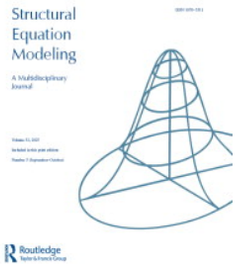
Routledge
Taylor & Francis Group



Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: www.tandfonline.com/journals/hsem20

Model Assumption Violations in Bayesian Latent Mediation Analysis: An Exploration of Bayesian SEM Fit Indices and PPP

Haiyan Liu, Ihnwhi Heo, Aleksandr Ivanov & Sarah Depaoli

To cite this article: Haiyan Liu, Ihnwhi Heo, Aleksandr Ivanov & Sarah Depaoli (2025) Model Assumption Violations in Bayesian Latent Mediation Analysis: An Exploration of Bayesian SEM Fit Indices and PPP, Structural Equation Modeling: A Multidisciplinary Journal, 32:5, 866-896, DOI: 10.1080/10705511.2025.2503789

To link to this article: https://doi.org/10.1080/10705511.2025.2503789

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
2025, VOL. 32, NO. 5, 866–896
https://doi.org/10.1080/10705511.2025.2503789

Model Assumption Violations in Bayesian Latent Mediation Analysis: An Exploration of Bayesian SEM Fit Indices and PPP

Haiyan Liu, Ihnwhi Heo, Aleksandr Ivanov, and Sarah Depaoli

University of California

ABSTRACT

This study investigates the effectiveness of Bayesian approximate fit indices and the posterior predictive *p*-value (PPP) in identifying model assumption violations in latent mediation analysis. We investigate three common forms of model-assumption violations: (1) errors in the measurement model of the mediator, (2) misspecifications in the measurement model of confounders, and (3) the omission of confounders. Additionally, we evaluate the sensitivity of these fit indices to prior specifications, comparing diffuse priors with weakly informative priors. Our findings demonstrate that BRMSEA, BCFI, BTLI, $\hat{\Gamma}$, $\hat{\Gamma}_{adj}$, and PPP effectively detect misfit with high certainty when sample size is large. Severe misspecifications, such as using standardized total scores for mediators, lead to significant misfit detected by all indices. Including true confounders mitigates the impact of mediator measurement model misspecifications, making these misfits more challenging to detect. Measurement model misspecifications for confounders of the mediator-to-outcome path result in detectable misfit using all indices. Furthermore, all fit indices effectively identify omitted confounders affecting the mediator-to-outcome path, particularly when the confounding effect is moderate or strong. Weakly informative priors have little impact on the variability of the fit indices. These results highlight the utility of Bayesian approximate fit indices and PPP in diagnosing model assumption violations in latent mediation analysis.

KEYWORDS

Bayesian fit indices; latent mediation analysis; model fit; model misspecification; omitting confounders

1. Introduction

Mediation analysis is a common framework used to explore and quantify how an independent variable influences a dependent variable (Baron & Kenny, 1986; Hayes, 2009). This analytical approach has gained increasing popularity across various research disciplines, including epidemiology, psychology, sociology, and related fields, due to its ability to untangle the underlying mechanisms of observed relationships (e.g., Fritz & MacKinnon, 2007; Richiardi et al., 2013). The core objective of mediation analysis is to determine whether the association between two variables is due, wholly or in part, to a third variable (i.e., *M*) that transmits the effect of the independent variable (i.e., *X*) on the dependent variable (i.e., *Y*; MacKinnon, 2012; MacKinnon et al., 2007). By identifying mediators, researchers can gain deeper insights into the causal pathways and processes that link independent and dependent variables, thus moving beyond mere associations to uncover potential causal mechanisms.

The mediation analysis typically involves decomposing an independent variable's total effect on a dependent variable into direct and indirect effects (MacKinnon, 2012). The direct effect represents the influence of the independent variable on the dependent variable that is not mediated by the third variable, while the indirect effect quantifies the pathway through the mediator. This decomposition is often conducted using regression-based approaches, where the relationships among the independent variable, mediator, and dependent variable are examined through a series of regression equations (Liu et al., 2021; Preacher & Kelley, 2011). However, it is important to recognize that obtaining a significant indirect effect in a tri-variate regression system does not necessarily confirm mediation. This significant indirect effect could also result from potential confounding factors, which may bias the estimates of the mediation effect (MacKinnon et al., 2000; Sweet, 2019; Valeri & VanderWeele, 2013). To draw valid causal inferences, it is essential to establish a theoretical basis for the causal pathway from *X* to *M* to *Y* and to consider potential confounders that may influence the relationships among these variables (Imai et al., 2010; Pearl, 2014). Moreover, the variables *X*, *M*, and *Y* should be measured without error to ensure accurate estimates. Recently, Liu and Wang (2021) systematically investigated the impact of the co-occurrence of measurement error and the omission of confounders on statistical inference in the simple mediation model without latent variables within the frequentist framework.

In the social and behavioral sciences, researchers often deal with variables that are not directly observable, such as

CONTACT Haiyan Liu hliu62@ucmerced.edu Psychological Sciences, University of California, Merced, 5200 N. Lake Road, Merced, CA 95343, USA.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/10705511.2025.2503789.

the teacher's role, students' emotions, and academic performance (Bollen, 1989). Measurement scales are typically employed to quantify these psychological constructs, involving multiple items or questions designed to capture different facets of a particular construct. However, these latent constructs are often measured with error. Confirmatory factor analysis (CFA) is widely used to gain insights into the structure of indicator data and latent traits by analyzing indicator data (Cattell, 1952). Importantly, CFA models account for measurement error, ensuring that the relationships between observed indicators and latent constructs are accurately represented, thus providing a more reliable assessment of the underlying constructs.

Latent mediation modeling is a framework that integrates latent variables into mediation pathways (Finch et al., 1997; Miočević, 2019). A typical latent mediation model comprises two key components: a measurement model and a structural model. The measurement model depicts the relationships between latent variables and their observable indicators, while the structural model captures direct and indirect effects among the latent variables. Latent mediation analysis is commonly conducted within the structural equation modeling (SEM) framework, enabling the simultaneous assessment of both the measurement and structural models. A latent mediation model can be estimated using various statistical approaches, including frequentist and Bayesian methods. Finch et al. (1997) investigated bias in estimating indirect effects and associated standard errors across various estimators within the frequentist framework. In a Bayesian context, Miočević et al. (2021) examined the roles of different types of priors on the parameter estimates for latent mediation models when the true model was fitted.

Accurately determining indirect and direct effects in a latent mediation model depends on meeting key assumptions. Specifically, it assumes the absence of unmeasured confounders influencing the relationships between the independent, mediator, and outcome variables, which is also required for the simple mediation analysis (Liu et al., 2021; VanderWeele, 2015). Additionally, there should be no misspecification in the measurement model of the latent variables, and the latent and manifest variables are assumed to be normally distributed, as in a traditional SEM model (Du & Bentler, 2022; Tong et al., 2014). These assumptions ensure that the specified structural and measurement components accurately reflect the underlying processes. However, these assumptions are easily violated in practice.

Deviation from these assumptions can introduce bias and compromise the validity of the inferred effects, underscoring the importance of careful consideration and validation in latent mediation analysis (Liu et al., 2025). Finch et al. (1997) explored how the degree of nonnormality in measured variables impacts the accuracy of model parameter estimates. Recently, Zhang and Wang (2024) proposed mitigating the consequences of measurement error and omitted confounders by considering the informativeness of the confounding effect under the mediation model. They concluded that omitting confounders and including the wrong latent factor as a mediator led to biased estimates of the mediation effect. Discerning measurement model misspecification and omitting confounders in latent mediation analysis remains a fundamental challenge for researchers.

Approximate fit indices for SEM have been extensively studied for their sensitivity to model misfit. For example, Hu and Bentler (1999) introduced widely used cutoff values for fit indices such as RMSEA, Tucker–Lewis Index (TLI), Comparative Fit Index (CFI), and Gamma Hat ($\hat{\Gamma}$). Fan and Sivo (2007) investigated the sensitivity of fit indices to model specification and model complexity. Savalei (2012) examined the behavior of the root mean square error of approximation (RMSEA) under different types of model misspecifications in a CFA model. Additionally, Yang et al. (2018) evaluated the performance of ten rescaled test statistics to reduce the false rejection of correctly specified models. These studies establish a foundation for understanding approximate fit indices within the frequentist framework.

Recently, SEM approximate fit indices have been adapted into their Bayesian counterparts (Depaoli et al., 2024; Garnier-Villarreal & Jorgensen, 2020; Hoofs et al., 2018). Hoofs et al. (2018) introduced Bayesian RMSEA and examined its sensitivity to model misspecification and prior specification in Bayesian confirmatory factor analysis (CFA). Garnier-Villarreal and Jorgensen (2020) extended this work by introducing seven chi-square-based approximate fit indices for Bayesian SEM using diffuse priors. Subsequent studies have further examined the performance of these indices across different modeling contexts. For example, Winter and Depaoli (2022) explored the influence of prior specification on Bayesian fit indices in linear growth curve modeling, while Edwards and Konold (2023) evaluated the performance of Bayesian RMSEA, TLI, and CFI in CFA models. Heo et al. (2024) compared Bayesian approximate fit indices with the posterior predictive *p*-value (PPP) in detecting model knot placements within piecewise growth curve modeling. Additionally, Cao et al. (2024) investigated the effectiveness of BRMSEA, CFI, and TLI in comparison to DIC and PPP for detecting structural misspecifications in SEM. These studies underscore the growing interest in Bayesian approximate fit indices and their applicability in diverse modeling frameworks.

Despite the increasing popularity of Bayesian approximate fit indices and the growing interest in latent mediation analysis, little research has examined their effectiveness in detecting model misspecifications arising from violations of key assumptions in latent mediation models. Given the critical role of accurate model specification in mediation analysis, understanding how these fit indices respond to assumption violations is essential to ensure valid inferences.

To fill in the current gap, this study systematically evaluates the effectiveness of Bayesian approximate fit measures and the PPP in detecting violations of model assumptions in the latent mediation analysis, including the measurement model misspecification and the omission of confounders. Specifically, we will compare the widely used Bayesian (approximate) fit indices for different types of misspecifications to provide a comprehensive assessment of their sensitivity and reliability. The indices examined in this study

include Bayesian root mean square error of approximation (BRMSEA), Bayesian comparative fit index (BCFI), Bayesian Tucker–Lewis index (BTLI), Bayesian $\hat{\Gamma}$ and $\hat{\Gamma}_{adj}$, and the PPP. These indices are commonly used by applied researchers and can be extracted from the output of popular SEM software, including M*plus* and `blavaan`. Given their accessibility and frequent application in empirical research, this study provides essential insights into their performance in detecting model misspecifications within latent mediation models, offering practical guidance for detecting model assumption violation in latent mediation analysis.

The remainder of this article is structured as follows. First, we introduce the latent mediation analysis framework. Next, we discuss relevant Bayesian methods that enhance this modeling approach. We then describe our simulation design, which incorporates various types of misspecifications that reflect common specification errors in applied latent mediation models. Following this, we present the results of the simulation study and provide a discussion highlighting key findings, recommendations on the use of fit indices, and future methodological research directions based on these insights.

2. Latent Mediation Analysis with Confounders

Within social and behavioral research, it is common for investigators to focus on unobserved traits and their connections. These unobserved traits are often captured through latent factors and are typically measured through measurement scales (or questionnaires). One advantage to using a latent variable setup is that measurement errors can be directly modeled within the analysis. The mediation model can be naturally extended into the latent variable framework if the key variables (e.g., the mediator or outcome) are not directly observed. To illustrate a mediation model with latent variables, we present the single mediator model with latent variables. As introduced by Finch et al. (1997) and Miočević et al. (2021), the primary latent mediation model comprises a measurement model for the independent variable, the mediator, and the outcome variable, along with a structural model for the indirect and direct effects among them. In the current study, we also consider potential latent confounders for the paths from the independent variable to the mediator and from the mediator to the outcome variable.

An example of the latent mediation model with three indicators for each latent variable is illustrated in Figure 1.

In this model, ξ is the latent independent variable, and η_M and η_Y are the latent mediator and latent outcome variable, respectively. As a critical substantive extension of this model, we also consider two latent confounders: ζ_{c1} for the path between ξ and η_M, and ζ_{c2} for the path between η_M and η_Y.

The measurement model for the independent latent variable ξ,[1] the two confounders ζ_{c1} and ζ_{c2}, and the

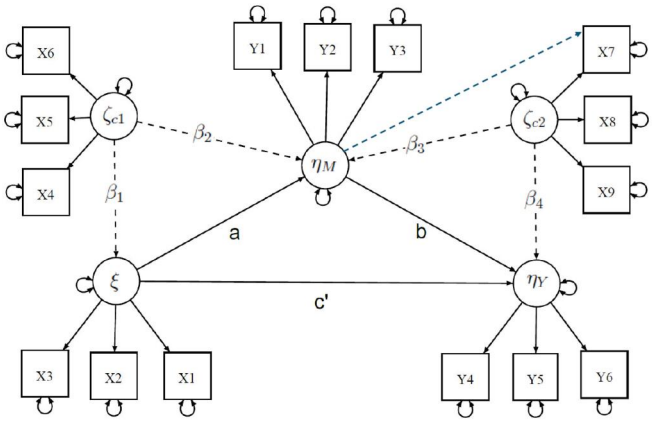


Figure 1. Latent mediation model with three indicators per latent variable. The dashed lines indicate either the potential cross-loading or the paths of the potential confounders.

measurement model for the mediator η_M[2] and the dependent variable η_Y are described as follows:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_9 \end{bmatrix} = \Lambda_x \begin{bmatrix} \xi \\ \zeta_{c1} \\ \zeta_{c2} \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_9 \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_6 \end{bmatrix} = \Lambda_y \begin{bmatrix} \eta_M \\ \eta_Y \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_6 \end{bmatrix} \tag{1}$$

The measurement model for the independent variable ξ and the two confounders has a 9×3 factor loading matrix Λ_x, with errors δ_i $(i = 1, \ldots, 9)$ following independent normal distributions $N(0, \sigma^2_{\delta_i})$. The mediator η_M and the dependent variable η_Y each have 3 indicators, with a 6×2 factor loading matrix Λ_y, and the error terms ε_j $(j = 1, \ldots, 6)$ following independent normal distributions $N(0, \sigma^2_{\varepsilon_j})$.

The mediation path among the latent variables is described by the following notation:

$$\begin{bmatrix} \xi \\ \eta_M \\ \eta_Y \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ a & 0 & 0 \\ c' & b & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \eta_M \\ \eta_Y \end{bmatrix} + \begin{bmatrix} \beta_1 & 0 \\ \beta_2 & \beta_3 \\ 0 & \beta_4 \end{bmatrix} \begin{bmatrix} \zeta_{c1} \\ \zeta_{c2} \end{bmatrix} + \begin{bmatrix} e_\xi \\ e_M \\ e_Y \end{bmatrix} \tag{2}$$

For reference, a is the coefficient linking the independent variable ξ to the mediator η_M, b and c' are the coefficients from the mediator η_M and the independent variable ξ to the dependent variable η_Y. The residuals or errors of the paths are e_ξ, e_M and e_Y, and they follow an independent normal distribution with mean 0 and variance parameters σ^2_ξ, σ^2_M, and σ^2_Y, respectively,

$$\begin{bmatrix} e_\xi \\ e_M \\ e_Y \end{bmatrix} \sim N_3 \left(\mathbf{0}, \begin{bmatrix} \sigma^2_\xi & 0 & 0 \\ 0 & \sigma^2_M & 0 \\ 0 & 0 & \sigma^2_Y \end{bmatrix} \right). \tag{3}$$

The indirect effect (or mediation effect) is denoted by the product of $a \cdot b$. This indirect effect captures the relationship between the independent and dependent variables

[1] Depending on whether there is a non-zero path from ζ_{c1} to ξ, the independent variable ξ could be exogenous or endogenous.

[2] The formulation assumes no cross-loading, but it can be generalized to include cross-loadings.

through the mediator η_M. That effect represents the extent to which changes in the independent variable affect the mediator (which in turn affects the dependent variable). The direct effect c' measures the influence of the independent variable on the dependent variable that is not mediated by η_M. When fitting this model to empirical data, the indirect and direct effects provide important insights into the relationship between the independent and dependent variables.

2.1. Model Assumptions

The simple and latent mediation models are both crucial tools for estimating the indirect effect and understanding the relationships between variables. However, the estimated indirect relationship effect ($\hat{a}\hat{b}$) may not represent the actual (true) mediation effect unless several assumptions are met, as discussed by VanderWeele and Vansteelandt (2009) and (VanderWeele, 2015). These assumptions include the absence of unmeasured confounders and accurate measurement of variables without error. The assumptions for simple mediation analysis have their counterparts in latent mediation analysis.

First, there must be an absence of unmeasured confounders influencing the relationships between the independent variable (ξ) and the mediator (η_M), the mediator (η_M) and the outcome variable ($\eta_{Y)}$, and the independent variable (ξ) and the outcome variable ($\eta_{Y)}$. These assumptions are crucial because unmeasured confounders can introduce bias, making it difficult to estimate the true mediation effect accurately. *Second*, the latent mediation model introduces additional complexity due to including latent variables. It is thus essential to ensure the absence of misspecification in the measurement models for all latent variables, including confounders. A misspecified measurement model leads to a biased understanding of the latent variables, which can, in turn, bias the estimates of the mediation effect.

Ensuring these assumptions are met can be challenging in practice. Unmeasured confounders are often difficult to identify and control for, and specifying the correct measurement model requires careful consideration and validation. Despite these challenges, adhering to these assumptions is vital for obtaining valid and reliable estimates of mediation effects in simple and latent mediation models. When the model assumptions are violated, the fit of the model to the empirical data is compromised, resulting in model misspecification. Therefore, the violation of model assumptions can be understood as an issue of model specification.

3. Bayesian Inference and Model Fit Assessment

Bayesian framework provides increased flexibility for models estimated within SEM, as demonstrated by B. Muthén and Asparouhov (2012). Additionally, it offers improved convergence, as noted by Kaplan and Depaoli (2012). These advantages have made Bayesian approach a preferred tool within SEM. Latent mediation models, encompassing both measurement models and the structural model describing the relationships among latent variables, can be fitted using Bayesian SEM methods. Miočević et al. (2021) evaluated the performance of Bayesian inference for latent mediation models under correct model specifications, highlighting its effectiveness in this context.

The Bayesian inference incorporates prior beliefs for each parameter and updates these beliefs using data collected through Bayes' theorem (e.g., Gelman, 2006; Kruschke, 2014; Liu et al., 2022):

$$P(\boldsymbol{\theta}|\text{data}) \propto P(\text{data}|\boldsymbol{\theta})P(\boldsymbol{\theta}). \tag{4}$$

Here, $\boldsymbol{\theta}$ represents the collection of all model parameters to be estimated, $P(\boldsymbol{\theta}|\text{data})$ is the posterior distribution, $P(\text{data}|\boldsymbol{\theta})$ is the likelihood, and $P(\boldsymbol{\theta})$ is the prior distribution of the model parameters. The prior encapsulates the initial belief about the model parameters, and it influences the posterior inference of these parameters.

3.1. Prior Specification

In Bayesian inference, the choice of priors plays a crucial role in guiding parameter estimation and ensuring stable model performance, particularly in complex models like latent mediation analysis. For continuous parameters such as factor loadings ($\lambda_{x/y}$) and path coefficients (a, b, c', and β_k) without sign restrictions, normal priors are commonly employed in both SEM and mediation analysis. This practice is well-documented in the literature, including studies by Asparouhov and Muthén (2010), Depaoli (2013), Yuan and MacKinnon (2009), and Zhao et al. (2024),

$$\lambda_{x/y} \sim \mathcal{N}\left[\mu_{\lambda_{x/y}}, \sigma^2_{\lambda_{x/y}}\right], \tag{5}$$

$$a, b, c', \beta_k \sim \mathcal{N}\left[\mu_{a/b/c'/\beta_k}, \sigma^2_{a/b/c\prime/\beta_j}\right]. \tag{6}$$

The accuracy of a normal prior is determined by its mean (μ). When μ is set to the true parameter value, the prior is considered accurate. If μ deviates from the true value, the prior becomes inaccurate, with the degree of inaccuracy increasing as the deviation grows. The variance (σ^2) of the prior controls its informativeness by determining its spread. An infinite variance (e.g., $\sigma^2 = \infty$) represents a diffuse prior, which provides minimal guidance and allows the data to primarily influence the estimation process. In contrast, a smaller variance (e.g., $\sigma^2 = 1$) results in a weakly informative prior that incorporates moderate prior certainty without being overly restrictive, thereby balancing prior knowledge and data-driven inference.

The residual variance parameters ($\sigma^2_{\xi/\eta_M/\eta_Y/\varepsilon_j}$), they take positive values, and inverse gamma priors are often used whose domain are nonnegative. An inverse gamma ($\mathcal{IG}$) prior can be utilized,

$$\sigma^2_{\xi/\eta_M/\eta_Y/\varepsilon_j} \sim \mathcal{IG}(\kappa, \nu), . \tag{7}$$

Here, κ and ν represent the shape and scale parameters, respectively. Adjusting these parameters influences the informativeness of the $\mathcal{IG}$ prior. For instance, setting both κ and ν to small values (e.g., 0.01) yields a weakly

informative prior, exerting minimal influence when prior knowledge about the variances is limited. Larger values for these parameters indicate greater certainty regarding the expected range of variance estimates. This flexibility allows the $\mathcal{IG}$ prior to accommodate varying levels of prior knowledge while ensuring computational stability (e.g., Depaoli et al., 2024; Liu et al., 2016; Merkle & Rosseel, 2015).

