STRUCTURAL EQUATION MODELING
2025, VOL. 32, NO. 2
ISSN: 1070-5511 | E-ISSN: 1532-8007

Evaluation of Structural Equation Model Forests' Performance to Identify Omitted Influential Covariates

John Alexander Silva Díaz[a], Moritz Heene[a], and Andreas M. Brandmaier[b,c,d]

[a]Ludwig Maximilian University of Munich; [b]Max Planck Institute for Human Development; [c]MSB Medical School Berlin; [d]Max Planck UCL Centre for Computational Psychiatry and Ageing Research

ABSTRACT

Model misspecification is typical in applied structural equation modeling (SEM). Traditional specification search methods, such as modification indices, search for misspecifications within the model's variables but overlook influential variables not initially included and fail to detect interactions. This study evaluates SEM forests as a complementary method to conduct SEM specification search related to omitted influential covariates. The omitted influential paths include unique, mixed, and interaction paths. SEM forests' performance is evaluated under different factor loading magnitudes, covariate path magnitudes, and sample sizes. Results show SEM forests accurately identify omitted influential covariates without falsely identifying non-influential covariates in large samples (1,000) with strong covariate-latent variable paths ($\beta = .5$).

KEYWORDS

Model fit; model specification search; SEM forests; structural equation modeling; variable importance

Structural equation modeling (SEM) is a widely used statistical tool in social sciences for establishing complex multivariate relations between latent variables while controlling for measurement error (Holbert & Stephenson, 2002; Mulaik, 2009). SEM is typically used as a confirmatory method where researchers, with a theory-driven approach, specify and test structural and measurement relations among variables. However, multivariate models in the social sciences often fail to account for all relevant variables and paths, leading to misspecified models, biased parameter estimates and incorrect interpretations (e.g., Kaplan, 1988; Mulaik, 2009; Tomarken & Waller, 2003).

Researchers often use a modification index-based approach for model specification search. This method estimates the increase in chi-square-based model fit when they sequentially add identified omitted relevant paths (e.g., regression weights) (MacCallum et al., 1992; Mulaik, 2009; Saris et al., 2009). However, modification indices inherit biases inherent to the chi-square test, such as the undesired influences of nuisance parameters like sample size and parameter magnitudes (Saris et al., 2009). They also assume the model is correctly specified in terms of included variables, failing to detect omitted variable bias. The latter occurs when omitted variables correlate with both included predictors and dependent variables, leading to biased coefficients of the included predictors and incorrect model fit indices (Kline, 2016; Tomarken & Waller, 2003; Wilms et al., 2021). Moreover, modification indices are not sensitive to misspecifications involving interaction effects (Brown & Templin, 2023; Mooijaart & Satorra, 2009] and their heuristic use can lead to overfitting (MacCallum et al., 1992). This is especially problematic in social science, where interactions reflect the contextualized effect of variables regulated by others (Morgan & Sonquist, 1963; Tomarken & Waller, 2003). Despite these limitations, latent models that account for measurement error are powerful for handling interaction effects, as they manage the increased error resulting from multiplying individual measures (Cortina et al., 2021).

How can we move beyond such limitations of modification indices in SEM for detecting model misspecification? Various alternatives for specification search in SEM have been proposed, such as Marcoulides and Drezner (2001) genetic algorithm search, Scheines et al. (1998) vanishing tetrads, Saris et al. (1987) expected parameter change (EPC), Marcoulides and Drezner (2003) ant colony optimization (ACO) algorithm, or Marcoulides et al. (1998) tabu search variable selection. Brandmaier et al. (2013, 2016) proposed SEM trees and SEM forests, the fusion of exploratory decision trees and random forests with SEM. These techniques enable a data-driven and yet theory-guided exploration of models, extending the confirmatory approach of SEM with the exploratory approach of decision trees. SEM forests allow the identification of possible covariates and covariate interactions that predict parameter heterogeneity in SEM, potentially informing about variables that provide additional predictive information. Moreover, tree and forest structures reduce the chance of overfitting, finding potential generalizable features in the data (Arnold et al., 2020; Brandmaier et al. 2013).

1. Decision Trees and SEM Trees

A decision tree is a nonparametric predictive model that recursively splits a data set to create groups with similar observations within each group and dissimilar observations between groups. Decision trees find locally optimal splits by maximizing an information criterion or applying statistical tests to determine the significance of the splits. They can be described as a data-driven, yet theory-constrained search in model space, starting with a hypothesized model and expanding it if the trees identify further relevant variables (Breiman et al., 1984). SEM trees, proposed by Brandmaier et al. (2013) and implemented in the R package semtree (Brandmaier et al., 2023), combine the nonparametric nature of decision trees with parametric SEMs as outcomes. They assume that observed data are drawn from different underlying multivariate normal distributions (Figure 1). SEM trees account for heterogeneity not explained by an initial theory-based model by hierarchically splitting a data set with respect to a potential set of covariate predictors and their interactions, maximizing differences in parameter estimates across the resulting groups while finding similar response patterns within groups (Brandmaier et al., 2016). Each node of the tree represents simultaneously a data partition with respect to a covariate and a SEM that induces the resulting subgroups. The partitions remain as long as relevant differences in the parameter estimates are discovered.

The R package *semtree* (Brandmaier et al., 2023) includes different criteria for variable and split-point selection, including the recently implemented score-guided test (Arnold et al., 2020). Score-guided tests use log-likelihood function derivatives giving scores sorted by the covariates under scrutiny and aggregated into a test statistic to evaluate SEM parameters homogeneity across all levels of a given covariate (Arnold et al., 2020). Larger deviations from the expected score (i.e., zero) indicate greater parameter instability. Arnold et al. (2020) reported that score-based tests are unbiased in covariates selection (i.e., there is no preference for any covariate when the null hypothesis is true) and have higher statistical power than original SEM tree methods (i.e., naïve and fair). Score-based tests are also computationally more efficient since only one model needs to be estimated, in contrast to the original likelihood ratio approach implemented in SEM trees, which needs to estimate models for each split candidate (Arnold et al., 2020).

SEM trees are especially suited for data with many covariates and unknown interactions (Brandmaier et al., 2013). However, SEM tree results may be unstable because each split may be affected by particularities of the sample at hand, and the effect of a small perturbation in a node would be magnified in a cascade effect down the tree, calling their generalizability into question (Brandmaier et al., 2016). As a more robust alternative to SEM trees, Brandmaier et al. (2016) proposed the use of ensembled SEM trees, called SEM forests.

2. Ensemble Methods and SEM Forests

An ensemble is a set of models generated from random samples of an original data set that is often more robust and accurate than individual models. Ensemble methods define a sampling scheme to generate random data and a combination scheme to aggregate individual model predictions into a final outcome. Basu et al. (2018) mentioned a series of decision trees ensembles to identify interaction effects, such as Random Forests (Breiman, 2001), Iterative Random Forests (Basu et al., 2018), and Node Harvest (Meinshausen, 2010). Basu et al. (2018) highlighted that random forests are the best option for detecting high-order interactions since other decision trees ensemble techniques grow shallow trees to prevent overfitting, but at the cost of predictive accuracy when detecting high-order interactions.

