Detecting Nonlinearities in Structural Equation Modeling Using Residual Plots

Laura A. Hildreth[*]

Abstract

The validity of analyses using structural equation modeling (SEM) relies on several assumptions, analogous to those in regression analysis. One assumption is that of linearity, implying that all equations are linear in both the (observed and latent) variables and the parameters. Violating this assumption is detrimental as it may cause biased parameter and standard error estimates and misleading model diagnostics and test statistics. Though several methods have been developed to estimate nonlinear effects they assume that the nature of the relationship is known a priori. Few diagnostics have been developed to assess the linearity assumption or to explore the nature of a nonlinear relationship. The purpose of this paper is to examine the use of residual plots to identify nonlinearities in SEM. The utility of these plots is assessed using simulated data. Results indicate that residual plots are able to detect nonlinear relationships and thus are a viable graphical diagnostic tool to evaluate the linearity assumption in SEM but the utility of the plots depends on how the residuals are calculated.

Key Words: structural equation modeling, residual analysis, residual plots, factor scores

1. Introduction

Structural equation modeling (SEM) is a commonly used statistical method in the social and behavioral sciences. This methodology is well-suited to model the types of relationships commonly observed and hypothesized in the social sciences as SEM allows for the modeling of systems of equations, the use of latent (unobserved) variables, and the use of variables measured with error. The validity of SEM analyses relies on several assumptions similar to those in regression (for more details see Bollen (1989)). One key assumption is linearity in that all equations are linear in both the variables and parameters. This can be a restrictive assumption as there are many theories in the social and behavioral sciences that hypothesize nonlinear relationships between and among variables such as the Yerkes-Dodson law in psychology (Yerkes and Dodson, 1908), chaos theory as applied in sociology, political science, and economics (Kiel and Elliott, 1996), and the law of diminishing marginal returns in economics.

The modeling of nonlinear relationships was first developed under the factor analytic framework to model nonlinear factor to item relationships (Etezadi-Amoli and McDonald, 1983; Gibson, 1959; McDonald, 1967). Modeling and estimating nonlinear relationships between factors in SEM was introduced in the seminal work of Kenny and Judd (1984) who developed an approach to estimate both quadratic and interaction effects. Because this method has several drawbacks (e.g. it is difficult to implement (Kelava and Brandt, 2009)) and the constraints only hold for normally distributed latent variables (Wall and Amemiya, 2001), several robust and distribution-free estimation procedures were developed. Such procedures include robust maximum likelihood methods (Jöreskog and Yang, 1996) and the two-staged least squares approach (Bollen, 1995; Bollen and Paxton, 1998). Because these methods do not perform well in finite samples (Brandt, Kelava, and Klein, 2014), methods have been developed that allow for the nonnormality that arises due to nonlinearities. These include maximum likelihood methods (Cudeck, Harring, and du Toit, 2009; Klein and

---

[*]Montana State University, Bozeman, MT 59717

Moosbrugger, 2000; Klein and Muthén, 2003; Lee and Zhu, 2002), the method of moments approach (Wall and Amemiya, 2003), Bayesian methods (Arminger and Muthén, 1998; Lee, Song, and Tang, 2007; Zhu and Lee, 1999), the unconstrained approach (Kelava and Brandt, 2009), and semiparametric methods (Bauer, 2005)

One major drawback of these models is that they assume the nature of relationships is known a priori and few diagnostics have been developed to evaluate the nature of the relationships in SEM (Bauer, Baldasaro, and Gottfredson, 2012). This is troublesome as Mooijaart and Satorra (2009) show that diagnostics of model fit are insensitive to the presence of nonlinear effects such that these commonly used diagnostics are not capable of identifying model misspecification due to nonlinear effects. Because of this, it is crucial diagnostics be developed to adequately and accurately assess the linearity assumption in SEM. Previous diagnostics include hypothesis tests developed by Klein and Schermelleh-Engel (2010) and Nestler (2015) and the application of structural equation mixture models to graphically identify nonlinear relationships (Bauer et al., 2012; Pek et al., 2015). The purpose of this paper is to evaluate the utility of residual plots in the SEM framework in detecting nonlinear relationships. The hypothesis tests developed by Klein and Schermelleh-Engel (2010) and Nestler (2015) are omnibus tests in that they detect a model misspecification due to a nonlinear relationship but are not able to specify where the nonlinearity is. The plots presented in this paper would allow for this identification to be made using a diagnostic tool familiar to SEM practitioners.

This paper is organized as follows. The next section presents the standard linear structural equation model and explains how the residuals can be used to detect nonlinear relationships. Section 3 briefly explains the construction of residual plots in SEM. A simulated example is provided in Section 4 and Section 5 concludes.

