsymbol{z} - \boldsymbol{\Lambda}\widehat{\boldsymbol{L}} = \boldsymbol{z}\boldsymbol{\Lambda}\boldsymbol{W}\boldsymbol{z} = (\boldsymbol{I} - \boldsymbol{\Lambda}\boldsymbol{W})\boldsymbol{z}. \tag{7}$$

Similarly, the *latent residuals* $\widehat{\boldsymbol{\zeta}}(\boldsymbol{\theta})$ are obtained by substituting the factor scores $\widehat{\boldsymbol{L}}$ for $\boldsymbol{L}$ in equation (6):

$$\widehat{\boldsymbol{\zeta}}(\boldsymbol{\theta}) \;\; = \;\; \boldsymbol{M}\widehat{\boldsymbol{L}} = \boldsymbol{M}\boldsymbol{W}\boldsymbol{z}. \tag{8}$$

Several choices for $\boldsymbol{W}$ exist and we consider the three most commonly used weight matrices. The approaches used to derive the weight matrices minimize a specific loss function which depends on the method used. We briefly explain these below (further details are available in the original works and McDonald and Burr (1967)). The *regression method* (Thurstone, 1935) utilizes the principles of ordinary least squares to derive the weight matrix $\boldsymbol{W}_r$. Under this approach, $\boldsymbol{W} = \boldsymbol{W}_r$ is the solution that minimizes the loss function $\text{Tr}[\text{E}\{(\boldsymbol{L}-\widehat{\boldsymbol{L}})(\boldsymbol{L}-\widehat{\boldsymbol{L}})^\top\}]$ where Tr denotes the trace of a matrix where:

$$\boldsymbol{W}_r = \boldsymbol{\Sigma}_{LL}\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}_{zz}^{-1}. \tag{9}$$

Estimates calculated using $\boldsymbol{W}_r$ are referred to as the *regression-method residuals*. Under *Bartlett's method* (Bartlett, 1937), based on the principles of weighted least squares, the loss function is $\text{Tr}[\text{E}\{[\boldsymbol{\Sigma}_{\nu\nu}^{-1/2}\boldsymbol{\Lambda}(\boldsymbol{L}-\widehat{\boldsymbol{L}})][\boldsymbol{\Sigma}_{\nu\nu}^{-1/2}\boldsymbol{\Lambda}(\boldsymbol{L}-\widehat{\boldsymbol{L}})]^\top\}]$ yielding:

$$\boldsymbol{W}_b = (\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}_{\nu\nu}^{-1}. \tag{10}$$

Residual estimates calculated using $\boldsymbol{W}_b$ are the *Bartlett's method residuals*. The method by *Anderson-Rubin* (Anderson and Rubin, 1956) is an extension of Bartlett's method with the constraint of an orthogonal factor model. That is, $\boldsymbol{W} = \boldsymbol{W}_{ar}$ minimizes the loss function $\text{Tr}[\text{E}\{[\boldsymbol{\Sigma}_{\nu\nu}^{-1/2}\boldsymbol{\Lambda}(\boldsymbol{L} - \widehat{\boldsymbol{L}})][\boldsymbol{\Sigma}_{\nu\nu}^{-1/2}\boldsymbol{\Lambda}(\boldsymbol{L} - \widehat{\boldsymbol{L}})]^\top\}]$ subject to the constraint $\text{E}[\widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}^\top] = \boldsymbol{I}$. The resulting weight matrix is:

$$\boldsymbol{W}_{ar} = \boldsymbol{A}^{-1}\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}_{\nu\nu}^{-1}, \tag{11}$$

where $\boldsymbol{A}^2 = (\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{\Lambda})$. The *Anderson-Rubin method residuals* are calculated using $\boldsymbol{W}_{ar}$.

The work by Bollen and Arminger (1991) introduced the regression method and Bartlett's method residuals. Subsequent work has primarily focused on the Bartlett's method residuals (e.g. Raykov and Penev, 2001; Sánchez et al, 2009; and Yuan and Hayashi, 2010) as the resulting residuals are orthogonal to the true factor scores (Yuan and Hayashi, 2010) and are unbiased estimates of the true residuals (Hildreth, 2013). To the best of our knowledge, the Anderson-Rubin residuals have not been used or evaluated in previous studies. This likely is because in SEM the condition $\text{E}[\widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}^\top] = \boldsymbol{I}$ is not practical as the latent variables are correlated. However, for comparison purposes and to provide more information regarding the effects of the choice of $\boldsymbol{W}$, we chose to include this weight matrix in our study. Though orthogonality and unbiasedness are desirable properties, there may be situations where these properties are of secondary importance to other properties such as variability. Consequently, we evaluate all three methods to better understand similarities and differences among the estimators.

In practice the parameters in the vector $\boldsymbol{\theta}$ are unknown and must be estimated. Though several methods exist, the most frequently used method is maximum likelihood estimation under the assumption of normally distributed error terms and latent variables (for more details see Bollen (1989)). We assume that maximum likelihood estimation is used to obtain $\widehat{\boldsymbol{\theta}}$ which are used in place of $\boldsymbol{\theta}$ in equations (8) and (7) resulting in the *estimated measurement residuals*,

$$\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}}) = (\boldsymbol{I} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}})\boldsymbol{z}$$

and the *estimated latent residuals*

$$\widehat{\boldsymbol{\zeta}}(\widehat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{M}}\widehat{\boldsymbol{W}}\boldsymbol{z}.$$

These estimated residuals are the values that are used to construct our proposed residual plots. We chose to use the MLEs in the construction of our plots because these estimates are most frequently used in practice. While maximum likelihood estimation is sensitive to outliers as noted by Yuan and Hayashi (2010), we follow the process that practitioners commonly use. Because the focus of this paper is to present a method to construct residual plots, we do not further explore the effects of using maximum likelihood estimation on the plots, though it is an area of future research.

