Structural Equation Modeling, 19:157–177, 2012

ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705511.2012.659612

Psychology Press
Taylor & Francis Group

Diagnostic Procedures for Detecting Nonlinear Relationships Between Latent Variables

Daniel J. Bauer, Ruth E. Baldasaro, and Nisha C. Gottfredson
Department of Psychology, University of North Carolina at Chapel Hill

Structural equation models are commonly used to estimate relationships between latent variables. Almost universally, the fitted models specify that these relationships are linear in form. This assumption is rarely checked empirically, largely for lack of appropriate diagnostic techniques. This article presents and evaluates two procedures that can be used to visualize and detect nonlinear relationships between latent variables. The first procedure involves fitting a linear structural equation model and then inspecting plots of factor score estimates for evidence of nonlinearity. The second procedure is to use a mixture of linear structural equation models to approximate the underlying, potentially nonlinear function. Targeted simulations indicate that the first procedure is more efficient, but that the second procedure is less biased. The mixture modeling approach is recommended, particularly with medium to large samples.

Keywords: structural equation models, nonlinear, diagnostics, linear, assumptions

Structural equation models are commonly used in the social, behavioral, and health sciences to evaluate the relationships among latent variables that cannot be observed directly or without measurement error. One traditional assumption of these models is that the latent variables (or factors) are linearly related to one another. This assumption is rarely overtly scrutinized, either empirically or theoretically. Indeed, many of the textbooks used to teach structural equation modeling (e.g., Bollen, 1989; Kaplan, 2000) provide little discussion of the linearity assumption, or of procedures for diagnosing potentially nonlinear relationships.[1] More attention is given

[1] Kline (2005) is an exception, suggesting that scatter plots between observed variables be inspected for potential nonlinear trends. The contribution of measurement error to the observed scores might, however, diminish the appearance of nonlinear trends. Diagnostic procedures that can be applied directly at the level of the latent variables are hence preferable.

Correspondence should be addressed to Daniel J. Bauer, Department of Psychology, CB#3270, University of North Carolina, Chapel Hill, NC 27599, USA. E-mail: dbauer@email.unc.edu

to modeling nonlinear trends using quadratic or product-interaction terms (a topic of intensive research; e.g., Kenny & Judd, 1984; Klein & Moosbrugger, 2000; Klein & Muthén, 2007, Lee, Song, & Tang, 2007; Marsh, Wen, & Hau, 2004; Schumacker & Marcoulides, 1998), but the use of these models presumes prior knowledge of the nature of the relationship. Further, tests of model fit are unlikely to prompt reexamination of the linearity assumption, as these tests are insensitive to the presence of nonlinear effects (Mooijaart & Satorra, 2009).

The notable absence of coverage of linearity diagnostics in structural equation modeling texts contrasts with the prominence of this topic in textbooks on regression models for observed variables (e.g., Cohen, Cohen, West, & Aiken, 2003; Neter, Kutner, Nachtsheim, & Wasserman, 1996). For instance, Cohen et al. (2003) noted that

> Unless there is strong theory that hypothesizes a particular form of nonlinear relationship, most researchers begin by specifying linear regression models.... However, there is no guarantee that the form of the relationship will in fact be linear. Consequently, it is important to examine graphical displays to determine if a linear relationship adequately characterizes the data. (p. 125)

In particular, the nonparametric locally weighted regression, or loess, procedure can be useful for visualizing and modeling nonlinear effects when the functional form of the relationship is unknown (Cleveland, 1993; Cleveland, Devlin, & Grosse, 1988).

There is little reason to believe that nonlinear effects are less common for latent than observed variables. If anything, nonlinear effects might be harder to visualize and detect with observed variables due to contamination by measurement error. It would therefore seem that the absence of discussion of techniques for detecting nonlinear effects in latent variable models has more to do with the fact that few techniques have been proposed or evaluated for this purpose. It is our goal here to present and compare two such techniques and to make recommendations for their use in practice.

The first technique we consider is to generate estimated scores for the latent variables, enabling the application of diagnostic procedures used with observed-variable regression models. This idea is hardly novel. A little over four decades ago, McDonald (1967) suggested visually inspecting plots of factor score estimates to judge whether indicators are linearly related to factors. A similar procedure could be used to judge whether factors are linearly related to one another. One would follow these steps:

1. Fit a linear structural equation model to the data.
2. Generate factor score estimates from the fitted model.
3. Plot the factor score estimates for each factor against the others.
4. Superimpose a scatterplot smoother (e.g., loess regression line) on the plots to aid in visually discriminating nonlinear trends.

We refer to this approach henceforth as FS-SM (factor scores with smoothing).

Although straightforward in implementation, there is reason to believe that FS-SM might produce misleading evidence in favor of adopting the linear model. Let us suppose that there is, in fact, a nonlinear relationship between two latent variables. The model fit in Step 1 incorrectly assumes the relationship to be linear, and it is this model that is used to generate factor score estimates in Step 2. The true nonlinear trend might then be attenuated in the factor score estimates due to the initial misspecification of the model as linear. Indeed, for at least some methods of factor score estimation, this is necessarily true.

Consider factor score estimates generated by the regression method (Thomson, 1936, 1951; Thurstone, 1935). As noted by Bartholomew and Knott (1999), the regression method yields scores that are equivalent to empirical Bayes's estimates. Like all empirical Bayes's estimates, the estimated scores are shrunken toward the empirical prior distribution; that is, the distribution of the latent variables in the fitted model. Because the fitted model assumes a linear relationship between the latent variables, the factor score estimates will be shrunken toward the linear prior (and away from the true nonlinear relationship). The extent to which the scores are shrunken toward the prior depends on the level of factor determination (i.e., the informativeness of the observed data). As such, regression method estimates can be expected to reveal nonlinear trends with greater fidelity as the number and communality of the measured variables for the factors increases. With less informative measured variables, however, nonlinear relationships can be occluded. Many other methods of factor score estimation have also been proposed, but these tend to generate scores that are highly correlated with regression method estimates (Fava & Velicer, 1992), and hence are likely to perform similarly to the regression method.

Ideally, a diagnostic procedure for detecting nonlinear effects should not be biased against finding them. We thus also consider a newer, model-based procedure for semiparametrically estimating nonlinear latent variable relationships of unknown form (Bauer, 2005; Pek, Sterba, Kok, & Bauer, 2009). This approach involves fitting a finite mixture of K linear structural equation models to the data. The relationship between the latent variables is specified as linear within each of the K mixing components (sometimes referred to as latent classes). Although the relationship is locally linear within each component of the mixture, aggregating across components provides a smoothed estimate of the underlying nonlinear trend. For instance, suppose that we were to estimate a two-component mixture and found that in the first class the latent variables have low means and a positive linear relationship, whereas in the second class the latent variables have higher means and no linear relationship. The smoothed function would then travel from the positively sloped relationship in Class 1 at low values to the null relationship in Class 2 at high values to produce an asymptotic curve. No prior knowledge of the shape of the function is required. Accordingly, we refer to this approach as semiparametric modeling (SPM).

Simulations suggest that SPM can accurately recover nonlinear relationships between latent variables. Thus, one potential advantage of SPM relative to FS-SM is that SPM does not produce a confirmation bias for the linear model. One potential disadvantage of SPM is that structural equation mixture models require the estimation of more parameters than conventional structural equation models, potentially resulting in higher sampling variability for SPM than FS-SM.

We compare the performance of these two possible diagnostic strategies using simulation methodology. Our specific hypotheses are as follows:

1. Factor score estimates will generally show greater bias than mixture estimates, although the performance of factor score estimates will improve as the number of manifest variables and their communalities increase.
2. Mixture estimates will often show greater sampling variability than factor score estimates.