2. Using Residuals to Detect Nonlinearities

Consider the linear structural equation model for the ith of N independent observations:

$$\boldsymbol{\eta}_i = \boldsymbol{\tau} + \boldsymbol{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\zeta}_i \qquad (1)$$

$$\boldsymbol{y}_i = \boldsymbol{\alpha}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i \qquad (2)$$

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

To avoid confusion from multiple subscripts, the subscript i is subsequently omitted from subsequent equations though all equations presented are at the observational level. The first time a vector/matrix is defined its dimensions are provided in the subscript. Equation (1) represents the latent variable model which describes the relationships among the m endogenous and n exogenous latent variables, $\boldsymbol{\eta}_{m\times 1}$ and $\boldsymbol{\xi}_{n\times 1}$, respectively, where $\mathrm{E}(\boldsymbol{\eta}) = \boldsymbol{\kappa}_\eta$, $\mathrm{E}(\boldsymbol{\xi}) = \boldsymbol{\kappa}_\xi$, $\mathrm{E}(\boldsymbol{\xi}\boldsymbol{\xi}^\top) = \boldsymbol{\Phi}_{n\times n}$,

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

and $\mathrm{E}(\boldsymbol{\zeta}\boldsymbol{\xi}^\top) = \boldsymbol{0}$. The matrix $\boldsymbol{B}_{m\times m}$ contains the coefficients relating the endogenous latent variables to one another where $(\boldsymbol{I} - \boldsymbol{B})$ is nonsingular and $\mathrm{diag}(\boldsymbol{B}) = \boldsymbol{0}$, and the matrix $\boldsymbol{\Gamma}_{m\times n}$ contains the coefficients linking the exogenous latent variables $\boldsymbol{\xi}$ to the endogenous latent variables $\boldsymbol{\eta}$. The vector $\boldsymbol{\zeta}_{m\times 1}$ represents the latent errors in equations with $\mathrm{E}(\boldsymbol{\zeta}) = \boldsymbol{0}$ and $\mathrm{E}(\boldsymbol{\zeta}\boldsymbol{\zeta}^\top) = \boldsymbol{\Psi}_{m\times m}$.

The measurement model described in Equations (2) and (3) relates the latent variables $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ to the observed variables in $\boldsymbol{y}_{p\times 1}$ and $\boldsymbol{x}_{q\times 1}$ respectively, where, by assumption, $\mathrm{E}(\boldsymbol{x}) = \boldsymbol{\alpha}_x + \boldsymbol{\Lambda}_x\boldsymbol{\kappa}_\xi$ and $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\alpha}_y + \boldsymbol{\Lambda}_y\boldsymbol{\kappa}_\eta$, and $\boldsymbol{\Theta}_{y,\, p\times p} = \mathrm{E}(\boldsymbol{y}\boldsymbol{y}^\top) = \boldsymbol{\Lambda}_y\boldsymbol{\Theta}_\eta\boldsymbol{\Lambda}_y^\top$ and $\boldsymbol{\Theta}_{x,\, q\times q} = \mathrm{E}(\boldsymbol{x}\boldsymbol{x}^\top) = \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x^\top$. $\boldsymbol{\Lambda}_{y,\, p\times m}$ and $\boldsymbol{\Lambda}_{x,\, q\times n}$ are matrices of factor loadings

and $\epsilon_{p\times 1}$ and $\delta_{q\times 1}$ are the measurement errors such that $E(\epsilon) = \mathbf{0}$, $E(\delta) = \mathbf{0}$, $\Theta_{\delta, q\times q} = E(\delta\delta^\top)$, and $\Theta_{\epsilon, p\times p} = E(\epsilon\epsilon^\top)$. It is assumed that ϵ are uncorrelated with η, ξ, and δ and that δ is uncorrelated with η, ξ, and ϵ.

For simplicity, using the approach of Bollen and Arminger (1991) let $z_{(p+q)\times 1} = [y^\top \ x^\top]^\top$, $\alpha_{(p+q)\times 1} = [\alpha_y^\top \ \alpha_x^\top]^\top$, $L_{(m+n)\times 1} = [\eta^\top \ \xi^\top]^\top$, $\nu_{(p+q)\times 1} = [\epsilon^\top \ \delta^\top]^\top$, and $\Lambda_{(p+q)\times(m+n)} = \mathrm{diag}(\Lambda_y, \Lambda_x)$ such that $E(z) = \alpha + \Lambda\kappa$ where $\kappa_{(m+n)\times 1} = [\kappa_\eta^\top \ \kappa_\xi^\top]^\top$. The measurement model is rewritten as:

$$z = \alpha + \Lambda L + \nu. \tag{4}$$

Associated with the measurement model in Equation (4) are the covariance matrix of measurement errors $\Theta_{\nu, (p+q)\times(p+q)} = \mathrm{diag}(\Theta_\epsilon, \Theta_\delta)$, the covariance matrix of the latent variables

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

and the covariance matrix of the observed variables

$$\Theta_{z\ (p+q)\times(p+q)} = E(zz^\top) = \begin{bmatrix} \Theta_y & \Lambda_x^\top\Phi\Gamma^\top(I-B)^{-\top}\Lambda_y^\top \\ \Lambda_y(I-B)^{-1}\Gamma\Phi\Lambda_x & \Theta_x \end{bmatrix}.$$

Rearranging Equation (1) the latent errors are defined as

$$\zeta = [(I-B)^{-1} - \Gamma]L - \tau = ML - \tau \tag{5}$$

where $M_{m\times(m+n)} = [(I-B)^{-1} - \Gamma]$. Similarly, earranging Equation (4) provides the measurement errors

$$\nu = z - (\alpha - \Lambda L). \tag{6}$$

A key assumption of the model in Equations (4) and (1) is that the relationships between the endogenous latent variables η and the exogenous latent variables ξ are linear and the relationships between the latent variables L and each of the observed variables z are also linear. Let's consider the consequences of fitting a linear structural equation model in the presence of nonlinear relationships. For illustrative purposes a simple example is considered in which there is one exogenous and one endogenous latent variable each with three indicators. To better understand the information contained in the error terms we consider two scenarios: (1) the relationship of a latent variable (either endogenous or exogenous) and one of its indicators is quadratic and (2) the relationship between the exogenous latent variable and the endogenous quadratic. We assume that the latent variables are normally distributed to allow for tractable calculations and that the error terms of means of 0. While there are nonlinear relationships that are not quadratic, this serves our purpose of showing the information that the errors contain.

