## Bias and efficiency of estimation in multilevel confirmatory factor models

R. Noah Padgett & Grant B. Morgan Baylor University

## Author Note

R. Noah Padgett, Department of Educational Psychology, Baylor University; Grant

B. Morgan, Department of Educational Psychology, Baylor University.

Correspondence concerning this article should be addressed to R. Noah Padgett, Department of Educational Psychology, One Bear Place # 97304, Baylor University, Waco, TX 76798. Contact: [noah\_padgett1@baylor.edu](mailto:noah_padgett1@baylor.edu)

## Abstract

Multilevel measurement models are more frequently applied to help answer questions when data arise from hierarchically structured multivariate data. We evaluated the efficiency of estimating model parameter among three estimation methods (MLR, ULSMV, WLSMV) for multilevel factor models. For nearly all sample size conditions, WLSMV resulted in more efficient parameter estimates, but MLR resulted in more efficient standard errors. Additionally, we found that all three estimation methods may overestimate the level-2 factor covariance matrix, but, on average, WLSMV resulted in the least amount of bias in parameter estimates. Our talk will include additional nuances of how parameters and standard errors are influenced by sample size conditions.

*Keywords:* ML-CFA, multilevel, categorical, efficiency, WLSMV

## Bias and efficiency of estimation in multilevel confirmatory factor models Objections and Purposes

One type of model utilized in social science that can directly account for the

complex dependencies fall under multilevel confirmatory factor analysis (ML-CFA, Goldstein & McDonald, [1988;](#_bookmark7) B. Muthén, [1994).](#_bookmark13) ML-CFA allows researchers to model hypotheses of why groups of participants vary on constructs of interest. For example, some hypotheses can be tested as to why teachers’ perceived school safety climate vary across schools and what school and teacher characteristics relate to these perceptions. In construct validation with multinational, cross-cultural instruments, ML-CFA can provides researchers with a unique lens to investigate their data with respect to how constructs potentially are influenced by group membership. The association of group membership and the latent construct can be investigating with ML-CFA when the factor structure is hypothesized to be invariant across levels (Stapleton et al., [2016,](#_bookmark15) Models 4-5, pg. 488).

The models are hypothesized such that group level constructs are aggregates of individual level construct. However, the hierarchical organization and coarse measurements associated with these data was difficult to appropriately account for in the model estimation.

However, limited information is available about the estimation performance of ML-CFA models, particularly under conditions when data are ordered categorical. Given

that data from educational settings typically arise from hierarchically structured contexts and use ordered categorical response formats, understanding how these models can be recovered in a variety of settings. Specifically, our research questions are the following:

1. Which estimation method (WLSMV, ULSMV, MLR) provides the least biased estimates of model parameters and standard errors (factor loadings, factor variances, factor covariances at different levels)?
2. Which estimation method is more efficient, especially when the sample size is relatively low?

We expect that WLSMV will yield the least biased and most efficient estimates across all parameters (Asparouhov & Muthén, [2007;](#_bookmark2) DiStefano & Morgan, [2014;](#_bookmark5) Hox et al., [2010).](#_bookmark11) But, when the number of level-2 units is below 100, some evidence points to sever bias in WLSMV estimates. However, some evidence suggests that ULS may be just as good as WLSMV for categorical data for single-level models (Forero et al., [2009),](#_bookmark6) we are hopeful that ULS parameter and standard errors estimates for ML-CFA will be comparable to that of WLSMV.

## Perspective, Theoretical Framework, and Background

ML-CFA is a decomposition of the observed covariance matrix into a level-1 (pooled within group) and level-2 (between group) specific covariance matrix. Because essentially two covariance matrices are being analyzed, two possibly unique models can be specified for each level of analysis. The two levels of analysis correspond to the individual and group level. With categorical observed data, these two effects directly influence the underlying response value for each item, as shown in Equation [1.](#_bookmark0) The underlying response value for an

individual (*y*\* ) is the composition of the individual component (*ywpig*) and a random effect

*pig*

of group (*y**bpg*). As with B. Muthén, [1984,](#_bookmark12) the latent response is assumed normally distributed.