This potential trade-off between bias and efficiency will be evaluated with an eye toward making recommendations for which procedure is most likely to be useful in practice. To test our hypotheses, we conducted two targeted simulation studies in which we manipulated the form and magnitude of the true nonlinear relationship, the number of manifest variables per factor, the communalities of the manifest variables, and the sample size. Study 1 evaluated the performance of the two procedures for quadratic functions, for which the magnitude of nonlinearity can be easily manipulated and quantified. To illustrate the general applicability of these methods, Study 2 focused on the recovery of two alternative, asymmetric nonlinear functions.

STUDY 1

As noted earlier, Study 1 examined the performance of FS-SM and SPM in recovering quadratic latent variable regression functions. Quadratic functions were chosen because they are perhaps the most common functions used to model nonlinearity in observed-variable regression models and because they have properties that are convenient for simulating data (e.g., analytically determined variance expressions, and quantifiable curvature parameters). The FS-SM and SPM approaches were applied to the simulated data without incorporation of any knowledge regarding the nature of the relationship between the factors. The goal was to recover the shape of the true, underlying function.

Simulation Design and Data Generation

Each model contained one latent predictor and one latent outcome variable, with 50% of the total variance in the latent outcome explained by the latent predictor via a quadratic relationship. We implemented a factorial design crossing the magnitude of the quadratic effect (medium or large), sample size ($N = 250$, 500, or 1,000), number of observed indicator variables per factor (three or six), and communalities of the indicators ($h^2 = .25$, .50, or .75). For each of the 36 conditions 250 data sets were generated in SAS.

Data generation proceeded in two steps. First, data were generated from the latent variable model

$$\eta_{1i} = \zeta_{1i}$$

$$\eta_{2i} = 4.85 + .5\eta_{1i} - .35\eta_{1i}^2 + \zeta_{2i}$$

or

$$\eta_{2i} = 5 - .5\eta_{1i}^2 + \zeta_{2i}$$

for the medium and large quadratic effect conditions, respectively, where $i = 1, 2, \ldots, N$. The disturbances ζ_{1i} and ζ_{2i} were sampled from a bivariate normal distribution with zero means and covariance matrix

$$\mathbf{\Psi} = \begin{pmatrix} 1 & 0 \\ 0 & .5 \end{pmatrix}$$

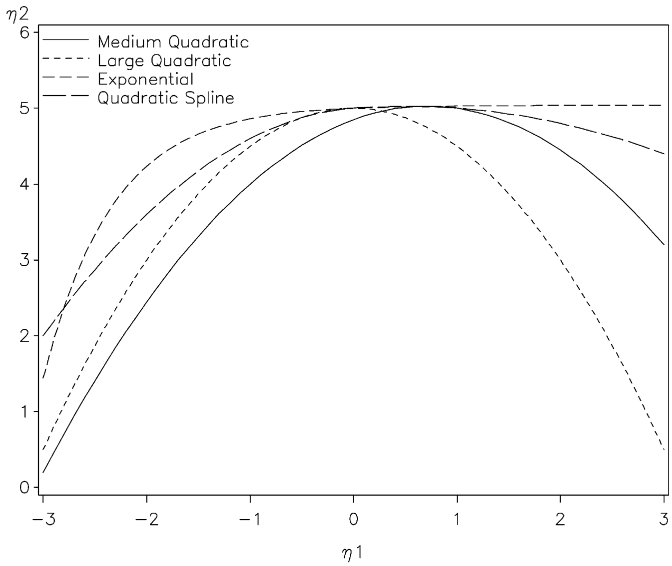

FIGURE 1 Target functions used in the simulation study.

In both the medium and large quadratic effect conditions, these values imply that η_1 and η_2 have marginal variances of 1 and that 50% of the variance in η_2 is explained by η_1. In the large quadratic condition, all of the explained variance is due to the quadratic effect, whereas in the medium quadratic effect condition only half of the explained variance is uniquely attributable to the quadratic effect. The two functions can be compared visually in Figure 1.

Second, for each factor, data were generated for three or six indicator variables via the following measurement model:

$$\mathbf{y}_{1i} = \mathbf{1}\eta_{1i} + \boldsymbol{\varepsilon}_{1i}$$

$$\mathbf{y}_{2i} = \mathbf{1}\eta_{2i} + \boldsymbol{\varepsilon}_{2i}$$

These equations imply that the intercepts for all of the indicators are zero and that each indicator loads on one and only one factor with a factor loading of one. The residuals (contained in $\boldsymbol{\varepsilon}_{1i}$ and $\boldsymbol{\varepsilon}_{2i}$) were normally distributed with zero means and a diagonal covariance matrix. The residual variance was held constant across indicators at 3, 1, or .333 to produce the desired communalities of $h^2 = .25$, .50, or .75, respectively.

Implementation of Diagnostic Procedures

For each data set, both the FS-SM and SPM approaches were used to evaluate the linearity of the relationship between η_1 and η_2. All models were fit in M*plus* 5.1. Maximum likelihood estimation for finite mixture models is prone to local solutions, hence solutions were obtained from 500 random starts for each of the first 50 replications per condition. The solution with the highest likelihood for each replication was selected, and then parameter estimates were averaged over replications (with care taken to avoid label switching across replications). The average values were then used as start values for the remaining 200 replications per condition.

FS-SM. Factor score estimates were computed by first fitting a standard structural equation model to the data in which the relationship between η_1 and η_2 was assumed to be linear. Each indicator variable was permitted to load only on the appropriate factor and the model was identified by setting the intercept and factor loading of the first indicator variable for each factor to zero and one, respectively, while estimating the remaining intercepts and factor loadings. The resulting parameter estimates (and observed data) were then used to produce factor score estimates via several different methods: the regression method (Thomson, 1936, 1951; Thurstone, 1935), Bartlett's generalized least squares method (Bartlett, 1937), and the constrained scores method of Anderson and Rubin (1956), as extended to correlated factor models by McDonald (1981; hereafter referred to as constrained covariance), with the idea of determining which set of estimates performs best when the goal is to detect nonlinear trends. The formulas used to generate the factor score estimates are provided in Appendix A, and were implemented within the SAS IML procedure (SAS Institute, 2008a). For each of the three sets of factor scores, we then ran a loess regression (Cleveland, 1993; Cleveland et al., 1988), as implemented within the LOESS procedure of SAS (SAS Institute, 2008b, pp. 3187–3254), to obtain a smoothed nonparametric estimate of the underlying regression function. The smoothing parameter was chosen by the default method, to minimize the corrected Akaike's Information Criterion (AIC; Hurvich, Simonoff, & Tsai, 1998).

SPM. The SPM method has been described in detail in Pek et al. (2009), with technical developments in Bauer (2005). To implement this method, structural equation mixture models with K components were fit to the data, where the within-component structural equation model specification was the same as described earlier for the FS-SM method. The smoothed nonlinear aggregate regression function was then generated from the model estimates obtained for each replication using formulas given by Bauer (2005), as shown in Appendix B. Pek et al. (2009) described online utilities that can be used to perform these computations for a given replication and generate plots of the estimated regression function.

For each replication, K was determined by fitting a sequence of models with increasing latent classes until the minimum AIC (Akaike, 1973) or Bayes's Information Criterion (BIC; Schwarz, 1978) was detected. Both AIC and BIC are commonly used for model selection with finite mixtures (McLachlan & Peel, 2000), with AIC generally favoring more classes. Of the two criteria, BIC is generally preferred in direct applications of finite mixtures (where the goal is to recover the number of true, discrete underlying subpopulations) because it is a consistent selector of the number of latent classes. In this instance, however, the application is indirect (there are no discrete subgroups to be found—the mixture is used as an approximation device) and hence BIC might not necessarily outperform AIC. Aggregate regression function estimates were thus obtained using both minimum AIC and BIC, with the goal of determining if one criterion outperforms the other for this type of application.

