**A small sample correction for factor score regression**

Jasper Bogaert, Wen Wei Loh, and Yves Rosseel

Department of Data Analysis

Ghent University

**Author Note**

Correspondence concerning this article should be addressed to Yves Rosseel,Department of Data Analysis, Ghent University, Henri Dunantlaan 2, 9000 Gent, Belgium.Email: Yves.Rosseel@UGent.be.

**Abstract**

Factor score regression (FSR) is widely used as a convenient alternative to traditional structural equation modeling (SEM) for assessing structural relations between latent variables. But when latent variables are simply replaced by factor scores, biases in the structural parameter estimates often have to be corrected, due to the measurement error in the factor scores. The method of Croon (MOC) is a well known bias correcting technique. However, its standard implementation can render poor quality estimates in small samples (e.g., less than 100). This paper aims to develop a small sample correction (SSC) that integrates two different modifications to the standard MOC. We conducted a simulation study to compare the empirical performance of (i) standard SEM, (ii) the standard MOC, (iii) naive FSR, and (iv) the MOC with the proposed SSC. In addition, we assessed the robustness of the performance of the SSC in various models with a different number of predictors and indicators. The results showed that the MOC with the proposed SSC yielded smaller mean squared errors than SEM and the standard MOC in small samples and performed similarly to naive FSR. However, naive FSR yielded more biased estimates than the proposed MOC with SSC, by failing to account for measurement error in the factor scores.

## A small sample correction for factor score regression

In the educational, behavioral, and social sciences, researchers often explore relationships between latent variables such as motivation and intelligence. The most obvious and widely adopted tool of choice to analyze these relationships is structural equation modeling (SEM). SEM is a statistical modeling procedure that estimates the measurement model and the structural relations simultaneously, and is (in combination with maximum likelihood estimation) often regarded to be the gold standard (Bollen, 1989). However, there are several disadvantages to (traditional) SEM. For example, a large sample size is a desideratum and a misspecification of the model might affect the estimation of all parameters (Bentler & Yuan, 1999; Boomsma, 1985; Nevitt & Hancock, 2004).

When substantive interest is primarily in assessing the structural relations between latent variables, a straightforward alternative to SEM is *factor score regression* (FSR) (Skrondal & Laake, 2001). In FSR, the parameters are estimated in two steps. In a first step, confirmatory factor analysis is used to estimate the parameters of the measurement model, and factor scores are predicted for each latent variable. In a second step, the factor scores are used in place of the latent variables to estimate the (linear) relationships among the latent variables. When the structural model is recursive (i.e., no feedback loops), the analysis in the second step boils down to a series of linear regressions (Devlieger et al., 2016). When the structural model is non-recursive, path analysis can be used (Devlieger & Rosseel, 2017). In this paper, we use the term factor score regression for both settings. Factor scores can be calculated in various ways, leading to different sets of observed factor scores. When using Regression (Thomson, 1934; Thurstone, 1935) or Bartlett factor scores (Bartlett, 1937), these are by construction always a linear combination of the observed indicators measuring the latent variable. Therefore, they still contain measurement error and naively using them in place of latent variables will result in biased regression coefficients (Bollen, 1989; Lastovicka & Thamodaran, 1991; Shevlin et al., 1997). We describe other issues that can arise when using this method in the Discussion section. In this paper, we refer to the method described above – without any form of bias-correction – as factor score regression (FSR).

It is nevertheless possible to remove the bias in the estimated regression coefficients. One possible approach is based on the work of Croon (2002), and termed the *method of Croon* (MOC) by Devlieger et al. (2016). The MOC proceeds by removing the measurement error

from the observed variance-covariance matrix of the factor scores in order to obtain a consistent estimator of the variance-covariance matrix of the latent variables. This (co)variance matrix can then be used to obtain asymptotically unbiased estimators of the regression coefficients. Devlieger et al. (2016) and Takane and Hwang (2018) show in their simulation studies that the MOC is a good performing alternative for SEM. As illustrated in Devlieger and Rosseel (2017), a major advantage of the MOC is its robustness against local misspecifications of the model.

While SEM and the MOC are useful methods, both methods can perform poorly in small samples ($N < 100$). Some of the known problems that may occur if SEM is used with (very) small samples are: (a) convergence problems, (b) inadmissible solutions (e.g., Heywood cases), (c) biased estimates for variance components and (d) incorrect confidence intervals and fit statistics (Bentler & Yuan, 1999; Boomsma, 1985; Nevitt & Hancock, 2004; Rosseel, 2020). The reason is that the (frequentist) estimation methods that are typically used in SEM (whether it be maximum likelihood estimation or generalized least squares) only work well if the sample size is sufficiently large. To tackle the problem of non-convergence and inadmissible solutions, De Jonckere and Rosseel (2022) suggest using *bounded estimation*. Their method effectively decreases the rate of non-converge in small samples, but the variability of the estimated regression coefficients in the structural model is still substantial. Smid and Rosseel (2020) show that the MOC, being a two-step method, performs slightly better, but still exhibits a lot of variability when the sample size is very small (say, $N < 50$). Nonetheless, as long as the models are correctly specified, the estimates of the regression coefficients obtained by SEM or the MOC remain unbiased, even in very small samples (Smid & Rosseel, 2020).

Indeed, it is well known that SEM and the MOC take into account the measurement error, and therefore give unbiased results. What is less known by the users of both methods, is the price to pay for these unbiased results. The bias-correction performed by SEM and the MOC leads to a large variability of the parameter estimates, in particular when the sample size becomes small. This is related to the concept of *bias-variance trade-off*, which implies that a decreasing bias will lead to an increasing variance and vice versa (Hastie et al., 2006; James et al., 2013). Therefore, in addition to bias, we also focus on the estimators' variability and mean squared error (MSE) in this paper. Cox and Hinkley (1979) stress that an estimate with a small bias and a small variance might be preferable to one with no bias but a large variance.

Small sample corrections have already been presented in the general SEM framework (e.g., see Ozenne et al. (2020)). However, this is not the case for the MOC. Hence, in this paper, we propose to modify the MOC with a small sample correction. Our modification is based on earlier work reported in the measurement error models literature (Carroll et al., 2006; Fuller, 1987; Wall & Amemiya, 2000). In the context of linear regression with measurement error in the predictors, Fuller (1980) proposes two modifications in order to improve small sample estimation. His aim is to reduce the variability by allowing some bias, resulting in a better MSE accordingly. The goal of this paper is to incorporate these two modifications into the MOC, and examine its performance.

The rest of this paper is structured as follows. First, we introduce a simple structural equation model and the notation. We then describe different estimation methods to analyze the relationships between latent variables: traditional SEM, FSR, and the MOC. Next, information concerning measurement error models and Fuller's modifications is given. Thereafter, we show how to integrate these different adjustments into a single small-sample correction for the MOC. Subsequently, two different simulation studies are described and the results are presented. Lastly, we discuss the performance of the estimators, the robustness of their performance in various models and provide some concluding thoughts.

**A simple structural equation model**

To facilitate the presentation of the correction techniques developed in this paper, we consider a structural equation model with one latent predictor (denoted by $\eta_x$) measured by $p$ indicators, and one latent dependent variable (denoted by $\eta_y$) measured by $q$ indicators. An example with $p = q = 3$ is given in Figure 1.    We use this simple model because it allows us to present many formulas in an easy to understand scalar form. For general settings, we refer the reader to Appendix A where all results are presented using matrix notation. The joint model consists of two models: a measurement model and a structural model. The measurement model can be written as

$$\boldsymbol{x} = \boldsymbol{\nu}_x + \boldsymbol{\lambda}_x \eta_x + \boldsymbol{\epsilon}_x \quad \text{and} \quad \boldsymbol{y} = \boldsymbol{\nu}_y + \boldsymbol{\lambda}_y \eta_y + \boldsymbol{\epsilon}_y, \tag{1}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the $p \times 1$ and $q \times 1$ vectors capturing the observed indicators measuring $\eta_x$ and $\eta_y$ respectively; $\boldsymbol{\nu}_x$ and $\boldsymbol{\nu}_y$ are the vectors of intercepts, $\boldsymbol{\lambda}_x$ and $\boldsymbol{\lambda}_y$ the vectors containing

the factor loadings, and $\boldsymbol{\epsilon}_x$ and $\boldsymbol{\epsilon}_y$ the vectors of the error variables. We assume that $\mathrm{E}(\boldsymbol{\epsilon}_x) = \mathrm{E}(\boldsymbol{\epsilon}_y) = \mathbf{0}$ and denote $\mathrm{Var}(\boldsymbol{\epsilon}_x) = \boldsymbol{\Theta}_x$ and $\mathrm{Var}(\boldsymbol{\epsilon}_y) = \boldsymbol{\Theta}_y$.

