STRUCTURAL EQUATION MODELING
2026, VOL. 33, NO. 3
ISSN: 1070-5511 | E-ISSN: 1532-8007

Evaluating a Bayesian Approach for Estimating Moderator Effects in Parameter-Based Meta-Analytic Structural Equation Modeling

Julian F. Lohmann[a], Oliver Lüdtke[a,b], and Alexander Robitzsch[a,b]

[a]Leibniz Institute for Science and Mathematics Education; [b]Centre for International Student Assessment (ZIB)

ABSTRACT

Meta-analytic structural equation modeling (MASEM) enables meta-analytic investigations of multivariate models. A common research objective in meta-analyses is identifying study-level moderators that explain heterogeneity across primary studies. Several MASEM approaches have been extended to include moderators; however, research evaluating and comparing these approaches remains scarce. The present study discusses several parameter-based moderated MASEM approaches covering one-stage and two-stage as well as fixed and random effects approaches. We implement these different MASEM approaches using a Bayesian modeling framework and evaluate them through two simulation studies. We compare bias, efficiency, coverage rates, and statistical power of moderator effects across different numbers of primary studies, sample sizes, random effect structures, and structural models. Results imply that different parameter-based MASEM approaches provide approximately unbiased estimates and appropriate coverage of study-level moderation effects. An empirical example illustrates the application of the different approaches. Finally, we provide an overall discussion of the findings and their implications.

KEYWORDS

Bayesian modeling; meta-analysis; meta-analytic structural equation modeling; simulation study; study-level moderators

1. Introduction

In recent years, several meta-analytic methods have been proposed to synthesize empirical findings on multivariate theoretical models across multiple primary studies that have measured the same or at least a similar set of variables. These meta-analytic methods allow researchers to move beyond the aggregation of univariate or bivariate statistics in meta-analyses. Many of these methods can be categorized as meta-analytic structural equation models (MASEM; e.g., Becker & Wu, 2007; Cheung & Cheung, 2016; Furlow & Beretvas, 2005; Ke et al., 2025; Viswesvaran & Ones, 1995). MASEM significantly expands the meta-analytic toolbox by enabling the estimation and testing of structural equation models (SEM; Bollen, 1989; Jöreskog et al., 2016) using data from multiple studies. This includes a broad range of multivariate models, such as multiple regression, factor analysis, mediation models, longitudinal models, or any combination of them.

Various approaches to MASEM have been developed that differ in theoretical assumptions and estimation procedures. A key distinction between different approaches concerns how between-study heterogeneity is conceptualized. Fixed effects MASEM assumes that a single structural model and a corresponding fixed set of population parameters hold across primary studies, and any observed differences in correlation or covariance structures are solely due to sampling error (Cheung & Cheung, 2016; Oort & Jak, 2016). By contrast, random effects approaches account for the possibility of systematic heterogeneity across studies, considering that correlations between variables or structural parameters may vary due to factors such as differences in study design features, sample characteristics, or other contextual moderators (Cheung & Cheung, 2016).

Two conceptually distinct approaches have been proposed for modeling random effects within MASEM: the correlation-based (e.g., Cheung & Chan, 2005, 2009; Furlow & Beretvas, 2005) and the parameter-based approach (e.g., Becker & Wu, 2007; Cheung & Cheung, 2016). Correlation-based approaches conceptualize random effects as deviations from a model-implied population correlation matrix, quantifying heterogeneity for each correlation coefficient between the included variables. Parameter-based approaches define random effects as deviations from the average population level structural parameters, quantifying heterogeneity for each freely estimated parameter in the structural model under investigation.

In order to explain heterogeneity across studies, many meta-analyses include examinations of study-level moderators. To expand MASEM for this purpose, both correlation-based and parameter-based MASEM have been proposed to be extended by moderator effects. Notably, in both approaches, these moderation effects are located at the level of structural parameters (Jak & Cheung, 2020; Ke et al., 2025). Thus, for applications of MASEM that prioritize interpreting SEM parameters, parameter-based approaches provide quantification of both between-study heterogeneity and moderation effects at the level of SEM parameters. This alignment can be a concrete advantage, for instance, when enriching moderation analyses through the calculation of R^2 measures to quantify explained heterogeneity for specific SEM parameters (Ke et al., 2025). Despite this potentially advantageous alignment between the quantification of heterogeneity and the effects of moderators, methodological evaluations of the statistical performance of parameter-based MASEM approaches remain scarce. This gap is particularly noteworthy given that the investigation of study-level moderators represents a primary research objective in many meta-analyses, including those using MASEM (see, e.g., Agostini & van Zomeren, 2021; Dapp et al., 2023; Möller et al., 2020; Scherer & Campos, 2022; Vasconcellos et al., 2025). To address this gap, the present study conducts two simulation studies to evaluate and compare different parameter-based MASEM approaches with a specific focus on moderation effects for the first time. Through this comparison, we aim to provide practical guidance for applied researchers interested in examining study-level moderators in the context of MASEM.

This article is structured as follows. After briefly introducing SEM notation, we provide an overview of MASEM approaches, including a detailed discussion of how study-level moderators can be included in MASEM. Next, two simulation studies are presented focusing on the performance of estimating study-level moderation effects across various conditions and parameter-based MASEMs. The first simulation study is based on a two-dimensional CFA model with study-level moderators. The second simulation study examines the flexibility of moderated parameter-based MASEM by implementing and evaluating a longitudinal meta-analytic model: the random-intercept first-order autoregressive (RI-AR(1)) model, also referred to as the univariate random-intercept cross-lagged panel model (Hamaker et al., 2015). We then illustrate the application of the different moderated parameter-based approaches by reanalyzing an empirical meta-analytic dataset. Finally, we discuss the simulation results, outline implications for applied research, and propose directions for future methodological developments and evaluations.

2. Structural Equation Modeling

SEM allows researchers to estimate and test models that imply complex relationships among manifest (observed) and latent (unobserved) variables (Bollen, 1989; McArdle & McDonald, 1984; Raykov & Marcoulides, 2012; Yuan & Bentler, 2006). SEM is typically represented in terms of two main components—a measurement model and a structural model (Bollen, 1989; Hoyle, 1995). The measurement model relates a vector of observed variables $\boldsymbol{y}$ (also referred to as indicators) to a vector of latent variables $\boldsymbol{\eta}$ and is represented as

$$\boldsymbol{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\Lambda}$ is a matrix of factor loadings and $\boldsymbol{\epsilon}$ is a vector of errors. Here, we assume $\mathrm{E}(\boldsymbol{y}) = \mathbf{0}$, $\mathrm{E}(\boldsymbol{\eta}) = \mathbf{0}$, $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$, meaning that all variables are mean-centered (if required, a mean structure can be included and estimated; see, e.g., Bollen, 1989; Hoyle, 1995). Measurement errors $\boldsymbol{\epsilon}$ are multivariate normally distributed with zero mean vector and covariance matrix $\boldsymbol{\Omega}$. In addition, measurement errors and latent variables are assumed to be uncorrelated.

The structural part, which represents the relationships between latent variables, is given by

$$\boldsymbol{\eta} = \mathrm{B}\boldsymbol{\eta} + \boldsymbol{\zeta}, \tag{2}$$

where B is a matrix of regression coefficients and $\boldsymbol{\zeta}$ is a vector of disturbances, which are again multivariate normally distributed with zero mean vector and covariance matrix $\boldsymbol{\Psi}$.

Adding the measurement part and assuming that $\boldsymbol{\zeta}$ and $\boldsymbol{\epsilon}$ are uncorrelated, the model-implied covariance matrix is derived as

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\mathbf{I}-\mathrm{B})^{-1}\boldsymbol{\Psi}\left[(\mathbf{I}-\mathrm{B})^{-1}\right]^{\mathrm{T}}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Omega}, \tag{3}$$

where $\boldsymbol{\theta}$ is a vector of transformed freely estimated model parameters contained in B, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Omega}$. In the following, we assume that appropriate parameter transformations are applied to ensure that optimization is performed over an admissible parameter space. More specifically, the logit transformation is used for parameters bounded between 0 and 1 (e.g., standardized loadings), the logarithmic transformation for parameters constrained to positive values (e.g., variances), and the Fisher z-transformation for parameters ranging between −1 and 1 (e.g., correlations). When estimating this model, the discrepancy between the empirical covariance or correlation matrix and the model-implied covariance (or standardized covariance) matrix is minimized (see, e.g., Bollen, 1989; Browne & Arminger, 1995; Cudeck, 1989; Yuan & Bentler, 2006).

3. Meta-Analytic Structural Equation Modeling

In the context of meta-analysis, researchers are interested in summarizing results across multiple primary studies rather than analyzing a single dataset from one study as assumed in the previous section. Consequently, instead of a single empirical covariance or correlation matrix, MASEM involves the integration and modeling of data from S studies, each contributing a sample correlation or covariance matrix (e.g., Cheung & Hafdahl, 2016; Viswesvaran & Ones, 1995). Typically, MASEM uses vectorized sample correlation matrices as reported (or computed) for each individual study as the primary input data (e.g., Becker, 1992; Cheung & Chan, 2005; Cheung & Cheung, 2016). Thus, let $\hat{\boldsymbol{\rho}}_i = (\hat{\rho}_{i1}, \ldots, \hat{\rho}_{iH})$ be a vector with H entries containing the bivariate estimated sample correlations of the variables of interest as measured in primary study i ($i = 1, 2, \ldots, S$). If the number of variables is p, then the length of $\hat{\boldsymbol{\rho}}_i$ is $H = \frac{p(p-1)}{2}$, representing all unique pairwise correlations among the p variables.

Using asymptotic theory, the distribution of the estimated correlations $\hat{\boldsymbol{\rho}}_i$ follows approximately a multivariate normal distribution, that is $\hat{\boldsymbol{\rho}}_i \sim MVN(\boldsymbol{\rho}_i, \boldsymbol{V}_i)$ (Hedges & Olkin, 1981; Olkin & Finn, 1995; Steiger & Hakstian, 1982). Here, the vector $\boldsymbol{\rho}_i$ includes the population correlation coefficients and $\boldsymbol{V}_i$ is the corresponding $H \times H$ sampling covariance matrix. Typically, an estimate $\hat{\boldsymbol{V}}_i$ of this covariance matrix is calculated as a function of the study sample size and the observed correlations $\hat{\boldsymbol{\rho}}_i$ (Olkin & Finn, 1995; see also Cheung & Chan, 2004).

3.1. Fixed Effects Approach

In the next step, an SEM is imposed to model the pattern of correlations among the observed variables across studies (e.g., Oort & Jak, 2016). As introduced in the previous section, the model-implied correlation matrix for a given structural model is a function of a vector of free model parameters $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. In the meta-analytic context, this formulation is extended to multiple studies by expressing each study's estimated vector of bivariate sample correlations as a function of the common structural model and its free model parameters written as

$$\hat{\boldsymbol{\rho}}_i = \boldsymbol{\rho}(\boldsymbol{\theta}_0) + \boldsymbol{e}_i, \tag{4}$$

where $\boldsymbol{\rho}(\boldsymbol{\theta}_0)$ denotes the vectorized form of the model-implied population average correlation coefficients reflecting the model structure imposed by a given structural model and its estimated population parameters included in $\boldsymbol{\theta}_0$. The residuals $\boldsymbol{e}_i$ follow a multivariate normal distribution with zero mean vector and sampling covariance matrix $\hat{\boldsymbol{V}}_i$. It needs to be emphasized that the fixed effects approach in Equation (4) assumes that all residual between-study variation in the observed correlations $\hat{\boldsymbol{\rho}}_i$ is attributable to sampling errors. However, in practical applications, it is often observed that the correlations reported across primary studies exhibit greater variability than can be explained by sampling errors alone (Cheung & Cheung, 2016; Hafdahl, 2008). This suggests the presence of additional sources of between-study heterogeneity, such as differences in study design features or sample characteristics. Consequently, random effects approaches to MASEM have been developed to account for heterogeneity in the observed correlations that cannot be explained by sampling error alone (Cheung & Cheung, 2016; Ke et al., 2025).

3.2. Random Effects Approach

Two general conceptualizations for incorporating random effects into MASEM can be distinguished (Cheung & Cheung, 2016). In the correlation-based approach (e.g., Jak & Cheung, 2020), Equation (4) is extended by a vector of random effects in the following way

$$\hat{\boldsymbol{\rho}}_i = \boldsymbol{\rho}(\boldsymbol{\theta}_0) + \boldsymbol{u}_{\rho i} + \boldsymbol{e}_i, \tag{5}$$

where $\boldsymbol{u}_{\rho i}$ represents the deviation of study i from the population average model-implied vector of correlation coefficients $\boldsymbol{\rho}(\boldsymbol{\theta}_0)$. Random effects $\boldsymbol{u}_{\rho i}$ have a multivariate normal distribution with zero means and a covariance matrix $\mathbf{T}_{\rho}$, that is $\boldsymbol{u}_{\rho i} \sim MVN(\mathbf{0}, \mathbf{T}_{\rho})$.

In the parameter-based approach (e.g., Cheung & Cheung, 2016; Ke et al., 2019), Equation (4) is extended as follows

$$\hat{\boldsymbol{\rho}}_i = \boldsymbol{\rho}(\boldsymbol{\theta}_0 + \boldsymbol{u}_{\theta i}) + \boldsymbol{e}_i, \tag{6}$$

where $\boldsymbol{u}_{\theta i}$ represents the deviation of study i from the population average vector of SEM parameters $\boldsymbol{\theta}_0$ with $\boldsymbol{u}_{\theta i} \sim MVN(\mathbf{0}, \mathbf{T}_{\theta})$. As is evident when comparing Equations (5) and (6), the correlation-based approach conceptualizes between-study heterogeneity (i.e., $\mathbf{T}_{\rho}$) as study-specific deviations from the population average correlation coefficients. In contrast, the parameter-based approach conceptualizes between-study heterogeneity (i.e., $\mathbf{T}_{\theta}$) as study-specific deviations from the population average SEM parameters.

3.3. Including Study-Level Moderation Effects

In many meta-analyses, a primary objective is not only to account for between-study heterogeneity but also to explain it through study-level moderators. Consequently, both types of random effects MASEM approaches have been extended by moderators. For example, Jak and Cheung (2020) incorporated moderators in the correlation-based approach. Using their proposed extension, Equation 5 is expanded as follows

$$\hat{\boldsymbol{\rho}}_i = \boldsymbol{\rho}(\boldsymbol{\theta}_0 + \boldsymbol{\Gamma}\boldsymbol{z}_i) + \boldsymbol{u}_{\rho i} + \boldsymbol{e}_i, \tag{7}$$

where $\boldsymbol{\Gamma}$ is a matrix of moderation effects and $\boldsymbol{z}_i$ is a vector of moderators. In a similar way, the parameter-based approach can be extended to include study-level moderators (Cheung & Cheung, 2016; Ke et al., 2025)

$$\hat{\boldsymbol{\rho}}_i = \boldsymbol{\rho}(\boldsymbol{\theta}_0 + \boldsymbol{\Gamma}\boldsymbol{z}_i + \boldsymbol{u}_{\theta i}) + \boldsymbol{e}_i. \tag{8}$$

Comparing Equations (7) and (8) reveals that moderated parameter-based MASEM aligns the quantification of heterogeneity at the level of SEM parameters with the location of the moderation terms. In contrast, moderated correlation-based MASEM assumes random effects at the level of correlation coefficients and moderation effects for SEM parameters (see Cheung, 2015).