Random forests are an ensemble method for tree-based methods like SEM trees. SEM forests are ensembled SEM trees that can guide the search for potential explanatory variables, assess their influence, determine which variables to control for, and generate new hypotheses about structural relations that are difficult to generate from a pure theory-driven approach in complex data sets. A key aim of SEM

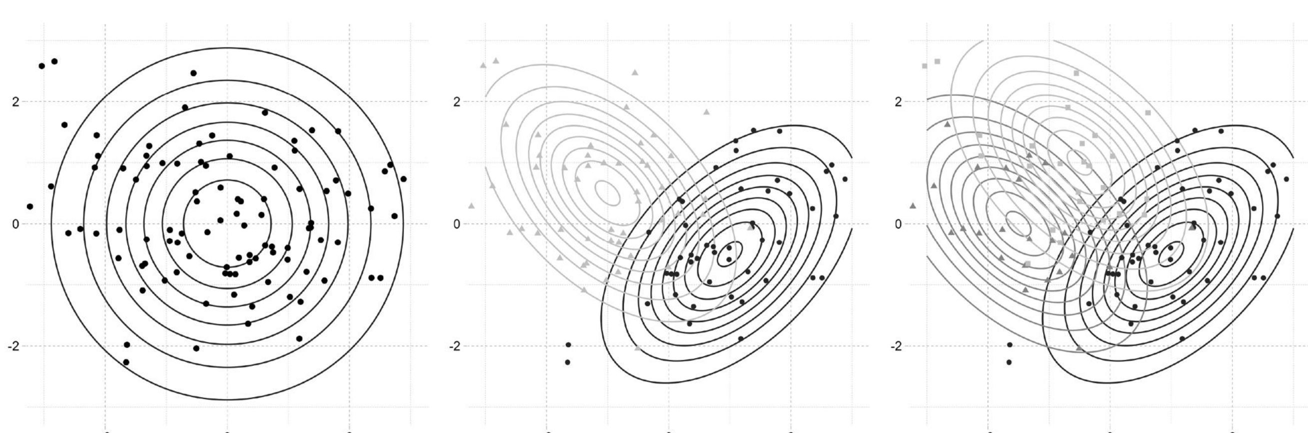


Figure 1. Illustration of SEM tree splitting process.
Note. A SEM represents a multivariate normal fit (illustrated as contours) to some data (illustrated as points). Left: A single Gaussian model describes the entire set. Center: Potential split into two subgroups for better fit. Right: Further split in left subgroup for improved model fit. Figure adapted from the Handbook of Structural Equation Modeling (2nd ed.), by A. M. Brandmaier and R. C. Jacobucci, 2023, Guilford Press. Copyright 2023 by Guilford Press.

forests is to determine the importance of covariates in predicting an outcome. If a covariate is important in predicting an outcome variable, it will be more consistently selected by the individual trees that included it in the random subset of covariates considered for splitting, indicating its stability as a predictor and helping to identify potential influential covariates.

To generate a SEM forest, resampled training data sets are created for each tree using either bootstrapping aggregating (bagging) or subsampling from the original data. Bagging generates new random samples with replacement, matching each new sample with the original sample size, while subsampling draws smaller new random samples without replacement. A SEM tree is grown from these resampled datasets keeping the observations that were not part of the training as out-of-bag samples for each tree. To enhance tree diversity, only a random subset of the total covariates is evaluated at each split, with subset size c defined heuristically (i.e., $c = log2(m)$, being m the total amount of covariates. Smaller c increases tree variability and independence but may miss relevant predictors and high-order interactions (Brandmaier et al., 2016).

Oshiro et al. (2012) explored the ideal number of trees in a random forest, based on the area under the ROC curve, which illustrates power against type-I error rate. They suggest using between 64 and 128 trees per forest, as adding more trees beyond this range only escalates computational cost without notable performance improvement. However, in the context of SEM forests, Brandmaier et al. (2016) noted that the optimal number of trees depends on various factors, including the number of predictors, their interactions, data heterogeneity, and model complexity. Although specific thresholds for the number of trees in SEM forests are yet to be established, Brandmaier et al. (2016) heuristically proposed setting the number of trees to a relatively large value, such as 2,000 trees.

2.1. SEM Forests Variable Importance Output

In SEM forests, a permutation-based *variable importance* is provided as a nonparametric estimator of the relative importance of a set of covariates and their interactions. Variable importance evaluates the decrease in fit due to the random permutation of a predictor, assuming that if an important predictor is randomly permuted its functional relation with the model-predicted distribution is broken (Brandmaier et al., 2016). Variable importance gives aggregate information about unmodeled variables that a single SEM tree cannot provide, quantifying the influence of potentially relevant covariates on the model's covariance structure (Brandmaier et al., 2016).

SEM forests variable importance is calculated as the average decrease in log-likelihood importance when a predictor is removed from the forest. This is estimated based on a simple resampling scheme, computing the likelihood of observing data for each tree in the forest, then randomly permuting the predictor of interest, removing all outcome information, and recalculating the likelihood for each tree. This process is repeated for each predictor (Brandmaier et al., 2016). The larger the drop in likelihood after the permutation, the more important the variable. Decisions regarding sampling parameters, such as resampling type, number of trees, and candidate predictors at each node, can impact variable importance. Additionally, missingness and categorical variable imbalance are other factors that may affect variable importance (Brandmaier et al., 2016). Then, informed definitions of those parameters are a key step to obtain accurate variable importance outcomes.

The inherent random variation of observed samples that entails the SEM forests generation process enhances the accuracy of variable importance estimates compared to single decision trees. That is, the instability disadvantage of individual trees becomes an advantage of forests, since random fluctuations allow an ensemble to be a better representation of the true partition of a sample. Increased diversity improves the performance of ensemble methods, with random forests outperforming other classification approaches such as generalized linear models and support vector machines (Bühlmann & Yu, 2002; Fernández-Delgado et al., 2014). However, forests improvement in accuracy comes at the expense of losing the straightforward interpretability of individual trees. Moreover, SEM forests do not specify variable relations, leaving the integration of covariates paths into the model open-ended. As with any data-driven technique, SEM forests do not provide a shortcut from data to theories and their outcomes should be tested on independent samples to validate their generalizability (Brandmaier et al., 2016).

3. Questions and Hypotheses

SEM forests have been successfully applied to explore heterogeneity in large-scale empirical data sets, allowing to identify influential predictors in diverse contexts: students' attitudes toward collaboration in PISA 2015 (Li et al., 2021), individual differences in episodic memory (Brandmaier et al., 2016), late-life well-being decline (Brandmaier et al., 2017), and early predictors of adolescent emotion regulation (Van Lissa et al., 2023). However, to our knowledge, there is a lack of studies that explored to what extent SEM forests' performance to identify relevant omitted predictors is influenced by either nuisance parameters such as factor loadings magnitude or sample size, or by data parameters related to the degree of misspecification, such as the magnitude of omitted covariate paths or interaction parameters. Usami et al. (2017, 2019) reported that SEM trees' ability to identify true classes explaining population heterogeneity in longitudinal data was influenced by the covariate's agreement with its true latent profile, sample size, and negatively affected by the number of true classes. However, Usami et al. (2017, 2019) examined SEM trees performance, but not SEM forests, and only in the context of longitudinal analysis.