The structural model is given by

$$\eta_y = \alpha + \beta\eta_x + \zeta, \tag{2}$$

where $\eta_x$ is the latent predictor and $\eta_y$ is the latent dependent variable; $\alpha$ is the intercept, $\beta$ the regression coefficient, and $\zeta$ the residual error term. We assume that $\mathrm{E}(\zeta) = 0$, $\mathrm{Cov}(\eta_x, \zeta) = 0$, and $\mathrm{Cov}(\boldsymbol{\epsilon}_x, \zeta) = \mathrm{Cov}(\boldsymbol{\epsilon}_y, \zeta) = \mathbf{0}$. We denote $\mathrm{Var}(\eta_x) = \Psi_x$ and $\mathrm{Var}(\zeta) = \Psi_y$.

Without loss of generality, we assume that the observed variables are centered and therefore, $\boldsymbol{\nu}_x = \boldsymbol{\nu}_y = \mathbf{0}$ and $\alpha = 0$. The remaining free parameters to be estimated in this model are the free elements of $\boldsymbol{\lambda}_x$, $\boldsymbol{\lambda}_y$, $\boldsymbol{\Theta}_x$, $\boldsymbol{\Theta}_y$, as well as the scalar parameters $\Psi_x$, $\Psi_y$, and $\beta$. We collect these free parameters in the parameter vector $\boldsymbol{\theta}$. The model-implied variance-covariance matrix is written as $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ to stress that it is a function of the parameter vector $\boldsymbol{\theta}$.

**Maximum likelihood estimation in structural equation modeling**

When all observed variables are continuous, a common estimator in structural equation modeling is maximum likelihood estimation (MLE). When using MLE, the parameter estimates are found by minimizing the discrepancy function

$$\mathrm{F}_{ML} = \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})^{-1}) - \log|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})| + \log|\mathbf{S}| - k, \tag{3}$$

with $\mathbf{S}$ the observed variance-covariance matrix of the observed indicators, $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ the model-implied variance-covariance matrix of the observed indicators, and $k$ the number of observed variables. Note that all the parameters in the parameter vector $\boldsymbol{\theta}$ are estimated simultaneously.

As mentioned before, SEM suffers from convergence issues in small samples. Nonconvergence occurs when the applied optimization method (e.g., quasi-Newton optimization) fails to acquire a solution satisfying certain criteria (Anderson & Gerbing, 1984; Boomsma, 1985). De Jonckere and Rosseel (2022) show that bounded estimation leads to a major decrease in the nonconvergence rate, without negative effects on the point estimates for

the unbounded parameters. Given the focus on small samples in this paper, bounded estimation (using standard bounds) is used for all estimators in this paper in order to avoid convergence problems. Importantly, applying standard bounds prevents the elements of $\boldsymbol{\Theta}_x$ and $\boldsymbol{\Theta}_y$ from being negative, thereby avoiding Heywood cases.

**Factor score regression**

FSR is a two-step method. In the first step, confirmatory factor analysis (CFA) is used to estimate the parameters of the measurement model. As with SEM, convergence issues may also occur here. Therefore, we again use MLE with bounded estimation for the CFA. Because the two measurement models (for $\boldsymbol{x}$ and $\boldsymbol{y}$) do not share any parameters (no cross-loadings or residual correlations), we can estimate both measurement models separately. Once the parameters of the measurement model are estimated, we can compute factor scores for both latent variables as follows:

$$\mathrm{f}_x = \mathbf{a}_x \boldsymbol{x} \quad \text{and} \quad \mathrm{f}_y = \mathbf{a}_y \boldsymbol{y}, \tag{4}$$

where $\mathbf{a}_x$ and $\mathbf{a}_y$ are the factor score vectors. There are multiple ways to compute the factor score vector $\mathbf{a}$, leading to different factor scores. The two most common predictors are the Regression (Thomson, 1934; Thurstone, 1935) and Bartlett predictor (Bartlett, 1937). Using the *Bartlett method*, the factor score vectors $\mathbf{a}_x^B$ and $\mathbf{a}_y^B$ are defined as:

$$\mathbf{a}_x^B = \left[ (\boldsymbol{\lambda}_x^T \boldsymbol{\Theta}_x^{-1} \boldsymbol{\lambda}_x) \boldsymbol{\lambda}_x^T \boldsymbol{\Theta}_x^{-1} \right]^T \quad \text{and} \quad \mathbf{a}_y^B = \left[ (\boldsymbol{\lambda}_y^T \boldsymbol{\Theta}_y^{-1} \boldsymbol{\lambda}_y) \boldsymbol{\lambda}_y^T \boldsymbol{\Theta}_y^{-1} \right]^T. \tag{5}$$

In the *Regression method*, the factor score vectors $\mathbf{a}_x^R$ and $\mathbf{a}_y^R$ are defined as:

$$\mathbf{a}_x^R = \left[ \mathrm{Var}(\eta_x) \boldsymbol{\lambda}_x^T \boldsymbol{\Sigma}_x^{-1} \right]^T \quad \text{and} \quad \mathbf{a}_y^R = \left[ \mathrm{Var}(\eta_y) \boldsymbol{\lambda}_y^T \boldsymbol{\Sigma}_y^{-1} \right]^T. \tag{6}$$

In the second step, we use the factor scores to estimate the regression coefficient $\beta$ of the structural model. The ordinary least-squares estimator of $\beta$ can be found from the variance-covariance matrix of the factor scores as follows:

$$\hat{\beta}_{fs} = \widehat{\mathrm{Var}(\mathrm{f}_x)}^{-1} \ \widehat{\mathrm{Cov}(\mathrm{f}_x, \mathrm{f}_y)}, \tag{7}$$

where the '$fs$' subscript indicates that the estimate is based on factor scores. This estimator is biased regardless of the sample size, except in the trivial setting where $\widehat{\text{Cov}(f_x, f_y)} = 0$, or when there is no measurement error ($\boldsymbol{\epsilon}_x = \boldsymbol{\epsilon}_y = \mathbf{0}$) (Fuller, 1987). But see also Skrondal and Laake (2001) where an unbiased estimate only under specific settings can be obtained using a bias-avoiding FSR method.

**The method of Croon**

The source of the bias when using naive FSR is that the variance–covariance matrix of the factor scores differs from the variance–covariance matrix of the latent variables. Hence, the MOC (Croon, 2002) performs a correction to the variance-covariance matrix of the factor scores so that the corrected variance-covariance matrix is consistent for that of the latent variables. For clarity, we call the elements of the initial variance-covariance matrix the (co)variances of the factor scores, in this case denoted by $\text{Var}(f_x)$, $\text{Var}(f_y)$, and $\text{Cov}(f_x, f_x)$. The elements of the corrected variance-covariance matrix are named the estimated (co)variances of the latent variables, denoted by $\text{Var}(\eta_x)_{Croon}$, $\text{Var}(\eta_y)_{Croon}$, and $\text{Cov}(\eta_x, \eta_y)_{Croon}$.

The variance of a latent variable based on the variance of the factor scores can be obtained by the formula:

$$\widehat{\text{Var}(\eta_x)}_{Croon} \;=\; \text{scale}_x^{-2}\left[\widehat{\text{Var}(f_x)} - \text{offset}_x\right] \;=\; (\boldsymbol{a}_x^T \boldsymbol{\lambda}_x)^{-2}\left[\widehat{\text{Var}(f_x)} - \boldsymbol{\lambda}_x^T \boldsymbol{\Theta}_x \boldsymbol{\lambda_x}\right]. \qquad (8)$$

Furthermore, the covariance of a latent variable based on the covariance of the factor scores can be obtained by:

$$\widehat{\text{Cov}(\eta_x, \eta_y)}_{Croon} \;=\; (\text{scale}_x \times \text{scale}_y)^{-1}\,\widehat{\text{Cov}(f_x, f_y)} \;=\; (\boldsymbol{a}_x^T \boldsymbol{\lambda}_x \; \boldsymbol{a}_y^T \boldsymbol{\lambda}_y)^{-1}\,\widehat{\text{Cov}(f_x, f_y)}. \qquad (9)$$

Thereafter, it is possible to obtain an unbiased estimate of the regression parameter using the corrected variance in Eq. (8) and covariance in Eq. (9):

$$\hat{\beta}_{Croon} = \widehat{\text{Var}(\eta_x)}_{Croon}^{-1}\,\widehat{\text{Cov}(\eta_x, \eta_y)}_{Croon}. \qquad (10)$$

The formulas above simplify when the Bartlett predictor is used to compute the factor scores, because in this case: $\text{scale}_x = \text{scale}_y = 1$. For a more comprehensive description of the MOC, we refer readers to Devlieger et al. (2016) and Devlieger and Rosseel (2017).