In many MASEM applications, researchers are primarily interested in examining and explaining the variability of SEM parameters such as standardized factor loadings, interfactor correlations, and regression coefficients. In these contexts, it can be helpful to align heterogeneity quantification and moderation effects at the SEM parameter level, for example, to calculate the explained heterogeneity for specific SEM parameters and moderators. However, despite these conceptual strengths, parameter-based MASEM has hardly been systematically evaluated to date, except for a simulation study provided by Ke et al. (2019). An evaluation regarding study-level moderation effects is completely missing so far. These gaps in the literature may help explain why parameter-based MASEM remains rarely used in substantive meta-analytic research. In light of these limitations and in response to recent calls for further methodological evaluations (see, e.g., Ke et al., 2019; 2025), we focus on parameter-based approaches in this article.

3.4. Estimation Procedures for the Parameter-Based Approach

There exist mainly two general strategies for estimating (moderated) parameter-based MASEMs: a one-stage (Ke et al., 2019) and a two-stage approach (e.g., Becker, 1992; Becker & Wu, 2007; Cheung & Cheung, 2016). In the two-stage approach, the target SEM is fitted to the correlation matrix (or dataset) of each primary study separately. After estimating the model of interest for each primary study, the study-specific vector of parameter estimates $\hat{\boldsymbol{\theta}}_i$ and their asymptotic sampling covariance matrix $\hat{\boldsymbol{V}}_{\boldsymbol{\theta}i}$ are kept from each primary study (e.g., Cheung & Cheung, 2016). The matrix $\hat{\boldsymbol{V}}_{\boldsymbol{\theta}i}$ reflects the uncertainty in $\hat{\boldsymbol{\theta}}_i$ due to sampling variability and is typically obtained directly from the estimation output of the SEM software used (e.g., via the inverse of the Fisher information matrix, or the robust sandwich estimator). Both, $\hat{\boldsymbol{\theta}}_i$ and $\hat{\boldsymbol{V}}_{\boldsymbol{\theta}i}$ are used as inputs for a random effects meta-regression model in the second step. More formally, the two-stage procedure can be represented as (e.g., Cheung & Cheung, 2016)

$$\begin{aligned}&\text{Stage 1:}\\&\text{fit } \boldsymbol{\rho}_{\mathrm{i}} = \boldsymbol{\rho}(\boldsymbol{\theta}_i) \text{ to } \hat{\boldsymbol{\rho}}_i \text{ and obtain } \hat{\boldsymbol{\theta}}_i \text{ and } \hat{\boldsymbol{V}}_{\boldsymbol{\theta}i} \text{ for study}\\&\quad i = 1, 2, \ldots, S\\&\text{Stage 2: } \hat{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_0 + \boldsymbol{\Gamma}\boldsymbol{z}_i + \boldsymbol{u}_{\boldsymbol{\theta}i} + \boldsymbol{e}_{\boldsymbol{\theta}i},\\&\quad\quad \text{with } \boldsymbol{e}_{\boldsymbol{\theta}i} \sim MVN(\mathbf{0}, \hat{\boldsymbol{V}}_{\boldsymbol{\theta}i}) \text{ and } \boldsymbol{u}_{\boldsymbol{\theta}i} \sim MVN(\mathbf{0}, \mathbf{T}_{\boldsymbol{\theta}}).\end{aligned} \tag{9}$$

where $\boldsymbol{u}_{\boldsymbol{\theta}i}$ is a vector of study-specific deviations from the average population parameters $\boldsymbol{\theta}_0$. Note that the residuals $\boldsymbol{e}_{\boldsymbol{\theta}i}$ reflect the uncertainty (i.e., $\hat{\boldsymbol{V}}_{\boldsymbol{\theta}i}$) that is due to estimating the SEM parameters in study i.

Ke et al. (2019) developed a one-stage approach to parameter-based MASEM that fits Equation (6) directly to the set of primary-study-specific correlation matrices. Compared to two-stage approaches, one-stage approaches can provide more efficient parameter estimates when the model is correctly specified. A particular challenge of this approach, however, is model estimation when many random effects are included, due to the large number of SEM parameters that must be estimated. To address the computational and numerical difficulties associated with the high-dimensional integration required by frequentist methods (e.g., maximum likelihood), Ke et al. (2019) proposed using Bayesian estimation via Markov Chain Monte Carlo (MCMC) sampling techniques. Simulation studies indicated that this one-stage approach produces approximately unbiased estimates and yields reliable inferences for MASEM parameters (in two-dimensional CFA models without study-level moderators) under realistic conditions. Comparing a fixed and a random effects model, Ke et al.'s simulation results implied accurate estimation of population average CFA model parameters (i.e., loadings and interfactor correlation) for both models but substantially reduced coverage rates for the fixed effects model when the data-generating model imposed between-study heterogeneity.

4. Present Study

In the present study, we aimed to evaluate four different moderated parameter-based MASEM approaches, with a particular focus on their performance in estimating moderation effects. First, we considered Ke et al.'s (2019) one-stage parameter-based random effects approach. In the model version used in Ke et al. (2019), the authors imposed an assumption of uncorrelated random effects to reduce the complexity of the model. In line with this, we implemented Equation (6) with $\mathbf{T}_{\boldsymbol{\theta}}$ being reduced to a diagonal matrix. In the following, this model is referred to as *OS-UCRE* (one-stage with uncorrelated random effects). Second, we also considered the same model with the full covariance matrix of the random effects being estimated. We refer to this model as *OS-CRE* (one-stage with correlated random effects). Third, we also implemented a fixed effects version of moderated one-stage MASEM (see Equation (4)). However, as known from the literature (Raudenbush & Bryk, 2002; see also Barr et al., 2013; Heisig & Schaeffer, 2019), ignoring heterogeneity—i.e., not modeling random effects when parameter heterogeneity is present in the data-generating model—can bias standard error (*SE*) estimation and thus statistical inference. This issue has also been demonstrated for one-stage MASEMs in the simulation studies by Ke et al. (2019). Nevertheless, given the possible scenario that researchers are not interested in quantifying parameter heterogeneity and instead focus on correct inference for moderation effects, a fixed effects MASEM with robust *SE* estimation methods can be a reasonable alternative to random effects modeling. Therefore, we included a fixed effect approach with moderator effects and resampling-based *SE* estimation (e.g., Rao & Wu, 1988) in our evaluations. This model is referred to as *OS-FE* in the following. Fourth, we also included the two-stage estimation procedure to parameter-based MASEM. This model is referred to as *TS* in the following.

We compared these approaches in two simulation studies. In Simulation Study 1, we used a frequently employed simulation scenario in MASEM, namely a two-dimensional CFA model with heterogeneous SEM parameters across primary studies, and extended it by study-level moderator effects. Simulation Study 2 demonstrates the flexibility of moderated parameter-based MASEM. For this purpose, we implemented MASEMs and generated data for a meta-analytic random-intercept autoregressive model (RI-AR(1); Skrondal & Rabe-Hesketh, 2008), also referred to as the univariate random-intercept cross-lagged panel model (Hamaker et al., 2015; see also Robitzsch & Lüdtke, 2025).

5. Simulation Study 1: Confirmatory Factor Analysis

The data-generating model of Simulation Study 1 was a two-dimensional CFA model with model parameters varying across primary studies. Two-dimensional CFA models have been frequently applied in prior evaluations of MASEM (Cheung & Chan, 2005; Furlow & Beretvas, 2005; Jak & Cheung, 2020; Ke et al., 2019). However, the current study is the first to include (continuous and dichotomous) study-level moderators.

5.1. Method

5.1.1. Data Generation

The data-generating model was based on a CFA model with two correlated latent factors and three indicators per factor (Figure 1). This model can be written as

$$\boldsymbol{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon} = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ \lambda_3 & 0 \\ 0 & \lambda_4 \\ 0 & \lambda_5 \\ 0 & \lambda_6 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}, \tag{10}$$

where $\boldsymbol{\Lambda}$ is the matrix of factor loadings and the vector $\boldsymbol{\epsilon}$ contains multivariate normally distributed residuals with covariance matrix $\boldsymbol{\Omega}$ and zero mean vector. The model-implied covariance matrix of observed variables is represented as

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Omega}. \tag{11}$$

Because correlation matrices are used as input data in MASEM, all observed and latent variables have a standardized metric here. This implies that this CFA model has seven free model parameters (e.g., Cudeck, 1989; Jak & Cheung, 2020), namely, six (standardized) factor loadings and the interfactor correlation. Latent factor variances are fixed to 1 and item residuals are a function of the item-specific standardized factor loadings.

However, as our data-generating model also imposed parameter heterogeneity across primary studies, study-specific parameters were sampled from a joint multivariate population distribution with a population mean vector $\boldsymbol{\theta}_0$ and covariance matrix $\mathbf{T}_{\boldsymbol{\theta}}$. As mentioned before, to avoid negative factor loadings and interfactor correlations outside the admissible range, we used logit transformations for loadings and Fisher z-transformations for the interfactor correlation coefficients. The logit-transformed standardized loadings were set to 0.9, 0.6, 0.3, 0.9, 0.6, and 0.3. These values are similar to the values used in the simulation study by Ke et al. (2019), and imply that observed variables were reliable indicators of their respective underlying latent factors with factor loadings (in the back-transformed metric) ranging from .57 to .71. The Fisher z-transformed interfactor correlation was set to 0.3 which corresponds to a positive medium interrelation of around .3 in the back-transformed metric.

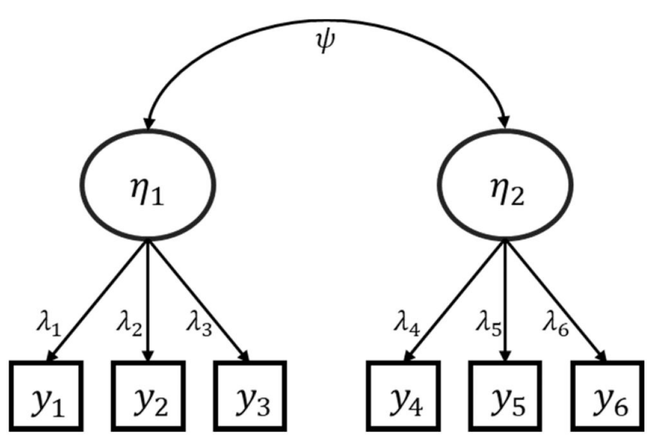

Figure 1. Path diagram of the confirmatory factor analysis model for simulation study 1.

The between-study heterogeneity in SEM parameters was determined by specifying the standard deviations of the random effects, which are given by the square root of the diagonal elements in $\mathbf{T}_{\boldsymbol{\theta}}$. The standard deviations were set to 0.4, 0.6, 0.8, 0.4, 0.6, and 0.8, for the (logit-transformed) loadings, and 0.083 for the residual standard deviation of the (Fisher z-transformed) correlation. These values correspond to small to medium heterogeneity in the back-transformed metric (e.g., Arend & Schäfer, 2019), ranging from around 0.1 to 0.2 for factor loadings, and was around 0.08 for the interfactor correlation. Figures of the random effect distributions in the raw parameter metrics can be found in the Supplemental Material.

We simulated a continuous and dichotomous moderator variable. The two moderators were assumed to be uncorrelated but related to the (Fisher z-transformed) interfactor correlation ψ. The regression coefficients were set to 0.05 for the continuous and −0.05 for the dichotomous moderator. These values represent a strong effect size for the continuous predictor explaining 25% of the between-study variance of the interfactor correlation and a small to medium effect size for the dichotomous predictor explaining 6.25% of the between-study variance (e.g., Hox et al., 2017)[1]. We ran 1000 replications for each simulation condition.

5.2. Design Factors

Three design factors varied across simulation conditions: (1) the number of studies S involved in the meta-analysis, (2) the sample size N of the primary studies involved in a meta-analysis, and (3) the covariance structure of the random effects. The number of studies (S) involved in a meta-analysis was set to 20, 50, 100, and 500. While a meta-analysis containing 20 primary studies represents a scenario for a relatively small research field (e.g., Wang et al., 2022), a typical number of studies for MASEM applications is between 50 and 100 (e.g., Deneault et al., 2023; Rnic et al., 2023; Scherer et al., 2019). A sample size of 500 will probably be attained by only a few empirical applications (but see, e.g., Agostini & van Zomeren, 2021; Möller et al., 2020).

[1]The standardized regression coefficient γ_{CM} of the continuous predictor was 0.5. For the dichotomous predictor, which was 0/1 coded with both categories equally likely, the standardized regression coefficient was $\gamma_{DM} = 0.25$. The residual SD of the interfactor correlation τ_ψ was calculated as $0.1 \cdot \sqrt{1 - (\gamma_{CM}^2 + \gamma_{DM}^2)} = 0.083$.

The sample sizes (N) of the primary studies involved in a simulated meta-analysis were set to 100, 300, and 1000. The sample sizes were not varied within a given meta-analysis. Although this is an unrealistic scenario—typically, in any meta-analysis, there are some primary studies with larger samples and some with smaller samples—these conditions allow us to approximate the effect of having mainly primary studies included that have a small versus medium or large sample.

The correlations of the random effects given by the covariance matrix $\mathbf{T}_{\boldsymbol{\theta}}$ were fixed to 0 and 0.5 (i.e., all off-diagonals of $\mathbf{T}_{\boldsymbol{\theta}}$ were either set to 0 or values that implied a correlation of 0.5). As many existing meta-analyses have shown that study characteristics can be substantially associated with the findings of primary studies, it is also likely that different parameter estimates are correlated across primary studies. By systematically varying the correlation among random effects, it is possible to examine the effect of model assumptions, expecting parameter estimates to be uncorrelated across studies, even though the data-generating model actually contains correlations of random effects that deviate from zero and vice versa.

5.3. Model Specifications and Estimation

The four meta-analytic models were implemented using the Bayesian modeling software Stan (Carpenter et al., 2017; Stan Development Team, 2024) via R (Version 4.4.2; R Core Team, 2024). The model code and additional files can be found on our OSF repository (https://osf.io/tvryz). Posterior means were used as Bayesian point estimates.