This study addressed the mentioned gaps in SEM forests performance research by examining their ability to consistently and accurately identify influential omitted covariate paths while avoiding false detection of non-influential ones in underspecified models under different data conditions. SEM forests' performance was evaluated under different standardized factor loadings, strength of covariate-latent variable relationships, and varying sample sizes. We hypothesized that larger factor loadings and stronger covariate effects would lead to more consistent detection of influential omitted covariate paths due to increased statistical power (Heene et al., 2011; Meyvis & Van Osselaer, 2018; Van Voorhis & Morgan, 2007). We also expected SEM forests to perform better with larger sample sizes, as they reduce sampling errors (Van Voorhis & Morgan, 2007), stabilize variable and split-point selection in individual trees, and trees in a forest can grow deeper. Additionally, we explored different covariate-latent variable relationships: covariates with no influential paths, mixed influential paths, unique influential paths, and a covariate interaction variable. We expected a particular high performance in detecting the omitted interaction path, as SEM trees and forests are especially suitable to detect interactions (Brandmaier et al., 2013; 2016).

4. Procedure

4.1. Population Structural Equation Model

We specified a population latent model inspired by Paxton et al. (2001), who deduced their model structure from a literature review of SEM studies published in psychological and sociological key journals in a period of five years. The population model included nine measured variables (indicators), three latent variables, and 10 exogenous variables (covariates) (Figure 2). Following Paxton et al. (2001), we defined a structural causality chain with mediator η_2 regressed on predictor η_1, and outcome variable η_3 regressed on mediator η_2. Additionally, each of the nine indicators loaded uniquely on one latent variable.

Our population model included 10 covariates, as SEM forests is a technique developed for larger sets of covariates. We regressed predictor η_1 on covariates X_8, X_9 and X_{10},

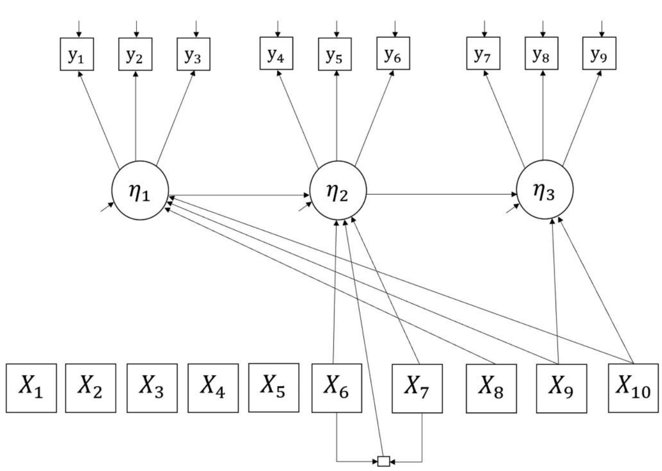


Figure 2. Path diagram of the population model.
Note. The population model has nine indicators *y* and three latent variables η_1 to η_3, with 10 exogenous variables X_1 to X_{10}. The box between X_6 and X_7 represents an interaction effect on η_2.

mediator η_2 on covariates X_6 and X_7, and outcome variable η_3 on covariates X_9 and X_{10}. The mediator η_2 was also regressed on the interaction between X_6 and X_7. With this set, the influential covariates X_6, X_7 and X_8 had unique paths (each covariate relates to only one latent variable), while the influential covariates X_9 and X_{10} had mixed paths (each covariate relates to two latent variables). Covariates X_1 through X_5 were defined as non-influential predictors with no paths to any latent variable. All paths between the latent variables η_1, η_2 and η_3, indicators y_1 to y_9, and latent variables X_1 to X_{10} were defined as linear relationships.

4.2. Experimental Design

We used a $3 \times 4 \times 3$ fully crossed design, totaling 36 conditions, with factors including measurement quality, magnitude of covariate paths, and sample size. For measurement quality we defined three levels representing low, medium, and high standardized factor loadings: $\lambda_1 = .4$, $\lambda_2 = .6$, and $\lambda_3 = .8$. For covariate path magnitudes, understood as the magnitude of standardized regression coefficients between covariates and latent variables, we defined four levels: null, low, medium, and high paths: $\beta_0 = 0$, $\beta_1 = .2$, $\beta_2 = .35$, and $\beta_3 = .5$. The interaction effect was fixed to .2 for all the conditions. For sample size, we used three levels corresponding to small, medium and large samples in the context of social sciences: $N_1 = 200$, $N_2 = 500$ and $N_3 = 1{,}000$.

4.3. Data Generation

We used the R package *simsem* (Pornprasertmanit et al., 2022) to generate sample data for each combination of our experimental conditions. For all the conditions, regression path coefficients between latent variables were set to .4, and measurement error correlations between indicators and the residual correlation among latent variables were both fixed to zero. To keep parameters in a standardized scale, indicator and exogenous latent variable variances were set to one. Noninfluential covariates X_1 to X_5 were randomly generated as continuous variables from a standard normal distribution. Influential covariates X_6 to X_{10} were continuous variables generated from a multivariate normal distribution with mean vector $\mu = 0$, and off-diagonal covariance matrix elements uniformly distributed within the range [0, .1]. The interaction variable X_6X_7 was computed as the product of X_6 and X_7. For each of the 36 possible combinations of the experimental conditions, 500 data sets were generated. Out of a total of 18,000 cases, 98.6% (17,751 cases) were successfully generated, while 1.4% (249 cases) failed during data generation. All the cases not correctly generated correspond to the largest covariate path condition ($\beta_3 = .5$), where 94.5% (4,251 of 4,500) were correctly generated, while 5.5% (249 of 4,500) were not generated.

4.4. SEM Forests Growth

For each of the 17,751 cases generated, we generated one SEM forest composed of 500 trees using the *semforest* function of the R package *semtree* (Brandmaier et al., 2023). We employed score-guided SEM trees for split selection, since this method is faster and more powerful than other split methods available in the *semtree* package (i.e., fair, naïve; see Arnold et al., 2020). Subsampling was used as the resampling procedure, with three candidate covariates considered at each node, based on the heuristic of randomly draw $\log_2 (m)$ variables for every split evaluation (Brandmaier et al., 2016) with 10 being the total number of covariates (m) for our data sets.

To promote robust forest growth and stable estimates, the minimum sample size per node was set to 100, ensuring no split attempts were made if a node had fewer than 100 observations. Additionally, the lower bound of the terminal nodes was set to 50, requiring at least 50 observations in a child node of a potential split to be valid. This decision came at the cost of having SEM forests with only two possible splits for our smallest sample size conditions ($N_1 = 200$). In random forests, it is often useful to let individual trees grow deep without applying a statistical cut-off. Thus, we kept the default package alpha level of 100% (i.e., never stop splitting based on the chi-square test statistic) as the most liberal statistical criterion to get deep trees. This increased power for detecting interactions but at the cost of Type I error, meaning a possible increase of the incorrect detection of the non-influential covariates in any given tree of a forest. However, even if individual trees overfit, the variable importance resampling approach averages across many trees helping to prevent individual trees from overfitting the data.

To evaluate SEM forests' performance to correctly identify omitted influence covariate paths and not falsely identify no-influence covariate paths, we used the variable importance analysis integrated in the R package *semtree* (Brandmaier et al., 2023). Higher variable importance indicates greater model misfit when a covariate is omitted, meaning that including those variables improves model fit.