**Fuller's modifications for small sample estimation**

We now present the two modifications for the small sample estimator suggested by Fuller (1980). Consider a simple linear regression model with one predictor and one independent variable. Using the notation of Fuller (1987), we can write this as:

$$Y_i = \alpha + \beta\, x_i + e_i, \quad i = 1, 2, \ldots, N, \tag{11}$$

where the $x_i$ are assumed to be independent drawings from a $N(\mu_x, \sigma_{xx})$ distribution, and the $e_i$ are independent $N(0, \sigma_{ee})$ random variables. It is also assumed that $e_i$ is independent of $x_i$. If we write $m_{xx}$ and $m_{xY}$ for the sample (co)variance of $x_i$ (and $Y_i$), then the least squares estimator

$$\hat{\beta}_{ols} = \left[ \sum_{i=1}^{N} (x_i - \bar{x})^2 \right]^{-1} \sum_{i=1}^{N} (x_i - \bar{x})(Y_i - \bar{Y}) = m_{xx}^{-1}\, m_{xY}, \tag{12}$$

is unbiased for $\beta$. The estimator $\hat{\beta}_{ols}$ is also the maximum likelihood estimator for $\beta$, and has the smallest variance of unbiased linear estimators (Fuller, 1987). Next, consider a measurement error model where it is impossible to observe $x_i$ directly. Instead, we observe the sum

$$X_i = x_i + u_i, \tag{13}$$

where $u_i$ is a $N(0, \sigma_{uu})$ random variable, indicating the measurement error. In the absence of measurement error, $\sigma_{uu}$ equals zero. This model is referred to as the *classical measurement error model* (Buonaccorsi, 2010; Carroll et al., 2006). In this measurement error model the least squares estimator

$$\hat{\beta}_{naive} = \left[ \sum_{i=1}^{N} (X_i - \bar{X})^2 \right]^{-1} \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y}) = m_{XX}^{-1}\, m_{XY}, \tag{14}$$

will now be biased. The impact of the measurement error is attenuation of the regression coefficient toward zero. However, if an estimate of $\sigma_{uu} > 0$ is available, it is possible to obtain an unbiased estimator for $\beta$ by using the method of moments (Buonaccorsi, 2010; Carroll et al., 2006; Fuller, 1980, 1987):

$$\hat{\beta}_{mm} = (m_{XX} - \sigma_{uu})^{-1}\, m_{XY}. \tag{15}$$

Note that $\sigma_{uu}$ in Eq. (15) plays the same role as offset$_x$ in Eq. (8). Fuller (1980, 1987) suggests an alternative estimator with better small sample properties. To construct this alternative estimator, Eq. (15) needs two adjustments. Fuller (1980, 1987) proposes the following alternative estimator

$$\hat{\beta}_{\lambda,\alpha} = \left[\hat{\text{H}}_{xx} + \alpha\,(N-1)^{-1}\,\sigma_{uu}\right]^{-1} m_{XY}, \tag{16}$$

where $\hat{\text{H}}_{xx}$ is an estimator of $\sigma_{xx}$, and $\alpha > 0$ is a fixed number to be specified. Furthermore,

$$\hat{\text{H}}_{xx} \quad = \quad \begin{cases} m_{XX} - \sigma_{uu} & \text{if} \quad \hat{\lambda} \geq 1 + (N-1)^{-1} \\[2mm] m_{XX} - [\hat{\lambda} - (N-1)^{-1}]\sigma_{uu} & \text{if} \quad \hat{\lambda} < 1 + (N-1)^{-1}, \end{cases} \tag{17}$$

with $\hat{\lambda}$ the root of the determinantal equation

$$\mid \boldsymbol{m}_{ZZ} - \lambda\,\text{diag}(0,\ \sigma_{uu})\mid = 0, \tag{18}$$

where $\boldsymbol{m}_{ZZ}$ is the sample variance-covariance matrix of both $X_i$ and $Y_i$, containing the elements $m_{XX}$, $m_{XY}$ and $m_{YY}$ in its upper triangular part.

Fuller's correction involves two modifications. First, a decision based on $\hat{\lambda}$ is implemented in Eq. (17) to ensure $\hat{\text{H}}_{xx}$ is positive definite. This modification is named the $\lambda$-correction from here on. Second, including the parameter $\alpha$ in Eq. (16) ensures a less right skewed distribution of the estimated regression coefficients, thereby reducing the variability at the cost of some bias as we explain in the next section. This modification is named the $\alpha$-correction throughout this paper. Both modifications have different functions and can be employed separately from each other. Together they form the alternative estimator presented by Fuller (1980, 1987) for better small sample estimation in the presence of measurement error.

**Combining Fuller's modifications and the method of Croon**

In this section, we show how the two modifications of Fuller (1980) can be incorporated into the MOC estimator. For the considered model, only Var$(\eta_x)_{Croon}$ needs to

be corrected. First, the $\lambda$-correction is performed:

$$\widehat{\mathrm{Var}}(\eta_x)_\lambda = \begin{cases} \widehat{\mathrm{scale}}_x^{-2}\left[\widehat{\mathrm{Var}}(\mathrm{f}_x) - \widehat{\mathrm{offset}}_x\right] & \text{if} \quad \hat{\lambda} \geq 1 + (N-1)^{-1} \\ \widehat{\mathrm{scale}}_x^{-2}\left[\widehat{\mathrm{Var}}(\mathrm{f}_x) - [\hat{\lambda} - (N-1)^{-1}]\widehat{\mathrm{offset}}_x\right] & \text{if} \quad \hat{\lambda} < 1 + (N-1)^{-1}, \end{cases}$$

(19)

where $\hat{\lambda}$ is the smallest root of the determinantal equation:

$$|\widehat{\mathrm{Var}}(\mathrm{f}_x) - \hat{\lambda}\,\widehat{\mathrm{offset}}_x| = 0. \tag{20}$$

This decision based on $\hat{\lambda}$ ensures that $\widehat{\mathrm{Var}}(\eta_x)_\lambda$ is positive. Thereafter, the $\alpha$-correction is performed:

$$\widehat{\mathrm{Var}}(\eta_x)_{ssc} = \widehat{\mathrm{Var}}(\eta_x)_\lambda + \widehat{\mathrm{scale}}_x^{-2}\,\alpha\,(N-1)^{-1}\,\widehat{\mathrm{offset}}_x, \tag{21}$$

where $\alpha > 0$ is a fixed number to be specified. In this model, the offset terms are not needed for the covariances, and therefore $\widehat{\mathrm{Cov}(\eta_x, \eta_y)}_{ssc} = \widehat{\mathrm{Cov}(\eta_x, \eta_y)}_{Croon}$. Now that we have obtained the corrected terms, we can again find an estimate of the regression parameter $\beta$ as usual:

$$\hat{\beta}_{ssc} = \widehat{\mathrm{Var}}(\eta_x)_{ssc}^{-1}\,\widehat{\mathrm{Cov}(\eta_x, \eta_y)}_{ssc}. \tag{22}$$

We call the adjusted estimator in Eq. (22) with the two modifications the method of Croon with a small sample correction (SSC). The two modifications can be employed separately by leaving out the decision based on $\hat{\lambda}$ or by setting $\alpha$ equal to zero. As a consequence, it is possible to implement solely the $\lambda$-correction into the MOC. This latter extension of the MOC is named the MOC-$\lambda$ method from here on.

If we assume $\hat{\lambda} \geq 1 + (N-1)^{-1}$, Eq. (21) can be rewritten differently, giving more insight into the rationale behind the $\alpha$-correction:

$$\widehat{\mathrm{Var}}(\eta_x)_{ssc} = \mathrm{scale}_x^{-2}\left[\widehat{\mathrm{Var}}(\mathrm{f}_x) - \left(1 - \alpha\,(N-1)^{-1}\right)\mathrm{offset}_x\right]. \tag{23}$$

The rationale of the correction is that the variance of the factor scores is subtracted with only a fraction $(1 - \alpha\,(N-1)^{-1})$ of the $\mathrm{offset}_x$ parameter. This results in less variability at the cost of a certain amount of bias. The amount of bias and reduction in variability depends on the value of $\alpha$. Meaningful values for $\alpha$ range from 0 to $N-1$, where using 0 is equivalent to

the MOC and using $N - 1$ is identical to FSR when working with the Bartlett predictor. Two suggestions have been made in the literature regarding the value of $\alpha$. On the one hand, Fuller (1980, 1987) proposes to use $\alpha = p + 1$ (with $p$ the number of predictors). On the other hand, Wall and Amemiya (2000) suggest using $\alpha = p + 5$. In our simulation study, we also consider $\alpha = (N - 1)/2$, which assumes a mid-position between FSR and the MOC. This way, the effect of the correction is larger and adjusts to the sample size at hand. On the other hand, the correction term no longer vanishes if the sample size increases, resulting in biased estimates even when the sample size is very large. This version of the correction is therefore only useful when the sample size is small.

### Simulation studies

Two simulation studies were performed. In the first study, we examined the operating characteristics of the MOC with SSC estimator as compared to other estimators. We have considered 12 different conditions varying in sample size and reliability of the indicators. Given the focus on small sample estimation, the following sample sizes were chosen: $N = 20$, 30, 50, 100, 200, and 2000. In the second study, we assessed the robustness of the performance of the three specifications of $\alpha$ in different models (varying in the amount of latent variables and indicators). We only considered models where the amount of observations is larger than the number of variables in the model. For this reason, we left the smallest sample size ($N = 20$) out in the second simulation study, resulting in 10 distinct conditions per model. The data generation and analysis were done using R (version 4.1.0) (R Core Team, 2020) and the lavaan package (version 0.6-10) (Rosseel, 2012).

