Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing Exploratory Structural Equation Modeling and Existing Approaches for Multiple Regression with Latent Variables. Structural Equation Modeling, 25(5), 737-749.

Comparing Exploratory Structural Equation Modeling and Existing Approaches for Multiple Regression with Latent Variables

Yujiao Mai

South China Normal University, China; University of Notre Dame, USA

Zhiyong Zhang

University of Notre Dame, USA

Zhonglin Wen

South China Normal University, China

Author Note

This research was funded by grants from the National Natural Science Foundation of China (31771245), the Graduate Research Fellowship Program of South China Normal University (2012kyjj106), and the U.S. Department of Education (R305D140037).

Correspondence should be addressed to Zhonglin Wen, School of Psychology, South China Normal University, Guangzhou, 510631, China. E-mail: wenzl@scnu.edu.cn

Abstract

Exploratory structural equation modeling (ESEM) is an approach for analysis of latent variables using exploratory factor analysis (EFA) to evaluate the measurement model. This study compared ESEM with two dominant approaches for multiple regression with latent variables, structural equation modeling (SEM) and manifest regression analysis (MRA). Main findings included: (1) ESEM in general provided the least biased estimation of the regression coefficients; SEM was more biased than MRA given large cross-factor loadings. (2) MRA produced the most precise estimation, followed by ESEM and then SEM. (3) SEM was the least powerful in the significance tests; statistical power was lower for ESEM than MRA with relatively small target-factor loadings, but higher for ESEM than MRA with relatively large target-factor loadings. (4) ESEM showed difficulties in convergence and occasionally created an inflated type I error rate under some conditions. ESEM is recommended when non-ignorable cross-factor loadings exist.

Keywords: exploratory structural equation modeling, latent variables, Monte Carlo simulation, multiple regression

Comparing Exploratory Structural Equation Modeling and Existing Approaches for Multiple Regression with Latent Variables

Introduction

Multiple regression is an essential methodological tool in modern social science, especially in psychological and educational research (Keith, 2014). For multiple regression with latent variables, there are two common modeling approaches. One is structural equation modeling (SEM) that typically assumes the latent variables have concise factor structures without cross-loadings, evaluating the measurement model by confirmatory factor analysis (CFA). The other is manifest regression analysis (MRA) that treats the latent variables as observed variables, usually scoring each latent variable with mean (or sum) item scores (e.g., Coffman & MacCallum, 2005; Stephenson & Holbert, 2003) or factor scores (e.g., Lu, Kwan, Thomas, & Cedzynski, 2011). Theoretically, SEM is preferred to MRA for analyzing latent variables given an adequate sample size because SEM allows the correction for the measurement errors, while observed variable approaches ignore the potential measurement errors (Bollen, 1989; P. Cohen, Cohen, Teresi, Marchi, & Velez, 1990; Rigdon, 1994). In practice, both SEM and MRA have strengths and weaknesses.

With respect to estimation accuracy for regression coefficients, the SEM approach in general outperforms MRA. Simulation studies (Coffman & MacCallum, 2005; Skrondal & Laake, 2001; Stephenson & Holbert, 2003) demonstrated that MRA underestimated the coefficients due to ignoring the measurement errors, and the underestimation became severer as the measurement reliability decreased (Ledgerwood & Shrout, 2011). Regarding estimation precision, the simulation study by Ledgerwood and Shrout (2011) showed that SEM produced larger standard errors than mean-item-score MRA. Similarly, the study by Devlieger, Mayer, and Rosseel (2016) found that SEM had larger empirical standard deviation but less biased standard errors of the coefficient estimates compared with factor-score MRA. One reason could be that the optimization for SEM involves a more complex sample covariance matrix and more parameters than MRA. To consider the trade-off between accuracy and precision, Ledgerwood and Shrout (2011) used figures to demonstrate that MRA outperformed SEM (sample size = 100) while Devlieger et al. (2016) employed mean square error to conclude that SEM worked better than most MRA approaches. For significance tests, MRA was found to have a higher power but an inflated type I error rate than SEM (Hoyle & Kenny, 1999; Ledgerwood & Shrout, 2011). In terms of model convergence and proper solutions, SEM was more likely to have problems than MRA particularly when the sample size was small (Devlieger et al., 2016; Devlieger & Rosseel, 2017; Ledgerwood & Shrout, 2011). While MRA worked well with a small sample (Devlieger et al., 2016; Ledgerwood & Shrout, 2011) SEM required a large sample (e.g., 10 cases per variable; Nunnally, 1978) to guarantee its good performance.

Since neither SEM nor MRA is satisfactory, improved approaches were introduced. For instance, the bias-correcting MRA proposed by Croon (2002) was found to have a higher standard error bias but a comparable bias, efficiency, mean square error, power, and type I error rate relative to SEM (Devlieger et al., 2016). However, the Croon method currently cannot analyze models with cross-loadings or correlated residual errors (Devlieger & Rosseel, 2017).

To better adjust for the cross-factor loadings, exploratory structural equation modeling (ESEM) was proposed as an alternative approach for latent variables analysis, which evaluates the measurement model of latent variables using exploratory factor analysis (EFA) instead of CFA (Marsh et al., 2009; Sass, 2011; Schmitt, 2011). Studies have demonstrated the impressive performance of ESEM compared to CFA in investigating the measurement structure of latent variables (Marsh, Liem, Martin, Morin, & Nagengast, 2011; Marsh et al., 2010; Marsh, Morin, Parker, & Kaur, 2014; Mattsson, 2012; Myers, Chase, Pierce, & Martin, 2011). In addition, ESEM instead of item-parcel methods was suggested to be a viable alternative to SEM for latent regression analysis when a number of cross-factor loadings exist (Marsh, Lüdtke, Nagengast, Morin, & Von Davier, 2013) because using item parcels can result in distorted relations among latent variables when the unidimensionality assumption of the items (see Little, Cunningham, Shahar, & Widaman, 2002) is violated by cross-factor loadings. Similarly, using item composite sores in MRA without adjusting for cross-factor loadings can also lead to distorted estimation.

Drawing insights from the above review, we expect that ESEM can be a viable alternative to SEM or MRA in multiple regression analysis of latent variables when there are substantive non-zero cross-factor loadings. This study aims to (a) compare ESEM with SEM and MRA in terms of estimation accuracy, estimation precision, statistical power, type I error rate, model convergence and goodness of fit, when applying to latent multiple regression and (b) provide an updated strategy for choosing modeling approaches in latent multiple regression. We will carry out a Monte Carlo simulation study to fulfill the aims. This simulation study will consider mean-item-score MRA as the representative MRA approach.

