SEM Estimation in the Context of Small Samples: Comparison of Latent Variable Models, Single Indicator, Regularized 2-Stage Least Squares and Observed Variable Models

W. Holmes Finch
Ball State University

Structural Equation Models (SEMs) are widely used in the social sciences. These models explicitly model statistical error, thereby yielding more accurate and efficient parameter estimates than do observed variable methods such as regression. However, SEM requires larger samples in order to work properly. Therefore, researchers working with small samples are faced with the choice of appropriately modeling measurement error with too few cases, or inappropriately ignoring it with a method that better accommodates smaller samples. To address this problem, alternatives have been developed, including the single indicator approach using reliability to set error variances for composite indicator variables, and regularized 2-stage least squares. This study compares these methods with one another, factor score regression, SEM, and observed variable models. Results revealed that the performance of the methods depended upon the underlying model being fit, as well as sample size and population reliability. Implications for practice are provided in the discussion.

Structural equation modeling (SEM) is widely used in the social and behavioral sciences in order to investigate relationships among latent constructs. One of the strengths of SEM is that it accounts for measurement error in a way that observed variable models, such as regression, do not (Kline, 2016). Despite its advantages however, SEM also has limitations that render it less useful in some circumstances. One of the more important of these limitations is the need for larger sample sizes than those required for alternative approaches such as observed variable regression (Hoyle, 2012). When samples are small, parameter estimation accuracy may be compromised (Ledgerwood & Shrout, 2011; MacCallum, Browne, & Cai, 2006), and standard errors will be inflated, in turn impacting hypothesis tests that are used to make determinations regarding relationships among variables (latent and/or observed). In such small sample situations it may be preferable for researchers to use observed variable methods (Ledgerwood & Shrout). The goal of the current study was to extend on earlier work in this area (Savalei, 2019; Miller, Finch, & Ballenger, 2019; Jung, 2013), by comparing several approaches to fitting SEMs with small samples, including single indicator methods, observed variable models, and regularized 2-stage least squares.

Structural Equation Models

The SEM relating latent variables to one another can be expressed as:

$$\eta = \mathrm{B}\eta + \Gamma\xi + \zeta \tag{1}$$

where: η = Vector of latent endogenous variables
B = Vector of coefficients linking the endogenous to one another
ξ = Vector of latent exogenous variables
Γ = Vector coefficients linking the latent exogenous variables to the latent endogenous variables
ζ = Covariance matrix of the structural error terms.

Latent variables in equation (1) consist of measurement models linking observed indicators to the latent construct, as below.

$$\eta = v + \lambda x + \varepsilon \tag{2}$$

where: x =Matrix of observed indicator variables
v =Vector of intercepts
λ =Factor loading matrix
ε =Covariance matrix of measurement errors.

The parameters in equations (1) and (2) can be estimated using a variety of approaches, including maximum likelihood (MLE), Bayesian modeling with the Markov Chain Monte Carlo (MCMC) approach, or 2-stage least squares (2SLS).

Structural Equation Modeling with Small Samples and Misspecified Models

As is evident from a perusal of equation (1) and (2), there can be a large number of parameters to estimate when fitting a SEM. Therefore, researchers using it are encouraged to obtain a relatively large sample, such as a bare minimum of 100 to 200 (Kline, 2016), or a minimum of 5-10 observations per parameter to be estimated (Bentler & Chou, 1987). Other authors have called for the conduct of formal a priori sample size calculations prior to gathering data and fitting a model (Brown, 2013; Muthèn & Muthèn, 2002). Regardless of these rules of thumb, or the results of a power analysis, in some situations small samples cannot be avoided. Researchers working with small and/or difficult to reach populations, such as adults with autism, people diagnosed with schizophrenia, or undocumented migrants may not be able to obtain samples that would meet either the ad hoc rules of thumb or more formal power calculations. And yet, in such situations there may still be a need to examine relationships between latent traits, such as cognitive functioning, personality, or mood.

Given the fact that small samples are sometimes unavoidable, there is a need for methods that allow for accurate estimation of the SEMs (Fok, Henry, & Allen, 2015). One approach that is commonly used in such cases involves the calculation of composite scores based on the indicator variables (e.g., scale items) that appear in equation (2). Such composites can then be used in standard regression or path analysis (PA) models. The primary advantage of using such a composite score approach is that the number of parameters that need to be estimated is much smaller than is the case in a standard SEM. For example, consider a simple scenario in which a researcher would like to examine the relationship between mathematics aptitude and executive functioning. She can measure math aptitude using a standardized test, and executive functioning using a scale such as the Kaufman Assessment Batter for Children (KABC; Kaufman & Kaufman, 2018). Each item on the standardized test is an indicator of the latent construct of math aptitude. Likewise, the items on the KABC are indicators of executive functioning. When fitting a full SEM, the measurement portions of each latent variable would involve estimating the parameters in equation (2), whereas the structural relationships between math aptitude and executive functioning would be estimated in equation (1). In contrast, researchers using an observed variable approach to addressing this research question could fit a regression model in which a composite of the items for each scale (e.g., sum, mean) represent the constructs, leaving only the intercept and regression (structural) coefficients to be estimated. This reduced number of parameters means that the sample size necessary to estimate the composite variables model is smaller than that needed for a full SEM involving latent variables.

Despite the apparent advantage of having fewer parameters to estimate, the composite variable approach carries with it some serious disadvantages in terms of estimation accuracy. Specifically, the fact that this method does not account for the error variance when estimating model parameters leads to biased parameter estimates (Savalei, 2019; Westfall & Yarkoni, 2016; Cole & Preacher, 2014). In addition, the use of observed variable models rather than latent variable approaches that appropriately model error has been shown to yield inflated Type I error rates for structural coefficients (Savalei; Westfall & Yarkoni). At the same time, as discussed above, standard latent variable SEM does not generally perform well in the context of small samples. Therefore, some alternative methodologies for fitting models when latent structure is present need to be employed by researchers.

Single Indicator Estimation

One approach that has been mentioned in the literature as a way to incorporate measurement error while at the same time using models based on composite variables involves the estimation of reliability for the latent trait, followed by the conversion of this value to an error estimate (e.g., Cole & Preacher, 2014). The composites then serve as single indicators for corresponding latent variables, and the error estimate is fixed for each in order to identify the model. The composite can be the sum or mean score, as described above in the context of regression and PA. The reliability value selected can be obtained using any one of the various estimates available, such as Cronbach's Alpha, McDonald's Omega, the greatest lower bound (GLB), split-half (SH), or some constant value (e.g., 0.7 or 0.8) that is based upon prior work with the scale in question. This single indicator (SI) technique can be conceptualized as lying between a purely observed variable approach for modeling relationships among constructs, such as regression or PA, and a fully latent SEM (Savalei, 2019).

The use of a single indicator for each latent variable carries with it the advantages associated with PA and small samples, in particular the fewer number of parameters that need to be estimated than is the case for a full SEM. At the same time, by incorporating the scale reliability (and thereby measurement error) into the model, the SI approach acknowledges the imperfection of the measurement in a way that PA does not. The result, theoretically, is less bias in model parameter estimates (Savalei, 2019). And indeed, this is precisely what Savalei found for estimation of the indirect effect in a mediation model, using a Monte Carlo simulation design. More precisely, in her study, Savalei discovered that using the SI approach with a fixed reliability value yielded more accurate (as measured by root mean square error) results than did SEM or SI with an estimate of Alpha, and also outperformed SI when Alpha was used for sample sizes of fewer than 200. In her manuscript describing this study, Savalei called for future research examining the performance of the SI method with other reliability estimates, which is one of the main purposes of the current research. The need to investigate the use of alternative reliability estimates is particularly trenchant, given the known problems associated with Alpha more generally (Sijtsma, 2009). Thus, one major purpose of the current simulation study is to include additional measures of reliability beyond Cronbach's Alpha in the conduct of SI modeling.

Regularized Two-Stage Least Squares Estimation

An alternative estimator to maximum likelihood estimation for SEM is 2SLS (Bollen, 1996). Like SI, 2SLS estimation is an observed variable technique. It has been shown to be particularly useful when model misspecification is a concern (Bollen & Bauer, 2004), with a regularized variant performing well in the context of small samples. 2SLS estimation reconfigures the structural portion of a SEM in terms of a model that directly relates the observed indicator variables with one another. More specifically, in the context of SEM, each latent variable can be expressed in terms of the indicator variable to which it is related, along with random error. In order to avoid the problem of indicators being correlated with error terms, an instrumental variables approach must be used, as described in Bollen (1996) and Bollen and Bauer (2004). The interested reader is referred to these publications for a full description of 2SLS.

The 2SLS approach described above can be extended to incorporate regularization methods, which have been shown to be effective in the context of small samples (Jung, 2013; Miller, Finch, & Ballenger, 2019). In general, regularization techniques are designed to reduce the number of non-zero parameter estimates in a model, which in turn makes them potentially useful with small samples and/or high dimensional data structures; i.e., when the number of parameters to be estimated approaches the sample size (Hastie, Tibshirani, & Wainwright, 2015). Previous research has found that standard 2SLS estimates can be biased when samples are small (Johnston, 1984). Jung proposed using regularization to identify a subset of the statistically most salient variables from among the full set, thereby reducing the number of parameters that need to be estimated. These methods rely on penalties to non-zero model coefficients that are applied to the minimization function. The three most common regularization penalties, and those used in the current study, are the ridge, the lasso, and the elastic net. A full description of these methods is not included here, but the interested reader is referred to Jung (2013) for a full description of them in the context of 2SLS estimation.

