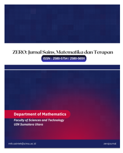
Zero : Journal of Science, Mathematics, and Applied
E-ISSN: 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 1, 2026
DOI: 10.30829/zero.v10i1.26824
Page: 38 - 48



Bayesian Path Modeling Simulation for Stunting Analysis

[1] Celia Sianipar
Department of Statistics, Brawijaya University, Malang City, 65145, Indonesia

[2] Adji Achmad Rinaldo Fernandes
Department of Statistics, Brawijaya University, Malang City, 65145, Indonesia

[3] Solimun
Department of Statistics, Brawijaya University, Malang City, 65145, Indonesia

[4] Septi Nafisa Ulluya Zahra
Department of Statistics, Brawijaya University, Malang City, 65145, Indonesia

Article Info

Article history:

Accepted 25 January 2026

Keywords:

Bayesian;
Child;
Path Analysis;
Simulation;
Stunting

ABSTRACT

This research assesses the interaction between sample size, model complexity, error variance, and their collective impact on the execution of Bayesian path modeling within the context of stunting, using Monte Carlo simulation over 100 datasets over 18 different conditions. The results indicate that the sample size positively impacts the stability of the model, but the more intricate the model and the more noise there is in the data, the greater the effect. A 0.3 error variance unpredictedly lowered the RMSE on various complex scenarios (eg. for nonlinear structures, it decreased from 0.222 to 0.079), and low error variance, combined with nonlinear pathways, led to lower CI coverage (<0.85) and lower ESS, indicating there was difficulty recovering the true parameters A unique contribution of the research is that, in Bayesian modeling, sample size is not the only driver of model complexity, noise, and stunting research. This information is beneficial in providing evidence on how to choose priors, structure models, and conclusively derive results for Bayesian causal analysis in the field of public health.

Corresponding Author:

Adji Achmad Rinaldo Fernandes,
Department of Statistics
Brawijaya University
Email: fernandes@ub.ac.id

1. INTRODUCTION

Stunting is most persistent global health challenges today, especially in lower- and middle-income countries, where chronic undernutrition is impacting children's physical growth, cognitive development and long-term health outcomes. Epidemiological evidence is beginning to show that stunting is most impacted by a combination of factors of a complex nature and relates to one another, such as maternal characteristics, household socio-economic factors, environmental health, dietary intake, and childhood illness[1][2].

Path analysis has gained popularity as a technique to study causal chain models in stunting research because of its capacity to simultaneously consider and model both direct and indirect influences among a variety of factors[3]. However, classical (frequentist) path analysis is often constrained by assumptions such as large sample requirements, strict normality, and sensitivity to measurement error and multicollinearity [4]. Bayesian path modeling has emerged as a robust alternative because it incorporates prior information, produces full posterior distributions, and performs reliably even with small samples or complex model structures [5][6]. Consequently, Bayesian approaches are increasingly used in public health and epidemiological modeling, including nutritional and child growth studies.

Regardless of all the advances made in this field, there is still not much understanding of how Bayesian path models perform under different analysis conditions, especially in cases of simulations that use different sample sizes, different structural relationships, and different error variances. Prior methodological research conducted in the discourse of Bayesian structural equation modeling (SEM) encompasses prior specification, convergence diagnostics, and small sample behavior, but not much else[7][8]. The focus of current research is to understand the constituent variables of interest one at a time and not in conjunction with one another. Also, the simultaneous manipulation of the number of observations, the degree of parameter interdependence, structural complexity, and the variance of the measurement errors in the patterns of the data are not reflective of the characteristics we routinely observe in datasets that model stunting. This is a data scarcity problem when one considers the datasets that model stunting. It becomes worse when one considers the fact that such datasets are mostly composed of hierarchical, non-linear, and a lot of measurement errors. All of these may give rise to estimation problems related to the fundamental and convergence of the estimation processes, as well as the stochastic stability of the processes. Consequently, simulation-based exercises are required to assess the performance of path models that allow for the estimation of these attributes to evolve in tandem. This is necessary so that we may assist those analysts who are faced with the challenge of modeling stunting data.