3. Construction of Residual Plots in Structural Equation Modeling

In regression analysis, there are several (scatter)plots that utilize the residuals to detect potential outliers and assumption violations. Kutner, Nachtsheim, Neter, and Li (2004) identify seven such plots: (1) plot of residuals against a predictor variable; (2) plot of absolute or squared residuals against a predictor variable; (3) plot of residuals against fitted or predicted values; (4) plot of residuals against time or another relevant sequence; (5) plot(s) of residuals against omitted predictor variables; (6) boxplots of residuals; and (7) normal probability plots of residuals. We propose extending the methodology of plotting the residuals against the fitted values to the SEM framework.

Consider the standard linear regression model $\boldsymbol{y} = \boldsymbol{X}^\top\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{y}$ is an $n \times 1$ vector of observed response variables, $\boldsymbol{X}$ is a $p \times n$ matrix of explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters,

and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of random errors term such that $\mathrm{E}(\boldsymbol{\epsilon}) = 0$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2$. Frequently the predicted values, $\widehat{\boldsymbol{y}} = \boldsymbol{X}^\top \widehat{\boldsymbol{\beta}}$ are plotted on the x-axis and the residuals, $\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{X}^\top \widehat{\boldsymbol{\beta}}$ on the y-axis. We propose extending this to the SEM framework as it is the most straightforward extension of this plot.

Under the SEM framework, the estimated measurement and latent residuals are:

$$\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}}) = (\boldsymbol{I} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}})\boldsymbol{z} \tag{12}$$

and

$$\widehat{\boldsymbol{\zeta}}(\widehat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{M}}\widehat{\boldsymbol{W}}\boldsymbol{z} \tag{13}$$

where $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{MLE}$ estimated under the assumption of normality.

Next, it is necessary to define the predicted values of $\boldsymbol{z}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{L} + \boldsymbol{\nu}$ and $\boldsymbol{\eta}(\boldsymbol{\theta}) = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$, denoted as $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\eta}}(\widehat{\boldsymbol{\theta}})$, respectively. As before we replace $\boldsymbol{L}$ with the factor scores $\widehat{\boldsymbol{L}} = \boldsymbol{W}\boldsymbol{z}$ and $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}}$ and define the predicted values as:

$$\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}}\boldsymbol{z} \tag{14}$$

and

$$\widehat{\boldsymbol{\eta}}(\widehat{\boldsymbol{\theta}}) = [\widehat{\boldsymbol{B}}\ \ \widehat{\boldsymbol{\Gamma}}]\widehat{\boldsymbol{W}}\boldsymbol{z}. \tag{15}$$

In regression analysis, the residuals $\boldsymbol{e}$ are plotted against the fitted values $\widehat{\boldsymbol{y}}$ because this plot displays random scatter if the appropriate model assumptions are met (Draper and Smith, 1981). Following this framework, we propose plotting the predicted values in equations (14) and (15) against the corresponding residuals in equations (12) and (13), respectively. However, under suitable regularity conditions (c.f. Miscellanea 10.6.2 of Casella and Berger (2003)), the estimated residuals and the predicted values are correlated in the SEM framework. (details are provided in the appendix). This poses a problem in the interpretation of the residual plots as the proposed residual plots would generally exhibit a linear trend even if all assumptions are met and no outliers are present. Consequently, the proposed plots cannot be interpreted analogously to residual plots in regression analysis. We propose transforming the data in such a way to make the estimated residuals and fitted values uncorrelated with one another using the inverse Cholesky transformation or decomposition (Greene, 2008).

The Cholesky decomposition is frequently used in financial and econometric applications to efficiently generate correlated random variables (e.g. Chan and Wong, 2013; Miranda and Fackler, 2002). Under the Cholesky decomposition, a positive definite matrix $\boldsymbol{\Sigma}$ is written as a product of a lower triangular matrix $\boldsymbol{L}$ and its transpose $\boldsymbol{U}$, an upper triangular matrix, such that $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{U}$ where $\boldsymbol{L}$ is the Cholesky factor of $\boldsymbol{\Sigma}$. For example, consider two independent variables X and Y. Using the matrix $\boldsymbol{L}$, it is possible to create two new random variables W and Z such that $\mathrm{Cov}[W, Z] = \boldsymbol{\Sigma}$. That is, each data point (x,y) is mapped to a new data point (w,z) by multiplying (x,y) by $\boldsymbol{L}$. Conversely suppose, as is done in this application, that the random variables W and Z are correlated with the covariance matrix $\boldsymbol{\Sigma}$. Then, it is possible to create two uncorrelated random variables X and Y using the matrix $\boldsymbol{L}^{-1}$. That is, each data point (w,z) is mapped to a new data point (x,y) by multiplying (w,z) by $\boldsymbol{L}^{-1}$.

The use of this transformation leads to several differences between the proposed residual plots in the SEM framework and the standard residual plots in regression analysis. First, the scale of the variables changes due multiplying the data by the matrix $\boldsymbol{L}^{-1}$. As a result the residuals, in general, will not be centered around 0 and the values of the predicted values, in general, are not necessarily values of the range of observed values in the data set. This necessitates a slight reorientation of reading a residual plot under the SEM framework. However, these plots should prove useful in the detection of outliers as outlying observations will fall far from the bulk of the data. Secondly, under this framework each SEM has a *set*

of residual plots associated with it. This implies that there are $p + q$ residual plots associated with the measurement model and m residual plots associated with the latent variable model.

The residual plots we propose differ markedly from previous diagnostics developed in the SEM framework to detect outliers. In general, previously developed diagnostics are useful in detecting potential outliers at the overall model level (e.g. Poon and Poon, 2002; Tanaka, Watadani, and Moon, 1991). That is, they are capable of detecting if an observation has a large influence on some element(s) of the model but they are not specific as to what portion(s) of the model the observation has a large influence. For example, Tanaka et al (1991) identify an observation as extremely influential but are not able to identify why that observation is influential. Instead, Tanaka et al (1991) compare this data point to the data set as a whole to speculate why this observation is influential. In contrast, our proposed residual plots are able to detect in what portion (equation) of the model an observation could be considered a potential outlier. Because each equation in the model has an associated residual plot, it is possible to detect what observation(s) is (are) potentially influential and in what equation the observation(s) is (are) influential. This provides considerably more information than previous diagnostics.