Evaluation of Diagnostic Procedures

The FS-SM and SPM diagnostic procedures were evaluated by comparing the accuracy and precision of estimation of the underlying regression function. Because both procedures are approximations, neither produces estimates of the parameters of the true function, hence performance cannot be judged by evaluating recovery of the function parameters. Instead, performance was judged with respect to recovery of the regression function as a whole, as measured by the bias, standard deviation, and root mean squared error (RMSE) of the function estimates. Bias measures accuracy, the standard deviation measures (in)efficiency, and the RMSE incorporates information about both. Computation of these criteria for the regression function estimates is detailed in Appendix C.

In addition to these numerical criteria, for some conditions we also provide a visual comparison of the performance of the FS-SM and SPM approaches. The average estimated regression function is overlaid on the true regression function (depicting bias) and the interval enclosing 90% of the estimates at any given point is also shown (depicting sampling variability). Note that this interval is not a confidence interval, but a sampling interval.

Results

Bias, standard deviation, and RMSE results for the two quadratic functions, with medium and large quadratic effects, are presented in Tables 1 and 2, respectively. For ease of interpretation, the lowest bias, standard deviation, and RMSE entries are bolded within Tables 1 and 2 for each condition. Overall, the pattern of findings was quite consistent across the two functions. First, one uniform trend was for bias, standard deviation, and RMSE to decrease as the communality of the indicators increased, and as the number of indicators per factor increased; that is, as factor determinacy increased. This trend was anticipated for FS-SM, but had not been anticipated for SPM. Second, as hypothesized, SPM was generally less biased than FS-SM, with the lowest bias in every condition achieved using SPM with the minimum AIC solution. Third, also as hypothesized, FS-SM was generally more efficient than SPM, with the lowest standard deviation obtained in nearly every condition using Bartlett's factor scores. Last, the method with the best RMSE varied across conditions depending primarily on sample size. At $N = 250$, the FS-SM approach resulted in the lowest RMSE in nearly all conditions, with the exception that, for the medium quadratic effect, the SPM resulted in lower RMSE when $h^2 = .50$. At $N = 500$, SPM produced a lower RMSE for both quadratic functions in the $h^2 = .50$ conditions, and also produced a lower RMSE when $h^2 = .25$ and there were six indicators per factor. Finally, at $N =$ 1,000, SPM produced a lower RMSE in every condition for both functions. This changeover in rank could be attributed to two sample size effects: More classes could be supported when using the SPM approach at higher sample sizes, reducing bias, and at the same time higher sample sizes dampened the relative contribution of sampling variability to RMSE.

In comparing within each diagnostic approach, two additional findings are worth noting. First, although Bartlett's method almost always resulted in the lowest standard deviation, no one factor score estimator consistently resulted in the lowest bias or RMSE. For the most part, this ambiguity is due to the quite similar performance of the different factor score estimators. Similarly, although SPM with the minimum AIC solution always resulted in the lowest bias, the minimum BIC solution was more efficient and quite often produced a lower RMSE. Given the

TABLE 1
Results for Quadratic Target Function: Medium Quadratic Effect

h^2	*Method*	*3 Indicators*			*6 Indicators*		
		Bias	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*
$N = 250$							
.25	FS-SM R	.293	.301	.420	.224	.226	**.318**
	FS-SM B	.350	**.201**	.404	.270	**.191**	.331
	FS-SM C	.270	.251	**.369**	.242	.208	.319
	SPM AIC	**.157**	.427	.455	**.053**	.376	.380
	SPM BIC	.306	.330	.450	.135	.337	.363
.50	FS-SM R	.177	.172	.246	.118	.158	.197
	FS-SM B	.213	**.150**	.260	.139	**.147**	.202
	FS-SM C	.222	.170	.280	.194	.168	.256
	SPM AIC	**.041**	.241	.245	**.029**	.227	.229
	SPM BIC	.077	.227	**.240**	.070	.181	**.194**
.75	FS-SM R	.087	.130	**.156**	.060	.128	.142
	FS-SM B	.093	**.127**	.157	.059	**.128**	**.141**
	FS-SM C	.183	.147	.235	.183	.146	.234
	SPM AIC	**.015**	.179	.180	**.020**	.152	.153
	SPM BIC	.065	.147	.161	.067	.143	.158
$N = 500$							
.25	FS-SM R	.290	.207	.356	.221	.176	.283
	FS-SM B	.375	**.136**	.399	.293	**.142**	.326
	FS-SM C	.281	.170	**.329**	.258	.155	.301
	SPM AIC	**.074**	.363	.371	**.043**	.342	.344
	SPM BIC	.190	.315	.368	.086	.220	**.236**
.50	FS-SM R	.171	.133	.216	.117	.119	.167
	FS-SM B	.221	**.111**	.248	.138	**.114**	.179
	FS-SM C	.229	.126	.262	.204	.124	.239
	SPM AIC	**.034**	.242	.245	**.019**	.164	.165
	SPM BIC	.072	.163	**.178**	.061	.141	**.154**
.75	FS-SM R	.082	.102	.131	.060	**.093**	.111
	FS-SM B	.095	**.099**	.138	.053	.093	**.107**
	FS-SM C	.191	.113	.222	.188	.105	.216
	SPM AIC	**.014**	.128	**.129**	**.015**	.107	.108
	SPM BIC	.057	.118	.131	.041	.112	.119
$N = 1{,}000$							
.25	FS-SM R	.307	.161	.346	.221	.135	.259
	FS-SM B	.385	**.103**	.399	.284	**.106**	.304
	FS-SM C	.287	.129	.314	.252	.120	.279
	SPM AIC	**.059**	.306	.312	**.038**	.242	.245
	SPM BIC	.091	.258	**.273**	.074	.165	**.181**
.50	FS-SM R	.168	.096	.194	.100	.092	.136
	FS-SM B	.231	**.086**	.246	.135	**.086**	.160
	FS-SM C	.236	.093	.254	.203	.092	.223
	SPM AIC	**.022**	.158	.160	**.010**	.120	.121
	SPM BIC	.064	.130	**.145**	.029	.117	**.120**
.75	FS-SM R	.077	.078	.109	.049	.074	.089
	FS-SM B	.095	**.074**	.121	.050	**.072**	.087
	FS-SM C	.196	.086	.213	.192	.092	.213
	SPM AIC	**.012**	.094	**.095**	**.010**	.080	**.081**
	SPM BIC	.028	.097	.101	.028	.081	.086

Note. h^2 = communality; *SD* = standard deviation; RMSE = root mean squared error; FS-SM R = factor scores with smoother, regression method; B = Bartlett's method; C = constrained covariance method; SPM AIC = semiparametric method, minimum Akaike's Information Criterion solution; BIC = minimum Bayes's Information Criterion solution. The entry with the lowest bias, SD, or RMSE within a given condition is shown in bold.