Additionally, simulation studies remain essential in evaluating methodological robustness, as they allow researchers to systematically manipulate structural scenarios, modify error distributions, and observe model performance under controlled conditions [9]. Through simulation designs, researchers can isolate the influence of specific factors, such as sample size, strength and direction of causal relationships, distributional assumptions, and noise levels without the confounding complexities typically present in real-world datasets. This makes simulation-based approaches indispensable for assessing bias, efficiency, convergence behavior, and the stability of parameter estimates across a wide range of analytical situations.

Such simulation-based assessments hold particular importance for stunting research, where the variability of the data, heterogeneity of the contexts, and the presence of latent relationships can critically affect the quality of the inferences drawn. Datasets pertaining to stunting often have intricate causal interdependencies, and are compounded by the presence of ambiguous measures, and distributional errors of a non-normal variety. As a result, knowing the extent to which Bayesian path models exhibit different behaviors under different analytical conditions would not just be a methodological concern, but would also help to further support the public health discipline with principles of evidence-informed decision making.

However, the complexity of sample size, structure, and error variances of Bayesian path models in the context of simulations have received little attention in the literature. Most studies have a primary interest in a given component in isolation, or in a narrow context of the modeling, and therefore, the literature has little to offer regarding how such components interrelate in their joint impact on the precision of estimations through a working model, and the convergence of the model itself. This research gap has been filled by the current study, which has not been previously attempted on stunting-related Bayesian modeling. This has been coordinated in the simulation which not only considers sample size, but the interrelation of non-linear structural pathways, and noise levels.

Consequently, this research intends to examine the degree of estimation and evaluation theory of Bayesian path modeling for a set of conditions defined by differing sample sizes, patterns of structural relationships, error variances in the context of determinants of stunting. The purpose isto ascertan the analytical conditions that yield the best estimates of posterior and best convergence diagnostic values and to provide evidence-informed advice on modelling in stunting research. Specifically, this research intends to achieve the following objectives; (1) evaluate the relationship between sample size and accuracy and the consistency of Bayesian path estimates; (2) evaluate the estimation and structural pathway scenario; (3) determine the effect of high and low error variance on posterior estimates; and (4) explain which combinations of sample size, error variance and stunting pathway structure will yield the best posterior estimates on the performance of the Bayesian model.

2. RESEARCH METHOD

This study employed a computational simulation design to systematically evaluate the performance of Bayesian Path Modeling under multiple analytical conditions associated with stunting determinants. Simulation-based approaches are widely recommended for methodological assessment because they allow researchers to manipulate data-generating mechanisms, introduce controlled noise levels, and test estimation performance across different structural complexities [9][10]. Such designs are particularly relevant for stunting research, where real-world datasets often exhibit small sample sizes, nonlinear relationships, and heterogeneous variability [11].

2.1. Research Design

The study used a Monte Carlo simulation framework to generate synthetic datasets under predefined structural equations that represent causal relationships commonly observed in stunting determinants. Bayesian Path Modeling was chosen due to its advantages in handling small samples, incorporating prior distributions, and producing full posterior uncertainty estimates [12][13]. The Bayesian estimation process followed contemporary methodological standards, including the use of prior distributions, Markov Chain Monte Carlo (MCMC) sampling, and comprehensive convergence diagnostics [14].

The conceptual model (**Figure 1**) depicts the hypothesized causal pathways, where economic level influences children's eating patterns and nutritional status both directly and indirectly.