Methods

Population Model

Figure 1 portrays the latent regression model employed to generate data. The structural equation is $\eta_1 = \gamma_1\xi_1 + \gamma_2\xi_2 + \zeta_1$, where η_1 regressions on ξ_1 and ξ_2 with regression coefficients γ_1 and γ_2, respectively. The variance of each latent variable is set to be one. The endogenous latent variable η_1 has three observed indicators, y_1, y_2, and y_3. The indicators have equal factor loadings of .7, corresponding to a scale composite reliability (Raykov, 1997) of .74 for η_1. The two exogenous latent variables, ξ_1 and ξ_2, are correlated with the Pearson correlation coefficient $\rho = .3$, corresponding to a medium size of correlation (J. Cohen, 1988). They have six indicators, $x_1 \sim x_6$. Specifically, ξ_1 is the target factor of x_1, x_2, and x_3; and ξ_2 is the target factor of x_4, x_5, and x_6. The corresponding target-factor loadings for the six indicators are λ_{11}, λ_{21}, λ_{31}, and λ_{42}, λ_{52}, λ_{62}, respectively. In addition, x_1, x_2, and x_3 each has a cross-factor loading, λ_{12}, λ_{22}, and λ_{32}, respectively, on ξ_1; and x_4, x_5, and x_6 each has a cross-factor loading, λ_{41}, λ_{51}, and λ_{61}, respectively, on ξ_2. The corresponding measurement error terms for the indicators $y_1 \sim y_3$ and $x_1 \sim x_6$ are $\varepsilon_1 \sim \varepsilon_3$ and $\delta_1 \sim \delta_6$, respectively. Their variances are denoted by $\theta_{\varepsilon_1} \sim \theta_{\varepsilon_3}$ and $\theta_{\delta_1} \sim \theta_{\delta_6}$, respectively. The model has the following assumptions: (1) The expectation of each error term equals zero; (2) The covariance of any two error terms equals zero, which means the error terms are independent from each other; and (3) Any of the error terms is independent from ξ_1, ξ_2, and η_1.

Experimental Design

In the simulation we manipulated six experimental factors as follows.

1. The standardized value of target-factor loadings: $\lambda_T = .55, .7, .84$, or .95.

2. The standardized value of cross-factor loadings: $\lambda_C = 0$, .05, .10, .15, .20, or .25. Note that $\lambda_C = 0$, .05, or .10 but neither .20 nor .25 when $\lambda_T = .95$ in our study.[1] Table 1 depicts the scale composite reliability (CR; Raykov, 1997) for the single latent variable ξ_1 (also called construct reliability in a measurement model; Hair, Black, Babin, Anderson, & Tatham, 2009) corresponding to different combinations of λ_T and λ_C.

3. The standardized value of regression coefficients $\gamma_{11} = \gamma_{12} = \gamma = .14$, .36, or .51. The corresponding coefficient of determination for the overall regression model R^2 is .05, .34, and .68, respectively. Being compared with only one predictor ξ_1 in the regression model, the change in R^2 (i.e., ΔR^2) is .03, .21, and .42, respectively.

4. The sample size: $N = 100$, 200, or 500. We used 100 as the minimum sample size following the suggestion by Boomsma (1982) for latent variable models.

5. The distribution of observed variables ($y_1 \sim y_3$ and $x_1 \sim x_6$) and latent variables (η_1, ξ_1, and ξ_2) can be normal or nonnormal. We used $\chi^2(df = 6)$ as the nonnormal distribution.

6. The modeling approaches included ESEM, SEM, and MRA. We used the maximum likelihood estimator (MLE) for all three approaches. Note that MRA using the Ordinal Least Square estimator will provide the same results as using MLE in normal case.

The first five experimental factors were between-subject factors resulting in 378 experimental conditions for generating data. The last one was a within-subject factor. That is, the three modeling approaches were separately employed to fit the model with the generated data for each experimental condition.

[1]Since λ_T and λ_C are standardized factor loadings of each of the indicators $x_1 \sim x_6$ on the correlated latent variables ξ_1 and ξ_2, respectively, $\lambda_T^2 + \lambda_C^2 + 2\rho\lambda_T\lambda_C \leq 1$ is a restriction to be satisfied. Given $\rho = .3$, the restriction is not satisfied when $\lambda_T = .95$ and $\lambda_C = .20$ or .25.

Simulation and Comparison Procedure

The procedure included three steps. First, given the predefined parameter values, we generated 1000 replicates of sample data for each experimental condition. Second, we fit ESEM, SEM, and MAR separately with each replicate of the generated data. Third, we compared the results from ESEM, SEM, and MRA using the criteria stated in the Results section. R program (R 2.1.5) was used for generating data and comparing the results and Mplus 7.0 (Muthén & Muthén, 1998-2012) was used for model estimation.

In data generation, we employed a sequential approach. For each experimental condition, we first generated random data of the exogenous latent variables ξ_1 and ξ_2 given ρ, as well as the error term ζ_1 in the structural equation. We then generated random data of the endogenous latent variable η_1 based on the structural equation given $\gamma_{11} = \gamma_{12} = \gamma$. Using the generated data of the latent variables and given factor loadings, we generated random data of the observed variables. We finally divided the data of the first indicator of each latent variable by its factor loading. For example, $\tilde{y}_1 = y_1/0.7$ (see Marsh, Hau, & Wen, 2004). By doing this, SEM or ESEM would result in the same scale of a latent variable by fixing the loading of its first indicator at one or by fixing the variance of the latent variable at one. Thus the estimated regression coefficients from both ways can be compared directly.

Results

We employed six criteria for comparisons: the rate of model convergence and convergence with proper solutions, goodness of fit, estimation accuracy, estimation precision, statistical power, and type I error rate. The last five criteria were calculated using only the replications that were convergent with proper solutions. In our results, we focused on the estimate of regression coefficient γ_{11}, since the population model is symmetrical with respect to the two predictors and the regression coefficients γ_{11} and γ_{12}.

Model Convergence and Convergence with Proper Solutions

Model-convergence rate (MCR) was calculated as the number of convergent replications divided by the total number of replications under each experimental condition. Convergence-with-proper-solution rate (CPSR) was calculated as the number of convergent replications having proper solutions divided by the total number of replications.

The results for both normal and nonnormal data showed that ESEM had problems with MCR and CPSR when target-factor loadings were small ($\lambda_T = .55$) and sample size was not large ($N = 100$, 200). The situation became worse as cross-factor loadings increased, with minimum MCR of 67%/65% and minimum CPSR of 49%/45%, for normal/nonnormal data, respectively. SEM also had problems (but less severe than ESEM) with CPSR when having small target-factor loadings and small sample size combined with very large cross-factor loadings($\lambda_C = .25$). It had a minimum CPSR of 81%.

Goodness of Fit

The indices of goodness of fit used in the study included χ^2/df, root mean square error of approximation (RMSEA), comparative fit index (CFI), standardized root mean square residual (SRMR), Akaike's information criterion (AIC), Bayesian information criterion (BIC), and T-size fit indexes RMSEA_t and CFI_t (Marcoulides & Yuan, 2017; Yuan, Chan, Marcoulides, & Bentler, 2016). For χ^2/df, RMSEA, RMSEA_t, SRMR, AIC, and BIC, the smaller is considered the better; while for CFI and CFI_t the larger is considered the better. Following the decision rules of Hu and Bentler (1999; also see

Marsh et al., 2004), model fitness is considered to be good when $\chi^2/df < 5$, CFI $> .95$, RMSEA $< .06$, and SRMR $< .08$. AIC and BIC are usually used to compare models either nested or not. The model with the smaller AIC and BIC is considered the better. In our case given sample size $N = 200$, RMSEA$_t < .085$ and CFI$_t > .890$ are considered to be good for SEM; while for ESEM the corresponding criteria are RMSEA$_t < .082$ and CFI$_t > .891$ (see calculation in Yuan et al., 2016). Note that we mainly compared ESEM and SEM in goodness of fit as MRA always has the smallest AIC/BIC and perfect values of other goodness-of-fit indices.

When the data were normally distributed, both ESEM and SEM performed well in terms of goodness of fit. ESEM consistently had smaller SRMR than SEM. Compared with SEM, ESEM in general had slightly larger values in CFI/CFI$_t$ and smaller values in RMSEA /RMSEA$_t$, but larger values in AIC/BIC. The differences between ESEM and SEM were more apparent as the cross-factor loadings became larger. Table 2 showed the detailed results for normal data with $\gamma = .14$, $\lambda_T = .55$, and $N = 200$. Similar results were observed for nonnormally distributed data (see the supplementary materials).