The previously described plots can be constructed using SAS 9.2. The general process is generalized in the following steps.

1. Center the data using `PROC STANDARD` [Note: the data need not be centered but this considerably simplifies later calculations].
2. Obtain parameter estimates $\widehat{\boldsymbol{\theta}}_{MLE}$ of the hypothesized model using `PROC CALIS`.
3. Calculate the following in `PROC IML`:
   (a) The estimated measurement residuals, $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})$, using equation (12). These estimates are then standardized by multiplying the estimates by the inverse square root of the appropriate diagonal element of the estimated covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\nu\nu} = (\boldsymbol{I} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}})\widehat{\boldsymbol{\Sigma}}_{zz}(\boldsymbol{I} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}})^{\top}$.
   (b) The estimated latent residuals, $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$, using equation (13). These estimates are also standardized by multiplying the estimates by the appropriate diagonal element of the estimated covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\zeta\zeta} = \widehat{\boldsymbol{M}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{\Sigma}}_{zz}\widehat{\boldsymbol{W}}^{\top}\widehat{\boldsymbol{M}}^{\top}$ where $\widehat{\boldsymbol{M}} = [\ (\boldsymbol{I} - \widehat{\boldsymbol{B}})^{-1} \quad -\widehat{\boldsymbol{\Gamma}}\ ]$.
   (c) The predicted values $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\eta}}(\widehat{\boldsymbol{\theta}})$ using equations (14) and (15), respectively.
4. Rotate the data in $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$ using the inverse Cholesky decomposition in `PROC IML` through the following steps:
   (a) Calculate the covariance matrix of the predicted values $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$, $\widehat{\boldsymbol{\Sigma}}_{\widehat{z}\widehat{z}} = \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{\Sigma}}_{zz}\widehat{\boldsymbol{W}}^{\top}\widehat{\boldsymbol{\Lambda}}^{\top}$.
   (b) Calculate the matrix of the covariances of the predicted values $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$ and and the estimated residuals $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})$ $\widehat{\boldsymbol{\Sigma}}_{\widehat{z}\nu} = \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{\Sigma}}_{zz}(\boldsymbol{I} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}})^{\top}$ such that the diagonal elements correspond to the covariance between predicted values and its corresponding estimated residuals.
   (c) For the set of predicted values $\widehat{\boldsymbol{z}}_j(\widehat{\boldsymbol{\theta}})$ and its corresponding estimated residuals $\widehat{\boldsymbol{\nu}}_j(\widehat{\boldsymbol{\theta}})$, where j denotes the equation such that $j = 1, \ldots, (p + q)$, construct the 2×2 covariance matrix $\boldsymbol{S}_{\widehat{z}\nu}$. The [1,1] diagonal element of $\boldsymbol{S}_{\widehat{z}\nu}$ is the jth diagonal element of $\widehat{\boldsymbol{\Sigma}}_{\widehat{z}\widehat{z}}$ (the variance of $\widehat{\boldsymbol{z}}_j(\widehat{\boldsymbol{\theta}})$), the [2,2] element of $\boldsymbol{S}_{\widehat{z}\nu}$ is set equal to 1 as the estimated residuals were standardized with a variance of 1, and the off-diagonal elements [1,2] and [2,1] of $\boldsymbol{S}_{\widehat{z}\nu}$ consist of the jth diagonal element from $\widehat{\boldsymbol{\Sigma}}_{\widehat{z}\nu}$ (the covariance of $\widehat{\boldsymbol{z}}_j(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\nu}}_j(\widehat{\boldsymbol{\theta}})$).

(d) Using $\boldsymbol{S}_{\widehat{z\nu}}$, obtain the inverse lower triangular matrix, $\boldsymbol{L}^{-1}$, such that $\boldsymbol{L}\boldsymbol{L}^{\top} = \boldsymbol{S}_{\widehat{z\nu}}$. Premultiply the data set $\widehat{\boldsymbol{z}}_j(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\nu}}_j(\widehat{\boldsymbol{\theta}})$ by $\boldsymbol{L}^{-1}$. This step can easily be done in SAS using the function `trisolv` in `PROC IML`. The resulting data set produces new variables, denoted as $\widehat{\boldsymbol{z}}_j(\widehat{\boldsymbol{\theta}})^R$ and $\widehat{\boldsymbol{\nu}}_j(\widehat{\boldsymbol{\theta}})^R$ such that $\text{Cov}[\widehat{\boldsymbol{z}}_j(\widehat{\boldsymbol{\theta}})^R, \widehat{\boldsymbol{\nu}}_j(\widehat{\boldsymbol{\theta}})^R] = \boldsymbol{I}$.

5. Rotate the data in $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$ using the inverse Cholesky decomposition in `PROC IML` through the following steps:

   (a) Calculate the covariance matrix of the predicted values $\widehat{\boldsymbol{\eta}}(\widehat{\boldsymbol{\theta}})$, $\widehat{\boldsymbol{\Sigma}}_{\widehat{\eta\eta}} = \widehat{\boldsymbol{D}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{\Sigma}}_{zz}\widehat{\boldsymbol{W}}^{\top}\widehat{\boldsymbol{D}}^{\top}$ where $\widehat{\boldsymbol{D}} = [\ \widehat{\boldsymbol{B}} \quad \widehat{\boldsymbol{\Gamma}}\ ]$.

   (b) Calculate the covariances matrix of the predicted values $\widehat{\boldsymbol{\eta}}(\widehat{\boldsymbol{\theta}})$ and and the estimated residuals $\widehat{\boldsymbol{\zeta}}(\widehat{\boldsymbol{\theta}})$ $\widehat{\boldsymbol{\Sigma}}_{\widehat{\eta\zeta}} = \widehat{\boldsymbol{M}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{\Sigma}}_{zz}\widehat{\boldsymbol{W}}^{\top}\widehat{\boldsymbol{D}}^{\top}$ such that the diagonal elements correspond to the covariance between predicted values and its corresponding estimated residual.