\*

*y*

*pig*

= *ybpg* + *ywpig* (1)

where *y*\* is the latent response underlying the observed categorical value *yp* for the *pth*

*pig*

item, *i* indexes across individuals within group *g*. Across *p* items, this general framework is easily incorporated into factor analysis by replacing the latent components *ywpig* and *ybpg* by the level specific factor loading matrices and latent variable vectors.

In models where the observed scores are categorical, as in this case, the item thresholds are used to relate the underlying response values (**y**\* ) to the observed vector of categorical responses (**y***ig*). These thresholds (*τpk*) are specific to each item (*p*) and are constant across groups. The total number of categories in the observed variable is defined

**ig**

as *k* = 2*,* 3*,* · · · *, K*; *ypig* = *k* ¢'? *τpk* < *y*\* < *τp*(*k*+1) where *τp*0 = -cx and *τpk*+1 = cx. The

*pig*

relationship between the observed categories and underlying response vector allows for the categorical nature of the observed data to be controlled for when estimating the

between-group variability.

The relative amount of between-group variability is expressed in two components of the model. First, amount of variability in each item that attributed to group differences.

And second, the amount of variability in the latent variable that is attributed to group differences. Both quantities can be estimated using an intraclass correlation (ICC). In ML-CFA with invariant factor structure and constrained loadings across levels, the ICC now represents the proportion of variability in the latent construct between groups (Heck & Thomas, [2015,](#_bookmark9) p. 155-171). This may provide key inferential outputs in applications of ML-CFA.

## Methods, techniques, or modes of inquiry

### *Model Estimation Methods*

All models were estimated using M*plus* (L. Muthén & Muthén, [2017,](#_bookmark14) version 8.2).

In this project, we focused on ordered categorical responses. We compared across three robust estimation methods: maximum likelihood with robust standard errors (MLR), unweighted least squares mean and variance adjusted (ULSMV), and weighted least squares mean and variance adjusted (WLSMV).

## Analysis of Monte Carlo Simulation

In Monte Carlo simulation studies the rates of admissible or usable replications should be checked prior to examination of results (Bandalos & Gagné, [2012).](#_bookmark3) Our working definition of usable replications is as follows. First, we identified converged solutions, where model estimation terminated. Secondly, we checked for usable solutions by whether the solution included all positive variances (i.e., absence of negative variances). This method of checking for usable replications is similar to previous simulation studies (DiStefano & Morgan, [2014;](#_bookmark5) Yang-Wallentin et al., [2010).](#_bookmark16) The non-converged and inadmissible

replications were removed from further analyses because they do not provide useful information (Bandalos & Leite, [2013).](#_bookmark4)

Parameter and standard error estimates were evaluated using relative bias and relative efficiency. We evaluated the difference between *θ*ˆ and *θ* using the average relative

bias (ARB),

ARB ˆ

*nr*  *θ*ˆ*j* - *θ* )

100 (2)

*θ* =

*j*=1 *θ*

/*nr* x

where, *nr* is the number of usable replications. ARB was computed for parameter and standard error estimates of factor loadings, factor covariances, level-2 factor (co)variances, level-2 residual variances, and ICC estimates. We evaluated the extend of RB as negligible for RB < 5% , as mild for 5% ::, RB < 10% , and unacceptable for RB > 10% (Hoogland & Boomsma, [1998).](#_bookmark10)

The RE of the estimates from different estimation methods was estimated using

Equation [3.](#_bookmark1)

*θ*ˆ*M LR* - *θ*

2

*RE* =

# \\

*j j*

ˆ*W LSM V*

2 (3)

# )

*j*

*θj* - *θ*

The RE was broken down across sample size conditions (level-1 and level-2 *N* ) to describe which estimation method yielded more efficient estimates as sample size decreased.