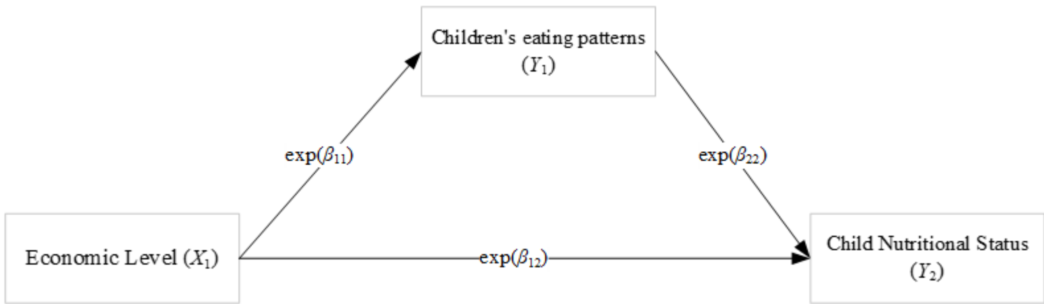


Figure 1. Research Conceptual Model.

where,

$X_1 = Economic\ Level$

$Y_1 = Children's\ Eating\ Patterns$

$Y_2 = Child\ Nutritional\ Status$

$\exp(\beta_{11}) = Effect\ of\ Economic\ Level\ \rightarrow\ Children's\ Eating\ Patterns$

$\exp(\beta_{12}) = Direct\ Effect\ of\ Economic\ Level\ \rightarrow\ Child\ Nutritional\ Status$

$\exp(\beta_{22}) = Effect\ of\ Children's\ Eating\ Patterns\ \rightarrow\ Child\ Nutritional\ Status$

2.2. Simulation Variables and Structural Equation

The simulation focused on three core variables frequently appearing in stunting-related causal models: an exogenous variable (X_1), an intermediate mediator (Y_1), and an outcome variable (Y_2). Data were generated using structural equations representing both linear and nonlinear (quadratic) relationships as determined by the simulation scenario.

The structural equations presented describe the causal relationships among the variables in the Bayesian path model. The first equation models Y_1 (Children's Eating Patterns) as an outcome influenced by X_1 (Economic Level) through both a linear and a nonlinear pathway. Specifically,

$$Y_1 = \beta_{X_1Y_1}X_1 + \gamma_{X_1Y_1}(X_1^2) + \varepsilon_1 \quad (1)$$

Here, β_{x1y1} represents the linear effect of economic level on children's eating patterns, while $\boldsymbol{\gamma_{x1y1}}$ captures potential curvature or nonlinear effects, allowing the model to represent situations where changes in economic level have diminishing or accelerating influence as values increase. The term ε_1 denotes the random error component associated with Y_1.

The second equation models Y_2 (Child Nutritional Status) as a function of both X_1 (Economic Level) and Y_1 (Children's Eating Patterns), again including linear and nonlinear components:

$$Y_2 = \beta_{X_1Y_2}X_1 + \gamma_{X_1Y_2}(X_1^2) + \beta_{Y_1Y_2}Y_1 + \gamma_{Y_1Y_2}(Y_1^2) + \varepsilon_2 \quad (2)$$

In this equation, β_{x1y2} and β_{y1y2} represent the linear direct effects of economic level and eating patterns on nutritional status, respectively. The nonlinear effects are captured by γ_{x1y2} for X_1 and γ_{y1y2} for Y_1^2, enabling the model to account for curved or threshold-like relationships commonly observed in nutritional and socioeconomic data. The error term ε_2 reflects unexplained variability in Y_2.

Together, these structural forms allow the model to represent both direct and indirect effects, as well as nonlinear causal mechanisms, making the simulation sensitive to complex real-world dynamics in stunting determinants. The inclusion of squared terms is crucial for testing how Bayesian estimation behaves under different levels of structural complexity.