Estimation Accuracy and Precision

To quantify the estimation accuracy, we used the relative bias of estimation. To evaluate the estimation precision, we used the standard deviation of the estimates. The relative bias for standard error was also presented. To consider the trade-off between estimation accuracy and precision, we employed the mean square error.

Relative bias of estimation. Relative bias was defined as the ratio of bias to the population value, where the bias was calculated by subtracting the population value from the average estimate across replications under each experimental condition.

Relative bias of estimation (RBEST) larger than zero implies overestimation while RBEST less than zero indicates underestimation. According to the recommendations of Hoogland and Boomsma (1998), RBEST is considered acceptable if its absolute value is smaller than .05.

For normal data, in the case of zero cross-factor loadings, the median RBEST is .048, .007, and $-.083$, with the range $[-.011, .081]$, $[-.035, .094]$, and $[-.263, -.027]$ for ESEM, SEM, and MRA, respectively. With respect to the absolute value of RBEST, SEM $<$ ESEM $<$ MRA in general. ESEM and SEM both had acceptable RBEST under most conditions while MRA systematically underestimated the regression coefficient. In the case of non-zero cross-factor loadings, the median RBEST is $-.021$, $-.146$, and $-.151$, with the range $[-.104, 0.058]$, $[-.318, -.016]$, and $[-.232, -.056]$ for ESEM, SEM, and MRA, respectively. Acceptable RBEST was observed for ESEM under most conditions except for very large cross-factor loadings ($\lambda_C = .25$). SEM and MRA both systematically underestimated the regression coefficient. In both cases, as target-factor loadings became larger RBEST became less severe for MRA, while it did not apparently change for ESEM or SEM. As cross-factor loadings became larger the absolute RBEST became larger for SEM, while it did not apparently change for MRA or ESEM. Figure 2 portrays the comparison in RBEST when $\gamma = .14$ and data were normally distributed. Similar patterns were observed for nonnormal data (see the supplementary materials).

Standard deviation of estimates. Standard deviation of the estimates (SD) was the standard deviation of the estimates under each experimental condition, which was the empirical standard error and treated as the population value of standard error in the study. For each experimental condition, the smaller SD indicates the more precise estimation.

Figure 3 depicts the comparison in SD when $\gamma = .14$ and data were normally distributed. In the case of zero cross-factor loadings, the median SD is .083, .088, and .072, with the range [.046, .148], [.048, .199], and [.044, .101] for ESEM, SEM, and MRA, respectively. In the case of non-zero cross-factor loadings, the median SD is .100, .113, and .077, with the range [.045, .245], [.046, .300], and [.045, .137] for ESEM, SEM, and MRA, respectively. In both cases, SD had the pattern MRA < ESEM < SEM in general. SD decreased for all three approaches as sample size increased. It became smaller for ESEM and SEM as target-factor loadings became larger. There were pronounced increases in SD associated with larger cross-factor loadings for ESEM and SEM, particularly when target-factor loadings were small ($\lambda_T = .55$); while the associated increases for MRA were not sizable. The differences in SD among the three approaches shrank toward zero as sample size and target-factor loadings became larger. For other population values of regression coefficient ($\gamma = .36, .51$) or for nonnormal data the results showed the similar patterns (see the supplementary materials).

Relative bias for standard error. To calculate the relative bias for standard error (RBSE), SD was treated as the population value of standard error. RBSE larger than zero implies overestimation while RBSE smaller than zero indicates underestimation. RBSE is considered acceptable if its absolute value is smaller than .10 (Hoogland & Boomsma, 1998).

For normal data and zero cross-factor loadings, the median RBSE is .002, −.015, and −.002, with the range [−.042, .044], [−.079, .049], and [−.026, .054] for ESEM, SEM, and MRA, respectively. RBSE was acceptable for all three approaches. Given non-zero cross-factor loadings, the median RBSE is −.007, −.001, and −.001, with the range [−.176, .077], [−.073, .380], and [−.065, .056] for ESEM, SEM, and MRA, respectively. RBSE was acceptable for MRA across all conditions and for ESEM and SEM under most conditions. When sample size was not large ($N = 100$, or 200), target-factor loadings were small, and cross-factor loadings were larger than .10, RBSE was not acceptable for ESEM or SEM. Specifically, ESEM underestimated the standard error while SEM overestimated the standard error. Figure 4 presents the comparison in RBSE given $\gamma = .14$ and normal data. Similar patterns were observed for nonnormal data (see the supplementary materials).

Mean square error. Mean square error was calculated as the average of the squared deviations, where the deviations were calculated by subtracting the population value from the parameter estimates under each experimental condition. The smaller MSE is considered the better.

Figure 5 portrays the results given $\gamma = .14$ and normal data. With zero cross-factor loadings, the median MSE is .007, .008, and .005, with the range [.002, .022], [.002, .039], and [.002, .011] for ESEM, SEM, and MRA, respectively. With non-zero cross-factor loadings, the median MSE is .010, .013, and .006, with the range [.002, .060], [.002, .091], and [.002, .019] for ESEM, SEM, and MRA, respectively. In general, MSE had the pattern MRA < ESEM < SEM. The patterns for MSE were similar to those for SD.

Statistical Power

Two-sided Z-tests along with $\alpha = .05$ were employed to test for non-zero regression coefficients. The test statistic was calculated as the estimate divided by its standard error for each replication under an experimental condition. The statistical power was calculated as the number of significant results divided by the number of replications.

Figure 6 depicts the comparison in statistical power when $\gamma = .14$ and data were normally distributed. In the case of zero cross-factor loadings, the median statistical power is .449, .379, and .442, with the range [.180, .879], [.089, .822], and [.204, .862] for ESEM, SEM, and MRA, respectively. In the case of non-zero cross-factor loadings, the median statistical power is .328, .222, and .343, with the range [.147, .895], [.023, .804], and [.152, .806] for ESEM, SEM, and MRA, respectively. In both cases, SEM had the lowest statistical power across conditions among the three approaches. Statistical power was higher for MRA than ESEM when target-factor loadings had small or median size (λ_T = .55, .7). It was higher for ESEM than MRA when target-factor loadings were large or very large (λ_T = .84, .95). For all three approaches, higher statistical power was associated with larger target-factor loadings, larger sample size, and smaller cross-factor loadings.

Statistical power was higher for larger regression coefficients and lower for nonnormal data than normal data. Detailed results on statistical power can be found in the supplementary materials.

Type I Error Rate

Type I error rate was calculated in the same way as statistical power but under the conditions with zero regression coefficient in the population ($\gamma = 0$). The acceptable range is [.025, .075] when $\alpha = .05$ (MacKinnon, Lockwood, & Williams, 2004; Williams & MacKinnon, 2008).

Figure 7 presents the results with normal data. Under most conditions, type I error rate was acceptable for all approaches. ESEM occasionally resulted in an inflated type I error rate under conditions with small target-factor loadings, non-zero cross-factor loadings, and not larger sample size. SEM produced a type I error rate lower than the acceptable lower band under conditions with small target-factor loadings, cross-factor loadings larger than .15, and not large sample size. Similar patterns of type I error rate were observed for nonnormal data.