## Data sources, evidence, objects, or materials Data Generating Process

Data Generating Process: Data for this project were generated using M*plus* v.8.2

(L. Muthén & Muthén, [2017)](#_bookmark14) utilizing the MONTECARLO command along with MplusAutomation (Hallquist & Wiley, [2018)](#_bookmark8) package in R.

### *Fixed Design Factors*

Data were generated from a factor model with 10 items and two correlated (0.3) factors at both levels. Across levels, the structure and factor loadings were specified to be invariant so that the construct is equivalent across levels. The factor loadings were set to

0.60. The observed indicators were generated to be ordinal with five categories. For this study, the indicators were set to be approximately normally distributed which is the best case possible for using ordered categorical data.

### *Manipulated Design Factors*

Factors that varied across simulation conditions are indicator ICC (ICC*O* = 0.1, 0.3, and 0.5), latent variable ICC (ICC*L* = 0.1, 0.5), number of groups (N2 = 30, 50, 100, 200), and number of level-1 units (N1 = 5, 10, 30).

## Results

Convergence was a problem in this study under all estimation methods when the sample size was low. We found convergence rates as low are 35.5% under WLSMV under low sample sizes, and that MLR and ULSMV had better convergence rate in all conditions. However, as sample size increased to at least 100 groups, differences among estimation methods become negligible.

## Parameter and Standard Error Bias

The average relative bias of parameters and intraclass correlation coefficients (latent and observed variable) across simulation conditions are shown in Figure [1a.](#_bookmark19) The average relative bias of standard errors across simulation conditions are shown in Figure [1b.](#_bookmark20) The effects of design factors on relative bias estimates are reported in the appendix in Table [1.](#_bookmark17) The results were mostly as expected in that estimation method was the most influential factor in relative bias estimates for all parameters and standard errors with partial-*ω*2 estimates ranging from 0.02 to 0.91. The only exception was under the estimation of factor loadings where we identified a partial-*ω*2 = 0*.*30 for the interaction between estimation method and observed variable ICC (ICCO). All other effects were below 0.14. Next, we described the patterns of bias in the different parameters in turn.

«FIGURE [1](#_bookmark19) ABOUT HERE»

## Relative Efficiency

The relative efficiency of parameter and standard error estimates among estimation methods are reported in Table [2.](#_bookmark18)

For parameter estimates, we found that WLSMV was generally the most efficient for all parameters among estimation methods included in this study. The only exception was for the level-2 residual variances in which ULSMV was slightly more efficient when the sample size was very large (N2 � 100 and N1 � 10). However, for estimation of the level-2 factor covariance matrix, ULSMV was only marginally less efficient across all sample sizes. IN generally, these results for the efficiency among estimation methods points to utilizing WLSMV as a first option and utilizing ULSMV as a potential alternative.

For standard error estimates, the story about the relationship among bias, sample size, and estimation method is a bit more complicated. We found that the maximum likelihood method, while treating five ordered categories as continuous, resulted in the most efficient standard errors in most conditions for all parameters. The standard errors for the elements of the level-2 factor covariance matrix were approximately equally efficient across all estimation methods. Exceptions were under the smallest sample size condition.

Estimation of standard errors appears to be highly variable among estimation methods at lower sample sizes.

## Scientific or Scholarly Significance

Nested data structures are abundant in educational research, and current software makes performing investigations of multilevel structures more widespread than ever before. Due to the growing use of these analyses, the burden lies on researchers to make decisions about how to connect theory and data. The benefits of these methods to a variety of disciplines in education, social sciences, and many other fields only come when these methods are used appropriately alongside a guiding theoretical framework. We investigated how well multilevel confirmatory factor analyses recover a known “true model” under a variety of potentially realistic conditions.