2.1 Scenario 1

First assume that there is a quadratic relationship between a latent variable and one of its indicators such that the true relationship is:

$$z = \alpha + \lambda L + \omega L^2 + \nu$$

where ω denotes the quadratic effect. Suppose the following linear model is erroneously fit to the data generated from the above relationship:

$$z = \alpha^* + \lambda^* L + \nu^*.$$

The superscript * is used to denote that these values will change due to the omission of the quadratic effect. We are concerned with examining the information contained in ν^*. Specifically it is of interest to determine if ν^* contains information about the nature of the true relationship between the latent variable and its indicator. Then,

$$
\begin{aligned}
\nu^* &= z - (\alpha^* + \lambda^* L) \\
&= \alpha + \lambda L + \omega L^2 + \nu - (\alpha^* + \lambda^* L) \\
&= \nu + [\alpha + \lambda L + \omega L^2 - (\alpha^* + \lambda^* L)].
\end{aligned}
$$

This shows that the errors of the misspecified linear model can be decomposed into two components: the true error (random error) and the discrepancy between the true and fitted models (model error). This shows that the error terms from a misspecified model contain valuable information about the nature of the model misspecification. It is also of interest to find $E(\nu^*)$ and $Var(\nu^*)$ which requires us to find λ^* and α^*. It can be shown that $\lambda^* = \lambda + 2\omega\kappa_L$ and $\alpha^* = \alpha + \omega\sigma_L - \omega\kappa_L^2$ where κ_L and σ_L are the population mean and standard deviation of the latent variable. Most notable from these results is that if the model is misspecified, these parameters are biased; however if $\kappa = 0$, as is the case when all latent variables are deviated from their means, the factor loadings are unbiased. Using this information it can be shown that $E(\nu^*) = 0$ and $Var(\nu^*) = \theta_\nu + 2\omega^2\sigma_L^2$ where θ_ν represents the variance of the error term ν. These results imply that model misspecifications do not affect the expected value but they do affect the variance such that the variance of the error terms from the misspecified model are more variable than the true variability of the error terms if the model were properly specified.

2.2 Scenario 2

Assume that the true relationship between the endogenous and exogenous latent variables is quadratic such that:

$$
\eta = \tau + \gamma\xi + \rho\xi^2 + \zeta
$$

where ρ is the quadratic effect. Suppose the following linear model is erroneously fit to the data generated by the above equation:

$$
\eta = \tau^* + \gamma^*\xi + \zeta^*
$$

where the superscript * again denotes that if the model is misspecified then the parameters from the linear model and the quadratic model differ. As before it is of interest to understand what information is contained in ζ^*. Then,

$$
\begin{aligned}
\zeta^* &= \eta - (\tau^* + \gamma^*\xi) \\
&= \tau + \gamma\xi + \rho\xi^2 + \zeta - (\tau^* + \gamma^*\xi) \\
&= \zeta + [\tau + \gamma\xi + \rho\xi^2 - (\tau^* + \gamma^*\xi)].
\end{aligned}
$$

As with the errors terms from a misspecified measurement model, the error terms from a misspecified latent variable model can also be decomposed into random error and model error that arises due to model misspecification. These error terms also contain valuable information regarding potential model misspecification that can be used to better understand the true relationship. As before we would like to find $E(\zeta^*)$ and $Var(\zeta^*)$. This requires us to calculate γ^* and τ^*. It can be shown that $\gamma^* = \gamma + 2\rho\kappa_\xi$ and $\tau^* = \tau + \rho\phi - \rho\kappa_\xi^2$ where ϕ is the variance of the η. These results imply that the path coefficient γ^* is biased unless the latent variables are deviated from their means. Using this information, it can be shown that $E(\zeta^*) = 0$ and $Var(\zeta^*) = \psi + 2\rho^2\phi^2$ where ψ is the variance of ζ. Similar to before,

the error terms from the misspecified model have a mean of 0 but the variance of the error terms are larger than the variance of the error terms from the true model.

In the derivations of the information contained in the measurement errors we did not consider endogenous and exogenous latent variables and their associated items separately as our results do not change if there is a model misspecification solely in the measurement model. One concern is that nonlinearities in the latent variable model would propagate through the measurement model such that the measurement errors associated with items of endogenous latent variable would contain information about model error in the latent variable model. Recall the relationship between an endogenous latent variable and one of its items: $y = \alpha + \lambda\eta + \epsilon$. Because $\epsilon = \alpha + \lambda\eta$ we were concerned that the effects of model misspecification in the latent variable model would be reflected in ϵ. However, it can be shown that ϵ does not contain information about misspecification in the latent variable model. This is useful in practice as this implies that the error terms contain information specific to misspecification in that particular equation of the model.

3. Predicted Residuals and Residual Plots

In regression analysis, model assumptions are frequently evaluated using residual plots of the residuals against the fitted values. This type of plot can be extended to the SEM framework by finding suitable ways to construct the residuals and the fitted values. First, we examine how to calculate the residuals in SEM.