Discussion

Summary

Regarding estimation accuracy, ESEM and SEM both performed well but MRA resulted in considerable underestimation when having zero cross-factor loadings. ESEM provided the least biased estimation while SEM and MRA resulted in systematical underestimation when having non-zero cross-factor loadings. SEM could be worse than MRA with relatively large cross-factor loadings (e.g., $\lambda_C \geq .10$). With respect to estimation precision, MRA produced the smallest standard deviation (SD) of the coefficient estimates, followed by ESEM and then SEM. The disparities in SD among the three approaches became smaller as sample size and target-factor loadings increased. All three approaches had an acceptable relative bias for standard error estimation under most conditions. Based on the trade-off between estimation accuracy and precision, MRA had the smallest mean square error (MSE), followed by ESEM and then SEM. The differences became much smaller as sample size and target-factor loadings became larger.

In terms of significance tests, SEM had the lowest statistical power across conditions. MRA was more powerful than ESEM under conditions with relatively small target-factor loadings (e.g., $\lambda_T = .55, .7$). Beyond our expectations, ESEM was slightly more powerful than MRA under conditions with relatively large target-factor loadings (e.g., $\lambda_T = .84, .95$). Overall, MRA was acceptable with respect to type I error rate. ESEM occasionally resulted in unacceptable inflations of type I error rate while SEM created flattened values under conditions with small target-factor loadings and small sample size.

Both ESEM and SEM had problems with model convergence and proper solutions when target-factor loadings were small and sample size was not large (e.g., $N < 500$).

Strategies for Choosing a Method

Taking all the criteria into consideration, we suggested the following strategies to choose among the three approaches when the sample size is no less than the minimum requirement for ESEM or SEM.

1. When cross-factor loadings are not ignorable (e.g., $\lambda_C \geq .10$), ESEM should be used for both estimation and significance test. Since ESEM involves relatively more complex models, large sample size is encouraged to avoid problems of non-convergence and inflated type I error rates, especially for the situation of small target-factor loadings and large cross-factor loadings.

2. When cross-factor loadings are close to zero (e.g., $\lambda_C < .10$) and the target-factor loadings are very large (e.g., $\lambda_T \geq .84$), corresponding to the situation of very high composite reliability, MRA is recommended for both estimation and significance test.

3. When cross-factor loadings are close to zero (e.g., $\lambda_C < .10$) and the target-factor loadings are not large (e.g., $\lambda_T < .84$), SEM is preferred for estimation but MRA is recommended for significance test if SEM fails to obtain significant results.

Unanticipated Findings and Implications

In addition to the anticipated findings such as the underestimation of the coefficients by MRA (also found by Coffman & MacCallum, 2005; Ledgerwood & Shrout, 2011; Skrondal & Laake, 2001; Stephenson & Holbert, 2003), the smallest SD of MRA (consistent with the studies by Devlieger et al., 2016; Ledgerwood & Shrout, 2011), and a lower statistical power for SEM than MRA (also shown by Hoyle & Kenny, 1999; Ledgerwood & Shrout, 2011), there were some findings beyond our expectations as discussed below.

First, ESEM had the highest statistical power instead of MRA under conditions with relatively large target-factor loadings, which was a surprise. As claimed by Ledgerwood and Shrout (2011), MRA was more powerful than SEM because the latent-variable approach tended to produce larger standard error (the estimate of SD) and MRA had slighter underestimation of the coefficients given higher reliability (because of larger target-factor loadings). Different from the previous study, we found that the disparities in SD among the three approaches no longer existed when sample size or target-factor loadings were large. Under these conditions, ESEM produced even smaller SD than MRA given non-ignorable cross-factor loadings. Furthermore, the results of MSE showed the similar pattern with SD. Thus, the statistical power for ESEM could be higher than MRA. These findings also suggested that large sample size may facilitate the good performance of ESEM.

Second, ESEM slightly overestimated the regression coefficients for population models without cross-factor loadings. This indicated that ESEM with the inclusion of all cross-factor loadings could be overly complex and contradict to the principle of parsimony when non-zero cross-factor loadings do not exist.

Third, when compared with SEM, MRA led to less biased standard error estimates under most conditions and smaller MSE regardless of the existence of non-zero cross-factor loadings. These results differed from those in the study by Devlieger et al. (2016) in which the focus was factor-score MRA but not mean-item-score MRA. The differences suggested that MRA approaches using various types of scores should be evaluated separately. Future studies could include factor-score MRA in comparisons.

Fourth, MRA provided less biased estimation of the coefficients than SEM as cross-factor loadings became larger. In fact, we have expected MRA to have more biased estimation compared with SEM, since MRA generally does not correct for the measurement errors (e.g., Bollen, 1989; Ledgerwood & Shrout, 2011) or non-zero cross-factor loadings. This unexpected result indicated that the distortion of structural relations by MRA could be more complex than commonly believed. This study only considered a balanced design with symmetric regression relations and factor loadings, which might be the reason for that the distortion effects by MRA were not revealed. Future studies should evaluate unbalanced designs to address this issue.

References

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

Boomsma, A. (1982). The robustness of lisrel against small sample sizes in factor analysis models. *Systems under Indirect Observation: Causality, Structure, Prediction*, *1*, 149–173.

Coffman, D. L., & MacCallum, R. C. (2005). Using parcels to convert path analysis models into latent variable models. *Multivariate Behavioral Research*, *40*(2), 235–259.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, P., Cohen, J., Teresi, J., Marchi, M., & Velez, C. N. (1990). Problems in the measurement of latent variables in structural equations causal models. *Applied Psychological Measurement*, *14*(2), 183–196.

Croon, M. (2002). Latent variable and latent structure modeling. In G. Marcoulides & I. Moustaki (Eds.), (pp. 195–223). Mahwah, NJ: Lawrence Erlbaum.

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement*, *76*(5), 741–770.

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM? *European Journal of Research Methods for the Behavioral & Social Sciences*, *13*(Suppl 1), 31–38. doi: 10.1027/a000001

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Multivariate data analysis: A global perspective* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling an overview and a meta-analysis. *Sociological Methods & Research*, *26*(3), 329–367.

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. *Statistical Strategies for Small Sample Research*, *1*, 195–222.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). New York, NY: Routledge.

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology*, *101*(6), 1174–1188.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 151–173.

Lu, I. R., Kwan, E., Thomas, D. R., & Cedzynski, M. (2011). Two new methods for estimating structural equation models: An illustration and a comparison with two established methods. *International Journal of Research in Marketing*, *28*(3), 258–268.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128.

Marcoulides, K. M., & Yuan, K.-H. (2017). New ways to evaluate goodness of fit: a note on using equivalence testing to assess structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(1), 148-153. doi: 10.1080/10705511.2016.1225260

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing hu and bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341.

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment*, *29*(4), 322–346.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*(3), 471–491.

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J., & Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right-amouflaging misspecification with item parcels in cfa models. *Psychological Methods*, *18*(3), 257–284.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85–110.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., &

Trautwein, U. (2009). Exploratory structural equation modeling, integrating cfa and efa: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 439–476.

Mattsson, M. (2012). Investigating the factorial invariance of the 28-item dbq across genders and age groups: An exploratory structural equation modeling study. *Accident Analysis & Prevention*, *48*, 379–396.

Muthén, L., & Muthén, B. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA, USA: Muthén & Muthén.

Myers, N. D., Chase, M. A., Pierce, S. W., & Martin, E. (2011). Coaching efficacy and exploratory structural equation modeling: A substantive-methodological synergy. *Journal of Sport and Exercise Psychology*, *33*(6), 779–806.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill Book Company.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173–184.