Prior work examining the performance of 2SLS in general shows that its estimates are less biased than those from maximum likelihood for misspecified models, and that neither works particularly well with small samples (Bollen et al., 2007; Bollen, 1996). One of the primary motivations behind Jung's (2013) development of the regularized 2SLS estimator was to improve its performance when sample sizes are small. Jung investigated the performance of regularized 2SLS with samples sizes of 5,10,20, 30, 40, and 50, and found that it produced more stable and accurate parameter estimates than did either standard 2SLS or maximum likelihood estimation. Miller, Finch, and Ballenger (2019) extended this work by comparing regularized 2SLS with regularized SEM based on maximum likelihood (Jacobucci, Grimm, & McArdle, 2016), as well as standard 2SLS and maximum likelihood, and found that regularized 2SLS yielded less biased and more efficient parameter estimates with samples of 150 or fewer, and when factor loadings were relatively small (0.4). The current study seeks to continue expanding on this line of research by comparing regularized 2SLS with other approaches that have been recommended for application in the small sample size case.

Factor Score Regression

Conceptually, factor scores are simply observed variable estimates of the latent trait that underlies a set of observed indicator variables (Gorsuch, 1983). The resulting scores can then be used in other analyses, such as regression models. Researchers have suggested the use of factor scores as an alternative means of incorporating information about the latent structure underlying a set of observed variables into regression (and other) models, while at the same time dealing with the issue of small sample size (Devlieger & Rosseel, 2017; Lu, Kwan, Thomas, Cedzynski, 2011). Such scores are individual variables, as with regression and PA, but incorporating information about error unlike those observed variable approaches.

There are multiple ways in which such factor scores can be calculated. Perhaps the most common of these approaches is the regression score method, which was proposed by Thurstone (1935). The observed variables are first standardized, and then equation (3) is applied to these standardized variables.

$$\hat{F} = Z\boldsymbol{R}^{-1}\boldsymbol{\Lambda\Phi} \tag{3}$$

where: $\hat{F}$ =Estimated factor score

Z =Standardized observed indicator scores

$\boldsymbol{R}$ =Correlation matrix for observed indicators

$\boldsymbol{\Lambda}$ =Factor pattern matrix

$\boldsymbol{\Phi}$ =Interfactor correlation matrix; for an orthogonal rotation the off-diagonal elements are 0.

The regression method yields a factor score for each latent trait for each individual in the sample by combining information from all of the observed indicators with non-zero factor loadings. Therefore, the score for a given factor is predominantly reflective of the variables that have the largest loadings associated with that factor, but all indicators with non-zero loadings will contribute something to each of the factor scores. In the context of confirmatory factor analysis, where loadings for indicators not theoretically associated with the factor are set to 0, the factor scores are representative of just the conceptually salient indicators.

One weakness of the regression approach to calculating factor scores is that the resulting values are biased (Gorsuch, 1983). An alternative technique that does not yield biased results is Bartlett's (1937) method:

$$\hat{F} = Z\boldsymbol{U}^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Lambda}'\boldsymbol{U}^{-1}\boldsymbol{\Lambda})^{-1} \tag{4}$$

where: $\boldsymbol{U}$ =Diagonal matrix of the variances of the factor scores.

Bartlett's method was designed to minimize the impact of the unique errors associated with individual indicators that is not accounted for by the factors.

Factor scores are popular tools for researchers to use in the context of fitting a SEM in the context of small samples (Devlieger, Mayer, & Rosseel, 2016). In this factor score regression (FSR), the factor scores are estimated for each latent variable, using one of the approaches described above, and then are subsequently included in a regression model. Thus, the ordinary least squares regression estimate for the slope relating two factor scores also serves as the estimate of the structural coefficient relating two latent variables in a SEM. However, the use of common methods for calculating factor scores, such as the regression approach, are known to be yield biased regression coefficients when used in the manner described above (e.g., Lewis & Linzer, 2005). Croon (2002) developed a technique designed to correct this bias whereby factor scores are first computed using a standard approach such as the regression method, after which the covariance matrix of these factor scores is used to calculate the covariance matrix of the actual latent variables. The regression coefficient linking the factor scores is then calculated using the true latent variable covariances and variances, thereby yielding an unbiased estimate of the SEM structural coefficient (Croon).

Although simulation work has shown that this corrected FSR method does indeed eliminate bias in the regression coefficient estimate (Devlieger, Mayer, & Rosseel, 2016), the standard errors for these estimates cannot be obtained from a standard regression analysis, as they are based on the uncorrected covariances and variances of the factor scores. Devlieger, et al. developed an alternative standard error for use with FSR that accounts for model error due to the regression, as well as to the factor scores themselves. In a simulation study, they showed that when used in conjunction with the bias correcting approach to conducting FSR, this approach performed well, yielding more accurate standard error estimates, Type I error rates that were under control, and the most optimal power values of the methods studied. This approach to fitting FSR has been incorporated in the R software `lavaan` package (Rosseel, 2012) through the `fsr` function, which was used in the current study.

Study Goals

The primary goal of this study was to extend upon the work of Savalei (2019) by examining an additional set of models, and more methods for estimating scale reliability, including GLB, Omega, and SH, in addition to Alpha and the constant reliability approach. In addition, the current work builds on research by Miller, Finch, and Ballenger (2019) through an investigation of the impact of scale reliability and model type on the performance of 2SLS regularized models, which have been shown to be particularly effective for small sample SEM problems. Of particular interest in the current study is to ascertain how the SI approach compares to these 2SLS regularized models for small samples, and whether there are optimal approaches for obtaining reliability values to apply in the SI context. In addition, these two sets of techniques were compared with standard SEM, as well as with PA and FSR. These latter approaches were included in the study as they represent standard practice (SEM or PA) in many cases, or because they offer an alternative paradigm for estimating relatively complex models with small samples (FSR). The methodology used to address these study goals is described below.

Methods

In order to address the study goals outlined above, a Monte Carlo simulation study design was used. For each combination of conditions (described below), a total of 1000 replications were generated. Data generation was done using Mplus, Version 8 (Muthèn & Muthèn, 2019), and data analysis was done using R, version 3.6.1 (R Core Team, 2019). Specific R libraries for the various analyses are described below. Data from 4 separate SEMs were simulated, where each factor had 5 observed indicator variables, and relationships among these latent variables differed by model. The indicators were generated from the multivariate normal distribution with mean of 0, and covariance matrix $\mathbf{\Sigma}$. The diagonal elements of $\mathbf{\Sigma}$ were set to 1, with the off-diagonal elements determined by the level of scale population reliability. The error terms of the indicators were related to the population reliability of their latent variables as:

$$\varepsilon = \frac{3(1-\rho)}{\rho} \quad (5).$$

Data Generating Models

Data were generated from four separate models, for each of which the simulation conditions described below were completely crossed. The data generating models were selected to represent examples that are frequently encountered in practice, and because they represent a variety ranging from relatively simple to more complex relationships among the latent variables. The models appear in the 4 panels of Figure 1. The first model represents a simple linear relationship among the latent variables, whereas model 2 involves partial mediation of the relationship between an exogenous and an endogenous variable, with 2 mediators. Model 3 was a multiple indicators multiple causes (MIMIC) model, with two latent variables, and two observed covariates. Finally, the fourth data generating model featured a nonlinear term in the form of an interaction between the two latent exogenous variables.

For each model one parameter was selected to be the focus of the discussion in the results section. In addition, parameter values were manipulated in order to allow for the investigation of model performance for cases when the parameter of interest was 0 in the population, as well as when it was larger than 0. For model 1, the structural coefficients linking F1 and F2 to F3 were the focus. These were set to either 0 or 0.5. For model 2, the indirect effects between F1 and F4 served as the focal parameters and took the values of either 0 or 0.25. In model 3, the focal parameters were the coefficients linking X1 to F1 and F2, and were set to either 0 or 0.5 in the population. Finally, for model 4 the moderator effect was the parameter of interest, and it was either 0 or 0.5 in the population.

Sample Size

One of the key manipulated variables in this study was the sample size, values of which were taken from Savalei (2019), and included 30, 50, 80, 100, and 200. These sample sizes were selected because the current study is an extension of this earlier work, and because these represent values ranging from what might be considered quite small (30) to moderate in size (200). The second key variable to be manipulated in this study was the level of reliability for the latent variables.

Population Reliability

The data were simulated so that the population reliability (ρ) values were either 0.65, 0.75, or 0.85. Once again, these reliability values were used by Savalei, and were selected in order to replicate her work.

