CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 904-920
p-ISSN: 2086-0382; e-ISSN: 2477-3344

A Simulation Study On The Robustness Of Bayesian Structural Equation Modeling Under Small Samples, Heavy Tails, and Collinearity

Agustina Susi Susanti Parung*, Ani Budi Astuti, and Rahma Fitriani

Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Malang, Indonesia

Abstract

Bayesian Structural Equation Modeling (BSEM) is increasingly used in district-level health and social research, where inference is often challenged by small sample size, heavy-tailed errors, and collinearity among determinants. This study evaluates the operating characteristics of BSEM in a fixed confirmatory full-mediation SEM with four latent variables and three reflective indicators per construct, estimated using `std.lv=TRUE` with indicator residual variances fixed to their data-generating values. The core design is a balanced $2 \times 2 \times 2$ Monte Carlo experiment varying sample size ($n \in \{22, 75\}$), exogenous correlation ($\rho \in \{0.30, 0.80\}$), and error family (Normal versus variance-matched Student-t with $\nu = 5$). Heavy tails are introduced only in the data-generating structural disturbances and measurement errors, while the fitted SEM retains the default Gaussian residual specification; thus, the Student-t scenarios serve as robustness tests under heavy-tailed error misspecification. For each scenario ($R = 50$), models are estimated in `blavaan` via Stan/NUTS and evaluated using posterior predictive p-values (PPP), sampler diagnostics, and Monte Carlo summaries of bias, RMSE, and 95% credible-interval coverage for structural paths, indirect effects, and ρ. Larger sample size improves posterior exploration and parameter recovery, whereas high exogenous correlation mainly degrades recovery of the parallel mediator regressions and their associated indirect effects. Student-t conditions primarily reduce PPP relative to the Normal-error cases. A feasibility study at $n = 10$ with two indicators per construct is reported descriptively.

Keywords: Bayesian structural equation modeling; collinearity; heavy tails; simulation study; small sample size.

1. Introduction

Stunting remains a major public health concern because of its long-term consequences for child development and human capital. In Indonesia, district-level evidence is often needed to support planning for nutrition and basic services. Recent work has applied predictive and exploratory approaches to stunting prevalence in East Nusa Tenggara (NTT) [1], while empirical and review studies consistently identify sociodemographic conditions, nutrition, and sanitation as important correlates of child growth outcomes [2, 3]. From a statistical perspective, district/city-level analyses are typically based on a limited number of observational units, which motivates careful evaluation of finite-sample operating characteristics in multivariate latent-variable models.

*Corresponding author. E-mail: agustinaparung@student.ub.ac.id

Submitted: February 24, 2026 Reviewed: March 21, 2026 Accepted: April 06, 2026
DOI: https://doi.org/10.18860/cauchy.v11i1.41328

District-level aggregated data commonly present three interacting challenges for structural equation modeling (SEM). First, the number of observational units is small, reducing the information content of the likelihood and increasing sensitivity to modeling assumptions. Second, residual behavior may depart from Gaussianity because of outlying districts, heterogeneous measurement quality, or unmodeled heterogeneity, producing heavy-tailed errors. Third, substantive determinants are often strongly correlated, creating collinearity among exogenous constructs. These conditions can weaken fit assessment and destabilize estimation in SEM [4–6]. In particular, strong exogenous correlation can reduce separability between parallel structural effects, whereas heavy-tailed errors can impair fit calibration and likelihood-based inference [7, 8]. Their joint presence therefore constitutes a practically relevant stress test for SEM-based inference in district-level settings.

Structural equation modeling (SEM) combines a measurement model, which links observed indicators to latent constructs, with a structural model, which relates latent constructs through directional paths [5, 9, 10]. Covariance-based SEM remains widely used because of established estimation routines and conventional fit indices [6, 11]. However, its performance is generally most reliable under approximate multivariate normality and moderate-to-large samples [4–6]. In district-level applications, where small samples, heavy-tailed errors, and strong exogenous dependence may occur simultaneously, these conditions can be difficult to satisfy [7, 8, 12]. This motivates Bayesian alternatives that may stabilize estimation through prior regularization while providing full posterior uncertainty quantification.

Bayesian structural equation modeling (BSEM) provides such a framework through posterior inference based on Markov Chain Monte Carlo (MCMC) [13, 14]. In small samples, priors may regularize weakly identified parameter directions, while posterior simulation supports direct uncertainty assessment and posterior predictive model checking [13, 15, 16]. Yet an important methodological gap remains. Existing simulation studies often examine small samples, non-normal errors, or strong exogenous correlations in isolation [17, 18], and many focus on simplified structural settings rather than confirmatory mediation structures that are common in substantive applications [19]. This matters because mediated effects are products of coefficients and may be especially sensitive to finite-sample variability, while high exogenous correlation can reduce separability of parallel structural effects and alter the posterior geometry explored by Hamiltonian Monte Carlo algorithms such as the No-U-Turn Sampler (NUTS).

To address this gap, we conduct a structured Monte Carlo simulation study for a specific confirmatory full-mediation BSEM setting motivated by district-level constraints. The model contains four latent variables: two exogenous constructs (X_1, X_2), one mediator (M), and one endogenous construct (Y), with three reflective indicators per construct and covariance allowed between X_1 and X_2. The core design is a balanced $2 \times 2 \times 2$ factorial experiment varying sample size ($n \in \{22, 75\}$), exogenous correlation ($\rho \in \{0.30, 0.80\}$), and error family (Normal versus variance-matched Student-t with $\nu = 5$), while holding the structural form and true coefficients fixed. Heavy tails are introduced only in the data-generating disturbances and measurement errors, whereas the fitted SEM retains the default Gaussian residual specification; accordingly, the Student-t conditions are interpreted as robustness tests under heavy-tailed error misspecification. For each core scenario, we generate $R = 50$ datasets and estimate the model in `blavaan` via Stan/NUTS. Performance is evaluated using posterior predictive p-values (PPP), standard Stan/NUTS diagnostics, and Monte Carlo summaries of bias, RMSE, and 95% credible-interval coverage for structural paths, indirect effects, and the exogenous correlation. A feasibility study at $n = 10$ under $\rho = 0.80$ with two indicators per construct is reported descriptively.

Contributions. This study makes three contributions. First, it provides a factorial robustness map for a confirmatory full-mediation BSEM under the joint stressors of small sample size, heavy-tailed errors, and exogenous collinearity. Second, it distinguishes sampler adequacy from inferential adequacy by evaluating Stan/NUTS diagnostics together with parameter-recovery

measures. Third, it treats the Student-t conditions explicitly as misspecification-based robustness tests, thereby interpreting PPP differences as fit-calibration behavior under heavy-tailed errors rather than as evidence of convergence failure.

2. Methods

This section describes the simulation framework used in the study, including the fixed SEM specification, scenario design, data-generating mechanism, prior structure, estimation settings, and performance measures.