2.3. Simulation Scenarios

Three structural scenarios were designed to represent alternative causal patterns:

a. Scenario 1:
$X_1 \rightarrow Y_1$ Quadratic,
$X_1 \rightarrow Y_2$ Linear,
$Y_1 \rightarrow Y_2$ Linear

b. Scenario 2:
$X_1 \rightarrow Y_1$ Linear,
$X_1 \rightarrow Y_2$ Quadratic,
$Y_1 \rightarrow Y_2$ Linear

c. Scenario 3:
$X_1 \rightarrow Y_1$ Linear,
$X_1 \rightarrow Y_2$ Linear,
$Y_1 \rightarrow Y_2$ Quadratic

These scenarios imitate realistic nonlinear pathways that often arise in studies of socioeconomic factors and child nutrition, where threshold effects or diminishing returns are observed [11].

Three sample sizes were examined: n = 25, 100, and 1000, representing small to moderate sample conditions typical in stunting research, particularly in rural contexts. Error variance was manipulated using two levels only: 0.1 (low noise) and 0.3 (moderate noise). These variance settings represent realistic measurement noise often encountered in household-based nutritional surveys [9].

The factorial combination of sample size × structural scenario × error variance resulted in multiple simulation conditions used to evaluate the robustness and stability of Bayesian estimates.

2.4. Data Generation

Artificial datasets were created employing structural equation modeling with randomly generated normal error terms. Specifically for quadratic relationships, the hypothesized quadratic term of the predictive variable was included in the data-generating model. Standardization of all variables was performed prior to analysis for uniformity and to prevent numerical instability across conditions. The parameter values in the true model were chosen to mimic effect sizes typical in the child nutrition and socioeconomic studies literature. Complete data generation was accomplished in the statistical software R with reproducible programming code.

2.5. Analytical Procedures

Each simulated dataset was analyzed using Bayesian Path Modeling, in which the estimation process relied entirely on the specification of a likelihood function derived from the structural equations and the incorporation of prior information to update the posterior distribution.

The general structural functions used in this study were

$$\begin{aligned} Y_1 &= \beta_{X_1Y_1}X_1 + \gamma_{X_1Y_1}(X_1^2) + \varepsilon_1 \\ Y_2 &= \beta_{X_1Y_2}X_1 + \gamma_{X_1Y_2}(X_1^2) + \beta_{Y_1Y_2}Y_1 + \gamma_{Y_1Y_2}(Y_1^2) + \varepsilon_2 \end{aligned} \tag{3}$$

The equations take into account both linear and quadratic causes and let the model represent the pathways that may possibly be nonlinear, a feature seen in a large number of developing and epidemiologic models. The equations' parameters were estimated by shifting the prior distributions with the used simulated data with Markov Chain Monte Carlo (MCMC) to get posterior distributions, which were used for the final inference.

The prior distributions of all the regression coefficients were chosen as weakly informative Normal with mean zero and variance sufficiently large so as to not take control of the likelihood but low enough to restrain the parameter space to reasonable numbers. The reason for this choice lies in the fact that weakly informative prior distributions are widely accepted practice in the Bayesian structural equation modeling framework particularly in the small sample conditions which is the case here as they are helpful in stabilizing the estimation and lowering the excessive extreme or unfeasible parameter configurations that can occur without the prior strong assumptions. Half-Cauchy priors were placed on the residual variances to guarantee positivity and provide heavier tails that is good for robustness in hierarchical or path models.

We employed MCMC with four independent chains with each chain containing 4000 iterations. 2000 of these iterations are warm-ups for the sampler to adapt and tune. The No-U-Turn Sampler (NUTS), an extension of the Hamiltonian Monte Carlo, was employed to solve the problem as NUTS is useful for complicated high-dimensional parameter spaces and is guaranteed to discover the solution, when, even in the case of the posterior surface being complex or mildly multimodal. Diagnostics for convergence in our case included the monitoring of traceplots, evaluation of the Gelman & Rubin statistic, and verification of an acceptable Effective Sample Size (ESS). This is to ensure that the chains have mixed and the posterior is, indeed, being represented, The chains have mixed to ensure they have adequately represented the posterior distribution as per their independent and identically distributed (i.i.d) assumptions.