In Bayesian analysis, the diffuse prior for path coefficients and factor loading parameters is often specified as $N(0, \infty)$, implying equal plausibility across all real numbers, and serves as the default in M*plus*. In practice, priors with a certain level of informativeness are often used to enhance the efficiency and convergence of MCMC estimation. However, the accuracy and informativeness of a normal prior depend on the true parameter value, making it essential to systematically examine these aspects. Given that model misspecifications can interact with prior choices, this study explores how varying levels of prior informativeness and accuracy influence Bayesian inference under both correctly specified and misspecified models. The specific normal priors used in the simulation design will be detailed in the corresponding section.

For the inverse Gamma prior, we adopt the default $\mathcal{IG}(-1, 0)$ in M*plus*, following the recommendations from Asparouhov and Muthén (2010). This prior is equivalent to a uniform distribution over the positive real line. Unlike the normal prior for factor loadings and path coefficients, which directly influence mediation effect estimates, the IG prior primarily governs residual variances. Since our study focuses on detecting and assessing the impact of model misspecifications, manipulating the IG prior would introduce an additional layer of complexity unrelated to our primary research question. To maintain a clear evaluation of how prior informativeness interacts with model misspecification, we keep the IG(−1, 0) for the variance parameter, while systematically varying the normal prior to assess its influence under different model conditions.

3.2. Set up for Markov Chain Monte Carlo

Posterior inference relies on samples drawn from the posterior distribution. Two Markov chains are generated for each parameter, each with a length of 20,000 iterations. The initial 50% of these samples (i.e., 10,000 iterations), known as the burn-in period, is discarded to reduce the initial values' influence and allow the chains to reach a stable distribution. Ensuring the convergence of these chains is crucial for reliable posterior inference. Various diagnostic tools, such as trace plots and the Gelman-Rubin ($\hat{R}$) statistic, can be employed to assess convergence. In the simulation study, a replication is considered "converged" if $\hat{R} < 1.1$ for all parameters. The number of iterations in the current setup was determined based on satisfactory evidence of convergence from these two diagnostic tools after running different chain lengths.

3.3. Bayesian Model Fit and Assessment Measures

A critical aspect of model evaluation is determining whether the specified model provides an adequate fit to the empirical data. Bayesian approximate fit and selection fit indices serve as valuable tools for detecting model misspecification and the violations of key assumptions. We evaluated BRMSEA, BCFI, BTLI, Bayesian $\hat{\Gamma}$ and $\hat{\Gamma}_{\text{adj}}$, and PPP. Those are the widely used fit indices that are either readily available or easily extractable in Bayesian SEM software such as M*plus* (Muthén & Muthén, 1998–2017) and the R `blavaan` package (Merkle & Rosseel, 2015).

3.3.1. Posterior Predictive p-Value

PPP for Bayesian SEM is calculated based on the chi-square statistic T_{ML} obtained for comparing the model fitted against the saturated model. The SEM discrepancy function T_{ML} is defined as:

$$T_{ML} = n\big[\log|\Sigma(\theta)| + \text{tr}\{S\Sigma(\theta)^{-1}\} - \log|S| - p\big],$$

where n is the sample size, S is the sample covariance matrix, $\Sigma(\theta)$ is the model-implied covariance matrix based on the parameter estimates θ, p is the number of observed variables, $|\cdot|$ denotes the determinant, and $\text{tr}\{\cdot\}$ denotes the trace of a matrix.

At each MCMC iteration, s, a simulated data set Y^s_{rep} of the same size as the observed dataset is generated from the model with parameter values set at θ^s. We then can calculate the $T_{ML}(Y_{obs}, \theta^s)$ and $T_{ML}(Y^s_{rep}, \theta^s)$, then

$$\text{PPP} = P\Big[T_{ML}(Y^s_{rep}, \theta^s) > T_{ML}(Y_{obs}, \theta^s)\Big] \tag{8}$$

A perfect-fitting model is expected to have a PPP value centering around 0.5 and spread evenly above and below, indicating that about 50% of the replicated datasets have discrepancy statistics greater than those of the observed data. A low PPP value near 0 suggests potential model misspecification (Asparouhov & Muthén, 2010).

3.3.2. Bayesian Approximate Fit Indices

The Bayesian RMSEA is calculated as follows:

$$\text{BRMSEA}_s = \sqrt{\max\left(0, \frac{T_{ML}(Y_{obs}, \theta^s) - p^*}{(p^* - p_D)n}\right)},$$

where $T_{ML}(Y_{obs}, \theta^s)$ is the discrepancy function for the observed data at iteration s, and p^* is the number of non-redundant sample moments. A small value of BRMSEA implies a good fit. A widely used value of the threshold is 0.05.

The Bayesian CFI is computed as follows:

$$\text{BCFI}_s = 1 - \frac{T^T_{ML}(Y_{obs}, \theta^s) - p^*}{T^B_{ML}(Y_{obs}, \theta^s) - p^*},$$

where $T^T_{ML}(Y_{obs}, \theta^s)$ is the discrepancy function (for the observed data) of the fitted model and $T^B_{ML}(Y_{obs}, \theta^s)$ is the discrepancy of the baseline model at iteration s.

The Bayesian TLI is computed as follows:

$$\text{BTLI}_s = \frac{\frac{T^B_{ML}(Y_{obs},\theta^s)-p_B}{p^*-p_B} - \frac{T^T_{ML}(Y_{obs},\theta^s)-p_T}{p^*-p_T}}{\frac{T^B_{ML}(Y_{obs},\theta^s)-p_B}{p^*-p_B} - 1},$$

where p_B is the model complexity term for the baseline model, and p_T is the target model under evaluation.

In addition to the standard fit indices provided by M*plus*, we also evaluate the Bayesian $\hat{\Gamma}$ and $\hat{\Gamma}_{adj}$, which have been shown to exhibit less variability across different model types and sample sizes Fan and Sivo (2007) and (Garnier-Villarreal & Jorgensen, 2020). The Bayesian $\hat{\Gamma}$ is computed as:

$$\hat{\Gamma}_s = \frac{p}{p + 2\frac{T^T_{ML}(Y_{obs},\theta^s)-p^*}{n}},$$

where p represents the number of observed variables in the model, $T^T_{ML}(Y_{obs},\theta^s)$ denotes the test statistic based on the maximum likelihood estimation for the observed data Y_{obs} and posterior draws θ^s, and n is the sample size. The adjusted version is

$$\hat{\Gamma}_{adj} = 1 - \frac{p*}{p*-pD}(1-\hat{\Gamma}),$$

BCFI, BTLI, $\hat{\Gamma}$, and $\hat{\Gamma}_{\text{adj}}$ values closer to 1.0 indicate a good fit. For a maximum likelihood method, a cutoff of 0.95 is practically used for them (Hu & Bentler, 1999).

The SEM discrepancy function is evaluated using the parameters at each iteration, the resulting T_{ML} values are used to calculate the approximate fit indices. These quantities form the posterior distribution of RMSEA, CFI, TLI, $\hat{\Gamma}$ and $\hat{\Gamma}_{\text{adj}}$. From these distributions, the median is reported as the point estimate, along with the corresponding credibility intervals (Asparouhov & Muthén, 2021).

4. Simulation Design

In this section, we will conduct a comprehensive simulation study to evaluate the performance of Bayesian approximate fit indices and the PPP in identifying violations of model assumptions (i.e., model misspecifications). We will examine various scenarios where these assumptions might be breached and assess the effectiveness of these indices in detecting such violations. Furthermore, we will delve into how the accuracy and informativeness of prior distributions influence the variability of these fit indices. By varying the priors values, we aim to understand the robustness of Bayesian RMSEA, CFI, TLI, $\hat{\Gamma}$, $\hat{\Gamma}_{\text{adj}}$ and PPP to prior specifications. This investigation is critical for providing insights into the sensitivity of Bayesian SEM fit indices to prior specifications and ensuring their reliable application in empirical research implementing the latent mediation model.

4.1. Population Model

The population model is a latent mediation model with confounders, depicted in Figure 1. The model includes a latent independent variable (ξ), a latent mediator (η_M), and a dependent variable ($\eta_{Y)}$. Each latent variable has three primary factor loadings for continuous items: ξ loads on Items X_1–X_3, η_M on Items Y_1–Y_3, and η_Y on Items Y_4–Y_6. Paths are specified from ξ to η_M and η_M to η_Y with coefficients a and b respectively, and a direct path from ξ to η_Y with coefficient c'.

To examine the impact of confounder omission, we include two confounders, ζ_{c1} (with primary indicators X_4–X_6) and ζ_{c2} (with primary indicators X_7–X_9). ζ_{c1} confounds the paths from ξ to η_M with coefficients β_1 and β_2, while ζ_{c2} confounds the paths from η_M to η_Y with coefficients β_3 and β_4. To reduce complexity, we restrict $\beta_1 = \beta_2 = \beta_3 = \beta_4$ but vary their magnitudes among 0, 0.14, and 0.39, representing no true confounding effect, a small confounding effect, and a medium-sized confounding effect, respectively.

The primary factor loadings of all items are set as 0.7. The non-zero cross-loading from Item X_7 to η_M is included, set at 0.5. The factor variance for ζ_{c1} and ζ_{c2} is set at 1. The variance or residual variance of ξ is set at $1-\beta_1^2$, taking values of 1, 0.9804, or 0.8479 depending on the magnitude of β_1, ensuring the variance of ξ equals 1.

The variances of η_M and η_Y include both the variance explained by their predictor factors and residual variance. The residual variance of η_M is $1-a^2-\beta_2^2-\beta_3^2$, and for η_Y it is $1-b^2-c'2-\beta_4^2$, making the total variances equal to 1. The residual variance of the indicators is set at 0.51, resulting in unit variance for each indicator.

4.2. Design Factors and Prior Specification

The simulation design factors include the manipulation of the mediation effect, the confounding effect, sample size, types of misspecification, and prior specifications.

4.2.1. Mediation Effect

In the population model, the coefficient c' is set at 0.14 consistently for all conditions. The coefficients a and b are constrained to equal and take values of 0.14, 0.39, and 0.59 to represent small, medium, and large mediation effects, respectively (Liu et al., 2021; Tofighi & Kelley, 2020).

4.2.2. Confounding Effect

The path coefficients from the confounders ζ_{c1} and ζ_{c2} to ξ and η_M, as well as η_M and $\eta_{Y,}$ are denoted as β_1, β_2, β_3, and β_4. These coefficients take values of 0, 0.14, and 0.39 to manipulate the degree of confounding effects. For simplicity, we set $\beta_1 = \beta_2 = \beta_3 = \beta_4$.

4.2.3. Sample Size

We consider three sample sizes ($n = 100$, 200, and 400) to represent a range commonly found in applied and methodological research (e.g., Miočević et al., 2021).

4.2.4. Model Specification

The diagram of the population model is shown in Figure 1. Each set of parameter values constitutes a "true model" for corresponding simulation conditions. We also constructed six misspecified models, each containing certain measurement or structural model errors. These include two types of measurement model misspecification for the mediator factor η_M: one omitting the cross-loading from Item X_7 to η_M (referred to as the "wrong measurement model for mediator factor") and the other using the standardized total score of the indicators for η_M (referred to as the "no measurement model for mediator factor").

Additionally, we consider the misspecification of the measurement models of the confounder factors with models ignoring the measurement structure of ζ_{c1} (referred to as "no measurement model for one confounder") and both ζ_{c1} and ζ_{c2} (referred to as "no measurement model for two confounders").

For the structural model misspecifications, we examine models that ignore the confounder for the path from ξ to η_M (referred to as "one ignored confounder factor") and models that omit both confounders ζ_{c1} and ζ_{c2} (referred to as "two ignored confounder factors").

4.2.5. Prior Specification

Priors for the residual variances were kept consistent with the M*plus* default diffuse settings, where residual variances received $\mathcal{IG}(-1, 0)$ priors. The diffuse prior for variance parameters ensures minimal influence on inference (Asparouhov & Muthén, 2010).

Since model misspecifications are introduced through factor loadings and path coefficients of the confounders, it is important to examine how prior specifications for these parameters interact with model specification. To achieve this, we vary the location and spread of the normal priors for factor loadings and path coefficients. In addition to diffuse priors, we include three weakly informative priors as defined in Depaoli (2013):

$$\text{Weakly informative-accurate}: \ \lambda_{x/y}, \beta, a, b, c' \sim \mathcal{N}(\text{true}, 0.5\text{true}), \quad (9)$$

$$\text{Weakly informative-inaccurate-1SD}: \ \lambda_{x/y}, \beta, a, b, c' \sim \mathcal{N}\left(\text{true} + \sqrt{0.1\text{true}}, 0.5\text{true}\right), \quad (10)$$

$$\text{Weakly informative-inaccurate-2SD}: \ \lambda_{x/y}, \beta, a, b, c' \sim \mathcal{N}\left(\text{true} + 2\sqrt{0.1\text{true}}, 0.5\text{true}\right), \quad (11)$$

where "true" refers to the true parameter values. These priors incorporate the scale of the true parameter values, ensuring that both the spread and the magnitude of inaccuracy are relative to the true value.

4.3. Summary of Simulation Conditions

Simulation and prior settings can be found in Table 1. In all, there were 756 simulation cells in this design, and we generated 500 datasets for each cell. The percentage of converged applications was, on average, 99.67%, with a median of 100% assessed using the Gelman-Rubin convergence diagnostic. A replication was considered "converged" if the Gelman-Rubin statistic ($\hat{R}$) fell below the threshold value of 1.1 for all parameters.

Table 1. Summary of simulation factors and prior specifications.

Simulation factor	Levels/specifications
Mediation effect	$a = b = 0.14$ (small) $a = b = 0.39$ (medium) $a = b = 0.59$ (large)
Confounding effect	$\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (none) $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.14$ (small) $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.39$ (medium)
Sample size	$n = 100, 200, 400$
Model specification	**True model** **Misspecified models:** - Wrong measurement model for mediator factor - No measurement model for mediator factor - No measurement model for one confounder - No measurement model for two confounders - One ignored confounder factor - Two ignored confounder factors
Prior specification	**Diffuse prior:** - $\mathcal{IG}(-1, 0)$ for residual variances **Weakly informative priors:** - Accurate: $\mathcal{N}(\text{true}, 0.5\text{true})$ - Inaccurate-1SD: $\mathcal{N}\left(\text{true} + \sqrt{0.1\text{true}}, 0.5\text{true}\right)$ - Inaccurate-2SD: $\mathcal{N}\left(\text{true} + 2\sqrt{0.1\text{true}}, 0.5\text{true}\right)$

4.4. Overview of the Simulation Results

We will assess the effectiveness of the Bayesian SEM approximate fit measures RMSEA, CFI, TLI, $\hat{\Gamma}$, $\hat{\Gamma}_{\text{adj}}$, and the PPP, in detecting violations of model assumptions. Specifically, we examine the misspecification in the measurement model and the omission of latent confounders. The results are structured into four sections: (1) misspecification of the measurement model of the mediator using diffuse priors, (2) misspecification of the measurement model of the confounders using diffuse priors, (3) omission of confounders under diffuse priors, and (4) prior sensitivity analysis.

We will investigate how the distribution of fit indices is affected by model misspecifications, sample size, the magnitude of indirect effects, and the influence of confounders. Additionally, we will evaluate the effectiveness of approximate fit measures in classifying model fit using a credible interval-based approach. For PPP, we will compute rejection rates across different cutoff values to examine how they vary under different thresholds.

4.4.1. Distribution of Fit Indices

We first examine the empirical distributions of the model fit indices—BRMSEA, BCFI, BTLI, $\hat{\Gamma}$, $\hat{\Gamma}_{\text{adj}}$ and PPP. In each replication, we extracted the value of each approximate fit indices evaluated at the posterior mean of the model parameters, and PPP. All these replicated values form the empirical distribution under each simulation condition.

All conditions to assess their sensitivity to model specification, sample size, magnitudes of the mediation effect, and

confounding effects. To visualize these distributions, we generate box-and-whisker plots representing the 5%, 25%, 50%, 75%, and 95% the percentiles of the replicated fit indices. In these plots, the whiskers indicate the 90% confidence intervals, the box represents the interquartile range (first and third quartiles), and the horizontal line inside the box denotes the median.

4.4.2. Credible Interval-Based Model Fit Classification

The approximate fit indices are intended to quantify the degree of model misfit. However, in practice, researchers often seek qualitative decisions about model adequacy despite the inherently quantitative nature of fit indices. To better reflect this need, we classify model fit base on the 90% credible intervals following the approach recommended by Asparouhov and Muthén (2021) and adopted in recent studies (e.g., Cao et al., 2024; Depaoli et al., 2023; Heo et al., 2024). Specifically, a model is classified into three categories. A model has a "good fit" if the entire 90% CI falls below 0.06 for BRMSEA or above 0.95 for BCFI, BTLI, $\hat{\Gamma}$, and $\hat{\Gamma}_{\text{adj}}$. In contrast, a model has a "poor fit" if the entire 90% CI is outside these cutoff values. If the cutoff value is within the 90% CI, the fit of the model is classified as "inconclusive."

To quantify each index's ability to detect model misspecification, we calculated the proportion of replications classified as "good fit," "inconclusive," or "poor fit." For correctly specified models, we expected a high proportion of "good fit" classifications. Conversely, for models misspecified, the fit index that produced the high proportion of "poor fit" classifications was deemed sensitive to detect misspecification.

We would like to note that the classification of model fit depends on the specific cutoff values used. In this study, we adopted the 0.95 threshold based on practical recommendations by Hu and Bentler (1999). If different thresholds were applied, classification results would vary. For example, using a 0.90 cutoff for BCFI or BTLI would increase the proportion of models classified as "good fit." We encourage readers to interpret these classifications in light of the chosen thresholds.

4.4.3. Evaluation of PPP

For PPP, there is no direct distributional comparison for competing models, and no universally agreed-upon cutoff values exist, despite the popular use of 0.05. To explore its utility in detecting model misspecification, we tested three threshold values—0.05, 0.10, and 0.15—and examined how often misspecified models were detected under each threshold, as done by Cain and Zhang (2019).

5. Simulation Result I: Misspecification of the Latent Mediator Measurement Model

This section focuses on misspecifications in the measurement model of the latent mediator η_M. Two types of misspecifications are examined. One ignores the cross-loading from x_7 to η_M (referred to as "the wrong measurement model of the mediator factor"), and the other ignores the measurement structure entirely, using the standardized total score of the indicators for η_M (referred to as "no measurement model of the mediator factor").

To assess chain convergence, we use Gelman-Rubin $\hat{R} < 1.1$ as the criterion, the percentage of converged replications was around 99.67%. The simulation results will be calculated based on only converged replications in the subsequent analyses.

To examine the distribution of the fit indices under different model specifications, we visualize their distributions using box-and-whisker plots displaying the 5%, 25%, 50%, 75%, and 90% percentiles. Furthermore, we present bar plots illustrating the proportion of models classified as "good fit," "inconclusive," and "poor fit" for (1) the correctly specified model, (2) the model with misspecification in the measurement model (i.e., ignoring the cross-loading), and (3) the model without a measurement model. The relative size of each category reflects the sensitivity of the fit indices to different types of misspecification. Furthermore, for PPP, we will explore different cutoff values—0.05, 0.10, and 0.15—to assess how varying thresholds influence the rates of detecting model misfit.

To focus on the behavior of the fit indices themselves, all reported results are based on the default diffuse priors implemented in M*plus*. The normal prior $\mathcal{N}(0, \infty)$ are used for factor loadings and path coefficients, and the $\mathcal{IG}(-1, 0)$ prior for variance parameters.

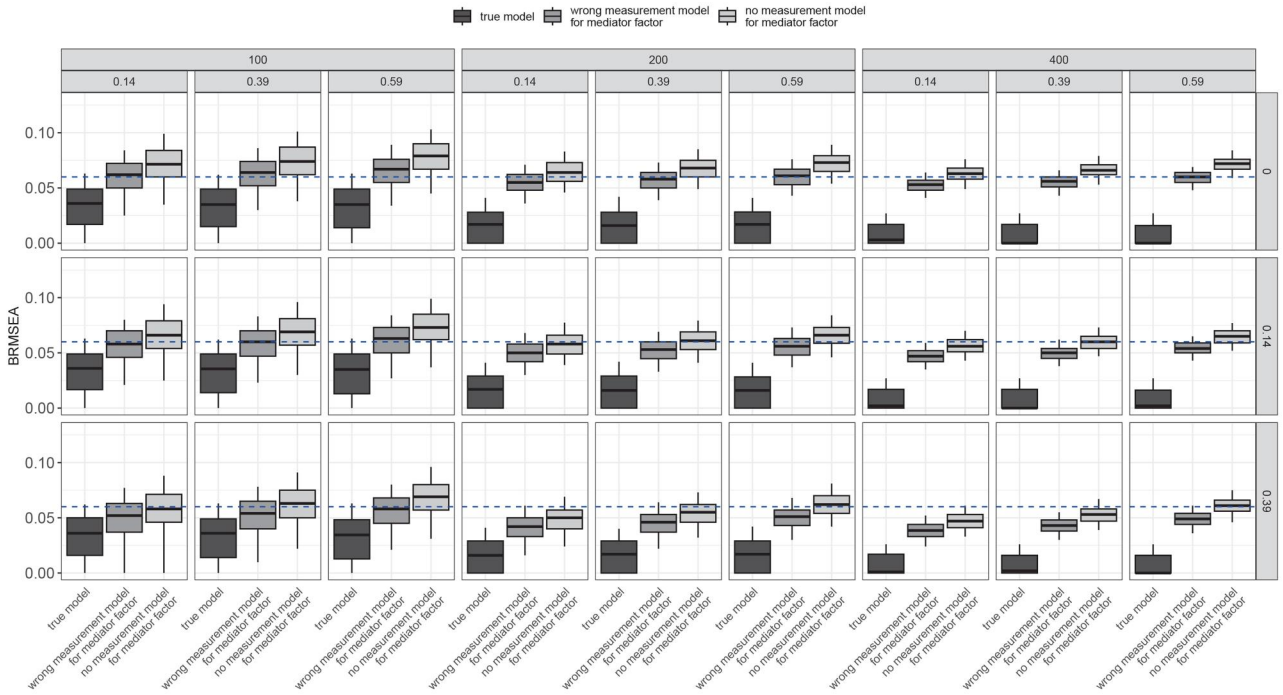
5.1. BRMSEA

Figure 2a and b contain the box-and-whisker plots of BRMSEA and the proportion of model classifications based on 90% credible intervals. The grid is organized with columns representing the three magnitudes of the mediation effects ($a = b = 0.14, 0.39, 0.59$) and sample sizes (100, 200, 400). The rows show the confounding effects (none = 0, small = 0.14, and medium = 0.39). The $x-$ axis represents different model specifications.