2.1. Overview

We conduct a Monte Carlo simulation study to evaluate the robustness of Bayesian structural equation modeling (BSEM) under district-level constraints. Robustness is assessed along three complementary dimensions: (i) *global fit calibration* using posterior predictive p-values (PPP), (ii) *posterior exploration and sampling stability* using standard Stan/NUTS diagnostics (e.g., $\widehat{R}$, effective sample size, divergent transitions, and tree-depth behavior), and (iii) *inferential accuracy* using Monte Carlo parameter-recovery summaries (bias, RMSE, and 95% credible-interval coverage). This joint evaluation is important because small samples, heavy-tailed errors, and strong exogenous dependence can simultaneously affect fit assessment, posterior uncertainty, and posterior geometry.

The workflow proceeds in six stages: (i) fixed model specification, (ii) factorial scenario design, (iii) data generation, (iv) priors and identification, (v) Bayesian estimation via Stan/NUTS, and (vi) diagnostic and parameter-recovery evaluation. The core design comprises eight scenarios (S1–S8) forming a balanced $2\times2\times2$ experiment that varies sample size ($n \in \{22, 75\}$), error family (Normal vs. variance-matched Student-t with $\nu = 5$), and exogenous correlation ($\rho \in \{0.30, 0.80\}$), while holding the SEM structure and true parameters fixed. Importantly, the Student-t conditions are imposed only in the *data-generating* error terms, whereas the fitted SEM retains the default Gaussian residual specification; thus, the Student-t scenarios serve as robustness tests under heavy-tailed error misspecification. The fixed true parameter values and design controls used in the data-generating process (DGP) are summarized in Table 1.

For each core scenario, we generate $R = 50$ independent datasets and fit the BSEM using `blavaan::bsem` with the Stan backend (NUTS). For reporting, we distinguish global fit (PPP), posterior exploration and sampling diagnostics ($\widehat{R}$, ESS, divergent transitions, and tree-depth behavior), and parameter recovery for the key structural coefficients (a_1, a_2, b), the indirect effects (IE_1, IE_2), and the exogenous correlation ρ, using posterior medians and 95% credible intervals within each replication.

Replication budget and Monte Carlo uncertainty. We use $R = 50$ replications per core scenario as a computationally feasible budget under Stan/NUTS estimation. This implies non-negligible Monte Carlo uncertainty in reported performance measures; for example, if $\widehat{p}$ denotes a coverage estimate, its Monte Carlo standard error is approximately $\sqrt{\widehat{p}(1-\widehat{p})/R}$ and is at most 0.071 when $R = 50$. Accordingly, small differences in RMSE or coverage across scenarios are interpreted conservatively, and Monte Carlo uncertainty summaries are reported for key measures.

Finally, we include a simplified feasibility setting at $n = 10$ (F1–F2) under worst-case collinearity ($\rho = 0.80$) with two indicators per construct to assess computational viability under extreme constraints. Because this feasibility design changes the measurement structure, its results are reported descriptively and are not pooled with the core factorial comparisons.

2.2. Model Specification

The structural form and measurement design are held fixed across scenarios so that cross-scenario differences can be interpreted primarily in relation to the manipulated factors (sample size, error family, and exogenous correlation), while recognizing that the variances of the endogenous latent variables are induced by the structural system and therefore vary with ρ. The specification is chosen to (i) represent a full-mediation mechanism common in determinant studies, (ii) retain a standard confirmatory measurement structure with reflective indicators and no cross-loadings, and (iii) remain estimable under small samples with correlated exogenous constructs.

Latent structural model (full mediation). Let $i = 1, \ldots, n$ index observational units. The latent structural relations are

$$M_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \zeta_{Mi}, \tag{1}$$
$$Y_i = \gamma M_i + \zeta_{Yi}, \tag{2}$$

The full-mediation structural relations are given in Eqs. (1)–(2), where X_1 and X_2 are exogenous latent constructs, M is a mediator, Y is an endogenous latent construct, and (ζ_{Mi}, ζ_{Yi}) are structural disturbances. Direct effects $X_1 \to Y$ and $X_2 \to Y$ are fixed to zero, so the total effects of (X_1, X_2) on Y operate entirely through M. The two mediated effects are

$$\mathrm{IE}_1 = \beta_1\gamma, \qquad \mathrm{IE}_2 = \beta_2\gamma,$$

which are treated as primary inferential targets because they are products of coefficients and are therefore especially sensitive to finite-sample uncertainty. For convenience in the recovery summaries, we write $(a_1, a_2, b) \equiv (\beta_1, \beta_2, \gamma)$.

Exogenous dependence is represented by

$$\mathrm{Corr}(X_1, X_2) = \rho,$$

which is controlled in the data-generating process to manipulate collinearity. In the fitted model, the corresponding covariance parameter $X_1 \leftrightarrow X_2$ is freely estimated, so recovery of ρ is also evaluated.

In the DGP, the exogenous latent variables are standardized to unit variance, whereas the variances of the endogenous latent variables are implied by the structural coefficients, the exogenous correlation, and the disturbance variances. Specifically,

$$\mathrm{Var}(M) = \beta_1^2\mathrm{Var}(X_1) + \beta_2^2\mathrm{Var}(X_2) + 2\beta_1\beta_2\mathrm{Cov}(X_1, X_2) + \psi_M, \tag{3}$$
$$\mathrm{Var}(Y) = \gamma^2\mathrm{Var}(M) + \psi_Y, \tag{4}$$

The scenario dependence of the endogenous latent variances follows directly from Eqs. (3)–(4), so $\mathrm{Var}(M)$ and $\mathrm{Var}(Y)$ are scenario-dependent rather than fixed to one in the DGP.

Measurement model (reflective, confirmatory). Each latent construct $C \in \{X_1, X_2, M, Y\}$ is measured by reflective indicators. For unit i and indicator j, the measurement model is

$$c_{C,ij} = \lambda_{C,j} C_i + \varepsilon_{C,ij}, \qquad j = 1, \ldots, p_C,$$

where $\lambda_{C,j}$ denotes the factor loading and $\varepsilon_{C,ij}$ the measurement error. The confirmatory structure imposes (i) no cross-loadings and (ii) no correlated measurement residuals, yielding a sparse loading matrix and a diagonal measurement-error covariance matrix. This structure is held fixed across the core scenarios. In the core design, each construct has three indicators; the feasibility setting uses two indicators per construct.

Identification and parameterization. In the fitted model, estimation uses the `std.lv=TRUE` parameterization, so latent scaling is handled by fixing latent variances to one within the estimation parameterization. Accordingly, the fitted SEM is described using latent-variance standardization rather than marker-variable identification. The loading pattern $(1, 0.80, 0.70)$ refers to the data-generating measurement design and is not treated as an additional identification constraint in estimation. Under `std.lv=TRUE`, the estimated exogenous covariance is directly interpretable as the exogenous correlation,

$$\mathrm{Cov}(X_1, X_2) = \rho,$$

which simplifies interpretation across scenarios. Latent means are set to zero in the DGP, and the fitted model adopts the standard SEM mean constraints because the study focuses on covariance and structural relations.

2.2.1. True parameter settings and design controls

All core scenarios share the same SEM structure and the same true structural coefficients; only (n, ρ) and the error family vary. Table 1 summarizes the fixed generating values and design controls used in the data-generating process (DGP), together with the scenario-varying exogenous correlation.