Rigdon, E. E. (1994). Demonstrating the effects of unmodeled random measurement error. *Structural Equation Modeling: A Multidisciplinary Journal*, *1*(4), 375–380.

Sass, D. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, *29*(4), 347–363.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304–321.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*(4), 563-575.

Stephenson, M. T., & Holbert, R. L. (2003). A monte carlo simulation of observable versus latent variable structural equation modeling techniques. *Communication Research*, *30*(3), 332–354.

Williams, J., & MacKinnon, D. P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(1), 23–51.

Yuan, K.-H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing structural equation models by equivalence testing with adjusted fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(3), 319-330. doi: 10.1080/10705511.2015.1065414

Table 1
Composite Reliability for A Single Latent Variable (e.g., ξ_1) in the Two-Factor Model

λ_T	λ_C					
	0	.05	.10	.15	.20	.25
.55	.565	.585	.603	.622	.640	.658
.7	.742	.758	.774	.789	.804	.818
.84	.878	.891	.903	.916	927	.939
.95	.965	.976	.986	–	–	–

Note. λ_T =Target-factor loading. λ_C = Cross-factor loading.

Table 2

Model-Convergence Rate, Convergence-with-Proper-Solution Rate, and Goodness of Fit under Normal Distribution When $\gamma = .14$, λ_L*=.55, and* $N = 200$

Conditions		Convergence		Goodness of Fit							
λ_C	Method	MCR	CPSR	χ^2/DF	CFI	CFI$_t$	RMSEA	RMSEA$_t$	SRMR	AIC	BIC
.00	ESEM	.99	.97	1.00	.997	.976	.015	.056	.029	5460	5572
	SEM	1.00	1.00	1.02	.997	.973	.016	.054	.035	5456	5555
	MRA	1.00	1.00	.00	1.000	1.000	.000	.000	.000	1704	1733
.05	ESEM	.98	.95	.99	.997	.977	.015	.055	.028	5435	5548
	SEM	1.00	1.00	1.01	.997	.974	.016	.054	.034	5432	5531
	MRA	1.00	1.00	.00	1.000	1.000	.000	.000	.000	1695	1725
.10	ESEM	.96	.92	.99	.997	.978	.015	.056	.028	5400	5512
	SEM	1.00	1.00	1.01	.997	.975	.016	.054	.033	5397	5496
	MRA	1.00	1.00	.00	1.000	1.000	.000	.000	.000	1683	1713
.15	ESEM	.95	.85	.97	.998	.980	.013	.054	.027	5361	5473
	SEM	1.00	1.00	1.03	.997	.976	.016	.055	.032	5358	5457
	MRA	1.00	1.00	.00	1.000	1.000	.000	.000	.000	1668	1697
.20	ESEM	.89	.79	.96	.998	.981	.013	.053	.025	5312	5424
	SEM	1.00	.99	1.03	.997	.977	.016	.055	.031	5308	5406
	MRA	1.00	1.00	.00	1.000	1.000	.000	.000	.000	1648	1678
.25	ESEM	.80	.67	.95	.998	.983	.012	.053	.023	5247	5359
	SEM	1.00	.95	1.03	.997	.979	.016	.055	.029	5244	5343
	MRA	1.00	1.00	.00	1.000	1.000	.000	.000	.000	1623	1653

Note. N = Sample size. γ = Regression coefficient. λ_T = Target-factor loading. λ_C = Cross-factor loading. MCR = Model-convergence rate. CPSR = Convergence-with-proper-solution rate. CFI = Comparative fit index. CFI$_t$ = T-size CFI. RMSEA = Root mean square error of approximation. RMSEA$_t$ = T-size RMSEA. SRMR = Standardized root mean square residual. AIC = Akaike's Information Criterion. BIC = Bayesian information criterion. ESEM = Exploratory structural equation modeling. SEM = Structural equation modeling.

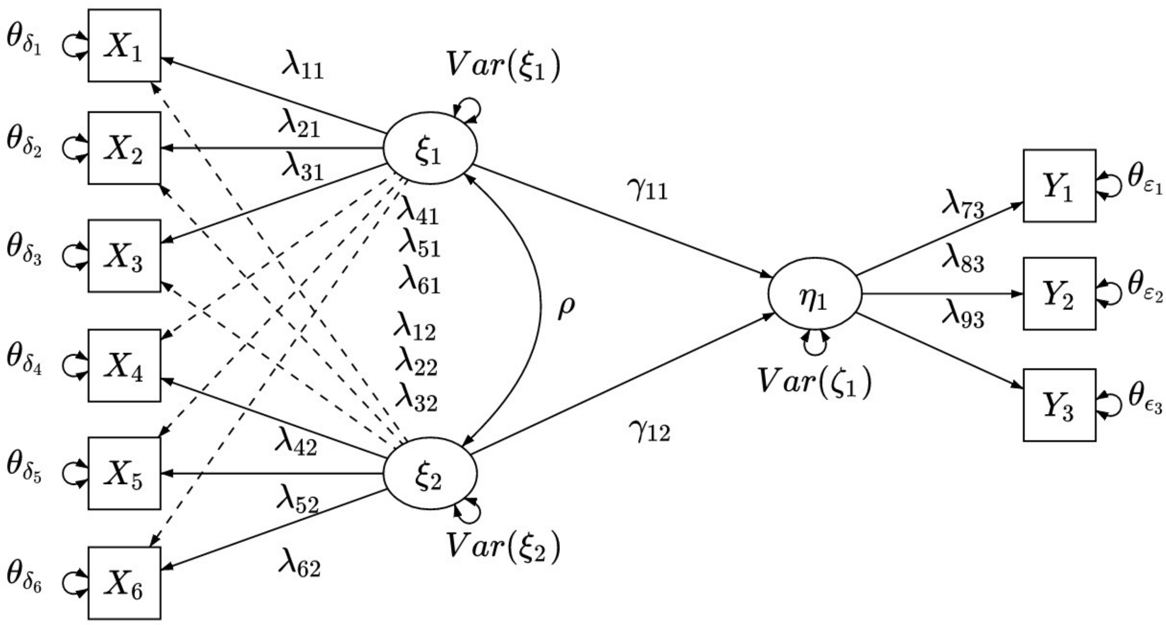

Figure 1. The Population Model.