In practice the latent and measurement errors ζ and ν are unknown and must be predicted. Both the latent and measurement errors are functions of model parameters and the values L. The model parameters can be estimated using a variety of different methods. For illustrative purposes we use maximum likelihood estimation under the assumption of normality (for more details see Bollen (1989)) to obtain parameter estimates. These estimates are then used in place of the model parameters in Equations (6) and (5).

The values of L in Equations (6) and (5) are, by definition, unknown and must be predicted. These values, $\widehat{L}$, are known as *factor scores* and frequently are constructed as linear weighted functions of the observed variables z, such that $\widehat{L} = W_{(m+n)\times(p+q)}$ where 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 $\widehat{L}$ for L in Equation (6) we define the (predicted) measurement residuals, denoted as $\widehat{\nu}$,:

$$
\begin{aligned}
\widehat{\nu} &= z - \widehat{\alpha} - \widehat{\Lambda}\widehat{L} &(7)\\
&= z - \widehat{\Lambda}\widehat{W}z - \widehat{\alpha} &(8)\\
&= (I - \widehat{\Lambda}\widehat{W})z - \widehat{\alpha}. &(9)
\end{aligned}
$$

Similarly, the (predicted) latent residuals $\widehat{\zeta}$ are obtained by substituting the factor scores $\widehat{L}$ for L in equation (5):

$$
\begin{aligned}
\widehat{\zeta} &= \widehat{M}\widehat{L} &(10)\\
&= \widehat{M}\widehat{W}z. &(11)
\end{aligned}
$$

Because it is assumed that we must use parameter estimates in place of parameters we omit the $\widehat{}$ from subsequent equations.

Several choices for 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 $\mathrm{Tr}[\mathrm{E}\{(\boldsymbol{L} - \widehat{\boldsymbol{L}})(\boldsymbol{L} - \widehat{\boldsymbol{L}})^\top]$ where Tr denotes the trace of a matrix such that:

$$\boldsymbol{W}_r = \boldsymbol{\Sigma}_{LL}\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}_{zz}^{-1}. \tag{12}$$

Residuals 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 minimized is $\mathrm{Tr}[\mathrm{E}\{[\boldsymbol{\Sigma}_{\nu\nu}^{-1/2}\boldsymbol{\Lambda}(\boldsymbol{L} - \widehat{\boldsymbol{L}})][\boldsymbol{\Sigma}_{\nu\nu}^{-1/2}\boldsymbol{\Lambda}(\boldsymbol{L} - \widehat{\boldsymbol{L}})]^\top\}]$ yielding:

$$\boldsymbol{W}_b = (\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}_{\nu\nu}^{-1}. \tag{13}$$

Residuals 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 $\mathrm{Tr}[\mathrm{E}\{[\boldsymbol{\Sigma}_{\nu\nu}^{-1/2}\boldsymbol{\Lambda}(\boldsymbol{L} - \widehat{\boldsymbol{L}})][\boldsymbol{\Sigma}_{\nu\nu}^{-1/2}\boldsymbol{\Lambda}(\boldsymbol{L} - \widehat{\boldsymbol{L}})]^\top\}]$ subject to the constraint $\mathrm{E}[\widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}^\top] = \boldsymbol{I}$. The resulting weight matrix is:

$$\boldsymbol{W}_{ar} = \boldsymbol{A}^{-1}\boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}_{\nu\nu}^{-1}, \tag{14}$$

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 $\boldsymbol{W}_{ar}$.

Next, we define the fitted values of $\boldsymbol{z} = \boldsymbol{\alpha} + \boldsymbol{\Lambda}\boldsymbol{L} + \boldsymbol{\nu}$ and $\boldsymbol{\eta} = \boldsymbol{\tau} + \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$, denoted as $\widehat{\boldsymbol{z}}$ and $\widehat{\boldsymbol{\eta}}$, respectively. As before we replace $\boldsymbol{L}$ with the factor scores $\widehat{\boldsymbol{L}}$ and parameter estimates are substituted for model parameters. The fitted values are defined as:

$$\widehat{\boldsymbol{z}} = \widehat{\boldsymbol{\alpha}} + \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{W}}\boldsymbol{z} \tag{15}$$

and

$$\widehat{\boldsymbol{\eta}} = \widehat{\boldsymbol{\tau}} + [\widehat{\boldsymbol{B}}\ \widehat{\boldsymbol{\Gamma}}]\widehat{\boldsymbol{W}}\boldsymbol{z}. \tag{16}$$

Next, the residuals can be plotted against their respective fitted values. In regression analysis this plot was chosen as it will display random scatter if all assumptions are satisfied (Draper and Smith, 1980). Unfortunately, under the SEM framework it is easy to show that, in general, the residuals and its respective fitted values are correlated (see Hildreth and Lorenz (2014) for details). Hildreth and Lorenz (2014) recommend "removing" this correlatedness using the Cholesky decomposition by premultiplying the residuals and their respective fitted values by the inverse Cholesky factor. For brevity we do not elaborate further on this process and refer the reader to Hildreth and Lorenz (2014) for details. The values obtained after premultiplying the residuals and its respective fitted values by the inverse Cholesky factor are then used when constructing the residual plot.