Table 1: True generating parameters and fixed design controls.

Component	Value
Structural paths (full mediation)	$a_1 = \beta_1 = 0.40,\ a_2 = \beta_2 = 0.30,\ b = \gamma = -0.50$
Indirect effects	$IE_1 = a_1 b = -0.20, \quad IE_2 = a_2 b = -0.15$
Exogenous correlation	$\rho = \mathrm{Corr}(X_1, X_2) \in \{0.30, 0.80\}$ (by scenario)
Structural disturbance variances	$\mathrm{Var}(\zeta_M) = \mathrm{Var}(\zeta_Y) = 1$
Measurement error variance	$\sigma_e^2 = 0.04$
Core loadings (per construct)	$(\lambda_1, \lambda_2, \lambda_3) = (1, 0.80, 0.70)$
Feasibility loadings (per construct)	$(\lambda_1, \lambda_2) = (1, 0.80)$
Heavy-tail parameter	Student-t degrees of freedom: $\nu = 5$ (variance-matched)

In the DGP, the exogenous latent variables are standardized to unit variance, whereas the variances of the endogenous latent variables are implied by the structural equations and therefore vary with ρ. Thus, the entries in Table 1 should be interpreted as fixed coefficients and variance controls in the generating mechanism, not as implying that all latent variances are constant across scenarios.

Fitted-model specification. In all core scenarios, the data-generating measurement-error variance is set to $\sigma_e^2 = 0.04$. The same value is imposed in the fitted SEM by fixing each observed indicator residual variance to 0.04, thereby controlling one source of measurement variability across scenarios. Estimation is conducted with `blavaan::bsem(target="stan")` under the default Gaussian residual specification; accordingly, the Student-t scenarios are treated as robustness tests under heavy-tailed error misspecification rather than as t-likelihood fits. Because the variances of the endogenous latent variables are induced by the structural system, fixing indicator residual variances does not imply that the effective indicator signal scale for M and Y is strictly invariant across scenarios.

Measurement design notes. In the core design, each construct has three reflective indicators with loadings $(1, 0.80, 0.70)$, providing moderate-to-high measurement reliability while keeping the measurement structure identical across scenarios. In the feasibility design, two indicators per construct with loadings $(1, 0.80)$ are used. Because this changes the measurement structure, feasibility results are reported descriptively and excluded from the factorial comparisons.

Implications for robustness under collinearity and small n. The parallel structural paths (β_1, β_2) are estimated conditional on correlated exogenous constructs. When ρ is high, the posterior for (β_1, β_2) may exhibit stronger dependence and reduced separability, particularly in small samples. This mechanism is expected to weaken recovery for β_1, β_2, and, consequently, for the mediated effects $(\beta_1\gamma, \beta_2\gamma)$. These effects are evaluated using Monte Carlo bias, RMSE, and coverage, and are interpreted jointly with sampler diagnostics such as $\widehat{R}$, ESS, divergences, and tree-depth summaries.

2.3. Simulation scenarios

The simulation scenarios are designed to evaluate BSEM robustness under three data conditions that commonly co-occur in district-level applications: (i) limited sample size, (ii) heavy-tailed errors, and (iii) collinearity among exogenous constructs. The SEM specification in Section 2.2 is held fixed across scenarios so that cross-scenario differences can be interpreted primarily in relation to the manipulated data conditions, while recognizing that some latent-scale features of the endogenous constructs are induced by the structural system and therefore vary with ρ.

Core factorial design (S1–S8). The core experiment follows a balanced $2 \times 2 \times 2$ factorial design with the following factors:

- **Sample size:** $n \in \{22, 75\}$. The smaller size reflects the limited number of district/city units typically available within a province-level setting, whereas $n = 75$ serves as a more informative benchmark for assessing the extent to which additional information mitigates small-sample and collinearity-related instability.
- **Error family (DGP stress test):** Normal (reference) versus Student-t with $\nu = 5$. The $t_{\nu=5}$ family provides a heavy-tailed stress test with finite variance; in the DGP, t errors are variance-matched to the Normal targets so that the comparison emphasizes tail behavior rather than marginal error scale.
- **Exogenous correlation:** $\rho \in \{0.30, 0.80\}$. The low-correlation condition represents mild dependence, whereas $\rho = 0.80$ represents strong collinearity that can reduce separability of the parallel structural effects (e.g., $X_1 \to M$ and $X_2 \to M$), especially under small n.

Importantly, the Student-t condition is imposed only in the *data-generating* error terms, whereas the fitted SEM retains the default Gaussian residual specification; accordingly, the Student-t scenarios are interpreted as robustness tests under heavy-tailed error misspecification. The factorial design therefore supports structured comparison across combinations of sample size, tail behavior, and exogenous correlation.

Feasibility design (F1–F2). In addition to the core factorial design, we include two feasibility scenarios at $n = 10$ under $\rho = 0.80$ to assess computational viability under extreme constraints. The feasibility setting uses a simplified measurement structure with two indicators per construct and is therefore reported descriptively only; it is not pooled with the core scenarios because it is not directly comparable in terms of measurement information.

Replication budget and Monte Carlo uncertainty. For each core scenario, we use $R = 50$ independent replications as a computationally feasible budget under Stan/NUTS estimation; the feasibility scenarios use $R = 30$. Because all reported performance measures are Monte Carlo estimates, small differences across scenarios are interpreted conservatively, and uncertainty summaries are reported for key measures (e.g., binomial standard errors for coverage and bootstrap uncertainty for RMSE; see Section 2.7).

Table 2 summarizes the eight core scenarios (S1–S8) and the feasibility cases (F1–F2).

Table 2: Core factorial scenarios (S1–S8) and feasibility cases (F1–F2).

Scenario	n	Error family	ν	ρ
S1	22	Normal	–	0.30
S2	22	Normal	–	0.80
S3	22	Student-t	5	0.30
S4	22	Student-t	5	0.80
S5	75	Normal	–	0.30
S6	75	Normal	–	0.80
S7	75	Student-t	5	0.30
S8	75	Student-t	5	0.80
F1	10	Normal	–	0.80
F2	10	Student-t	5	0.80

2.4. Data-generating mechanism

For each replication $r = 1, \ldots, R$ within a scenario, data are generated in the following order: (i) exogenous latent variables, (ii) structural disturbances, (iii) endogenous latent variables, (iv) measurement errors, and (v) reflective indicators. This ordering separates the exogenous dependence structure, which is controlled by ρ, from the distributional stress test, which is introduced through the error terms. The fixed coefficients and design controls referenced below are summarized in Table 1. Unless stated otherwise, all random draws are independent across units i, and measurement errors are independent across indicators, constructs, and units.

Exogenous constructs. For each unit $i = 1, \ldots, n$, the exogenous latent vector $\mathbf{X}_i = (X_{1i}, X_{2i})^\top$ is generated from a bivariate standard Normal distribution,

$$\mathbf{X}_i \sim \mathcal{N}_2(\mathbf{0}, \Sigma_\rho), \qquad \Sigma_\rho = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

so that $\mathrm{Var}(X_{1i}) = \mathrm{Var}(X_{2i}) = 1$ and $\mathrm{Corr}(X_{1i}, X_{2i}) = \rho$ is fixed by the scenario. Heavy-tailed behavior is not introduced in the exogenous constructs; it is imposed only on the error terms to target robustness to non-Gaussian residual behavior.