5. Results

Figure 3 shows that influential covariates (X_6 to X_{10} and the interaction variable X_6X_7) had higher importance values than non-influential ones (X_1 to X_5, and condition $\beta_0 = 0$). Covariates with mixed paths (X_9 and X_{10}) also had higher importances than those with unique paths (X_6, X_7 and X_8) and the interaction variable X_6X_7, which had similar importance scores. Noninfluential covariates had variable importances slightly negative but still centered around zero, indicating unbiased estimates.

There was a notable dispersion in the importance values for both influential and non-influential covariates, indicating that SEM forests may sometimes misidentify covariates. To calculate SEM forests' probabilities of correctly selecting influential covariates and falsely selecting non-influential ones, we used the absolute value of the largest negative importance of each covariate as a threshold (Strobl et al., 2009). Importances above this threshold were classified as influential, while those below as noninfluential. The probability of correctly identifying influential covariates increased with larger covariate path magnitudes, factor loading magnitudes or sample sizes (Table 1). The probability of falsely identifying noninfluential covariates was close to zero across all covariate path magnitudes, factor loading magnitudes, and sample sizes.

Specifically, probabilities to identify influential covariates were nearly 100% for covariates with medium or high path magnitudes ($\beta_2 = .35$, $\beta_3 = .5$) combined with medium or large samples ($N_2 = 500$, $N_3 = 1{,}000$), and medium or high factor loadings ($\lambda_2 = .6$, $\lambda_3 = .8$), with a few exceptions (Table 1). Under other conditions, the probabilities of identifying influential covariates were close to zero or varied widely for each covariate. Influential covariates with mixed paths (X_9 and X_{10}) typically had higher identification probabilities than those with unique paths (X_6, X_7 and X_8) and the interaction variable (X_6X_7). Although the interaction variable X_6X_7 had its path fixed at .2 for every condition, its identification probability increased when other influential covariates had higher path magnitudes, with higher factor loading magnitudes, and with larger sample sizes (Table 1)

Table 2 shows how factor loading magnitude, covariate path magnitude, and sample size influenced the SEM forest variable importance averages. Larger covariate path and factor loading magnitudes led to higher variable importance scores. For sample size, small and medium samples ($N_1 = 200$, $N_2 = 500$) had similar importance scores, while the large sample ($N_3 = 1{,}000$) had smaller importance scores. For non-influential covariates, variable importance averages were negative but close to zero across all conditions, except for the smallest sample size, which had a more negative variable importance average (-.182).

Figure 4 shows the influential and non-influential covariate mean differences per experimental condition. The largest mean differences between influential and non-influential covariates were from conditions with the large covariate path ($\beta_3 = .5$), especially when combined with medium and large factor loadings ($\lambda_2 = .6$ and $\lambda_3 = .8$). Conditions with null ($\beta_0 = 0$) and small covariate paths ($\beta_1 = .2$) had negligible mean differences, regardless of the different factor loadings or sample sizes.

Appendices A and B provide a detailed graphical comparison of how each experimental condition affected the SEM forests' performance for influential and non-influential covariates. Influential covariates with mixed paths (X_9 and X_{10}) were most frequently identified across all conditions, while the interaction variable X_6X_7 had similar importance scores to those with unique paths (X_6, X_7 and X_8) (see Appendix A). Among experimental conditions, covariate path magnitude had the greatest impact on influential variables importance scores, with importance scores near zero for small paths ($\beta_1 = .2$), and scores between .5 and 1 for large paths ($\beta_3 = .5$). Higher factor loadings also increased importance scores, though less strongly than covariate paths. Sample sizes showed similar average importance scores, but with reduced score dispersions in larger samples. Factor loading and covariate path magnitudes did not influence variable importance dispersion (see Appendix A). Variable

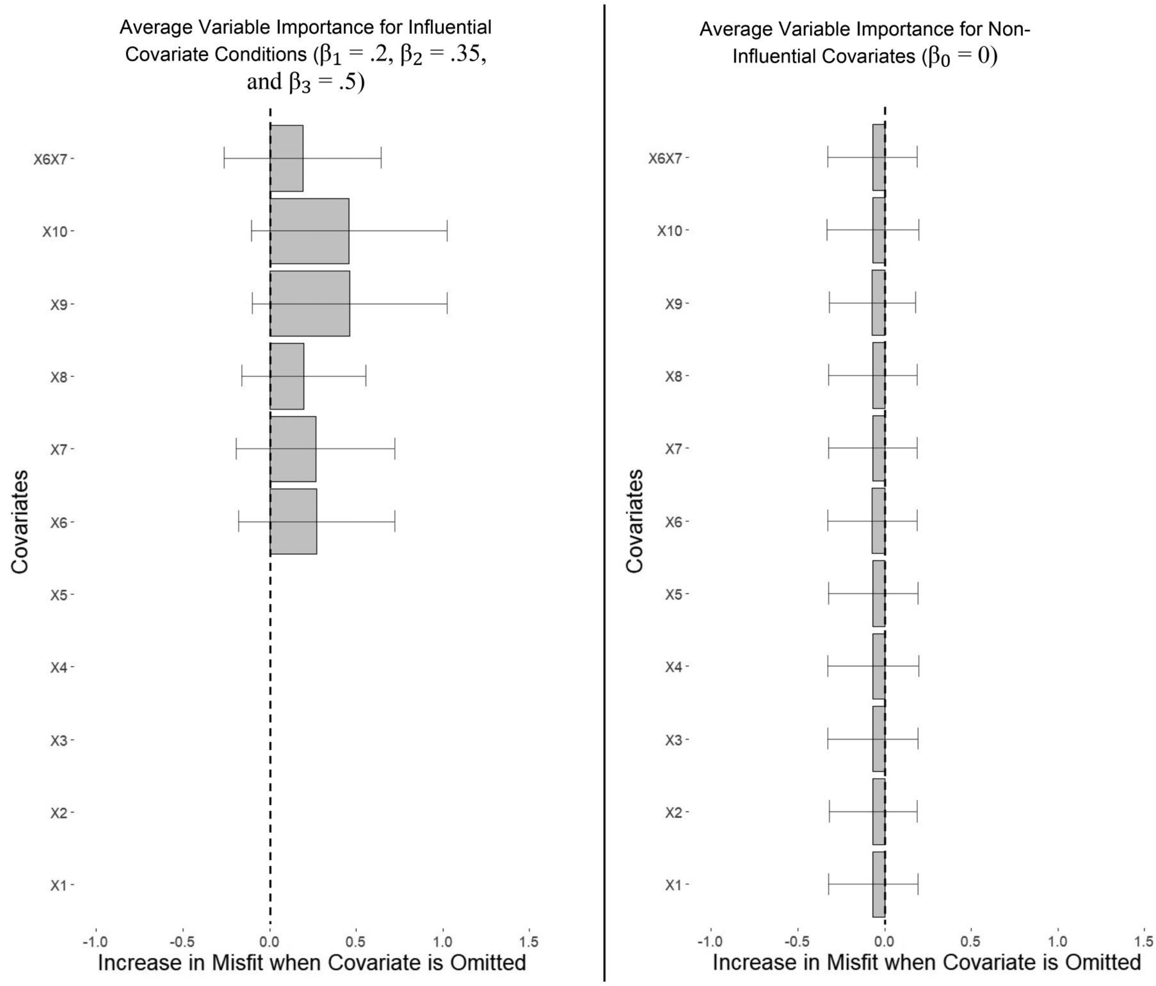


Figure 3. Average Variable importances: Influential vs. Non-influential covariates.
Note. On the left, X_6X_7 is an interaction variable; X_6, X_7, and X_8 have unique influence paths (paths to only one latent variable); X_9 and X_{10} have mixed influence paths (paths to two latent variables); and X_1 to X_5 are non-influential covariates. On the right, all covariates are non-influential.

importance scores to identify non-influential covariates were slightly negative and close to zero for all factor loadings and sample sizes, though the small sample had more negative and dispersed scores (see Appendix B).