Our results indicated that the performance of the three estimation methods considered (MLR, ULSMV, and WLSMV) depended on which parameter was of interest and sample size. The model parameters were biased in different ways for different estimation methods, similar to the results of (DiStefano & Morgan, [2014)](#_bookmark5) for single level CFA models. An unexpected finding was that MLR, where the 5 ordered categorical data were treated as continuous, resulted in the most consistently biased across all sample sizes and parameters. In general, the parameter estimated under MLR were negatively biased except when ICCs were low.

When researchers are planning studies that utilize ML-CFA models, one consideration for data collection that will be to carefully consider the sample size in terms of number of groups and number of participants within groups. We found that there is a small trade-off between number of groups versus within group sample size. Which means that researcher may be able to successfully recover model parameters, even level-2 parameters, when the number of participants within group is increased. For inferences about variances and covariances, when researchers are constrained by the number of groups they have access to (e.g., less than 50) increasing the within group sample size to as many as possible (30+) may help to more accurately estimates one’s model. However, gaining access to more groups would result in more gains in accuracy.

## References

Asparouhov, T., & Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models, In *Joint statistical meeting: Section on statistics in epidemiology*, Salt Lake City, Utah.<https://www.statmodel.com/download/JSM2007000746.pdf>

Bandalos, D. L., & Gagné, P. (2012). Simulation methods in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 92–108). New York, NY, Guilford Press.

Bandalos, D. L., & Leite, W. (2013). The use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd, pp. 625–666). Charlotte, NC, Information Age Publishing.

DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *21* (3), 425–438.<https://doi.org/10.1080/10705511.2014.915373>

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS sstimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16* (4), 625–641.<https://doi.org/10.1080/10705510903203573>

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, *53* (4), 455–467.

<https://doi.org/0033-3123/88/0900-9022500.75/0>

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *25* (4), 621–638. <https://doi.org/10.1080/10705511.2017.1402334>

Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus.* (Third). New York, NY, Routledge.<https://doi.org/10.1177/0146621609344848>

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling an overview and a meta-analysis. *Sociological Methods and Research*, *26* (3), 329–367. <https://doi.org/10.1177/0049124198026003003>

Hox, J., Maas, C. J., & Brinkhuis, M. J. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, *64* (2), 157–170.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49* (1), 115–132. <https://doi.org/10.1007/BF02294210>

Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods &* *Research*, *22* (3), 376–398. <https://doi.org/10.1177/0049124194022003006>

Muthén, L., & Muthén, B. (2017). *Mplus User’s Guide* (8th). Los Angeles, CA, Muthén & Muthén.

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics*, *41* (5), 481–520.<https://doi.org/10.3102/1076998616646200>

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling: A Multidisciplinary Journal*, *17* (3), 392–423. <https://doi.org/10.1080/10705511.2010.489003>

## Table 1

*Design effects on relative bias of parameter and standard error estimation (partial-ω*2 *estimates)*

Loadings L1 Factor Cov. L2 Factor Cov. L2 Residual Var.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Effect | Est. | SE |  | Est. | SE |  | Est. | SE |  | Est. | SE |  |
| N1 | .00 | .11 |  | .00 | .03 |  | .00 | .00 |  | .00 | .00 |  |