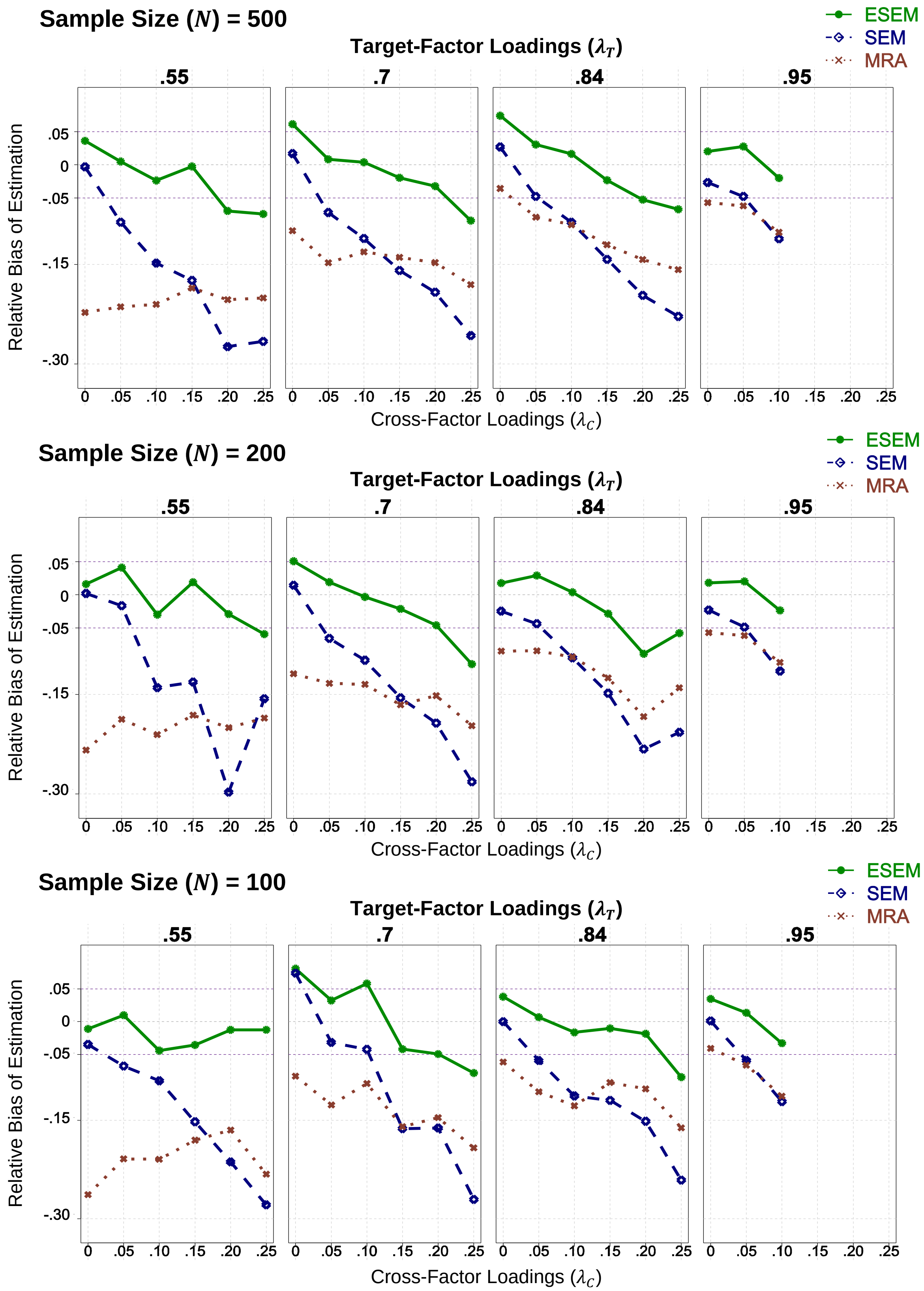

Figure 2. Relative Bias of Estimation with $\gamma = .14$ and Normal Data.

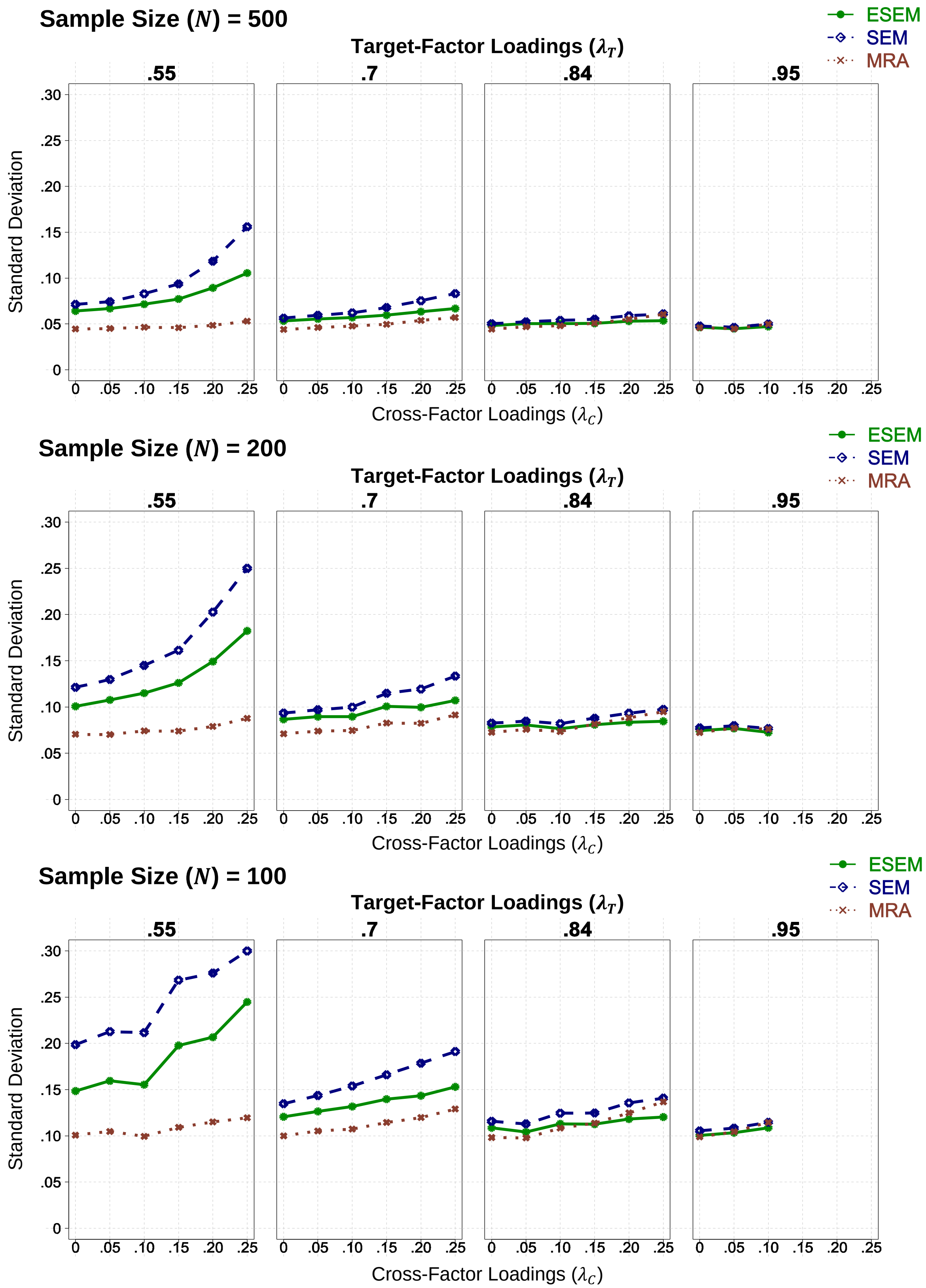

Figure 3. Standard Deviation of the Estimates for $\gamma = .14$ and Normal Data.