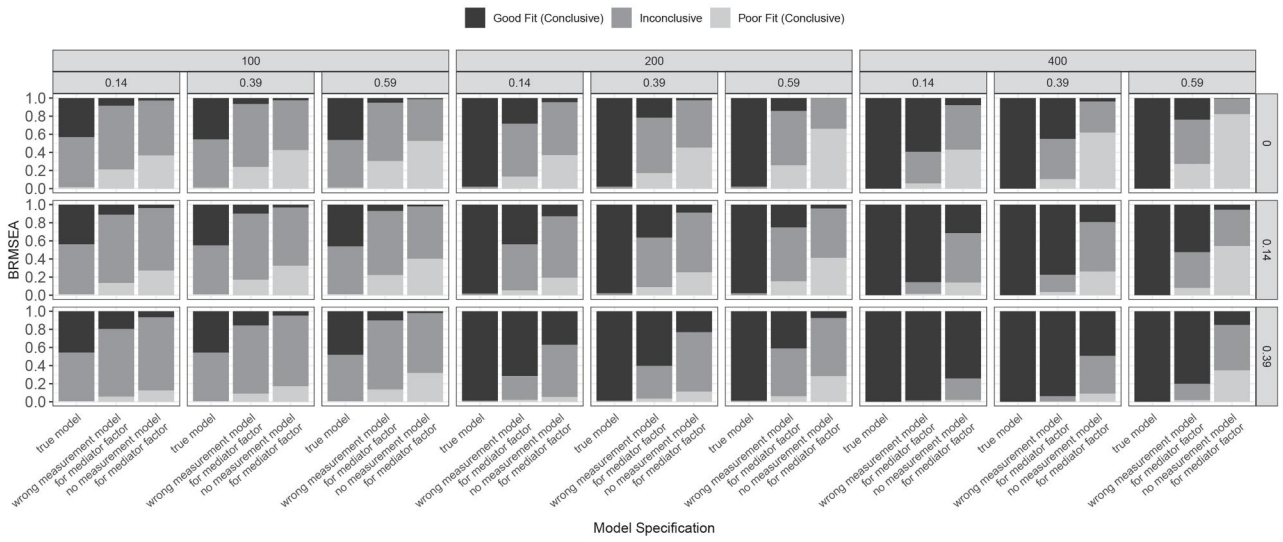
5.1.1. Distribution of BRMSEA

In Figure 2a, the horizontal line refers to the value of 0.06. The box plots reveal that the distributions of the BRMSEA values are influenced by the model specification of the mediator η_M, sample size, and mediation effects.

The BRMSEA results consistently show that the true model maintains a BRMSEA value below the traditional threshold of 0.06 under all conditions even with a sample size 100. Both the wrong measurement model (omitting the cross-loading from Item X_7) and the model using the standardized total score of the indicators (no measurement model) result in inflated BRMSEA values. The use of standardized total scores has more severe consequences for model fit than the omission of a cross-loading.



(a) *Boxplots of BRMSEA across different simulation conditions.*



(b) *BRMSEA 90% credible interval rejection rates*

Figure 2. Bayesian RMSEA for different model specifications of the mediator factor using the diffuse priors.

Moreover, the presence of confounders may slightly mitigate the impact of a misspecified measurement model, as increasing confounding effects lead to a slight decrease in BRMSEA values. For the same type of misspecification, BRMSEA values tend to increase as the mediation paths strengthen. Additionally, sample size influences both the median and interquartile range of BRMSEA values. Larger sample sizes generally reduce both the variability and median BRMSEA values across all models, though the magnitude of this reduction varies depending on the degree of misspecification.

5.1.2. Model Fit Classification Using the 90% Credible Intervals of BRMSEA

Figure 2b presents stacked bar charts that depict the proportion of replications classified as "good," "inconclusive," or "poor fit" model fit based on the 90% credible intervals of BRMSEA values.

For the true model, BRMSEA consistently classified the model as "good fit" for sample sizes of 200 and above. When the sample size was 100, approximately 50% of the replications fell into the "good" fit category, while the remaining 50% were classified as "inconclusive," with little

to no replications categorized as "poor fit." This suggests that smaller sample sizes introduce greater uncertainty in model fit evaluation but do not lead to systematic misclassification of well-specified models as a "poor fit" model.

In contrast, models with misspecified measurement structures exhibited a noticeable shift toward "inconclusive" or "poor fit." Specifically, when the measurement model for the mediator was misspecified (ignoring the cross-loading), the proportion of incorrect classification as a "good fit" decreases, and meanwhile, the "poor fit" classifications increases. For instance, the "good fit" class is only around 20–40%, when the confounding effect is null. This pattern suggests that BRMSEA detects the minor misfit such as ignoring a cross-loading but with some uncertainty.

The misfit becomes even more pronounced when no measurement model is used for the mediator, leading to a substantial decrease in the proportion of "good fit", but increase in the proportion of "poor fit." The proportion of "good fit" is less than 10%. This pattern indicates that BRMSEA is highly sensitive to the omission of a measurement model.

Furthermore, the strength of the mediation effect influenced the sensitivity of BRMSEA to detect misspecifications. As the mediation effect increased, misspecified measurement models for the latent mediator are more frequently classified as "poor fit" or "inconclusive." However, this sensitivity declined when the confounding effect increased, suggesting a potential interaction between confounding and model fit assessment. This indicates that strong confounding effects may obscure the detection of mediation model misspecifications, which warrants careful consideration when interpreting fit indices in the presence of confounders.

5.2. BCFI and BTLI

Since BCFI and BTLI exhibit similar performance, we present them within the same subsection. Our primary focus will be on interpreting the results of BCFI while highlighting key distinctions between BTLI and BCFI.

Figure 3a and b demonstrate the distributions of Bayesian CFI and the rates to detect misfit using the 90% credible intervals of BCFI. The columns represent the three magnitudes of the mediation effects ($a = b = 0.14, 0.39, 0.59$) and three sample sizes (i.e., 100, 200, 400). The rows show the confounding effects (no effect = 0, small = 0.14, and medium = 0.39). The $x-$ axis is the model specifications.

5.2.1. Distribution of BCFI

In each panel of Figure 3a, the horizontal dashed line represents a BCFI of 0.95.

For the true model, more than 95% of the replications produced a BCFI greater than 0.95 when the sample size was 200 or 400, confirming a good model fit. With a sample size of 100, approximately 75% of the replications had BCFI greater than 0.95.

With a misspecified measurement model for the mediator (i.e., ignoring cross-loadings), BCFI values are consistently lower than those of the correctly specified model in all conditions. This indicates a poorer model fit when the measurement model is not correctly specified. In the model that uses the standardized total score for the mediator, BCFI values are even lower than those observed in the cross-loading misspecification scenario. This suggests that replacing the latent mediator with its total score leads to the poorest fit among the evaluated conditions.

When comparing across different sample sizes, BCFI values increased for the correctly specified model and decreased slightly for the misspecified model as the sample size increased. This demonstrated the expected benefits of larger samples. In addition, the variability of BCFI values decreases as the sample size increases, suggesting that larger samples lead to more stable estimates of the model fit.

Finally, as the magnitude of the indirect effect increased, the BCFI for misspecified models declined, making them more distinguishable from correctly specified models. However, in the presence of strong confounders, the impact of measurement model misspecification was reduced, reinforcing the stand that confounding effects can obscure model misspecifications, potentially leading to misleading conclusions about model fit.

5.2.2. 90% Credible Interval Rejection Rates

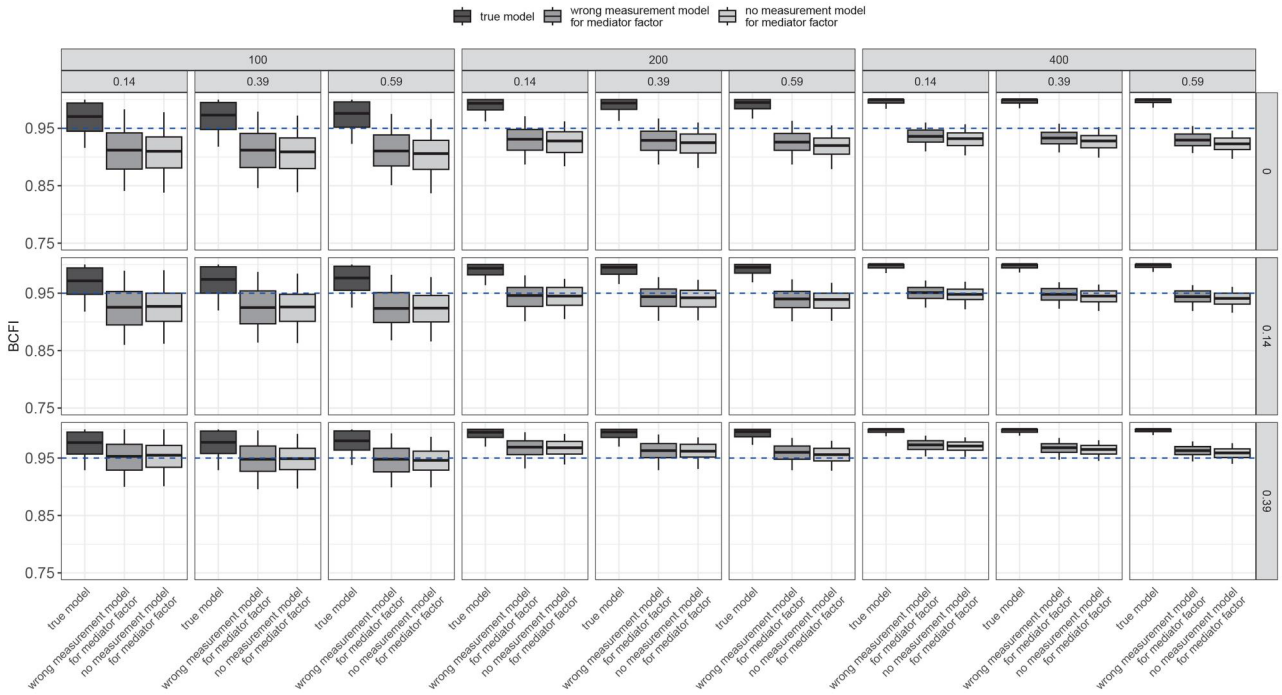
Figure 3b provides a comprehensive depiction of the model fit classification based on the 90% credible intervals of BCFI.

For the correctly specified model, BCFI consistently indicated a "good fit," particularly when the sample size is 200 or 400, with almost 100% replications falling into this category. This result demonstrates that BCFI effectively identifies correctly specified models when the sample size is adequate. However, with a sample size 100, the proportion of "inconclusive" classification increased to approximately 75% but no "poor fit" classification, suggesting that the small sample size introduced more uncertainty in the model fit classification.

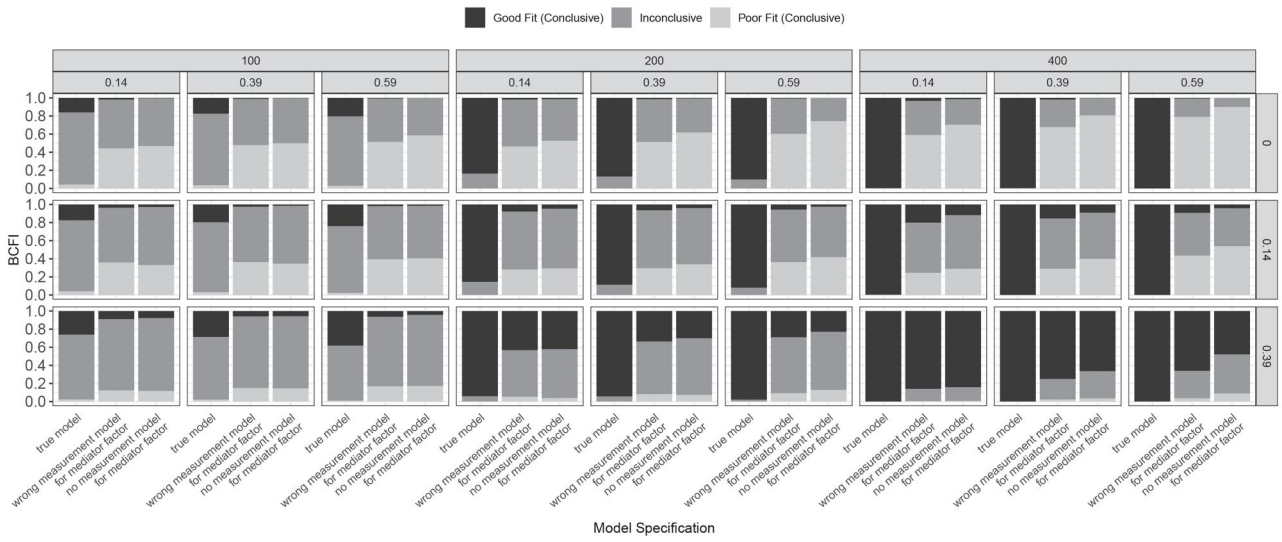
In contrast, models with errors in the measurement model exhibited a noticeable shift toward "inconclusive" or "poor fit" classifications. When the mediator measurement model is misspecified (e.g., ignoring cross-loadings), the proportion of incorrectly classifying the model as having a "good fit" decreased, with the proportion of "good fit" classification falls below 5% when no confounding effect, below 10% when the confounding effect is small, and around 40% when the confounding effect is medium.

The omission of a measurement model for the mediator further amplified the misfit of the model, leading to a substantial increase in "poor fit" classifications and notable decrease in "good fit" classification.

For misspecified models, a larger sample size enhanced BCFI's ability to detect misfit in the measurement model of the latent mediator, as evidenced by an increasing proportion of "poor fit" classifications in all conditions.



(a) *box-and-whisker plots of BCFI values.*



(b) *Model fit classification using 90% credible intervals of BCFI*

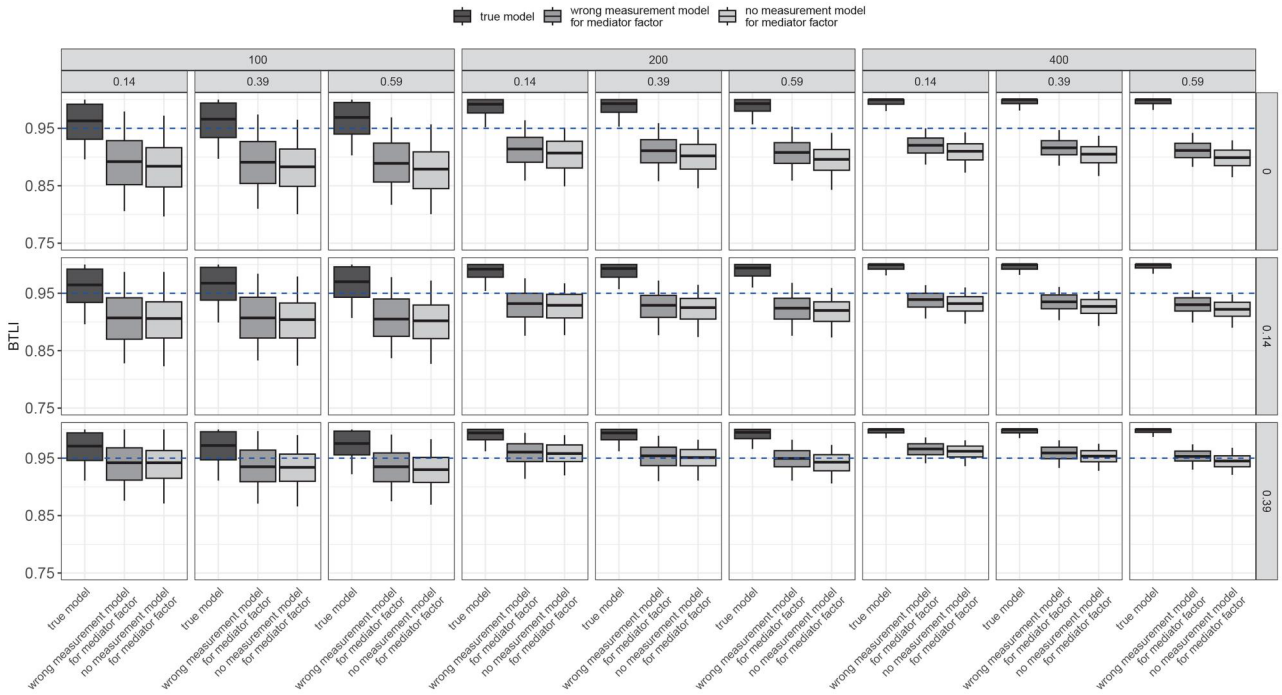
Figure 3. BCFI for different model specifications of the mediator factor using the diffuse priors.

Additionally, as the mediation effect strengthened, the BCFI became more sensitive to measurement model misspecification, resulting in a higher portion of "poor fit." However, stronger confounding effects reduced the proportion of "poor fit", suggesting that increased confounding can obscure mediator misspecifications, making them more challenging to detect.

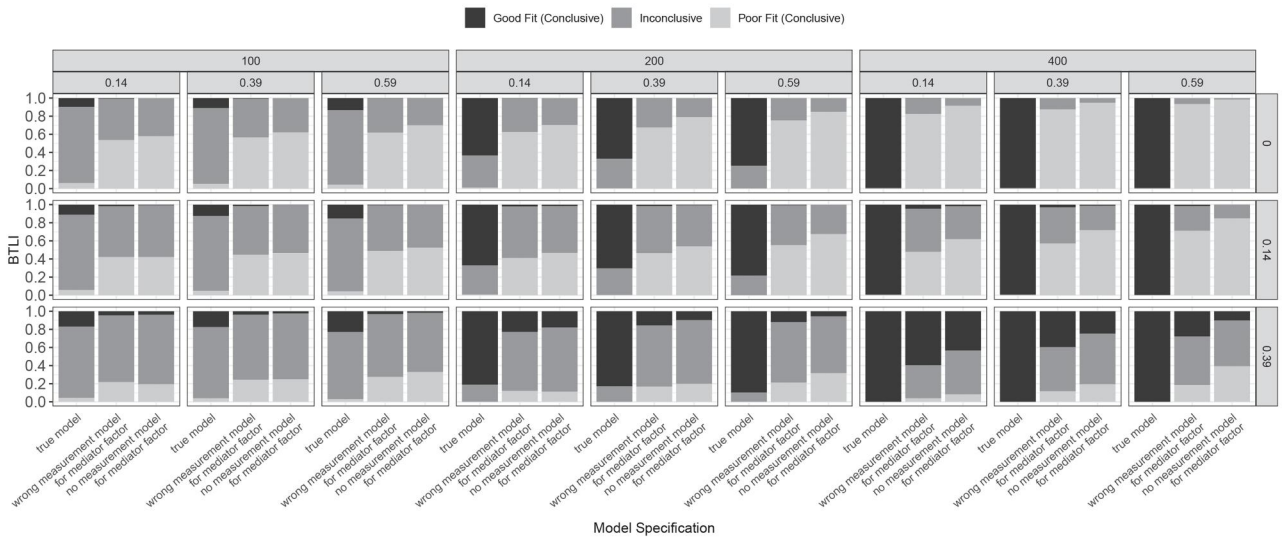
5.2.3. BTLI vs. BCFI

The distribution of BTLI and the model fit classification are presented in Figure 4a and b.

We observed that their patterns followed a similar trend to those of BCFI. However, BTLI values tend to be lower on average than BCFI, with the entire distribution shifting downward toward 0.50 for both correctly specified and misspecified models. This shift increases its ability in detecting misspecifications, resulting in a higher proportion of "poor fit." Additionally, it reduces the chance of falsely treating a misspecified model as "good fit." However, this downward shift also inflates the proportion of "inconclusive" classifications for correctly specified models when the sample size was small.



(a) *Box-and-Whisker plots of BTLI values*



(b) *Model fit classification using 90% credible interval of BTLI*

Figure 4. Bayesian TLI for different model specifications of the mediator factor using the diffuse priors.