N2 .01 .01 .00 .06 .02 .00 .02 .08

ICCO .03 .09 .00 .03 .00 .01 .04 .03

ICCL .02 .02 .00 .02 .02 .06 .02 .00

Estimation **.73 .21** .00 **.20** .02 .02 **.91 .23**

N1:N2 .00 .05 .00 .05 .00 .01 .01 .01

N1:ICCO .00 .04 .00 .02 .00 .02 .01 .04

N1:ICCL .00 .01 .00 .00 .00 .01 .00 .00

N1:Estimation .00 .09 .00 .03 .00 .02 .02 .02

N2:ICCO .00 .03 .00 .01 .00 .00 .01 .03

N2:ICCL .00 .04 .00 .02 .02 .01 .01 .00

N2:Estimation .00 .01 .00 .05 .00 .00 .03 .11

ICCO:ICCL .00 .10 .00 .04 .01 .03 .01 .01

ICCO:Estimation .05 **.30** .00 .09 .00 .00 .11 .01

ICCL:Estimation .02 .07 .00 .03 .00 .01 .06 .01

*Note.* The effects greater than .14 (large effect) are in bold. The meaning of each value can be interpreted as follows. For example, for the effect of the level-1 sample size (N1), 11% of the variability in estimates of relative bias in the standard errors of factor loadings can be attributed to the number of level-1 units were sampled per group after controlling for all other design factors (i.e., N2, ICCO, ICCL, estimation method, and bivariate interactions).

## Table 2

*Relative efficiency indicates WLSMV is better under most sample size conditions*

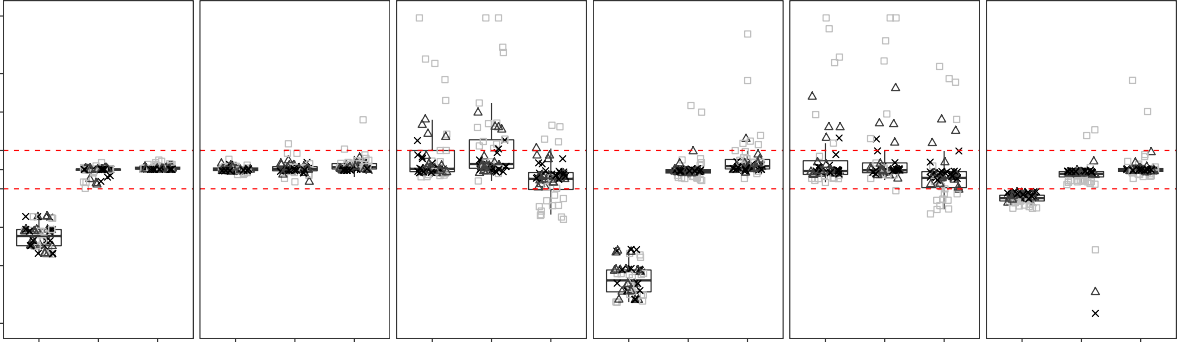
Loadings L1 Factor Cov. L2 Factor Cov. L2 Residual Var. N2 N1 MU MW UW MU MW UW MU MW UW MU MW UW