6. Discussion

This study examined SEM forests' performance to estimate variable importance scores for influential covariates, covariate interactions, and non-influential covariates, considering data conditions such as factor loadings magnitudes, covariate path magnitudes, and sample size. Using the permutation-based variable importance measure from the *semtree* R package (Brandmaier et al., 2023), results showed that SEM forests are sensitive to omitted influential covariates and provide unbiased importance scores centered around zero for noninfluential covariates across the explored conditions.

SEM forests' probability of correctly selecting influential covariates increases with larger covariate paths, factor loading magnitudes, or sample sizes, while the probability of falsely selecting non-influential covariates is nearly zero across all conditions. Specifically, the probability of identifying omitted influential covariates was nearly 100% for covariates with high path magnitudes combined medium or large samples, and for medium path magnitudes combined with large samples and medium or large factor loadings. SEM forests also perform better at detecting omitted influential covariates with mixed paths (covariates related to two latent variables) than those with unique paths (covariates related to one latent variable).

Larger covariate paths and factor loadings resulted in higher variable importance scores, while sample sizes showed similar scores but less dispersed for larger samples. SEM forests more frequently identified omitted influential covariates with higher regression paths since larger effect sizes increase statistical power (Meyvis & Van Osselaer, 2018; Van Voorhis & Morgan, 2007). The better and more consistent SEM forests' performance with larger factor loadings is also in line with previous results showing that larger factor loadings have smaller standard errors (Heene et al., 2011). Less dispersed variable importances closer to the true value for larger samples are in line with previous results showing that larger samples decrease sampling error, allowing more stable statistics (Van Voorhis and Morgan, 2007). Thus, the effect of both factor loading magnitude and

Table 1. SEM Forests' probabilities of correctly selecting influential covariates and falsely selecting non-influential covariates.

Probability of Correctly Selecting an Influential Covariate

	$\beta_1 = .2$									$\beta_2 = .35$								
	N_1			N_2			N_3			N_1			N_2			N_3		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
X_6	0	0	0	.02	.01	.23	.07	.44	.51	.01	.03	.04	.38	.9	.97	.73	1	1
X_7	0	0	0	.02	.06	.24	.27	.54	.74	0	.01	.02	.13	.88	.88	.93	1	1
X_8	0	0	0	.04	.07	.11	.13	.34	.29	0	0	.05	.09	.44	.79	.81	1	1
X_9	0	0	0	.06	.21	.46	.21	.77	.97	0	.01	.28	.62	1	1	1	1	1
X_{10}	.01	0	0	.03	.29	.13	.32	.73	.98	0	.31	.08	.82	.98	1	1	1	1
X_6X_7	0	0	0	.01	.07	.1	.1	.22	.29	0	.01	.02	.1	.27	.63	.26	.64	.99

	$\beta_3 = .5$								
	N_1			N_2			N_3		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
X_6	0	.41	.86	.72	1	1	1	1	1
X_7	.04	.38	.96	.93	1	1	1	1	1
X_8	.01	.07	.37	.94	1	1	1	1	1
X_9	.3	.59	.94	1	1	1	1	1	1
X_{10}	.3	.31	.97	1	1	1	1	1	1
X_6X_7	.02	.12	.85	.25	.99	1	.87	1	1

Probability of Falsely Selecting a Non-Influential Covariate

	$\beta = 0$								
	N_1			N_2			N_3		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
X_1	0	0	0	0	.01	.01	.02	.01	.01
X_2	0	0	0	0	0	.01	.01	.01	.01
X_3	0	0	0	0	0	.01	.01	.01	0
X_4	0	0	0	0	0	0	.01	.01	.01
X_5	0	0	0	0	.01	.01	.01	.01	0
X_6	0	0	0	0	0	0	.01	.02	.02
X_7	0	0	0	0	.02	.01	.01	.02	.01
X_8	0	0	0	.01	.02	.01	.02	.01	.01
X_9	0	0	0	.01	.01	.01	.02	.02	.03
X_{10}	0	0	0	0	.01	.01	.04	.02	.01
X_6X_7	0	0	0	0	0	.01	.02	.02	.02

Note. Sample size levels: $N_1 = 200$, $N_2 = 500$ and $N_3 = 1{,}000$; factor loading levels: $F_1 = .4$, $F_2 = .6$, and $F_3 = .8$. For the probability of falsely selecting a non-influential covariate, covariates X_1 to X_5 are averaged across all experimental conditions since they were defined as non-influential, while covariates X_6 to X_{10} and interaction variable X_6X_7 correspond to the condition with no covariate effect ($\beta_0 = 0$).

Table 2. Average Variable importance per condition.

	Average importance scores of influential covariates			Average importance scores of non-influential covariates		
	Small	Medium	Large	Small	Medium	Large
Factor Loadings	.129 (.285)	.279 (.381)	.515 (.609)	−.059 (.244)	−.069 (.251)	−.078 (.282)
Sample Size	.323 (.687)	.321 (.349)	.279 (.287)	−.182 (.414)	−.021 (.101)	−.003 (0.42)
Covariate Paths	.031 (.264)	.241 (.322)	.671 (.541)		Null ($\beta_0 = 0$) −.069 (.26)	

Note. Factor loading levels: small = .4, medium = .6, and large = .8; sample size levels: small = 200, medium = 500, and large = 1,000; covariate path levels: null = 0, small = .2, medium = .35, and large = .5. For non-influential covariates, importance scores are averaged from all covariates corresponding to the experimental condition with no covariate effect ($\beta_0 = 0$).

sample size shows that such nuisance parameters can affect the performance of variable importance measures.

We expected SEM forests to detect the omitted interaction variable, as Brandmaier et al. (2013, 2016) reported that SEM trees and forests are especially suited to detect predictor interactions. While the influential interaction variable was not the most detected omitted path, SEM forests identify its omission even though it was fixed to a small path of .2, when used with a large sample (1,000) and when other omitted influential covariates have beta paths above .35. We hypothesized that larger beta interactions would be even more detectable for SEM forests, but further studies are needed. Detecting omitted interactions is crucial since they are rather common in social sciences, and traditional methods used for model specification search, such as modification indices, are not sensitive to them (Mooijaart & Satorra, 2009). Despite the prevalence of interaction effects in social sciences, Cortina et al. (2021) noted that researchers often avoid latent variable models when dealing with interactions, possibly due to unfamiliarity and methodological obstacles. In addition, the lack of a reliable method to detect influential omitted interactions in latent models, such as SEM, may deter their use. While indices for detecting interactions exist for diagnostic classification models