To implement the one-stage approach with correlated random effects (OS-CRE), Equation 11 is vectorized using the vech() operator, which stacks the lower triangular of matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ in a column vector (e.g., Ke et al., 2019). In this vectorized form, it is inserted in Equation 8. The prior distributions for the population mean parameters ($\boldsymbol{\theta}_0$) of the loadings and factor correlation, as well as the regression coefficients of the moderators, were set to uninformative normal distribution priors with a mean of zero and standard deviation (SD) of 100. Weakly informative half-Cauchy priors with a location parameter of 0 and a scale parameter of 2.5 were used for the standard deviations of the random effects, and Lewandowski-Kurowicka-Joe (LKJ) priors (Lewandowski et al., 2009) with a shape parameter of 2.0 were used for the correlation matrix of the random effects.

MCMC via Stan's NUTS sampler was used for model estimation (Stan Development Team, 2024). A small preliminary simulation study on model convergence revealed that most of the runs reached a potential scale reduction (PSR; e.g., Gelman et al., 2013) measure below 1.05 and an effective sample size of around 1000 with two parallel chains within 2500 iterations, including 500 warm-up iterations, which implied that chains reached convergence (Zitzmann & Hecht, 2019). Therefore, we relied on this setting for the complete simulation study and evaluated all runs.

The OS-UCRE differs from the OS-CRE only in that the random effects are assumed to be uncorrelated. Consequently, the model specification is almost identical to OS-CRE, with the key simplification that the off-diagonals of the matrix $\mathbf{T}_{\boldsymbol{\theta}}$ in Equation (8) are fixed to zero.

To implement the OS-FE, the vectorized form of Equation (11) was inserted in Equation (4). As this model assumes parameter homogeneity, only population means for item loadings, interfactor correlation, and moderation effects were estimated. Accordingly, prior distributions were only specified for the loadings, the factor correlation, and the regression coefficients of the moderators. The same settings were used as for OS-UCRE and OS-UCRE. Furthermore, we used resampling-based *SE*s for inference on moderation effects in the OS-FE. Since resampling can be computationally demanding and time-consuming under MCMC estimation, we implemented a jackknife procedure based on Bayesian penalized maximum likelihood estimation using Stan's optimization function with default settings to derive the resampling-based *SE*s (Stan Development Team, 2024). For simulation conditions with 100 or 500 studies, we additionally applied the delete-d jackknife procedure (Shao, 1989; Zitzmann et al., 2023) to reduce the number of required replications.

For the TS approach (Equation (9)), we used OpenMx (Boker et al., 2011) to fit the CFA model to each primary-study-specific correlation matrix (i.e., the first-stage analysis). The parameter estimate for the interfactor correlation and its corresponding *SE* were extracted from each of these separately conducted primary-study-specific analyses and kept for the second stage analyses. For the second stage, we implemented a (univariate) random-effects meta-regression model in Stan (see Equation (9), Stage 2). The primary-study-specific correlation estimates served as outcomes and were regressed on the moderators. Again, MCMC estimation was used with 2500 iterations per chain and 500 warm-up iterations. The same specifications for the prior distributions were applied.

5.4. Evaluation Measures

We evaluated the estimated moderation effects from the four MASEM approaches by examining bias, efficiency, coverage rates, and statistical power. Additionally, bias and efficiency measures for all other model parameters can be found on OSF. To evaluate bias, we used a robust bias measure that was computed by subtracting the true simulated population moderation effect from the median estimated moderation effect across all replications for a given model and within a given simulation condition. To facilitate interpretation, we standardized this deviation by dividing it by the true population effect and multiplying it by 100, yielding a relative bias measure. We considered absolute values equal to or greater than 10% as substantial, and values below 10% as negligible bias (Hoogland & Boomsma, 1998; Muthén & Muthén, 2002).

Efficiency was assessed by taking the absolute value of the deviations between the estimated moderation effect and the true simulated effect, and taking the median value of these deviations for a given condition and model. This represents a robust indicator of variability and is less influenced by extreme outliers than the root mean square error

(RMSE). We used this robust RMSE in the present study. To facilitate comparisons of the efficiency of the four MASEM approaches, we computed a relative robust RMSE by expressing each method's efficiency as a percentage of a reference method. In our case, we chose OS-CRE as the reference method. Relative RMSE values greater than 100% indicate lower efficiency compared to OS-CRE, while values below 100% reflect higher efficiency for the respective alternative method under the given simulation condition.

Coverage was evaluated using 95% symmetric credible intervals (CIs) calculated as $\gamma \pm 1.96 \cdot SD_{\gamma}$, where γ denotes the estimated posterior mean of the moderation effect and SD_{γ} represents its posterior *SD* (analogous to the standard error in the frequentist framework, e.g., Gelman et al., 2013). For each condition and model, we calculated the relative frequency that the true parameter fell within the CI. A coverage rate between 91% and 98% is generally considered acceptable (Muthén & Muthén, 2002). As a further robustness check, we also tested Bayesian quantile-based CIs and report the results in Appendix A.

Statistical power was determined by calculating the proportion of replications in which the 95% CI did not include zero across the 1000 replications. A power of at least 80% is typically regarded as adequate in psychological research and social sciences.

5.5. Results

Table 1 shows relative bias, RMSE, and coverage for the moderation effect of the continuous moderator. Overall, bias decreased as both the number of studies and the within-study sample size increased. Across all conditions and estimators, bias remained below 10%. For OS-CRE and OS-UCRE, bias was even consistently below 5%. For OS-FE and TS, a few cells showed mild bias between 5% and 10% when within-study sample sizes were small ($N = 20$) or moderate ($N = 50$).

Efficiency was evaluated using a relative RMSE, which was computed as a ratio of an alternative method's RMSE and the RMSE of OS-CRE. The RMSE for OS-CRE was therefore fixed at 100 across conditions. For the alternative methods (i.e., OS-UCRE, TS, and OS-FE), values below 100 indicate greater efficiency than OS-CRE; values above 100 indicate lower efficiency. Under uncorrelated random effects, OS-UCRE exhibited an almost consistent efficiency advantage over OS-CRE. TS and OS-FE also tended to be less efficient than OS- UCRE in this setting. When random effects were correlated, a reverse pattern was observed. Here, OS-CRE was generally the most efficient method, and the other estimators mostly exhibited values above 100, with only a few exceptions at smaller sample sizes and a small number of studies. Furthermore, coverage was within the desired range for the four methods across all conditions.

Table 2 presents relative bias, relative RMSE, and coverage for the dichotomous moderator. In contrast to the continuous moderator results, bias exceeded 10% in two conditions for OS-UCRE and one condition for OS-FE. However, these substantial biases only occurred when the number of studies was small ($S = 20$). For medium and large numbers of studies, bias was consistently negligible across models.

Regarding RMSE, we observed the same general patterns as for the continuous moderator: OS-UCRE outperformed OS-CRE under the uncorrelated random effects condition,

Table 1. Simulation study 1: relative bias, relative RMSE, and coverage rates for the moderation effects of the continuous moderator.

			Relative Bias				Relative RMSE				Coverage			
Corr	*N*	*S*	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS
0	100	20	−1.7	−0.5	−5.4	−9.1	100	99.7	**98.9**	100.4	96.0	95.9	94.2	96.9
		50	0.0	1.3	−1.4	−3.2	100	**99.0**	109.8	103.9	94.8	95.5	96.2	95.5
		100	1.5	1.1	3.0	−4.0	**100**	**100.0**	106.0	105.8	94.4	95.8	95.0	94.5
		500	−0.1	0.3	−0.7	−4.6	100	**93.6**	102.8	107.7	94.4	95.0	95.8	94.3
	300	20	1.9	2.5	2.4	1.4	100	**88.7**	98.3	92.3	96.0	95.8	94.3	95.9
		50	−2.8	−3.5	−5.7	−4.5	100	**99.5**	101.7	100.4	93.9	95.8	94.5	95.1
		100	−1.3	0.2	−0.4	−1.1	100	**99.5**	101.3	100.3	94.7	95.2	95.6	95.5
		500	−0.1	0.0	−0.2	−1.4	100	**94.9**	103.6	99.5	93.9	94.2	93.3	94.3
	1000	20	−0.9	0.1	−5.0	−1.5	100	97.8	99.8	**97.7**	93.2	95.2	93.3	95.2
		50	4.0	3.1	1.3	2.8	100	**93.6**	98.8	95.6	92.9	94.1	92.5	93.9
		100	0.6	0.2	0.8	0.1	100	**96.0**	102.1	101.2	94.1	94.2	92.9	94.1
		500	0.9	0.6	−0.1	0.8	100	100.1	108.6	**99.0**	94.8	95.9	94.8	95.7
0.5	100	20	3.9	5.0	3.9	−1.3	**100**	102.0	104.4	114.5	97.0	97.2	93.4	96.6
		50	3.2	3.3	2.0	0.3	100	**98.7**	106.6	112.4	95.6	96.4	94.3	94.3
		100	1.1	−1.3	0.3	−3.2	**100**	102.1	104.7	120.0	94.3	94.1	94.2	93.5
		500	−0.7	−1.0	−1.5	−3.6	100	**97.9**	108.0	149.6	96.6	95.6	95.6	92.8
	300	20	−2.2	−1.2	−3.7	−2.6	**100**	109.6	104.9	111.0	94.5	94.2	92.9	93.3
		50	1.7	1.1	−0.6	0.9	**100**	114.0	102.7	113.8	95.6	94.7	93.8	95.2
		100	0.4	−1.0	−0.3	−1.3	**100**	114.3	105.9	111.9	95.3	95.0	94.1	94.3
		500	0.3	−0.3	−0.1	−0.4	**100**	109.4	107.3	113.8	96.7	95.5	95.8	96.0
	1000	20	2.6	2.8	−1.1	3.6	**100**	108.1	104.0	108.0	95.1	95.6	94.7	96.1
		50	1.3	0.4	0.5	0.6	**100**	118.1	109.4	120.2	94.0	92.3	92.6	93.2
		100	−0.5	0.7	−0.6	1.0	**100**	113.6	117.5	114.5	94.1	95.0	93.7	95.6
		500	−0.1	−0.4	−0.7	0.0	**100**	120.0	114.1	117.3	93.8	94.6	95.9	94.8

Note. Corr = correlation of the RE; *N* = sample size within each primary study; *S* = number of primary studies included in the meta-analysis. OS-CRE = one-stage parameter-based MASEM assuming correlated random effects; OS-UCRE one-stage parameter-based MASEM assuming uncorrelated random effects; OS-FE = one-stage fixed effects MASEM with resampling-based standard errors; TS = two-stage parameter-based MASEM.
Relative bias above the threshold of 10% and the lowest relative RMSE for each condition representing best efficiency are printed in bold.

Table 2. Simulation study 1: relative bias, relative RMSE, and coverage rates for the moderation effects of the dichotomous moderator.

			Relative Bias				Relative RMSE				Coverage			
Corr	*N*	*S*	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS
0	100	20	−8.9	**−12.5**	1.5	−5.2	100	**96.4**	107.3	107.5	97.0	97.3	96.1	96.3
		50	−2.3	−3.6	−0.2	1.3	100	**99.0**	104.2	103.3	94.0	95.1	94.8	95.3
		100	3.0	3.5	2.3	2.8	100	**96.7**	101.8	106.5	95.3	94.8	94.7	95.3
		500	1.5	1.1	2.2	6.8	100	**97.7**	107.4	112.0	93.8	94.4	94.1	95.0
	300	20	−8.7	**−11.7**	**−12.1**	−8.6	100	**94.1**	104.4	100.8	94.4	95.0	94.1	94.1
		50	−2.8	−7.5	−6.3	−6.1	100	**93.6**	104.2	95.7	92.2	94.2	94.7	94.2
		100	1.8	3.2	2.3	4.0	100	95.8	98.7	**94.5**	93.6	95.0	91.3	94.7
		500	−0.5	−0.7	−1.4	0.6	**100**	103.8	110.6	100.9	95.5	96.6	95.2	95.4
	1000	20	−3.0	−2.6	2.7	0.1	100	**94.2**	99.5	94.4	91.7	95.2	92.6	94.9
		50	4.9	5.0	6.2	5.9	100	98.5	100.8	**98.1**	94.2	94.6	95.2	94.7
		100	−2.4	−0.9	0.4	−0.2	100	92.7	92.1	**91.5**	92.4	94.0	94.0	94.8
		500	0.9	1.0	1.0	0.9	100	98.3	102.1	**97.3**	95.5	94.8	95.6	95.4
											96.6	96.9	95.7	96.7
0.5	100	20	5.6	8.7	5.6	8.2	100	**99.4**	106.3	108.5	96.6	97.0	96.0	97.6
		50	−5.6	−4.2	−3.2	−1.6	**100**	105.7	102.5	119.9	95.7	95.5	96.6	96.4
		100	1.1	2.7	2.7	0.2	100	**97.4**	105.4	117.9	94.6	95.8	95.2	96.4
		500	−0.7	0.3	−1.3	3.0	**100**	106.4	106.5	158.3	96.2	96.5	94.7	95.8
	300	20	−5.8	0.0	−1.1	−0.5	**100**	105.6	106.5	109.0	95.1	95.6	95.6	95.3
		50	−2.5	−3.4	−1.0	−8.5	**100**	106.5	100.8	107.4	94.7	94.2	93.5	95.2
		100	2.8	2.2	3.2	2.8	**100**	107.8	106.7	111.0	95.3	95.8	95.1	95.4
		500	0.3	1.1	0.7	−0.1	**100**	104.8	108.2	116.1	94.4	94.7	94.0	94.6
	1000	20	2.7	0.8	1.7	−0.7	100	107.9	**94.4**	105.1	94.6	95.4	96.0	95.8
		50	1.6	2.7	1.6	2.9	**100**	113.5	108.5	108.8	96.0	94.3	93.7	94.5
		100	−1.4	0.7	1.0	0.6	**100**	119.9	104.8	121.6	96.6	95.9	95.5	96.0
		500	−0.1	−0.6	0.1	−0.7	**100**	114.3	108.4	117.2	97.0	97.3	96.1	96.3

Note. Corr = correlation of the RE; *N* = sample size within each primary study; *S* = number of primary studies included in the meta-analysis. OS-CRE = one-stage parameter-based MASEM assuming correlated random effects; OS-UCRE one-stage parameter-based MASEM assuming uncorrelated random effects; OS-FE = one-stage fixed effects MASEM with resampling-based standard errors; TS = two-stage parameter-based MASEM.
Relative bias above the threshold of 10% and the lowest relative RMSE for each condition representing best efficiency are printed in bold.

whereas OS-CRE outperformed OS-UCRE under the correlated random effects condition. Compared to the one-stage random effects approaches, TS and OS-FE showed similar patterns as for the continuous moderator—disadvantages under uncorrelated random effects compared to OS-UCRE; disadvantages under correlated random effects compared to OS-CRE. Coverage was again consistently within the expected range for all estimators and conditions.