Parameter Estimates

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 30 | 5 | 1.84 | 2.18 | 1.57 | 1.81 | 1.83 | 1.13 | 1.83 | 2.22 | 1.50 | 1.41 | 1.42 | 1.50 |
| 30 | 10 | 1.68 | 2.33 | 1.54 | 1.08 | 1.28 | 1.19 | 1.10 | 1.42 | 1.30 | 1.43 | 1.52 | 1.05 |
| 30 | 30 | 2.27 | 3.75 | 2.04 | 0.92 | 1.16 | 1.31 | 1.03 | 1.25 | 1.20 | 1.56 | 1.64 | 1.06 |
| 50 | 5 | 1.81 | 2.37 | 1.44 | 1.15 | 1.33 | 1.20 | 1.31 | 1.64 | 1.26 | 1.50 | 1.58 | 1.05 |
| 50 | 10 | 1.78 | 2.81 | 1.79 | 0.97 | 1.15 | 1.21 | 1.05 | 1.25 | 1.18 | 1.64 | 1.66 | 1.00 |
| 50 | 30 | 2.84 | 4.27 | 1.95 | 0.94 | 1.04 | 1.13 | 0.99 | 1.10 | 1.10 | 1.88 | 1.89 | 1.03 |
| 100 | 5 | 2.02 | 2.71 | 1.45 | 1.04 | 1.15 | 1.14 | 1.10 | 1.20 | 1.09 | 1.82 | 1.85 | 1.01 |
| 100 | 10 | 2.79 | 3.64 | 1.63 | 0.96 | 1.07 | 1.14 | 1.01 | 1.11 | 1.09 | 2.18 | 2.17 | 0.99 |
| 100 | 30 | 4.03 | 5.90 | 2.17 | 0.86 | 1.01 | 1.31 | 0.99 | 1.03 | 1.05 | 2.63 | 2.55 | 0.95 |
| 200 | 5 | 2.91 | 3.53 | 1.40 | 0.97 | 1.05 | 1.08 | 1.01 | 1.07 | 1.05 | 2.40 | 2.37 | 0.98 |
| 200 | 10 | 3.86 | 4.86 | 1.54 | 0.96 | 1.01 | 1.07 | 0.99 | 1.03 | 1.04 | 3.02 | 2.95 | 0.96 |
| 200 | 30 | 5.92 | 8.24 | 2.39 | 0.89 | 1.00 | 1.27 | 0.99 | 1.01 | 1.02 | 3.77 | 3.60 | 0.94 |
| Standard Errors | | | | | | | | | | | | | |
| 30 | 5 | 0.78 | 0.89 | 41.18 | 1.34 | 1.26 | 1.00 | 2.55 | 3.17 | 1.83 | 0.34 | 0.34 | 106.74 |
| 30 | 10 | 0.29 | 0.45 | 2.76 | 0.60 | 0.73 | 1.42 | 1.06 | 1.47 | 1.41 | 0.22 | 0.24 | 1.39 |
| 30 | 30 | 0.20 | 0.39 | 3.33 | 0.63 | 0.81 | 1.54 | 0.84 | 1.21 | 1.55 | 0.23 | 0.24 | 1.78 |
| 50 | 5 | 0.34 | 0.52 | 2.50 | 0.80 | 0.84 | 1.39 | 1.35 | 1.78 | 1.32 | 0.23 | 0.26 | 1.41 |
| 50 | 10 | 0.25 | 0.48 | 4.43 | 0.69 | 0.82 | 1.70 | 1.09 | 1.32 | 1.20 | 0.25 | 0.26 | 1.14 |
| 50 | 30 | 0.23 | 0.37 | 4.24 | 0.80 | 0.86 | 1.23 | 0.92 | 1.12 | 1.25 | 0.27 | 0.26 | 4.20 |
| 100 | 5 | 0.28 | 0.44 | 3.58 | 1.04 | 1.14 | 1.59 | 1.21 | 1.29 | 1.08 | 0.23 | 0.24 | 1.55 |
| 100 | 10 | 0.32 | 0.47 | 5.01 | 0.84 | 0.90 | 1.75 | 1.07 | 1.19 | 1.11 | 0.29 | 0.28 | 1.00 |
| 100 | 30 | 0.27 | 0.40 | 4.56 | 0.62 | 0.82 | 3.29 | 0.96 | 1.05 | 1.10 | 0.31 | 0.29 | 0.98 |
| 200 | 5 | 0.33 | 0.42 | 4.23 | 0.71 | 0.86 | 1.76 | 1.08 | 1.13 | 1.06 | 0.25 | 0.25 | 1.02 |
| 200 | 10 | 0.35 | 0.48 | 6.13 | 1.01 | 1.14 | 1.87 | 1.03 | 1.08 | 1.05 | 0.29 | 0.28 | 1.00 |
| 200 | 30 | 0.31 | 0.45 | 8.76 | 0.75 | 0.90 | 4.39 | 0.97 | 1.02 | 1.06 | 0.32 | 0.30 | 0.96 |

*Note.* MU = relative efficiency of MLR compared to ULSMV; MW = relative efficiency of MLR compared to MLSMV; and UW = relative efficiency of ULSMV compared to MLSMV.