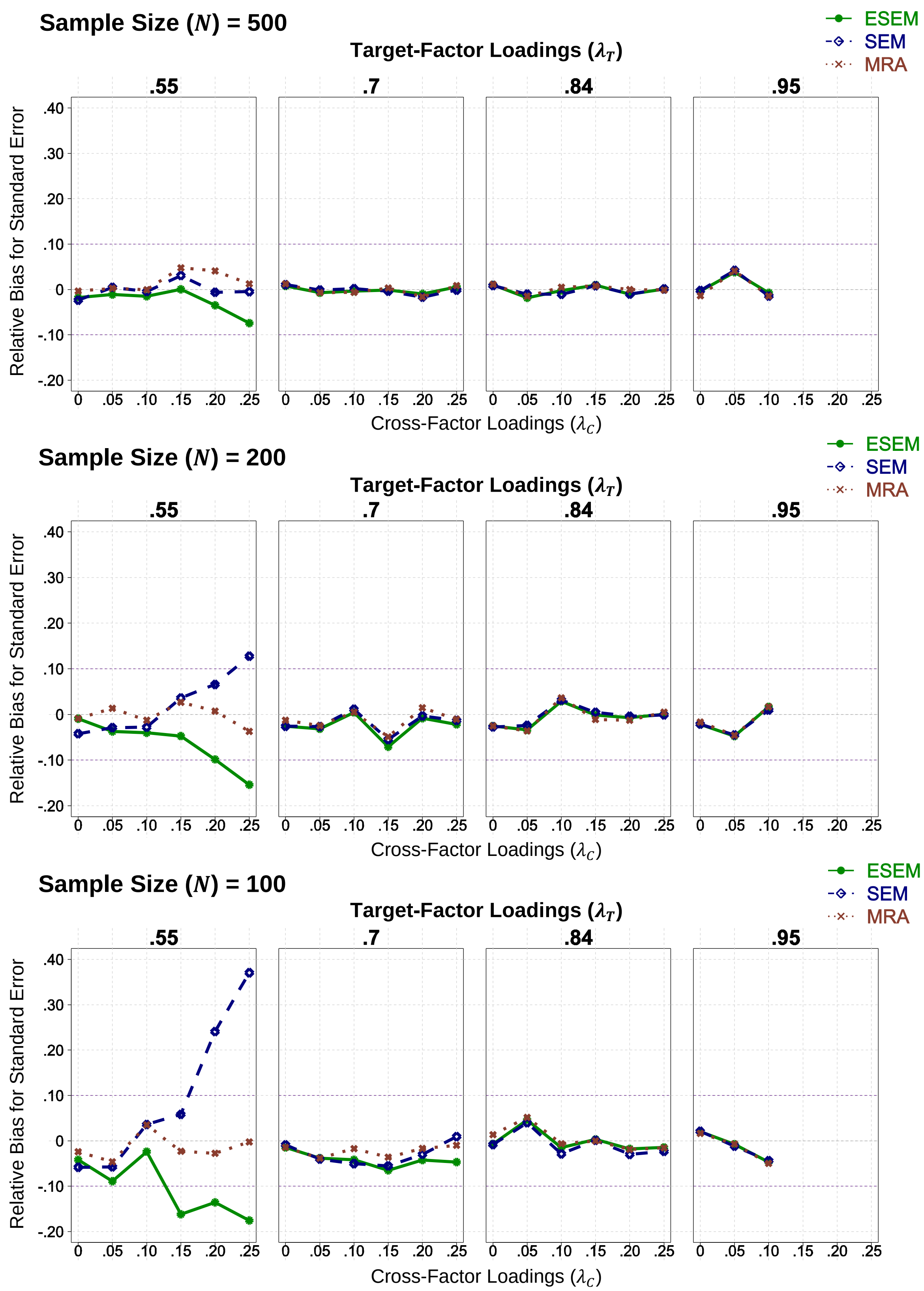

Figure 4. Relative Bias for Standard Error with $\gamma = .14$ and Normal Data.

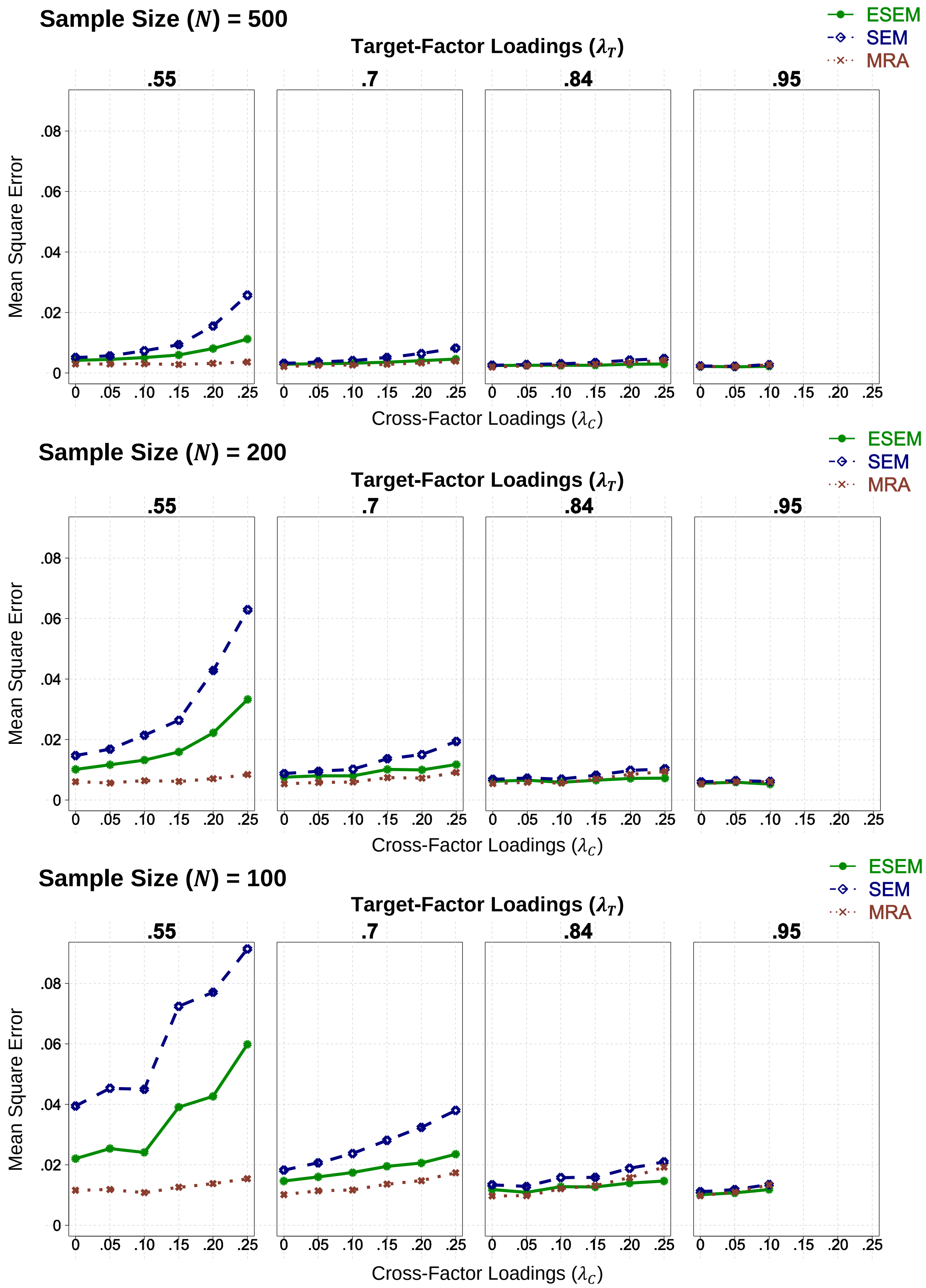

Figure 5. Mean Square Error for Estimation with $\gamma = .14$ and Normal Data.