5.3. Bayesian $\hat{\Gamma}$ and $\hat{\Gamma}_{adj}$

Since $\hat{\Gamma}$ and $\hat{\Gamma}_{adj}$ exhibit similar behavior, we will focus on interpreting $\hat{\Gamma}_{adj}$, which demonstrates slightly better performance in detecting model misfit. The results about $\hat{\Gamma}$ are available as online supplementary materials.[3]

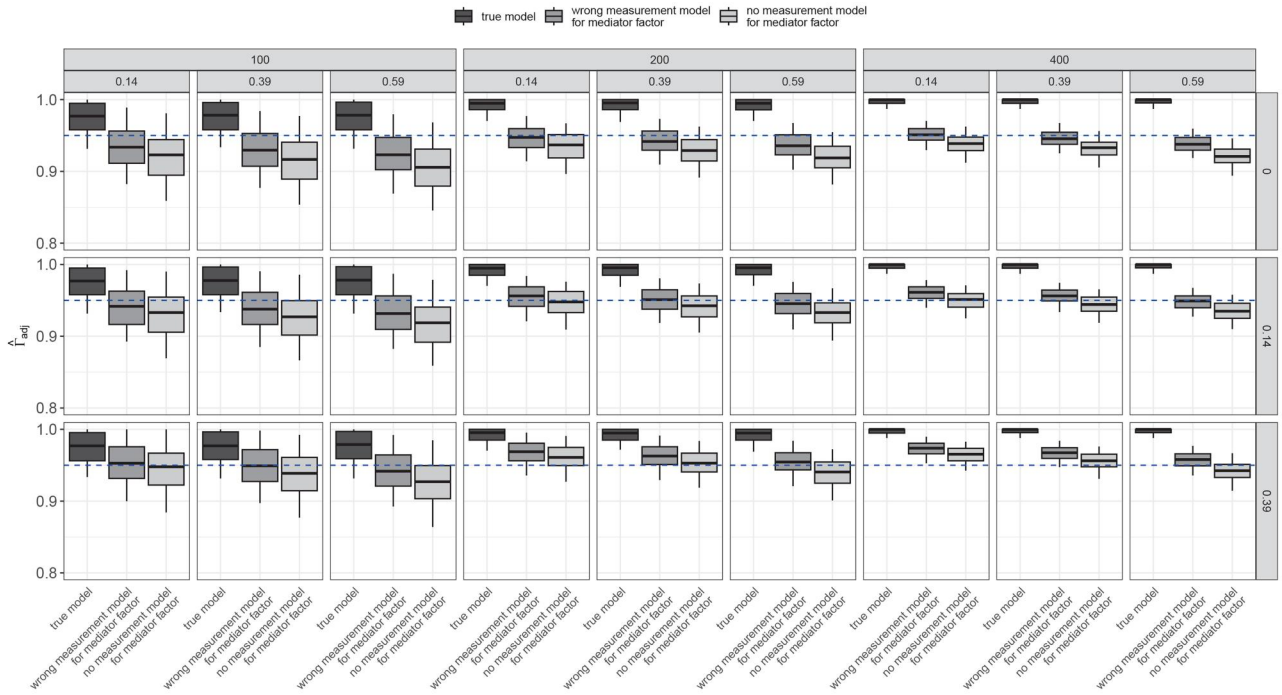
Figure 5a and b display the box-and-whisker plots for $\hat{\Gamma}_{adj}$ alongside stacked bar plots illustrating model fit classification based on their 90% credible intervals.

5.3.1. Distribution of $\hat{\Gamma}_{adj}$

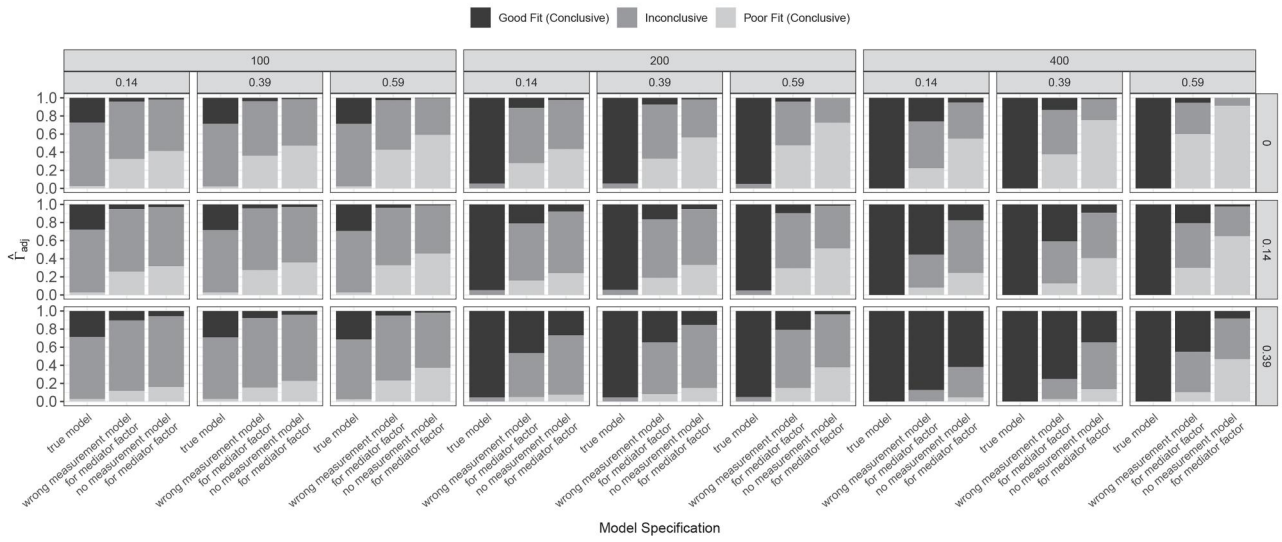
The distribution of $\hat{\Gamma}_{adj}$ remained consistently high across all model specifications. For the correctly specified model, the $\hat{\Gamma}_{adj}$ values clustered around 0.95.

When the measurement model is incorrectly specified by ignoring a cross-loading, $\hat{\Gamma}_{adj}$ values decreases. The decline

[3]The $\hat{\Gamma}$ is available at https://osf.io/q8vpm/?viewonly=fec9e23cafdb4f1babefe56eb4230da4



(a) *Box-and-Whisker plots of* $\hat{\Gamma}_{adj}$



(b) *Model fit classification based on the* 90% *credible intervals of* $\hat{\Gamma}_{adj}$

Figure 5. Bayesian $\hat{\Gamma}_{adj}$ for the different model specifications of the latent mediator factor.

is even more pronounced when the standardized total score is used for the latent mediator, suggesting that using standardized total scores constituted a more severe form of misspecification compared to ignoring a cross-loading.

Additionally, as the sample size increases, the central tendency (the middle line of the box) of the distribution is pretty stable, however, the variability of $\hat{\Gamma}_{adj}$ decreases.

5.3.2. Model Fit Classification Using the 90% Credible Interval

Figure 5b contains the stacked bar charts of the three model fit classification: "good fit", "inconclusive", and "bad fit."

The $\hat{\Gamma}_{adj}$ performed exceptionally well in classifying correctly specified models as a "good fit," particularly when the sample size is 200 or 400, where the vast majority of replications fall into this category. However, with a smaller sample size of 100, around $80\% - 90\%$ of the replications have "inconclusive" fit.

When a cross-loading is omitted (i.e., incorrect measurement model), $\hat{\Gamma}_{adj}$ values decrease, resulting in a noticeable increase in the proportion of replications classified as "poor fit" and a significant reduction in the "good fit" classifications. For example, when no confounder is included in the

model, fewer than 5% of replications are classified as "good fit." The misfit is even more pronounced when the standardized total score is used for the latent mediator, leading to a higher proportion of "poor fit" classifications and an even lower occurrence of "good fit."

As the mediation effect increased, the likelihood of incorrectly classifying a misspecified model as a "good fit" decreased, while the proportion of correctly identifying a "poor fit" increased. However, as the confounding effect grew stronger, the proportion of "poor fit" classifications dropped.

5.4. PPP

The box-and-whisker plots of PPP and the corresponding rejection rates are shown in Figure 6a and b. To assess the robustness of PPP under different cutoff values, we tested three thresholds: 0.05, 0.10, and 0.15. A model was classified as having poor fit and "rejected" if its PPP value fell below the respective cutoff. In the figures, the columns and rows represent the levels of mediation effects, sample size, and confounding effects, respectively.

5.4.1. Distribution of PPP

Figure 6a illustrated the distribution of PPP values across different model specifications, mediation path levels, and confounding effects. The horizontal dashed line is the reference line with PPP being 0.05.

For the correctly specified model, PPP values remain around 0.5 in all conditions, consistently indicating a good fit regardless of the mediation path strength and confounding effects. In contrast, the central tendency of PPP values for the model with a misspecified measurement model for the mediator is generally lower than that of the correctly specified model, reflecting a decline in model fit. The model that omits the measurement model for the mediator has even smaller PPP values.

As the sample size increases, the box plots for misspecified models shift downward toward zero. With a sample size of 200 or 400, all PPP values fall below 0.05, indicating poor model fit.

Both the variability and central tendency of the PPP distribution decrease slightly as the strength of the indirect path increases but show a slight increase as the confounding effect strengthens.

5.4.2. Rejection Rates Using PPP with Different Cutoff Values

Figure 6b presents the rejection rates of a model using PPP. The rate of rejection is defined as the proportion of replications with PPP values smaller than a given cutoff value. In the current study, we computed the rejection rates using three different cutoff values: 0.05, 0.10, and 0.15.

For the correctly specified model (i.e., true model), the rejection rates remain below 0.05 when using cutoff values of 0.05 and 0.10. When the cutoff increases to 0.15, the rejection rate rises to approximately 10%, indicating a slightly higher false rejection at this threshold.

The rejection rates of the misspecified model increases as the cutoff values increases from 0.05 to 0.15. In addition, rejection rates demonstrate greater discrepancies between different sample size conditions. With a threshold of 0.05, a sample size of 400 ensured rejection rates exceeded 80%. When the threshold increased to 0.10, a sample size of 200 was sufficient to achieve similar rejection rates.

The most pronounced misfit occurs in the no measurement model for the mediator factor, where the rejection rates were around 90%, especially for the conditions with large mediation effects and small confounding effects.

5.5. Summary

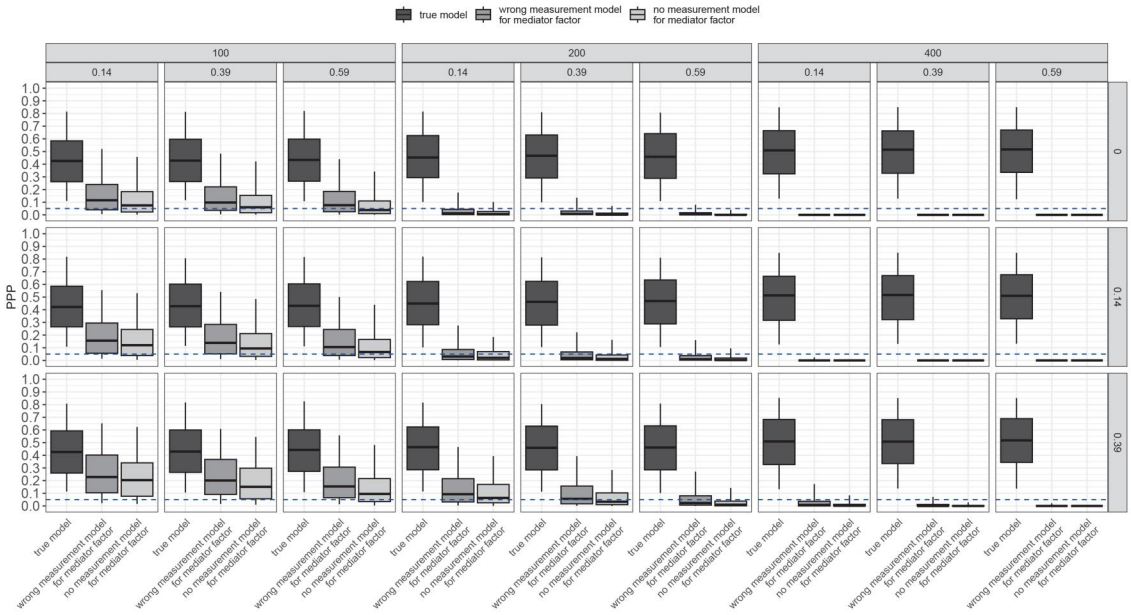
For detecting measurement model misspecifications in a latent mediator, we recommend using BCFI and BTLI as the primary indicators due to their strong sensitivity to misspecifications. BRMSEA can serve as a supplementary measure, particularly in cases of severe misspecification. The $\hat{\Gamma}_{\text{adj}}$ demonstrated similar perform as BCFI and BTLI. PPP may be useful for larger sample size conditions, but its effectiveness depends on cutoff selection. A threshold 0.10 helps balance the false rejection of a correctly specified model and the correct selection of a misspecified model.

6. Simulation Result II: Misspecification of the Confounder Measurement Model

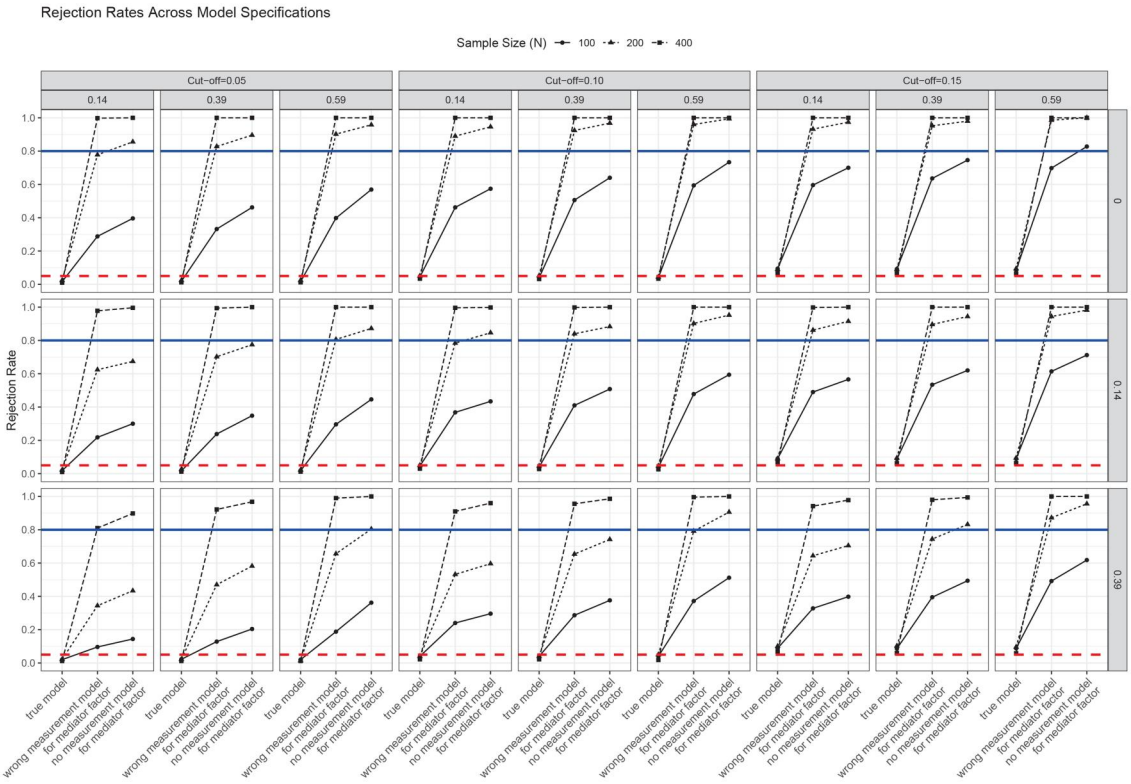
In this section, we evaluate the impact of ignoring the measurement model and using standardized total scores of the indicators for the latent confounders ζ_{c1} and ζ_{c2} on the model fit. We will examine how Bayesian approximate fit indices and PPP detect the absence of a measurement model for these confounders. We consider two scenarios with varying degrees of misspecification. The first scenario involves ignoring the measurement model for the confounder ζ_{c1}, which affects the path between the independent variable ξ and the mediator η_M. The second scenario involves omitting the measurement model for both confounders, impacting the paths from ξ to η_M and from η_M to the dependent variable η_Y. For reference, the results also include the correctly specified model.

6.1. BRMSEA

The box-and-whisker plots of BRMSEA values and the proportion of model fit classifications based on credible intervals are displayed in Figure 7a and b. The columns are organized so that within each sample size condition, the three magnitudes of the mediation effects are nested. The rows represent the three magnitudes of the confounding effects. Across the plots, the x-axis denotes different model specifications.



(a) *Box-and-whisker plots of PPP values*

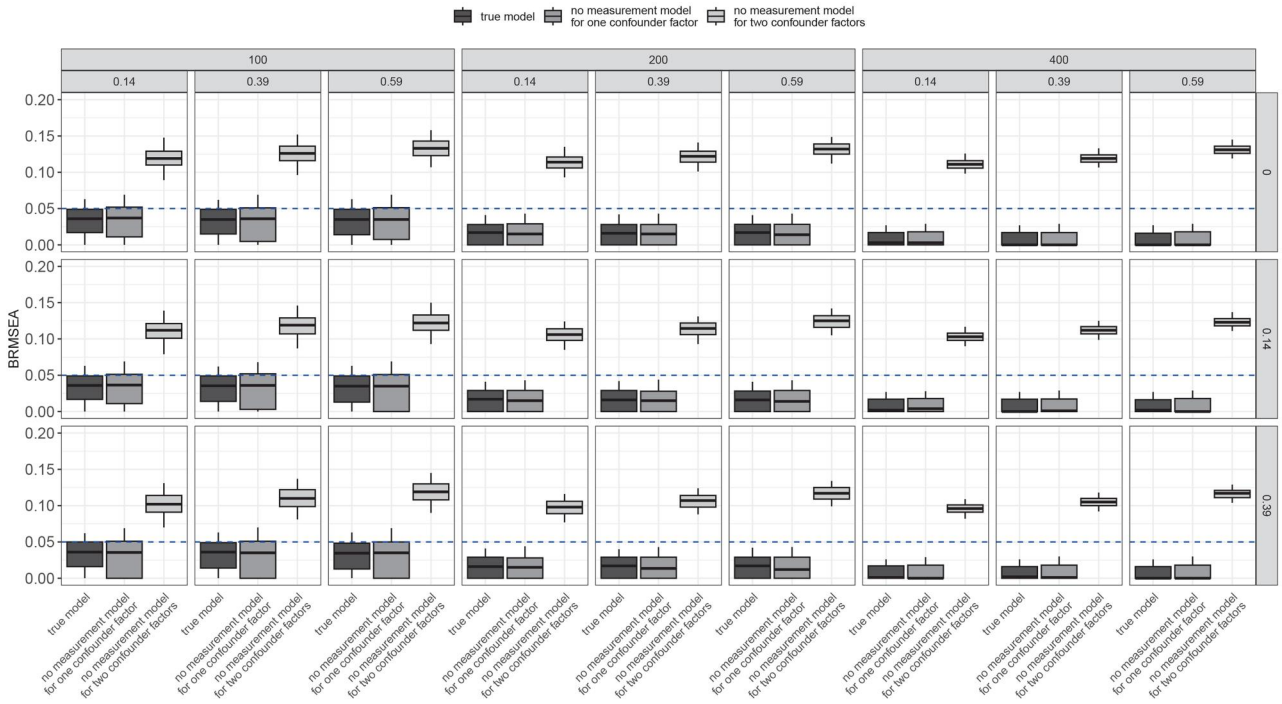


(b) *Rate for detecting the misfit using PPP with cutoff values 0.05, 0.10, and 0.15*

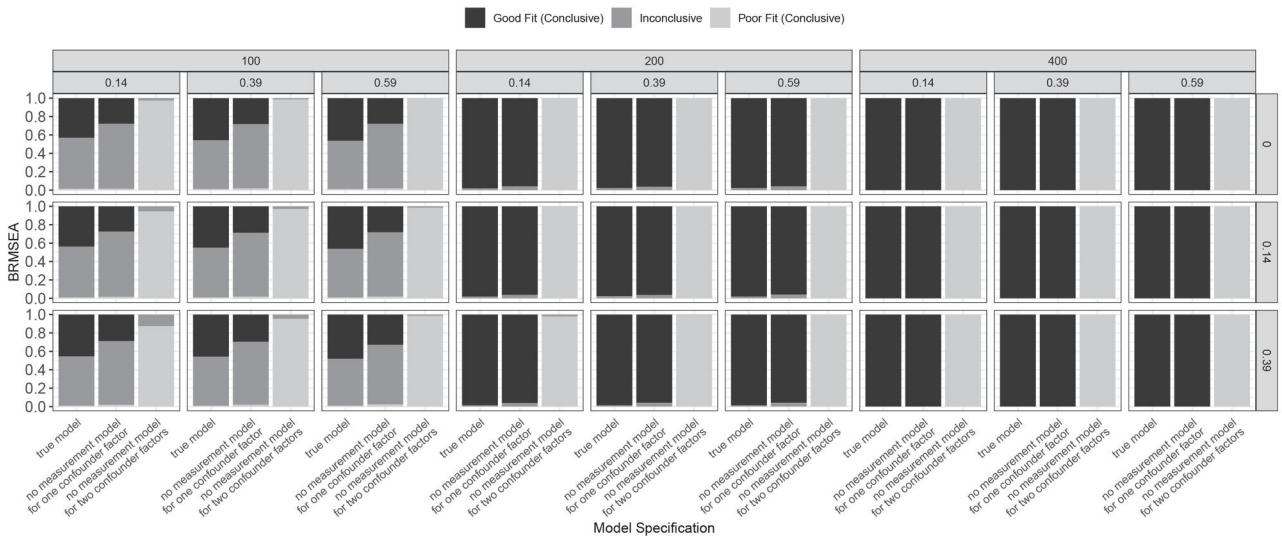
Figure 6. Bayesian PPP for different model specifications of the mediator factor using the diffuse priors.