Structural disturbances and endogenous constructs. Structural disturbances (ζ_{Mi}, ζ_{Yi}) are generated independently across units, independently of $\mathbf{X}_i$, and independently of each other within unit. Under Normal scenarios,

$$\zeta_{Mi} \sim \mathcal{N}(0, 1), \qquad \zeta_{Yi} \sim \mathcal{N}(0, 1).$$

Under Student-t scenarios, heavy-tailed disturbances are generated from centered Student-t distributions with $\nu = 5$ degrees of freedom and then variance-matched to the same target variances (details below). Endogenous latent variables are constructed as

$$M_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \zeta_{Mi},$$
$$Y_i = \gamma M_i + \zeta_{Yi}.$$

By design, the structural disturbance variances are fixed across scenarios at

$$\mathrm{Var}(\zeta_M) = \mathrm{Var}(\zeta_Y) = 1,$$

so that differences between Normal and Student-t conditions reflect tail behavior rather than changes in residual scale. However, the variances of the endogenous latent variables are induced by the structural equations and therefore vary with ρ; they are not fixed to one in the DGP.

Measurement errors and reflective indicators. For each construct $C \in \{X_1, X_2, M, Y\}$ and indicator j, measurement errors are generated independently across indicators and units with target variance $\sigma_e^2 = 0.04$:

$$\varepsilon_{C,ij} \sim \begin{cases} \mathcal{N}(0, \sigma_e^2), & \text{Normal scenarios,} \\ \text{variance-matched Student-}t \text{ with } \nu = 5 \text{ and target variance } \sigma_e^2, & \text{Student-}t \text{ scenarios.} \end{cases}$$

Indicators are then generated from the reflective measurement model

$$c_{C,ij} = \lambda_{C,j}\, C_i + \varepsilon_{C,ij}.$$

In the core design, the DGP loading pattern is $(\lambda_{C,1}, \lambda_{C,2}, \lambda_{C,3}) = (1, 0.80, 0.70)$ for each construct; in the feasibility design, $(1, 0.80)$ is used. Fixing σ_e^2 and the loading pattern across scenarios controls one source of measurement variability, but it does not imply that the effective signal scale of indicators for the endogenous constructs is strictly invariant across scenarios, because $\mathrm{Var}(M)$ and $\mathrm{Var}(Y)$ are scenario-dependent.

Variance-matched Student-t errors. To ensure fair comparisons between Normal and Student-t conditions, all t-distributed error terms are rescaled to match the corresponding target variances. For $\nu > 2$, if $u \sim t_\nu$ then

$$\mathrm{Var}(u) = \frac{\nu}{\nu - 2}.$$

Define the standardized draw

$$u^* = \frac{u}{\sqrt{\nu/(\nu-2)}} = u\sqrt{\frac{\nu-2}{\nu}},$$

so that $\mathrm{Var}(u^*) = 1$. A target-variance draw with variance σ^2 is then obtained via

$$u_\sigma = \sigma\, u^*,$$

where $\sigma = 1$ for structural disturbances and $\sigma = \sqrt{0.04}$ for measurement errors. This variance matching ensures that differences between Normal and Student-t scenarios are attributable to tail heaviness rather than to changes in marginal error variance.

Interpretation as a robustness stress-test. In the Student-t scenarios, heavy tails are introduced only in the data-generating error terms while maintaining the same target variances as in the Normal scenarios. The fitted BSEM retains the default Gaussian residual specification as implemented by `blavaan`/Stan; therefore, the Student-t conditions evaluate robustness of posterior inference under heavy-tailed error misspecification rather than fitting a t-likelihood model. Consequently, PPP behavior in the Student-t scenarios reflects how well the Gaussian-error SEM reproduces the chosen discrepancy under heavy-tailed data generation.

2.5. Priors and identification

Default priors (Stan target). All analyses use the default `blavaan` prior specification for the Stan backend via `dp = dpriors()` [20, 21]. We do not override these defaults, so the simulation does not introduce additional variation through scenario-specific prior retuning. The hyperparameters reported below are those returned by `dpriors()` in our implementation under `target="stan"`.

The default prior classes returned by `dpriors()` in this implementation are

$$\lambda \sim \mathcal{N}(0, 10), \qquad \beta \sim \mathcal{N}(0, 10), \qquad r \sim \mathrm{Beta}(1, 1),$$

where r denotes correlation-type parameters in the fitted model. Variance-type parameters use Gamma priors with the `[sd]` modifier, so the prior is placed on the standard-deviation scale:

$$\theta \sim \text{Gamma}(1, 0.5)\,[sd], \qquad \psi \sim \text{Gamma}(1, 0.5)\,[sd].$$

Here, λ denotes factor loadings, β denotes structural coefficients, r denotes fitted correlation parameters associated with covariance relations, and θ and ψ denote manifest-level and latent-level variance classes in the `blavaan` parameter taxonomy. Because the study focuses on covariance and structural relations, intercept- and mean-related priors are not substantively central to the operating characteristics examined here.

For covariance relations, `blavaan` places the default prior on the associated correlation rather than directly on the covariance. The Beta prior is defined on $(0, 1)$ and internally transformed to support $(-1, 1)$, providing a numerically stable default for fitted correlation parameters.

Active vs. inactive priors under design controls. Because the observed indicator residual variances are fixed at $\sigma_e^2 = 0.04$ in the fitted model, those measurement-error variance terms are not estimated freely. Accordingly, the default `dpriors()` specification is active only for parameter classes that remain free under the fitted parameterization, including factor loadings, structural coefficients, the exogenous correlation, and any remaining variance-type parameters.

Identification and parameterization. In the fitted model, latent scaling is handled using `std.lv=TRUE`. Accordingly, the fitted SEM is described using latent-variance standardization rather than marker-variable identification. The loading pattern $(1, 0.80, 0.70)$ refers to the data-generating measurement design and is not imposed as an additional identification constraint in estimation. Under `std.lv=TRUE`, the estimated covariance parameter $X_1 \leftrightarrow X_2$ is directly interpretable as the exogenous correlation,

$$\text{Cov}(X_1, X_2) = \rho,$$

which aligns the fitted-model parameter with the scenario-controlled exogenous dependence used in the DGP. Latent means are set to zero in the data-generating process, and the fitted model adopts the standard SEM mean constraints because the study focuses on covariance and structural relations.

2.6. Bayesian estimation (Stan/NUTS)

All models are estimated in R using `blavaan::bsem` with `target="stan"`, which fits the SEM in Stan using Hamiltonian Monte Carlo with the No-U-Turn Sampler (NUTS) [20, 21].