   (c) For the set of predicted values $\widehat{\boldsymbol{\eta}}_k(\widehat{\boldsymbol{\theta}})$ and its corresponding estimated residuals $\widehat{\boldsymbol{\zeta}}_k(\widehat{\boldsymbol{\theta}})$ where k denotes the equation such that $k = 1, \ldots, m$, construct the 2×2 covariance matrix $\boldsymbol{S}_{\widehat{\eta\zeta}}$. The [1,1] diagonal element of $\boldsymbol{S}_{\widehat{\eta\zeta}}$ is the kth diagonal element of $\widehat{\boldsymbol{\Sigma}}_{\widehat{\eta\eta}}$ (the variance of $\widehat{\boldsymbol{\eta}}_k(\widehat{\boldsymbol{\theta}})$), the [2,2] element of $\boldsymbol{S}_{\widehat{\eta\zeta}}$ is set equal to 1 as the estimated residuals were standardized with a variance of 1, and the off-diagonal elements [1,2] and [2,1] of $\boldsymbol{S}_{\widehat{\eta\zeta}}$ consist of the kth diagonal element from $\widehat{\boldsymbol{\Sigma}}_{\widehat{\eta\zeta}}$ (the covariance between $\widehat{\boldsymbol{\eta}}_k(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\zeta}}_k(\widehat{\boldsymbol{\theta}})$).

   (d) Using $\boldsymbol{S}_{\widehat{\eta\zeta}}$, obtain the inverse lower triangular matrix, $\boldsymbol{L}^{-1}$, such that $\boldsymbol{L}\boldsymbol{L}^{\top} = \boldsymbol{S}_{\widehat{\eta\zeta}}$. Premultiply the data set consisting of $\widehat{\boldsymbol{\eta}}_k(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\zeta}}_k(\widehat{\boldsymbol{\theta}})$ by $\boldsymbol{L}^{-1}$. As before, this step is easily done in SAS using the function `trisolv` in `PROC IML`. The resulting data set produces new variables, denoted as $\widehat{\boldsymbol{\eta}}_k(\widehat{\boldsymbol{\theta}})^R$ and $\widehat{\boldsymbol{\zeta}}_k(\widehat{\boldsymbol{\theta}})^R$ such that $\text{Cov}[\widehat{\boldsymbol{\eta}}_k(\widehat{\boldsymbol{\theta}})^R, \widehat{\boldsymbol{\zeta}}_k(\widehat{\boldsymbol{\theta}})^R] = \boldsymbol{I}$.

6. The residual plot for the jth equation of the measurement model is then obtained by plotting $\widehat{\boldsymbol{\zeta}}_j(\widehat{\boldsymbol{\theta}})^R$ on the y-axis against $\widehat{\boldsymbol{\eta}}_j(\widehat{\boldsymbol{\theta}})^R$ on the x-axis.

7. The residual plot for the jth equation of the latent variable model is then obtained by plotting $\widehat{\boldsymbol{\nu}}_k(\widehat{\boldsymbol{\theta}})^R$ on the y-axis against $\widehat{\boldsymbol{z}}_k(\widehat{\boldsymbol{\theta}})^R$ on the x-axis.

4. Example: Mardia Exam Data (1979)

The data for our example are from Mardia, Kent, and Bibby (1979). The data set contains exam scores from $n = 88$ students on five subjects in mathematics including mechanics (`mec`), vectors (`vec`), algebra (`alg`), analysis (`ana`), and statistics (`sta`). The mechanics and vectors exams were closed-book while the remaining three exams were open-book. Previous research (e.g. Mardia et al, 1979) indicates that a two factor model adequately fits the data where the first factor represents (latent) scores on closed-book exams and the second factor represents (latent) scores on open-book exams. These two factors are correlated though the error terms of the items (the exam scores for the five subjects) are not. Due to the adequacy of this model, we also use this two factor model to construct our proposed residual plots. The SAS code used to construct these plots is available from the first author.

These data have been used extensively in previous research as an empirical example to illustrate the utility of proposed diagnostics to detect outliers (c.f. Cadigan, 1995; Lee and Wang, 1996; Poon, Lew, and Poon, 2000; Poon and Poon, 2002; Tanaka et al, 1991; Yuan and Hayashi, 2010). We refer to these works for further details regarding the construction of these diagnostics. These previous studies indicate that several observations are potential outliers. Observations 81, 87, and 88 are frequently identified in these studies as outlying while observations 1, 2, and 3 are identified somewhat less frequently as outliers. When assessing the utility of our proposed residual plots we will primarily focus on these observations to evaluate how our plots compare to previous diagnostics. To aid in this, these observations are plotted in color (1 = orange, 2 = yellow, 3 = purple, 81 = red, 87 = blue, 88 = green).

4.1 Assessing the Effects of Using the Inverse Cholesky Decomposition

We first briefly examine the effects using the inverse Cholesky decomposition has on our proposed residual plots. Using the procedure from Section 3, we first calculated $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$ for each of the three weight matrices $\boldsymbol{W}_r$, $\boldsymbol{W}_b$, and $\boldsymbol{W}_{ar}$ defined in (9)-(11), respectively. We refer to these values as the *unrotated data* and the plots constructed using these data as the *unrotated residual plots*. Next, we applied the inverse Cholesky decomposition, as described previously, to obtain $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})^R$ and $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})^R$. These values are referred to as the *rotated data* and the subsequent plots constructed from these data as the *rotated residual plots*. We present the residual plots associated with the mechanics exam in Figure (4.1) (results were similar for the other four exams and thus were omitted). The plots in the left column correspond to the residual plots of $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})$ against $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$ constructed using each of the weight matrices under consideration while the plots in the right column are of $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})^R$ against $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})^R$ to illustrate the effect of rotating the data.