TABLE 2
Results for Quadratic Target Function: Large Quadratic Effect

		3 Indicators			6 Indicators		
h^2	*Method*	*Bias*	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*
$N = 250$							
.25	FS-SM R	.352	.288	.455	.293	.246	.383
	FS-SM B	.359	**.201**	**.411**	.291	**.208**	**.358**
	FS-SM C	.360	.226	.425	.292	.214	.362
	SPM AIC	**.231**	.466	.520	**.054**	.542	.544
	SPM BIC	.387	.347	.520	.090	.401	.411
.50	FS-SM R	.210	.208	.295	.143	.190	.238
	FS-SM B	.245	**.169**	.298	.145	**.175**	**.227**
	FS-SM C	.232	.176	**.292**	.147	.178	.230
	SPM AIC	**.048**	.320	.324	**.032**	.262	.264
	SPM BIC	.092	.288	.303	.066	.230	.239
.75	FS-SM R	.111	.149	.186	.076	.142	.161
	FS-SM B	.117	**.141**	**.183**	.070	**.138**	**.154**
	FS-SM C	.122	.142	.187	.093	.158	.184
	SPM AIC	**.028**	.203	.205	**.028**	.167	.170
	SPM BIC	.065	.192	.202	.062	.174	.184
$N = 500$							
.25	FS-SM R	.351	.210	.409	.276	.179	.329
	FS-SM B	.377	**.136**	.401	.312	**.147**	.345
	FS-SM C	.357	.155	**.389**	.284	.151	.322
	SPM AIC	**.116**	.559	.571	**.033**	.372	.373
	SPM BIC	.206	.528	.567	.065	.289	**.297**
.50	FS-SM R	.213	.149	.260	.138	.145	.200
	FS-SM B	.253	**.124**	.281	.160	**.131**	.207
	FS-SM C	.224	.126	.257	.145	.133	.196
	SPM AIC	**.028**	.231	.233	**.019**	.194	.196
	SPM BIC	.048	.214	**.219**	.044	.186	**.191**
.75	FS-SM R	.099	.116	**.152**	.060	.113	.128
	FS-SM B	.118	**.109**	.161	.063	**.110**	**.127**
	FS-SM C	.106	.121	.161	.066	.128	.144
	SPM AIC	**.022**	.153	.154	**.023**	.129	.131
	SPM BIC	.046	.149	.156	.047	.128	.137
$N = 1{,}000$							
.25	FS-SM R	.354	.155	.387	.267	.144	.304
	FS-SM B	.400	**.105**	.413	.326	**.114**	.345
	FS-SM C	.366	.114	.384	.280	.124	.306
	SPM AIC	**.038**	.394	.396	**.026**	.293	.294
	SPM BIC	.074	.332	**.340**	.050	.231	**.236**
.50	FS-SM R	.208	.114	.237	.128	.110	.169
	FS-SM B	.270	**.093**	.286	.168	**.099**	.195
	FS-SM C	.225	.095	.244	.137	.103	.172
	SPM AIC	**.029**	.177	.180	**.020**	.141	**.142**
	SPM BIC	.049	.163	**.170**	.045	.136	.143
.75	FS-SM R	.097	.088	.131	.058	.087	.104
	FS-SM B	.121	**.083**	.147	.066	**.084**	.107
	FS-SM C	.103	.090	.137	.061	.100	.117
	SPM AIC	**.021**	.109	**.111**	**.022**	.094	**.010**
	SPM BIC	.046	.110	.119	.041	.102	.110

Note. h^2 = communality; SD = standard deviation; RMSE = root mean squared error; FS-SM R = factor scores with smoother, regression method; B = Bartlett's method; C = constrained covariance method; SPM AIC = semiparametric method, minimum Akaike's Information Criterion solution; BIC = minimum Bayes's Information Criterion solution. The entry with the lowest bias, SD, or RMSE within a given condition is shown in bold.

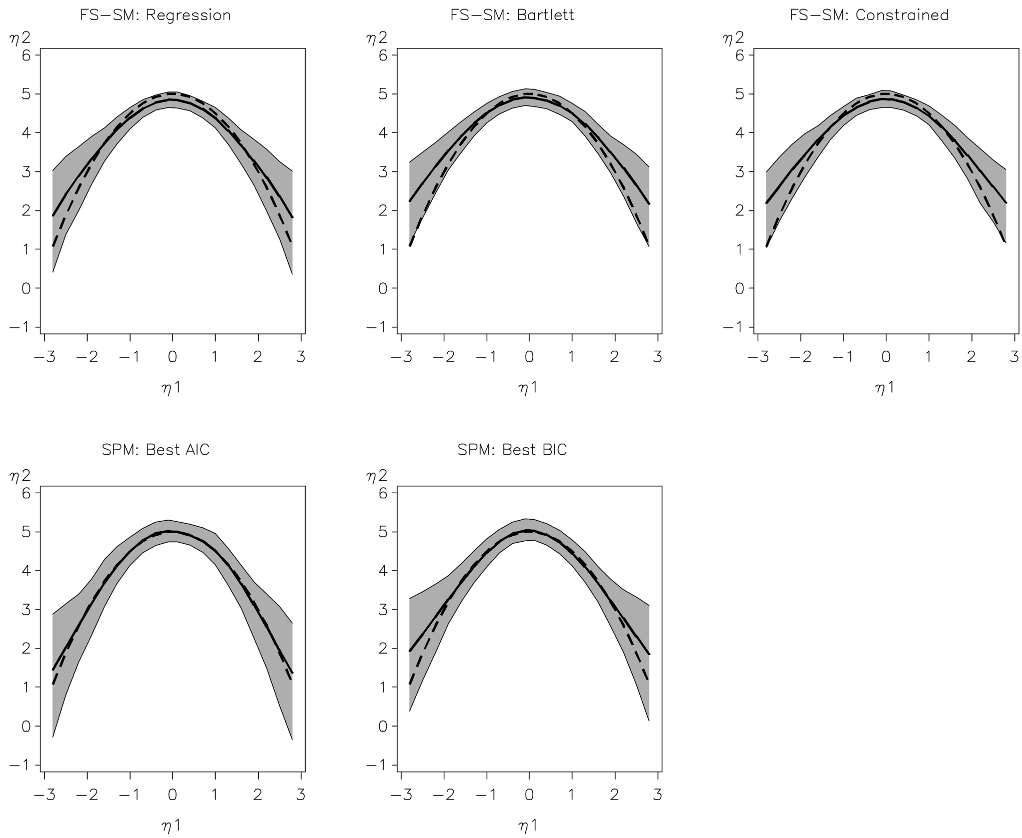

FIGURE 2 Target function is quadratic with large quadratic trend, $N = 250$, six indicators, and communality of .50; the shaded region is the 90% sampling interval, the bold solid line is the mean estimate, and the bold dashed line is the true function. FS-SM = factor scores with smoothing; SPM = semiparametric modeling using a finite mixture of linear structural equations; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion.

more conservative nature of the BIC criterion, this is not surprising. With fewer classes selected by BIC than AIC, fewer parameters are estimated, resulting in less flexibility to capture the underlying function (more bias) but also less sampling variability (lower standard deviation).

Figures 2 and 3 communicate these findings visually. Figure 2 depicts the results from the large quadratic function with $h^2 = .50$, six indicators per factor, and $N = 250$, whereas Figure 3 is for the same conditions at $N = 1{,}000$. Note that the mean estimated regression function (bold line) follows the underlying function (dashed line) quite closely when using SPM with minimum AIC, the method with the least bias. SPM with minimum BIC is the second least biased method, because the mean estimated regression function follows the true function quite closely between -2 and 2, where 95% of the mass of the predictor distribution lies. The FS-SM approach is most biased because the mean estimated regression function is more linear than it should be, shrinking both the tails and the peak of the quadratic function.