Core scenarios (S1–S8). For each core scenario, we run 2 independent chains with 500 warm-up iterations and 1000 post-warm-up draws per chain, using `start="simple"` to promote stable initialization under small-sample conditions. Default control settings are `adapt_delta=0.995` and `max_treedepth=12`. For the most challenging core condition (S8: $\rho = 0.80$ under Student-t error generation), tuning is tightened to `adapt_delta=0.999` and `max_treedepth=15` as a precaution against divergences and tree-depth saturation.

Feasibility scenarios (F1–F2). For the feasibility settings ($n = 10$ with two indicators per construct under $\rho = 0.80$), we use 2 chains with 800 warm-up iterations and 1200 post-warm-up draws per chain, together with `adapt_delta=0.999` and `max_treedepth=15`. These more conservative settings reflect the more severe information constraints in the feasibility design.

Computation and recorded diagnostics. For each fitted model, we monitor posterior predictive p-values (PPP), $\widehat{R}$, effective sample size (ESS), divergent transitions, and tree-depth summaries. Although two chains are used as a computationally feasible default, $\widehat{R}$ and ESS are checked in every scenario so that posterior summaries are interpreted only when posterior exploration is adequate. Because S8 is estimated under tighter tuning than the other core scenarios, computational diagnostics are interpreted as evidence of feasible posterior exploration under the scenario-specific tuning regime used here rather than as a strict like-for-like comparison of computational behavior across all scenarios.

2.7. Diagnostics and performance measures

Global fit via posterior predictive checking. Global model fit is evaluated using posterior predictive p-values (PPP) computed from the default chi-square discrepancy used for covariance-structure models. Let y denote the observed data and let y^{rep} denote a replicated dataset generated from the posterior predictive distribution:

$$y^{\text{rep}} \sim p(y^{\text{rep}} \mid \theta), \qquad \theta \sim p(\theta \mid y)$$

Given a discrepancy statistic $T(\cdot)$, the posterior predictive p-value is defined as [15]

$$\text{PPP} = \Pr\{T(y^{\text{rep}}) \geq T(y) \mid y\}\,, \tag{5}$$

Global fit is evaluated using the posterior predictive p-value defined in Eq. (5). In this study, $T(\cdot)$ is the chi-square discrepancy used for SEM covariance fitting, as implemented by `blavaan`/`lavaan` for the Stan target [20, 21]. PPP values near 0.5 indicate better calibration, whereas values near 0 or 1 suggest potential global misfit. Because PPP is a Bayesian calibration measure rather than a frequentist p-value, it is interpreted jointly with sampler diagnostics and parameter-recovery summaries.

Convergence, effective information, and sampler pathologies. Convergence and sampling quality are assessed using $\widehat{R}$ and effective sample size (ESS), together with Stan/NUTS diagnostics including divergent transitions and tree-depth behavior. For each replication, we report

$$\widehat{R}_{\max} = \max_k \widehat{R}_k,$$
$$\text{ESS}_{\min} = \min_k \text{ESS}_k,$$

where k indexes monitored parameters across the fitted model. Divergent transitions indicate potential numerical integration problems in HMC, while tree-depth summaries indicate whether NUTS is approaching its exploration limit. These diagnostics are extracted directly from Stan output and summarized at the scenario level.

Parameter recovery. Parameter recovery is evaluated for the structural effects (a_1, a_2, b), indirect effects $\text{IE}_1 = a_1 b$ and $\text{IE}_2 = a_2 b$, and the exogenous correlation ρ. For replication $r = 1, \ldots, R$, point estimates are posterior medians and 95% credible intervals are defined by the 2.5% and 97.5% posterior quantiles. Let $\hat{\theta}_r$ denote the replication-r posterior median for parameter θ, and let CI_r denote the associated 95% credible interval. The recovery summaries are computed using the definitions in Eqs. (6)–(8):

$$\text{Bias}(\theta) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_r - \theta), \tag{6}$$

$$\text{RMSE}(\theta) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_r - \theta)^2}, \tag{7}$$

$$\mathrm{Cov}(\theta) = \frac{1}{R}\sum_{r=1}^{R} \mathbb{I}\{\theta \in \mathrm{CI}_r\}, \tag{8}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function.

Scenario-level aggregation. Scenario-level summaries are obtained by aggregating replication-level outputs within each scenario. We report medians for PPP, $\widehat{R}_{\max}$, and $\mathrm{ESS}_{\min}$ to reduce sensitivity to occasional extreme replications, and we report total divergent transitions together with related NUTS summaries to characterize sampler behavior and computational burden. Because the most difficult scenario is estimated under tighter tuning than the remaining core conditions, these computational summaries are interpreted cautiously and not as a strict like-for-like comparison across all scenarios.

Monte Carlo uncertainty. Because all performance measures are estimated from a finite number of replications ($R = 50$ for core scenarios and $R = 30$ for feasibility), reported bias, RMSE, and coverage are subject to Monte Carlo error. For coverage, $\widehat{\mathrm{Cov}}(\theta)$ is a sample proportion with Monte Carlo standard error

$$\mathrm{SE}\{\widehat{\mathrm{Cov}}(\theta)\} \approx \sqrt{\frac{\widehat{\mathrm{Cov}}(\theta)\{1 - \widehat{\mathrm{Cov}}(\theta)\}}{R}}.$$

For bias, letting $e_r = \hat{\theta}_r - \theta$, the Monte Carlo standard error is

$$\mathrm{SE}\{\widehat{\mathrm{Bias}}(\theta)\} \approx \frac{\mathrm{sd}(e_r)}{\sqrt{R}}.$$

For RMSE, uncertainty is summarized using a nonparametric bootstrap over replications by resampling $\{e_r\}_{r=1}^{R}$ with replacement, recomputing RMSE, and using the 2.5% and 97.5% quantiles to form a Monte Carlo interval. Accordingly, small cross-scenario differences are interpreted conservatively when they are comparable to Monte Carlo variability.

2.8. Reproducibility

All simulations are run under fixed seed control to support reproducibility of both the data-generating process and Stan/NUTS estimation. For each scenario, the analysis pipeline stores replication-level outputs together with aggregated scenario summaries of recovery measures and diagnostic statistics, including PPP, $\widehat{R}$, ESS, divergence counts, tree-depth summaries, bias, RMSE, and coverage. The implementation is based on R using `blavaan` with the Stan backend, and software-version details used for the reported results are documented in the project files. Simulation scripts, scenario definitions, and software-version information are available from the corresponding author upon reasonable request.

3. Results and Discussion

We report (i) global fit and sampler behavior using posterior predictive checks and Stan/NUTS diagnostics, and (ii) parameter-recovery performance using Monte Carlo bias, RMSE, and 95% credible-interval coverage. Results are organized to compare performance across the factorial combinations of sample size ($n \in \{22, 75\}$), exogenous collinearity ($\rho \in \{0.30, 0.80\}$), and error tail-heaviness (Normal vs. variance-matched Student-t with $\nu = 5$) under the fixed full-mediation SEM studied here.

3.1. Global fit and sampler diagnostics

Table 3 summarizes posterior predictive p-values (PPP) and Stan/NUTS diagnostics for the core scenarios, using scenario medians over replications and total divergences across replications. The columns PPP_{low} and PPP_{high} report the proportions of replications with PPP < 0.05 and PPP > 0.95, respectively, and provide a simple calibration check for extreme PPP behavior.