**Simulation study 1**

For the first simulation study, two data generating models based on Eq. (25) were considered. They differed only in terms of the reliability of the indicators. The first model had a reliability of 0.5 for all indicators, while in the second model the reliability was set to 0.8. The conditions corresponded respectively to low and high reliability or, in other words, a condition with a higher and a lower amount of measurement error (Brunner & Austin, 2009; Nunnally & Bernstein, 1994). Both models contained three correlated (latent) predictors ($\eta_u$, $\eta_v$, $\eta_w$) and one (latent) dependent variable ($\eta_y$). All latent variables were measured by three observed indicators. In this study, the parameters of interest were the regression coefficients.

The true parameter values $\boldsymbol{\beta}$ for both models were: $\beta_u = 0.5$, $\beta_v = 2$, and $\beta_w = 4$. The true standardized parameter values $\boldsymbol{\beta^z}$ for both models were: $\beta_u^z = 0.072$, $\beta_v^z = 0.302$, and $\beta_w^z = 0.502$. Figure 2 depicts the data-generating model with the true parameter values. For each simulated dataset, we fitted the same model in Figure 2. We then compared seven different estimators: (1) traditional SEM (SEM-MLE), (2) the standard MOC, (3) the MOC-$\lambda$, (4) FSR (using Bartlett factor scores), and (5-7) the SSC with one of the three suggested values for $\alpha$, being: $\alpha = p + 1 = 4$, $\alpha = p + 5 = 8$, and $\alpha = (N-1)/2$. Bounded estimation was used for all estimators to obtain a higher number of converged solutions. Note that SEM-MLE refers to the estimation method (using MLE) and not the modelling procedure. The MOC-$\lambda$ method was added to the simulation to see whether the $\lambda$-correction is beneficial for the MOC.

To evaluate each estimator, the following performance criteria were considered: the proportion of simulated datasets where the method converged (from here on named the convergence rate), the relative mean bias, the empirical standard deviation (ESD), and the mean squared error (MSE) of the estimators. Table 1 summarizes the performance criteria. We simulated 10000 samples, for each of the $6 \times 2 = 12$ settings.

**Simulation study 2**

In the second simulation study, six data generating models were considered. They differed in terms of the reliability of the indicators, as well as in the number of latent predictors and indicators. The first three models had a reliability of 0.5 for all indicators, while in the last three models the reliability was set to 0.8. Hence, there was again a condition with low and high reliability (Brunner & Austin, 2009; Nunnally & Bernstein, 1994). Model 1 and 4 both contained three correlated (latent) predictors ($\eta_u$, $\eta_v$, $\eta_w$) and one (latent) dependent variable ($\eta_y$). All latent variables were measured by three observed indicators. Model 2 and 5 included three correlated (latent) predictors ($\eta_u$, $\eta_v$, $\eta_w$) and one (latent) dependent variable ($\eta_y$) as well. However, in contrast to the other models, all latent variables were measured by six observed indicators. Model 3 and 6 contained six correlated (latent) predictors ($\eta_r$, $\eta_s$, $\eta_t$, $\eta_u$, $\eta_v$, $\eta_w$) and one (latent) dependent variable ($\eta_y$). As in model 1 and 4, all latent variables were measured by three observed indicators. As before, the parameters of interest were the regression coefficients. For convenience the true parameter values $\boldsymbol{\beta}$ for all

models were set to one: $(\beta_r = 1, \beta_s = 1, \beta_t = 1,)$ $\beta_u = 1$, $\beta_v = 1$, and $\beta_w = 1$. The true standardized parameter values $\boldsymbol{\beta^z}$ for all models were: $(\beta_r^z = 0.134, \beta_s^z = 0.155, \beta_t^z = 0.161,)$ $\beta_u^z = 0.134$, $\beta_v^z = 0.155$, and $\beta_w^z = 0.161$. A visualisation of the models and their specific parameter values can be found in the online supplementary material.

Analogous to the first simulation study, we have compared the same seven estimators. Bounded estimation was again used for all estimators to obtain a higher number of converged solutions. In contrast to the first simulation study, the interest lies here in the robustness of the performance of the different specifications of $\alpha$ in the SSC. In order to asses the performance, we used the averaged mean squared error: the mean squared error was averaged over the regression coefficients for every estimator in each model separately. We simulated 10000 samples, for each of the 10 settings (5 different samples sizes x 2 reliabilities) per model.

**Results simulation study 1**

Based on the bias-variance trade-off, a certain pattern in the results was anticipated. On the one hand, we expect SEM and the MOC to be the best performing estimators with respect to bias, but the worst regarding variability. On the other hand, we foresee FSR and the SSC with the largest $\alpha$-value to have the largest bias, but the lowest variability. Moreover, we expect the bias and variability to converge towards zero as the sample size grows larger. Except for FSR and the SSC with the largest $\alpha$-value we anticipate the bias to remain, even in large sample sizes. Finally, it is expected that the patterns will be less distinct when the reliability is high. Tables with the values of the considered performance criteria for the various estimators and settings can be found in the online supplementary material.

*Convergence rate*

Because methods (2)-(7) were adaptations of FSR, they would all either converge or fail to converge for the same dataset. Hence, we grouped these methods together only when assessing convergence rates, and compared SEM-MLE and FSR. Table 2 reveals a high overall proportion of successful replications. The lowest convergence rates (99.30% and 99.89%) are detected for SEM-MLE in the conditions with low reliability and the smallest sample sizes. In fact, a converged solution is acquired 9930 out of 10000 times when $N = 20$, and 9989 out of 10000 times when $N = 30$, both in case of low reliability. From a sample size of 50 onwards, the convergence rates reach approximately 100% for all conditions.

*Bias*

As expected, the results plotted in Figures 3 and 4 display that the mean bias decreases when the sample size grows. Similarly, the differences between the estimators become smaller, but remain larger with low reliability. In contrast to SEM-MLE, the MOC, and MOC-$\lambda$, the bias of FSR and the SSC does not converge to zero, as anticipated. Figures 3 and 4 show that the bias remains largest for FSR and the SSC with the largest $\alpha$-value. This is in particular true for the model with low reliability.

The findings show identical results for the MOC and MOC-$\lambda$ regarding (relative) bias in most conditions. The $\lambda$-correction was used (which means $\widehat{\text{Var}(\boldsymbol{\eta})}_{Croon}$ was not positive definite) substantially when the reliability was low and the sample size was small. The correction was used about 4.43% when $N = 20$ and 1.58% when $N = 30$. It was used less frequently in the following three conditions: 0.20% when the reliability was low and $N = 50$, 0.01% when the reliability was low and $N = 100$, and 0.06% when the reliability was high and $N = 20$. The findings suggest that the $\lambda$-correction does not have a beneficial effect for the (relative) bias.

Interestingly, the (relative) bias seems to depend on how large the regression coefficient is. If the true value is small (e.g., $\beta_u = 0.5$), then the bias is positive. If the true value becomes larger (e.g., $\beta_w = 4$), then the bias is negative.

*Empirical standard deviation*

As expected, the results suggest that the empirical standard deviation reduces as the sample sizes grows. The differences between the estimators become smaller as well, especially in case of high reliability. The findings reveal no major differences between MOC-$\lambda$, FSR, and the SSC. The lowest ESD value is found for FSR and the SSC with the largest $\alpha$-value. Figures 5 and 6 display similar results for all estimators from a sample size of 100 onwards.

The results imply that SEM-MLE is the worst performing estimator in the smaller sample sizes ($N \leq 50$), in particular when the reliability was low. Figures 5 and 6 show that the MOC performed well in most conditions, except in the condition with the smallest sample size ($N = 20$) and low reliability. In this condition, MOC-$\lambda$ outperformed the MOC. Therefore, suggesting the $\lambda$-correction contributes to a lower variability of the estimator. As mentioned above, the $\lambda$-correction is only used substantially when the reliability is low for the

two smallest sample sizes. In all other conditions, the MOC and MOC-$\lambda$ have identical results.

### *Mean squared error*

As expected, the findings show decreasing mean squared error values as the sample size grows larger. Figures 7 and 8 indicate that the differences between the estimators become smaller as well, even more so in case of high reliability.

The overall worst performing method is SEM-MLE, specifically in the conditions with low reliability and small sample sizes. From a sample size of 100 onwards, SEM-MLE performs similar to the other estimators. The MOC shows comparable results to the other FSR estimators. However, its performance is poor in the conditions with low reliably and a sample size of 20 or 30. Hence, the $\lambda$-correction contributes to a lower MSE in these conditions.

In most conditions, the best performing method in terms of MSE is FSR, followed by the SSC with the largest $\alpha$-value. On the other hand, FSR is the worst performing estimator when considering the MSE of $\hat{\beta}_w$ in the two largest sample sizes ($N = 200, 2000$).

### Results simulation study 2

As in the first simulation study, the findings show decreasing average mean squared error values as the sample size grows larger. Figures 9 and 10 indicate that the differences between the estimators become smaller as well, especially in case of high reliability.