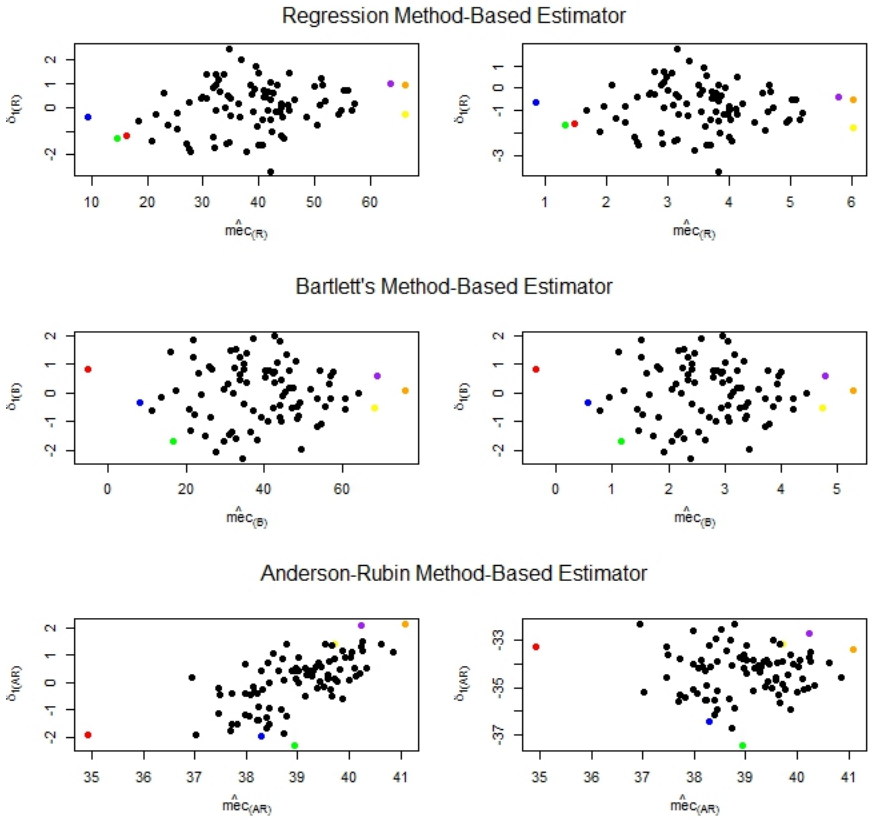
There are several features of note in these plots. First is that when rotating the data, as expected, the range of the y-axes corresponding with the residuals changes with the exception of the plot constructed using the Bartlett's method-based estimator. Similarly, the range of x-axes (the fitted values) also changes. Also as expected, the residuals may not be centered around 0 and the fitted values may not correspond to values in the range of the original data. Secondly, rotating the data essentially serves to rotate the shape seen in the residual plots, though this does not alter what observations would be detected as outliers. For example, for the plots constructed using the Anderson-Rubin method-based estimator, there is a strong linear relationship in the unrotated residual plot while the rotated residual plot does not display any discernible trend. However, both the unrotated and rotated residual plots detect observation 81 (shown in red) as outlying. Similar results are found in the unrotated and rotated residual plots for the other two methods (differences among the methods will be discussed shortly). This indicates that the inverse Cholesky decomposition is successful in creating residual plots that are similar to what we would expect in regression analyses while not affecting the utility of the plots in detecting potential outliers.

4.2 Assessing Utility to Detect Potential Outliers

We next assess the utility of the proposed residual plots to detect potential outliers and examine differences in the plots due to the choice of estimator. Figure (4.2) displays the rotated residual plots for each equation of the model and each estimator. First we compare these plots to previous studies using this data set to assess diagnostics for detecting outliers and then we evaluate differences among the plots in terms of the estimator used.

These plots indicate several potential outliers which differs depending on the residual of interest. In general the observations that have been frequently considered outliers are also outlying in our proposed residual plots. For example, observation 81 is frequently considered an outlier based on previous diagnos-

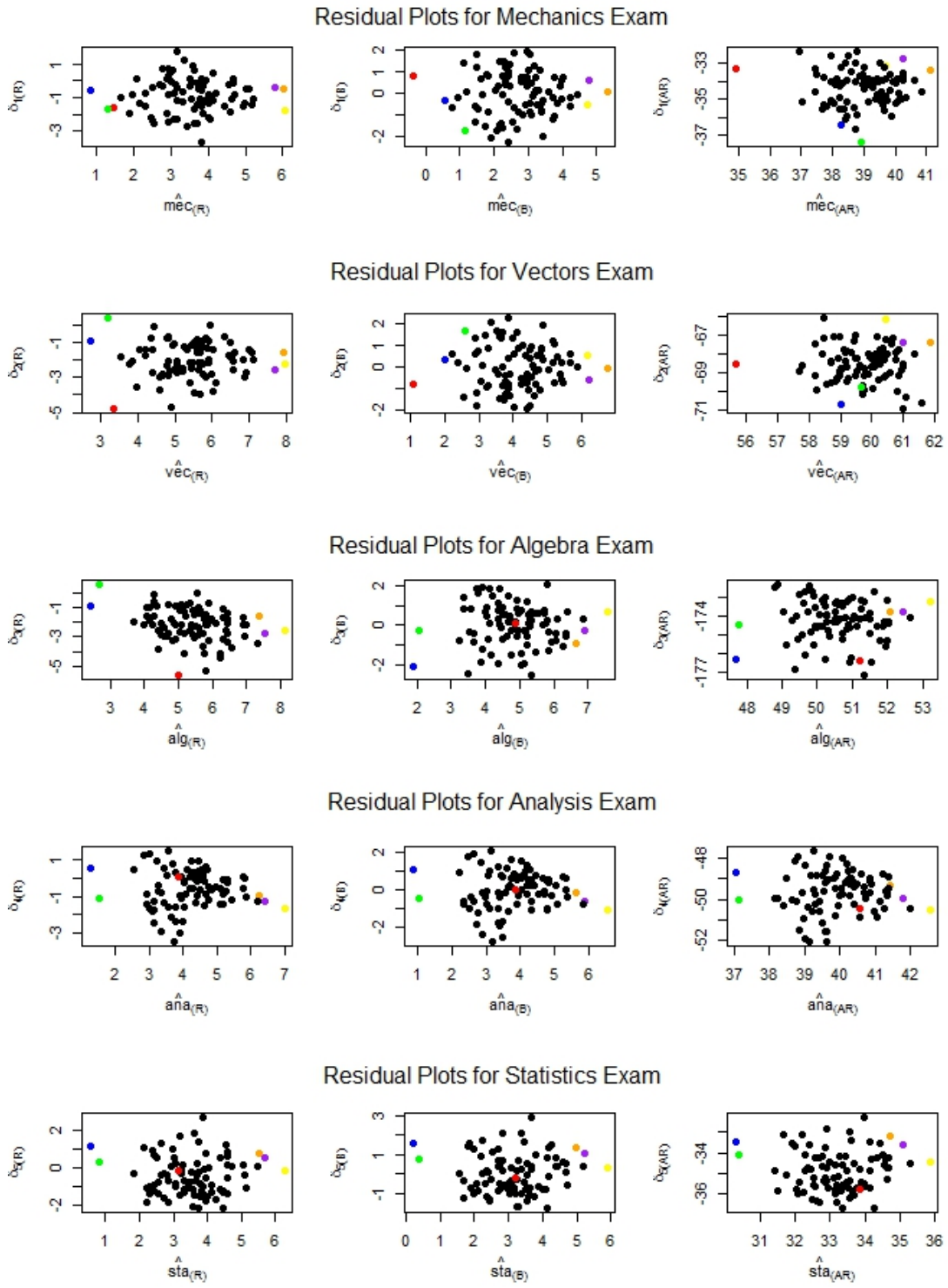
Figure 1: Residual Plots of the Unrotated and Rotated Data for the Mechanics Exam