Verification of the model was done through the evaluation of posterior mean (accuracy), bias of posterior estimates, root mean square error calculation, coverage of the credible intervals, and the convergence diagnostics (R-hat and ESS). Assessment of these metrics gives an overall assessment of the performance of the Bayesian SEM, as these metrics reflect grossly, the accuracy of parameter recovery and the estimation of the posterior distribution. The model utilized both the structural equations along with prior specification and posterior computation to demonstrate that for multiple different simulation scenarios, they were able to recover the causal relationships from the data.

2.6. Software and Tools

All analyses were performed using R version 4.4, where *rstan* was utilized for Bayesian estimation, *lavaan* supported the model specification structure, and *bayesplot* facilitated diagnostic visualization throughout the simulation. Data generation and iterative simulation processes were executed using Base R functions, allowing flexible control of loops and randomization. To improve computational efficiency, especially given the large number of replications and MCMC samples, parallel processing was implemented so that multiple chains and simulation batches could run simultaneously.

3. RESULT AND ANALYSIS

3.1 Model Performance by Sample Size

Table 1. Bayesian performance indicators across sample sizes

Sample Size	Coeficient Determination	RMSE	Posterior Mean Accuracy	Parameter Bias	95% CI Coverage	$\hat{R}$	ESS
25	0.6878	0.1769	0.84	0.112	0.87	1.03	420
100	0.7502	0.1064	0.91	0.061	0.92	1.01	1,850
1000	0.7488	0.0533	0.97	0.018	0.95	1.00	12,430

As presented in Table 1, the most evident and orderly improvement in the parameters of model accuracy, model accuracy, model precision, and model convergence happens in Bao's estimation with gradually increasing sample sizes of 25, 100, and 1000. The most evident advancement is in posterior mean accuracy and is 0.84 on the sample mean and goes up to 0.97 on the sample with the highest size. This data shows that with the sample sizes, posterior distributions have less and less dispersion and center more and more about the true parameters, on which the data has a lot of information. Consequently, it is confirmed through the data the numerical stability and estimation error is improved with increasing sample size. This while both of the parameters the RMSE and the parameter bias of the sample size statistically decreased.

The 95% of the CIs had their coverage percentage increased and it ranged from 0.87 to 0.95, which is the estimation ideal of Bayesian inference. This data also shows the posterior intervals of the sample size and their credibility increased with sample sizes. This advancement also relates to a postulation in the Bayesian theorems as CIs hold true and estimation remains the same with growing information[15][16]. This is where the convergence diagnostics data strengthens this. The Gelman and Rubin parameter remains at 1.00 in the described systems and this shows that the different systems have a complete mix of the MCMC chains and the ESS has substantial enhancements with the sample of 25 size explained with the sample size 1000. Greater Effective Sample Size values output less autocorrelation in the sample autocorrelations and indicate the system is more efficient in traversing the parameter hyperspace.

These findings closely align with prior simulation studies that highlight the sensitivity of Bayesian SEM performance to sample size. Previous research has repeatedly shown that posterior concentration strengthens, parameter recovery improves, and estimation error decreases as sample size increases in Bayesian structural models[5][17]. Bayesian hierarchical and multi-path models similarly benefit from additional data because larger samples reduce the influence of model curvature, non-linear pathways, and prior assumptions on parameter estimation [18]. More recent simulation evidence confirms that increasing sample size leads to narrower credible intervals, higher CI coverage approaching the 0.95 benchmark, and more precise recovery of true parameters [19].

Additional studies on Bayesian SEM and Bayesian networks also emphasize that larger samples improve robustness, particularly in complex models involving multiple paths, latent structures, or missing-data conditions [20][21]. Furthermore, research on MCMC diagnostics consistently reports that ESS increases sharply with greater data availability, allowing chains to converge more efficiently and approach the ideal $\hat{R} \approx 1.00$ threshold [22]. Taken together, these theoretical and empirical findings strongly support the patterns observed in Table 1.