Regarding the best and worst performing methods, the same results as in the first simulation study can be observed. The worst performing methods regarding average MSE are SEM-MLE and the MOC. The best performing method is FSR, followed by the SSC. From a sample size of 100 onwards, the methods perform more alike. Additionally, the results reveal once more the relevance of the $\lambda$-correction in the smallest sample sizes ($N = 30, 50$) when reliability is low. This can be verified in Figures 9 and 10, which summarize the average MSE values for the three models with low and high reliability (respectively).

In line with the earlier findings, FSR is the best performing method with respect to average MSE in most conditions (i.e., when $N \leq 200$). To assess the robustness of the results obtained from the three specifications for $\alpha$ specifically, we examine the average MSE of the methods relative to FSR. By using this relative average MSE, the differences in performance compared to other methods can be easily illustrated. The relative average MSE values are shown in Tables 3 and 4 for the three models with low and high reliability (respectively).

As expected, the largest relative average MSE values can be found for SEM-MLE and MOC(-$\lambda$). The differences become smaller when the sample size grows and when the reliability of the indicators is high. In the results, the same patterns can be observed across the three different models. After FSR, the SSC with $\alpha = (N-1)/2$ has the lowest average MSE values, followed by the specifications $\alpha = p + 5$ and $\alpha = p + 1$ (respectively). Despite of the different number of latent predictors and indicators, the findings suggest that the performance of the SSC with the three specifications for $\alpha$ is similar. Furthermore, the results indicate that the model with more indicators per latent variable has lower average MSE values.

**Summary of simulation studies**

In the first simulation study, the overall convergence rates were high for both SEM-MLE and the FSR methods. This high number of converged solutions suggests a beneficial effect from the use of bounded estimation (in this case standard bounds), as discovered earlier by De Jonckere and Rosseel (2022). The $\lambda$-correction was often used in the smallest sample sizes when reliability is low. Nevertheless, it had a profound impact on the ESD and MSE in the condition with the very small sample sizes. The results for the FSR methods were acquired using the Bartlett predictor. However, identical results can be obtained using the Regression predictor.

As mentioned above, the SSC performed well in terms of MSE in the considered conditions. The rationale of the SSC is to introduce some bias for less variability. This trade-off has a beneficial effect for the MSE in small sample estimation. Out of the three suggestions, the estimator with $\alpha = (N-1)/2$ performed best in terms of MSE. Although it had the largest bias, combined with the smaller ESD it resulted in a lower MSE for most cases. We suggested the value $(N-1)/2$ as it adjusts to the sample size at hand and assures a mid-position between FSR and the MOC, but for use in small samples only. Additionally, the SSC with the suggestions from Fuller (1980, 1987) and Wall and Amemiya (2000) surpassed SEM-MLE, MOC, and MOC-$\lambda$ in all conditions.

SEM-MLE and the MOC were the worst performing estimators regarding ESD and MSE, particularly in the smaller sample sizes. However, both estimators outperformed FSR and the SSC regarding bias in most conditions. SEM-MLE and the MOC were particularly superior in terms of bias in case of low reliability and larger sample sizes. The results also

showed decreasing bias of SEM-MLE and the MOC as the sample size grows. Furthermore, the ESD will also reduce, which will eventually (in lager sample sizes) result in a better performance compared to FSR and the SSC.

For the second simulation study, in each of the considered models, highly similar patterns were present in the results. This illustrates the robustness of the performance of the three $\alpha$-values for the SSC in models varying in the number of latent predictors and indicators. Furthermore, the estimators performed slightly better in the model where latent variables were measured by more indicators.

## Discussion

The goal of this paper was to (a) evaluate the performance of (i) standard SEM, (ii) the standard MOC, (iii) naive FSR, and (iv) the MOC with the proposed SSC in small samples sizes and (b) assess the robustness of the performance of the different specification for the SSC in different models. We assessed the estimators in terms of convergence rates, relative mean bias, empirical standard deviation, and mean squared error. The comparison was performed in 12 different conditions, with varying sample sizes and reliability of the indicators.

The findings reveal FSR as the best overall performing predictor, followed by the SSC with the largest $\alpha$-value. Only for the conditions where $N = 200$ or $N = 2000$ and reliability was low, FSR was not the best performing estimator for the regression coefficient $\beta_w$. In fact, FSR was then the worst performing predictor in terms of MSE in both cases. Therefore, the results suggest that both the reliability and sample size at hand play an important role. Our findings are in line with those of Devlieger et al. (2016) where FSR was the worst performing estimator (compared to SEM-MLE and the MOC) regarding MSE in sample sizes above 300. Furthermore, the findings demonstrated that the performances of the different specifications of $\alpha$ were similar across the considered models. The average MSE values of all methods were lowest in the model with more indicators.

Although the FSR and SSC estimators had the best performance in most of the conditions,they do not take the measurement error (entirely) into account. This results in a lower MSE in small samples. However, despite the beneficial effect on the MSE, the consequences of not taking measurement into account should not be neglected. On the one hand, the regression parameters can be over- or underestimated depending on the correlation

structure between the latent variables (Cole & Preacher, 2014). On the other hand, an inflation of the type I error rate may occur under certain circumstances (Brunner & Austin, 2009). Moreover, the more complex the model becomes, the worse the consequences of both problems.

When analyzing relationships between latent variables, we encourage researchers to not choose a method ill-considered. Instead, we recommend to consider the sample size, complexity of the model, reliability of the indicators at hand, and to use the best method for the situation. If the sample size is large enough ($N \geq 200$), one might be better off using SEM-MLE or the MOC instead of FSR or the SSC. However, if the sample size is small ($N \leq 100$), one might consider using FSR or the SSC instead of SEM-MLE or the MOC. Additionally, the goal of the research should be taken into account. If bias is important, unbiased methods such as SEM and the MOC are favored in larger sample sizes, or the SSC with the smallest $\alpha$-value ($= p + 1$) if the sample size is small. If bias is not of primary importance (and a lower variability is), methods such as FSR and the SSC with the largest $\alpha$-value ($= (N-1)/2$) are favored. The small sample correction lets researchers choose a compromise between ignoring measurement error (i.e., using FSR) and taking into account measurement error (i.e., using SEM or the MOC).

Like all simulation studies, our study was not without any limitations. The first limitation of this study is that only certain conditions and models were examined. Although these conditions and models may be commonly found in the literature, different settings may lead to different results. The second limitation is that we have not considered the type I error and the power as performance criteria. Future research should look into hypothesis testing to see how well the SSC and other estimators perform in (small) sample sizes, preferably using a different simulation model to see how general the results obtained in this paper are.

# References

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*(2), 155–173.

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

Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate behavioral research*, *34*(2), 181–197.

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

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, *50*(2), 229–242.

Brunner, J., & Austin, P. C. (2009). Inflation of type i error rate in multiple regression when independent variables are measured with error. *Canadian Journal of Statistics*, *37*(1), 33–46.

Buonaccorsi, J. P. (2010). *Measurement error: Models, methods, and applications*. Chapman; Hall/CRC.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective*. CRC press.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological methods*, *19*(2), 300.

Cox, D. R., & Hinkley, D. V. (1979). *Theoretical statistics*. CRC Press.

Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 195–223). Lawrence Erlbaum.

De Jonckere, J., & Rosseel, Y. (2022). Using bounded estimation to avoid nonconvergence in small sample structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–16.

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM? *Methodology*, *13*, 31–38.

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. https://doi.org/10.1177/0013164415607618

Fuller, W. A. (1980). Properties of some estimators for the errors-in-variables model. *The Annals of Statistics*, *8*(2), 407–422.

Fuller, W. A. (1987). *Measurement error models.* John Wiley & Sons.

Hastie, T., Tibshirani, R., & Friedman, J. (2006). *The elements of statistical learning: Data mining, inference, and prediction.* Springer.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Lastovicka, J. L., & Thamodaran, K. (1991). Common factor score estimates in multiple regression problems. *Journal of Marketing Research*, *28*(1), 105–112.

Nevitt, J., & Hancock, G. R. (2004). Evaluating Small Sample Approaches for Model Test Statistics in Structural Equation Modeling. *Multivariate Behavioral Research*, *39*(3), 439–478. https://doi.org/10.1207/S15327906MBR3903\_3

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (third edition).* McGraw-Hill, Inc.

Ozenne, B., Fisher, P. M., & Budtz-Jørgensen, E. (2020). Small sample corrections for wald tests in latent variable models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *69*(4), 841–861.

R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://www.jstatsoft.org/v48/i02/

Rosseel, Y. (2020). Small sample solutions for structural equation modeling. In R. van de Schoot & M. Miocević (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 226–238). Routledge.

Shevlin, M., Miles, J. N., & Bunting, B. P. (1997). Summated rating scales. a monte carlo investigation of the effects of reliability and collinearity in regression models. *Personality and Individual Differences*, *23*(4), 665–676.

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

Smid, S. C., & Rosseel, Y. (2020). SEM with small samples: Two-step modeling and factor score regression versus Bayesian estimation with informative priors. In R. van de Schoot & M. Miocević (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 239–254). Routledge.

Takane, Y., & Hwang, H. (2018). Comparisons among several consistent estimators of structural equation models. *Behaviormetrika*, *45*(1), 157–188.

Thomson, G. (1934). The meaning of i in the estimate of g. *British Journal of Psychology*, *25*, 92–99.