4. Simulated Example

We simulate data from a simple model with one exogenous and one endogenous factor each having three indicators that do not cross-load. We assume that the relationship between ξ and η is a moderate quadratic relationship such that:

$$\eta = 0.5\eta_1 - 0.35\eta_1^2 + \zeta \qquad y_1 = 1.0\eta_1 + \epsilon_1 \qquad x_1 = 1.0\xi_1 + \delta_1$$
$$y_2 = \lambda_2\eta_1 + \epsilon_2 \qquad x_2 = \lambda_5\xi_1 + \delta_2$$
$$y_3 = \lambda_3\eta_1 + \epsilon_3 \qquad x_3 = \lambda_6\xi_1 + \delta_3 \tag{17}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\nu}} = \boldsymbol{I}_6$, $\boldsymbol{\Psi} = [1]$, $\boldsymbol{\Phi} = [1]$, $\gamma_{11} = 0.5$, and $\lambda_2 = \lambda_3 = \lambda_5 = \lambda_6 = 0.7$ with a sample size of $n = 250$.

Figure 1: Residual Plots
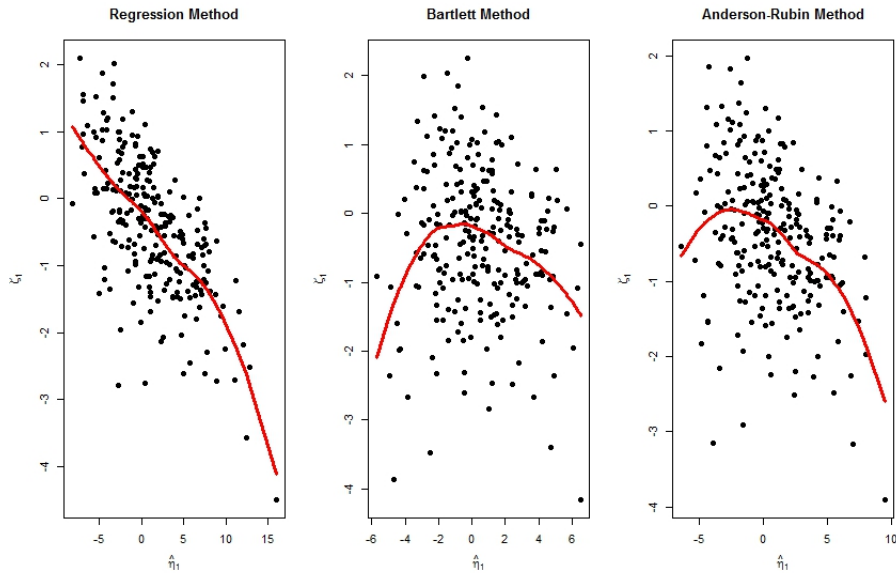
To construct the residual plots, we fit the linear structural equation model in R using the `lavaan` package (all R code is provided in Appendix A):

$$\eta = \gamma_{11}\xi + \zeta \qquad\qquad y_1 = 1.0\eta_1 + \epsilon_1 \qquad\qquad x_1 = 1.0\xi_1 + \delta_1$$
$$y_2 = \lambda_2\eta_1 + \epsilon_2 \qquad\qquad x_2 = \lambda_5\xi_1 + \delta_2$$
$$y_3 = \lambda_3\eta_1 + \epsilon_3 \qquad\qquad x_3 = \lambda_6\xi_1 + \delta_3. \qquad (18)$$

Using the parameter estimates, we calculate $\widehat{\zeta}$ and $\widehat{\eta}$ using each of the three methods for constructing factor scores. We then obtain the inverse Cholesky factor and premultiply the residuals and fitted values by this matrix. We use these values in the construction of our residual plots. For these proposed plots to be interpreted analogously to those in regression analysis, we would expect to see an inverse U shape.

Figure 4 displays the residual plots using the three methods for constructing factor scores for one simulated data set. Because results are similar for other simulated data sets we provide one set of plots. The red curve in each plot corresponds to the loess curve associated with the regression of $\widehat{\zeta}$ on $\widehat{\eta}$ which allows for a better inspection of any potential patterns in the residual plot. it is clear that the plot constructed using the regression method-based factor scores does not produce the type of nonlinear relationship we expect to see. The plots constructed using the Bartlett method-based and the Anderson-Rubin method-based factor scores, however, do indicate the type of relationship we expect to see. From these plots, it appears that the Bartlett method-based factor scores result in a residual plot that is closest to what we expect to as the right tail of the loess curve is longer than we would expect in the plot constructed using the Anderson-Rubin method-based factor scores.

This example suggests that there is potential for the use of residual plots in SEM to detect nonlinear relationships and that plots constructed using the Bartlett method-based factor scores produce a residual plot that can be interpreted analogously to those used in regression analysis. While this is only one example, it establishes the viability of residual plots in SEM and indicates that future research in this area is warranted.

5. Discussion and Conclusions

Many theories in the social and behavioral sciences hypothesize nonlinear relationships leading to the development of several methods for nonlinear SEM. One concern with these methods is that they assume that the nature of the relationship is known a priori. Consequently, it would be valuable to develop diagnostics to detect nonlinear relationships and the nature of these nonlinearities. Few diagnostics have been developed to check any of the assumptions in SEM as difficulties arise due to the use of latent variables. Theoretical results from the simple example of a two factor model with a quadratic relationship, show that the error terms contain information about model misspecification. This paper presented a method for constructing residual plots in SEM and illustrated their utility through a simulated example. This example also suggests that the best method for constructing factor scores is Bartlett's method as the resulting residual plot produces the type of relationship we expected to see.