6.1.1. Distribution of BRMSEA

Figure 7a shows that the model with no measurement model for one confounder factor has slightly higher and more spread-out BRMSEA values compared to the correctly specified model. The absence of a measurement model for both confounder factors significantly increases BRMSEA values, indicating the poorest fit among the models. BRMSEA values rise slightly as the mediation effect strengthens and the confounding level decreases, suggesting that a stronger mediation effect amplifies the misfit caused by the misspecified measurement model of confounders.



(a) *Box-and-whisker plots of BRMSEA values*



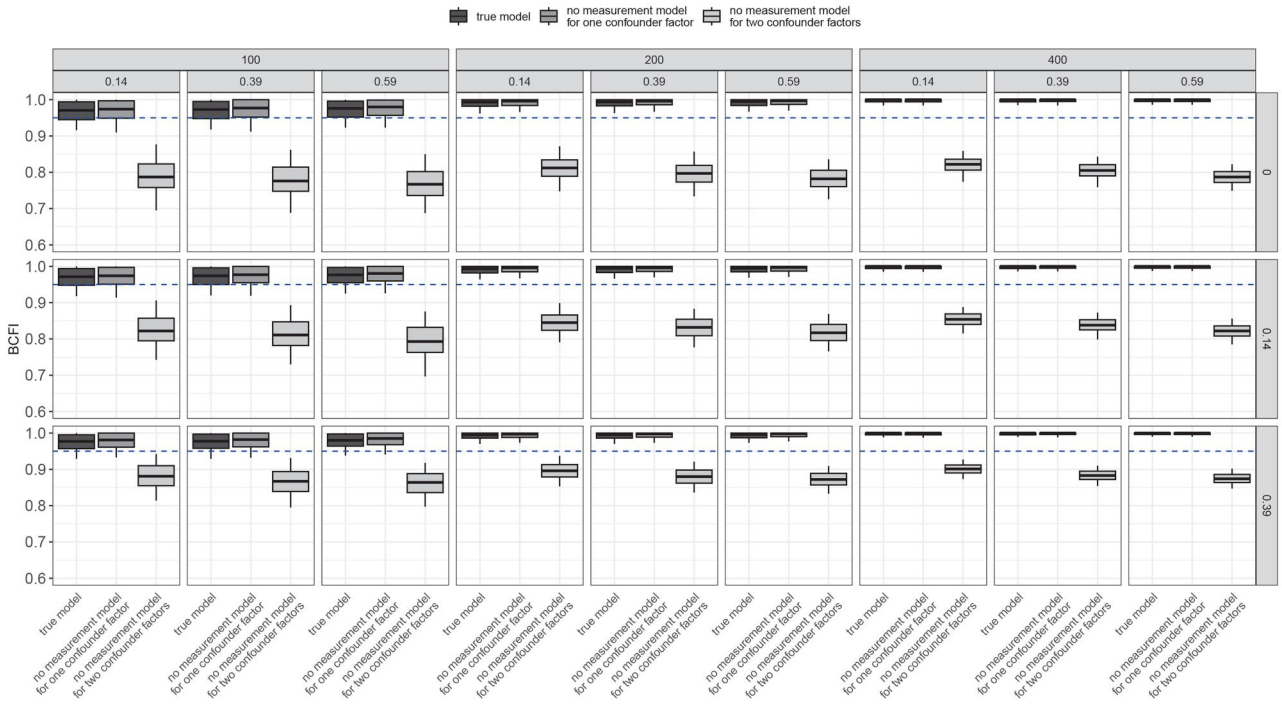
(b) *Model fit classification using BRMSEA 90% credible intervals.*

Figure 7. Bayesian RMSEA for different model specifications of the confounding factors using the diffuse priors.

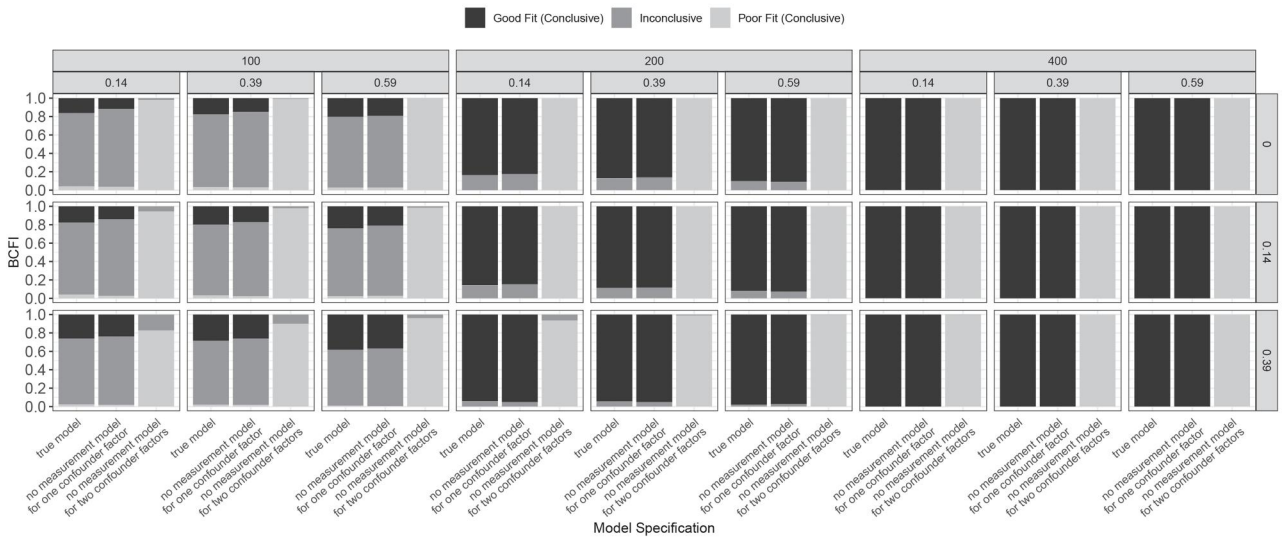
6.1.2. Model Fit Classification Using 90% Credible Intervals

Figure 7b indicates that when the sample size is 100, the proportion of replications classified as having a "good fit" is approximately 50% even for the true model across different conditions; on the other hand, the remaining replications are classified as "inconclusive." When only one confounder has no measurement model, about one-third of replications are still incorrectly classified as having a "good fit," with the rest being "inconclusive" mostly. Only when both confounders have no measurement models are almost all or all replications classified as having a "poor fit."

When the sample size increases to 200 and 400, regardless of the magnitudes of the mediation or confounder paths, all replications for the correctly specified model and the model without a measurement model for one confounder are classified as conclusively having a "good fit." On the other hand, for the misspecified model with no measurement models for two confounder factors, replications are dominated by conclusively "poor fit"



(a) *Box-and-whisker plots of BCFI values*



(b) *Model fit classification using 90% credible intervals of BCFI*

Figure 8. Bayesian CFI for different model specifications of the confounding factor using the diffuse priors.

classifications. These patterns suggest that the credible interval-based methods can correctly classify the true model when the sample size is at least 200, whereas they may not be effective at detecting model misfit when the measurement model for one confounder factor is ignored.

6.2. BCFI and BTLI

Because the BCFI and BTLI share similar patterns and they are organized in one subsection. The box-and-whisker plots and 90% credible interval-based misfit detection rates of the Bayesian CFI (BCFI) values are reported in Figure 8a and b. The columns represent the magnitudes of the mediation paths nested within sample sizes. The rows correspond to the magnitudes of the confounding effects.

6.2.1. Distribution of BCFI

Figure 8a presents the box-and-whisker plots of the BCFI for models with misspecified measurement models of confounders.

The analysis reveals several key patterns. Ignoring the measurement model of the confounder ζ_{c1}, which influences

the path between the independent variable ξ and the mediator η_M, results in minimal misfit. The BCFI values for this scenario are comparable to those of the correctly specified model, indicating a good fit. However, when the measurement models for both confounders are neglected, the BCFI values drop significantly, suggesting a poor fit. This decrease in BCFI highlights the detrimental effect of ignoring the measurement structure for multiple confounders on the overall model fit.

6.2.2. Model Fit Classification Using 90% Credible Intervals

According to the model misfit detection rates shown in Figure 8b, when the sample size is the smallest, the proportion of replications classified as "good fit" is approximately 20% for the true model when the mediation paths are 0.14 or 0.39, but this percentage increases slightly when the mediation path is 0.59 and the confounder path is 0.39. A similar pattern is found when the measurement model of only one confounder is ignored, indicating the poor performance of credible interval-based methods. However, when the measurement models of two confounder factors are ignored, most of the replications are classified as "poor fit."

When the sample size increases to 200 and 400, the patterns for the correctly specified model and the model where one confounder has no measurement model remain very similar, with most replications classified as "good fit," meaning that the misspecified model where one confounder's measurement model is ignored cannot be detected. In contrast, when both confounder factors have no measurement models, the credible interval-based methods classify the model as "poor fit."

6.2.3. BTLI vs. BCFI

The box plots and the rates of detecting model misfit using the Bayesian TLI (BTLI) values are shown in Figure 9a and b.

According to Figure 9a, the BTLI values for the severely misspecified models, where the measurement models of both confounder factors were ignored, are lower than the corresponding BCFI values. This suggests that BTLI can detect such severely misspecified models. However, the values and patterns remain similar between the correctly specified model and the misspecified model where the measurement model for one confounder is ignored.

Figure 9b further illustrates that, compared to BCFI, BTLI classifies a higher proportion of replications as "inconclusive," particularly when the sample size is 200. In this case, approximately 20% to 30% of replications fall into the "inconclusive" category—higher than the corresponding proportion for BCFI.

6.3. $\hat{\Gamma}$ and $\hat{\Gamma}_{\text{adj}}$

In this subsection, we examine the performance $\hat{\Gamma}_{\text{adj}}$ in detecting model misspecifications in the latent confounders. The two fit indices have similar pattern, We thus detail the results about $\hat{\Gamma}_{\text{adj}}$.

Figure 10a and b present the box-and-whisker plots and the proportions of model fit classification based on credible intervals of $\hat{\Gamma}_{\text{adj}}$. The horizontal dashed line represents a value of 0.95, serving as a reference for assessing model fit. In both figures, the columns represent different sample sizes (i.e., 100, 200, 400) nested within varying magnitudes of the mediation effects (i.e., $a = b = 0.140.39, 0.59$), while the rows correspond to different levels of confounding effects (no = 0, small = 0.1, medium = 0.39).

6.3.1. Distribution of $\hat{\Gamma}_{adj}$

According to the box-and-whisker plots in Figure 10b, $\hat{\Gamma}_{\text{adj}}$ consistently exhibits high values, clustering above 0.95 for both the correctly specified model and the model with a misspecified measurement model for a single confounder.

However, when the measurement models for both confounders are omitted, $\hat{\Gamma}_{\text{adj}}$ values show a notable decrease, with the entire box plots below 0.95. This shift suggests that $\hat{\Gamma}_{\text{adj}}$ becomes more responsive when the measurement models for both confounders are ignored.

Additionally, for misspecified models, $\hat{\Gamma}_{\text{adj}}$ values exhibit a slight downward trend as the mediation effect strengthens, indicating a marginal increase in its sensitivity to model misfit under these conditions.

6.4. Model Fit Classification Based on 90% Credible Intervals

Figure 10b shows that when the sample size is 100, most replications are classified as "good fit" or "inclusive" for both the true model and the model without a measurement model for one confounder factor.

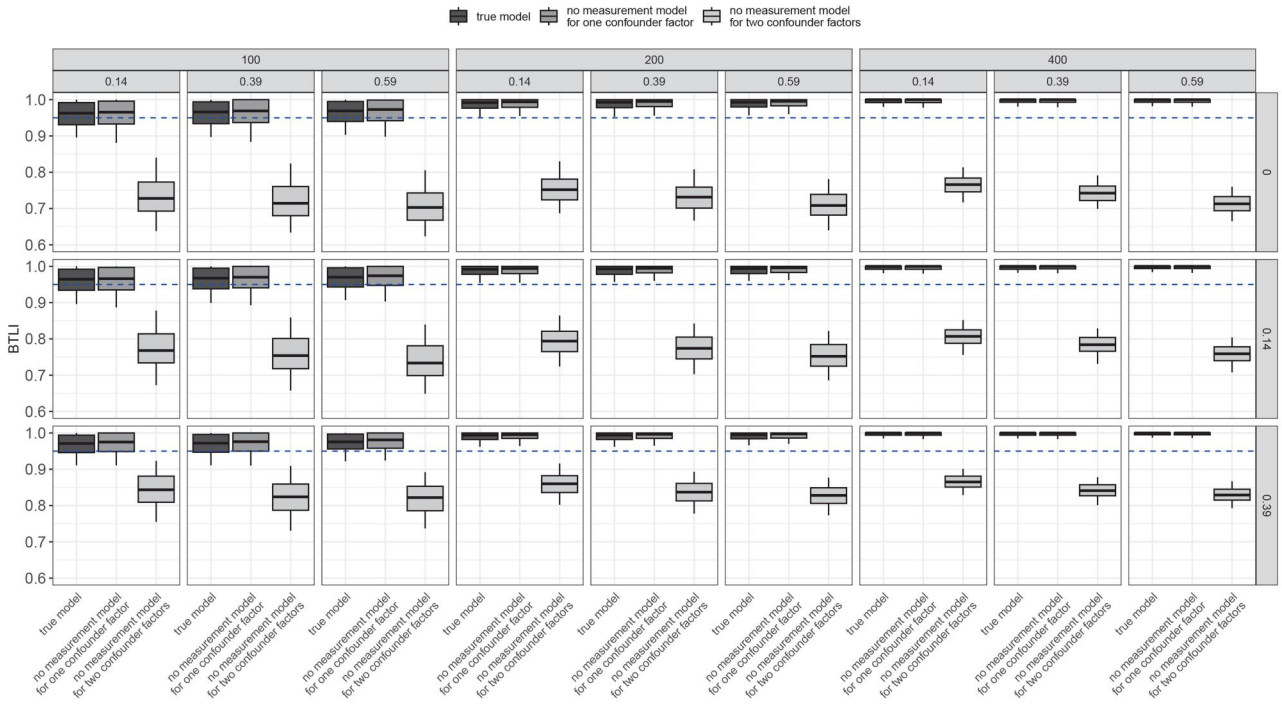
For the model without measurement models for two confounder factors, more than 90% replications are classified as "poor fit." The observed patterns remain generally similar for larger sample sizes of 200 and 400.

6.5. PPP

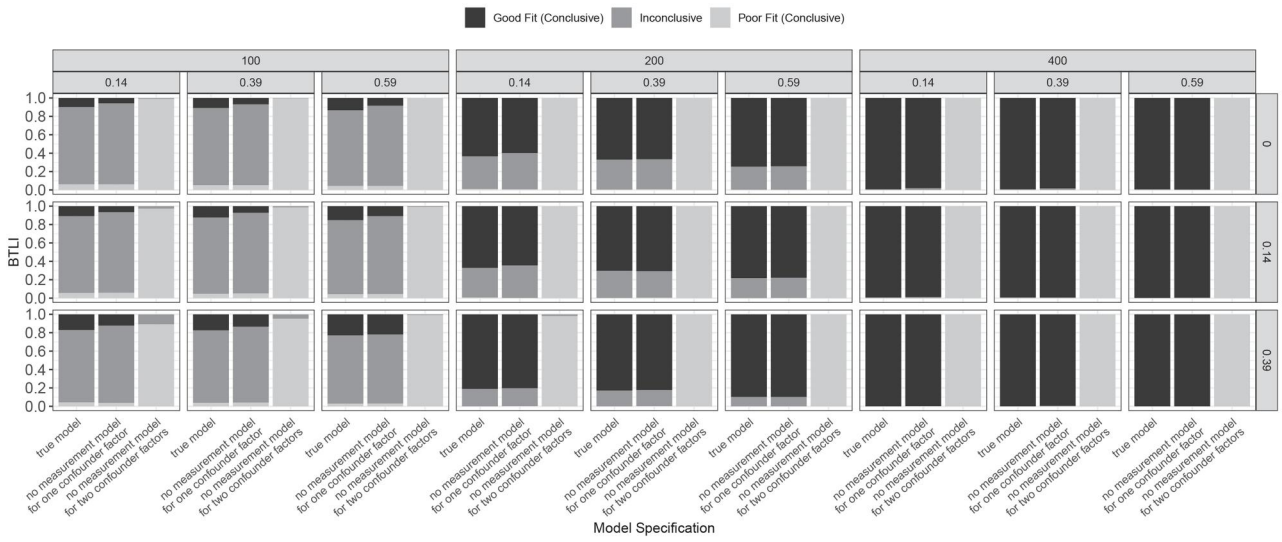
We report the box-and-whisker plots and rejection rates based on different cutoffs for PPP values in Figure 11a and b. In Figure 11a, the columns represent different sample sizes and magnitudes of the mediation effect, and the rows indicate different magnitudes of the confounder effects. In Figure 11b, the columns represent three different cutoff values (0.05, 0.10, and 0.15) with the magnitudes of the mediation effects, and the rows indicate different magnitudes of the confounder effects.

6.5.1. Distribution of PPP

The patterns in Figure 11b demonstrate that both the correctly specified model and the model omitting the measurement model for a single confounder consistently yield PPP



(a) *Box plots of BTLI values*



(b) *Rate for detecting the misfit using BTLI*

Figure 9. Bayesian TLI for different model specifications of the confounders using the diffuse priors.

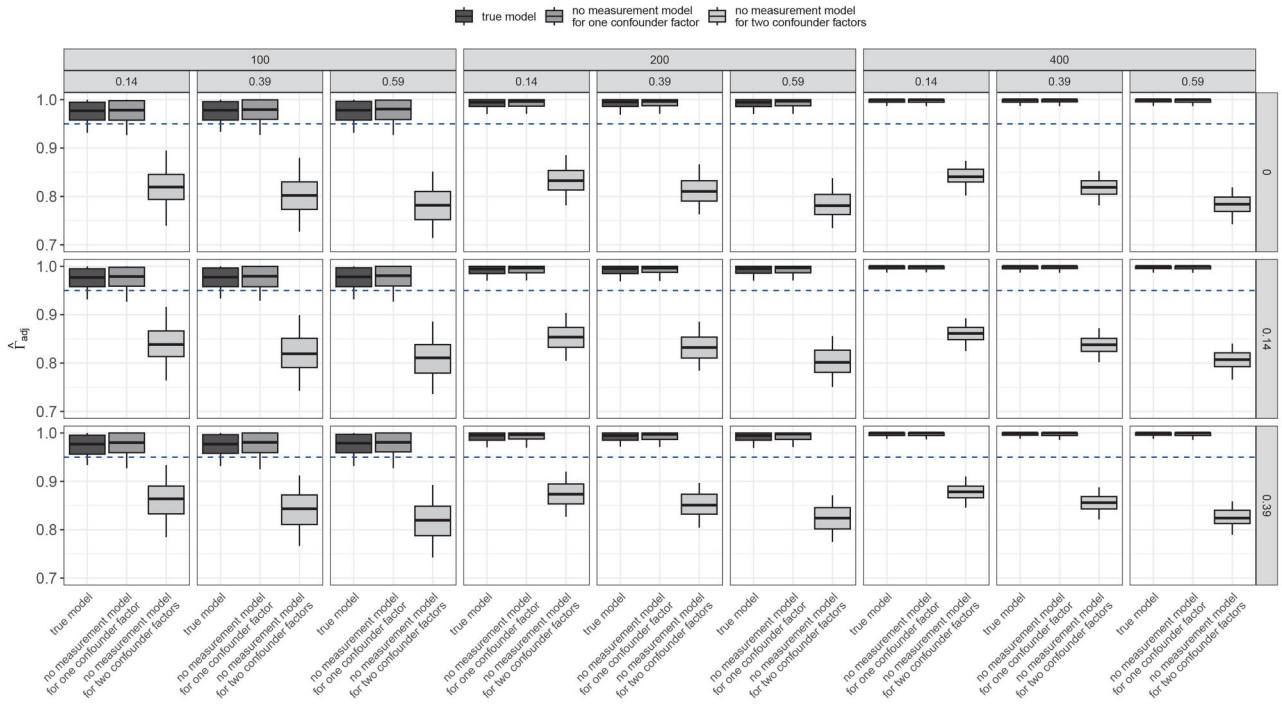
values around 0.5 across all conditions. This stability persists regardless of the mediation path strength or the magnitude of confounding effects, making it challenging to detect the misspecification of a single confounder.

In contrast, the most severe model misspecification, where the measurement models for both confounder factors are omitted, results in PPP values dropping below 0.05 across all conditions. The severity of misfit becomes more pronounced as the mediation effect strengthens and the confounding effect decreases, further lowering PPP values and indicating a substantial deviation from a well-specified model.