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

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

## Appendix A: formulas in matrix notation

**The statistical model in structural equation modeling**

Consider a structural equation model with $p$ observed variables and $m$ latent variables. The measurement model is defined as

$$\boldsymbol{y} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{24}$$

where $\boldsymbol{y}$, $\boldsymbol{\nu}$, and $\boldsymbol{\epsilon}$ are $p \times 1$ vectors of respectively the observed variables, intercepts, and residual errors. The $p \times m$ matrix $\boldsymbol{\Lambda}$ contains the factor loadings relating the latent to the observed variables. We denote $\text{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Theta}$. The structural model is defined as

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta}, \tag{25}$$

where $\boldsymbol{\eta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\zeta}$ are $m \times 1$ vectors containing the latent variables, intercepts, and residual error terms respectively. $\mathbf{B}$ is an $m \times m$ matrix containing regression coefficients. The diagonal elements of $\mathbf{B}$ must be zero, and $(\mathbf{I} - \mathbf{B})$ should be invertible. We assume that $\text{E}(\boldsymbol{\epsilon}) = \mathbf{0}$, $\text{E}(\boldsymbol{\zeta}) = \mathbf{0}$, $\text{Cov}(\boldsymbol{\eta}, \boldsymbol{\epsilon}) = \mathbf{0}$, and $\text{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\zeta}) = \mathbf{0}$. The model-implied variance-covariance matrix $\text{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\text{Var}(\boldsymbol{\eta})\boldsymbol{\Lambda}^T + \boldsymbol{\Theta}, \tag{26}$$

where

$$\text{Var}(\boldsymbol{\eta}) = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})^{-T}. \tag{27}$$

**Factor score regression**

The factor scores can be obtained by

$$\mathbf{f} = \mathbf{A}\boldsymbol{y}, \tag{28}$$

with $\mathbf{A}$ the $m \times p$ factor score matrix. Using the Bartlett method (Bartlett, 1937, 1938), the factor score matrix is defined as:

$$\mathbf{A}^B = (\boldsymbol{\Lambda}^T\boldsymbol{\Theta}^{-1}\boldsymbol{\Lambda})\boldsymbol{\Lambda}^T\boldsymbol{\Theta}^{-1}. \tag{29}$$

In the Regression method (Thomson, 1934; Thurstone, 1935), the factor score matrix is defined as:

$$\mathbf{A}^R = \text{Var}(\boldsymbol{\eta})\boldsymbol{\Lambda}^T \boldsymbol{\Sigma}^{-1}. \tag{30}$$

In the second step (once the factor scores are calculated), an estimate of the regression parameters can be obtained as follows:

$$\hat{\boldsymbol{\beta}}_{fs} = \widehat{\text{Var}(\mathbf{f_x})}^{-1} \widehat{\text{Cov}(\mathbf{f_x}, \mathbf{f_y})}. \tag{31}$$

This estimator is biased regardless of the sample size, except in the trivial setting where $\widehat{\text{Cov}(\mathbf{f_x}, \mathbf{f_y})} = \mathbf{0}$, or when there is no measurement error ($\boldsymbol{\epsilon} = \mathbf{0}$) (Fuller, 1987).

**The method of Croon**

Croon (2002) presented his correction formulas in scalar form. We provide here the formulas in matrix form:

$$\text{Var}(\boldsymbol{\eta})_{Croon} = \mathbf{D}^{-1}\left[\text{Var}(\mathbf{f}) - \mathbf{E}\right]\mathbf{D}^{-T} \tag{32}$$

with $\mathbf{E}$ an $m \times m$ matrix defined as

$$\mathbf{E} = \mathbf{A\Theta A}^T, \tag{33}$$

and where $\mathbf{D}$ is an $m \times m$ matrix defined as

$$\mathbf{D} = \mathbf{A\Lambda}. \tag{34}$$

Note that when Bartlett factor scores are used, $\mathbf{D} = \mathbf{I}$, but this is not true for regression factor scores. Using the Croon corrected (co)variance matrix in Eq. (32), we can obtain unbiased estimates for the regression parameters:

$$\hat{\boldsymbol{\beta}}_{Croon} = \widehat{\text{Var}(\boldsymbol{\eta}_x)}^{-1}_{Croon} \ \widehat{\text{Cov}(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y)}_{Croon}. \tag{35}$$

**Small sample correction**

To ensure that the small sample correction formulas are valid for both Bartlett and regression factor scores, we first define

$$\text{Var}(\mathbf{f})^{\star} = \mathbf{D}^{-1} \text{Var}(\mathbf{f}) \, \mathbf{D}^{-T} \tag{36}$$

and

$$\mathbf{E}^{\star} = \mathbf{D}^{-1} \mathbf{E} \, \mathbf{D}^{-T}. \tag{37}$$

Eq.(32) can then be rewritten as

$$\text{Var}(\boldsymbol{\eta})_{Croon} = \text{Var}(\mathbf{f})^{\star} - \mathbf{E}^{\star}. \tag{38}$$

The $\lambda$-correction can be expressed as follows:

$$\widehat{\text{Var}}(\boldsymbol{\eta})_{\lambda} = \begin{cases} \widehat{\text{Var}(\mathbf{f})}^{\star} - \widehat{\mathbf{E}}^{\star} & \text{if} \quad \hat{\lambda} \geq 1 + (N-1)^{-1} \\ \widehat{\text{Var}(\mathbf{f})}^{\star} - \left[\, \hat{\lambda} - (N-1)^{-1} \,\right] \widehat{\mathbf{E}}^{\star} & \text{if} \quad \hat{\lambda} < 1 + (N-1)^{-1}, \end{cases} \tag{39}$$

where $N$ is the sample size, and $\hat{\lambda}$ is the smallest root of the determinantal equation $|\widehat{\text{Var}(\mathbf{f})} - \hat{\lambda} \, \widehat{\mathbf{E}}| = 0$. This decision based on $\hat{\lambda}$ ensures that $\widehat{\text{Var}}(\boldsymbol{\eta})_{\lambda}$ is positive definite. The $\alpha$-correction then proceeds as follows:

$$\widehat{\text{Var}}(\boldsymbol{\eta})_{ssc} = \widehat{\text{Var}}(\boldsymbol{\eta})_{\lambda} + \alpha \, (N-1)^{-1} \, \widehat{\mathbf{E}}^{\star}, \tag{40}$$

where $\alpha$ is a fixed number to be specified. Once the elements of the variance-covariance matrix are corrected, it is possible to obtain estimates of the regression parameters $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}_{ssc} = \widehat{\text{Var}}(\boldsymbol{\eta}_x)_{ssc}^{-1} \, \widehat{\text{Cov}(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y)}_{ssc}. \tag{41}$$

| Performance criteria | Formulas |
|---|---|
| Convergence rate | $\frac{\#\ \text{estimator converged}}{R}$ |
| Mean bias | $\frac{1}{R}\sum_{i=1}^{R}(\hat{b}_i - b)$ |
| Relative mean bias | $\frac{1}{R}\sum_{i=1}^{R}\left(\frac{\hat{b}_i}{b} - 1\right)$ |
| Empirical standard error | $\sqrt{\frac{1}{R-1}\sum_{i=1}^{R}(\hat{b}_i - \bar{b})^2}$ |
| | with $\bar{b} = \frac{1}{R}\sum_{i=1}^{R}\hat{b}_i$ |
| Mean squared error | $\sum_{i=1}^{R}(\hat{b}_i - b)^2$ |

**Table 1**

*Overview of performance criteria. For a given simulated dataset, the estimator is denoted by $\hat{b}_i$ and its true value denoted by b. The number of simulations is denoted by R.*

| | low reliability | | high reliability | |
|---|---|---|---|---|
| $N$ | SEM-MLE | FSR | SEM-MLE | FSR |
| 20 | 99.30 | 99.84 | 99.96 | 99.96 |
| 30 | 99.89 | 99.94 | 100 | 100 |
| 50 | 100 | 99.97 | 100 | 100 |
| 100 | 100 | 100 | 100 | 100 |
| 200 | 100 | 100 | 100 | 100 |
| 2000 | 100 | 100 | 100 | 100 |