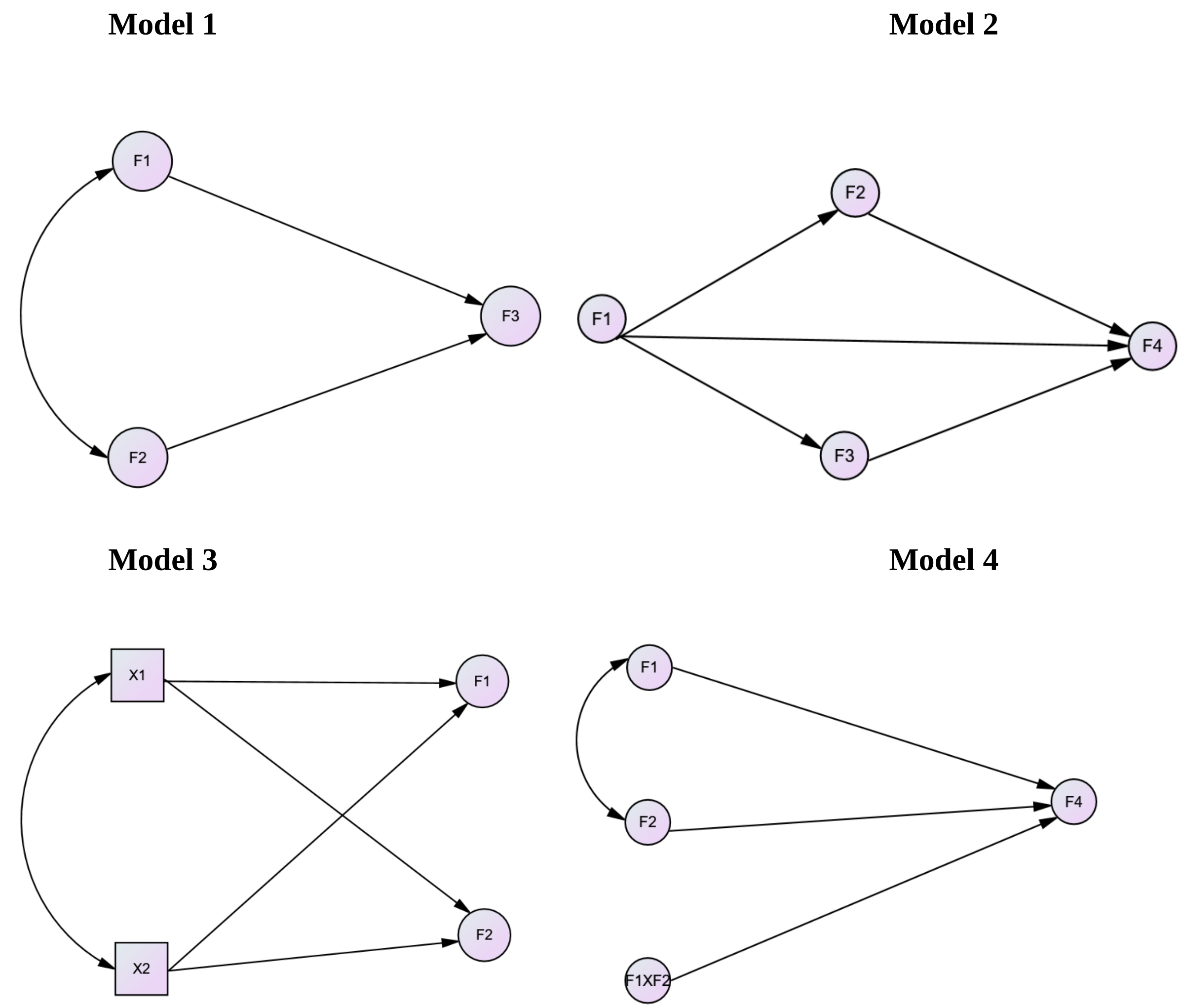

Figure 1. Data Generating Models

For one set of conditions, each latent variable had the same level of reliability, so that in the 0.75 condition, all factors were generated to have reliability of 0.75. An additional condition involved simulating the data such that each latent trait had a different level of reliability, i.e., one had reliability of 0.65, another 0.75, and the third 0.85. This represents an extension from earlier work in which all scales had a common population reliability.

Estimation Method

The data were fit using each of the models described above, including SI with Cronbach's Alpha, McDonald's Omega, the GLB, SH, and a constant reliability value of 0.8. For all but the constant approach, reliability estimates were obtained from the data and then applied as described above. In the case of Omega, a single factor model was fit to each of the latent variable indicator sets separately. Likewise, Alpha, GLB, and SH were also obtained for the individual latent variables individually. Reliability estimates were obtained from the data using functions from the R `Psych` package (Revelle, 2018). The reliability estimates were then used to determine the error terms for each observed indicator, and these errors were set to this specific value in the manner described above for SI model estimation. The models were fit using the `lavaan` R library (Rosseel, 2012). In addition to these 5 SI reliability models, FSR PA, and standard SEM were also used to fit the models. PA involved the calculation of composite scores through averaging the indicators for each of the latent variables. These composites were then related to one another using PA. For each model, standard SEM was also used in which the factors were fit to the data and then related to one another with a structural model. FSR was fit to the data using the `fsr` function in the `lavaan` R library, whereas PA and SEM were both fit to the data using `lavaan`. The regularized 2SLS estimators were fit using the glmnet R package (Friedman, Hastie, Tibshirani, Simon, Narasimhan, & Qian, 2018), with leave-one-out cross-validation being used to determine the optimal values of λ and α (Molinaro, Simon, & Pfeiffer, 2005). A total of 100 values of the tuning parameters were fit to the data, with the optimal model being the one that minimized the mean cross-validated error based on the leave-one-out cross validation. This model was then determined to be optimal, and its estimates were reported.

Study Outcomes

The outcomes of interest in the study were the absolute estimation bias of the key coefficients for each model (as defined above), the empirical standard errors of the key coefficients, the observed coverage rates of the nominal 95% confidence interval for these coefficients, and the Type I error and power rates for the key coefficients. For each key value, the absolute bias was calculated as:

$$|b-\beta| \quad (6)$$

where: `b =Sample estimate and β =Population data generating value.

The empirical standard error was the standard deviation of the b values across study replications. Coverage rates were calculated as the proportion of replications for which the 95% confidence interval calculated using b contained the population value β. The Type I error rates were calculated as the proportion of replications for which the 95% confidence interval calculated using b was found not to contain 0, when in fact β was 0 in the population. Conversely, power was the proportion of replications for which the 95% confidence interval did not include 0 when β was not 0 in the population.

Results

Model 1

The structural coefficient parameter estimation bias by level of reliability, sample size, and estimation method appears in Table 1. Across methods, bias was more extreme for smaller samples and lower levels of reliability. The combined impact of these two variables was most notable for the FSR, SEM, and Alpha estimation approaches, which had the highest rates of bias for reliability of 0.65 and a sample size of 80 or less. The lowest bias when reliability was 0.65 and samples were 80 or fewer was associated with the constant reliability approach, PA, and GLB. Under other conditions simulated here, bias was comparable across the methods. Finally, when population reliability differed across the latent variables, the estimation method bias results were between those of reliability of 0.75 and 0.85 for all estimation methods.

Table 1: Structural coefficient estimation bias by estimation method, sample size, and reliability: Model 1

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	2.03	0.66	1.10	1.59	0.39	2.96	0.31	3.96	0.44	0.56	0.61
	50	0.17	0.05	0.07	0.06	0.03	0.25	0.03	0.19	0.08	0.16	0.16
	80	0.20	0.00	0.01	0.00	0.00	0.11	0.00	0.12	0.01	0.01	0.02
	100	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.01	0.01
	200	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
.75	30	0.08	0.06	0.07	0.07	0.06	0.07	0.05	0.08	0.06	0.07	0.06
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
	80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	200	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
.85	30	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.03	0.03	0.03
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
	200	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Differ	30	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.06	0.05	0.04
	50	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	80	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	200	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 2: Structural coefficient standard error by estimation method, sample size, and reliability: Model 1.

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	3.02	0.62	1.27	2.97	0.35	3.89	0.27	0.19	0.60	1.56	0.92
	50	0.61	0.23	0.29	0.27	0.18	0.54	0.15	0.25	0.36	0.44	0.46
	80	0.48	0.18	0.22	0.25	0.15	0.28	0.12	0.11	0.24	0.26	0.25
	100	0.41	0.16	0.19	0.20	0.13	0.20	0.10	0.09	0.22	0.23	0.22
	200	0.21	0.11	0.13	0.13	0.09	0.14	0.07	0.07	0.16	0.15	0.15
.75	30	0.36	0.22	0.25	0.26	0.24	0.26	0.19	0.17	0.30	0.33	0.32
	50	0.24	0.18	0.20	0.20	0.19	0.20	0.15	0.14	0.24	0.24	0.24
	80	0.18	0.14	0.15	0.15	0.14	0.15	0.11	0.11	0.18	0.18	0.18
	100	0.15	0.13	0.13	0.14	0.13	0.14	0.10	0.10	0.16	0.16	0.16
	200	0.11	0.09	0.09	0.09	0.09	0.10	0.07	0.07	0.11	0.11	0.11
.85	30	0.26	0.22	0.23	0.23	0.24	0.23	0.19	0.18	0.28	0.29	0.28
	50	0.19	0.17	0.17	0.17	0.18	0.17	0.14	0.14	0.21	0.21	0.21
	80	0.15	0.13	0.14	0.14	0.14	0.14	0.11	0.11	0.16	0.16	0.16
	100	0.13	0.12	0.12	0.12	0.13	0.12	0.10	0.10	0.15	0.14	0.14
	200	0.09	0.08	0.09	0.09	0.09	0.09	0.07	0.07	0.10	0.10	0.10
Differ	30	0.27	0.21	0.23	0.25	0.22	0.24	0.17	0.13	0.35	0.42	0.39
	50	0.20	0.15	0.17	0.17	0.16	0.16	0.12	0.08	0.24	0.25	0.25
	80	0.15	0.12	0.13	0.13	0.12	0.13	0.10	0.07	0.19	0.19	0.19
	100	0.13	0.11	0.12	0.12	0.11	0.11	0.09	0.06	0.17	0.17	0.17
	200	0.09	0.08	0.08	0.08	0.08	0.08	0.06	0.04	0.12	0.12	0.12

The average standard errors for the structural coefficient parameters by reliability, sample size, and estimation method appear in Table 2. Standard errors were largest for all methods when reliability was 0.65, and samples were 80 or fewer. The largest standard errors in these conditions belonged to Alpha, with the smallest being associated with FSR, PA, and constant reliability. Indeed, across levels of reliability the standard errors for these methods was the lowest for samples of 100 or fewer individuals. The regularization methods yielded somewhat larger standard errors than did the other approaches (except for Alpha) across conditions. As with estimation bias, when reliability values differed, the standard errors were between those of reliability 0.75 and 0.85.

Coverage rates for the structural coefficient parameter values appear in Table 3. Across conditions, coverage rates for all estimation methods except FSR was above 0.90, and generally at or above the nominal 0.95 level. As noted, FSR was the exception for a reliability value of 0.65 and sample sizes of 30 and 50, as well as reliability of 0.75 and a sample size of 30. These low coverage rates appear to be due to the relatively high parameter estimation bias and low standard errors for FSR under these combinations of conditions. Across conditions, all methods held the Type I error rate below the nominal 0.05 level, except for FSR and PA. For these estimators, the error rate was greater than 0.10 for combinations of sample size of 100 or more, and population reliability of 0.75 or less. The other methods exhibited Type I error rates that were at or below 0.05 regardless of sample size or reliability value. With respect to power (Table 4), all methods exhibited higher rates in conjunction with larger samples and higher latent variable reliability values. FSR had the highest power rates, but these must be interpreted with great caution given the inflated Type I errors discussed above. Among the other methods, the highest power rates for the structural coefficient were associated with the GLB, Omega, and constant reliability approaches. With the exception of the largest sample size condition, the regularization methods generally exhibited the lowest power rates.