Our theoretical results also suggest that the error terms contain information about issues with heteroskedasticity. Research on how heteroskedasticity can be detected in SEM using residual analysis is needed as heteroskedastic data are common in the social and behavioral sciences. The results of this paper indicate that residuals are useful in detecting model misspecifications as well as heteroskedasticity and consequently more attention on how to use the information contained in the residuals in warranted.

One concern with the construction of residual plots in SEM is that it could potentially lead to a large number of plots. Because of this, future research should focus on developing an omnibus test or graphical display to detect if there are issues with nonlinearities in the model. Based on this test, residual plots could be constructed if issues with nonlinearities are detected. The residual plots could then detect in what portion of the model nonlinearities are present and allow the researcher to more adequately model the data.

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.

Arminger, G. and Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and metropolis-hastings algorithm. *Pyschometrika*, *63(3)*, 271–300.

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

Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *12(4)*, 513–535.

Bauer, D. J., Baldasaro, R. E., Gottfredson, N. C. (2012). Diagnostic procedures for detecting nonlinear relationships between latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *19(2)*, 157-177.

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

Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables: A least squares estimator. *Sociological Methodology*, *25*, 223–251.

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

Bollen, K. A. and Paxton, P. (1998). Interactions of latent variables in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *5(3)*, 267–293.

Brandt, H., Kelava, A., and Klein, A. (2014). A Simulation Study Comparing Recent Approaches for the Estimation of Nonlinear Effects in SEM Under the Condition of Nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal*, *21(2)*, 181–195.

Cudeck, W. H., Harring, J. R., and du Toit, S. H. C. (2009). Marginal maximum likelihood estimation of a latent variable model with interactions. *Journal of Educational and Behavioral Statistics*, *34*, 131–144.

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.

Etezadi-Amoli, J. and McDonald, R. P. (1983). A second generation of nonlinear factor analysis. *Psychometrika*, *48(3)*, 315–342.

Gibson, W.A. (1959). Three multivariate models: Factor analysis, latent struture analysis, and latent profile analysis. *Psychometrika*, *24(3)*, 229–252.

Hildreth, L.A. and Lorenz, F.O. Residual plots to identify outliers in structural equation modeling, *2014 Joint Statistical Meetings Proceedings*, Boston, MA, August 2014.

Jöreskog, K. G., and Yang, F. (1996). Non-linear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides and R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (57–88). Mahway, NJ: Lawrence Erlbaum Associates.

Kenny, D. A. and Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96(1)*, 201–210.

Kelava, A. and Brandt, H. (2009). Estimation of nonlinear latent structural equation models using the extended unconstrained approach. *Review of Psychology*, *18(2)*, 123–131.

Kiel, L. D. and Elliott, E. W. (1996). *Chaos Theory in the Social Sciences: Foundations and Applications*. USA: University of Michigan Press.

Klein, A. and Moosbrugger, H. (2000). Maximum likelihood estimation of interaction effects with the LMS method. *Psychometrika*, *65(4)*, 457–474.

Klein, A. G. and Muthén, B. O. (2007), Quasi-maximum likelihood estimation of structural equation models with multiple interactions and quadratic effects. *Multivariate Behavioral Research*, *42(4)*, 647–673.

Klein, A. G., and Schermelleh-Engel, K. (2010). Introduction of a new measure for detecting poor fit due to omitted non-linear terms in SEM. *Advances in Statistical Analysis*, *94*, 157-166.

Lee, S. Y., Song, X. Y., and Tang, N. S. (2007). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *14(3)*, 404–434.

Lee, S. Y., and Zhu, H. T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika*, *67(2)*, 189–210.

McDonald, R. P. (1967). Numerical methods for polynomial models in nonlinear factor analysis. *Psychometrika*, *32(1)*, 77–112.

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

Mooijaart, A. and Satorra, A. (2009). On insensitivity of the chi-square model test to nonlinear misspecification in structural equation models. *Psychometrika*, *74(3)*, 443–455.

Nestler, S. (2015). A specification error test that uses instrumental variables to detect latent quadratic and latent interaction effects. *Structural Equation Modeling: A Multidisciplinary Journal*, *forthcoming*, published online 26 June 2015.

Pek, J., Chalmers, R. P., Kok, B. E., and Losardo, D. (2015). Visualizing Confidence Bands for Semiparametrically Estimated Nonlinear Relations Among Latent Variables. *Journal of Educational and Behavioral Statistics*, *40(4)*, 402-423.

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

Wall, M. M. and Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association*, *95(451)*, 929–940.

Wall, M. M., and Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology*, *56(1)*, 47–63.

Yerkes R. M. and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459-482.

Zhu, H. T., and Lee, S.'Y. (1999). Statistical analysis of nonlinear factor analysis models. *British Journal of Mathematical and Statistical Psychology*, *52(2)*, 225–242.