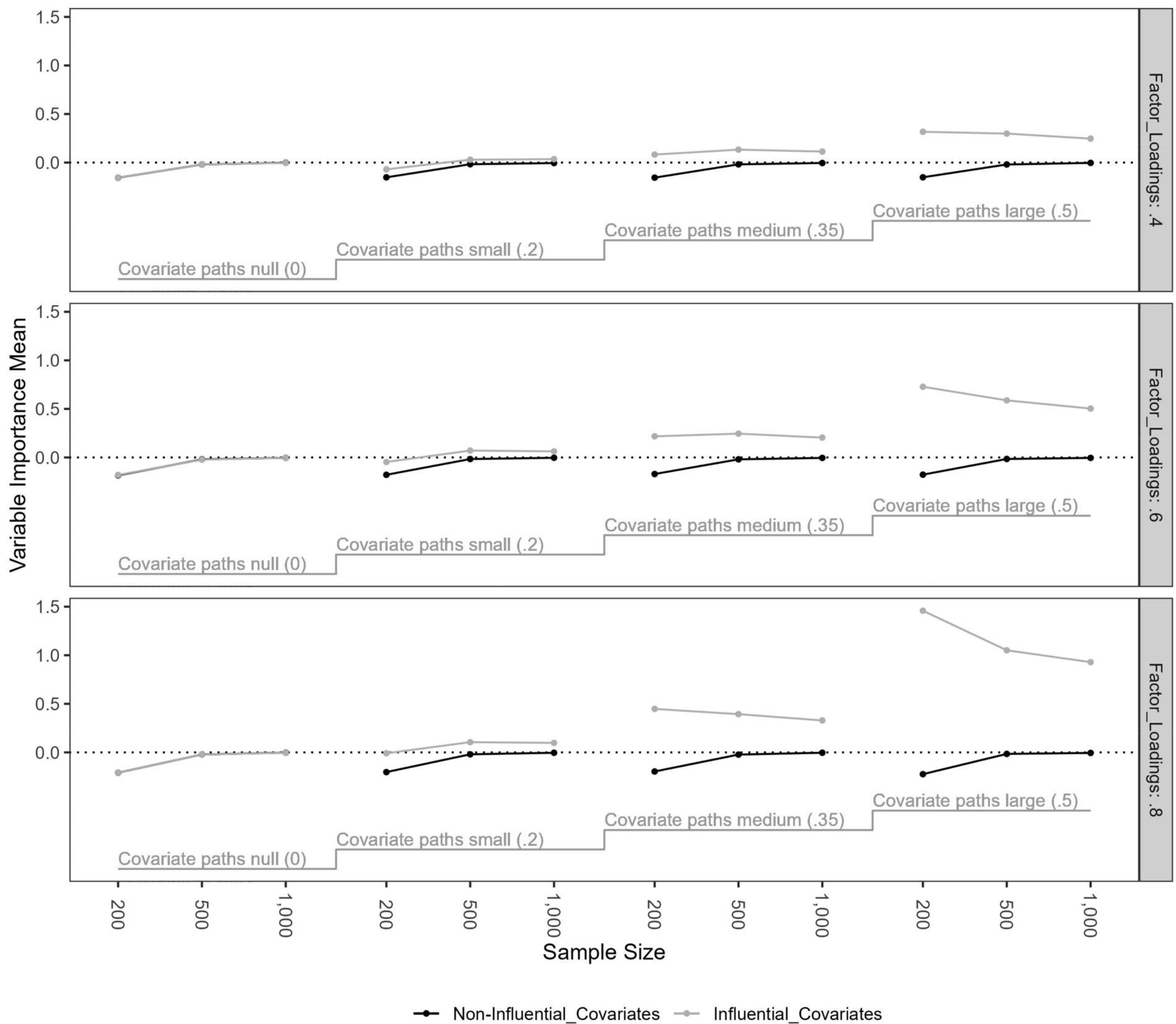


Figure 4. Variable importance mean differences per condition: Influential vs. Non-influential covariates.
Note. On the x-axis, sample size levels: $N_1 = 200$, $N_2 = 500$ and $N_3 = 1{,}000$; on the y- axis, factor loading levels: $F_1 = .4$, $F_2 = .6$, and $F_3 = .8$; and on the steps, covariate path levels: $\beta_0 = 0$, $\beta_1 = .2$, $\beta_2 = .35$, and $\beta_3 = .5$. Under null covariate paths, non-influential and influential covariates showed nearly identical means, visually indistinguishable on the graph.

(Brown & Templin, 2023), to our knowledge such indices for SEM are still needed. Thus, using SEM forests to detect omitted covariates and covariate interactions could encourage the inclusion of interaction effects in latent variable models.

A fundamental drawback of SEM forests as a modification search method is their inability to specify where and how (i.e., with which functional form) to include detected omitted influential covariates and interactions in the model. Thus, practitioners may know through SEM forests *that* a set of covariates is relevant and should be included in their models, but they still should decide *where* and *how* to include those covariates or interactions. However, SEM forests include partial dependence plots, allowing users to conduct an exploratory analysis of how a given important predictor or set of predictors influences a given model parameter. Moreover, Brandmaier et al. (2016) provided an example of model modification based on SEM forests variable importance in a single-factor analysis with the Wechsler Adult Intelligence–Revised. As a result of an SEM forest analysis with 1,000 trees on a single-factor model that hypothesizes one latent factor for verbal cognitive ability, authors included the effect of the most important predictor detected by the importance analysis (i.e., education) on the factor structure. Brandmaier et al. (2016) tested the hypothesis that education predicted differences in mean verbal performance, including education as an exogenous predictor and restricting its effect to zero. The model fit including the zero constraint was unacceptable, but when the zero constraint was freeing, model fit improved, a result that was confirmed by a likelihood ratio test. Brandmaier et al. (2016) exemplify how to include influential covariates in simple models, but the challenge remains for complex models with multiple latent variables. Further tutorial papers with step-by-step guidance on using SEM forests and incorporating influential covariates in complex models are still needed.

The conclusions presented in this study are limited to the simulation parameters and should not be generalized beyond these conditions. Further studies are needed to explore the impact of model complexity and non-linear relationships on SEM forests' performance, and to determine how and where to include omitted influential covariates. Random forests are particularly suitable for non-linear relationships since they are ensembles of decision trees that split the data recursively based on predictor values, creating partitions without assuming linearity (Breiman, 2001). Moreover, the ensemble nature of SEM forests and the non-parametric nature of their model search enhance their ability to capture diverse patterns, including non-linear relationships (Brandmaier & Jacobucci, 2023). While random forests are ideal for non-linear relationships, SEM forests can also handle linear relationships, as the results of this study suggest. Potentially, practitioners might use SEM forests to detect both linear and non-linear omitted covariate-latent variable paths, however, more research is needed to confirm SEM forests' performance with non-linear omitted influential covariates. On the other hand, SEM forests' potential ability to detect both linear and non-linear relationships would pose a challenge in determining the functional form to include a covariate in a model, as covariates may relate linearly or non-linearly to one or more latent variables.

Practitioners need to consider that while SEM forests are able to identify measured omitted influential covariates, they cannot detect unmeasured influential ones. Mixture multi-group factor analysis (MMG-FA) (De Roover, 2021; De Roover et al., 2022) offers an alternative by identifying latent clusters that correspond to unmeasured covariates. Specifically, MMG-FA clusters model groups according to a specific level of measurement invariance (e.g., equal factor loadings, equal intercepts) and allows exploring potential measured and unmeasured covariates that explain cluster memberships (De Roover, 2021; De Roover et al., 2022). Tree-based analysis can partially capture the effect of unmeasured influential covariates if they are highly correlated with measured covariates selected for splitting, with those correlated measured covariates serving as proxies and capturing some effect of the unmeasured ones, as suggested by Strobl et al. (2015) with Rasch trees. In these cases, predictive accuracy might hold, but interpretability could suffer, since those measured covariates selected for splitting could simply reflect the true effect of a correlated unmeasured influential covariate, and interpreting those proxies as having a direct causal impact on the model would be misleading (Strobl et al., 2015).