Model 2

As noted above (Figure 1), the second model from which data were simulated in this study included two mediators. The focus of this portion of the simulation study is on estimation of the mediator effects. The estimation bias results appear in Table 5. The largest bias for each of the methods was associated with smaller sample sizes and lower population reliability values. For the lowest reliability condition, the greatest bias in the mediation effects was associated with FSR, followed by SEM and Alpha for a sample of 30, with PA yielding a similar degree of bias for larger sample sizes at this lowest reliability level. As the sample size increased, bias declined for most of the methods, with this effect being least marked for PA

Table 3: Structural coefficient coverage rate by estimation method, sample size, and reliability: Model 1.

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00
	50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00
	80	1.00	0.95	0.95	0.95	0.92	0.95	0.93	0.91	0.95	0.95	0.95
	100	0.98	0.95	0.95	0.95	0.93	0.95	0.93	0.95	0.92	0.92	0.92
	200	0.96	0.95	0.96	0.96	0.95	0.95	0.95	0.97	0.94	0.94	0.94
.75	30	0.96	0.93	0.93	0.93	0.93	0.90	0.93	0.68	0.94	0.94	0.94
	50	0.96	0.95	0.95	0.95	0.95	0.95	0.96	0.92	0.95	0.95	0.95
	80	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.94	0.94	0.94	0.94
	100	0.94	0.94	0.94	0.94	0.94	0.94	0.93	0.94	0.96	0.96	0.96
	200	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.95	0.96	0.96	0.96
.85	30	0.95	0.94	0.94	0.94	0.94	0.94	0.94	0.97	0.94	0.95	0.95
	50	0.94	0.94	0.94	0.94	0.94	0.93	0.94	0.96	0.95	0.95	0.95
	80	0.95	0.95	0.95	0.95	0.95	0.94	0.95	0.96	0.96	0.96	0.96
	100	0.94	0.94	0.94	0.94	0.94	0.95	0.95	0.95	0.95	0.95	0.95
	200	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Differ	30	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.85	1.00	1.00	1.00
	50	0.97	0.97	0.97	0.97	0.97	0.96	0.97	0.94	0.94	0.94	0.94
	80	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.95
	100	0.94	0.94	0.94	0.94	0.94	0.93	0.94	0.96	0.96	0.96	0.96
	200	0.96	0.96	0.95	0.95	0.96	0.95	0.95	0.95	0.96	0.95	0.95

Table 4: Structural coefficient power rate by estimation method, sample size, and reliability: Model 1.

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.09	0.12	0.12	0.08	0.15	0.07	0.22	0.36	0.23	0.23	0.23
	50	0.25	0.75	0.75	0.75	0.75	0.50	0.75	0.78	0.50	0.50	0.50
	80	0.40	0.80	0.80	0.80	0.80	0.60	0.80	0.80	0.40	0.40	0.40
	100	0.44	0.74	0.74	0.74	0.78	0.52	0.74	0.93	0.52	0.52	0.52
	200	0.95	0.98	0.98	0.98	0.99	0.94	0.98	1.00	0.89	0.89	0.89
.75	30	0.57	0.67	0.67	0.56	0.64	0.56	0.68	0.75	0.33	0.33	0.33
	50	0.70	0.73	0.72	0.72	0.73	0.70	0.74	0.79	0.56	0.55	0.55
	80	0.90	0.91	0.90	0.90	0.91	0.89	0.91	0.88	0.77	0.77	0.77
	100	0.94	0.95	0.95	0.95	0.95	0.94	0.95	0.92	0.85	0.84	0.84
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99
.85	30	0.61	0.72	0.71	0.63	0.73	0.62	0.76	0.80	0.36	0.33	0.33
	50	0.80	0.81	0.81	0.81	0.80	0.81	0.81	0.92	0.67	0.66	0.66
	80	0.93	0.94	0.93	0.94	0.93	0.93	0.94	0.97	0.85	0.85	0.85
	100	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.99	0.93	0.93	0.93
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Differ	30	0.18	0.25	0.24	0.15	0.25	0.18	0.38	0.09	0.13	0.13	0.13
	50	0.63	0.67	0.67	0.67	0.66	0.60	0.67	0.86	0.47	0.46	0.46
	80	0.83	0.84	0.84	0.83	0.83	0.80	0.85	0.97	0.71	0.71	0.71
	100	0.90	0.90	0.90	0.90	0.90	0.85	0.90	0.98	0.83	0.83	0.83
	200	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	0.98	0.98	0.98

and FSR. The methods that yielded the least biased parameter estimates were GLB, and Omega. Having said this, when reliability was 0.75 or higher, and the sample size was 100 or larger, GLB, Omega, Split-half, Constant, and SEM all performed very similarly in terms of estimation bias for the indirect effect in the mediation model. Bias results for the regularization approaches were generally between those of the best performing methods (i.e., GLB and Omega), and those of the worst performers (PA, FSR, and Alpha).

The standard errors associated with the indirect effect, by estimation method, population reliability, and sample size appear in Table 6. When population reliability was 0.65 or 0.75 and sample sizes were 50 or

Table 5. Structural coefficient estimation bias by estimation method, sample size, and reliability: Model 2

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.08	0.18	0.13	0.11	0.34	0.14	0.50	0.22	0.21	0.25	0.08
	50	0.08	0.08	0.11	0.09	0.12	0.14	0.15	0.17	0.17	0.17	0.08
	80	0.05	0.01	0.08	0.08	0.01	0.14	0.13	0.05	0.08	0.05	0.05
	100	0.05	0.01	0.03	0.09	0.07	0.13	0.12	0.04	0.04	0.03	0.05
	200	0.02	0.03	0.01	0.07	0.00	0.13	0.10	0.07	0.06	0.05	0.02
.75	30	0.05	0.03	0.06	0.08	0.10	0.08	0.09	0.15	0.20	0.17	0.05
	50	0.02	0.01	0.01	0.05	0.01	0.08	0.09	0.04	0.03	0.03	0.02
	80	0.01	0.01	0.01	0.04	0.00	0.08	0.08	0.04	0.03	0.02	0.01
	100	0.00	0.01	0.01	0.03	0.01	0.08	0.08	0.03	0.03	0.02	0.00
	200	0.01	0.00	0.00	0.03	0.00	0.07	0.08	0.02	0.02	0.02	0.01
.85	30	0.01	0.03	0.03	0.01	0.03	0.06	0.08	0.04	0.05	0.05	0.01
	50	0.01	0.00	0.00	0.01	0.00	0.06	0.06	0.03	0.03	0.04	0.01
	80	0.00	0.01	0.01	0.01	0.00	0.05	0.06	0.02	0.02	0.02	0.00
	100	0.00	0.01	0.01	0.01	0.00	0.05	0.06	0.02	0.02	0.02	0.00
	200	0.00	0.01	0.01	0.00	0.00	0.05	0.06	0.02	0.02	0.02	0.00
Differ	30	0.09	0.11	0.10	1.48	0.03	0.10	0.14	0.14	1.46	0.14	0.09
	50	0.05	0.05	0.07	0.07	0.03	0.09	0.12	0.05	0.06	0.05	0.05
	80	0.04	0.03	0.07	0.07	0.03	0.09	0.13	0.04	0.04	0.05	0.04
	100	0.03	0.03	0.07	0.07	0.02	0.09	0.13	0.03	0.04	0.04	0.03
	200	0.03	0.02	0.08	0.08	0.01	0.10	0.13	0.03	0.02	0.02	0.03

Table 6. Structural coefficient standard error by estimation method, sample size, and reliability: Model 2.

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.84	0.27	0.57	0.60	0.30	0.55	0.14	0.12	0.31	0.31	0.32
	50	0.53	0.14	0.23	0.39	0.11	0.34	0.07	0.10	0.27	0.29	0.27
	80	0.23	0.13	0.21	0.17	0.10	0.15	0.06	0.09	0.25	0.25	0.25
	100	0.19	0.12	0.15	0.13	0.08	0.12	0.05	0.07	0.22	0.23	0.22
	200	0.11	0.08	0.11	0.12	0.06	0.12	0.04	0.06	0.16	0.16	0.16
.75	30	0.71	0.16	0.21	0.22	0.20	0.20	0.12	0.12	0.30	0.34	0.31
	50	0.42	0.13	0.16	0.16	0.15	0.18	0.09	0.10	0.25	0.26	0.25
	80	0.17	0.10	0.11	0.12	0.11	0.12	0.07	0.09	0.20	0.19	0.19
	100	0.14	0.09	0.10	0.10	0.10	0.10	0.06	0.08	0.17	0.17	0.17
	200	0.09	0.07	0.07	0.07	0.07	0.07	0.04	0.06	0.12	0.12	0.12
.85	30	0.25	0.16	0.19	0.19	0.23	0.19	0.13	0.16	0.30	0.31	0.30
	50	0.16	0.12	0.13	0.13	0.16	0.13	0.09	0.11	0.23	0.23	0.23
	80	0.12	0.10	0.10	0.10	0.12	0.10	0.07	0.09	0.18	0.18	0.17
	100	0.10	0.09	0.09	0.09	0.11	0.09	0.07	0.08	0.16	0.16	0.16
	200	0.07	0.06	0.06	0.06	0.07	0.06	0.05	0.05	0.11	0.11	0.11
Differ	30	0.54	0.13	0.18	0.15	0.46	0.16	0.09	0.09	0.34	0.37	0.35
	50	0.15	0.09	0.11	0.11	0.13	0.10	0.07	0.06	0.26	0.26	0.25
	80	0.10	0.07	0.08	0.08	0.09	0.08	0.05	0.05	0.20	0.20	0.20
	100	0.09	0.07	0.07	0.07	0.08	0.07	0.05	0.04	0.18	0.17	0.17
	200	0.06	0.05	0.05	0.05	0.05	0.05	0.03	0.03	0.12	0.12	0.12

less, the largest standard errors were associated with the Alpha estimator. For samples of 80 or more, the largest standard errors were produced by the regularization methods, regardless of the population reliability. On the other hand, PA and FSR consistently yielded estimates with the lowest standard errors, across study conditions. Among the reliability based estimation methods excluding Alpha, there was not a clearly discernible pattern with respect to the relative magnitude of the standard errors.