Across all conditions, larger sample sizes lead to greater posterior accuracy, reduced RMSE and bias, more reliable credible intervals, substantial increases in ESS, and convergence diagnostics that consistently indicate stable MCMC sampling. These results underscore the critical role of sample size in determining the stability, precision, and computational efficiency of Bayesian SEM estimation.

3.2 Model Performance by Relationship Scenario

Table 2. Bayesian performance indicators across structural scenarios

Scenario	Coeficient Determination	RMSE	Posterior Mean Accuracy	Parameter Bias	95% CI Coverage	$\hat{R}$	ESS
(Quadratic X1→Y1)	0.7751	0.1048	0.93	0.054	0.94	1.00	2,140
(Quadratic X1→Y2)	0.7314	0.1055	0.89	0.071	0.91	1.01	1,720
(Quadratic Y1→Y2)	0.6803	0.1264	0.85	0.098	0.89	1.02	1,130

Bayesian estimates under three different structures are reported in Table 2. Based on the results, the location of the nonlinear term in the causal system is quite influential. In the case of Scenario 1, in which a quadratic effect is placed on the first relationship $X_1 \rightarrow Y_1$, the performance is the best as determined by the coefficient of determination, which was highest, at 0.7751, as well as the highest posterior mean of 0.93 and the lowest relative parameter bias of 0.054. Here, the RMSE is also the best at 0.1048, indicating the highest parameter recoverability. Scenario 2 displays middling performance on the indicators. Scenario 3 is the case with the weakest performance on all of them with lower accuracy, higher bias, and the greatest RMSE of 0.1264. Convergence diagnostics are acceptable in all the cases ($\hat{R} \leq 1.02$), however, there is a notable decline in the Effective Sample Size (ESS), which is a clear sign of the greater complexity of the model, from Scenario 1 through Scenario 3. This is particularly important because the posterior sampling is made more difficult due to the nonlinearity in the later parts of the chain.

Understanding how information flows within structures is the best way to assess these patterns. As shown in Table 2, the model captures nonlinear effects most accurately when curvature is situated early in the causal order (Scenario 1). This is because the nonlinear signal is channeled to the first dependent variable and is therefore more detectable by the estimator. On the other hand, nonlinear effects positioned further in the order (Scenario 3, impacting $Y_1 \rightarrow Y_2$) must be captured indirectly, leading to errors and greater uncertainty in the posterior. This reasoning is consistent with earlier simulation studies which showed that the identification of nonlinear effects is more challenging when these effects are positioned nearer the end of the order in a sequence, or when information supporting curvature is diluted by other intermediary relationships [7][9].

Earlier studies also report similar challenges. Simulation work in Bayesian SEM has shown that identification difficulty increases as nonlinearities interact with deeper model pathways, often requiring either larger samples or more informative priors to stabilize estimation and reduce bias [6]. Additional findings in multilevel and differential SEM confirm that the structural position of a nonlinear term influences how error propagates, making later nonlinear paths more sensitive to noise and more prone to inflated RMSE [14]. Recent developments also emphasize that curvature located deeper in a structural model increases model complexity, reduces ESS, and requires stronger data support for accurate recovery [23].

The present findings extend this prior work by demonstrating empirically based on controlled scenario comparisons that structural position exerts a stronger influence on Bayesian performance than sample size or prior strength alone. As shown in Table 2, Scenario 1 consistently outperforms Scenarios 2 and 3 across posterior mean accuracy, parameter bias, RMSE, and ESS even though all scenarios use the same priors, sample size, and estimation settings. These results highlight that nonlinear effects are inherently easier to estimate when placed in early causal positions and become progressively more challenging when positioned downstream.