Figure 2 depicts the statistical power to detect the two moderation effects. While a strong effect was simulated for the continuous moderator, a small to medium effect was simulated for the dichotomous moderator. Power increased with both the within-study sample size and the number of studies. Using 80% as the benchmark, adequate power was reached for the continuous moderator when at least 50 studies were included in meta-analyses and sample size was medium or high. However, under small sample sizes, 80% was only reached in the conditions with at least 100 primary studies per meta-analysis. For the dichotomous moderator, the threshold of 80% was only exceeded when at least 100 studies were included and sample size was high. Under small or medium sample sizes, this threshold was only reached under $S = 500$. Across conditions, TS showed slightly lower power than the OS methods, particularly when the within-study sample size was small ($N = 100$).

5.6. Summary

The results of this simulation study imply that the different parameter-based MASEM approaches can provide approximately unbiased estimates of study-level moderation effects. A considerable bias only occurred for the dichotomous moderator under a small number of primary studies. In these cases, the absolute value of the moderation effect tended to be moderately to substantially underestimated. Comparing the four methods in terms of bias across conditions and types of moderators, no systematic advantages occurred in favor of one approach.

Coverage rates were within the desired range for all models across conditions. This conclusion also held when using Bayesian quantile-based credible intervals instead of symmetric intervals calculated using the posterior *SD* (Appendix A).

The power to detect the strong moderation effect of the continuous moderator exceeded the 80% threshold only when at least 50 primary studies were included in meta-analyses. When the sample size within studies was small, 80% was only reached under 100 and 500 primary studies per meta-analysis. For the small to medium effect of the dichotomous moderator, results showed that adequate power was reached under 100 primary studies and large sample size. However, when the sample sizes within studies were small or medium, 80% were only reached under 500 primary studies. Comparing the different models, slight disadvantages occurred for TS, which were most prominent in the small sample size conditions.

Furthermore, the results showed that efficiency gains of the one-stage approach depend on the correct specification of the random effects structure. Specifically, the one-stage approach with uncorrelated random effects (OS-UCRE) tended to be more efficient when the correlations of the random effects were indeed simulated to be zero. By

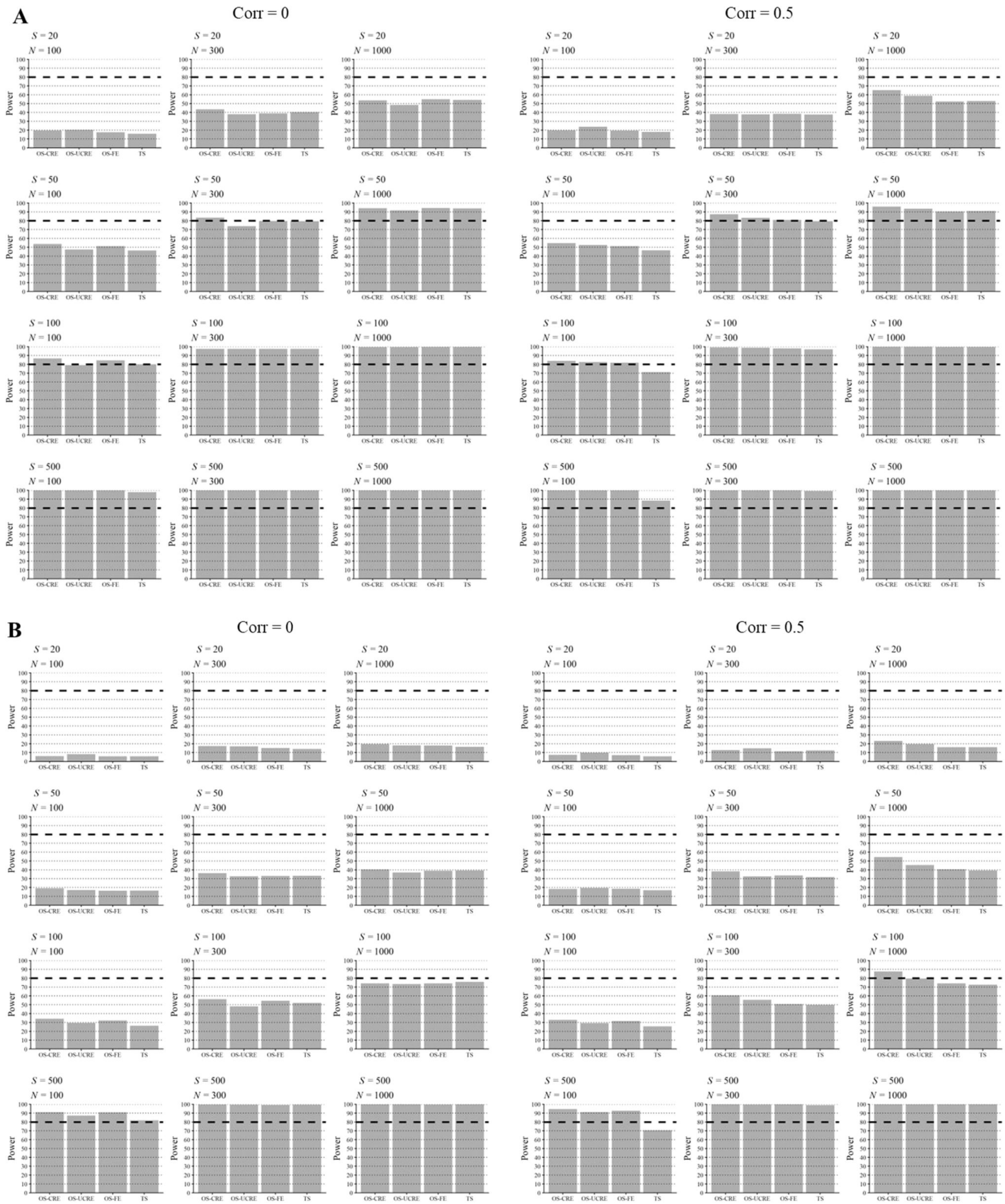

Figure 2. Power to detect the effect of the continuous (A) and the dichotomous moderator (B).

contrast, OS-CRE was more efficient when the random effects were correlated. The same patterns—however less consistent—were apparent when comparing OS-FE and TS against OS-CRE and OS-UCRE, respectively. These results highlight a typical tradeoff: parsimony in the specification of the random effects structure pays when it matches the data-generating structure.

However, the random-effects correlations in the main simulation were relatively unrealistic, as all off-diagonal elements were fixed to either 0 or .50. We therefore conducted an additional simulation varying the random-effects correlation structure in two more realistic ways (see Appendix B for further details): (1) heterogeneous correlations, with off-diagonal elements sampled from a uniform distribution (either [−.20,.20] or [.30,.70]); and (2) sparse correlations, where only 6 of the 21 off-diagonal elements were set to .50 (all others to 0). Results were consistent with the main findings: OS-UCRE was more efficient when random-effects correlations were close to zero, whereas OS-CRE was more efficient when correlations were substantial. Under the sparse-correlation condition, OS-UCRE still tended to be more efficient, suggesting that assuming uncorrelated random effects can be adequate under mild to moderate violations of this assumption.

Moreover, the main simulation assumed homogeneous primary-study sample sizes within each meta-analysis. To assess robustness under more realistic conditions, we conducted an additional simulation with heterogeneous sample sizes drawn from a truncated log-normal distribution (see Appendix C). Conclusions regarding bias, coverage, and relative efficiency remained unchanged under this condition.

In addition, we examined the impact of correlated moderators. Whereas the main simulation assumed independent moderators, we repeated the simulation with correlated moderators by setting the point-biserial correlation to .30 and .70 while holding the raw moderation effects constant (but the R^2 changed accordingly). Results remained largely unchanged with respect to bias, efficiency, and coverage (Appendix D)—the power increased.

6. Simulation Study 2: Random-Intercept First-Order Autoregressive Model

To further demonstrate the flexibility of moderated parameter-based MASEMs, we focused on a longitudinal model in the second simulation study. One particular challenge of the meta-analysis of longitudinal studies is that the design (e.g., number and spacing of measurement waves) affects the interpretation of the study-specific parameters (e.g., stability of variables). Previous work has highlighted the great potential of a parameter-based continuous-time approach for meta-analytically integrating results from different longitudinal studies (Dormann et al., 2020; Kuiper & Ryan, 2018). In the following, we generated data and implemented MASEMs for an RI-AR(1) model (e.g., Skrondal & Rabe-Hesketh, 2008, see also Hamaker et al., 2015). The RI-AR(1) model allows researchers to examine the stability of constructs under a random-intercept variance decomposition.

Many longitudinal meta-analyses aim at quantifying the stability of constructs over time (e.g., Bleidorn et al., 2022; Bühler & Orth, 2022; d'Huart et al., 2023; Scherrer et al., 2025). Consequently, methodological choices about how stability is defined and estimated are directly relevant for the substantive conclusions drawn from such meta-analyses. In recent years, longitudinal modeling has increasingly emphasized the distinction between stable, trait-like between-person differences and within-person fluctuations. Random-intercept approaches such as the RI-CLPM explicitly implement this variance decomposition and thereby allow to study within-person stability and dynamics (Hamaker et al., 2015; see also Usami et al., 2019). This development has motivated a series of re-analyses and comparative investigations showing that conclusions based on traditional panel models may differ once stable between-person variance is controlled (e.g., Hübner et al., 2023; Lüdtke & Robitzsch, 2022; Orth et al., 2021). From a meta-analytic perspective, these developments imply a growing need for models that can synthesize evidence across heterogeneous longitudinal designs while targeting stability with a clear within-person interpretation.

6.1. Method

6.2. Data Generation

The underlying data-generating model was an univariate RI-AR(1) model (see Figure 3). For a longitudinal design with three waves, the measurement part for this model is given as

$$\boldsymbol{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix}, \tag{12}$$

where the matrix $\boldsymbol{\Lambda}$ relates the latent stable trait η_1 and the latent states (η_2, η_3, and η_4) for each wave to the observed measures. This equation reflects that each measurement occasion is decomposed into a stable between-person part and a time-varying within-person part. Note that the errors ($\boldsymbol{\epsilon}$) are set to zero in the measurement part.

The corresponding structural part is defined as

$$\boldsymbol{\eta} = \mathrm{B}\boldsymbol{\eta} + \zeta = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & \beta & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{bmatrix}, \tag{13}$$

where the matrix B contains the (first-order) autoregressive effects β that relate the within-person states over time. As we again assume standardized mean-centered variables, this

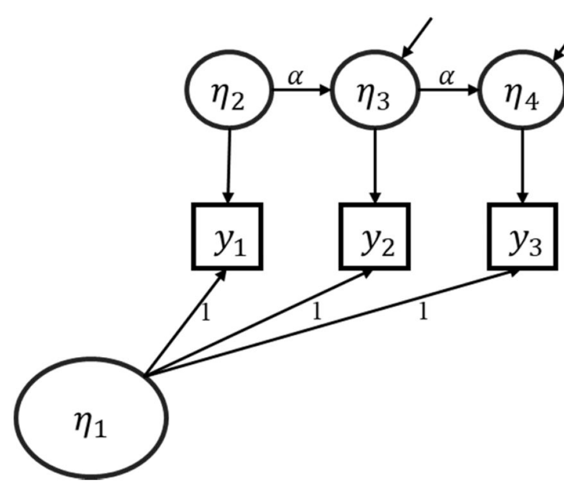

Figure 3. Path diagram of a three-wave random-intercept autoregressive model for simulation study 2.

model has two free model parameters: (1) the trait variance $Var(\zeta_1) = \psi$ representing stable between-person differences across time and (2) the (first-order) autoregressive effect β describing within-person carry-over effects from one measurement occasion to the next (e.g., Usami et al., 2019). Note that the within-person state variance is a function of the between-person variance, that is $Var(\zeta_2) = 1 - \psi$, and the within-level residual variances are a function of the autoregressive effect, that is $Var(\zeta_3) = (1 - \beta^2)\ (1 - \psi)$, and $Var(\zeta_4) = (1 - \beta^2)\ (1 - \psi)$.

In line with Simulation Study 1, these two model parameters were simulated to have population means and population (co-)variances representing parameter heterogeneity across primary studies. The trait variance has an admissible range of [0, 1], where a value of zero indicates that there are no stable between-person differences and a value of one implies that all variance in y is due to stable between-person differences. To limit the trait variance to this admissible range, we used a logit transformation.

We used an exponential transformation for modeling the autoregressive parameter. In meta-analyses of longitudinal studies, different primary studies typically have different time lags (i.e., distances between measurement occasions)—a specific challenge for longitudinal meta-analyses (e.g., Dormann et al., 2020; Kuiper & Ryan, 2020). In these contexts, the autoregressive effect β can be assumed to be a function of the time lag and an estimated underlying constant parameter β^* (Voelkle et al., 2012). This exponential function typically describes decreasing autoregressive effects with growing time lags (i.e., two measurement occasions tend to be less correlated when the time interval in between is larger). As is known from continuous-time modeling, for many dynamic processes, time-varying autoregressive effects can be assumed to follow the exponential function $\beta = e^{\beta^* \Delta t}$ (e.g., Ryan et al., 2018; Voelkle et al., 2012), where Δt represents the distance between two subsequent measurement occasions. A negative value for β^* implies that the association between two measurement occasions is positive and tends to be smaller when time lags increase and asymptotically approaches zero. Smaller estimates of the underlying constant β^* indicate that autoregressive effects tend to decrease faster with growing time lags (Figure 4).

The population mean of the logit-transformed stable trait variance was set to 0, which corresponds to a trait variance of 0.5. Thus, on average, half of the variance was assumed to be due to stable between-person differences and the other half of the variance were intraindividual fluctuations. The autoregressive effect (β) was set to be 0.5 for a time lag Δt of one. This corresponds to a continuous-time constant $\beta^* = -0.7$.

The population *SD* quantifying the between-study differences in the two structural parameters across primary studies was set to 0.364 for logit-transformed stable trait variances, and 0.091 for the autoregressive effect. These values approximately correspond to *SD* in the raw metric of 0.09 and 0.05, respectively, representing small between-study heterogeneity.