Coverage rates by estimation method, population reliability, and sample size appear in Table 7. When the population reliability was 0.65 and sample size was 50 or fewer, coverage rates for all methods was 1.0.

Table 7. Structural coefficient coverage rate by estimation method, sample size, and reliability: Model 2.

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00
	50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	80	1.00	0.89	0.93	0.93	0.90	0.93	0.57	0.56	1.00	1.00	1.00
	100	0.88	0.95	0.96	0.94	0.93	0.96	0.67	0.65	1.00	0.96	1.00
	200	0.97	0.94	0.96	0.96	0.97	0.94	0.73	0.74	0.98	0.98	0.99
.75	30	1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.75	1.00	1.00	1.00
	50	0.96	0.88	0.91	0.92	0.90	0.90	0.75	0.77	0.99	0.99	0.99
	80	0.97	0.94	0.94	0.94	0.93	0.93	0.83	0.84	0.99	0.99	0.99
	100	0.97	0.93	0.95	0.95	0.93	0.94	0.87	0.89	0.99	0.99	0.99
	200	0.95	0.95	0.93	0.93	0.92	0.92	0.95	0.94	0.98	0.99	0.99
.85	30	0.95	0.89	0.92	0.92	0.95	0.90	0.76	0.78	0.99	0.99	0.99
	50	0.95	0.91	0.92	0.92	0.94	0.92	0.83	0.81	0.99	0.99	0.99
	80	0.97	0.94	0.95	0.95	0.96	0.95	0.85	0.83	0.99	0.99	1.00
	100	0.97	0.95	0.96	0.95	0.96	0.95	0.91	0.91	0.99	1.00	1.00
	200	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.95	1.00	1.00	1.00
Differ	30	0.85	0.75	0.77	0.75	0.77	0.73	0.60	0.54	0.98	0.98	0.98
	50	0.87	0.74	0.79	0.78	0.81	0.73	0.49	0.56	0.99	0.99	0.99
	80	0.87	0.69	0.74	0.75	0.78	0.66	0.38	0.55	0.99	0.99	1.00
	100	0.85	0.68	0.73	0.74	0.77	0.63	0.28	0.58	1.00	1.00	1.00
	200	0.82	0.57	0.62	0.63	0.69	0.44	0.07	0.48	0.99	0.99	0.99

Table 8. Structural coefficient power rate by estimation method, sample size, and reliability: Model 2.

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	80	0.04	0.19	0.07	0.04	0.26	0.07	0.33	0.38	0.19	0.19	0.19
	100	0.08	0.46	0.19	0.23	0.46	0.12	0.58	0.57	0.12	0.12	0.12
	200	0.29	0.92	0.87	0.85	0.96	0.76	0.97	0.95	0.44	0.43	0.42
.75	30	0.00	0.25	0.25	0.26	0.26	0.00	0.26	0.25	0.00	0.00	0.00
	50	0.15	0.33	0.27	0.26	0.29	0.25	0.41	0.41	0.10	0.10	0.09
	80	0.51	0.68	0.66	0.65	0.67	0.61	0.76	0.77	0.20	0.19	0.18
	100	0.73	0.85	0.83	0.82	0.85	0.80	0.91	0.95	0.30	0.28	0.27
	200	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.65	0.63	0.62
.85	30	0.10	0.28	0.27	0.27	0.27	0.17	0.25	0.25	0.07	0.07	0.07
	50	0.36	0.44	0.41	0.41	0.36	0.41	0.49	0.47	0.15	0.14	0.14
	80	0.75	0.81	0.80	0.79	0.75	0.78	0.86	0.93	0.26	0.24	0.23
	100	0.87	0.92	0.91	0.91	0.88	0.90	0.95	0.97	0.35	0.33	0.33
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.81	0.80	0.80
Differ	30	0.04	0.08	0.08	0.06	0.02	0.10	0.10	0.10	0.04	0.04	0.04
	50	0.21	0.30	0.27	0.26	0.23	0.24	0.34	0.35	0.10	0.10	0.10
	80	0.56	0.65	0.63	0.62	0.58	0.55	0.71	0.71	0.16	0.15	0.15
	100	0.70	0.77	0.75	0.75	0.72	0.70	0.83	0.86	0.23	0.21	0.21
	200	0.98	0.99	0.99	0.99	0.98	0.97	1.00	1.00	0.61	0.58	0.58

This result was due to the large standard errors associated with all methods under these conditions. Indeed, the regularization techniques exhibited the largest standard errors under most of the study conditions, as described above, and also had the highest coverage rates. The lowest coverage rates were associated with PA and FSR except for 0.65 reliability sample size of 50 or fewer conditions, and for the reliability of 0.75 or higher and sample size of 200 conditions. The coverage rates for GLB, Omega, Split-half, and constant reliability approaches were above 0.93 for samples of 80 or more with reliability of 0.75 or higher. In addition, for reliability of 0.65, coverage was 0.93 or higher when the sample size was 100 or more. Finally, the Alpha estimation approach yielded coverage rates of 0.93 or higher except for population reliability of 0.65 and a sample size of 100.

Type I error rates for all methods were under the nominal 0.05 level across sample size and population reliability conditions, except for PA and FSR, which yielded error rates above 0.2 for samples of 100 or greater across all population reliability conditions. Power for detecting an indirect effect when was present in the population appear in Table 8, by estimation method, sample size, and population reliability.

For all methods, power increased concomitantly with increases in sample size. In addition, power was higher across methods for larger population reliability, except for sample sizes of 30 and 50 coupled with reliability of 0.75 and 0.85. In those cases, power was comparable across population reliability values. Otherwise, however, higher reliability was associated with greater power for detecting the mediation effect. In general, PA and FSR exhibited the highest power rates of the methods studied here. Again, however, given the Type I error inflation associated with these methods such increased power values need to be interpreted with great caution. The lowest power across methods was associated with the Alpha approach when reliability was 0.65. For the other reliability conditions, the regularization methods yielded the lowest power rates.

Model 3

The third model to be examined in this study was the MIMIC (Figure 1), involving two latent variables and two observed variables. The primary focus was on the coefficients linking the observed and latent variables. Mean parameter estimation bias for these coefficients appears in Table 9. As with models 1 and 2, estimation bias was greater when the population reliability of the scales was lower, and for smaller samples. This combined effect of reliability and sample size was most notable for FSR, SEM, Alpha, SH, and Omega. These estimators produced the most biased coefficient estimates for small samples and low reliability, but yielded estimates that were among the least biased for large samples coupled with high reliability. The exception to this latter result was Alpha, which consistently had among the most biased estimates across conditions. In contrast, constant reliability, PA, and the regularization techniques appear to have been the least impacted by sample size within each level of scale reliability, in terms of estimation bias. However, as with the other approaches, these methods exhibited less bias for scales with higher levels of reliability. When reliability was 0.85, the regularization models yielded the most biased parameter estimates. The method exhibiting the lowest degree of estimation bias for the coefficients linking the latent and observed variables at a reliability of 0.65 was GLB. Indeed, across conditions, GLB consistently yielded among the least biased parameter estimates.

Table 10 contains the standard errors for the MIMIC model coefficients. Higher standard errors for all estimators were associated with lower population reliability and smaller samples. In addition, PA and FSR had the smallest standard errors across conditions. In contrast, for reliability of 0.65, Alpha yielded the largest standard errors across all conditions, followed by the regularization methods, SEM, SH, and Omega. The pattern of standard errors among the methods when population reliability was 0.75 was similar to that for reliability of 0.65, though differences among the methods were somewhat less pronounced. For the highest reliability condition, the approaches based upon the reliability estimates all performed similarly to one another, and to the regularization methods, across sample size conditions. Again, PA and FSR had the lowest standard error estimates in the highest reliability setting.

Coverage rates for the estimation methods by population reliability and sample size appear in Table 11. In general, higher coverage rates were associated with higher population reliability and larger samples. For samples of 50 or more, coupled with population reliability of 0.75, 0.85, and the different reliability condition, coverage rates for Alpha, GLB, Omega, split-half, constant, and SEM were all 0.935 or higher. In contrast, coverage rates for FSR, and the regularization methods were above 0.93 only for population reliability of 0.85 and samples of 80 or more. PA had very similar coverage rates to FSR across study conditions. Finally, when the population reliability was 0.65, coverage rates for all methods were below the nominal 0.95 level for samples of 30 or 50. However, for samples of 80 or more, GLB, Omega, split-half, Constant, and SEM all exhibited coverage rates of 0.935 or higher.