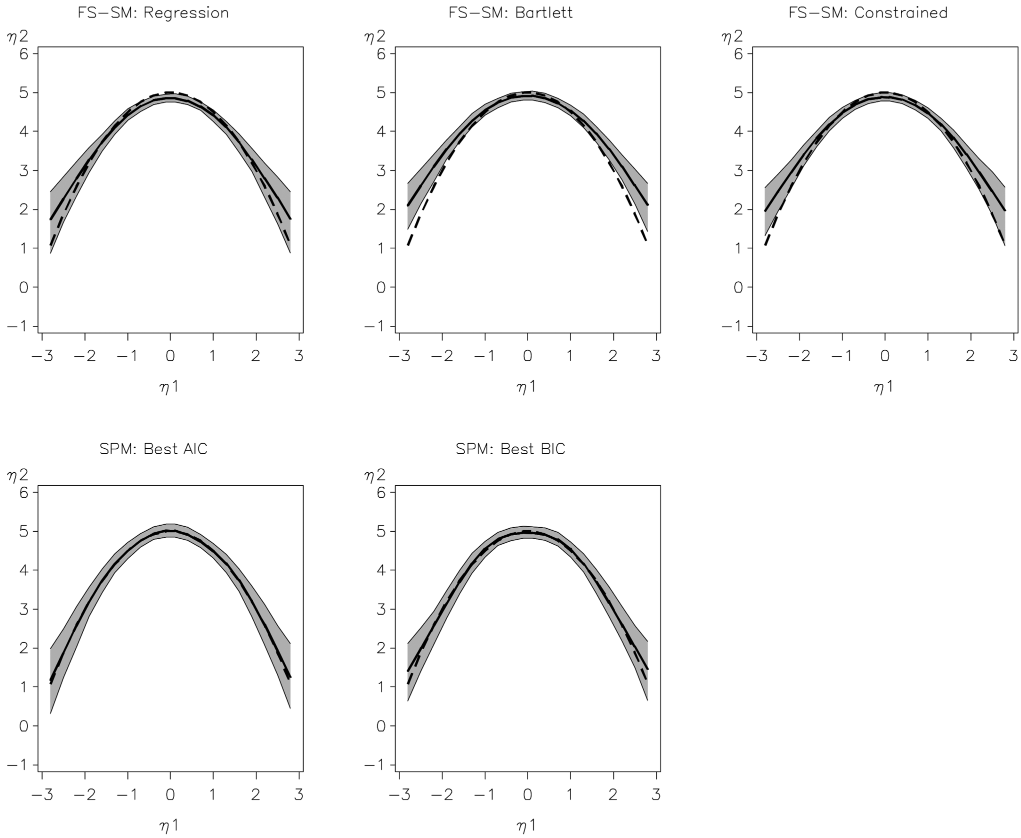

FIGURE 3 Target function is quadratic with large quadratic trend, $N = 1{,}000$, six indicators, and communality of .50; the shaded region is the 90% sampling interval, the bold solid line is the mean estimate, and the bold dashed line is the true function. FS-SM = factor scores with smoothing; SPM = semiparametric modeling using a finite mixture of linear structural equations; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion.

Also note that the intervals for the estimates are narrower for the FS-SM methods, indicating their greater efficiency. For $N = 250$, depicted in Figure 2, this greater efficiency was sufficient to make FS-SM with Barlett's method the lowest RMSE approach, whereas for $N = 1{,}000$, depicted in Figure 3, SPM with the minimum AIC solution produced the lowest RMSE.

Overall, the bias–efficiency trade-off is clearly evident in these results. The SPM method using the minimum AIC solution produces the least bias but is also least efficient. The FS-SM approach provides greater efficiency, but is the most biased. Finally, the SPM method using the minimum BIC solution occupies a middle ground, more biased than AIC but less so than FS-SM, and less efficient than FS-SM but more efficient than SPM with AIC. With small sample sizes, the FS-SM approach tends to result in lower RMSE, whereas with large samples the SPM approach always produces the lowest RMSE.

STUDY 2

To probe the generality of the results obtained in Study 1 beyond solely quadratic functions, Study 2 was conducted using simulated data from two asymmetric nonlinear functions, a negatively accelerated exponential function and a quadratic spline. Because Study 1 indicated that sample size was the single largest determinant of the relative ranking of FS-SM and SPM approaches, sample size was again varied in Study 2, but the number of observed indicator variables and their communalities were held constant at six indicators per factor with $h^2 = .5$.

Simulation Design and Data Generation

Data were generated as described for Study 1, but with the exception that the function relating η_2 to η_1 was either specified as a negatively accelerated exponential function

$$\eta_{2i} = 5 + .04(1 - e^{-1.5\eta_{1i}}) + \zeta_{2i}$$

or as a quadratic spline producing an asymmetric concave-down curve:

$$\eta_{2i}(\eta_{1i} < 0) = 5 + .1\eta_{1i} - .3\eta_{1i}^2 + \zeta_{2i}$$

$$\eta_{2i}(\eta_{1i} > 0) = 5 + .1\eta_{1i} - .1\eta_{1i}^2 + \zeta_{2i}$$

For visual comparison, these two functions are depicted alongside the quadratic functions in Figure 1. As in Study 1, 50% of the variance in the latent outcome was explained by the latent predictor. Again, 250 samples were generated for each condition ($N = 250$, 500, or 1,000). FS-SM and SPM procedures were applied to each sample as described in Study 1, and similarly evaluated on the basis of bias, standard deviation, and RMSE, in addition to visual depictions of performance.

Results

Tables 3 and 4 present the results for the negatively accelerated exponential function and quadratic spline function, respectively. Consistent with Study 1, for both functions, SPM with the minimum AIC solution consistently yielded the lowest bias. For the quadratic spline function, the FS-SM approach also again consistently produced the lowest standard deviation, although in this case the constrained covariance factor scores outperformed Bartlett's scores at $N = 500$ and $N = 1{,}000$. In contrast, for the negatively accelerated exponential, the lowest standard deviation was obtained at all sample sizes using SPM with the minimum BIC solution. For both functions, SPM tended to produce the lowest RMSE. For the quadratic spline, the lowest RMSE was obtained at all sample sizes using SPM with the minimum BIC solution. For the negative exponential, the lowest RMSE was obtained using SPM with the minimum AIC solution, except at $N = 250$, in which case FS-SM with Bartlett's scores produced a lower RMSE.

The performance of the different methods for the negative exponential and quadratic spline functions can also be assessed visually in Figures 4 and 5, respectively, plotted at $N = 500$. As can be seen, for both functions SPM with minimum AIC is least biased, but this

TABLE 3
Results for the Negatively Accelerated Exponential Function

	Performance Criteria		
Method	*Bias*	*SD*	*RMSE*
$N = 250$			
FS-SM R	.054	.150	.159
FS-SM B	.050	**.110**	**.120**
FS-SM C	.070	.148	.164
SPM AIC	**.034**	.124	.129
SPM BIC	.068	.107	.127
$N = 500$			
FS-SM R	.058	.110	.125
FS-SM B	.063	.081	.102
FS-SM C	.087	.132	.158
SPM AIC	**.043**	.086	**.096**
SPM BIC	.071	**.072**	.101
$N = 1,000$			
FS-SM R	.062	.085	.105
FS-SM B	.068	.072	.099
FS-SM C	.091	.067	.113
SPM AIC	**.043**	.062	**.075**
SPM BIC	.067	**.055**	.086

Note. SD = standard deviation; RMSE = root mean squared error; FS-SM R = factor scores with smoother, regression method; B = Bartlett's method; C = constrained covariance method; SPM AIC = semiparametric method, minimum Akaike's Information Criterion solution; BIC = minimum Bayes's Information Criterion solution. The entry with the lowest bias, SD, or RMSE within a given condition is shown in bold.

TABLE 4
Results for the Quadratic Spline Function

	Performance Criteria		
Method	*Bias*	*SD*	*RMSE*
$N = 250$			
FS-SM R	.072	.091	.116
FS-SM B	.081	**.084**	.117
FS-SM C	.124	.087	.151
SPM AIC	**.017**	.135	.137
SPM BIC	.041	.103	**.111**
$N = 500$			
FS-SM R	.065	.068	.094
FS-SM B	.079	.063	.101
FS-SM C	.129	**.063**	.144
SPM AIC	**.012**	.098	.098
SPM BIC	.038	.077	**.086**
$N = 1,000$			
FS-SM R	.065	.053	.084
FS-SM B	.081	.050	.096
FS-SM C	.134	**.048**	.142
SPM AIC	**.010**	.072	.073
SPM BIC	.028	.065	**.070**

Note. SD = standard deviation; RMSE = root mean squared error; FS-SM R = factor scores with smoother, regression method; B = Bartlett's method, C = constrained covariance method; SPM AIC = semiparametric method, minimum Akaike's Information Criterion solution; BIC = minimum Bayes' Information Criterion solution. The entry with the lowest bias, SD, or RMSE within a given condition is shown in bold.