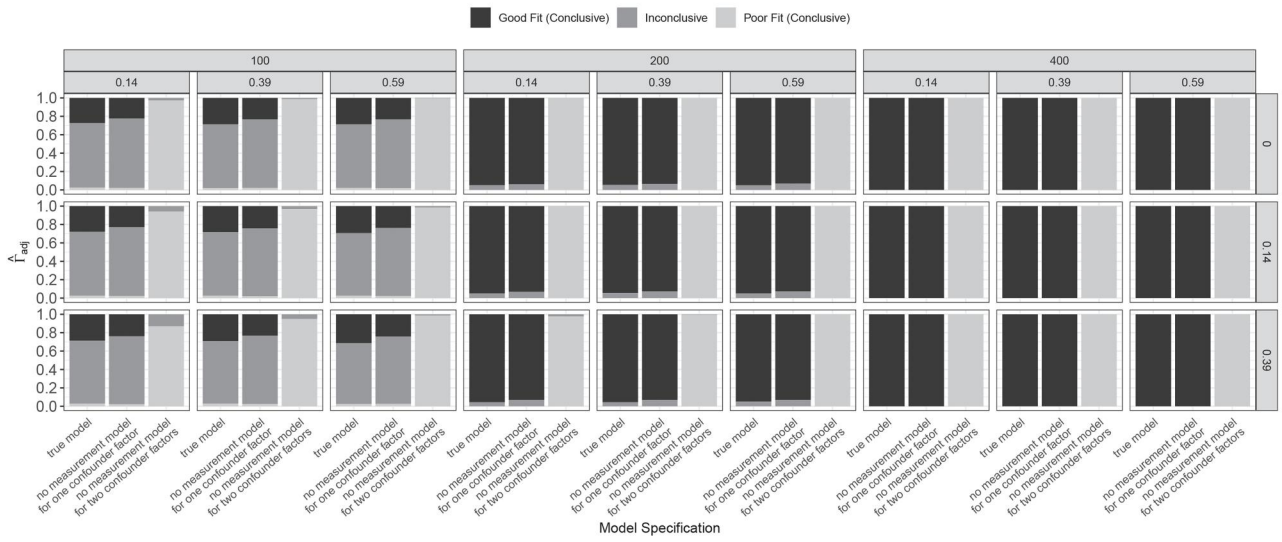
6.6. Rejection Rates of PPP with Different Cutoffs

Figure 11b presents the rejection rates for detecting model misfit using PPP.

For models where both confounders are misspecified, the rejection rate exceeds 0.95 across all conditions and cutoff values, indicating that PPP is highly effective in identifying severe misspecifications. However, when only a single confounder is misspecified, the rejection rates remain unacceptably low, suggesting limited sensitivity to minor model misspecifications.



(a) *Box plots of* $\hat{\Gamma}_{adj}$ *values*



(b) *Rate for detecting the misfit using* $\hat{\Gamma}_{adj}$

Figure 10. $\hat{\Gamma}_{adj}$ for different model specifications of the confounders using the diffuse priors.

These findings indicate that PPP is more responsive to severe misspecifications in latent confounders but struggles to detect more subtle forms of model misfit.

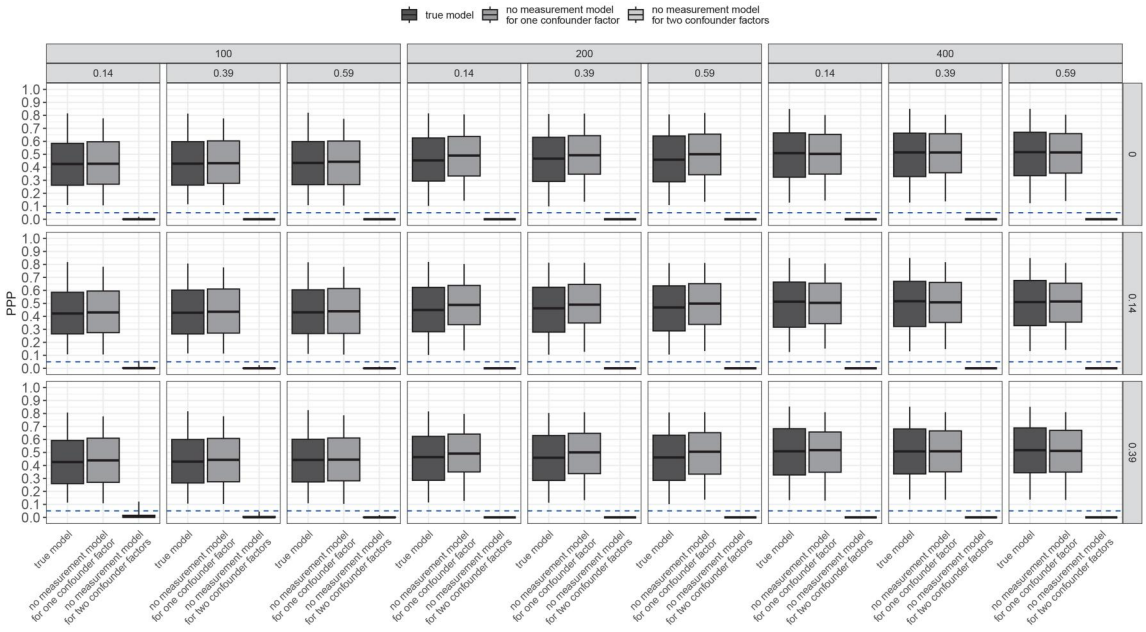
6.7. Summary

When only the confounder ξ_{c1}, which affects the path from the independent variable η_{ξ} to the mediator η_M, is misspecified by ignoring its measurement model, none of the fit indices under investigation effectively detect the misspecification. In this case, PPP is particularly less effective compared to the Bayesian SEM approximate fit indices in identifying the model misfit.

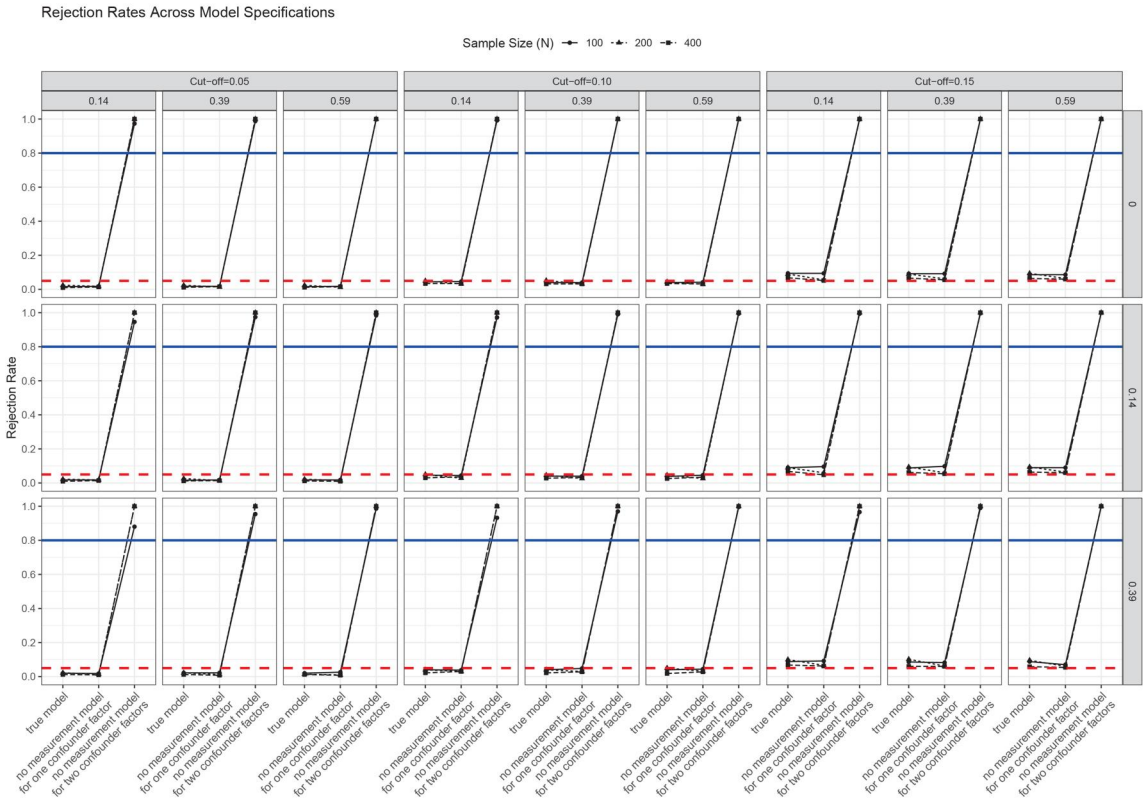
However, when both confounders, ξ_{c1} and ξ_{c2}, which influence the paths from η_{ξ} to η_M and from η_M to the outcome variable $\eta_{Y,}$ are misspecified, all fit indices, including PPP, exhibit strong sensitivity in detecting the misfit.

7. Simulation Result III: Ignoring True Confounders

This section examines how BRMSEA, BCFI, BTLI, $\hat{\Gamma}$, $\hat{\Gamma}_{adj}$, and PPP detect the omission of true confounders in



(a) *Box plots of PPP values*



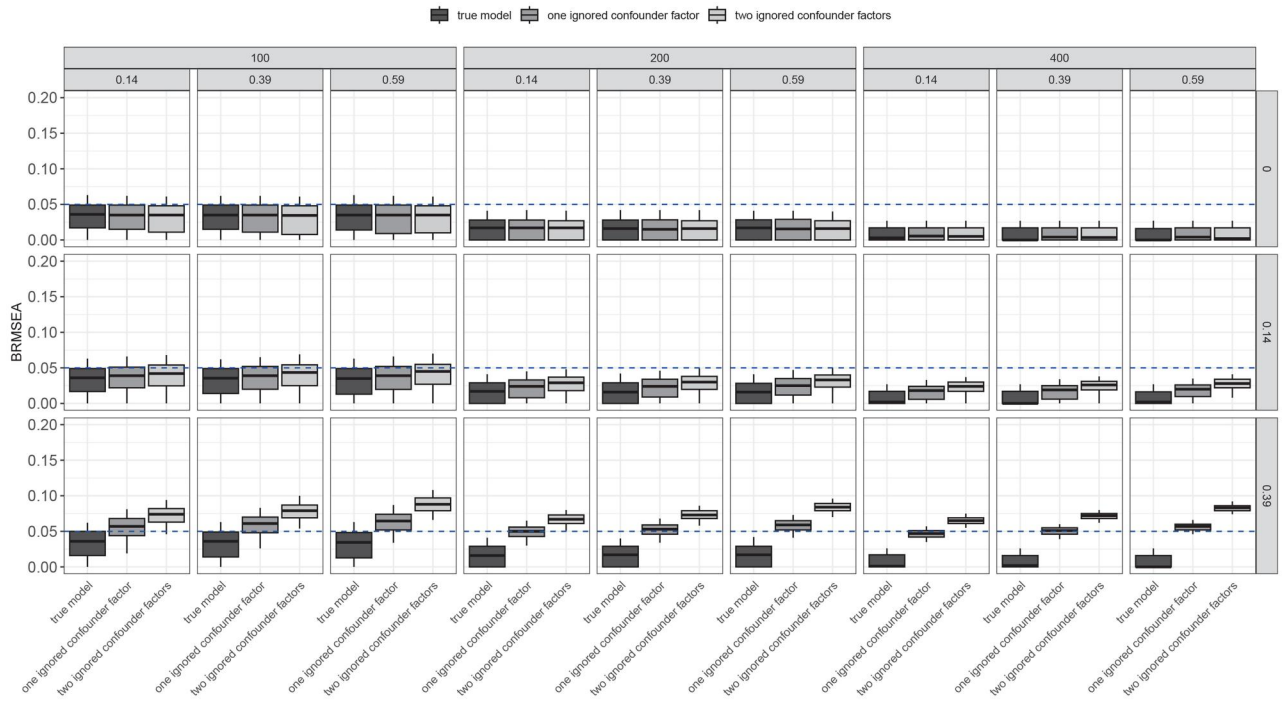
(b) *Rate for detecting the misfit using PPP*

Figure 11. PPP for different model specifications of the confounders using the diffuse priors.

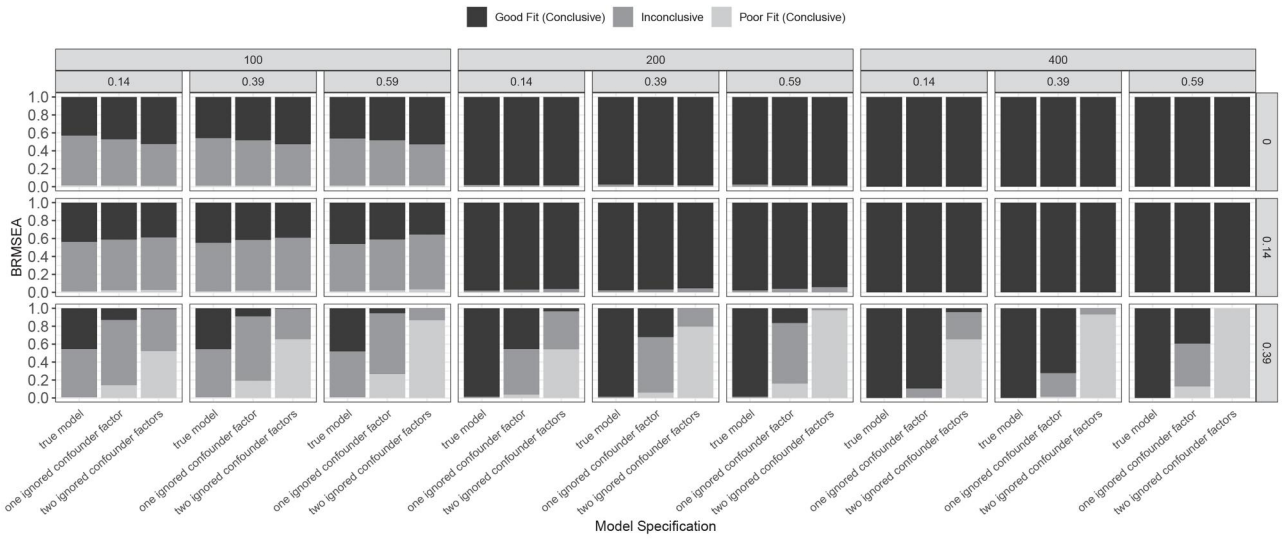
mediation analysis. We analyze the distribution of these fit indices using box-and-whisker plots that display the 5%, 25%, 50%, 75%, and 95% percentiles of their empirical distributions. In addition, we report the proportions of models classified as "good fit," "inconclusive," and "poor fit" based on the 90% credible interval, providing insights into the effectiveness of these indices in identifying confounder omission.

7.1. BRMSEA

The box-and-whisker plots of the values of BRMSEA and proportion of model fit classification are demonstrated in



(a) *Box-and-whisker plots of BRMSEA values*



(b) *Model fit classifications based on the 90% credible intervals of BRMSEA*

Figure 12. Bayesian RMSEA for models omitting confounders using the diffuse priors.

Figure 12a and b. The columns reflect the levels of the mediation paths and the sample sizes, and the rows reflect the levels of the confounding effects (no = 0, small = 0.14, medium = 0.39).

7.1.1. Distribution of BRMSEA

Based on the box plots in Figure 12a, when no confounding effect is present (i.e., confounding effect = 0), omitting confounders does not lead to model misspecification. Under these conditions, the 90% credible interval (CI) for BRMSEA remains below the threshold of 0.05, particularly for sample sizes of 200 and 400. However, with a smaller sample size (e.g., 100), approximately 25% of replications yield BRMSEA values above 0.05, suggesting greater variability in fit assessment.

As the confounding effect increases, omitting a confounder results in higher BRMSEA values. When the confounding effect is small (e.g., 0.14), the BRMSEA values are slightly elevated compared to the correctly specified. With a medium confounding effect (e.g., 0.39), the BRMSEA values increase more substantially, and the entire box plot shifts above 0.06, indicating poor model fit. This trend becomes more pronounced as the mediation effect strengthens and the sample size increases.

7.1.2. Model Fit Classification Based on 90% Credible Interval of BRMSEA

Figure 12b are the stacked bar plots of the proportions of "good fit", "inconclusive", and "poor fit" groups based on the 90% credible intervals of the BRMSEA.

When the confounding effect is 0 or small (i.e., 0.14), the proportion of "good fit" classifications remains close to one for sample sizes of 200 and 400. With a sample size 100, around 50% replications are incorrectly classified as having a "good fit" with the remaining as "inclusive." This suggests that the interval-based approach using BRMSEA effectively identifies well-specified models while struggling to detect the omission of confounders with minimal influence on the mediation pathway.

As the confounding effect increases to a medium level (i.e., 0.39), a notable shift occurs. The proportion of models omitting one confounder being classified as a "good fit" drops significantly-falling to around 40% with a moderate mediation path and to less than 10% with a strong mediation path. When both confounders are omitted, the proportion of "good fit" classifications is nearly zero. Meanwhile, the proportion of "poor fit" classifications rises sharply, reaching approximately 100% when the sample size is 200 or 400. This pattern suggests an interaction between the magnitude of the mediation effect and the detectability of the omission of confounders, where stronger mediation effects enhance the sensitivity of BRMSEA.

7.2. BCFI and BTLI

Since BCFI and BTLI function similarly, we present them within the same subsection, focusing primarily on the interpretation of BCFI results while highlighting key distinctions between BTLI and BCFI.

The box plots and rejection rates using the BCFI values are reported in Figure 13a and b. The columns and rows reflect the levels of the mediation paths and sample sizes, and the confounding effects. The horizontal dash line lies at 0.95.

7.2.1. Distribution of BCFI

The box plots in Figure 13a illustrate the distributions of BCFI.

For the correctly specified model and models omitting confounders with a small effect (i.e., 0 and 0.14), BCFI values remain above 0.95 in most cases. However, at a sample size of 100, the lower whisker falls below 0.95, indicating increased variability. This suggests that BCFI has limited sensitivity in detecting the omission of confounders with minimal influence on the mediation path.

As the confounding effect increases to a medium level (0.39), BCFI values decline significantly for models omitting confounders. The median BCFI falls below 0.95 when one confounder is omitted, while the entire distribution drops below 0.95 when both confounders are omitted, indicating a clear deterioration in model fit. As the sample size increases, the variability of BCFI decreases.

7.2.2. Model Fit Classification Based on 90% Credible Interval of BCFI

The model fit classification using BCFI is illustrated in Figure 13b.

For the correctly specified model and models omitting confounders with no or small effects (i.e., 0 or 0.14), most replications fall into the "good fit" category when the sample size is 200 or larger. However, with a sample size of 100, a greater proportion are classified as "inconclusive," indicating that BCFI has limited sensitivity to the omission of minor confounders.

As the confounding effect increases to 0.39 (medium-sized), the "good fit" classification decreases when confounders are omitted. When one confounder is omitted, approximately 40–50% of replications are classified as "poor fit," with the remaining as "inconclusive." When two confounders are omitted, nearly 100% of replications indicate "poor fit," highlighting BCFI's effectiveness in detecting severe confounder omissions.

Smaller sample sizes introduce more uncertainty, leading to a higher proportion of "inconclusive" classifications, even for correctly specified models.

7.2.3. Comparison of BTLI and BCFI

The results of the BTLI are presented in Figure 14a and b.

The distribution and model fit classification of BTLI follow a similar pattern to BCFI. However, BTLI shows slightly greater sensitivity to confounder omission. Its values shift downward more noticeably, indicating a stronger response to model misspecifications. Additionally, a higher proportion of replications are classified as "poor fit" with BTLI compared to BCFI, suggesting BTLI is more effective in detecting the negative impact of confounder omission on model fit.

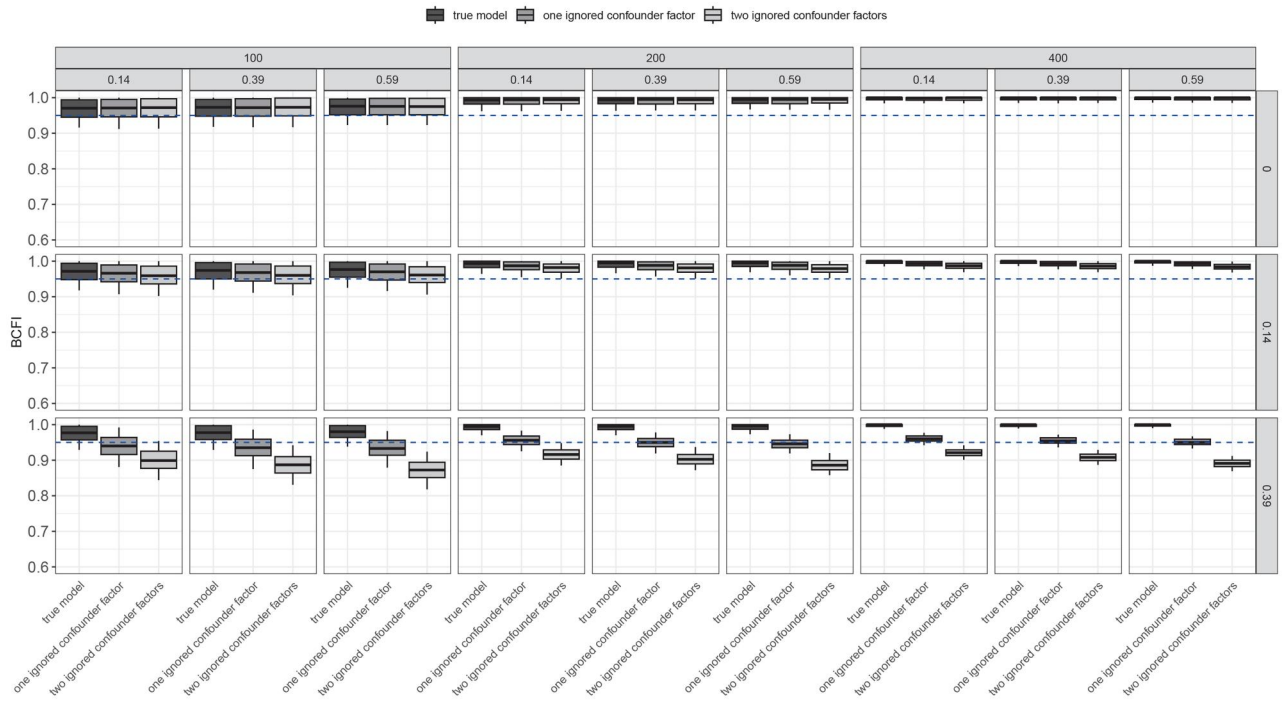
7.3. $\hat{\Gamma}_{adj}$

Figure 15a and b illustrate the distribution and model fit classifications based on $\hat{\Gamma}_{adj}$, respectively. The plots are structured according to the magnitude of the indirect paths, sample size, and levels of confounding effects, providing a comprehensive view of how these factors influence model fit evaluation.