In keeping with previously described patterns, across sample size and population reliability values, Type I error rates were held in control for all estimators except PA and FSR. For these two approaches, the error rates were greater than 0.1 for samples of 50 or more in conjunction with population reliability of 0.75 or less. In addition, for a population reliability of 0.85, PA and FSR exhibited Type I error rates in excess

Table 9. Structural coefficient estimation bias by estimation method, sample size, and reliability: Model 3

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.30	0.10	0.20	0.36	0.11	0.28	0.13	0.29	0.28	0.29	0.28
	50	0.16	0.07	0.13	0.26	0.09	0.22	0.10	0.27	0.27	0.28	0.27
	80	0.13	0.01	0.05	0.15	0.06	0.17	0.11	0.17	0.26	0.28	0.26
	100	0.10	0.04	0.04	0.11	0.08	0.07	0.11	0.05	0.21	0.26	0.20
	200	0.08	0.03	0.04	0.06	0.07	0.02	0.04	0.21	0.16	0.13	0.13
.75	30	0.28	0.04	0.03	0.02	0.08	0.00	0.10	0.16	0.15	0.19	0.17
	50	0.14	0.02	0.02	0.03	0.09	0.03	0.08	0.15	0.14	0.15	0.15
	80	0.11	0.00	0.03	0.03	0.08	0.03	0.07	0.13	0.14	0.14	0.14
	100	0.08	0.01	0.02	0.02	0.08	0.02	0.08	0.13	0.14	0.14	0.14
	200	0.07	0.00	0.02	0.02	0.07	0.01	0.08	0.13	0.13	0.14	0.14
.85	30	0.18	0.01	0.00	0.03	0.04	0.02	0.07	0.09	0.10	0.11	0.10
	50	0.06	0.02	0.00	0.01	0.03	0.00	0.06	0.08	0.10	0.10	0.10
	80	0.05	0.01	0.00	0.00	0.03	0.00	0.06	0.08	0.11	0.11	0.11
	100	0.05	0.01	0.00	0.00	0.03	0.00	0.06	0.08	0.11	0.11	0.11
	200	0.04	0.00	0.00	0.00	0.02	0.00	0.05	0.08	0.10	0.10	0.10
Differ	30	0.15	0.02	0.05	0.04	0.06	0.05	0.06	0.11	0.13	0.13	0.12
	50	0.09	0.00	0.02	0.02	0.05	0.02	0.04	0.06	0.13	0.13	0.13
	80	0.06	0.01	0.01	0.01	0.04	0.01	0.05	0.05	0.12	0.12	0.12
	100	0.05	0.00	0.01	0.01	0.04	0.01	0.05	0.05	0.11	0.11	0.10
	200	0.05	0.00	0.01	0.01	0.04	0.01	0.05	0.05	0.11	0.11	0.11

Table 10. Structural coefficient standard error by estimation method, sample size, and reliability: Model 3

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.81	0.30	0.38	0.42	0.27	0.40	0.21	0.23	0.53	0.60	0.53
	50	0.51	0.24	0.39	0.34	0.21	0.44	0.17	0.17	0.31	0.37	0.33
	80	0.47	0.19	0.23	0.24	0.16	0.25	0.13	0.12	0.24	0.25	0.24
	100	0.40	0.18	0.21	0.22	0.15	0.22	0.12	0.11	0.21	0.22	0.21
	200	0.21	0.13	0.15	0.15	0.10	0.15	0.08	0.08	0.15	0.15	0.15
.75	30	0.35	0.26	0.29	0.30	0.27	0.29	0.21	0.21	0.31	0.34	0.32
	50	0.27	0.21	0.22	0.23	0.21	0.23	0.17	0.17	0.24	0.24	0.24
	80	0.20	0.17	0.18	0.18	0.17	0.18	0.13	0.14	0.18	0.18	0.18
	100	0.18	0.15	0.16	0.16	0.15	0.16	0.12	0.12	0.16	0.16	0.16
	200	0.13	0.11	0.11	0.11	0.11	0.11	0.08	0.09	0.12	0.11	0.11
.85	30	0.29	0.24	0.26	0.26	0.27	0.26	0.22	0.22	0.28	0.30	0.29
	50	0.23	0.20	0.21	0.21	0.22	0.21	0.17	0.18	0.22	0.22	0.22
	80	0.17	0.16	0.16	0.16	0.17	0.16	0.13	0.14	0.17	0.17	0.17
	100	0.16	0.14	0.14	0.14	0.15	0.14	0.12	0.12	0.15	0.15	0.15
	200	0.11	0.10	0.10	0.10	0.11	0.10	0.08	0.09	0.11	0.10	0.10
Differ	30	0.26	0.20	0.22	0.21	0.22	0.21	0.17	0.13	0.33	0.37	0.35
	50	0.17	0.15	0.16	0.16	0.17	0.15	0.13	0.09	0.24	0.24	0.24
	80	0.13	0.12	0.12	0.12	0.13	0.12	0.10	0.07	0.19	0.19	0.19
	100	0.12	0.11	0.11	0.11	0.12	0.11	0.09	0.07	0.17	0.16	0.16
	200	0.08	0.08	0.08	0.08	0.08	0.07	0.07	0.05	0.12	0.12	0.12

of 0.12 for samples of 80 or larger. In other conditions, Power rates appear in Table 12. Generally speaking, power for all estimation methods increased concomitantly with larger samples and higher population reliability. In addition, power for detecting a MIMIC model effect was at or near 0 for sample sizes of 30 and 50 coupled with a reliability of 0.65, and for reliability of 0.75 and a sample size of 30. The method consistently exhibiting the highest power rates across conditions was FSR, followed by PA. The lowest power rates were associated with the regularization methods and Alpha. More specifically, when reliability was 0.65, Alpha had the lowest reliability values across methods studied here. For reliability values of 0.75

Table 11. Structural coefficient coverage rate by estimation method, sample size, and reliability: Model 3

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.74	0.80	0.80	0.75	0.80	0.79	0.61	0.62	0.77	0.77	0.76
	50	0.83	0.93	0.90	0.90	0.88	0.92	0.84	0.84	0.84	0.86	0.85
	80	0.84	0.94	0.97	0.93	0.97	0.94	0.91	0.91	0.78	0.76	0.79
	100	0.85	0.95	0.95	0.96	0.95	0.94	0.93	0.92	0.83	0.80	0.82
	200	0.86	0.96	0.96	0.95	0.95	0.96	0.94	0.93	0.90	0.89	0.90
.75	30	0.90	0.90	0.89	0.90	0.89	0.90	0.87	0.89	0.76	0.79	0.72
	50	0.94	0.94	0.95	0.95	0.95	0.94	0.91	0.90	0.83	0.82	0.84
	80	0.94	0.95	0.94	0.95	0.95	0.95	0.91	0.91	0.91	0.83	0.84
	100	0.95	0.96	0.96	0.96	0.97	0.96	0.95	0.91	0.91	0.82	0.82
	200	0.95	0.95	0.95	0.95	0.95	0.95	0.98	0.96	0.91	0.91	0.91
.85	30	0.94	0.93	0.93	0.93	0.93	0.92	0.92	0.88	0.86	0.84	0.86
	50	0.95	0.94	0.94	0.94	0.95	0.94	0.92	0.89	0.89	0.89	0.90
	80	0.95	0.94	0.95	0.94	0.95	0.94	0.94	0.92	0.94	0.95	0.96
	100	0.94	0.95	0.96	0.95	0.95	0.95	0.97	0.95	0.94	0.95	0.95
	200	0.95	0.95	0.95	0.95	0.95	0.95	0.97	0.97	0.97	0.95	0.96
Differ	30	0.84	0.84	0.83	0.83	0.85	0.89	0.85	0.84	0.82	0.83	0.83
	50	0.87	0.91	0.92	0.89	0.90	0.93	0.91	0.91	0.89	0.88	0.89
	80	0.94	0.94	0.94	0.94	0.93	0.94	0.93	0.93	0.92	0.90	0.92
	100	0.96	0.97	0.96	0.95	0.96	0.96	0.94	0.95	0.92	0.92	0.21
	200	0.97	0.96	0.95	0.95	0.96	0.96	0.95	0.95	0.92	0.93	0.93

Table 12. Structural coefficient power rate by estimation method, sample size, and reliability: Model 3

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	80	0.00	0.19	0.11	0.07	0.16	0.04	0.48	0.51	0.15	0.19	0.19
	100	0.06	0.50	0.35	0.42	0.48	0.15	0.65	0.68	0.15	0.15	0.16
	200	0.27	0.92	0.87	0.83	0.85	0.74	0.96	0.95	0.43	0.40	0.40
.75	30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
	50	0.13	0.31	0.26	0.25	0.27	0.23	0.40	0.42	0.10	0.11	0.10
	80	0.49	0.65	0.62	0.62	0.63	0.57	0.72	0.76	0.20	0.19	0.19
	100	0.73	0.87	0.85	0.84	0.86	0.81	0.91	0.95	0.28	0.26	0.25
	200	0.98	0.99	0.99	0.99	1.00	0.99	1.00	1.00	0.66	0.64	0.63
.85	30	0.07	0.13	0.11	0.09	0.05	0.12	0.17	0.25	0.08	0.11	0.10
	50	0.37	0.47	0.44	0.44	0.38	0.43	0.53	0.56	0.12	0.12	0.11
	80	0.73	0.79	0.78	0.78	0.74	0.77	0.83	0.88	0.22	0.21	0.21
	100	0.86	0.91	0.90	0.90	0.87	0.89	0.93	0.97	0.32	0.30	0.29
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.78	0.76	0.75
Differ	30	0.05	0.09	0.09	0.08	0.04	0.05	0.13	0.13	0.06	0.06	0.06
	50	0.22	0.32	0.29	0.28	0.23	0.24	0.38	0.44	0.09	0.09	0.09
	80	0.61	0.69	0.68	0.68	0.64	0.61	0.75	0.84	0.19	0.18	0.17
	100	0.76	0.82	0.81	0.81	0.78	0.74	0.87	0.90	0.28	0.26	0.25
	200	0.98	0.99	0.99	0.99	0.98	0.97	1.00	1.00	0.64	0.62	0.61

and 0.85, Alpha estimation yielded higher power for detecting the MIMIC effect than did the regularization approaches across sample sizes. However, power for Alpha remained below that of the other estimators except for a sample size of 200. Among the other reliability based estimation methods, GLB exhibited slightly higher power than the other such approaches, except for a sample size of 200 and reliability of 0.75, and sample sizes of 100 or 200 and reliability of 0.85.