tics. From our plots, observation 81 would be considered an outlier as well but through this analysis it is possible to determine in what regard observation 81 is considered an outlier. The residual plots indicate that observation 81 is an outlier in the equations for the `mec` and `vec` exams but none of the open book exams. In particular, the predicted scores on the mechanics and vector exams are comparatively small owing to the fact that this student performed poorly on the closed-book exams. Tanaka et al (1991) reach this conclusion by comparing the data associated with observation 81 to the other observations in the data set. For this application with five indicators and 88 observations it is possible to use this approach, however for larger data sets and more complex models this is not a viable option. Consequently, the proposed residual plots have utility in detecting potential outliers and are able to detect why these observations are outliers. This is a major advantage over previous diagnostics designed to detect outliers as these previous diagnostics are global diagnostics. In contrast, our proposed plots are able to identify potential outliers and to identify in what portion of the model the observation is outlying and why the observation may be outlying. This provides invaluable information to SEM practitioners and would allow for improved model assessment.

We are also interested in the effects the choice of residual estimator has on the residual plots. Interestingly, observations detected as potential outliers in the residual plots associated with closed book exams (`mec` and `vec`) depend on the estimator used while the residual plots associated with the open book exams (`alg`, `ana`, and `sta`) tend to detect the same observations as potential outliers. For example, in the residual plots for `mec`, the plots constructed using the Bartlett's method-based estimator and the Anderson-Rubin

Figure 2: Residual Plots for the Mardia Exam Data



method-based estimator both indicate that observation 81 (in red) is a potential outlier as this observation lies far to the left of the bulk of the observations. However, in the plot constructed using the regression method-based estimator observation 81 is closer to the bulk of the observations while observation 87 is further to the left and may be considered outlying. In contrast, in the residual plots for `alg`, `ana`, and `sta`, observations 87 and 88 (in blue and green, respectively) are far to the left of the bulk of the data and observation 2 (in yellow) is far to the right. Aside from what observations are detected as potential outliers, it is also notable that potential outliers tend to be easier to detect using the Anderson-Rubin method-based estimator plots as these potential outliers lie considerably further from the bulk of the data in comparison to the plots constructed under the other two estimators. Though the reason for this is not immediate, one potential explanation is that under the Anderson-Rubin method factor scores are obtained assuming an orthogonal factor model such that the Anderson-Rubin method-estimator is a type of canonical factor analysis (Jackson, 1991). Under this type of analysis, canonical variables (linear combinations of observed variables) are formed when the canonical correlation (correlation between two canonical variables) is maximized. According to Jackson (1991), a small number of outliers may lead to canonical correlations and plots of canonical variables or any rotation of these variables, such as the residuals and predicted values obtained using the Anderson-Rubin method-based estimator, can then easily detect the outlying observations (Jackson, 1991). This suggests that the Anderson-Rubin estimator may be particularly useful in isolating a small number of potential outliers.

5. Conclusions and Future Work

The purpose of this paper was to provide a method to construct residual plots in SEM and conduct an exploratory analysis of the utility of these plots using a well-known data set. Through this work we have shown that our proposed residual plots have the utility to detect potential outliers and are thus are worth pursuing in future research. There are several areas that need to be further explored to better develop the use of these residual plots. The next step in developing these plots is to conduct a thorough simulation study to better understand the properties of these plots. In particular, this simulation study would provide insight on how severely outlying an observation must be to be detected as a potential outlier as well as the most appropriate choice of estimator. Another area that needs to be addressed is that while we used the Cholesky inverse decomposition to "remove" the correlatedness between the residuals and predicted values, this rotation of the data does not take into account that residuals associated with different items or latent variables are correlated (more details are found in Hildreth, 2013). This is of concern as the identification of an outlier may be an artifact of problems in other equations of the model. A simulation study would be able to address this concern which would help to further the development of residual plots in SEM to detect outliers and also provide information for other purposes of residual plots such as the ability to detect assumption violations. Lastly, because we used maximum likelihood estimation a thorough simulation study would address the impacts of this estimation choice, in particular potential masking effects.

Though there are several areas that must be addressed in regards to our proposed residual plots, this work lays the foundation for the development of residuals plots and subsequently residual analysis in the SEM framework. Through this work we have proposed a useful method to calculate residuals and predicted values in SEM and how to plot them to construct residual plots analogous to those used in other statistical methods. This work indicates the promise of residual analysis in SEM.