In addition, we simulated a continuous moderator that was related to both structural parameters. The two moderation effects were set to 0.12 and 0.03—both representing a

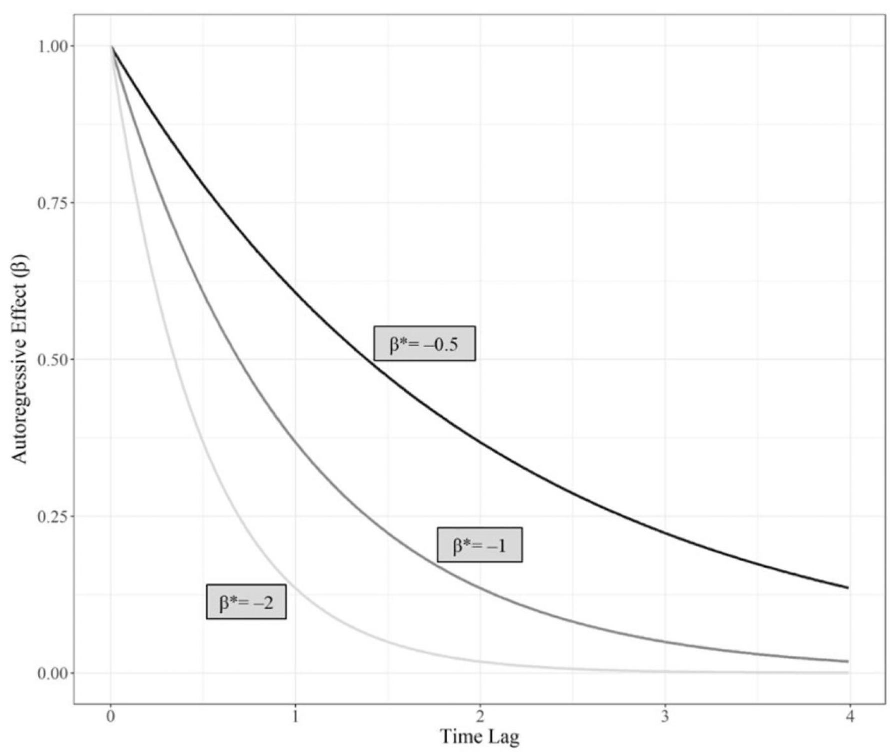

Figure 4. Autoregressive effects as a function of the time lags and the estimated constant β^*.

medium standardized regression coefficient explaining around 10% of the heterogeneity in trait variance and autoregressive effect, respectively (Arend & Schäfer, 2019).

6.3. Design Factors

In line with Simulation Study 1, three main design factors varied across simulation conditions: (1) the number of studies involved in the meta-analysis set to 20, 60, 100, and 500, (2) the sample size N of the primary studies involved in the meta-analysis set to 300 and 1000, (3) and the correlation of the random effects was set to 0 and 0.5.

In addition, we also varied the number of measurement occasions and the length of the time lags between measurement occasions. However, both factors were varied within replications, so that each replication (i.e., meta-analysis) contained primary studies with all combinations of the number of measurement occasions and time lags. The number of measurement occasions was set to 3, 4, 5, and 6, and time lags to 0.5, 0.75, 1, 2, and 3, leading to 20 possible combinations. Systematically crossed, each combination occurred with equal frequency within each meta-analysis. Thus, for example, in the condition with 20 primary studies per meta-analysis (the lowest number of studies condition), each combination occurred once; in the condition with 60 primary studies, each meta-analysis contained three primary studies per combination.

6.4. Model Specifications and Estimation

The same four moderated parameter-based MASEM approaches as tested in Simulation Study 1 were examined: (1) OS-CRE, (2) OS-UCRE, (3) OS-FE, and (4) TS. We used the same implementations as in Study 1; the full model implementations can be found on OSF (https://osf.io/tvryz). Again, we first explored convergence properties in a few preliminary test runs. These tests revealed that PSR < 1.05 and an effective sample size above 1000 were reached for almost all runs after 2500 iterations with two chains, including 500 warm-up iterations (Gelman et al., 2013; Zitzmann & Hecht, 2019). We used this setting for the 1000 replications per model and condition throughout the simulation study and analyzed the results of all runs in the evaluations afterwards. The same evaluation measures were used, as in Simulation Study 1, namely, relative bias, relative RMSE, coverage rates, and power.

6.5. Results

The results for the moderation of the trait variance are summarized in Table 3. Parameter bias generally decreased as both the within-study sample size and the number of studies increased. With one exception, bias was below the threshold of 10%. Specifically, in the condition with correlated random effects, $N = 1000$, and 100 studies, OS-FE exhibited a negative bias of -11.2%.

In terms of efficiency, TS, OS-FE, and OS-UCRE almost consistently yielded values below 100 under the condition that the random effects are uncorrelated. This indicates that these three methods were more efficient than the reference method OS-CRE. Under correlated random effects, OS-CRE showed minor efficiency advantages over OS-UCRE, but disadvantages compared to TS and OS-FE.

Coverage was largely within the expected range across estimators and conditions. A notable exception was that TS occasionally slightly exceeded the lower bound of acceptable coverage when the number of studies was high or very high and random effects were correlated.

Table 4 shows bias, RMSE, and coverage for the moderation effect of the autoregressive effect. No bias above the threshold of 10% was evident when the data-generating model included uncorrelated random effects. However, when comparing the different estimators, TS exhibited larger bias compared to the one-stage approaches.

Several substantial biases emerged for the moderation of the autoregressive effect across estimators under correlated random effects. Primarily, bias occurred when the number

Table 3. Simulation study 2: relative bias, relative RMSE, and coverage rates for the moderation of the trait variance.

			Relative Bias				Relative RMSE				Coverage			
Corr	*N*	*S*	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS
0	300	20	4.0	3.4	4.4	1.4	100	99.7	100.2	**98.7**	96.4	94.9	87.6	95.3
		50	4.1	3.4	−1.5	0.4	100	99.7	**97.3**	98.2	95.5	94.8	92.8	96.0
		100	−0.6	0.0	−3.3	−2.0	100	100.3	**98.6**	99.3	94.5	93.7	93.2	94.2
		500	1.1	0.9	−2.4	−0.6	100	99.9	**98.2**	99.1	95.5	95.0	93.8	94.7
	1000	20	4.8	4.8	−0.3	2.4	100	100.0	**97.6**	98.9	94.2	94.1	89.0	94.3
		50	2.2	2.1	3.7	0.9	100	99.9	100.7	**99.3**	93.5	93.0	91.6	94.4
		100	1.0	1.0	−1.8	0.3	100	100.0	**98.6**	99.6	95.4	95.2	92.9	94.7
		500	0.2	−0.3	−4.2	−1.2	100	99.8	**97.8**	99.3	94.3	94.1	92.6	91.3
0.5	300	20	−0.6	−1.8	−3.3	−9.7	100	99.4	98.7	**95.5**	94.3	94.4	94.4	94.3
		50	6.1	6.4	−0.4	0.1	100	100.2	**96.8**	97.1	96.1	95.9	96.6	94.4
		100	0.6	0.1	−2.9	−4.2	100	99.7	98.2	**97.6**	92.3	94.9	93.5	94.8
		500	2.3	2.4	0.2	0.3	100	100.0	**98.9**	99.0	97.3	97.2	97.1	90.5
	1000	20	−2.9	0.5	3.5	−0.2	**100**	101.7	103.2	101.4	94.2	91.7	94.6	92.9
		50	−0.9	−0.3	−2.0	−3.9	100	100.3	99.4	**98.5**	95.7	96.1	95.5	95.8
		100	−0.6	−0.2	**−11.2**	−4.6	100	100.2	**94.7**	98.0	96.7	94.7	93.1	90.7
		500	0.8	1.3	1.5	1.3	**100**	100.3	100.3	100.3	96.6	96.7	93.7	89.2

Note. Corr = correlation of the RE; N = sample size within each primary study; S = number of primary studies included in the meta-analysis. OS-CRE = one-stage parameter-based MASEM assuming correlated random effects; OS-UCRE one-stage parameter-based MASEM assuming uncorrelated random effects; OS-FE = one-stage fixed effects MASEM with resampling-based standard errors; TS = two-stage parameter-based MASEM.
Relative bias above the threshold of 10% and the lowest relative RMSE for each condition representing best efficiency are printed in bold.

Table 4. Simulation study 2: relative bias, relative RMSE, and coverage rates for the moderation of the autoregressive effect.

			Relative Bias				Relative RMSE				Coverage			
Corr	*N*	*S*	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS
0	300	20	0.7	−2.4	−2.6	0.2	100	**96.8**	121.1	125.9	96.5	96.2	91.7	95.4
		50	−0.4	3.2	−1.1	−6.7	100	**98.0**	118.5	118.9	95.0	94.2	92.8	93.7
		100	−4.7	−4.1	−8.7	**−15.8**	100	**99.8**	124.5	133.5	95.4	94.7	93.5	94.3
		500	−3.3	−1.5	−4.6	−9.4	100	**99.7**	125.3	136.3	94.6	94.3	93.1	91.9
	1000	20	5.3	2.7	−5.5	−0.2	**100**	102.1	136.3	120.1	95.4	95.5	93.5	94.9
		50	−0.1	1.0	−7.3	−7.5	100	**98.1**	136.8	140.1	94.1	94.1	92.2	94.2
		100	−1.9	−1.5	2.1	−9.3	**100**	100.1	135.4	151.3	95.5	95.0	93.5	93.0
		500	1.7	2.5	2.2	−4.9	**100**	101.8	146.6	146.2	94.4	94.1	92.1	92.5
0.5	300	20	**−11.8**	**−11.1**	**−11.2**	−0.5	**100**	101.4	113.7	127.4	97.2	96.5	94.4	96.1
		50	8.3	7.2	6.2	−7.2	**100**	102.1	121.1	145.6	94.0	94.6	95.7	93.7
		100	0.4	−0.1	−2.4	−5.6	**100**	100.4	111.3	121.4	94.4	94.7	94.9	94.8
		500	−0.6	−0.5	−5.2	−4.6	100	**99.1**	119.5	155.8	94.1	94.3	93.9	92.7
	1000	20	−0.5	−1.0	**13.1**	−4.0	100	**97.4**	142.8	127.9	95.2	95.3	95.2	95.5
		50	−2.6	−3.7	−0.5	**−10.5**	100	**99.7**	133.2	135.3	95.7	95.8	95.2	94.7
		100	2.3	1.2	−1.2	−1.6	100	**99.3**	138.1	141.2	95.0	95.7	94.5	94.1
		500	−0.1	−1.1	−3.5	−3.8	100	**99.4**	128.0	145.5	93.4	94.2	93.9	91.3

Note. Corr = correlation of the RE; N = sample size within each primary study; S = number of primary studies included in the meta-analysis. OS-CRE = one-stage parameter-based MASEM assuming correlated random effects; OS-UCRE one-stage parameter-based MASEM assuming uncorrelated random effects; OS-FE = one-stage fixed effects MASEM with resampling-based standard errors; TS = two-stage parameter-based MASEM.
Relative bias above the threshold of 10% and the lowest relative RMSE for each condition representing best efficiency are printed in bold.

of studies was small or medium. For TS, some considerable biases persisted even when the number of studies was high, whereas the OS estimators were essentially unbiased once the number of studies was high or very high.

Furthermore, efficiency patterns for the autoregressive effect also differed from those for the trait variance. OS-CRE and OS-UCRE consistently outperformed TS and OS-FE across conditions. In addition, there were no systematic efficiency differences between the condition with correlated random effects and the condition with uncorrelated random effects. Comparing the two one-stage random-effects approaches, OS-UCRE showed a small but consistent efficiency advantage over OS-CRE.

Coverage rates were generally acceptable across models for the moderation of the autoregressive effect, with a few exceptions in which coverage was modestly reduced.

Figure 5 shows the statistical power to detect a medium-sized moderation of the trait variance and the autoregressive effect, respectively. As could have been expected, power increased with both the number of studies and the within-study sample size. For the trait variance, roughly 100 studies were sufficient to achieve adequate power to detect this medium-sized moderation effect. In contrast, detecting the moderation of the autoregressive parameter required considerably more information. Here, adequate power was only attained under $N = 1000$ and 500 studies for OS-UCRE and OS-CRE.

6.6. Summary

This simulation study evaluated moderated parameter-based MASEM in a longitudinal application with the RI-AR(1) model, focusing on the moderation of the trait variance and the autoregressive parameter. Overall, the results show that moderated MASEM can cover moderation effects with acceptable bias and sufficient coverage under many conditions. However, performance depended strongly on which parameter was moderated and how many primary studies were included in a meta-analysis. All estimators partly exhibited notable bias for the moderation of the autoregressive parameter when the number of studies was small or medium and the random effects were correlated. Across conditions, TS exhibited a somewhat larger bias compared to one-stage approaches, with some substantial biases persisting even under many primary studies per meta-analysis.

Efficiency turned out to be higher for TS and OS-FE across conditions with respect to the moderation of the trait variance. By contrast, for the moderation of the autoregressive effect, OS-CRE and OS-UCRE were more efficient. Comparing the two random effects one-stage approaches, the results were less consistent than in Simulation Study 1. For the moderation of the autoregressive effect, OS-UCRE even achieved efficiency advantages over OS-CRE under correlated random effects.

Finally, power analyses highlighted the data demands of detecting moderation in dynamic parameters. While roughly 100 studies were sufficient to detect a medium-sized moderation of the trait variance reliably, satisfactory power for the moderation of the autoregressive effect was only achieved in the most informative condition we examined ($N = 1{,}000$ and 500 studies).

7. Empirical Example

We reanalyzed an existing meta-analysis to illustrate the application of the four parameter-based approaches evaluated in the simulation studies. Specifically, we used data from the meta-analytic factor analysis of the Multidimensional Scale of Perceived Social Support (MSPSS; Zimet et al., 1988) conducted by Koğar and Yılmaz Koğar (2024). The MSPSS assesses individuals' perceived social support and is designed to measure three dimensions: Family (FAM), Friends (FRI), and Significant Others (SO). In their meta-analysis of the MSPSS factor structure, Koğar and Yılmaz Koğar (2024) found support for the implied three-factor model.

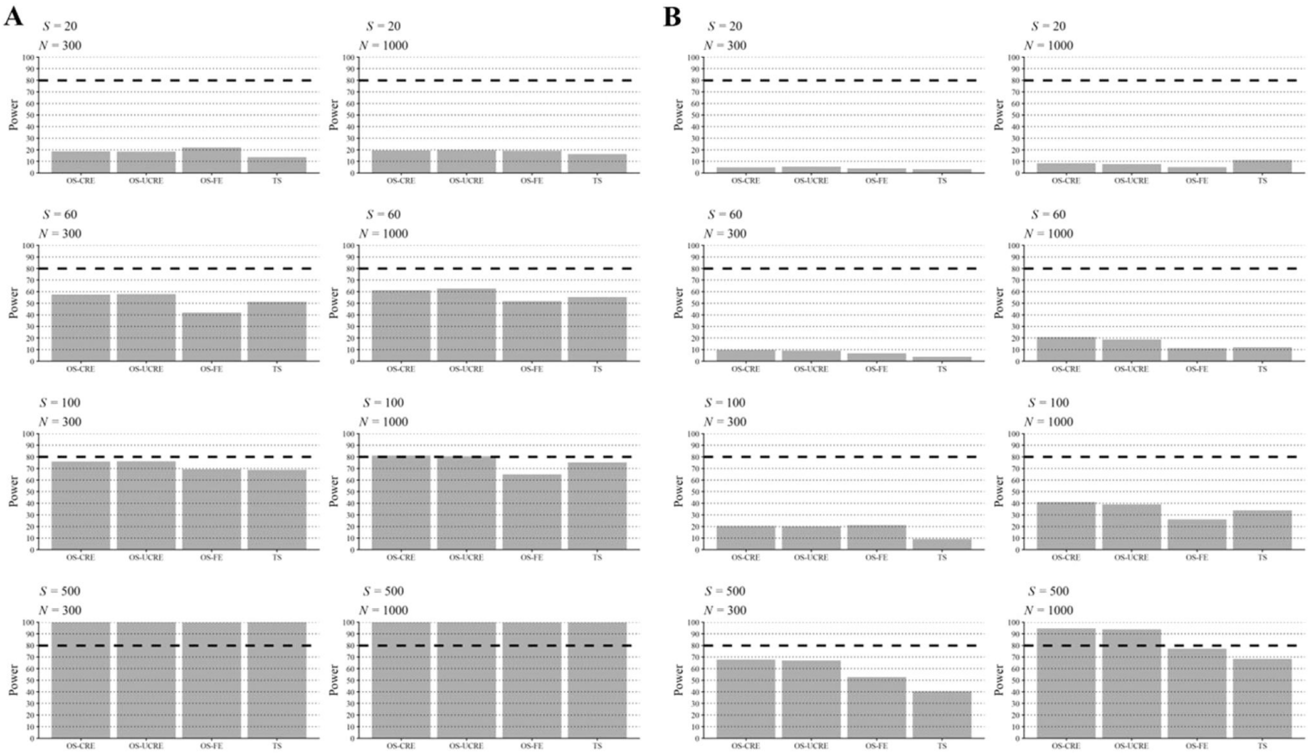

Figure 5. Statistical power to detect a medium sized moderation effect of the trait variance (A) and the autoregressive effect (B).