Table 13. Structural coefficient estimation bias by estimation method, sample size, and reliability: Model 4

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.52	0.40	0.46	0.46	0.51	0.42	0.29	0.33	0.33	0.33	0.36
	50	0.39	0.22	0.28	0.27	0.32	0.25	0.19	0.22	0.28	0.28	0.29
	80	0.26	0.12	0.14	0.15	0.26	0.21	0.14	0.13	0.15	0.16	0.15
	100	0.15	0.10	0.12	0.12	0.15	0.16	0.11	0.11	0.10	0.12	0.12
	200	0.13	0.07	0.09	0.10	0.12	0.13	0.06	0.08	0.05	0.06	0.06
.75	30	0.40	0.23	0.27	0.29	0.34	0.30	0.21	0.22	0.19	0.21	0.21
	50	0.27	0.13	0.16	0.17	0.21	0.25	0.11	0.11	0.09	0.11	0.06
	80	0.24	0.11	0.12	0.16	0.22	0.20	0.08	0.09	0.08	0.09	0.09
	100	0.15	0.06	0.06	0.08	0.11	0.15	0.05	0.05	0.05	0.07	0.06
	200	0.10	0.05	0.05	0.07	0.08	0.08	0.04	0.05	0.04	0.05	0.06
.85	30	0.32	0.17	0.18	0.20	0.28	0.28	0.16	0.16	0.15	0.15	0.15
	50	0.20	0.11	0.14	0.14	0.17	0.18	0.11	0.11	0.08	0.10	0.10
	80	0.19	0.09	0.12	0.12	0.15	0.13	0.08	0.08	0.07	0.08	0.08
	100	0.12	0.05	0.05	0.06	0.10	0.10	0.05	0.04	0.05	0.05	0.06
	200	0.05	0.01	0.01	0.02	0.04	0.05	0.01	0.02	0.01	0.01	0.01
Differ	30	0.45	0.32	0.34	0.34	0.37	0.36	0.24	0.25	0.25	0.26	0.26
	50	0.37	0.26	0.26	0.26	0.29	0.29	0.16	0.16	0.16	0.16	0.16
	80	0.25	0.16	0.16	0.16	0.18	0.16	0.14	0.14	0.15	0.16	0.16
	100	0.22	0.09	0.09	0.09	0.12	0.10	0.07	0.07	0.08	0.08	0.09
	200	0.16	0.06	0.06	0.07	0.08	0.07	0.04	0.05	0.06	0.06	0.06

Table 14. Structural coefficient standard error by estimation method, sample size, and reliability: Model 4

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	1.15	1.03	1.04	1.07	1.08	1.17	0.90	0.95	1.16	1.19	1.14
	50	0.90	0.75	0.75	0.79	0.81	0.89	0.64	0.69	0.87	0.87	0.85
	80	0.82	0.73	0.72	0.71	0.75	0.81	0.57	0.59	0.82	0.85	0.84
	100	0.69	0.65	0.67	0.67	0.68	0.69	0.49	0.57	0.68	0.68	0.68
	200	0.27	0.25	0.26	0.27	0.27	0.25	0.19	0.19	0.26	0.25	0.25
.75	30	0.86	0.78	0.81	0.80	0.82	0.84	0.68	0.74	0.86	0.86	0.85
	50	0.81	0.64	0.67	0.67	0.73	0.79	0.52	0.57	0.75	0.74	0.75
	80	0.52	0.44	0.44	0.47	0.47	0.49	0.36	0.42	0.72	0.72	0.74
	100	0.49	0.41	0.41	0.42	0.44	0.47	0.34	0.38	0.52	0.52	0.52
	200	0.19	0.15	0.16	0.17	0.16	0.18	0.12	0.13	0.20	0.20	1.99
.85	30	0.60	0.57	0.55	0.61	0.64	0.59	0.53	0.54	0.68	0.69	0.68
	50	0.56	0.54	0.55	0.58	0.59	0.54	0.44	0.42	0.63	0.63	0.63
	80	0.37	0.35	0.38	0.36	0.38	0.34	0.32	0.37	0.42	0.42	0.42
	100	0.36	0.33	0.33	0.35	0.35	0.32	0.31	0.35	0.42	0.42	0.42
	200	0.18	0.15	0.15	0.15	0.16	0.16	0.14	0.15	0.16	0.16	0.16
Differ	30	0.80	0.70	0.69	0.66	0.70	0.70	0.59	0.64	0.74	0.75	0.75
	50	0.75	0.60	0.60	0.60	0.63	0.74	0.45	0.59	0.70	0.67	0.67
	80	0.45	0.41	0.41	0.45	0.49	0.42	0.38	0.40	0.53	0.52	0.52
	100	0.41	0.40	0.40	0.43	0.48	0.41	0.30	0.30	0.52	0.52	0.52
	200	0.13	0.11	0.11	0.11	0.11	0.12	0.11	0.12	0.15	0.14	0.14

Model 4

The fourth model that was considered in this study, involved a moderation effect between two latent variables, as seen in Figure 1, panel 4. The absolute bias results for the moderation effect by estimation method, population reliability, and sample size appear in Table 13. For all of the estimation methods, bias declined with increasing sample sizes and higher reliability values, which has been seen consistently across the models. With respect to the estimation method itself, PA exhibited the lowest degree of bias, followed by FSR and the regularization methods. On the other hand, the largest bias was exhibited by Alpha. Among

Table 15. Structural coefficient coverage rate by estimation method, sample size, and reliability: Model 4

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.64	0.69	0.69	0.69	0.69	0.68	0.81	0.80	0.73	0.72	0.73
	50	0.72	0.88	0.85	0.86	0.83	0.82	0.88	0.87	0.90	0.88	0.90
	80	0.86	0.88	0.86	0.86	0.86	0.88	0.92	0.93	0.92	0.92	0.94
	100	0.88	0.91	0.90	0.87	0.90	0.88	0.94	0.94	0.94	0.94	0.94
	200	0.92	0.93	0.93	0.92	0.92	0.94	0.95	0.95	0.98	0.97	0.98
.75	30	0.71	0.79	0.75	0.78	0.74	0.77	0.84	0.79	0.88	0.88	0.87
	50	0.76	0.86	0.83	0.88	0.86	0.89	0.93	0.92	0.90	0.90	0.90
	80	0.82	0.89	0.87	0.89	0.86	0.89	0.94	0.95	0.96	0.95	0.92
	100	0.93	0.95	0.94	0.95	0.93	0.93	0.95	0.96	0.95	0.95	0.95
	200	0.94	0.96	0.95	0.95	0.96	0.96	0.96	0.96	0.96	0.95	0.95
.85	30	0.78	0.86	0.85	0.85	0.82	0.83	0.88	0.90	0.91	0.91	0.91
	50	0.80	0.94	0.95	0.95	0.94	0.96	0.95	0.95	0.95	0.93	0.93
	80	0.86	0.95	0.95	0.95	0.94	0.96	0.95	0.96	0.96	0.94	0.94
	100	0.95	0.96	0.94	0.95	0.95	0.96	0.96	0.97	0.95	0.95	0.93
	200	0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.97	0.96	0.96	0.96
Differ	30	0.73	0.85	0.84	0.85	0.83	0.86	0.90	0.91	0.91	0.90	0.91
	50	0.75	0.86	0.86	0.86	0.91	0.88	0.92	0.93	0.94	0.94	0.94
	80	0.85	0.94	0.93	0.92	0.93	0.95	0.95	0.94	0.96	0.95	0.96
	100	0.64	0.69	0.69	0.69	0.69	0.68	0.81	0.80	0.73	0.72	0.73
	200	0.72	0.88	0.85	0.86	0.83	0.82	0.88	0.87	0.90	0.88	0.90

Table 16. Structural coefficient power rate by estimation method, sample size, and reliability: Model 4

Reliability	N	Alpha	GLB	Omega	SH	constant	SEM	PA	FSR	Ridge	Lasso	Net
.65	30	0.14	0.27	0.23	0.25	0.20	0.25	0.36	0.38	0.27	0.27	0.27
	50	0.25	0.38	0.33	0.33	0.28	0.30	0.48	0.45	0.34	0.35	0.33
	80	0.31	0.46	0.45	0.44	0.36	0.45	0.61	0.60	0.52	0.52	0.51
	100	0.35	0.52	0.53	0.52	0.45	0.53	0.78	0.80	0.65	0.66	0.67
	200	0.76	0.90	0.88	0.91	0.81	0.89	0.94	0.95	0.87	0.87	0.87
.75	30	0.22	0.40	0.38	0.36	0.33	0.37	0.52	0.51	0.42	0.41	0.44
	50	0.33	0.46	0.45	0.46	0.35	0.43	0.63	0.63	0.50	0.49	0.51
	80	0.59	0.81	0.79	0.82	0.67	0.76	0.85	0.88	0.85	0.85	0.86
	100	0.77	0.89	0.87	0.90	0.82	0.89	0.94	0.95	0.94	0.95	0.93
	200	0.91	0.99	0.99	0.99	0.93	0.99	1.00	1.00	1.00	1.00	1.00
.85	30	0.45	0.50	0.50	0.50	0.46	0.48	0.61	0.60	0.56	0.55	0.55
	50	0.57	0.66	0.64	0.63	0.58	0.60	0.74	0.77	0.67	0.67	0.67
	80	0.82	0.94	0.94	0.95	0.86	0.88	0.96	0.95	0.96	0.97	0.98
	100	0.91	0.98	0.98	0.98	0.93	0.98	1.00	1.00	0.99	1.00	0.99
	200	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Differ	30	0.25	0.33	0.34	0.32	0.27	0.27	0.40	0.39	0.40	0.40	0.40
	50	0.35	0.54	0.55	0.54	0.38	0.53	0.63	0.64	0.61	0.62	0.61
	80	0.66	0.79	0.80	0.77	0.69	0.75	0.84	0.86	0.85	0.84	0.86
	100	0.74	0.86	0.87	0.87	0.76	0.83	0.92	0.95	0.92	0.94	0.94
	200	0.90	1.00	0.99	1.00	0.92	1.00	1.00	1.00	1.00	1.00	1.00

the other reliability based approaches, the least bias was associated with GLB, followed by Omega and SH. Finally, when the population reliability was 0.85, and the sample size was 200, GLB, Omega, SH, PA, FSR, and the regularization methods all performed very similarly in terms of bias.

The standard errors of the interaction coefficient estimate by method, population reliability, and sample size appear in Table 14. Smaller standard errors were associated with larger samples and higher population reliability values. With regard to methods, PA and FSR exhibited the smallest standard errors, whereas the largest were associated with the regularization estimators.