Factor Loadings L1 Factor Covariance L2 Factor (co)Variance L2 Residual Variance ICC LV ICC OV



●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

● ●

●

●

●

●

●

●

●

●

●

●

●

●

● ●

●●●

●●

●

●●●●●●●● ●●●●●●●●

● ●

●● ●●●

● ●

●●

●

●

● ●

●

● ●●●●

● ●

●

●●●●●●● ●●●●●

●● ● ● ●

●

●

●●●●●●●●

●

●●●●

●●● ●●

●●●● ●

●

●●●

●●●

●●

●

● ●●●●

●●

●

●●●

●●●●●●●● ●

●●

●

●●●●●●

●

●● ●

●

●● ●●●

●●●

● ●●

●●

● ●

●

●●● ●

●● ● ●

●

●●

●

●● ●●

● ●●

●●●●●●●●

●

● ● ●

●●

●●

●●

Level−2 N

30

* 50

100

200

●

●

●●●●

●●

●

●●●

●

●

●

●

●●●●

●● ●

●

● ●● ●

●●

●●●●●

●

80

50

30

Average Relative Bias

10

0

−10

−30

−50

−80

MLR ULSMV WLSMV

MLR ULSMV WLSMV

MLR ULSMV WLSMV

MLR ULSMV WLSMV

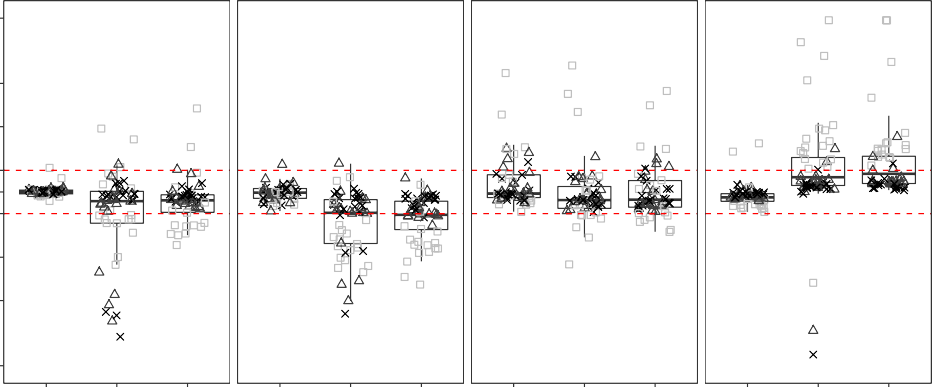
MLR ULSMV WLSMV

MLR

ULSMV WLSMV

1. *Relative bias of parameter estimates showed distinct differences among estimation methods*

Factor Loadings L1 Factor Covariance L2 Factor (co)Variance L2 Residual Variance



●

●

●

●

●

●

●

●

●

●

● ●

●

●

●

●

●

●

●

●

●

●●●

●

●

● ●

●●

●● ●

●

●●●●●●● ●

● ● ●

● ●●

●

●●

●●●●

● ●

● ●

●

●

●

● ●

●● ●

● ●●

●●●

●●● ●

●

● ●●●

●●

●● ●●

● ●

●

●●

●

●

●

● ●●

●

●●● ●

●

●

●●

●

●●●

●● ●

●

●

●

●

● ●

●●●●

● ●●

●

●●

●●

● ●●

● ●●●

●●● ●●

● ●

●

● ●●

● ●●

●●

●●●● ●

●

●●●

● ●●

●

●

●

●

● ●

● ●

●

●●● ●

●●

●

●

●

●

●

Level−2 N

30

* 50

100

200

80

50

30

Average Relative Bias

10

0

−10

−30

−50

−80

MLR ULSMV WLSMV

MLR ULSMV WLSMV

MLR ULSMV WLSMV

MLR ULSMV WLSMV

1. *Standard errors were biased under most low sample size conditions*

## Figure 1

*Bias associated with parameter and standard error estimates. (a) Parameter estimates including intraclass correlation estimates; (b) Standard error estimates.*