Table 5. Parameter estimates of the factor correlations and moderator effects in the empirical example (perceived social support scale).

	OS-CRE		OS-UCRE		OS-FE		TS	
Parameter	B (B_{raw})	CI	B (B_{raw})	CI	B (B_{raw})	CI	B (B_{raw})	CI
Intercept								
FAM-FRI	**0.60 (0.54)**	[0.52, 0.68]	**0.58 (0.52)**	[0.50, 0.66]	**0.59 (0.53)**	[0.52, 0.66]	**0.56 (0.51)**	[0.49, 0.63]
FAM-SO	**0.75 (0.64)**	[0.62, 0.89]	**0.72 (0.61)**	[0.58, 0.85]	**0.72 (0.62)**	[0.61, 0.83]	**0.60 (0.54)**	[0.51, 0.70]
FRI-SO	**0.73 (0.63)**	[0.62, 0.85]	**0.72 (0.62)**	[0.61, 0.83]	**0.77 (0.65)**	[0.60, 0.77]	**0.62 (0.55)**	[0.52, 0.73]
Year								
FAM-FRI	−0.01	[−0.08, 0.07]	0.01	[−0.07, 0.09]	−0.05	[−0.14, 0.05]	0.01	[−0.05, 0.07]
FAM-SO	0.05	[−0.07, 0.17]	0.06	[−0.07, 0.19]	0.05	[−0.10, 0.20]	**0.10**	[0.02, 0.18]
FRI-SO	−0.05	[−0.16, 0.05]	−0.08	[−0.19, 0.03]	−0.16	[−0.41, 0.08]	−0.05	[−0.15, 0.04]
Clinical								
FAM-FRI	−0.03	[−0.17, 0.14]	−0.02	[−0.18, 0.15]	−0.05	[−0.27, 0.16]	−0.07	[−0.23, 0.09]
FAM-SO	0.02	[−0.22, 0.29]	0.06	[−0.23, 0.36]	−0.02	[−0.32, 0.29]	−0.01	[−0.21, 0.20]
FRI-SO	0.17	[−0.08, 0.41]	0.15	[−0.11, 0.41]	0.04	[−0.36, 0.43]	0.17	[−0.06, 0.42]

Note. B_{raw} = Factor correlation back-transformed from Fisher z metric. CI = Credible intervals. OS-CRE = one-stage parameter-based MASEM assuming correlated random effects; OS-UCRE = one-stage parameter-based MASEM assuming uncorrelated random effects; OS-FE = one-stage fixed effects MASEM with resampling-based standard errors; TS = two-stage parameter-based MASEM. FAM = Family; FRI = Friends; SO = significant others.

The dataset comprised 59 correlation matrices from 54 primary studies published between 1988 and 2022. For illustration, we fit the three-factor model using the four parameter-based MASEM approaches examined in the simulation studies. In addition, we considered two study-level moderators—year of publication and clinical versus nonclinical sample—and examined whether they moderated the correlations among the three factors (see Simulation Study 1). We included year of publication as a continuous moderator and clinical (coded 1) versus non-clinical (coded 0) sample as a dichotomous moderator.

Not all primary studies reported complete correlation matrices (or they did not report cross-loadings when performing exploratory factor analyses), resulting in missing cells in several study-specific matrices. Rather than imputing zeros for these cells (Koğar & Yılmaz Koğar, 2024), we used a full-information likelihood that conditions on all observed correlations (Jak & Cheung, 2020), while retaining Bayesian MCMC estimation for all parameters in the one-stage models. For TS, we restricted the analysis to 45 correlation matrices without missing correlation coefficients.

The core results are summarized in Table 5. Across all approaches, the estimated correlations between three factors were positive and high. Across MASEM approaches, the correlation between the subscales FRI and FAM was considerably lower compared to the other two correlations involving the factor SO. However, the TS approach obtained noticeably smaller point estimates for the three-factor correlations than the one-stage models.

Across the one-stage MASEMs, none of the moderation effects differed statistically significantly from zero. Small differences in point estimates were observed across the one-stage models. More pronounced discrepancies emerged for TS: The effect of publication year on the FAM-SO factor correlation was statistically significant, indicating higher factor correlations in more recently published studies. All other TS moderation effects were, however, also nonsignificant.

Overall, this reanalysis shows that the different parameter-based approaches can yield broadly similar substantive conclusions but may also produce meaningful differences. A key contributor to the divergence between TS and the one-stage approaches is the handling of missing correlations. The one-stage models used a full-information likelihood and therefore retained studies with partially missing correlation matrices, whereas TS required complete matrices and consequently analyzed a smaller subset of studies. In meta-analyses with more extensive missingness, this distinction may have an even larger impact. The smaller differences observed among the one-stage approaches likely reflect their differing assumptions about the random effects structure (i.e., correlated random effects, uncorrelated random effects, or fixed effects). Accordingly, applied researchers should select the approach that best matches their assumptions about the data-generating process and consider sensitivity analyses (e.g., fitting a less restrictive model allowing correlated random effects) to evaluate the robustness of their conclusions.

8. Discussion

In this article, we have elaborated on parameter-based MASEM as a suitable framework for examining study-level moderators—a primary research objective in many meta-analyses (e.g., Agostini & van Zomeren, 2021; Möller et al., 2020; Scherer et al., 2019; Vasconcellos et al., 2025). So far, the estimation of moderator effects within MASEMs has been systematically evaluated only rarely through simulation studies. To address this gap, we conducted two simulation studies to compare different moderated parameter-based MASEMs across various conditions and under two different structural models.

Overall, the results of our simulation studies revealed that all implemented parameter-based MASEMs provide appropriate coverage rates for moderation effects (and all other model parameters, see the additional materials on OSF) with only a very few exceptions for single models and conditions in Simulation Study 2. Furthermore, Simulation Study 1 showed that all four approaches yield (approximately) unbiased estimates for moderation effects under moderate to large numbers of studies included in meta-analyses. Only for the dichotomous moderator, a few considerable biases occurred when the number of studies was small ($S = 20$). In these cases, the absolute value of the moderation effect tended to be moderately to substantially underestimated. This pattern is in line with small-sample instabilities of binary predictors, where imbalanced category frequency can limit the available information and thus induce bias.

The findings of Simulation Study 2 largely support our findings from the first study in terms of bias. The moderation effects of the trait variance parameter did not exhibit any substantial bias. However, results showed some substantial biases for the moderation effects of the autoregressive effect under correlated random effects. For TS, considerable biases even occurred when the number of studies was large. The observed biases in the moderation of the autoregressive parameter could be explained in different ways.

First, the interpretation of the observed biases must consider the parameterization and scaling of the moderation effects. Although we simulated medium-sized standardized effects for both moderation effects, the raw effects were scaled to the heterogeneity (i.e., *SD* of the random effects) of the target parameter (Arend & Schäfer, 2019). The *SD* of the random effects differed considerably between the two structural parameters: the *SD* of the random effects was larger for the (logit-transformed) trait variance than for the (exponentially transformed) autoregressive parameter ($SD = 0.4$ vs. 0.1). Therefore, the raw moderation effect for the autoregressive parameter was numerically much smaller. Consequently, a given percentage bias (e.g., 10% relative bias) corresponds to a raw bias value that is four times smaller for the moderation of the autoregressive parameter than for the moderation of the feature variance.

Second, previous simulation studies on longitudinal state-traits models have pointed out that estimates of the trait variance and the autoregressive effect can be highly dependent, particularly in conditions with a substantial autoregressive effect (Lüdtke et al., 2018; Usami et al., 2019). Thus, when biases in the trait variance occur, this also likely imposes biases on the autoregressive effect, with particular difficulties in reliably estimating the autoregressive effect when the trait variance is close to its boundaries (i.e., close to zero or one). Similar patterns can be found for the point estimates of the autoregressive effect in our simulation results (see the additional materials on OSF). The stronger biases observed under the correlated random effects condition in our simulation study further underscores this interpretation and might have been exacerbated by the fact that our data-generating model included just one moderator with the same positive effect on both parameters.

Comparing the different estimators in terms of power, slight disadvantages occurred for TS, which were most prominent in the small sample size conditions. This is consistent with the results of other simulation studies comparing TS and OS approaches in the context of multilevel models (Liu, 2017; Lohmann et al., 2025; Lüdtke et al., 2008). Under several conditions, one-stage methods can perform better than TS methods because they can use hierarchical information more efficiently.

Finally, one-stage approaches to MASEM have attracted considerable attention in recent years as they combine the possibility of examining categorical and continuous moderators of structural parameters with the capability of accounting for missing correlation coefficients in certain primary studies (Ke et al., 2025). Thereby, one-stage approaches allow for less restrictive inclusion criteria in meta-analyses. In the present study, we focused on parameter-based OS-

MASEM as it aligns quantification of heterogeneity and moderation at the level of structural parameters, allowing researchers to calculate the relative importance of a moderator for specific structural parameters via a coefficient of determination. This advantage does not extend to fixed effects models that assume homogeneity. Since homogeneity is often implausible in practice, we have equipped the fixed effect approach with a resampling strategy to enable robust standard errors and valid statistical conclusions even in the presence of heterogeneity. Our simulation results underscore the potential of this approach.

9. Limitations and Future Directions

While the current article provides an in-depth evaluation of moderated parameter-based MASEM, some limitations should be acknowledged. These can also help to identify several possible pathways for future methodological development and evaluations in the context of (moderated) MASEM.

First, in the simulation studies presented, we limited ourselves to four different versions of parameter-based MASEMs. Correlation-based OS-MASEM also allows for analyzing continuous and categorical study-level moderators (Jak & Cheung, 2020), and researchers might be interested in comparisons between these two methods. However, as noted by Jak and Cheung (2020), both types of MASEM assume different data-generating models, posing challenges for direct comparisons between the two approaches. One possibility for future research could be to generate data from both models and evaluate the robustness of the approaches. Another possibility pertains to model development, which could explore whether both models can be unified under a general conceptualization of MASEM (Cheung, 2015; Ke et al., 2025).

Second, we did not examine the impact of missing correlation coefficients in the simulation studies. Such missing correlations can arise, for example, when variables relevant to the model of interest were not measured in every primary study (see our empirical example). One-stage approaches (Ke et al., 2019) allow direct estimation of MASEM under the assumption that correlations are missing at random. Alternatively, multiple imputation methods can be used to impute missing correlation coefficients using an appropriate model (e.g., a multivariate normal model), after which the TS approach can be applied to the multiply imputed study-specific correlation matrices. Initial simulation results regarding the effect of missing correlation coefficients in (unmoderated) OS-MASEM versus TS are available in the literature (Jak & Cheung, 2020). A related issue concerns the treatment of missing values in moderator variables (Lee & Beretvas, 2023). A promising topic for future research is the extension of one-stage approaches to estimate moderator effects in MASEMs in the presence of incomplete correlation matrices and missing moderator variables (Enders, 2022; Gelman et al., 2013).

Third, we aimed at exploring and demonstrating the flexibility of parameter-based MASEM for meta-analytic longitudinal research. In recent years, several debates have addressed the robustness of longitudinal evidence under different longitudinal models (Hamaker, 2023; Lüdtke & Robitzsch, 2022; Orth et al., 2021; Usami et al., 2019). Meta-analytically reanalyzing longitudinal studies with different longitudinal models might be a highly relevant endeavor in this context. In Simulation Study 2, we set up such a longitudinal meta-analytic model for an univariate RI-AR(1), which accounts for differences in the number of waves and the length of time lags between primary studies (Kuiper & Ryan, 2020). The results of our simulation study have provided initial evidence for the possibility of specifying and estimating advanced longitudinal meta-analytic models with parameter-based MASEM. However, some results patterns indicated that bias can occur, and further evaluations should be performed in future research. Furthermore, to fully exploit the potential for longitudinal meta-analyses, multivariate extensions such as a meta-analytic RI-CLPM (Dormann et al., 2020; see also Wiese et al., 2025) or the STARTS model (Kenny & Zautra, 1995) could be developed and evaluated in future research.

Fourth, providing extensive simulation studies, we addressed the current lack of systematic evaluations of (moderated) parameter-based MASEMs. However, the lack of systematic evaluation could be just one of the reasons why parameter-based approaches have been little used in applied research to date. Another obstacle to a broader application of parameter-based approaches might be a more complex model implementation. While there is the MetaSEM package (Cheung, 2014) for the implementation of correlation-based approaches, there is no comparable alternative for parameter-based MASEMs.