REFERENCES

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the third Berkeley Symposium on mathematical statistics and probability*, *5*, 111–150.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, *28*, 97–104.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. and Arminger, G. (1991). Observational residuals in factor analysis and structural equation models. *Sociological Methodology*, *21*, 235–262.

Cadigan, N. G. (1995). Local influence in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *2(1)*, 13–30.

Campbell, N. A. (1980). Robust procedures in multivariate analysis I: Robust covariance estimation. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *29(3)*, 231–237.

Casella, G. and Berger, R. L. (2002). *Statistical inference* (2nd Edition). New York: Duxbury Press.

Chan, N. H. and Wong, H. Y. (2013). *Handbook of Financial Risk Management: Simulations and Case Studies*. USA: Wiley.

Coffman, D. L. and Millsap, R. E. (2006). Evaluating latent growth curve models using individual fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, *13(1)*, 1–27.

DiStefano, C., Zhu, M., and Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research and Evaluation*, *14(20)*.

Draper, N. and Smith, H. (1981). *Applied Regression Analysis* (2nd edition). USA: Wiley.

Ferrando, P. J. (2007). Multidimensional factor-analysis-based procedures for assessing scalability in personality measurements. *Structural Equation Modeling: A Multidisciplinary Journal*, *16(1)*, 109–133.

Ferrando, P. J. (2010). Some statistics for assessing person-fit based on continuous response models. *Applied Psychological Measurement*, *34(4)*, 219–237.

Greene, W. H. (2003). *Econometric Analysis* (6th Edition). New Jersey: Prentice Hall.

Hildreth, L. A. (2013). *Residual analysis for structural equation modeling* (Doctoral dissertation). Retrieved from ProQuest, UMI Dissertations Publishing (3597323).

Jackson, J. E. (1991). *A User's Guide to Principal Components*. Hoboken, New Jersey: Wiley.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004). *Applied Linear Statistical Models* (Fifth edition). New York, New York: McGraw-Hill/Irwin.

Lee, S. and Lu, B. (2003). Case-deletion diagnostics for nonlinear structural equation models. *Multivariate Behavioral Journal*, *38(3)*, 375–400.

Lee, S., Lu, B., and Song, X. (2006). Assessing local influence for nonlinear structural equation models with ignorable missing data. *Computational Statistics and Data Analysis*, *50*, 1356–1377.

Lee, S. and Tang, N. (2004). Local influence analysis of nonlinear structural equation models. *Psychometrika*, *69(4)*, 573-592.

Lee, S. and Wang, S. J. (1996). Sensitivity analysis of structural equation models. *Psychometrika*, *61(1)*, 93–108.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

McDonald, R. P. and Burr, E. J. (1967). A comparison of four methods of constructing factor scores. *Psychometrika*, *32(4)*, 381–401.

Miranda, M. J. and Fackler, P. L. (2002). *Applied Computational Economics and Finance*. Cambridge, MA: MIT Press.

Mullen, M. R., Milne, G. R., and Doney, P. M. (1995). An international marketing application of outlier analysis for structural equations: A methodological note. *Journal of International Marketing*, *3(1)*, 45–62.

Poon, W., Lew, S., and Poon, Y. S. (2002). A local influence approach to identifying multiple multivariate outliers. *British Journal of Mathematical and Statistical Psychology*, *53*, 255–273.

Poon, W., and Poon, Y. S. (2002). Influential observation in the estimation of the mean vector and covariance matrix. *British Journal of Mathematical and Statistical Psychology*, *55*, 177–192.

Poon, W., Wang, S, and Lee, S. (1999). Influence analysis of structural equation models with polytomous variables. *Psychometrika*, *64(4)*, 461–473.

Raykov, T. and Penev, S. (2001). Equivalent models and individual residuals. In G. A. Marcoulides and R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (297–321). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Rensvold, R. B., and Cheung, G. W. (1999). Identification of influential cases in structural equation models using the jackknife method. *Organizational Research Methods*, *2(3)*, 295–308.

Sánchez, B. N., Houseman, E. A. and Ryan, L. M. (2009). Residual-based diagnostics for structural equation models. *Biometrics*, *65*, 104-115.

Tanaka, Y., Watadani, S. and Moon, S. H. (1991). Influence in covariance structure analysis: with an application to confirmatory factor analysis. *Communications in Statistics–Theory and Methods*, *20(12)*, 3805–3821.

Thurstone, L. L. (1935). *The Vectors of the Mind*. Chicago: University of Chicago Press.

Yuan, K., Fung, W. K., and Reise, S. P. (2004). Three Mahalanobis distances and their role in assessing unidimensionality. *British Journal of Mathematical and Statistical Psychology*, *57*, 151–165.

Yuan, K., and Hayashi, K. (2010). Fitting data to model: Structural equation modeling diagnosis using two scatter plots. *Psychological Methods*, *15(4)*, 335–351.

Yuan, K. and Zhang, Z. (2012). Structural equation modeling diagnostics using R package semdiag. *Structural Equation Modeling: A Multidisciplinary Journal*, *19(4)*, 683–702.