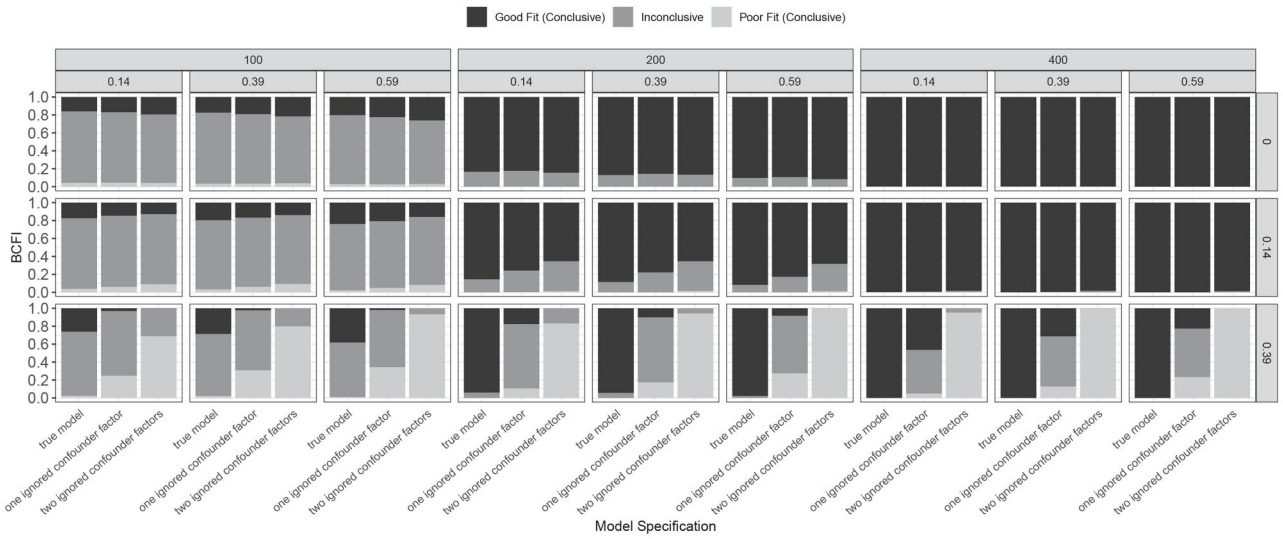
7.3.1. Distribution of $\hat{\Gamma}_{adj}$

When the confounding effect is 0 or 0.14 (small), $\hat{\Gamma}_{adj}$ values remain consistently high for both the correctly specified model and the model omitting confounders, with minimal variation. As a result, the entire distribution of $\hat{\Gamma}_{adj}$ stays above 0.95 for sample sizes of 200 and 400. However, for a sample size of 100, the lower whisker falls below 0.95, indicating increased variability and reduced sensitivity in smaller samples.

When the confounding effect reaches a medium level (0.39), the distributions of $\hat{\Gamma}_{adj}$ become more distinct across model conditions. The central tendency of $\hat{\Gamma}_{adj}$ drops below



(a) *Box-and-whisker plots of the BCFI values*



(b) *Model fit classification based on the 90% credible intervals of BCFI.*

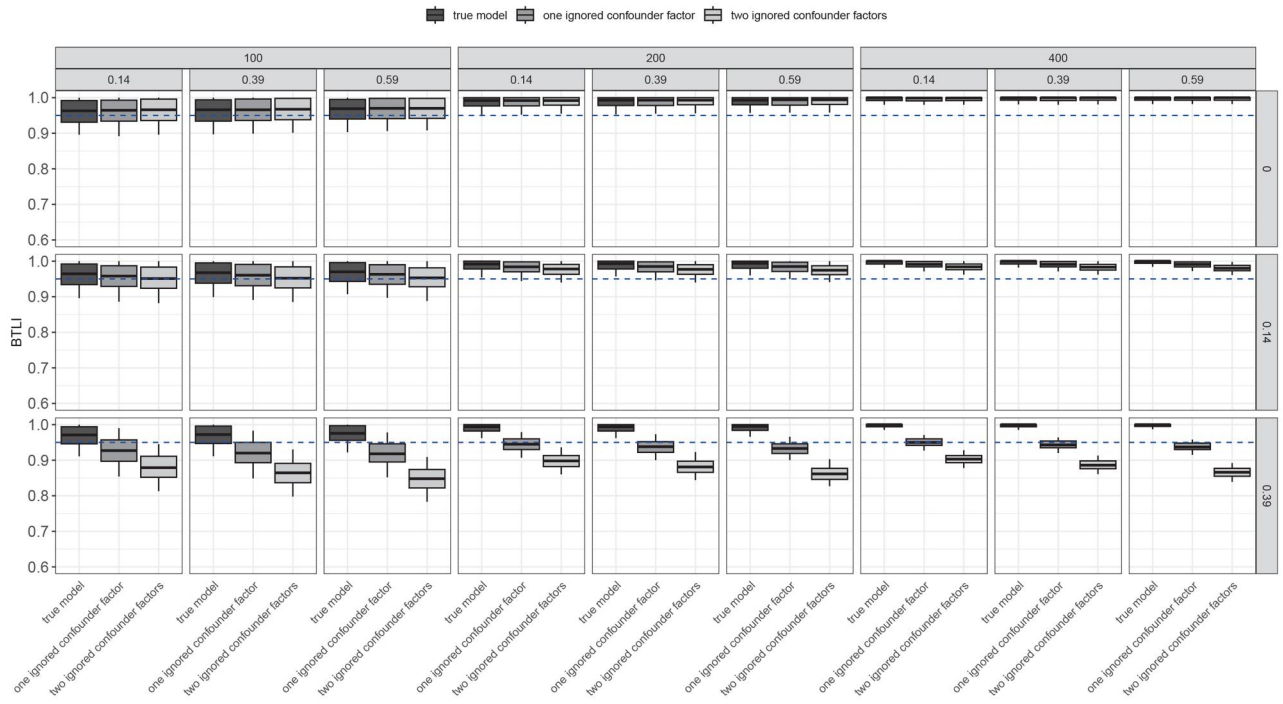
Figure 13. Bayesian CFI for omitting confounders using the diffuse priors.

0.95 when omitting one confounder and falls further below 0.90 when omitting two confounders, reflecting greater sensitivity to model misspecification. Additionally, as the sample size increases, the variability of $\hat{\Gamma}_{\text{adj}}$ decreases, leading to more stable estimates.

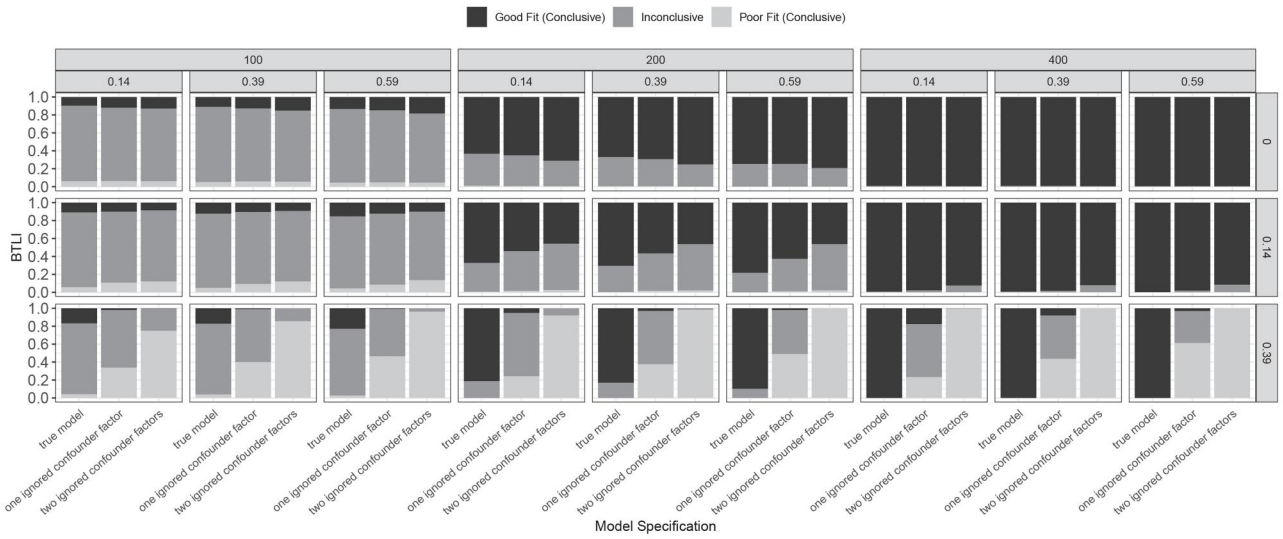
7.3.2. Model Classification Using the 90% Credible Interval of $\hat{\Gamma}_{adj}$

When the confounding effect is absent or small (i.e., 0.14), both the correctly specified model and the model omitting confounders are predominantly classified as "good fit" for sample sizes of 200 and 400. However, with a sample size of 100, a larger proportion of replications fall into the "inconclusive" category, reflecting increased uncertainty due to smaller sample sizes.

When the confounding effect is medium (i.e., 0.39), fewer than 5% of replications are classified as "good fit" when one confounder is omitted. When both confounders are omitted, the misfit becomes so severe that nearly 100% of replications are classified as "poor fit." With a small sample size of 100, a larger proportion of cases fall into the "inconclusive" category, due to the large variability in the posterior distribution.



(a) *Box-and-whisker plots of BTLI values*



(b) *Model fit classification using the 90% credible intervals of BTLI.*

Figure 14. Bayesian TLI for models omitting confounders using the diffuse priors.

7.4. PPP

Figure 16a and b demonstrate the distribution of PPP and the rate for detecting misfit using different cutoff values.

7.4.1. Distribution of PPP

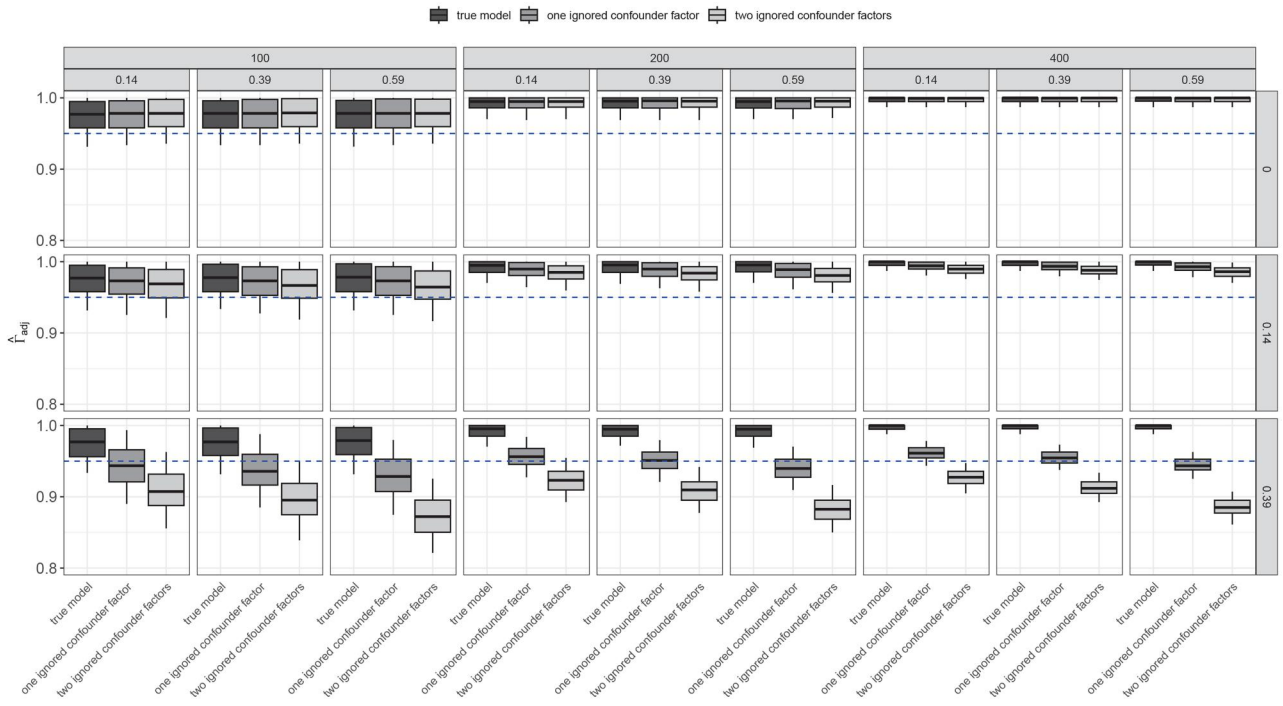
For correctly specified models, PPP values remain nicely around 0.5 across all conditions. This confirms PPP correctly identifies well-specified models.

When the confounding effect is small (i.e., 0.14), PPP values for the misspecified model shift downward but generally remain above 0.05. However, when the confounding effect is medium (i.e., 0.39), PPP values drop further below 0.05 for models omitting confounders, particularly when the sample size is 200 or 400 or when both confounders are omitted.

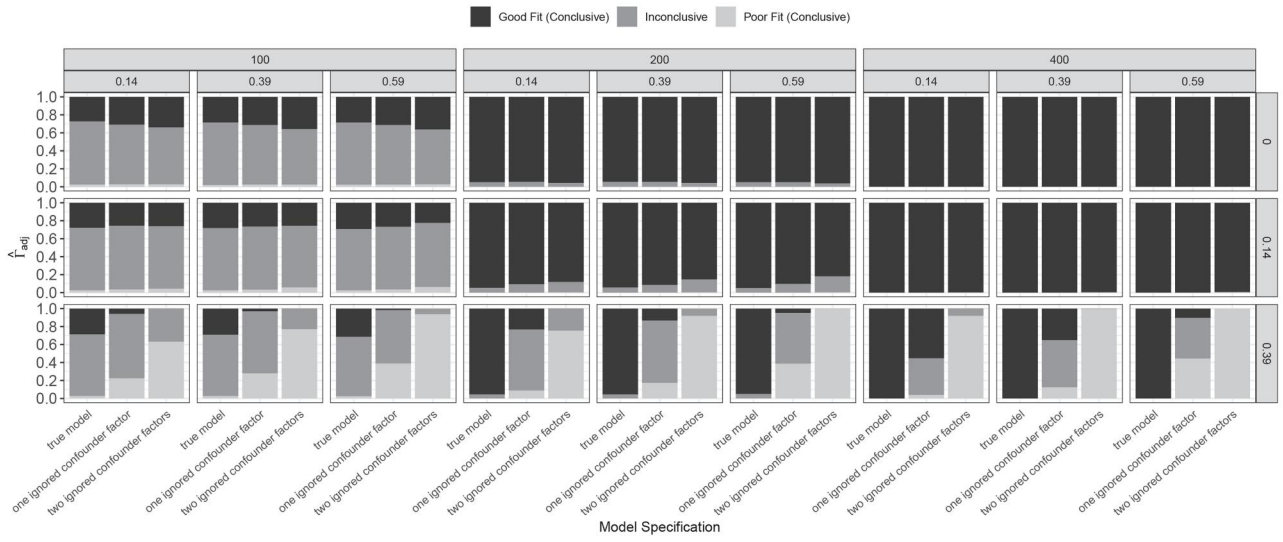
7.4.2. Rejection Rate Using PPP with Different Cutoffs

Figure 16b illustrates the misfit detection rates based on PPP.

A model is classified as lacking a good fit if its PPP falls below a specified cutoff value.



(a) *Box-and-whisker plots of the* $\hat{\Gamma}_{adj}$



(b) *Model fit classification based on the* 90% *credible intervals of* $\hat{\Gamma}_{adj}$

Figure 15. $\hat{\Gamma}_{adj}$ for models ignoring confounders.

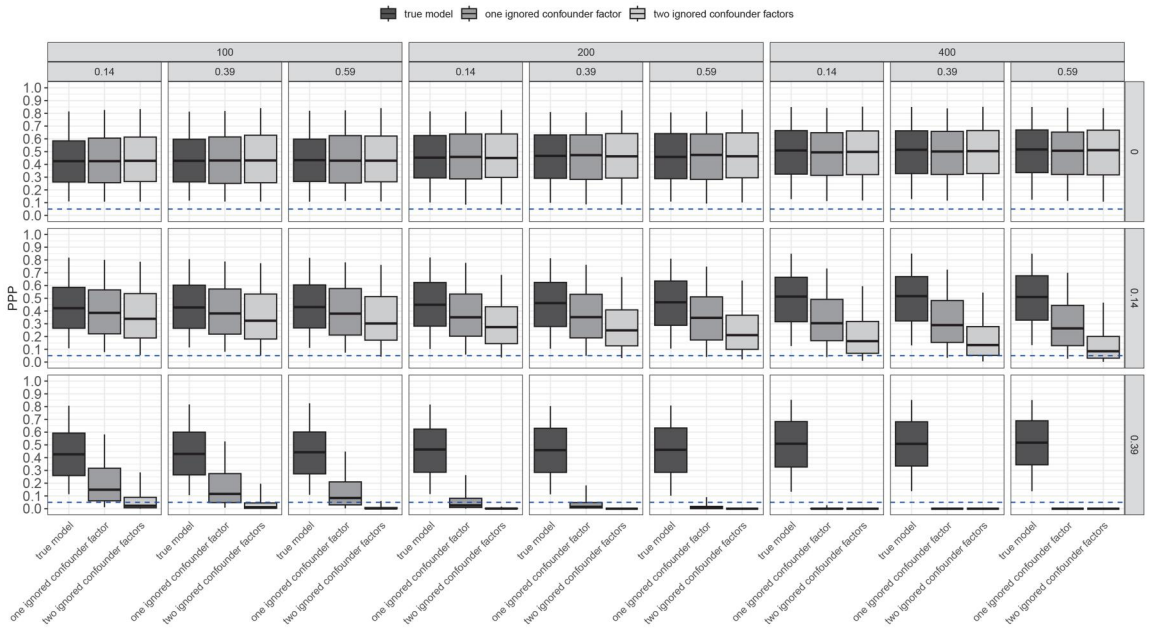
The rejection rate for a correctly specified model remains below 5% when using thresholds of 0.05 and 0.10, and increases to approximately 10% with a threshold of 0.15.

For models that ignore confounders, rejection rates rise as sample size increases, mediation effects strengthen, and confounding effects become more pronounced.

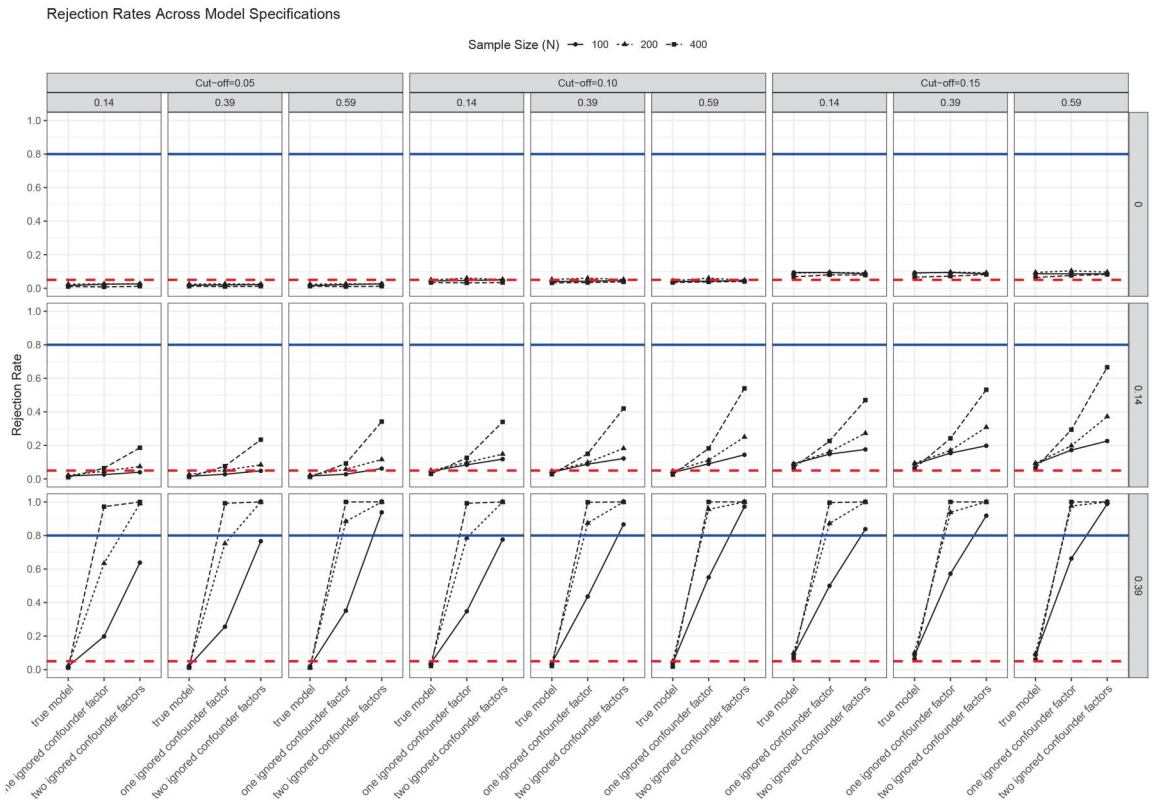
Ignoring two confounders leads to more severe model misspecification than omitting a single confounder. When the confounding effect is small (i.e., 0.14), the rejection rates for models ignoring two confounders range from 50% to 70% with a sample size of 400 and thresholds of 0.10 and 0.15.