10. Conclusion

In this article, we discussed and evaluated different moderated parameter-based MASEM approaches. The findings of two simulation studies highlight that all tested models can yield unbiased estimates and appropriate coverage rates under realistic conditions. However, results also imply that bias and power requirements can vary substantially depending on the sample size conditions, the structural model, and the model parameters being moderated. One-stage random effects models showed comparably strong performance across conditions and model parameters. Considering further strengths of one-stage parameter-based MASEM discussed in the article, it holds considerable potential for future multivariate meta-analyses, including longitudinal studies. We hope that our article will support practical applications of moderated parameter-based MASEM in the future.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Agostini, M., & van Zomeren, M. (2021). Toward a comprehensive and potentially cross-cultural model of why people engage in collective action: A quantitative research synthesis of four motivations and structural constraints. *Psychological Bulletin*, *147*, 667–700. https://doi.org/10.1037/bul0000256

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods*, *24*, 1–19. https://doi.org/10.1037/met0000195

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Becker, B. J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics*, *17*, 341–362. https://doi.org/10.3102/10769986017004341

Becker, B. J., & Wu, M.J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science*, *22*, 414–429. https://doi.org/10.1214/07-STS243

Bleidorn, W., Schwaba, T., Zheng, A., Hopwood, C. J., Sosa, S. S., Roberts, B. W., & Briley, D. A. (2022). Personality stability and change: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *148*, 588–619. https://doi.org/10.1037/bul0000365

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P., & Fox, J. (2011). Openmx: An open source extended structural equation modeling framework. *Psychometrika*, *76*, 306–317. https://doi.org/10.1007/s11336-010-9200-6

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley. https://doi.org/10.1002/9781118619179

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences.* (pp. 185–249). Springer US. https://doi.org/10.1007/978-1-4899-1292-3_4

Bühler, J. L., & Orth, U. (2022). Rank-order stability of relationship satisfaction: A meta-analysis of longitudinal studies. *Journal of Personality and Social Psychology*, *123*, 1138–1165. https://doi.org/10.1037/pspp0000430

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1–32. https://doi.org/10.18637/jss.v076.i01

Cheung, M. W. L. - (2014). Metasem: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, *5*, 1521. https://doi.org/10.3389/fpsyg.2014.01521

Cheung, M. W. L. - (2015). *Meta-analysis: A structural equation modeling approach*, Wiley. https://doi.org/10.1002/9781118957813

Cheung, M. W. L., & Chan, W. (2004). Testing dependent correlation coefficients via structural equation modeling. *Organizational Research Methods*, *7*, 206–223. https://doi.org/10.1177/1094428104264024

Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, *10*, 40–64. https://doi.org/10.1037/1082-989X.10.1.40

Cheung, M. W. L., & Chan, W. (2009). A two-stage approach to synthesizing covariance matrices in meta-analytic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 28–53. https://doi.org/10.1080/10705510802561295

Cheung, M. W. L., & Cheung, S. F. (2016). Random-effects models for meta-analytic structural equation modeling: Review, issues, and illustrations. *Research Synthesis Methods*, *7*, 140–155. https://doi.org/10.1002/jrsm.1166

Cheung, M. W. L., & Hafdahl, A. R. (2016). Special issue on meta-analytic structural equation modeling: Introduction from the guest editors. *Research Synthesis Methods*, *7*, 112–120. https://doi.org/10.1002/jrsm.1212

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, *105*, 317–327. https://doi.org/10.1037/0033-2909.105.2.317

Dapp, L. C., Krauss, S., & Orth, U. (2023). Testing the bottom-up and top-down models of self-esteem: A meta-analysis of longitudinal studies. *Journal of Personality and Social Psychology*, *124*, 1111–1131. https://doi.org/10.1037/pspp0000444

Deneault, A. A., Duschinsky, R., van IJzendoorn, M. H., Roisman, G. I., Ly, A., Fearon, R. P., & Madigan, S. (2023). Does child-mother attachment predict and mediate language and cognitive outcomes? A series of meta-analyses. *Developmental Review*, *70*, 101093. https://doi.org/10.1016/j.dr.2023.101093

d'Huart, D., Seker, S., Bürgin, D., Birkhölzer, M., Boonmann, C., Schmid, M., & Schmeck, K. (2023). The stability of personality disorders and personality disorder criteria: A systematic review and meta-analysis. *Clinical Psychology Review*, *102*, 102284. https://doi.org/10.1016/j.cpr.2023.102284

Dormann, C., Guthier, C., & Cortina, J. M. (2020). Introducing continuous time meta-analysis (CoTiMA). *Organizational Research Methods*, *23*, 620–650. https://doi.org/10.1177/1094428119847277

Enders, C. K. (2022). *Applied missing data analysis*. Guilford Publications.

Furlow, C. F., & Beretvas, S. N. (2005). Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data. *Psychological Methods*, *10*, 227–254. https://doi.org/10.1037/1082-989X.10.2.227

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis. Chapman and Hall/CRC, https://doi.org/10.1201/b16018

Hafdahl, A. R. (2008). Combining heterogeneous correlation matrices: Simulation analysis of fixed-effects methods. *Journal of Educational and Behavioral Statistics*, *33*, 507–533. https://doi.org/10.3102/1076998607309472

Hamaker, E. L. (2023). The within-between dispute in cross-lagged panel research and how to move forward. *Psychological Methods*, Advance online publication. https://doi.org/10.1037/met0000600

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116. https://doi.org/10.1037/a0038889

Hedges, L. V., & Olkin, I. (1981). The asymptotic distribution of commonality components. *Psychometrika*, *46*, 331–336. https://doi.org/10.1007/BF02293739

Heisig, J. P., & Schaeffer, M. (2019). Why you should always include a random slope for the lower-level variable involved in a cross-level interaction. *European Sociological Review*, *35*, 258–279. https://doi.org/10.1093/esr/jcy053

Hoogland, J. J., & Boomsma, A. N. N. E. (1998). Robustness studies in covariance structure modeling. *Sociological Methods & Research*, *26*, 329–367. https://doi.org/10.1177/0049124198026003003

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis*, Routledge. https://doi.org/10.4324/9781315650982

Hoyle, R. H. (Ed.) (1995). *Structural equation modeling: Concepts, issues, and applications*. Sage Publications, Inc.

Hübner, N., Wagner, W., Zitzmann, S., & Nagengast, B. (2023). How strong is the evidence for a causal reciprocal effect? Contrasting traditional and new methods to investigate the reciprocal effects model of self-concept and achievement. *Educational Psychology Review*, *35*, 1–45. https://doi.org/10.1007/s10648-023-09724-6

Jak, S., & Cheung, M. W. L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods*, *25*, 430–455. https://doi.org/10.1037/met0000245

Jöreskog, K. G., Olsson, U. H., & Wallentin, F. Y. (2016). *Multivariate analysis with LISREL*. Springer.

Ke, Z., Du, H., Cheung, R. Y. M., Liang, Y., Liu, J., & Chen, W. (2025). Quantifying and explaining heterogeneity in meta-analytic structural equation modeling: Methods and illustrations. *Behavior Research Methods*, *57*, 131. https://doi.org/10.3758/s13428-025-02647-w

Ke, Z., Zhang, Q., & Tong, X. (2019). Bayesian meta-analytic SEM: A one-stage approach to modeling between-studies heterogeneity in structural parameters. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 348–370. https://doi.org/10.1080/10705511.2018.1530059

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, *63*, 52–59. https://doi.org/10.1037/0022-006X.63.1.52

Koğar, H., & Yılmaz Koğar, E. (2024). The structure of the multidimensional scale of perceived social support: A meta-analytic confirmatory factor analysis. *Global Mental Health (Cambridge, England)*, *11*, e126. https://doi.org/10.1017/gmh.2024.118

Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 809–823. https://doi.org/10.1080/10705511.2018.1431046

Kuiper, R. M., & Ryan, O. (2020). Meta-analysis of lagged regression models: A continuous-time approach. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*, 396–413. https://doi.org/10.1080/10705511.2019.1652613

Lee, J., & Beretvas, S. N. (2023). Comparing methods for handling missing covariates in meta-regression. *Research Synthesis Methods*, *14*, 117–136. https://doi.org/10.1002/jrsm.1585

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*, 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Liu, S. (2017). Person-specific versus multilevel autoregressive models: Accuracy in parameter estimates at the population and individual levels. *The British Journal of Mathematical and Statistical Psychology*, *70*, 480–498. https://doi.org/10.1111/bmsp.12096

Lohmann, J. F., Machts, N., Möller, J., & Zitzmann, S. (2025). A more comprehensive, more reliable multilevel approach for assessing and modeling teacher judgment accuracy using latent variables. *Educational Psychology Review*, *37*. https://doi.org/10.1007/s10648-025-10029-z

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203–229. https://doi.org/10.1037/a0012869

Lüdtke, O., & Robitzsch, A. (2022). A comparison of different approaches for estimating cross-lagged effects from a causal inference perspective. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*, 888–907. https://doi.org/10.1080/10705511.2022.2065278

Lüdtke, O., Robitzsch, A., & Wagner, J. (2018). More stable estimation of the STARTS model: A Bayesian approach using Markov chain Monte Carlo techniques. *Psychological Methods*, *23*, 570–593. https://doi.org/10.1037/met0000155

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *The British Journal of Mathematical and Statistical Psychology*, *37*, 234–251. https://doi.org/10.1111/j.2044-8317.1984.tb00802.x

Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A meta-analysis of relations between achievement and self-concept. *Review of Educational Research*, *90*, 376–419. https://doi.org/10.3102/0034654320919354

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Olkin, I., & Finn, J. D. (1995). Correlations redux. *Psychological Bulletin*, *118*, 155–164. https://doi.org/10.1037/0033-2909.118.1.155

Oort, F. J., & Jak, S. (2016). Maximum likelihood estimation in meta-analytic structural equation modeling. *Research Synthesis Methods*, *7*, 156–167. https://doi.org/10.1002/jrsm.1203

Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*, *120*, 1013–1034. https://doi.org/10.1037/pspp0000358

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.4.2) [Computer software]. https://www.R-project.org/

Rao, J. N. K., & Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, *83*, 231–241. https://doi.org/10.1080/01621459.1988.10478591

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage Publications, Inc.

Raykov, T., & Marcoulides, G. A. (2012). *A first course in structural equation modeling*. Routledge. https://doi.org/10.4324/9780203930687

Rnic, K., Santee, A. C., Hoffmeister, J. A., Liu, H., Chang, K. K., Chen, R. X., Neufeld, R. W. J., Machado, D. A., Starr, L. R., Dozois, D. J. A., & LeMoult, J. (2023). The vicious cycle of psychopathology and stressful life events: A meta-analytic review testing the stress generation model. *Psychological Bulletin*, *149*, 330–369. https://doi.org/10.1037/bul0000390

Robitzsch, A., & Lüdtke, O. (2025). A note on the occurrence of the illusory between-person component in the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal*, *32*, 36–45. https://doi.org/10.1080/10705511.2024.2379495

Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). A continuous-time approach to intensive longitudinal data: What, why, and how? In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous time modeling in the behavioral and related sciences.* (pp. 27–54). Springer International Publishing. https://doi.org/10.1007/978-3-319-77219-6_2

Scherer, R., & Campos, D. G. (2022). Measuring those who have their minds set: An item-level meta-analysis of the implicit theories of intelligence scale in education. *Educational Research Review*, *37*, 100479. https://doi.org/10.1016/j.edurev.2022.100479

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, *128*, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009

Scherrer, V., Breit, M., & Preckel, F. (2025). The stability of students' academic achievement in school: A meta-analysis of longitudinal studies. *Educational Research Review*, *48*, 100687. https://doi.org/10.1016/j.edurev.2025.100687

Shao, J. (1989). The efficiency and consistency of approximations to the jackknife variance estimators. *Journal of the American Statistical Association*, *84*, 114–119. https://doi.org/10.1080/01621459.1989.10478745

Skrondal, A., & Rabe-Hesketh, S. (2008). Multilevel and related models for longitudinal data. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis.* (pp. 275–299). Springer New York. https://doi.org/10.1007/978-0-387-73186-5_7

Stan Development Team. (2024). *RStan: the R interface to Stan* (Version 2.32.6) [Computer software]. https://mc-stan.org/

Steiger, J. H., & Hakstian, A. R. (1982). The asymptotic distribution of elements of a correlation matrix: Theory and application. *British Journal of Mathematical and Statistical Psychology*, *35*, 208–215. https://doi.org/10.1111/j.2044-8317.1982.tb00653.x

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*, 637–657. https://doi.org/10.1037/met0000210

Usami, S., Todo, N., & Murayama, K. (2019). Modeling reciprocal effects in medical research: Critical discussion on the current practices and potential alternative models. *PLoS One*, *14*, e0209133. https://doi.org/10.1371/journal.pone.0209133

Vasconcellos, R. P., Sanders, T., Lonsdale, C., Parker, P., Conigrave, J., Tang, S., Del Pozo Cruz, B., Biddle, S. J. H., Taylor, R., Innes-Hughes, C., Salmela-Aro, K., Vasconcellos, D., Wilhite, K., Tremaine, E., Booker, B., & Noetel, M. (2025). Electronic screen use and children's socioemotional problems: A systematic review and meta-analysis of longitudinal studies. *Psychological Bulletin*, *151*, 513–543. https://doi.org/10.1037/bul0000468

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, *48*, 865–885. https://doi.org/10.1111/j.1744-6570.1995.tb01784.x

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*, 176–192. https://doi.org/10.1037/a0027543

Wang, R., Liu, Q., & Zhang, W. (2022). Coping, social support, and family quality of life for caregivers of individuals with autism: Meta-analytic structural equation modeling. *Personality and Individual Differences*, *186*, 111351. https://doi.org/10.1016/j.paid.2021.111351

Wiese, C. W., Dormann, C., Vaziri, H., Tay, L., Wille, B., Chen, J., Moran, L. H., & Li, Y. (2025). Happy work, happy life? A replication and comparison of the longitudinal effects between job and life satisfaction using continuous time meta-analysis. *Journal of Organizational Behavior*, *46*, 487–511. https://doi.org/10.1002/job.2861

Yuan, KH., & Bentler, P. M. (2006). Structural equation modeling. *In Handbook of Statistics. Psychometrics*, *26*, 297–358. (Elsevier. https://doi.org/10.1016/S0169-7161(06)26010-3

Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The multidimensional scale of perceived social support. *Journal of Personality Assessment*, *52*, 30–41. https://doi.org/10.1207/s15327752jpa5201_2

Zitzmann, S., & Hecht, M. (2019). Going beyond convergence in Bayesian estimation: Why precision matters too and how to assess it. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 646–661. https://doi.org/10.1080/10705511.2018.1545232

Zitzmann, S., Weirich, S., & Hecht, M. (2023). Accurate standard errors in multilevel modeling with heteroscedasticity: A computationally more efficient jackknife technique. *Psych*, *5*, 757–769. https://doi.org/10.3390/psych5030049

Appendix A:

Coverage Rates Based on Bayesian Quantile-based 95% Credible Intervals

Table A1. Bayesian quantile-based 95% credible intervals.