6. Appendix: R Code

```
#require needed packages
require(MASS)
require(lavaan)
#simulate data
set.seed(1202) #can
zeta1<-rnorm(250,0,0.5)
xi1<- rnorm(250,0,1)
eps<-t(mvrnorm(250,mu=c(0,0,0),Sigma=diag(3)))
delta<-t(mvrnorm(250,mu=c(0,0,0),Sigma=diag(3)))
Lambda<-matrix(c(1,1,1),3,1,byrow=TRUE)

##quadratic effect
#medium effect
eta1.qm<-0.5*xi1-0.35*xi1^2+zeta1
y.qm<-t(Lambda%*%eta1.qm+eps)
x.qm<-t(Lambda%*%xi1+delta)
```

```
#create data set
data.qm<-data.frame(y.qm,x.qm)
colnames(data.qm)<-c("y1","y2","y3","x1","x2","x3")
z.qm<-cbind(y.qm,x.qm)

#estimate the model
model.qm<-'
xi1=~x1+x2+x3
eta1=~y1+y2+y3
eta1~xi1'
qm.fit<-sem(model.qm,data=data.qm)
summary(qm.fit)
#make the plot
Lambdahat.qm<-matrix(c(1,0,1.073,0,0.982,0,0,1,0,1.111,0,1.144),byrow=TRUE,6,2)
Thetahat.qm<-matrix(c(.957,0,0,0,0,0,
                      0,.996,0,0,0,0,
                      0,0,.884,0,0,0,
                      0,0,0,1.223,0,0,
                      0,0,0,0,1.142,0,
            0,0,0,0,0,0.703),byrow=TRUE,6,6)
#Bartlett's method
Wb.qm<-solve(t(Lambdahat.qm)%*%solve(Thetahat.qm)%*%Lambdahat.qm)%*%t(Lambdahat.qm)%*%solve(Thetahat.qm)
nuhat.qm.b<-t(t(z.qm)-Lambdahat.qm%*%Wb.qm%*%t(z.qm))
colnames(nuhat.qm.b)<-c("e1hat.qm.b","e2hat.qm.b","e3hat.qm.b","d1hat.qm.b","d2hat.qm.b","d3hat.qm.b")
mhat.qm<-matrix(c(1,-0.376),1,2)
zetahat.qm.b<-t(mhat.qm%*%Wb.qm%*%t(z.qm))
names(zetahat.qm.b)<-"zetahat.qm.b"
zhat.qm.b<-t(Lambdahat.qm%*%Wb.qm%*%t(z.qm))
colnames(zhat.qm.b)<-c("y1hat.qm.b","y2hat.qm.b","y3hat.qm.b","x1hat.qm.b","x2hat.qm.b","x3hat.qm.b")
dhat.qm<-matrix(c(0,.376),1,2)
etahat.qm.b<-t(dhat.qm%*%Wb.qm%*%t(z.qm))
names(etahat.qm.b)<-"etahat.qm.b"
zetahat.etahat.qm.b<-cbind(zetahat.qm.b,etahat.qm.b)
Sigma.z.qm<-cov(z.qm)
Sigma.zetahat.qm.b<-mhat.qm%*%Wb.qm%*%Sigma.z.qm%*%t(Wb.qm)%*%t(mhat.qm)
Sigma.etahat.qm.b<-dhat.qm%*%Wb.qm%*%Sigma.z.qm%*%t(Wb.qm)%*%t(dhat.qm)
Sigma.etahatzetahat.qm.b<-mhat.qm%*%Wb.qm%*%Sigma.z.qm%*%t(Wb.qm)%*%t(dhat.qm)
Sigma.LV.qm.b<-matrix(c(Sigma.zetahat.qm.b,Sigma.etahatzetahat.qm.b,
Sigma.etahatzetahat.qm.b,Sigma.etahat.qm.b),2,2,byrow=TRUE)
LVRotated.qm.b<-t(chol2inv(t(chol(Sigma.LV.qm.b)))%*%t(zetahat.etahat.qm.b))
#regression method
Sigma.L.qm<-matrix(c(0.724,0,0,1.102),2,2,byrow=TRUE)
Wr.qm<-Sigma.L.qm%*%t(Lambdahat.qm)%*%solve(Sigma.z.qm)
nuhat.qm.r<-t(t(z.qm)-Lambdahat.qm%*%Wr.qm%*%t(z.qm))
colnames(nuhat.qm.r)<-c("e1hat.qm.r","e2hat.qm.r","e3hat.qm.r","d1hat.qm.r","d2hat.qm.r","d3hat.qm.r")
mhat.qm<-matrix(c(1,-0.376),1,2)
zetahat.qm.r<-t(mhat.qm%*%Wr.qm%*%t(z.qm))
names(zetahat.qm.r)<-"zetahat.qm.r"
zhat.qm.r<-t(Lambdahat.qm%*%Wr.qm%*%t(z.qm))
colnames(zhat.qm.r)<-c("y1hat.qm.r","y2hat.qm.r","y3hat.qm.r","x1hat.qm.r","x2hat.qm.r","x3hat.qm.r")
dhat.qm<-matrix(c(0,.376),1,2)
etahat.qm.r<-t(dhat.qm%*%Wr.qm%*%t(z.qm))
names(etahat.qm.r)<-"etahat.qm.r"
zetahat.etahat.qm.r<-cbind(zetahat.qm.r,etahat.qm.r)
Sigma.zetahat.qm.r<-mhat.qm%*%Wr.qm%*%Sigma.z.qm%*%t(Wr.qm)%*%t(mhat.qm)