Users of SEM forests should also consider that the *semtree* R package (Brandmaier et al., 2023) calculates marginal variable importance, which can be biased when predictors are correlated (Strobl et al., 2008). Strobl et al. (2008) recommend using conditional variable importance for random forests, which are not yet included in the *semtree* R package (Brandmaier et al., 2023). This study set predictor correlations between zero and .1 to avoid convergence issues and making marginal variable importance applicable. Moreover, covariate correlations were set low considering that random forests marginal variable importance tends to overselect correlated predictors (Strobl et al., 2008), and that with highly correlated predictors, importance is often spread across them rather than attributed to a single predictor, leading to potential misinterpretation of which variables are most important (Breiman, 2001). Although predictors in psychological research tend to show low to moderate correlations, ranging from .20 to .50 (Cohen, 1988), the covariate correlations specified in this study are smaller than the typical predictor correlations found in social sciences. Thus, practitioners need to consider predictor correlations when interpreting SEM forests variable importance results, and further studies that explore the impact of higher covariate correlations on SEM forests variable importance performance are needed. Finally, practitioners should note that since SEM forests use likelihood ratio tests for split selection, they may inherit biases proper of the chi-square test, such as undesired influences of sample size and parameter magnitudes (cf. Saris et al., 2009). However, as it is common practice with random forests, we allow individual trees to grow deep without applying a statistical cut-off, setting the alpha level to 100% (i.e., never stop splitting based on the chi-square test statistic).

In summary, practitioners can benefit from using SEM forests to address model misspecification due to omitted influential covariates, particularly with samples above 1,000, or around 500 if the omitted covariate-latent variable paths are high (around .5). SEM forests are also particularly effective in detecting omitted influential covariates with paths to multiple latent variables. However, given the lack of clear guidance on where and how to include influential covariates detected by SEM forests, we suggest that their inclusion should be guided by causal inquiries. An interpretation guided by theoretical frameworks and that considers different levels of causation (e.g., counterfactual, intervention, association) (Pearl, 2009) can help users to understand the model structure, hypothesize the direct or indirect effects of predictors, and determine what type of formal relationships are expected in the model.

References

Arnold, M., Voelkle, M. C., & Brandmaier, A. M. (2020). Score-guided structural equation model trees. *Frontiers in Psychology*, *11*, 564403. https://doi.org/10.3389/fpsyg.2020.564403

Basu, S., Kumbier, K., Brown, J. B., & Yu, B. (2018). Iterative random forests to discover predictive and stable high-order interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 1943–1948. https://doi.org/10.1073/pnas.171123611

Brandmaier, A. M., Prindle, J. J., Arnold, M., & Van Lissa, C. J. (2023). Semtree: Recursive Partitioning for Structural Equation Models. R package version 0.9.19. https://github.com/brandmaier/semtree

Brandmaier, A. M., & Jacobucci, R. C. (2023). Machine learning approaches to structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling.* (2nd ed.), (pp. 722–739). Guilford.

Brandmaier, A. M., Prindle, J. J., McArdle, J. J., & Lindenberger, U. (2016). Theory-guided exploration with structural equation model forests. *Psychological Methods*, *21*, 566–582. https://doi.org/10.1037/met0000090

Brandmaier, A. M., Ram, N., Wagner, G. G., & Gerstorf, D. (2017). Terminal decline in well-being: The role of multi-indicator constellations of physical health and psychosocial correlates. *Developmental Psychology*, *53*, 996–1012. https://doi.org/10.1037/dev0000274

Brandmaier, A. M., von Oertzen, T., McArdle, J. J., & Lindenberger, U. (2013). Structural equation model trees. *Psychological Methods*, *18*, 71–86. https://doi.org/10.1037/a0030001

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Chapman and Hall/CRC, https://doi.org/10.1201/9781315139470

Brown, C., & Templin, J. (2023). Modification indices for diagnostic classification models. *Multivariate Behavioral Research*, *58*, 580–597. https://doi.org/10.1080/00273171.2022.2049672

Bühlmann, P., & Yu, B. (2002). Analyzing bagging. *The Annals of Statistics*, *30*, 927–961. https://doi.org/10.1214/aos/1031689014

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cortina, J. M., Markell-Goldstein, H. M., Green, J. P., & Chang, Y. (2021). How are we testing interactions in latent variable models? Surging forward or fighting shy? *Organizational Research Methods*, *24*, 26–54. https://doi.org/10.1177/1094428119872531

De Roover, K. (2021). Finding clusters of groups with measurement invariance: Unraveling intercept non-invariance with mixture multigroup factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 663–683. https://doi.org/10.1080/10705511.2020.1866577

De Roover, K., Vermunt, J. K., & Ceulemans, E. (2022). Mixture multigroup factor analysis for unraveling factor loading noninvariance across many groups. *Psychological Methods*, *27*, 281–306. https://doi.org/10.1037/met0000355

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems. ?*The Journal of Machine Learning Research*, *15*, 3133–3181. https://doi.org/10.5555/2627435.2697065

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, *16*, 319–336. https://doi.org/10.1037/a0024917

Holbert, R. L., & Stephenson, M. T. (2002). Structural equation modeling in the communication sciences, 1995–2000. *Human Communication Research*, *28*, 531–551. https://doi.org/10.1111/j.1468-2958.2002.tb00822.x

Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research*, *23*, 69–86. https://doi.org/10.1207/s15327906mbr2301_4

Saris, W. E., Satorra, A., & Sorbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological Methodology*, *17*, 105–129. https://doi.org/10.2307/271030

Saris, W. E., Satorra, A., & Van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 561–582. https://doi.org/10.1080/10705510903203433

Scheines, R., Spirtes, P., Glymour, C., Meek, C., & Richardson, T. (1998). The tetrad project: Constraint based aids to causal model specification. *Multivariate Behavioral Research*, *33*, 65–117. https://doi.org/10.1207/s15327906mbr3301_3

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, 307. https://doi.org/10.1186/1471-2105-9-307

Strobl, C., Kopf, J., & Zeileis, A. (2015). Rasch trees: A new method for detecting differential item functioning in the Rasch model. *Psychometrika*, *80*, 289–316. https://doi.org/10.1007/s11336-013-9388-3

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. (4th ed.). Guilford.