Coverage rates for the 95% confidence interval of the interaction effect by estimation method, sample size, and population reliability appear in Table 15. Coverage rates increased concomitantly with sample size and population reliability value. When the population reliability was 0.65, only PA, FSR, and the regularization methods yielded coverage rates that were at or above 0.93, and then only for sample sizes of 100 or more. At this lowest reliability value, GLB also had coverage rates that were greater than 0.93 for a sample size of 200. When population reliability was 0.75 the coverage rates were 0.93 or higher for PA, FSR, and the regularization methods for samples of 80 or larger, and for all methods when N was 100 or more. Finally, for population reliability of 0.85, all estimators except for Alpha, yielded coverage rates of 0.93 or larger for sample sizes of 50 or more. When the reliability values differed, the regularization methods yielded in control coverage rates for samples of 50 or more, whereas PA, FSR, SEM, GLB, Omega, and the constant approaches required samples of 80 or more in order to have coverage approaching the nominal level. The Alpha estimator consistently exhibited the lowest coverage rates of the methods studied here.

In keeping with previously described patterns, the Type I error rates for all methods except PA and FSR were at or below the nominal 0.05 level across study conditions. PA and FSR each yielded error rates that were above 0.1 for samples of 50 or greater in conjunction with population reliability of 0.75 or less. When population reliability was 0.85, PA and FSR had Type I error rates of 0.1 or greater for samples of 80 or fewer.

Power rates by estimator, population reliability, and sample size appear in Table 16. PA and FSR had the highest power rates across conditions included in this study. However, these must be interpreted with caution given the elevated Type I error rates reported above. Among the methods that were able to control the Type I error rates, the regularization estimators yielded the highest power for sample sizes of 50 to 100 and reliability of 0.75 and 0.85. For the largest sample size condition, as well as for N=100 and reliability of 0.85, these regularization methods had comparable power to GLB, Omega, and SH. The lowest power rates were associated with Alpha and the constant reliability approaches.

Discussion

Summary of Results

Given the complexity of the results reported above, a summary is warranted. First, these results demonstrated that the optimal approach to use when estimating SEMs with small samples and low reliability values differs depending on the type of model being fit to the data, and the primary goal of the researcher. For example, if a researcher is working with a moderation model (Model 4), and the primary goal is to estimate the moderation effect as accurately as possible, then PA, FSR, or a regularization approach might be optimal. On the other hand, if the sample size is small and the researcher is working with a doubly mediated model (Model 2) then GLB or Omega would likely yield the least biased estimates of the indirect effects. Similarly, for the MIMIC model (Model 3) GLB exhibited the least degree of bias. Finally, for relatively simple models with no indirect effects or latent interactions (Model 1), several methods would likely yield relatively unbiased results.

A second major finding of this study is that both PA and FSR consistently exhibited inflated Type I error rates for the effects of interest. This result held true across sample size and population reliability. Therefore, if a researcher is particularly sensitive to not making a Type I error, neither of these approaches would likely be optimal, regardless of the model that is used. A third major finding of this study is that the Alpha approach generally did not perform as well as most of the other techniques studied here. In other words, researchers employing SI are best suited to avoid using Alpha as the reliability estimate in favor of GLB, SH, or in most cases Omega. The constant reliability approach generally did not perform as well as these other three methods, and thus would not be preferred either. A fourth finding of this study is that GLB consistently yielded reasonably accurate results, in terms of estimation bias, coverage rates, Type I error, and power. It was not always the optimal method, but it was typically among the better performers in terms of parameter estimation accuracy, efficiency of the estimate, coverage rate, and Type I error control. Such could not be said of many of the other methods, such as FSR, PA, regularization, and Omega, which yielded the optimal results for one model, but not for another.

A fifth result of interest to come out of this study is that the regularization methods should sometimes be the method of choice, but not in all cases. Prior research focused on relatively simple latent variable models, equivalent to Model 1 in this study. In addition, prior work did not examine the impact of underlying scale reliability on the performance of these methods, as was the case here. Results of the current study showed that the regularization approach should definitely be considered for small sample cases with models involving moderation (Model 4) or simple direct effects models, but not for models involving indirect effects (Model 2), or observed predictors of latent variables (Model 3).

Limitations and Directions for Future Research

As with any single study, there are limitations to the current work that should be addressed in future studies. First, although this study expanded on the number and variety of models examined in the context of modeling SEMs with small samples, it did not vary the factor structure or the number of indicators per factor. Thus, future work should consider cases in which the number of indicators per factor is both smaller and larger than 5. In particular, investigating the performance of these estimators with fewer indicators would be quite interesting. In addition, this study involved models that were correctly specified. However, in reality researchers will not know the actual structure of their model, and therefore might well fit one that is misspecified. Given that 2SLS is designed to ameliorate bias associated with model misspecification, and that regularized 2SLS has been shown to work well under such conditions (Jung, 2013; Miller, Finch, & Ballenger, 2019), future work should include model misspecification as a condition. Indeed, it is not known how the SI approaches in particular might work when the model fit to the data is not the same as the population model that generated the data. A third area for future research is with respect to the distribution underlying the indicators. In this study, the indicators were generated from the multivariate normal distribution. However, in practice scales made up of ordinal or nominal variables in the form of items might be used. Thus, future work should investigate the performance of the methods examined here when the indicator variables are categorical in nature.

Conclusions

The results of this study highlight the fact that when fitting SEMs with small samples there is not a single optimal estimator, but rather the choice of approach is to some extent dependent on the structure of the model itself. Having said that, it does seem clear that PA and FSR are likely to have elevated Type I error rates, regardless of the underlying model. In addition, though not always the absolute best approach, using SI in conjunction with a GLB, Omega, or SH estimate of reliability is likely to yield relatively accurate and efficient model parameter estimates. In this respect, the results of the current study mirror those of Savalei (2019), who also found that the SI approach yielded better Type I error control, in particular, than did observed variable methods such as PA. The current study extends on this earlier work by revealing that when using SI, there are better options than Alpha for estimating scale reliability. Furthermore, the use of traditional latent variable SEM with a maximum likelihood estimator is generally not recommended for samples of fewer than 100, based on the results of this study. Finally, the regularized 2SLS estimators were found to be particularly effective for fitting moderation models when the sample size was small. Taking these findings together, researchers who need to fit structural models involving latent variables with samples of 100 or fewer individuals should consider the nature of the model when selecting the estimator, and if in doubt about the optimal approach, should consider SI with GLB, Omega, or SH estimates of reliability.

References

Aggarwal, C.C. (2016) *Outlier analysis, 2nd edn*. Springer, Berlin.

Breunig, M.M., Kriegel, H.P., Ng, R.T., & Sander, J. (2000). LOF: identifying density-based local outliers. *Association for Computing Machinery Sigmod Rec 29,* 93–104.

Brodinova, S., Filzmoser, P., Ortner, T., Breiteneder, C., & Rohm, M. (2019). Robust and sparse *k*-means clustering for high-dimensional data. *Advances in Data Analytics and Classification, 13(4),* 905-932.

Candel, M.J. & van Breukelen, G.J. (2010). Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression. *Statistics in Medicine, 29(14),* 1488-1501.

Ester, M., Kriegel, H-P, Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96).* AAAI Press. 226– 231.

Evans, K., Love, T., & Thurston, S.W. (2015). Outlier identification in model-based cluster analysis. *Journal of Classification, 32(1),* 63-84.

Fraley, C. & Raftery, A. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association, 97,* 611–631.

Galimberti, G., Manisi, A., & Soffritti, G. (2018). Modelling the role of variables in model- based cluster analysis. *Statistical Computing, 18(1),* 145–169

Garcia-Escudero, L.A., & Gordaliza, A. (1999). Robustness properties of k-means and trimmed k-means. *Journal of the American Statistical Association, 94(447),* 956–969.

Hahs-Vaughn, D. (2017). *Applied Multivariate Statistical Concepts.* New York: Routledge, Taylor & Francis group.

Hajnal, I. & Loosveldt, G. (1998). The evaluation of the sensitivity of some clustering techniques to irrelevant variables. *Advances in Methodology, Data analysis, and Statistics, 14*, 61-75.

Henry, D.B., Tolan, P.H., & Gorman-Smith, D. (2005). Cluster analysis in family psychology research. *Journal of Family Psychology, 19(1)*, 121-132.

Holden, J. E., Finch, W. H., & Kelley, K. (2011). A Comparison of Two-Group Classification Methods. *Educational and Psychological Measurement, 71,* 870-901.

Kaufman, L. & Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley.

Kaufman, L. & Rousseeuw, P.J. (1987). Clustering by means of medoids, in Y. Dodge (Ed.),
Statistical Data Analysis based on the L1 Norm, 405–416.

Lewandowski, K.E., Sperry, S.H., Cohen, B.M., & Ongur, D. (2014). Cognitive variability in psychotic disorders: A cross-diagnostic cluster analysis. Psychological Medicine, 44(15), 3239-3248.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 281–297.

Neykov, N., Filzmoser, P., Dimova, R., & Neytchev, P. (2007). Robust fitting of mixtures using the trimmed likelihood estimator. *Computational. Statistics and Data Analysis. 52(1),* 299–308.

Prokasky, A., Rudasill, K., Molfese, V.J., Putnam, S., Gartstein, M., & Rothbart, M. (2017). Identifying child temperament types using cluster analysis in three samples. *Journal of Research in Personality, 67*, 190-201.

R Core Team. (2018). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Raftery, A.E., & Dean, N. (2006)Variable selection for model-based clustering. *Journal of the American Statistical Association, 101(473)*:168– 178

Rehm, F., Klawonn, F., & Kruse, R. (2007). A novel approach to noise clustering for outlier detection. *Soft Computing, 11(4),* 489-494.

Rocke, D.M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *Annals of Statistics, 24(3),* 1327–1345.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B, 58(1),* 267-288.

Viswanath, P. & Babu, S. (2009). Rough-DBSCAN: A fast hybrid density based clustering method for large data sets. *Pattern Recognition Letters, 30(16),* 1477-1488.

Send correspondence to: W Holmes Finch
Ball State University
Email: whfinch@bsu.edu