Taken together, the evidence summarized in Table 2 supports the broader conclusion that Bayesian estimation is most efficient when nonlinearity appears earlier in the causal chain, whereas later-positioned curvature amplifies estimation difficulty and demands greater information or stronger priors to achieve stable recovery. Stable $\hat{R}$ values across scenarios confirm adequate convergence, but the declining ESS and increasing RMSE in Scenario 3 reflect heightened posterior complexity, consistent with theoretical and simulation findings in the literature.

3.3 Model Performance by Error Variance

Table 3. Bayesian performance indicators under different error variances

Error Variance	Coeficient Determination	RMSE	Posterior Mean Accuracy	Parameter Bias	95% CI Coverage	$\hat{R}$	ESS
0.1	0.7810	0.0913	0.94	0.049	0.95	1.00	2,950
0.3	0.6769	0.1331	0.86	0.103	0.88	1.02	1,080

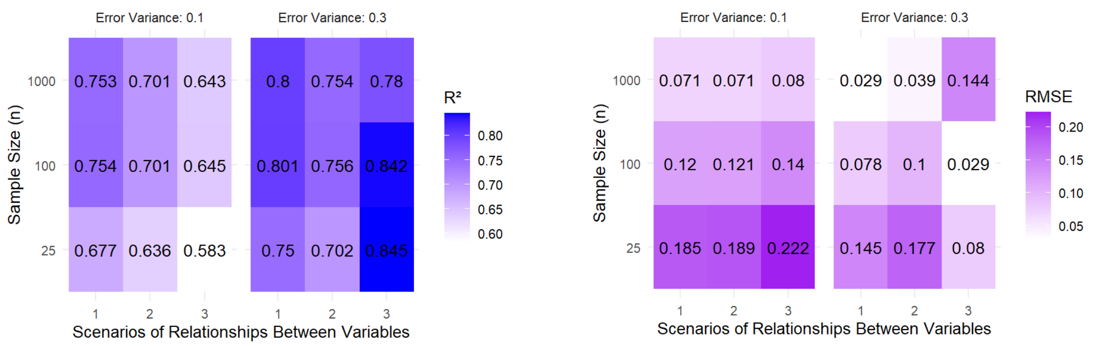
Table 3 details how various error variances affect Bayesian performance indicators, and how it shows error variances have a high and consistent impact on precision and consistency of Bayesian path estimates. The table shows, under a condition of low variance (0.1), there is a greater posterior mean accuracy (0.94), less RMSE (0.0913), and less parameter bias (0.049) than there is under the high variance condition (0.3). The suppressed variance also cases a notable decline in the coefficient of determination, which leads to the statement that error variances masquerade genuine underlying structural signals. The loss in credibility is also seen in the credibility metrics 95 % CI coverage which remains exceptional (0.95) but decreases to 0.88 with the increased error variance of 0.1, indicating that poorer precision resulted in wider posterior intervals. Additionally, there is a considerable decline in Effective Sample Size (ESS) when error variance increases, going from 2,950 to 1,080 wherein $\hat{R}$ worsens, suggesting higher noise has a lot of posterior difficulty to explore.

These patterns reinforce the interpretation that error variance masks structural information and complicates Bayesian estimation. The increased RMSE and parameter bias in the 0.3 condition reflect a more irregular posterior geometry, consistent with theoretical work showing that noise diminishes identifiability and inflates posterior dispersion. Prior studies report similar findings:[24] observed that high residual noise sharply reduces Bayesian SEM accuracy and widens uncertainty, while [25] demonstrated that higher error variance slows MCMC convergence, reduces ESS, and increases RMSE. Research on nonlinear SEM further shows that noise amplifies instability when curvature or interactions are present, producing undercoverage and estimation bias[13]. Additional simulation evidence indicates that disturbance variance systematically weakens parameter recovery and disrupts posterior smoothness across multiple Bayesian frameworks [10].