When the confounding effect is medium (i.e., 0.39), the rejection rate for models omitting one confounder exceeds 80% when the sample size is 200 or larger. For models ignoring two confounders, the rejection rate surpasses 80% even with a sample size of 100 and reaches 95% when the sample size is 200 or above.

The choice of cutoff value impacts the rejection rates. Stricter thresholds (0.05) result in lower rejection rates, potentially overlooking moderate misfit, whereas more lenient thresholds (0.10,



(a) *Box plots of PPP values*



(b) *Rate for detecting the misfit using PPP*

Figure 16. PPP for omitting confounders using the diffuse priors.

0.15) increase sensitivity to misfit but also raise the likelihood of mistakenly rejecting a correctly specified model.

Additionally, stronger mediation effects amplify the detectability of confounder omission, particularly in smaller samples, highlighting an interaction between mediation strength and model misfit detection.

7.5. Summary

The effectiveness of fit indices in detecting the omission of confounders varies in sensitivity and performance.

BRMSEA, BCFI, BTLI, and $\hat{\Gamma}_{\text{adj}}$ effectively identify confounder omission when the confounding effect is substantial

(0.39), with a higher proportion of "poor fit" classifications as sample size increases. BTLI is slightly more sensitive than BCFI, showing a greater downward shift in values and a higher proportion of "poor fit" when confounders are omitted. $\hat{\Gamma}_{adj}$ is more reliable for confirming correctly specified models but slightly less effective in detecting model misspecification than BTLI.

PPP performs well in detecting confounder omission, particularly when the confounding effect is moderate (0.39), where rejection rates exceed 95% for sample sizes of 200 and above. However, it is less sensitive to smaller confounding effects or limited sample sizes.

Overall, BRMSEA, BCFI, BTLI, and $\hat{\Gamma}_{adj}$ are robust in identifying moderate to severe confounder omissions, while PPP excels in detecting extreme cases.

8. Prior Sensitivity Analysis

To evaluate the sensitivity of each index to prior specification, we conducted factorial ANOVA to estimate the percentage of sampling variance η^2 attributable to prior specification. Table 2 values for prior specification, model specification, and their interaction, based on a sample size of 100.

Based on the results in Table 2, prior specification contributes a very small proportion of variance across all fit indices, with values consistently below 0.003. This suggests that different priors have little impact on the variation in fit indices compared to model specification.

The model specification consistently explains the largest proportion of variance in all fit indices across conditions. The variation accounted for by the model specification increases as the mediation effect increases but decreases as the confounding effect (i.e., β_k) increases.

Among all the indices under investigation, the greater portion of the variation in $\hat{\Gamma}_{adj}$ is explained by the model specification than other indices.

9. Discussion and Conclusion

Mediation analysis is a widely used approach for examining causal relationships between independent and dependent variables. This framework has been extended to incorporate latent variables within mediation paths, allowing for a more comprehensive representation of complex psychological and social processes. However, valid inferences in latent mediation analysis hinge on two critical assumptions: the correct specification of the measurement model for latent variables and the absence of unmeasured confounders. Violations of these assumptions introduce model misspecifications, potentially distorting results and undermining the credibility of findings.

This raises a fundamental question: How can approximate fit indices and PPP serve as diagnostic tools to help researchers evaluate model fit and detect these violations? Addressing this question is essential for improving the reliability of mediation analysis in applied research.

9.1. Highlights of the Results

In response to the practical need for detecting model assumption violations in latent mediation models, this study systematically evaluates the performance of Bayesian approximate fit measures and the posterior predictive p-value (PPP) in identifying model misspecifications in the presence of latent confounders. We examine various sources of misspecification, including errors in the measurement model of the mediator, misspecifications in the measurement model of latent confounders, and the omission of confounders. Additionally, we assess the sensitivity of these fit measures to different prior specifications. Our simulation study employs a comprehensive design, varying mediation effects, confounding effects, sample sizes, and prior specifications to provide a robust assessment of fit index performance across diverse conditions.

The study revealed that ignoring the mediator's measurement structure and using the standardized total score led to

Table 2. Percentage of variance η^2 in fit indices explained by model specification (MS) and prior specification (prior) and their interaction when the sample size is 100.

	$\beta_k = 0$			$\beta_k = 0.14$			$\beta_k = 0.39$		
	MS	Prior	Interaction	MS	Prior	Interaction	MS	Prior	Interaction
					$a = b = 0.14$				
CFI	0.708	0.002	0	0.629	0.002	0	0.531	0.002	0
TLI	0.716	0.003	0.001	0.644	0.003	0.001	0.536	0.003	0
RMSEA	0.672	0.003	0	0.609	0.003	0	0.554	0.002	0
ppp	0.481	0.002	0.001	0.411	0.002	0.001	0.44	0.002	0.001
$\hat{\Gamma}_{adj}$	0.763	0.002	0.001	0.727	0.002	0.001	0.640	0.001	0.001
					$a = b = 0.39$				
CFI	0.752	0.001	0	0.679	0.001	0	0.603	0.001	0
TLI	0.761	0.002	0.001	0.695	0.002	0.001	0.608	0.002	0
RMSEA	0.707	0.002	0	0.65	0.002	0	0.619	0.002	0
ppp	0.502	0.002	0.001	0.431	0.002	0.001	0.49	0.001	0.001
$\hat{\Gamma}_{adj}$	0.798	0.001	0.001	0.764	0.001	0.001	0.700	0.001	0.001
					$a = b = 0.59$				
CFI	0.801	0.001	0	0.738	0.001	0.001	0.696	0.001	0
TLI	0.809	0.001	0.001	0.751	0.002	0.002	0.698	0.001	0
RMSEA	0.749	0.001	0	0.695	0.001	0	0.698	0.001	0
ppp	0.531	0.002	0.001	0.459	0.002	0.001	0.556	0.001	0.001
$\hat{\Gamma}_{adj}$	0.836	0.001	0.001	0.803	0.000	0.001	0.757	0.001	0.001

MS: model specification; Prior: prior specification; Interaction: MS × Prior; *a*, *b* represent the mediation paths, and β_k is the magnitude of the confounding effect.

severe misfits, which were effectively detected by BRMSEA, BCFI, BTLI, and $\hat{\Gamma}_{adj}$ with increased certainty with a larger sample size. When the sample size is 200 and above, the PPP is highly sensitive to misspecification of the measurement model with a high proportion of "poor fit." Additionally, omitting a cross-loading for the latent mediator also led to detectable misfit, as identified by both the Bayesian SEM approximate fit indices and PPP. However, with a sample size of 100, the increased uncertainty in model fit classification resulted in a larger proportion of replications being classified as "inconclusive."

Interestingly, the inclusion of true confounders reduces the misfit introduced by an incorrect measurement model for the mediator, making it more difficult to detect such misspecifications, especially when the confounding effect is moderate. Additionally, as the strength of the mediation path increases, the impact of measurement model misspecification on overall model fit becomes more pronounced, leading to greater misfit.

Ignoring the measurement model of the confounders also leads to misfits, which can be detected by the Bayesian approximate fit indices and PPP. The misfit is substantial only when the confounding effect is medium-sized or larger. When the confounding effect is absent or small, ignoring the measurement model of the confounder has little impact on the model misfit. Additionally, the misfit of the confounder to the mediator-outcome path has a stronger influence on the overall model fit.

The omission of the confounder of the mediator-outcome path also resulted in substantial misfits, particularly when the confounding effect was medium or large. BRMSEA, BCFI, BTLI, $\hat{\Gamma}_{adj}$, and PPP demonstrated high sensitivity to these omissions. The rates for classifying a misspecified model as "good fit" decrease as the coefficient of the indirect path increases.

Regarding the indices under investigation, the PPP is more sensitive to the influence of the sample size and performs outstandingly when the sample size is 200 or above. When the sample size is small (e.g., 100), the Bayesian approximate fit indices show better performance in detecting misfit. Regarding prior sensitivity, the choice of weakly informative prior has minimal effect on the variability of the fit indices.

9.2. Recommendation to Practitioners

For practitioners, our results provide clear guidance on choosing fit indices based on the type of model misspecification in latent mediation analysis. When detecting measurement model misspecifications, BTLI was the most sensitive, followed by BCFI, $\hat{\Gamma}_{adj}$, and BRMSEA, with PPP serving as a supporting indicator. When the confounder in the mediator-outcome path was misspecified, BTLI, BCFI, $\hat{\Gamma}_{adj}$, and BRMSEA performed similarly for sample sizes of 200 or larger. However, with smaller samples, BTLI and BCFI were more reliable than BRMSEA and $\hat{\Gamma}_{adj}$, as they had a lower tendency to falsely classify misspecified models as "good fit." In this scenario, PPP showed high sensitivity, achieving high rejection rates even at a sample size of 100. When detecting omitted confounders, all fit indices effectively identified the omission when the confounder had a moderate or strong effect. BTLI and BCFI were preferred over BRMSEA and $\hat{\Gamma}_{adj}$. PPP performed well but was most reliable when the sample size was at least 200.

We would like to note that while BCFI and BTLI showed strong and consistent performance across the conditions evaluated in the current study, their sensitivity to certain model features (e.g., degrees of freedom, model complexity) is well documented in the literature. For example, BTLI (or TLI) is sensitive to model complexity and tends to underestimate model fit in more complex models, whereas BCFI (or CFI) is designed to be less sensitive to model complexity but can still be influenced under certain conditions (Fan & Sivo, 2007).

All the fit indices and the PPP evaluated in the current study are available or extractable from widely used software such as M*plus* (Muthén & Muthén, 1998–2017) and the R `blavaan` package (Merkle & Rosseel, 2015). Given this accessibility, when applying these fit indices in practice, we recommend that researchers avoid relying on a single fit index. Instead, they should consider multiple indices to obtain a more comprehensive assessment of model fit. Specifically, BCFI and BTLI can be examined for their sensitivity to model misspecifications, while BRMSEA provides additional information on model complexity. $\hat{\Gamma}_{adj}$ can serve as a supplementary measure. PPP can be used as a supporting index, particularly when the sample size is sufficient (e.g., 200 or more), as it becomes more effective in detecting misfit under these conditions. By considering multiple indices together, researchers can make more informed decisions about the model adequacy and potential specification errors.

9.3. Impact of Model Complexity on Bayesian Fit Indices

The effective number of parameters (*pD*) serves as an indicator of model complexity and the computational effort required for estimation. Due to the incorporation of prior information, *pD* is typically smaller than the actual number of parameters, leading to a difference between Bayesian and maximum likelihood (ML) degrees of freedom. This discrepancy directly affects the behavior of Bayesian fit indices.

BRMSEA tends to be inflated in more complex models due to reduced degrees of freedom, even when model misfit is minimal. BTLI is highly sensitive to degrees of freedom—specifically, the difference in model complexity between the baseline and target models—making it more likely than other indices to classify simpler models as having good fit given the same level of misfit. Compared to BTLI, BCFI is less influenced by Bayesian model complexity, as it quantifies improvements in chi-square statistics relative to the null model. In contrast, $\hat{\Gamma}_{adj}$ is less effective in detecting misfit in simpler models as its values tend to be larger in such cases. Finally, PPP is highly sensitive to sample size, making it less reliable for small samples. The performance of

individual indices will depend on the complexity of the model. Therefore, performance issues should be interpreted in light of model-specific features.

In this study, the effective number of parameters (pD) was derived using the DIC-based method, consistent with the implementation in M*plus*. We acknowledge that DIC-based pD can underestimate model complexity in certain cases, which may, in turn, affect the behavior of approximate fit indices, as discussed by Garnier-Villarreal and Jorgensen (2020). Alternative approaches, such as LOO- or WAIC-based pD calculations available in `blavaan`, may offer improved estimates of model complexity under specific conditions.

9.4. Future Methodological Considerations

This study focused on a specific latent mediation model with defined paths and measurement structures. Although this approach provides valuable information, the findings may not generalize to other model configurations. Exploring different types of mediation models, including more complex structures with multiple mediators and outcomes, would enhance the generalizability of the results.

Another limitation lies in the assumption of normally distributed latent variables and errors. This assumption may not always hold in practical scenarios, potentially affecting the robustness of the findings. Future research should investigate the performance of Bayesian fit indices under violations of normality, including scenarios with skewed or heavy-tailed distributions.

While we examined the impact of diffuse and weakly informative priors, it may also be beneficial to examine a fuller spectrum of prior distributions that might be used in Bayesian analysis. Considering a wider variety of prior distributional forms, including informative priors that incorporate substantive knowledge, would provide a more comprehensive understanding of their influence on Bayesian fit indices in the context of latent mediation modeling.

Additionally, Bayesian SEM fit indices account for model complexity through the effective number of parameters (pD). In the present study, we adopted the DIC-based approach for estimating pD, which is known to underestimate model complexity in certain scenarios. Other approaches, such as LOO-based and WAIC-based methods, offer alternative calculations of pD and may yield different behavior in fit indices (Garnier-Villarreal & Jorgensen, 2020). Future studies could explore the impact of these alternative pD calculations to further refine the evaluation of model fit in Bayesian latent mediation analysis.

We believe that the current investigation helps to shed light on the performance of these common fit measures for the latent mediation model. This knowledge will help applied researchers have a more complete understanding of the impact of misspecifications when implementing this model in practice.

ORCID

Haiyan Liu http://orcid.org/0000-0002-4272-9399
Ihnwhi Heo http://orcid.org/0000-0002-6123-3639
Sarah Depaoli http://orcid.org/0000-0002-1277-0462

References

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis of latent variable models using mplus [Computer software manual]. Retrieved from https://www.statmodel.com/download/BayesAdvantages6.pdf

Asparouhov, T., & Muthén, B. (2021). Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*, 1–14. https://doi.org/10.1080/10705511.2020.1764360

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

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

Cain, M. K., & Zhang, Z. (2019). Fit for a Bayesian: An evaluation of PPP and DIC for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 39–50. https://doi.org/10.1080/10705511.2018.1490648

Cao, C., Lugu, B., & Li, J. (2024). The sensitivity of Bayesian fit indices to structural misspecification in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*, 477–493. https://doi.org/10.1080/10705511.2023.2253497

Cattell, R. B. (1952). *Factor analysis: An introduction and manual for the psychologist and social scientist*. Harper.

Depaoli, S. (2013). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods*, *18*, 186–219. https://doi.org/10.1037/a0031609

Depaoli, S., Jia, F., & Heo, I. (2023). Detecting model misspecifications in Bayesian piecewise growth models. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*, 574–591. https://doi.org/10.1080/10705511.2022.2144865

Depaoli, S., Winter, S. D., & Liu, H. (2024). Under-fitting and over-fitting: The performance of Bayesian model selection and fit indices in sem. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*, 604–625. https://doi.org/10.1080/10705511.2023.2280952

Du, H., & Bentler, P. M. (2022). 40-year old unbiased distribution free estimator reliably improves SEM statistics for nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*, 872–887. https://doi.org/10.1080/10705511.2022.2063870

Edwards, K. D., & Konold, T. R. (2023). Impact of informative priors on model fit indices in Bayesian confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*, 272–283. https://doi.org/10.1080/10705511.2022.2126359

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*, 509–529. https://doi.org/10.1080/00273170701382864

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*, 87–107. https://doi.org/10.1080/10705519709540063

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*, 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Garnier-Villarreal, M., & Jorgensen, T. D. (2020). Adapting fit indices for Bayesian structural equation modeling: Comparison to maximum likelihoods. *Psychological Methods*, *25*, 46–70. https://doi.org/10.1037/met0000224

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*, 515–534. https://doi.org/10.1214/06-BA117A

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*, 408–420. https://doi.org/10.1080/03637750903310360

Heo, I., Jia, F., & Depaoli, S. (2024). Performance of model fit and selection indices for Bayesian piecewise growth modeling with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*, 455–476. https://doi.org/10.1080/10705511.2023.2264514

Hoofs, H., van de Schoot, R., Jansen, N. W. H., & Kant, I. (2018). Evaluating model fit in Bayesian confirmatory factor analysis with large samples: Simulation study introducing the BRMSEA. *Educational and Psychological Measurement*, *78*, 537–568. https://doi.org/10.1177/0013164417709314

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. https://doi.org/10.1080/10705519909540118

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*, 309–334. https://doi.org/10.1037/a0020761

Kaplan, D., & Depaoli, S. (2012). *Bayesian structural equation modeling*. Guilford Press.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and STAN*. Academic Press.

Liu, H., Depaoli, S., & Marvin, L. (2022). Understanding the deviance information criterion for SEM: Cautions in prior specification. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*, 278–294. https://doi.org/10.1080/10705511.2021.1994407

Liu, H., Heo, I., Depaoli, S., & Ivanov, A. (2025). Parameter recovery for misspecified latent mediation models in the Bayesian framework. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–20. https://doi.org/10.1080/10705511.2025.2475490

Liu, H., Jin, I. H., Zhang, Z., & Yuan, Y. (2021). Social network mediation analysis: A latent space approach. *Psychometrika*, *86*, 272–298. https://doi.org/10.1007/s11336-020-09736-z

Liu, H., Zhang, Z., & Grimm, K. J. (2016). Comparison of inverse wishart and separation-strategy priors for Bayesian estimation of covariance parameter matrix in growth curve analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 354–367. https://doi.org/10.1080/10705511.2015.1057285

Liu, X., & Wang, L. (2021). The impact of measurement error and omitting confounders on statistical inference of mediation effects and tools for sensitivity analysis. *Psychological Methods*, *26*, 327–342. https://doi.org/10.1037/met0000345

MacKinnon, D. P. (2012). *Introduction to statistical mediation analysis*. Routledge.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*, 593–614. https://doi.org/10.1146/annurev.psych.58.110405.085542

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*, 173–181. https://doi.org/10.1023/a:1026595011371

Merkle, E. C., & Rosseel, Y. (2015). blavaan: Bayesian structural equation models via parameter expansion. *arXiv preprint arXiv: 1511.05604*.

Miočević, M. (2019). A tutorial in Bayesian mediation analysis with latent variables. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *15*, 137–146.

Miočević, M., Levy, R., & MacKinnon, D. P. (2021). Different roles of prior distributions in the single mediator model with latent variables. *Multivariate Behavioral Research*, *56*, 20–40. PMID: 32003232) https://doi.org/10.1080/00273171.2019.1709405

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. https://doi.org/10.1037/a0026802

Muthén, L., & Muthén, B. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods*, *19*, 459–481. https://doi.org/10.1037/a0036434

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*, 93–115. https://doi.org/10.1037/a0022658

Richiardi, L., Bellocco, R., & Zugna, D. (2013). Mediation analysis in epidemiology: Methods, interpretation and bias. *International Journal of Epidemiology*, *42*, 1511–1519. https://doi.org/10.1093/ije/dyt127

Savalei, V. (2012). The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models. *Educational and Psychological Measurement*, *72*, 910–932. https://doi.org/10.1177/0013164412452564

Sweet, T. M. (2019). Modeling social networks as mediators: A mixed membership stochastic blockmodel for mediation. *Journal of Educational and Behavioral Statistics*, *44*, 210–240. https://doi.org/10.3102/1076998618814255

Tofighi, D., & Kelley, K. (2020). Improved inference in mediation analysis: Introducing the model-based constrained optimization procedure. *Psychological Methods*, *25*, 496–515. https://doi.org/10.1037/met0000259

Tong, X., Zhang, Z., & Yuan, K.-H. (2014). Evaluation of test statistics for robust structural equation modeling with nonnormal missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 553–565. https://doi.org/10.1080/10705511.2014.919820

Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, *18*, 137–150. https://doi.org/10.1037/a0031034

VanderWeele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press.

VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*, *2*, 457–468. https://doi.org/10.4310/SII.2009.v2.n4.a7

Winter, S. D., & Depaoli, S. (2022). Sensitivity of Bayesian model fit indices to the prior specification of latent growth models. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*, 667–686. https://doi.org/10.1080/10705511.2022.2032078

Yang, M., Jiang, G., & Yuan, K.-H. (2018). The performance of ten modified rescaled statistics as the number of variables increases. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 414–438. https://doi.org/10.1080/10705511.2017.1389612

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*, 301–322. https://doi.org/10.1037/a0016972

Zhang, Q., & Wang, Q. (2024). Handling measurement error and omitted confounders considering informativeness of the confounding effect under mediation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*, 1083–1103. https://doi.org/10.1080/10705511.2023.2270167

Zhao, S., Zhang, Z., & Zhang, H. (2024). Bayesian inference of dynamic mediation models for longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*, 14–26. https://doi.org/10.1080/10705511.2023.2230519