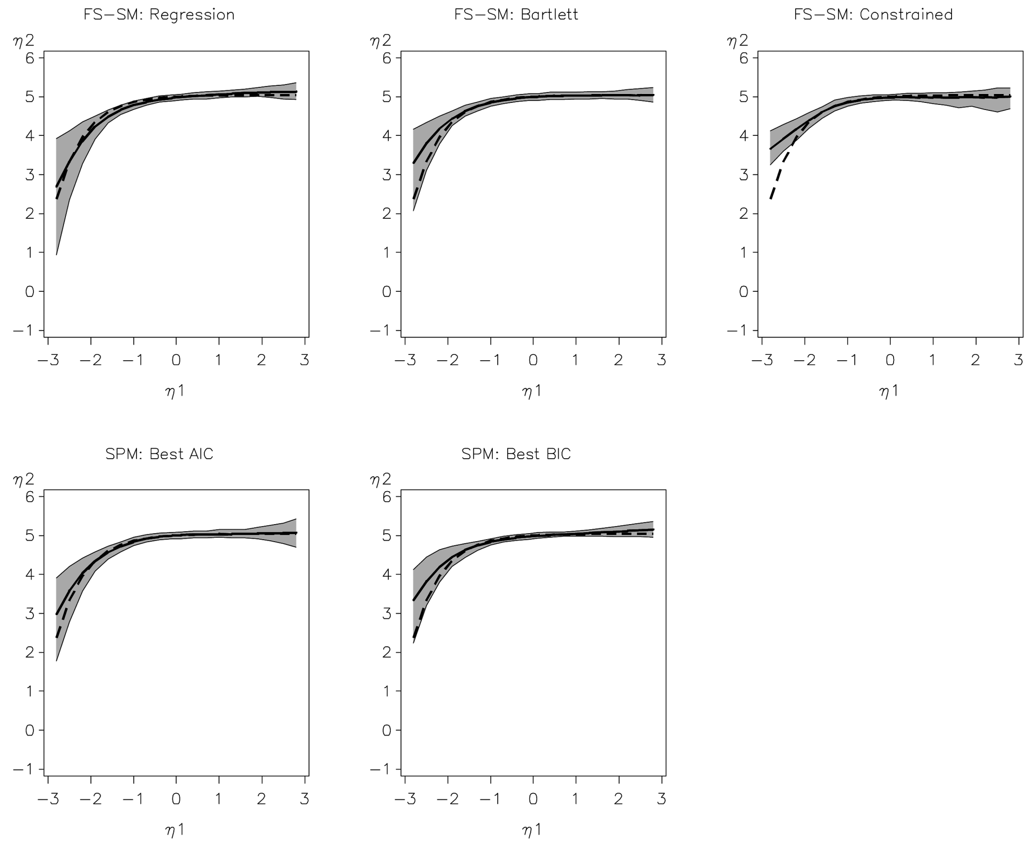

FIGURE 4 Target function is negatively accelerated exponential function, $N = 500$; the shaded region is the 90% sampling interval, the bold solid line is the mean estimate, and the bold dashed line is the true function. FS-SM = factor scores with smoothing; SPM = semiparametric modeling using a finite mixture of linear structural equations; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion.

method produces wider sample estimate intervals than SPM with minimum BIC. For the negative exponential, the FS-SM approach did not yield appreciably greater efficiency than SPM, although this was the case for the quadratic spline function, as reflected in the relative widths of the sample estimate intervals. The FS-SM approach generally had greater difficulty approximating the quadratic spline function.

As in Study 1, there was no single best method of generating factor score estimates, in terms of bias, standard deviation, or RMSE metrics of regression function recovery. Also consistent with the earlier findings, among the two SPM solutions, the minimum AIC resulted in lower bias but also lower efficiency than the minimum BIC. In some conditions, lower bias trumped higher efficiency, resulting in a lower RMSE for the minimum AIC solution, and in other conditions higher efficiency trumped lower bias, resulting in a lower RMSE for the minimum BIC solution.

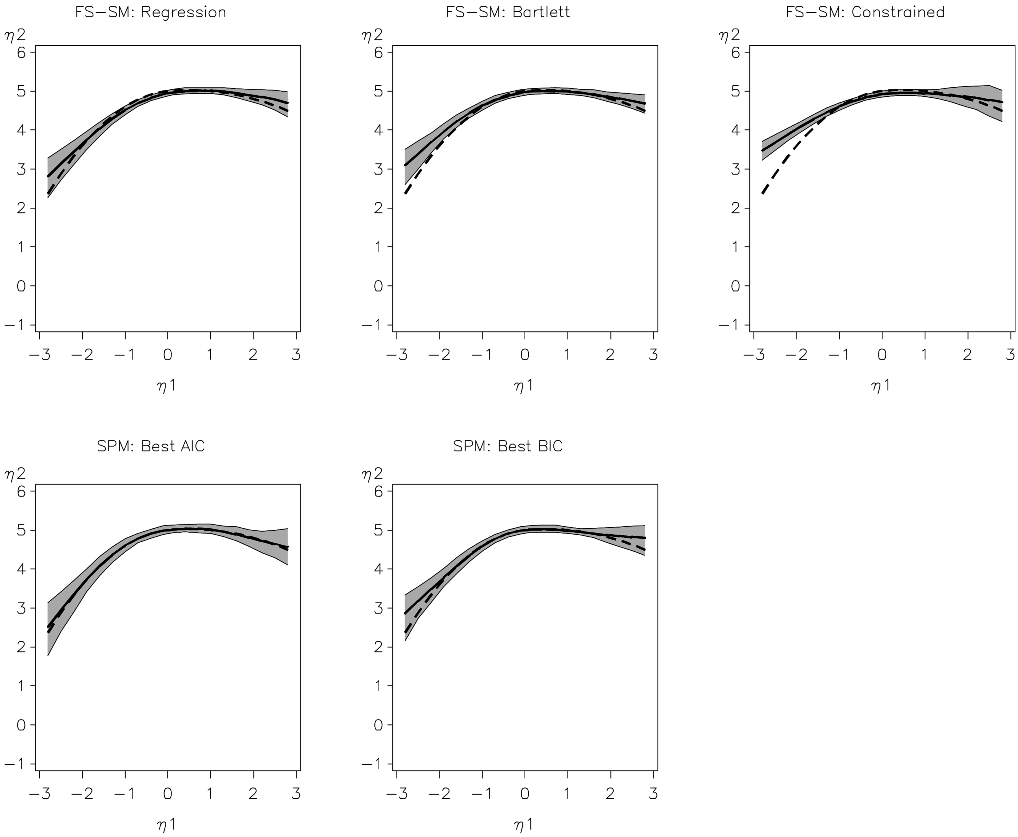

FIGURE 5 Target function is quadratic spline function, $N = 500$; the shaded region is the 90% sampling interval, the bold solid line is the mean estimate, and the bold dashed line is the true function. FS-SM = factor scores with smoothing; SPM = semiparametric modeling using a finite mixture of linear structural equations; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion.

CONCLUSIONS

Summing over our results, we can offer the following conclusions. First, as hypothesized, the FS-SM approach tends to be more biased than the SPM approach, due to shrinkage of nonlinear relationships toward a straight line in factor score estimation. As expected, plots of factor score estimates were more likely to accurately reflect nonlinear trends when factor score determinacy was high, that is, with high communality factor indicators, and more factor indicators.

Interestingly, the SPM approach also seems to benefit from higher factor determinacy. In retrospect, the latter finding is perhaps not surprising. The estimation of latent classes in the SPM approach relies on nonnormality of the multivariate marginal distribution of the observed indicators (Bauer & Curran, 2003, 2004). The multivariate nonnormality that exists among the latent factors due to their nonlinear relationship is transmitted to the marginal observed distribution more strongly as the communality of the indicators increases, providing more information from which to estimate the latent classes and thereby improve the SPM approximation procedure. Additional indicators would similarly increase the information in the multivariate observed distributions from which to estimate classes. More simply put, increasing the quality and amount of data is always a good thing.