Sigma.etahat.qm.r<-dhat.qm%*%Wr.qm%*%Sigma.z.qm%*%t(Wr.qm)%*%t(dhat.qm)
Sigma.etahatzetahat.qm.r<-mhat.qm%*%Wr.qm%*%Sigma.z.qm%*%t(Wr.qm)%*%t(dhat.qm)
Sigma.LV.qm.r<-matrix(c(Sigma.zetahat.qm.r,Sigma.etahatzetahat.qm.r,
Sigma.etahatzetahat.qm.r,Sigma.etahat.qm.r),2,2,byrow=TRUE)
LVRotated.qm.r<-t(chol2inv(t(chol(Sigma.LV.qm.r)))%*%t(zetahat.etahat.qm.r))
#Anderson-Rubin method
Asq.qm<-t(Lambdahat.qm)%*%solve(Thetahat.qm)%*%Sigma.z.qm%*%solve(Thetahat.qm)%*%Lambdahat.qm
#find the inverse square root of Asq using spectral decomposition
mat.sqrt.inv<-function(A)
{
    ei<-eigen(A)
    d<-ei$values
    d<-(d+abs(d))/2
    d2<-1 / sqrt(d)
    d2[d == 0]<-0
    ans<-ei$vectors %*% diag(d2) %*% t(ei$vectors)
    return(ans)
}
A.sqinv.qm<-mat.sqrt.inv(Asq.qm)
War.qm<-A.sqinv.qm%*%t(Lambdahat.qm)%*%solve(Thetahat.qm)
nuhat.qm.ar<-t(t(z.qm)-Lambdahat.qm%*%War.qm%*%t(z.qm))
colnames(nuhat.qm.ar)<-c("e1hat.qm.ar","e2hat.qm.ar","e3hat.qm.ar","d1hat.qm.ar","d2hat.qm.ar","d3hat.qm.ar")
mhat.qm<-matrix(c(1,-0.376),1,2)
zetahat.qm.ar<-t(mhat.qm%*%War.qm%*%t(z.qm))
names(zetahat.qm.ar)<-"zetahat.qm.ar"
zhat.qm.ar<-t(Lambdahat.qm%*%War.qm%*%t(z.qm))
colnames(zhat.qm.ar)<-c("y1hat.qm.ar","y2hat.qm.ar","y3hat.qm.ar","x1hat.qm.ar","x2hat.qm.ar","x3hat.qm.ar")
dhat.qm<-matrix(c(0,.376),1,2)
etahat.qm.ar<-t(dhat.qm%*%War.qm%*%t(z.qm))
names(etahat.qm.ar)<-"etahat.qm.ar"
zetahat.etahat.qm.ar<-cbind(zetahat.qm.ar,etahat.qm.ar)
Sigma.zetahat.qm.ar<-mhat.qm%*%War.qm%*%Sigma.z.qm%*%t(War.qm)%*%t(mhat.qm)
Sigma.etahat.qm.ar<-dhat.qm%*%War.qm%*%Sigma.z.qm%*%t(War.qm)%*%t(dhat.qm)
Sigma.etahatzetahat.qm.ar<-mhat.qm%*%War.qm%*%Sigma.z.qm%*%t(War.qm)%*%t(dhat.qm)
Sigma.LV.qm.ar<-matrix(c(Sigma.zetahat.qm.ar,Sigma.etahatzetahat.qm.ar,
Sigma.etahatzetahat.qm.ar,Sigma.etahat.qm.ar),2,2,byrow=TRUE)
LVRotated.qm.ar<-t(chol2inv(t(chol(Sigma.LV.qm.ar)))%*%t(zetahat.etahat.qm.ar))

par(mfrow=c(1,3))
plot(LVRotated.qm.r[,2],LVRotated.qm.r[,1],main="Regression Method",pch=16)

##add loess line
```

```
lw1.qm.r <- loess(LVRotated.qm.r[,1] ~ LVRotated.qm.r[,2])
Rotated.qm.r<-data.frame(LVRotated.qm.r)
names(Rotated.qm.r)<-c("zetahat","etahat")
order.qm.r <- order(Rotated.qm.r$etahat)
lines(Rotated.qm.r$etahat[order.qm.r],lw1.qm.r$fitted[order.qm.r],col="red",lwd=3)

plot(LVRotated.qm.b[,2],LVRotated.qm.b[,1],main="Bartlett Method",pch=16)

lw1.qm.b <- loess(LVRotated.qm.b[,1] ~ LVRotated.qm.b[,2])
Rotated.qm.b<-data.frame(LVRotated.qm.b)
names(Rotated.qm.b)<-c("zetahat","etahat")
order.qm.b <- order(Rotated.qm.b$etahat)
lines(Rotated.qm.b$etahat[order.qm.b],lw1.qm.b$fitted[order.qm.b],col="red",lwd=3)

plot(LVRotated.qm.ar[,2],LVRotated.qm.ar[,1],main="Anderson-Rubin Method",pch=16)

lw1.qm.ar <- loess(LVRotated.qm.ar[,1] ~ LVRotated.qm.ar[,2])
Rotated.qm.ar<-data.frame(LVRotated.qm.ar)
names(Rotated.qm.ar)<-c("zetahat","etahat")
order.qm.ar <- order(Rotated.qm.ar$etahat)
lines(Rotated.qm.ar$etahat[order.qm.ar],lw1.qm.ar$fitted[order.qm.ar],col="red",lwd=3)
```