A. Supplementary Material

First consider the covariance between $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$: $\text{Cov}[\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})] = \text{E}[\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})^\top]$. In general there are not closed-form solutions for the elements of $\widehat{\boldsymbol{\theta}}$ (Bollen, 1989) and thus it is not possible to explicitly derive this covariance. Instead, assume that the elements in $\boldsymbol{\theta}$ are known. Because the elements in $\boldsymbol{\theta}$ are estimated using maximum likelihood estimation, by the consistency property of MLEs (c.f. Theorem 10.1.6 of Casella and Berger, 2003), $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}}) \xrightarrow{d} \widehat{\boldsymbol{\nu}}(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}}) \xrightarrow{d} \widehat{\boldsymbol{z}}(\boldsymbol{\theta})$. Consequently, $\text{Cov}[\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})] \xrightarrow{d} \text{Cov}[\widehat{\boldsymbol{\nu}}(\boldsymbol{\theta}), \widehat{\boldsymbol{z}}(\boldsymbol{\theta})]$ under the assumption that the regularity conditions in Miscellanea 10.6.2 of Casella and Berger (2003) are satisfied. Then,

$$\begin{aligned} \text{Cov}[\widehat{\boldsymbol{\nu}}(\boldsymbol{\theta}), \widehat{\boldsymbol{z}}(\boldsymbol{\theta})] &= \text{E}[\widehat{\boldsymbol{\nu}}(\boldsymbol{\theta})\widehat{\boldsymbol{z}}(\boldsymbol{\theta})^\top] \\ &= \text{E}((\boldsymbol{I} - \boldsymbol{\Lambda}\boldsymbol{W})\boldsymbol{z}(\boldsymbol{\Lambda}\boldsymbol{W}\boldsymbol{z})^\top) \\ &= \text{E}((\boldsymbol{I} - \boldsymbol{\Lambda}\boldsymbol{W})\boldsymbol{z}\boldsymbol{z}^\top\boldsymbol{W}^\top\boldsymbol{\Lambda}^\top) \\ &= (\boldsymbol{I} - \boldsymbol{\Lambda}\boldsymbol{W})\boldsymbol{\Sigma}_{zz}\boldsymbol{W}^\top\boldsymbol{\Lambda}^\top \end{aligned} \tag{16}$$

such that = $\text{Cov}[\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})] = \boldsymbol{0}$ when one of the following three conditions hold: $\boldsymbol{\Lambda}\boldsymbol{W} = \boldsymbol{I}$, $\boldsymbol{\Lambda}\boldsymbol{W} = \boldsymbol{0}$, or $\boldsymbol{W} = \boldsymbol{\Lambda}^\top(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top)^{-1}$ [Note: $\boldsymbol{\Lambda}$ is an $(p+q) \times (m+n)$ matrix of rank $(m+n)$ where $(m+n) < (p+q)$, $\boldsymbol{\Lambda}^{-1}$ does not exist]. Because it is assumed that the elements of $\boldsymbol{\theta}$ are real-valued and, with the exception of $\boldsymbol{B}$, are not equal to $\boldsymbol{0}$ the condition $\boldsymbol{\Lambda}\boldsymbol{W} = \boldsymbol{0}$ is not possible and is thus disregarded. None of the weight matrices in (9)–(11) satisfy the condition $\boldsymbol{W} = \boldsymbol{\Lambda}^\top(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top)^{-1}$ and is disregarded for this paper. As is easily shown, the condition $\boldsymbol{\Lambda}\boldsymbol{W} = \boldsymbol{I}$ is satisfied when $\boldsymbol{W} = \boldsymbol{W}_b$ implying that $\widehat{\boldsymbol{\nu}}(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{z}}(\boldsymbol{\theta})$ are uncorrelated when using $\boldsymbol{W} = \boldsymbol{W}_b$. These results show that in general $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{z}}(\widehat{\boldsymbol{\theta}})$ are correlated except for when using the Bartlett's method estimator or if using a weight matrix where $\boldsymbol{W} = \boldsymbol{\Lambda}^\top(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top)^{-1}$.

Next, consider the covariance between $\widehat{\boldsymbol{\zeta}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\eta}}(\widehat{\boldsymbol{\theta}})$: $\text{Cov}[\widehat{\boldsymbol{\zeta}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{\eta}}(\widehat{\boldsymbol{\theta}})] = \text{E}[\widehat{\boldsymbol{\zeta}}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\eta}}(\widehat{\boldsymbol{\theta}})^\top]$. As noted before, closed-form solutions, in general, do not exist for $\widehat{\boldsymbol{\theta}}$. We use the analogous argument as above that under the necessary regularity conditions, by the consistency of MLEs, $\widehat{\boldsymbol{\zeta}}(\widehat{\boldsymbol{\theta}}) \xrightarrow{d} \widehat{\boldsymbol{\zeta}}(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\eta}}(\widehat{\boldsymbol{\theta}}) \xrightarrow{d} \widehat{\boldsymbol{\eta}}(\boldsymbol{\theta})$ implying that $\text{Cov}[\widehat{\boldsymbol{\zeta}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{\eta}}(\widehat{\boldsymbol{\theta}})] \xrightarrow{d} \text{Cov}[\widehat{\boldsymbol{\zeta}}(\boldsymbol{\theta}), \widehat{\boldsymbol{\eta}}(\boldsymbol{\theta})]$. Then,

$$\begin{aligned} \text{E}(\widehat{\boldsymbol{\zeta}}(\boldsymbol{\theta})\widehat{\boldsymbol{\eta}}(\boldsymbol{\theta})^\top) &= \text{E}(\boldsymbol{M}\boldsymbol{W}\boldsymbol{z}([\boldsymbol{B}\ \boldsymbol{\Gamma}]\boldsymbol{W}\boldsymbol{z})^\top) \\ &= \text{E}(\boldsymbol{M}\boldsymbol{W}\boldsymbol{z}\boldsymbol{z}^\top\boldsymbol{W}^\top[\boldsymbol{B}\ \boldsymbol{\Gamma}]^\top) \\ &= \boldsymbol{M}\boldsymbol{W}\boldsymbol{\Sigma}_{zz}\boldsymbol{W}^\top[\boldsymbol{B}\ \boldsymbol{\Gamma}]^\top. \end{aligned} \tag{17}$$

Under the assumption that the elements of $\boldsymbol{\theta}$ are real-valued and, with the exception of the matrix $\boldsymbol{B}$ are not equal to $\boldsymbol{0}$, $\text{Cov}[\widehat{\boldsymbol{\zeta}}(\boldsymbol{\theta}), \widehat{\boldsymbol{\eta}}(\boldsymbol{\theta})] = \boldsymbol{0}$ is not possible. Consequently, regardless of the choice of $\boldsymbol{W}$, $\widehat{\boldsymbol{\zeta}}(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\eta}}(\boldsymbol{\theta})$ are correlated.