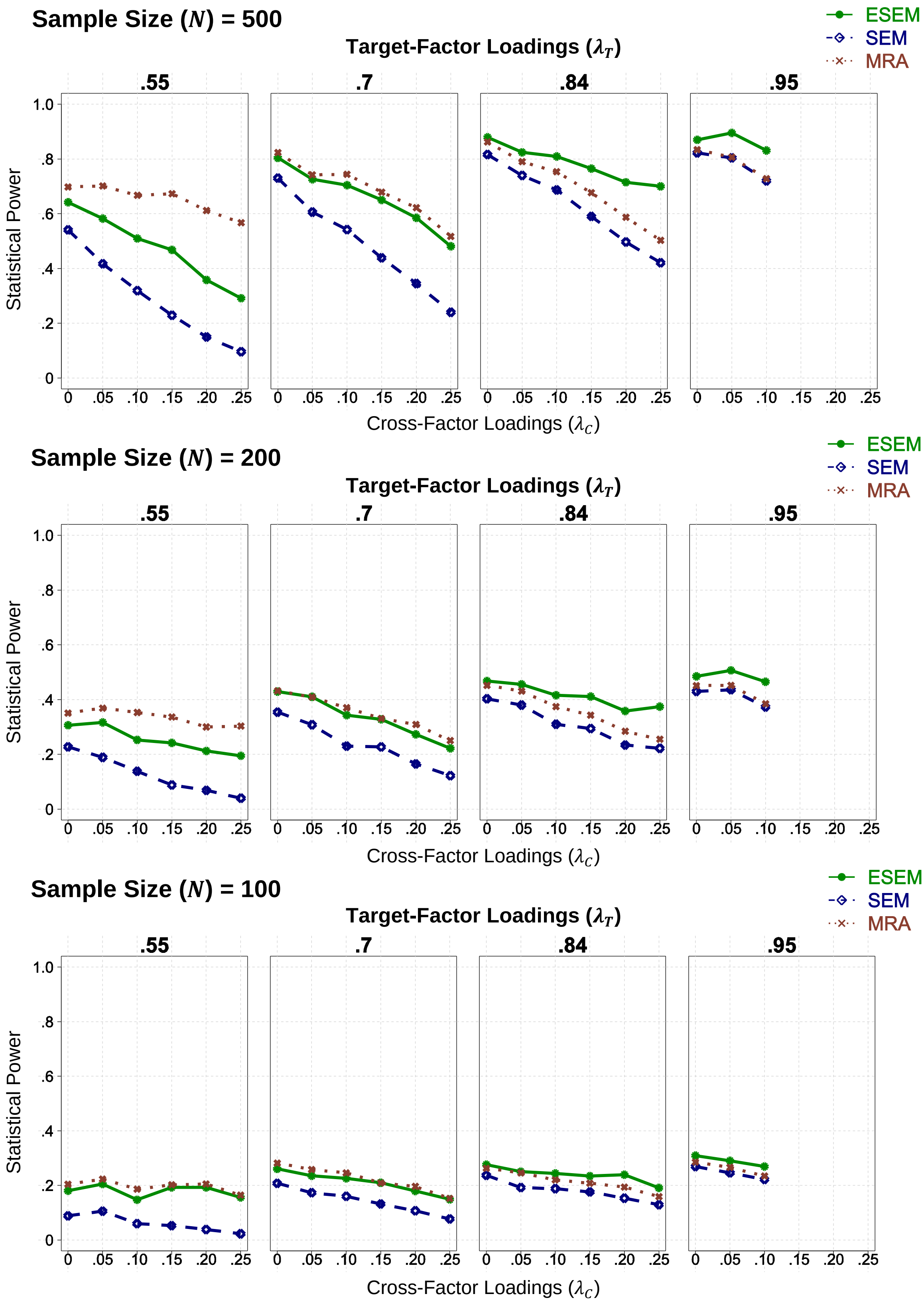

Figure 6. Statistical Power with $\gamma = .14$ and Normal Data.

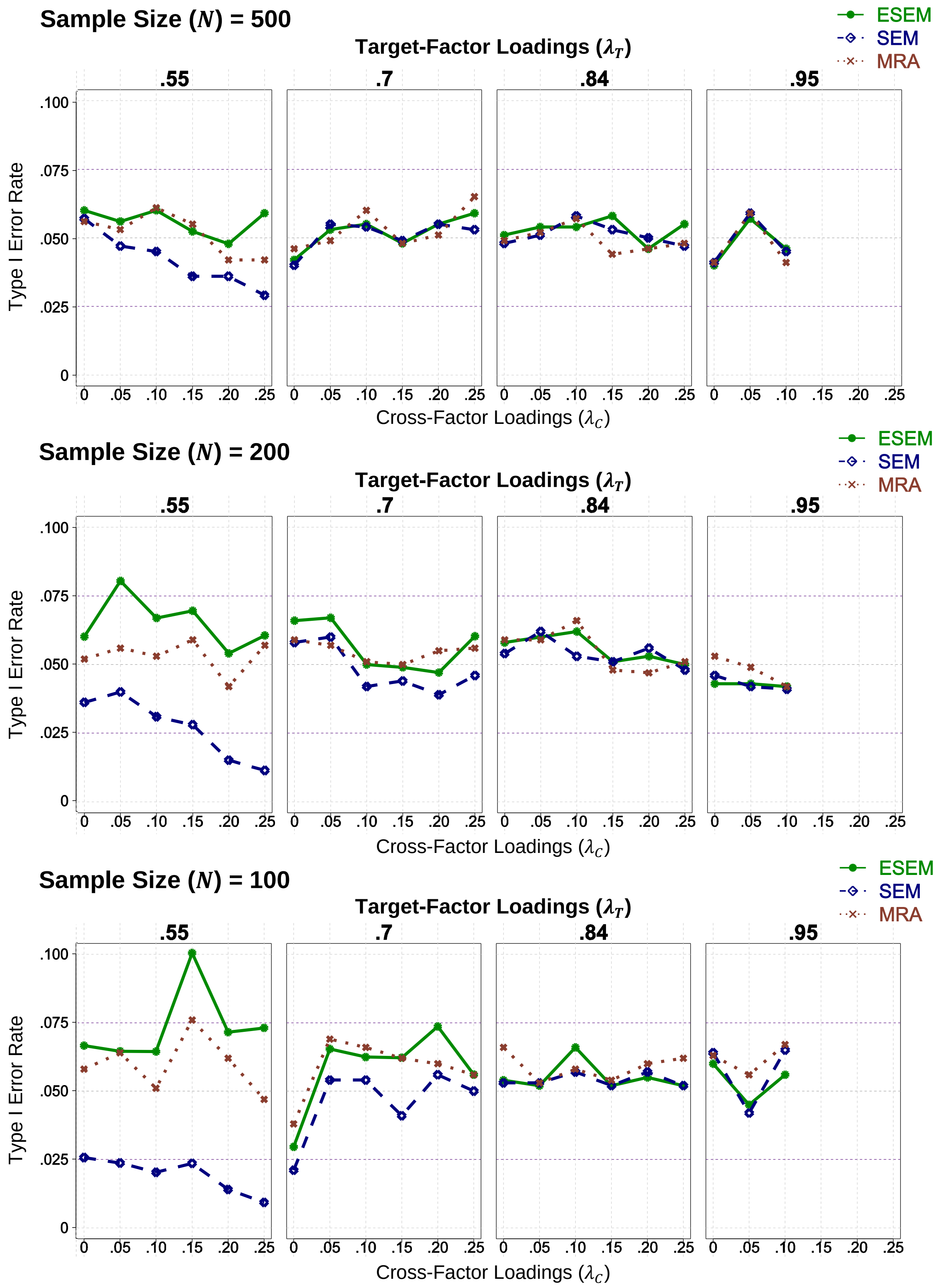

Figure 7. Type I Error Rate with Normal Data.