Table 3: Global fit and Stan/NUTS diagnostics across core scenarios. Values are scenario medians over replications, except divergences (summed). PPP_{low} and PPP_{high} are the proportions with PPP < 0.05 and PPP > 0.95.

Scenario	n	Error	ρ	PPP_{med}	PPP_{low}	PPP_{high}	$\widehat{R}_{\text{max,med}}$	$\text{ESS}_{\text{min,med}}$	$\text{treedepth}_{\text{med}}$	Div_{sum}
S1	22	Normal	0.30	0.428	0.06	0.00	1.012	517.5	8	0
S2	22	Normal	0.80	0.500	0.10	0.00	1.014	429.7	8	0
S3	22	$t(5)$	0.30	0.275	0.06	0.00	1.011	511.5	8	0
S4	22	$t(5)$	0.80	0.275	0.06	0.00	1.012	417.9	8	0
S5	75	Normal	0.30	0.455	0.02	0.00	1.007	850.6	8	0
S6	75	Normal	0.80	0.460	0.02	0.00	1.008	617.2	8	0
S7	75	$t(5)$	0.30	0.275	0.14	0.00	1.007	831.6	8	0
S8	75	$t(5)$	0.80	0.268	0.12	0.02	1.008	640.9	8	0

Across all core scenarios, no divergent transitions were observed ($\text{Div}_{\text{sum}} = 0$), indicating feasible posterior exploration under the tuning settings used in this study. Median tree-depth usage is stable at 8 across the core scenarios, while $\widehat{R}_{\text{max,med}}$ remains close to 1 (approximately 1.007–1.014) and $\text{ESS}_{\text{min,med}}$ ranges from about 418 to 851. Taken together, these diagnostics are consistent with generally stable posterior exploration, although effective information is lower in the smaller-sample and higher-correlation settings. PPP is interpreted as a Bayesian calibration measure rather than a frequentist p-value. Relative to the Normal-error conditions, PPP_{med} is lower in the Student-t scenarios, consistent with the intended heavy-tailed misspecification stress test under a Gaussian fitted SEM. The proportions PPP_{low} and PPP_{high} provide a simple descriptive check for extreme PPP behavior, although small differences in these proportions should be interpreted cautiously because they are based on $R = 50$ replications. Because the most difficult scenario (S8) is estimated under tighter tuning than the remaining core scenarios, these computational diagnostics should be interpreted as evidence of feasible exploration under the scenario-specific tuning regime used here rather than as a strict like-for-like comparison across all scenarios.

3.2. Parameter recovery for structural and mediated effects

Table 4 reports recovery summaries for the structural paths, indirect effects, and the exogenous correlation. RMSE is computed across replications using posterior medians, and coverage denotes the proportion of 95% credible intervals containing the true value. Across the core design, the clearest recovery patterns are improved accuracy with larger n and substantial deterioration under high exogenous collinearity, especially for the parallel paths into the mediator and their associated indirect effects.

3.3. Cross-scenario interpretation

Sample-size contrast. Increasing n from 22 to 75 generally improves recovery across the core design. Under Normal errors and low collinearity, RMSE(a_1) decreases from 0.285 (S1) to 0.141 (S5), while RMSE(IE_2) decreases from 0.174 to 0.066. Under high collinearity and Normal errors, RMSE(a_1) decreases from 0.524 (S2) to 0.214 (S6), and RMSE(a_2) from 0.555 to 0.194. Similar improvements are observed under Student-t conditions, indicating that additional information primarily improves separation of the parallel mediator regressions and stabilizes the associated indirect effects.

Table 4: Parameter recovery across core scenarios (S1–S8) and feasibility settings (F1–F2). RMSE uses posterior medians across replications; Coverage is the proportion of 95% credible intervals containing the true value. Feasibility settings are descriptive only (two indicators per construct).

Scenario	n	Error	ρ	RMSE(a_1)	Cov	RMSE(a_2)	Cov	RMSE(b)	Cov	RMSE(IE_1)	Cov	RMSE(IE_2)	Cov	Bias(ρ)	RMSE(ρ)	Cov(ρ)
S1	22	Normal	0.30	0.285	0.98	0.321	0.92	0.261	0.96	0.169	0.96	0.174	0.94	0.033	0.227	0.94
S2	22	Normal	0.80	0.524	0.98	0.555	0.92	0.255	0.96	0.284	0.98	0.285	0.96	0.024	0.092	0.94
S3	22	$t(5)$	0.30	0.377	0.88	0.286	0.96	0.270	0.94	0.181	0.96	0.159	1.00	0.034	0.217	0.96
S4	22	$t(5)$	0.80	0.598	0.92	0.521	0.96	0.256	0.96	0.285	0.96	0.276	0.98	0.025	0.084	0.98
S5	75	Normal	0.30	0.141	0.94	0.118	0.98	0.115	0.96	0.086	0.92	0.066	0.96	0.003	0.113	0.94
S6	75	Normal	0.80	0.214	0.94	0.194	0.98	0.115	0.96	0.116	0.94	0.098	1.00	0.004	0.045	0.94
S7	75	$t(5)$	0.30	0.141	0.94	0.127	1.00	0.128	0.96	0.076	0.98	0.069	0.98	-0.002	0.113	0.96
S8	75	$t(5)$	0.80	0.196	0.94	0.184	1.00	0.131	0.92	0.100	0.98	0.100	0.96	0.003	0.045	0.96
Feasibility (descriptive only; $R = 30$, two indicators per construct)																
F1	10	Normal	0.80	1.064	0.967	1.036	0.933	0.366	0.933	0.379	1.000	0.345	1.000	-0.043	0.196	0.967
F2	10	$t(5)$	0.80	1.158	0.933	0.998	0.900	0.346	0.967	0.554	0.967	0.418	1.000	-0.055	0.239	0.933

Effect of exogenous collinearity. Exogenous collinearity is the dominant source of recovery deterioration for the parallel paths into the mediator. Holding sample size and error family fixed, moving from $\rho = 0.30$ to $\rho = 0.80$ consistently inflates RMSE for a_1, a_2, IE_1, and IE_2. Under Normal errors at $n = 22$, RMSE(a_1) increases from 0.285 (S1) to 0.524 (S2), and RMSE(a_2) from 0.321 to 0.555; under Student-t errors, the corresponding increases are from 0.377 to 0.598 for a_1 and from 0.286 to 0.521 for a_2 (S3 vs. S4). The same pattern remains visible at $n = 75$ (S5 vs. S6; S7 vs. S8), although with smaller absolute RMSE. These contrasts support the interpretation that strong exogenous dependence reduces separability of the two parallel effects on the mediator and propagates uncertainty to the mediated effects.

Effect of heavy-tailed DGP errors. Heavy-tailed errors mainly aggravate recovery in the smaller-sample settings. At $n = 22$ and $\rho = 0.30$, RMSE(a_1) increases from 0.285 (S1) under Normal errors to 0.377 (S3) under Student-t errors; at $\rho = 0.80$, the corresponding increase is from 0.524 (S2) to 0.598 (S4). For several targets, the additional impact of heavy tails is more modest at $n = 75$, suggesting that increased sample information partly offsets misspecification-related instability under the Gaussian fitted model.