Second, also as hypothesized, the efficiency of the FS-SM approach often exceeded the efficiency of the SPM approach. That is, the SPM approach generally exhibited greater sampling variability (a higher standard deviation). Only for the negatively accelerated exponential function did the SPM approach perform more efficiently than the FS-SM approach. This bias–efficiency trade-off resulted in FS-SM often producing lower RMSE at low sample sizes (i.e., $N = 250$), and SPM always producing lower RMSE at high sample sizes (i.e., $N = 1{,}000$).

Third, although no explicit hypotheses were posited about performance differences across functions, some interesting trends were observed. In particular, the degree of bias for FS-SM seemed to decrease as the monotonicity of the underlying function increased. That is, the ordering of the functions from least to most biased, as estimated by FS-SM, is negative exponential, quadratic spline, medium quadratic, and large quadratic. FS-SM thus appears to produce less bias as the target function becomes more linear, a predictable result of the fact that FS-SM regression function estimates are shrunken toward a straight-line relationship. But this trend also implies that the FS-SM approach would likely fail to detect subtler nonlinear effects. This bias trend was not observed for the SPM approach. For SPM a similar trend was, however, observed for *sampling variability:* The standard deviation of SPM estimates generally decreased as the monotonicity of the target function increased, likely because fewer classes are needed to approximate such functions.

RECOMMENDATIONS

Investigators seeking to diagnose nonlinear trends among latent variables would do well to employ either of the methods investigated here. As is clear from Figures 2 through 5, both approaches are capable of identifying nonlinear trends, albeit with different degrees of accuracy and precision. In choosing between the two approaches, therefore, other considerations are relevant.

The ease of implementation of the FS-SM approach is its chief advantage. Structural equation modeling programs generally offer the option to output factor score estimates, although the available types of factor score estimates (e.g., regression method or Bartlett's scores) might be limited. Given these estimates, it is straightforward to use any standard statistical analysis program to produce a scatterplot with a smoothed regression function estimate. In contrast, the SPM approach is somewhat more difficult to implement, requiring the estimation of multiple models (i.e., mixture structural equation models with $K = 1, 2, 3$, etc., classes) and considerable postprocessing of model results. Fortunately, Pek et al. (2009) recently provided computing resources, including an online utility and an R package, which read in the results of a fitted mixture structural equation model and automatically generate the smoothed regression function estimate.[2]

[2] Available at http://www.unc.edu/psychology/dbauer/plotSEMM.htm

The primary advantage of the SPM approach is that it is less biased than the FS-SM approach. The minimum BIC solution will produce an estimate of the regression function that is unbiased over much of the range of the data. The minimum AIC solution improves reproduction of the tails of the function where data are sparse, but at the cost of increased sampling variability. The downside is that both SPM estimates are less efficient than FS-SM. In examining Figures 2 through 5, these differences in efficiency are noticeable in the width of the estimate intervals, but not so large as to dissuade use of the SPM approach or the minimum AIC solution in particular. These efficiency differences are less consequential at large sample sizes, at which the SPM approach consistently outperforms FS-SM in terms of RMSE. Thus, particularly for large samples, SPM is the better approach.

An additional advantage of the SPM approach is that it is entirely model-based, so that it can also be used inferentially, whereas FS-SM is entirely descriptive. Pek, Losardo, and Bauer (2011) recently evaluated two methods for generating nonsimultaneous confidence bands around the estimated regression function obtained from the SPM approach. Both delta-method and parametric bootstrap confidence intervals performed quite well over most of the range of the estimated function (with coverage rates slipping somewhat in the tails). Thus, whereas both the FS-SM and SPM approaches provide estimates of underlying regression function, only the SPM approach can also provide information about the uncertainty of the estimate, permitting inferences about the population from which the sample was drawn.

In sum, therefore, we recommend use of the SPM procedure for detecting and visualizing unknown nonlinear functions, particularly when fitting structural equation models to large samples. The FS-SM approach is not, however, entirely without merit, and the use of either approach would represent an advance over current practice. With the availability of these procedures, diagnostic testing of the linearity assumption routinely made in structural equation modeling should become a matter of course.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (Award SES-0716555 to Daniel J. Bauer).

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petroc & B. F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Academiai Kiado.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability* (pp. 111–150). Berkeley, CA: University of California Press.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). London, UK: Arnold.

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, 4,* 513–535.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for over-extraction of latent trajectory classes. *Psychological Methods, 8,* 338–363.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9,* 3–29.
Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: Wiley.
Cleveland, W. S. (1993). *Visualizing data.* Summit, NJ: Hobart Press.
Cleveland, W. S., Devlin, S. J., & Grosse, E. (1988). Regression by local fitting. *Journal of Econometrics, 37,* 87–114.
Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analyses for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Fava, J. L., & Velicer, W. F. (1992). An empirical comparison of factor, image, component, and scale scores. *Multivariate Behavioral Research, 27,* 301–322.
Hurvich, C. M., Simonoff, J. S., & Tsai, C. L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike Information Criterion. *Journal of the Royal Statistical Society B, 60,* 271–293.
Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions.* Thousand Oaks, CA: Sage.
Kendall, M. G., & Stuart, A. (1969). *The advanced theory of statistics.* London, UK: Griffin.
Kenny, D., & Judd, C. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96,* 201–210.
Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65,* 457–474.
Klein, A., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research, 42,* 647–673.
Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford.
Lee, S., Song, X., & Tang, N. (2007). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling, 14,* 404–434.
Marsh, H., Wen, Z., & Hau, K. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9,* 275–300.
McDonald, R. P. (1967). *Nonlinear factor analysis* (Psychometric Monograph No. 15). Richmond, VA: William Byrd Press.
McDonald, R. P. (1981). Constrained least squares estimators of oblique common factors. *Psychometrika, 46,* 337–341.
McLachlan, G., & Peel, D. (2000). *Finite mixture models.* New York, NY: Wiley.
Mooijaart, A., & Satorra, A. (2009). On insensitivity of the chi-square model test to nonlinear misspecifications in structural equation models. *Psychometrika, 74,* 443–455.
Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). Boston, MA: McGraw-Hill.
Pek, J., Losardo, D., & Bauer, D. J. (2011). Confidence intervals for a semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling: A Multidisciplinary Journal, 18,* 537–553.
Pek, J., Sterba, S. K., Kok, B. E., & Bauer, D. J. (2009). Estimating and visualizing nonlinear relations among latent variables: A semiparametric approach. *Multivariate Behavioral Research, 44,* 407–436.
SAS Institute. (2008a). *SAS/IML 9.2 user's guide.* Cary, NC: Author.
SAS Institute. (2008b). *SAS/STAT 9.2 user's guide.* Cary, NC: Author.
Schumacker, R. E., & Marcoulides, G. A. (1998). *Interaction and nonlinear effects in structural equation modeling.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.
Thomson, G. H. (1936). Some points of mathematical technique in the factorial analysis of ability. *Journal of Educational Psychology, 27,* 37–54.
Thomson, G. H. (1951). *The factorial analysis of human ability* (5th ed.). Boston, MA: Houghton Mifflin.
Thurstone, L. L. (1935). *Vectors of the mind.* Chicago, IL: University of Chicago Press.