**Table 2**
*Overview of convergence rates in percentage. Bounded estimation was used for all estimators. The latent variables were handled separately in the FSR methods, implying that convergence was only achieved if the factor analysis was successful for all four measurement models.*

| Estimators | N | Model 1 | | | | | Model 2 | | | | | Model 3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 30 | 50 | 100 | 200 | 2000 | 30 | 50 | 100 | 200 | 2000 | 30 | 50 | 100 | 200 | 2000 |
| FSR | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SSC ($\alpha = (N-1)/2$) | | 1.29 | 1.30 | 1.29 | 1.23 | 0.72 | 1.22 | 1.20 | 1.19 | 1.16 | 0.91 | 1.43 | 1.42 | 1.38 | 1.31 | 0.86 |
| SSC ($\alpha = p+5$) | | 1.51 | 1.67 | 1.75 | 1.69 | 0.79 | 1.35 | 1.39 | 1.41 | 1.38 | 1.00 | 1.63 | 1.92 | 2.00 | 1.94 | 1.09 |
| SSC ($\alpha = p+1$) | | 1.72 | 1.79 | 1.81 | 1.72 | 0.79 | 1.45 | 1.45 | 1.43 | 1.40 | 1.00 | 1.99 | 2.17 | 2.10 | 1.98 | 1.09 |
| MOC-$\lambda$ | | 2.09 | 1.95 | 1.88 | 1.75 | 0.79 | 1.57 | 1.51 | 1.46 | 1.41 | 1.00 | 4.79 | 3.13 | 2.30 | 2.06 | 1.09 |
| MOC | | 2.36 | 1.95 | 1.88 | 1.75 | 0.79 | 1.57 | 1.51 | 1.46 | 1.41 | 1.00 | 7926.05 | 4.03 | 2.30 | 2.06 | 1.09 |
| SEM | | 12.17 | 3.68 | 2.16 | 1.86 | 0.80 | 2.08 | 1.65 | 1.51 | 1.43 | 1.00 | 32.00 | 15.84 | 2.83 | 2.21 | 1.09 |

**Table 3**
*Relative average mean squared error in case of low reliability.*

| Estimators | N | Model 4 | | | | | Model 5 | | | | | Model 6 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 30 | 50 | 100 | 200 | 2000 | 30 | 50 | 100 | 200 | 2000 | 30 | 50 | 100 | 200 | 2000 |
| FSR | | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| SSC $(\alpha = (N-1)/2)$ | | 1.09 | 1.08 | 1.08 | 1.07 | 0.93 | 1.05 | 1.05 | 1.04 | 1.04 | 1.00 | 1.12 | 1.10 | 1.08 | 1.08 | 0.96 |
| SSC $(\alpha = p+5)$ | | 1.13 | 1.15 | 1.15 | 1.15 | 0.96 | 1.07 | 1.08 | 1.08 | 1.08 | 1.04 | 1.15 | 1.16 | 1.16 | 1.15 | 1.00 |
| SSC $(\alpha = p+1)$ | | 1.16 | 1.16 | 1.16 | 1.15 | 0.96 | 1.09 | 1.09 | 1.09 | 1.08 | 1.04 | 1.19 | 1.18 | 1.17 | 1.16 | 1.00 |
| MOC-$\lambda$ | | 1.19 | 1.18 | 1.17 | 1.16 | 0.96 | 1.10 | 1.09 | 1.09 | 1.08 | 1.04 | 1.27 | 1.22 | 1.19 | 1.17 | 1.00 |
| MOC | | 1.19 | 1.18 | 1.17 | 1.16 | 0.96 | 1.10 | 1.09 | 1.09 | 1.08 | 1.04 | 1.27 | 1.22 | 1.19 | 1.17 | 1.00 |
| SEM | | 4.43 | 1.24 | 1.19 | 1.17 | 0.96 | 1.13 | 1.11 | 1.09 | 1.09 | 1.04 | 29.30 | 1.29 | 1.21 | 1.17 | 1.00 |

**Table 4**
*Relative average mean squared error in case of high reliability.*
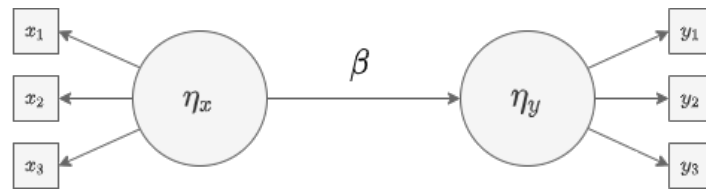
**Figure 1**

*A simple structural equation model.*
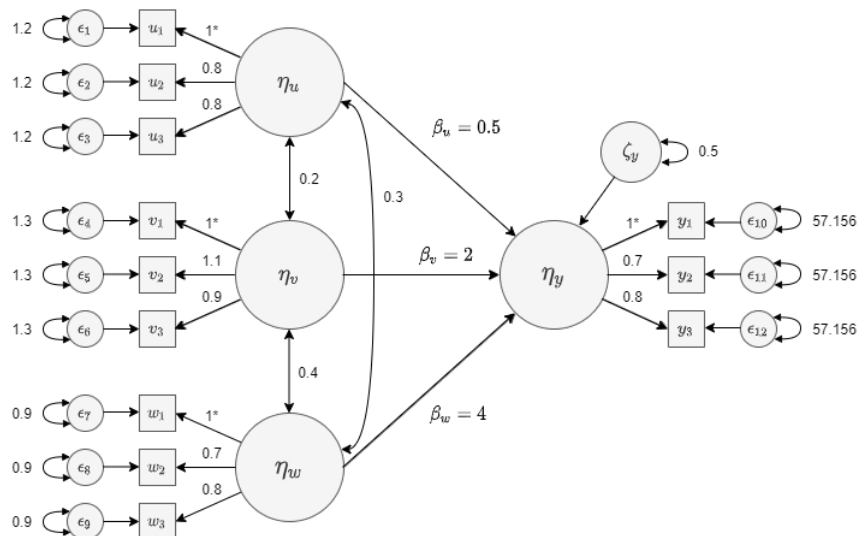
**Figure 2**

*Simulation model. The variances of the latent predictors varied in both models. For the model with low reliability, these were:* $\mathrm{Var}(\eta_u) = 1.2$, $\mathrm{Var}(\eta_v) = 1.3$, *and* $\mathrm{Var}(\eta_w) = 0.9$. *For the model with high reliability, these were:* $\mathrm{Var}(\eta_u) = 4.8$, $\mathrm{Var}(\eta_v) = 5.2$, *and* $\mathrm{Var}(\eta_w) = 3.6$.
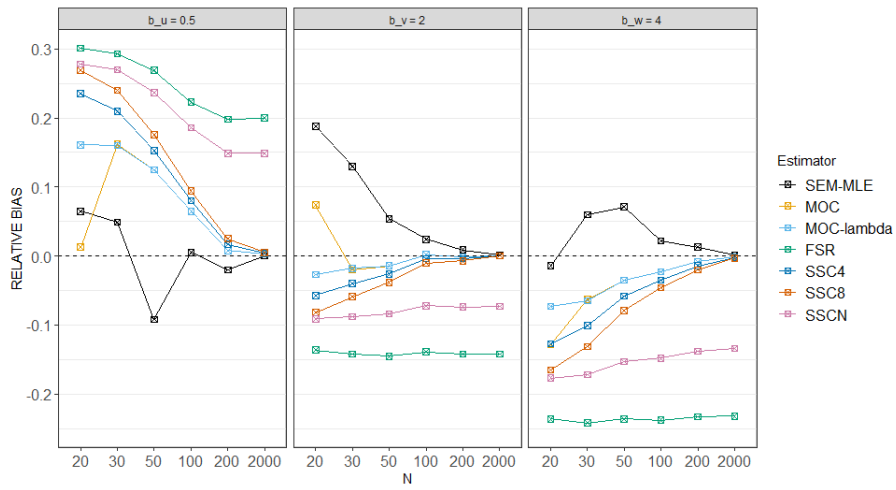
**Figure 3**

*Relative bias in case of low reliability.*
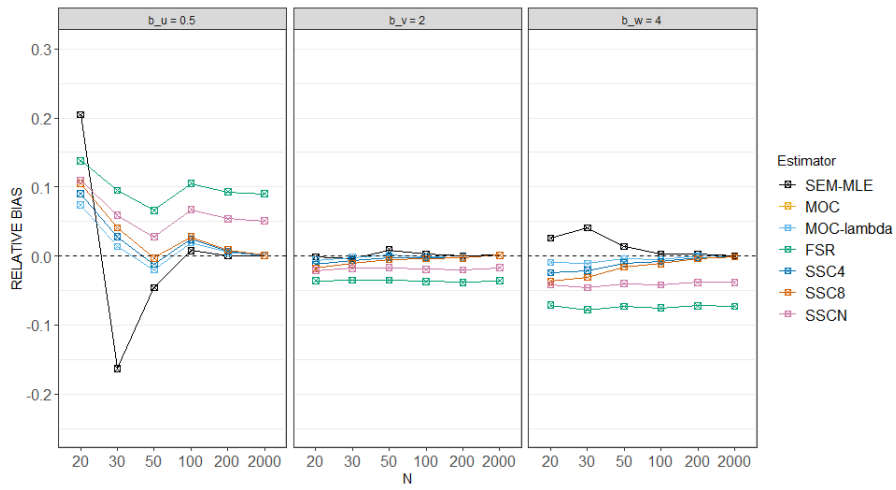
**Figure 4**