			Continuous Moderator				Dichotomous Moderator			
Corr	*N*	*S*	OS-CRE	OS-UCRE	OS-FE[1]	TS	OS-CRE	OS-UCRE	OS-FE[1]	TS
0	100	20	96.1	96.0	94.2	96.9	97.3	97.1	96.1	96.4
		50	94.6	95.6	96.2	95.6	94.0	95.4	94.8	95.6
		100	93.4	95.6	95.0	94.2	94.6	95.2	94.7	95.3
		500	94.4	95.4	95.8	95.1	93.8	94.3	94.1	94.4
	300	20	95.4	96.6	94.3	96.6	94.3	95.4	94.1	94.3
		50	93.9	95.9	94.5	95.4	91.6	93.9	94.7	94.1
		100	94.3	95.3	95.6	96.2	95.0	96.2	91.3	95.5
		500	93.5	94.1	93.3	94.7	95.7	96.3	95.2	95.1
	1000	20	93.0	95.1	93.3	94.8	92.1	95.1	92.6	95.3
		50	93.0	94.8	92.5	94.5	93.8	94.2	95.2	94.2
		100	94.4	94.7	92.9	94.6	93.1	94.0	94.0	94.6
		500	94.8	95.8	94.8	95.6	95.2	95.1	95.6	95.2
0.5	100	20	95.9	95.9	93.4	96.1	96.2	96.4	95.7	96.4
		50	95.2	96.0	94.3	94.3	96.3	96.6	96.0	97.5
		100	93.8	93.8	94.2	94.1	96.0	96.2	96.6	96.5
		500	95.4	95.4	95.6	93.1	93.8	94.4	95.2	96.1
	300	20	93.9	94.4	92.9	93.9	96.2	97.0	94.7	96.5
		50	95.5	94.5	93.8	94.8	95.1	95.7	95.6	95.2
		100	94.5	94.7	94.1	94.4	93.6	93.5	93.5	94.1
		500	96.4	95.3	95.8	96.1	95.1	95.8	95.1	95.3
	1000	20	95.0	95.1	94.7	95.4	95.0	95.3	94.0	95.3
		50	94.4	92.9	92.6	94.0	94.0	95.0	96.0	95.3
		100	94.3	95.1	93.7	95.1	96.1	95.1	93.7	94.8
		500	93.5	95.0	95.9	93.8	95.9	96.0	95.5	96.2

[1]Coverage rates for OS-FE calculated via symmetric Cis as derived from resampling.

Appendix B:

Additional Simulation Results under Heterogenous Correlations of the Random Effects

Table A2. Relative bias, relative RMSE, and coverage rates.

				Relative Bias				Relative RMSE				Coverage			
S	*N*	Cor	*RE*	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS
20	300	full	0	−2.2	−1.2	−3.7	−2.6	**100**	109.6	104.9	111.0	94.5	94.2	92.9	93.3
			0.1	−1.5	0.0	2.4	−1.5	**100**	110.5	108.1	104.5	96.0	96.4	94.8	95.9
		block	0	2.7	−0.1	0.9	1.5	100	97.0	**94.6**	96.6	95.3	96.0	95.2	96.1
			0.1	0.7	−1.0	−0.4	0.5	100	**93.9**	97.0	96.6	94.2	95.0	93.4	95.0
	1000	full	0	2.6	2.8	−1.1	3.6	**100**	108.1	104.0	108.0	95.1	95.6	94.7	96.1
			0.1	−5.0	−3.7	−4.4	−4.5	**100**	118.6	113.9	122.4	96.0	95.0	94.5	94.9
		block	0	−2.4	−2.4	−5.9	−1.6	100	89.9	93.0	**89.8**	93.5	94.9	93.7	95.5
			0.1	−1.9	−0.8	−1.1	−1.7	100	**91.0**	93.0	93.2	92.9	94.4	92.6	94.5
50	300	full	0	1.7	1.1	−0.6	0.9	**100**	114.0	102.7	113.8	95.6	94.7	93.8	95.2
			0.1	−0.3	−2.1	−1.8	−1.8	**100**	107.0	104.7	106.8	95.3	95.4	94.1	95.3
		block	0	−0.7	−3.2	−4.1	−2.5	100	99.3	98.5	**97.0**	92.3	94.5	94.3	94.5
			0.1	−1.9	−4.2	−4.1	−4.3	100	97.8	98.6	**95.5**	93.1	95.5	94.4	95.4
	1000	full	0	1.3	0.4	0.5	0.6	**100**	118.1	109.4	120.2	94.0	92.3	92.6	93.2
			0.1	1.2	0.6	0.0	1.5	**100**	128.6	116.0	127.4	94.9	94.4	95.2	94.8
		block	0	1.6	1.0	−1.2	1.1	100	**94.9**	103.6	97.9	92.9	94.1	94.4	94.8
			0.1	−2.4	−1.9	−4.8	−1.7	100	**91.8**	95.7	96.4	93.7	94.2	93.5	94.6
100	300	full	0	0.4	−1.0	−0.3	−1.3	**100**	114.3	105.9	111.9	95.3	95.0	94.1	94.3
			0.1	−1.2	−2.0	−1.7	−2.4	**100**	110.5	107.2	111.2	95.4	95.0	93.8	94.8
		block	0	0.8	−1.4	−1.5	−1.0	100	**92.4**	100.6	94.6	94.8	95.6	94.4	95.5
			0.1	−0.5	−1.9	−2.2	−2.1	100	**94.4**	102.0	97.2	93.7	95.1	94.3	95.2
	1000	full	0	−0.5	0.7	−0.6	1.0	**100**	113.6	117.5	114.5	94.1	95.0	93.7	95.6
			0.1	0.1	0.0	−0.3	0.7	**100**	126.9	128.0	127.8	95.3	94.6	94.8	95.3
		block	0	0.6	−1.4	−3.2	−1.1	100	**98.6**	107.8	98.8	94.5	95.8	94.1	95.8
			0.1	−0.3	−1.9	−3.3	−1.3	**100**	104.0	108.1	102.7	94.2	95.6	93.9	95.8

Note. Cor = structure of the correlation matrix of the random effects (full = all cells have the same value/are sampled from the same value range; block = only correlations between item loading of item belonging to the same latent factor have correlations that deviate from zero); RE = indicates whether all cells in the correlation matrix were fixed to a certain value (i.e., either 0 or .50) or whether values were sampled from a given range of values (i.e., either [−.20,.20] or [.30,.70]). S = number of primary studies per meta-analysis; OS-CRE = one-stage parameter-based MASEM assuming correlated random effects; OS-UCRE one-stage parameter-based MASEM assuming uncorrelated random effects; OS-FE = one-stage fixed effects MASEM with resampling-based standard errors; TS = two-stage parameter-based MASEM.

Appendix C:

Additional Simulation Results under Heterogenous Sample Sizes

Table A3. Relative bias, relative RMSE, and coverage rates under heterogenous sample sizes.

			Relative Bias				Relative RMSE				Coverage			
Mod	*S*	SD_N [1]	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS
CM	20	0	−1.7	−0.5	−5.4	−9.1	100	99.7	**98.9**	100.4	96.0	95.9	94.2	96.9
		100	−2.5	−1.1	−1.0	−2.8	100	**98.9**	100.1	101.4	94.0	95.0	95.0	95.5
		240	−1.2	−0.7	−0.3	−2.4	100	99.7	107.2	**98.4**	94.1	95.6	93.6	95.7
	50	0	0.0	1.3	−1.4	−3.2	100	**99.0**	109.8	103.9	94.8	95.5	96.2	95.5
		100	0.2	−0.3	−0.2	−1.1	100	97.1	107.3	**96.4**	93.5	95.3	93.9	95.3
		240	0.2	1.0	−0.5	−0.3	100	**97.2**	104.9	97.6	93.2	95.4	94.0	95.2
	100	0	1.5	1.1	3.0	−4.0	**100**	**100.0**	106.0	105.8	94.4	95.8	95.0	94.5
		100	1.1	0.8	−0.2	−0.8	**100**	100.1	111.3	102.8	95.1	95.4	95.1	95.4
		240	0.3	0.4	−0.3	−1.0	100	97.7	107.2	**96.9**	93.4	93.9	94.1	93.5
DM	20	0	−2.2	−1.2	−3.7	−2.6	**100**	109.6	104.9	111.0	94.5	94.2	92.9	93.3
		100	−5.3	−5.5	−6.5	−7.0	100	**97.1**	102.9	96.1	94.5	95.5	94.2	95.8
		240	−9.7	−6.1	−10.2	−9.8	100	101.0	105.1	**95.3**	93.9	94.9	94.1	95.2
	50	0	1.7	1.1	−0.6	0.9	**100**	114.0	102.7	113.8	95.6	94.7	93.8	95.2
		100	3.4	3.2	4.0	5.4	100	98.3	100.5	**93.9**	92.2	94.3	94.4	93.8
		240	0.7	−2.7	−0.9	−0.9	100	**94.6**	104.9	96.2	92.5	94.8	95.5	95.0
	100	0	0.4	−1.0	−0.3	−1.3	**100**	114.3	105.9	111.9	95.3	95.0	94.1	94.3
		100	−0.8	−0.4	−0.6	0.6	100	**95.9**	100.2	96.0	94.4	95.7	95.2	95.3
		240	−1.3	2.1	2.0	2.9	100	100.3	99.8	**99.6**	92.9	94.5	94.6	94.7

Note. Mod = Moderator; CM = continuous moderator; DM = dichotomous moderator; *S* = number of primary studies per meta-analysis; SD_N = between-study sample size heterogeneity in standard deviation. OS-CRE = one-stage parameter-based MASEM assuming correlated random effects; OS-UCRE one-stage parameter-based MASEM assuming uncorrelated random effects; OS-FE = one-stage fixed effects MASEM with resampling-based standard errors; TS = two-stage parameter-based MASEM.

[1]Study-specific sample sizes were randomly drawn from a truncated log-normal distribution with mean = 300, lower bound = 100, and $SD_{N_{log}}$ = .33, .61 (i.e., $SD_{N_{raw}}$ = 100, 240).

Appendix D:

Additional Simulation Results under Correlated Moderators

Table A4. Relative bias, relative RMSE, and coverage rates under correlated moderators.

				Relative Bias				Relative RMSE				Coverage			
Cor_{RE}	*N*	*S*	Cor_{Mod}	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS	OS-CRE	OS-UCRE	OS-FE	TS
0.0	100	20	0.3	4.9	4.0	4.4	1.8	**100**	100.1	108.9	105.3	96.5	96.5	92.6	96.5
	100	20	0.7	2.7	1.4	−0.5	−2.9	100	**97.5**	109.3	103.3	95.8	95.8	93.0	96.0
	100	50	0.3	2.8	1.5	−1.4	−4.1	100	**93.6**	103.4	100.4	95.6	96.5	94.4	96.8
	100	50	0.7	1.7	1.1	0.3	−4.5	100	**98.9**	103.0	99.0	94.7	96.5	94.5	95.9
	100	100	0.3	−0.1	−0.2	−0.4	−6.8	100	**96.2**	97.8	106.0	95.2	96.1	94.3	94.1
	100	100	0.7	−0.8	0.4	0.0	−4.8	100	**99.1**	107.9	104.9	92.5	94.1	94.8	93.6
	300	20	0.3	2.6	4.1	3.5	2.3	100	**91.6**	99.0	95.0	95.3	96.3	95.8	96.3
	300	20	0.7	−0.8	−0.3	−0.6	−3.1	100	**98.3**	103.4	100.7	92.6	94.2	93.1	93.7
	300	50	0.3	−1.1	−0.6	−2.0	−3.3	100	**97.4**	102.6	99.0	95.1	95.9	95.9	95.8
	300	50	0.7	−1.1	−0.2	−0.9	−0.5	100	99.4	105.9	**97.8**	94.5	96.3	96.8	95.6
	300	100	0.3	−0.6	−0.8	−0.3	−1.9	100	100.5	**99.3**	105.8	94.5	95.2	94.5	94.3
	300	100	0.7	−0.2	−0.1	−1.1	−1.2	100	**95.2**	95.9	96.5	94.1	95.2	93.8	95.2
	1000	20	0.3	2.4	0.6	−2.2	0.3	100	95.6	97.3	**92.7**	92.4	94.7	93.5	94.6
	1000	20	0.7	4.0	3.0	1.8	3.7	100	**88.0**	88.0	89.9	93.0	95.3	93.3	95.1
	1000	50	0.3	0.3	0.7	0.6	−0.2	100	**92.7**	103.9	96.2	92.4	95.2	92.0	94.9
	1000	50	0.7	0.1	0.1	−1.1	−0.6	**100**	102.2	108.4	100.6	92.4	94.9	94.9	94.9
	1000	100	0.3	0.6	0.0	−0.4	−0.7	100	**98.4**	102.7	97.9	94.3	95.7	93.6	95.3
	1000	100	0.7	−1.0	−1.4	−2.3	−1.7	100	**97.7**	100.6	98.0	94.8	95.5	95.2	95.5
0.5	100	20	0.3	1.2	1.9	2.5	−1.1	**100**	101.1	100.8	100.1	96.5	96.5	93.7	96.5
	100	20	0.7	−1.7	−0.2	5.2	−4.1	100	98.5	**95.4**	107.1	96.5	95.8	94.7	95.6
	100	50	0.3	0.6	0.5	1.3	−2.2	**100**	100.7	101.3	112.5	93.6	93.3	94.0	93.7
	100	50	0.7	−1.9	−3.1	−4.2	−7.6	**100**	113.3	104.2	118.4	95.2	95.2	93.6	96.1
	100	100	0.3	−0.1	0.4	−0.3	−2.6	100	**99.1**	103.2	109.1	95.9	95.0	94.3	95.4
	100	100	0.7	−3.3	−2.8	−1.2	−6.5	**100**	105.3	108.9	112.0	94.8	94.7	93.8	95.2
	300	20	0.3	−0.7	−0.4	−0.5	−0.9	**100**	102.2	101.4	103.1	95.6	95.3	93.7	95.6
	300	20	0.7	−2.5	−2.5	−1.2	−4.2	100	102.4	**95.4**	102.5	95.3	95.3	94.4	95.6
	300	50	0.3	−0.3	−0.8	−0.8	−0.7	**100**	114.7	107.8	114.3	95.8	94.7	94.9	95.9
	300	50	0.7	0.4	0.4	−1.5	−0.7	100	101.0	**98.1**	110.1	94.7	96.3	94.3	95.6
	300	100	0.3	−0.1	0.2	0.8	0.1	**100**	106.8	109.8	110.0	95.0	94.5	94.8	93.9
	300	100	0.7	1.8	0.7	0.2	0.2	**100**	100.9	102.6	103.6	94.3	93.8	93.2	93.8
	1000	20	0.3	0.7	1.4	1.7	2.7	**100**	111.7	101.0	112.1	95.1	96.5	93.1	96.0
	1000	20	0.7	1.5	−0.6	−0.4	0.9	**100**	106.5	101.1	110.0	92.8	94.0	93.7	94.0
	1000	50	0.3	1.8	1.6	0.9	2.3	**100**	104.2	106.7	109.7	93.3	94.3	93.5	94.3
	1000	50	0.7	1.3	−1.6	−1.6	−0.9	**100**	104.9	104.4	111.8	93.8	95.0	93.4	94.9
	1000	100	0.3	−1.0	−1.5	−1.4	−0.4	100	109.7	**99.6**	109.5	96.2	96.4	95.2	97.0
	1000	100	0.7	−1.1	−0.3	−1.3	0.0	**100**	122.3	113.5	126.4	94.1	92.5	93.0	92.5

Note. Cor_{RE} = correlation of the RE; Cor_{Mod} = point-biserial correlation of the two moderators; *N* = sample size within each primary study; *S* = number of primary studies per meta-analysis.