Coverage, graphical summary, and feasibility results. Coverage is generally close to nominal across the core scenarios, although some undercoverage remains in the more demanding small-sample settings, such as Cov(a_1)=0.88 in S3. Fig. 1 summarizes RMSE patterns for the key structural and mediated targets across the core design and visually reinforces the dominant roles of sample size and exogenous collinearity in recovery performance. The feasibility settings F1–F2, which use a different measurement structure, are markedly less stable for a_1, a_2, and the indirect effects and are therefore interpreted descriptively only rather than as part of the core factorial comparison.

Monte Carlo uncertainty. Because $R = 50$ replications are used per core scenario, small cross-scenario differences should not be overinterpreted when they are comparable to Monte Carlo variability. For example, Cov(a_1)=0.88 in S3 implies a binomial Monte Carlo standard error of approximately 0.046. Detailed supplementary summaries report updated scenario-level RMSE results and coverage Monte Carlo standard errors (Tables S1–S2), while Figure 1 displays bootstrap 95% intervals for the principal RMSE contrasts.

3.4. Discussion and practical implications

Taken together, the diagnostic summaries (Table 3) and the Monte Carlo recovery results (Table 4) show that, for the specific model class and estimation setup studied here, BSEM implemented in `blavaan` via Stan/NUTS remains computationally feasible across the core scenarios, whereas inferential accuracy depends strongly on sample size and exogenous collinearity. In particular,

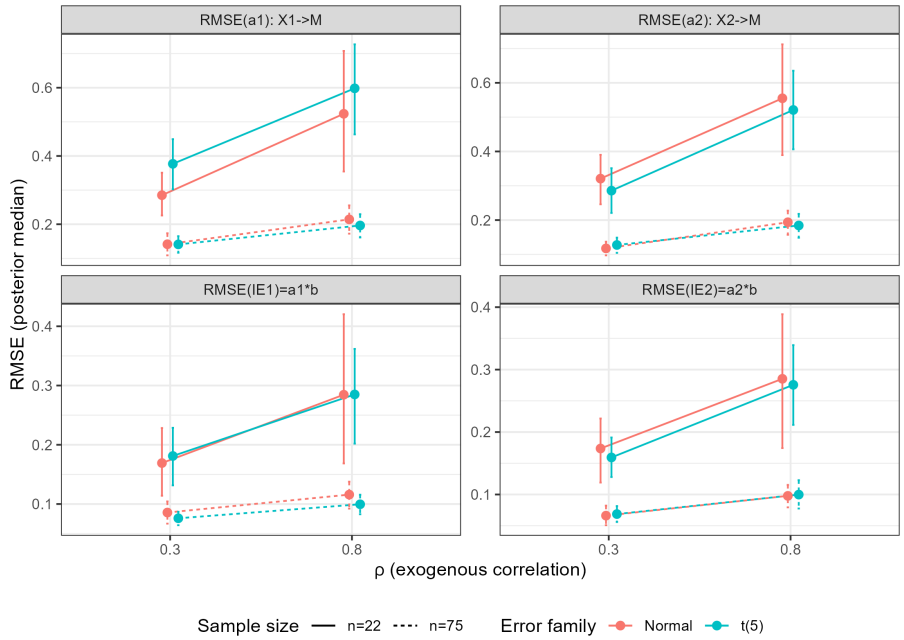

Fig. 1: RMSE for key targets (a_1, a_2, IE_1, IE_2) across the core $2 \times 2 \times 2$ design. Colors denote error family, line types denote sample size, and error bars are bootstrap 95% intervals. Panels use free y-scales for readability.

high exogenous correlation reduces separability between the parallel effects $X_1 \rightarrow M$ and $X_2 \rightarrow M$, which inflates uncertainty and weakens recovery for (a_1, a_2) and, consequently, for the mediated effects ($IE_1 = a_1 b$, $IE_2 = a_2 b$). This mechanism is most consequential at $n = 22$, where likelihood information is weakest and the trade-off between a_1 and a_2 is most pronounced.

Measurement-model recovery is treated as secondary because the manipulated factor ρ primarily affects separability of the parallel structural effects rather than within-construct loading relations. Under the fixed confirmatory measurement design, loading-recovery summaries can be computed analogously but are omitted from the main text to maintain focus on structural robustness.

From a computational perspective, no divergent transitions were observed in the core scenarios, median tree-depth usage remained stable at 8, and convergence summaries stayed close to 1, with $\widehat{R}_{\text{max,med}}$ approximately 1.007–1.014. $\text{ESS}_{\text{min,med}}$ ranged from about 418 to 851 and was consistently smaller under stronger exogenous collinearity, indicating reduced effective information in the more difficult scenarios. Because the most difficult scenario (S8) was estimated under tighter tuning than the remaining core scenarios, these diagnostics are interpreted as evidence of feasible posterior exploration under the scenario-specific tuning regime used here rather than as a strict like-for-like comparison of computational behavior across conditions.

PPP provides complementary information about global fit calibration. Under Normal-error scenarios, PPP_{med} remains moderate, whereas PPP is lower and extreme PPP behavior is somewhat more common under Student-t data generation. Because the Student-t scenarios introduce heavy tails only in the DGP errors while the fitted SEM retains the Gaussian residual specification, lower PPP is interpreted as weaker calibration of the fitted discrepancy under heavy-tailed misspecification rather than as evidence of convergence failure. Accordingly, PPP should be interpreted jointly with sampler diagnostics and recovery summaries.

The finite replication budget also implies non-negligible Monte Carlo uncertainty in reported performance measures. Accordingly, closely spaced differences in RMSE or coverage should be interpreted cautiously when they are comparable to Monte Carlo variability; supplementary tables provide updated scenario-level RMSE summaries and coverage Monte Carlo standard errors (Tables S1–S2).

Two practical implications follow for district-level applications under settings similar to those studied here. First, analysts should report a compact diagnostic bundle—PPP, $\widehat{R}_{\max}$, $\mathrm{ESS}_{\min}$, tree depth, and divergence counts—together with recovery-relevant summaries for structural and mediated effects. Second, mediated effects should be interpreted cautiously when exogenous constructs are highly correlated, especially at very small n, where separability is weakest and uncertainty inflation is most severe. These conclusions are specific to the full-mediation BSEM and estimation regime studied here, in which indicator residual variances are fixed to their data-generating values.

4. Conclusion

This study evaluated the operating characteristics of Bayesian structural equation modeling (BSEM) for a district-level simulation setting characterized by small sample size, heavy-tailed errors, and exogenous collinearity. Using a fixed confirmatory full-mediation SEM and a balanced $2 \times 2 \times 2$ Monte Carlo design, we examined global fit calibration, sampler diagnostics, and parameter recovery under variation in $n \in \{22, 75\}$, $\rho \in \{0.30, 0.80\}$, and Normal versus variance-matched Student-t error generation with $\nu = 5$.