*Relative bias in case of high reliability.*
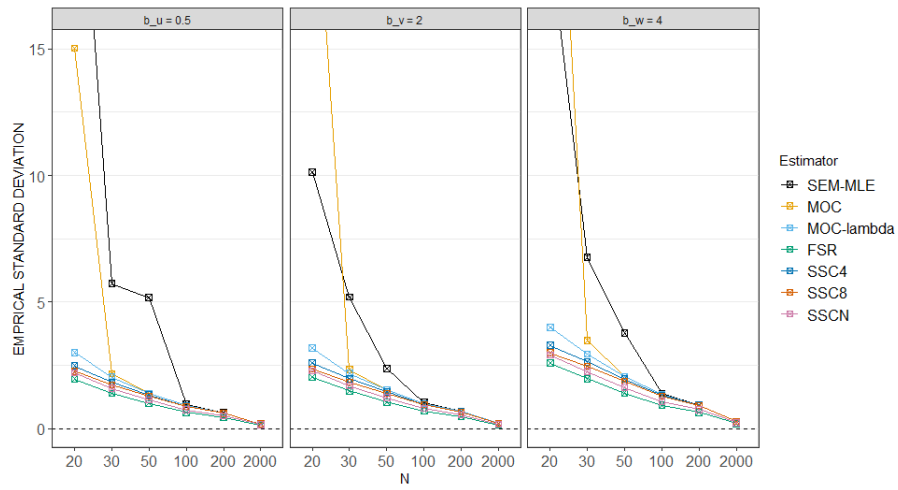
**Figure 5**

*Empirical standard deviation in case of low reliability (some values lie out of scale, these can be found in the corresponding table in the online supplemental material).*
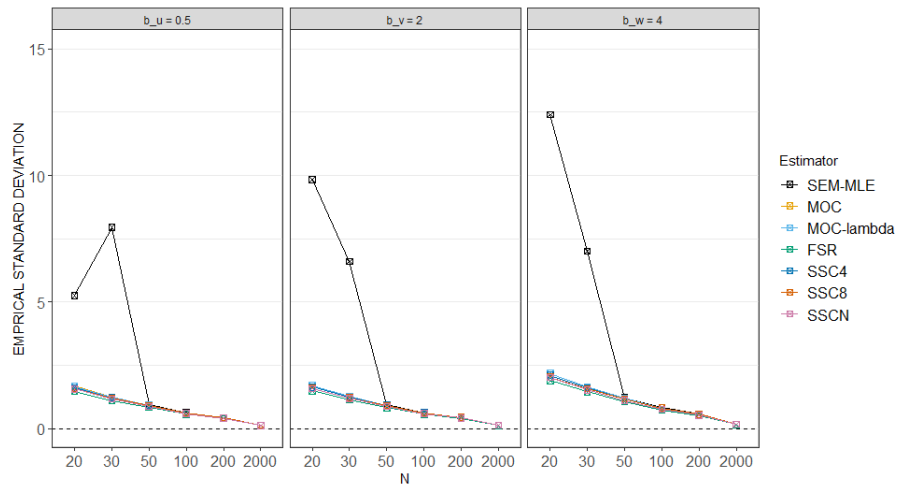
**Figure 6**

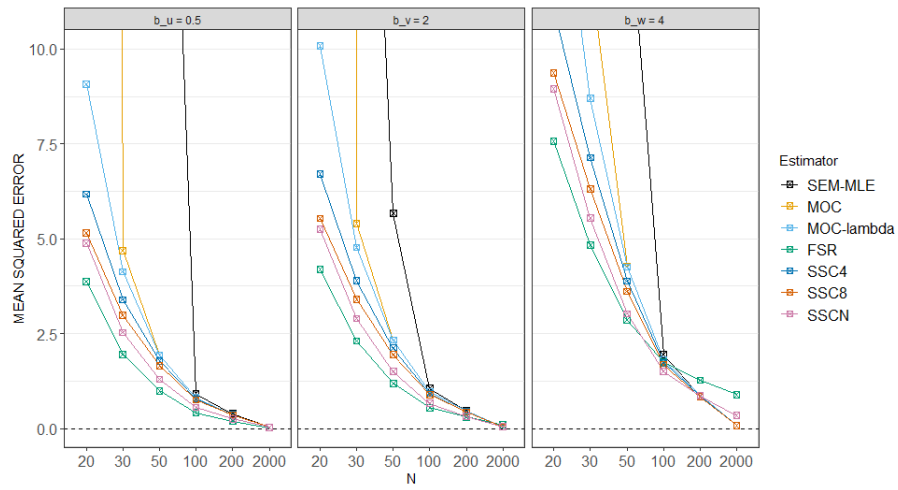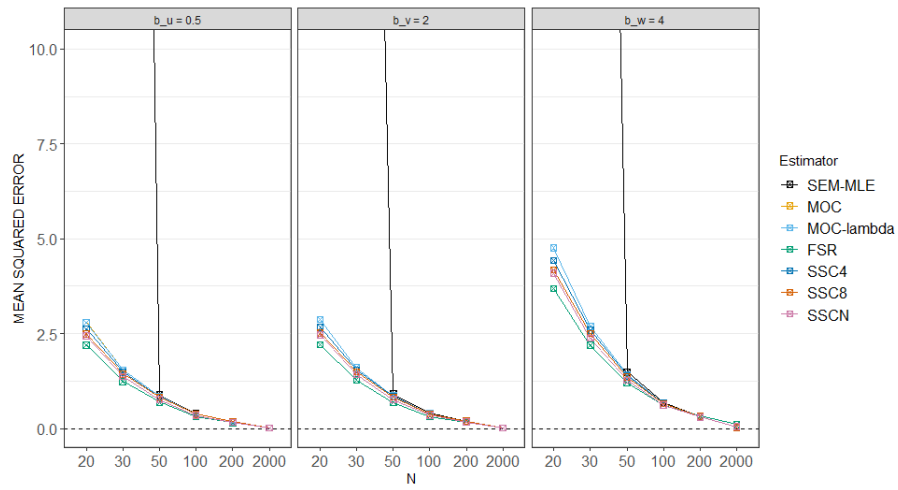*Empirical standard deviation in case of high reliability.*

**Figure 7**

*Mean squared error in case of low reliability (some values lie out of scale, these can be found in the corresponding table in the online supplemental material).*

**Figure 8**

*Mean squared error in case of high reliability (some values lie out of scale, these can be found in the corresponding table in the online supplemental material).*
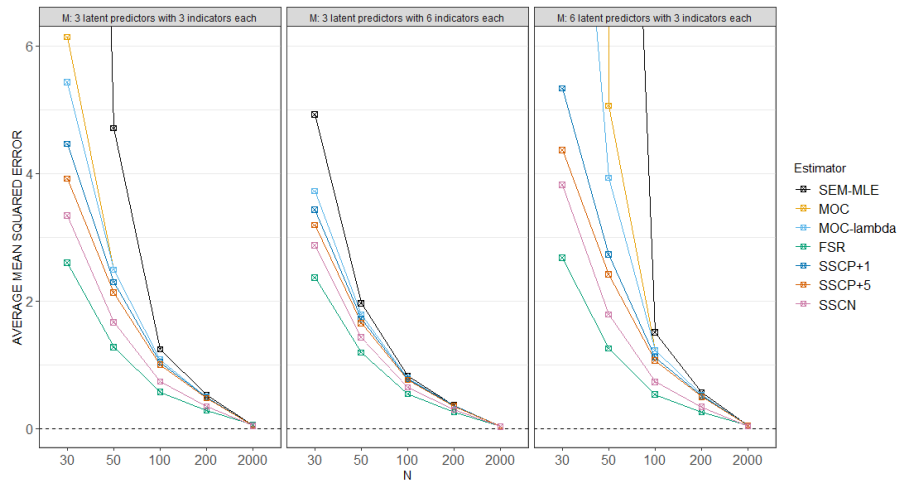
**Figure 9**

*Average mean squared error in case of low reliability (some values lie out of scale, more information regarding these values can be found in Table 3).*
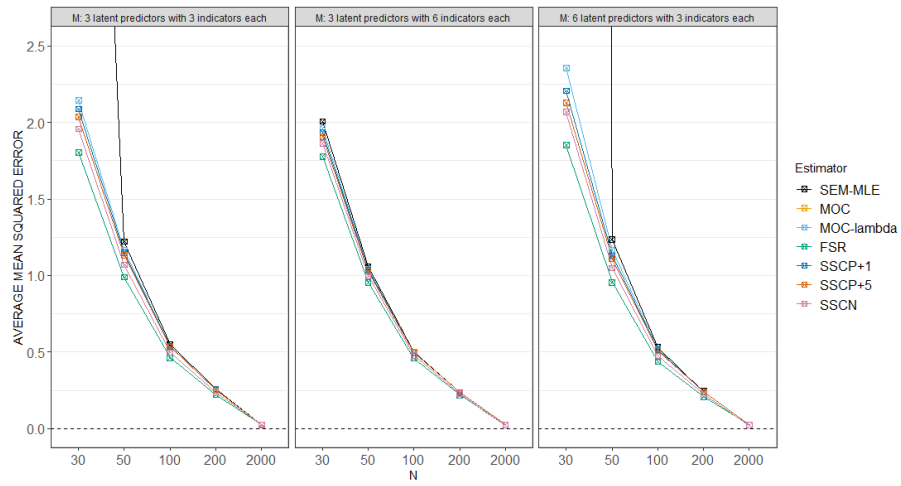
**Figure 10**

*Average mean squared error in case of high reliability (some values lie out of scale, more information regarding these values can be found in Table 4).*