APPENDIX A
FACTOR SCORE ESTIMATES

A general expression for the measurement and latent variable models of the structural equation model is

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \qquad \boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i$$

where i indexes individual and $\mathbf{y}$ is a $p \times 1$ vector of measured variables; $\boldsymbol{\nu}$ is a $p \times 1$ vector of intercepts for the measured variables; $\boldsymbol{\Lambda}$ is a $p \times q$ matrix of factor loadings; $\boldsymbol{\eta}$ is a $q \times 1$ vector of latent variables; $\boldsymbol{\varepsilon}$ is a $p \times 1$ vector of residuals for the measured variables; $\boldsymbol{\alpha}$ is a $q \times 1$ vector of intercepts for the latent variables; $\mathbf{B}$ is a $q \times q$ matrix of regression coefficients for the latent variables; and $\boldsymbol{\zeta}$ is a $q \times 1$ vector of residuals for the latent variables.

Further, define $\boldsymbol{\Theta}$ to be a $p \times p$ covariance matrix for the residuals of the measured variables and $\boldsymbol{\Psi}$ to be a $q \times q$ covariance matrix for the residuals of the latent variables.

To compute the factor scores, we shall additionally make use of the following moment matrices:

$\boldsymbol{\mu}_{\boldsymbol{\eta}} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha}$ is the $q \times 1$ model-implied mean vector of $\boldsymbol{\eta}$
$\boldsymbol{\Sigma}_{\boldsymbol{\eta\eta}} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}[(\mathbf{I} - \mathbf{B})^{-1}]'$ is the $q \times q$ model-implied covariance matrix of $\boldsymbol{\eta}$
$\boldsymbol{\mu}_{\mathbf{y}}$ is the $p \times 1$ mean vector of $\mathbf{y}$
$\boldsymbol{\Sigma}_{\mathbf{yy}}$ is the $p \times p$ covariance matrix of $\mathbf{y}$

Finally, define $\mathbf{C}$ to be the Cholesky root of $\boldsymbol{\Sigma}_{\boldsymbol{\eta\eta}}$ such that $\boldsymbol{\Sigma}_{\boldsymbol{\eta\eta}} = \mathbf{CC}'$.

We can now write the factor score formulas for the three methods used in the simulation. Note that, in practice, matrices are populated with sample-specific estimates to compute the scores.

The regression method (Thomson, 1936, 1951; Thurstone, 1935):

$$\mathbf{f}_{Ri} = \boldsymbol{\Sigma}_{\boldsymbol{\eta\eta}}\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_{\mathbf{yy}}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{\mathbf{y}}) + \boldsymbol{\mu}_{\boldsymbol{\eta}}$$

Bartlett's method (Bartlett, 1937):

$$\mathbf{f}_{Bi} = (\boldsymbol{\Lambda}'\boldsymbol{\Theta}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Theta}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{\mathbf{y}}) + \boldsymbol{\mu}_{\boldsymbol{\eta}}$$

The constrained covariance method (McDonald, 1981):

$$\mathbf{f}_{Ci} = \mathbf{C}'[(\mathbf{C}\boldsymbol{\Lambda}'\boldsymbol{\Theta}^{-1}\boldsymbol{\Sigma}_{\mathbf{yy}}\boldsymbol{\Theta}^{-1\prime}\boldsymbol{\Lambda}\mathbf{C}')^{-\frac{1}{2}}]'\mathbf{C}\boldsymbol{\Lambda}'\boldsymbol{\Theta}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_{\mathbf{y}}) + \boldsymbol{\mu}_{\boldsymbol{\eta}}$$

where $-\frac{1}{2}$ indicates the inverse of the Cholesky root of the matrix.

APPENDIX B
SEMIPARAMETRIC MODELING

To produce regression estimates via the SPM approach, a finite mixture of K linear structural equation models is fit, and then a smoothed estimate of the regression estimate is obtained by averaging over the mixing components. For the simulation, the latent variable model for component k was specified as

$$\eta_{1i} = \alpha_{1k} + \zeta_{1i}$$

$$\eta_{2i} = \alpha_{2k} + \beta_{12k}\eta_{1i} + \zeta_{2i}$$

To aid in model estimation, we assumed that the covariance matrix of ζ_{1i} and ζ_{2i} was equal across mixing components or

$$\boldsymbol{\Psi}_k = \begin{pmatrix} \psi_{11} & 0 \\ 0 & \psi_{22} \end{pmatrix}$$

We designate the mixing probabilities for the components as $\pi_1, \pi_2, \ldots, \pi_K$, where $\sum_{k=1}^{K} \pi_k = 1$.

The aggregate regression function for this model is implied to be

$$E(\eta_2|\eta_1) = \sum_{k=1}^{K} (\pi_k|\eta_1)(\alpha_{2k} + \beta_{12k})$$

where $\pi_k|\eta_1$ are conditional mixing weights that smooth the regression function and are computed from Bayes's theorem as

$$(\pi_k|\eta_1) = \frac{\pi_k \phi(\eta_1; \alpha_{1k}, \psi_{11})}{\sum_{k=1}^{K} \pi_k \phi(\eta_1; \alpha_{1k}, \psi_{11})}$$

APPENDIX C
PERFORMANCE CRITERIA

Using results from Kendall and Stuart (1969, p. 51, Equation 2.37), the mean squared error (MSE) for the regression function for replication r can be defined as

$$MSE_r = E([\hat{g}_r(\eta_1) - g(\eta_1)]^2) = \int [\hat{g}_r(\eta_1) - g(\eta_1)]^2 f(\eta_1) d\eta_1$$

where $\hat{g}_r(\eta_1)$ denotes the estimated regression function returning the predicted value of η_2 given η_1, generated either from the FS-SM or SPM method, as detailed in Appendices A and B. Likewise, $g(\eta_1)$ is the predicted value of η_2 returned by the true regression function. Finally, $f(\eta_1)$ is the probability density function (PDF) of η_1. Integrating the squared error over the probability density of η_1 implicitly weights errors by the density of observations in the region in which they occur. That is, small discrepancies in areas dense with data might be more impactful than larger discrepancies in sparse regions.

Because, in this case, this integral cannot be solved analytically, it was approximated numerically using a Monte Carlo procedure. First, 10,000 values for η_1 were generated from $f(\eta_1)$, the distribution used to draw values of η_1 for the simulation study; that is, the standard normal distribution. These 10,000 values of η_1 were then used to calculate predicted values of η_2 from the true regression function $g(\eta_1)$ and from the estimated function $\hat{g}_r(\eta_1)$. Averaging the squared difference between the true and estimated values over the simulated distribution of

η_1 produces an approximate MSE for each replication.

$$MSE_r \approx \sum_{m=1}^{10,000} [\hat{g}_r(\eta_{1m}) - g(\eta_{1m})]^2/10,000$$

Averaging the squared error within a given condition and taking the expected value over the function produces an estimate of the overall MSE:

$$MSE \approx \sum_{m=1}^{10,000} \left\{ \sum_{r=1}^{250} \frac{[\hat{g}_r(\eta_{1m}) - g(\eta_{1m})]^2}{250} \right\} /10,000$$

The overall MSE was also decomposed into components reflecting bias and sampling variance, as described by Kendall and Stuart (1969, p. 21):

$$MSE = B^2 + V$$

where B^2 represents the squared bias component and V represents the variance of the estimates.

Squared bias was approximated by Monte Carlo as

$$B^2 \approx \sum_{m=1}^{10,000} [\overline{g}(\eta_{1m}) - g(\eta_{1m})]^2/10,000$$

where $E(\hat{g}_r(\eta_1))$ is replaced by the across-replications within-condition average $\overline{g}(\eta_{1m})$, defined as

$$\overline{g}(\eta_{1m}) = \sum_{r=1}^{250} \hat{g}_r(\eta_{1m})/250$$

Similarly, the variance can be approximated as

$$V \approx \sum_{m=1}^{10,000} \left\{ \sum_{r=1}^{250} \frac{[\hat{g}_r(\eta_{1m}) - \overline{g}(\eta_{1m})]^2}{250} \right\} /10,000$$

Taking the root of each of these values puts the results back into the scale of the dependent variable. Thus, for interpretation, the RMSE, bias, and standard deviation are reported as opposed to MSE, B^2, and V.