Across the studied settings, larger sample size improved posterior exploration and parameter recovery, whereas strong exogenous collinearity was the main source of deterioration for the parallel mediator regressions and their associated indirect effects. Heavy-tailed data generation primarily reduced PPP relative to the Normal-error cases, consistent with misspecification-based fit-calibration stress under a Gaussian fitted SEM. Under the scenario-specific tuning used here, all core scenarios showed feasible posterior exploration and no divergent transitions were observed; however, these computational results should not be interpreted as a strict like-for-like comparison across all conditions because the most difficult scenario required tighter tuning.

These findings are specific to the model class and estimation regime studied and should not be generalized to Bayesian SEM broadly. In particular, the fitted SEM uses `std.lv=TRUE` and fixes indicator residual variances to their data-generating values. At the same time, the variances of the endogenous latent variables are induced by the structural equations and therefore vary with ρ, so the effective signal scale for the endogenous indicators is not strictly invariant across all conditions. In addition, performance summaries are based on a finite replication budget, so small differences in RMSE or coverage should be interpreted cautiously when they are comparable to Monte Carlo variability.

For district-level applications under settings similar to those studied here, analysts should report PPP together with key posterior exploration diagnostics and recovery-relevant summaries for structural and mediated effects. Mediated effects should also be interpreted cautiously when exogenous constructs are highly correlated, especially at very small n, where separability is weakest and uncertainty inflation is most severe.

CRediT authorship contribution statement

Agustina Susi Susanti Parung: Conceptualization; Methodology; Software; Formal analysis; Investigation; Data curation; Writing – original draft; Visualization.
Ani Budi Astuti: Supervision; Methodology; Writing – review & editing; Validation.
Rahma Fitriani: Supervision; Writing – review & editing; Validation.

Declaration of Generative AI and AI-assisted technologies

During manuscript preparation, generative AI tools were used solely for language editing improving clarity and organization of prose, and assisting with LATEX formatting. No generative AI tool was used to generate, analyze, or interpret data, run simulations, or draw scientific conclusions. All authors reviewed, verified, and edited the manuscript and take full responsibility for the content.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research received no external funding. The authors thank the Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, for academic support and constructive feedback during the development of this work.

Data and Code Availability

The simulation code (data-generating scripts, Bayesian estimation routines, and post-estimation summarization code) is available from the corresponding author upon reasonable request. Requests should be addressed to `agustinaparung@student.ub.ac.id`.

References

[1] S. P. Manongga, H. Hendry, and D. H. F. Manongga. "Predicting a stunting prevalence using semi-supervised learning models in East Nusa Tenggara". In: *International Journal of Innovative Computing, Information and Control* 19.4 (2023), pp. 1073–1086. DOI: 10.24507/ijicic.19.04.1073.

[2] M. Ariani. "Determinan Penyebab Kejadian Stunting Pada Balita: Tinjauan Literatur". In: *Dinamika Kesehatan: Jurnal Kebidanan dan Keperawatan* 11.1 (July 2020), pp. 172–186. DOI: 10.33859/dksm.v11i1.559.

[3] T. A. E. Permatasari et al. "The association of sociodemographic, nutrition, and sanitation on stunting in children under five in rural area of West Java Province in Indonesia". In: *Journal of Public Health Research* 12.3 (July 2023). DOI: 10.1177/22799036231197169.

[4] T. A. Whittaker and R. E. Schumacker. *A Beginner's Guide to Structural Equation Modeling*. New York: Routledge, 2022. DOI: 10.4324/9781003044017.

[5] R. H. Hoyle, ed. *Handbook of Structural Equation Modeling*. London and New York: The Guilford Press, 2023.

[6] J. C. Westland. *Structural Equation Models*. Vol. 22. Studies in Systems, Decision and Control. Cham: Springer International Publishing, 2019. DOI: 10.1007/978-3-030-12508-0.

[7] Alexander Robitzsch. "Comparing the Robustness of the Structural after Measurement (SAM) Approach to Structural Equation Modeling (SEM) against Local Model Misspecifications with Alternative". In: *Stats* 3.2 (2022), pp. 631–672. DOI: 10.3390/stats3020009.

[8] L. J. Jobst, M. Auerswald, and M. Moshagen. "The Effect of Latent and Error Non-Normality on Measures of Fit in Structural Equation Modeling". In: *Educational and Psychological Measurement* (2022). DOI: 10.1177/00131644211046201.

[9] M. E. Civelek. *Essentials of Structural Equation Modeling*. Zea Books. Mar. 2018. DOI: 10.13014/K2SJ1HR5.

[10] N. W. S. Wardhani et al. "Structural Equation Modeling (SEM) Analysis with WarpPLS Approach Based on Theory of Planned Behavior (TPB)". In: *Mathematics and Statistics* 8.3 (2020), pp. 311–322. DOI: 10.13189/ms.2020.080310.

[11] J. J. Thakkar. *Structural Equation Modelling*. Vol. 285. Studies in Systems, Decision and Control. Singapore: Springer Singapore, 2020. DOI: 10.1007/978-981-15-3793-6.

[12] W. H. Nugroho et al. "Robust Regression Analysis Study for Data with Outliers at Some Significance Levels". In: *Mathematics and Statistics* 8.4 (2020), pp. 373–381. DOI: 10.13189/ms.2020.080401.

[13] S. Depaoli. *Bayesian Structural Equation Modeling*. New York, United States of America: The Guilford Press, 2021.

[14] H. Liu et al. "A New Bayesian Structural Equation Modeling Approach with Priors on the Covariance Matrix Parameter". In: *Journal of Behavioral Data Science* 2.2 (2022), pp. 1–24. DOI: 10.35566/jbds/v2n2/p2.

[15] A. Gelman et al. *Bayesian Data Analysis*. Third. CRC Press, 2014.

[16] A. A. Johnson, M. Q. Ott, and M. Dogucu. *Bayes Rules!* Boca Raton: Chapman and Hall/CRC, 2022. DOI: 10.1201/9780429288340.

[17] L. Ben. *A Student's Guide to Bayesian Statistics*. 1st ed. London: SAGE Publications, 2018.

[18] X. Y. Song and S. Y. Lee. "A tutorial on the Bayesian approach for analyzing structural equation models". In: *Journal of Mathematical Psychology* 56.3 (2012), pp. 135–148. DOI: 10.1016/j.jmp.2012.02.001.

[19] R. Jacobucci and K. J. Grimm. "Comparison of Frequentist and Bayesian Regularization in Structural Equation Modeling". In: *Structural Equation Modeling: A Multidisciplinary Journal* 25.4 (July 2018), pp. 639–649. DOI: 10.1080/10705511.2017.1410822.

[20] Edgar C. Merkle et al. "Efficient Bayesian Structural Equation Modeling in Stan". In: *Journal of Statistical Software* 100.6 (2021), pp. 1–22. DOI: 10.18637/jss.v100.i06.

[21] Edgar C. Merkle and Yves Rosseel. "blavaan: Bayesian Structural Equation Models via Parameter Expansion". In: *Journal of Statistical Software* 85.4 (2018), pp. 1–30. DOI: 10.18637/jss.v085.i04.