Li, J., Zhang, M., Li, Y., Huang, F., & Shao, W. (2021). Predicting students' attitudes toward collaboration: Evidence from structural equation model trees and forests. *Frontiers in Psychology*, *12*, 604291. https://doi.org/10.3389/fpsyg.2021.604291

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490–504. https://doi.org/10.1037/0033-2909.111.3.490

Marcoulides, G. A., & Drezner, Z. (2001). Specification searches in structural equation modeling with a genetic algorithm. In Marcoulides, G. A. & Schumacker, R. E. (Eds.), *New developments and techniques in structural equation modeling.* (pp. 247–268). Psychology Press. https://doi.org/10.4324/9781410601858

Marcoulides, G. A., & Drezner, Z. (2003). Model specification searchers using ant colony optimization algorithms. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*, 154–164. https://doi.org/10.1207/S15328007SEM1001_8

Marcoulides, G. A., Drezner, Z., & Schumacker, R. E. (1998). Model specification searches in structural equation modeling using tabu search. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*, 365–376. https://doi.org/10.1080/10705519809540112

Meinshausen, N. (2010). Node harvest. *The Annals of Applied Statistics*, *4*, 2049–2072. https://doi.org/10.1214/10-AOAS367

Meyvis, T., & Van Osselaer, S. M. (2018). Increasing the power of your study by increasing the effect size. *Journal of Consumer Research*, *44*, 1157–1173. https://doi.org/10.1093/jcr/ucx110

Mooijaart, A., & Satorra, A. (2009). On insensitivity of the chi-square model test to nonlinear misspecification in structural equation models. *Psychometrika*, *74*, 443–455. https://doi.org/10.1007/s11336-009-9112-5

Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, *58*, 415–434. https://doi.org/10.2307/2283276

Mulaik, S. A. (2009). *Linear causal modeling with structural equations*. Chapman and Hall/CRC. https://doi.org/10.1201/9781439800393

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest?. In Perner, P. (Ed.), *Machine learning and data mining in pattern recognition. MLDM 2012. Proceedings. 8* (pp. 154–168). Springer. https://doi.org/10.1007/978-3-642-31537-4_13

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*, 287–312. https://doi.org/10.1207/S15328007SEM0802_7

Pearl, J. (2009). *Causality: models, reasoning, and inference.* (2nd ed.). Cambridge University. https://doi.org/10.1017/CBO9780511803161

Pornprasertmanit, S., Miller, P., Schoemann, A. M., & Jorgensen, T. D. (2022). Simsem: SIMulated structural equation modeling. R package version 0.5-16.909. Retrieved from https://CRAN.R-project.org/package=simsem

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*, 323–348. https://doi.org/10.1037/a0016973

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, *112*, 578–598. https://doi.org/10.1037/0021-843X.112.4.578

Usami, S., Hayes, T., & McArdle, J. (2017). Fitting structural equation model trees and latent growth curve mixture models in longitudinal designs: The influence of model misspecification. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*, 585–598. https://doi.org/10.1080/10705511.2016.1266267

Usami, S., Jacobucci, R., & Hayes, T. (2019). The performance of latent growth curve model-based structural equation model trees to uncover population heterogeneity in growth trajectories. *Computational Statistics*, *34*, 1–22. https://doi.org/10.1007/s00180-018-0815-x

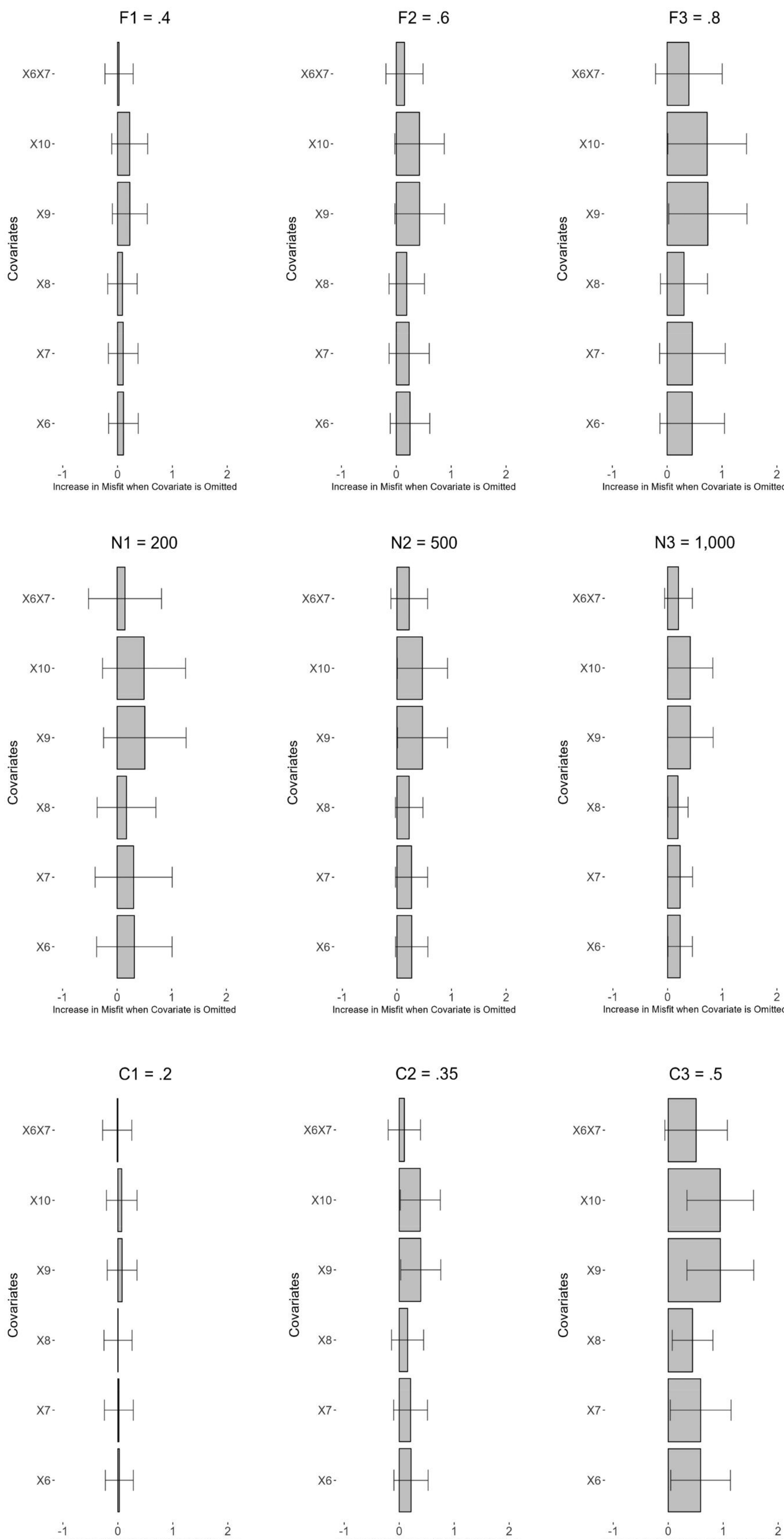
Van Lissa, C. J., Beinhauer, L., Branje, S., & Meeus, W. H. J. (2023). Using machine learning to identify early predictors of adolescent emotion regulation development. *Journal of Research on Adolescence: The Official Journal of the Society for Research on Adolescence*, *33*, 870–889. https://doi.org/10.1111/jora.12845

Van Voorhis, C. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, *3*, 43–50. https://doi.org/10.20982/tqmp.03.2.p043

Wilms, R., Mäthner, E., Winnen, L., & Lanwehr, R. (2021). Omitted variable bias: A threat to estimating causal relationships. *Methods in Psychology*, *5*, 100075. https://doi.org/10.1016/j.metip.2021.100075

Appendix A

Average Variable Importance per Experimental Condition for Influential Covariates



Appendix B

Average Variable Importance per Experimental Condition for Non-Influential Covariates