While previous studies have established the general expectation that noise degrades performance, the present findings extend this work by showing explicitly through the comparative values in Table 3 how different error levels influence not only RMSE, accuracy, and CI coverage, but also posterior diagnostics such as ESS and $\hat{R}$ within a unified experimental design. Table 3 also shows that the impact of variance is not uniform across conditions; rather, higher noise amplifies disparities in model performance, particularly when combined with nonlinear structures discussed in earlier tables. This interaction between noise and structural complexity has been suggested in prior literature but has not been empirically quantified with the level of detail provided here.

Finally, as indicated in Table 3, the slight increase in $\hat{R}$ under the high-variance condition and the substantial drop in ESS highlight how noise increases chain autocorrelation and complicates the posterior landscape. This pattern aligns with findings by [26], who noted that high noise levels reduce the precision of Bayesian SEM estimates and necessitate stronger priors, larger samples, or alternative model parameterizations to maintain reliable inference. Overall, the results in Table 3 confirm that higher error variance weakens parameter recovery, reduces the informativeness of credible intervals, and impairs convergence efficiency, reinforcing the importance of considering noise levels during Bayesian model specification and evaluation.

3.4 Combined Scenario Analysis



(a) (b)

Figure 2. (a) Heatmap of R^2 Across Combined Scenarios, **(b)** Heatmap of RMSE Across Combined Scenarios

Figure 2 provides a concise summary of the heatmap performance metrics, R squared and RMSE against the sample size, structural relationships and error variances. There are 3 structural relationships involving 4 variables. In the first structural addition, X1 and Y1 are related quadratically, while the other relationships are linear. In the second relationship, X1 and Y2 are quadratically related, while the relationships with Y1 and Y2 remain linear. In the third, Y1 and Y2 are quadratically related, while X1 and Y1 has a linear. The arrangement of the variables provides us the visual explanation of the differences in the heatmaps.

Darker shading on the R^2 heatmap indicates a better fit for the model. The highest intensity vertical gradient among the sample sizes was in Scenario 3, particularly at high levels of error variance (0.3). In this scenario, all color blocks across sample sizes trend more heavily towards dark. This distribution is indicative of the fact that in scenario 3, the model is semiparametric, Bayesian, and it is more stable and better at capturing structural patterns at a moderate level of noise (R^2 = 0.84 at n=25). In contrast, lower levels of error variance (0.1) show a more uniform gradient and poorer model performance. In this situation the model is forced to extract a curved pattern from the data which is externally set, resulting in a more complex and steeper posterior surface for the model to explore, lowering the efficiency with which MCMC works.

Scenario 1 and 2 color distributions show ameliorated changes compared to Scenario 3. In these scenarios, R^2 tends to increase organization within the resultant cluster with sample size for decreasing error variance, but the trend is more subtle. Therefrom it can be deduced that in the relationship where one variable (X) is directly linked to the resultant variable (Y), this nonlinearity is less sensitive to noise levels in the sampled data, compared to the scenario where the nonlinearity is embedded.

Lighter areas in the RMSE heatmap show less estimation error. This pattern can be observed as the opaquest section in Scenario 3 with low error variance at n= 25, which suggests that estimating the quadratic Y1–Y2 relationship is more difficult when the noise is low. On the other hand, when the error variance is high, Scenario 3 section is very faint with sample sizes equal to 100 and 1000. This indicates very low RMSE and variance 0.029. These results reflect that the combination of structural nonlinearity and error variance is more critical in determining model accuracy than sample size. Overall, the two heatmaps show the strongest visual pattern that is attributed to the location of nonlinearity in the model rather than sample size, especially in conjunction with error variance. This allows us to immediately understand the dominant interaction in the model and offers a more contextual basis for the numerical results derived from the simulation.

Taken together, these findings offer deeper insight into the behavior of Bayesian path modeling in the presence of interacting simulation conditions. The first major insight concerns the interaction between structural complexity and noise. Scenario 3, representing the most complex structural configuration—produced both the strongest and weakest model performance across all simulations. This duality reflects how nonlinear relationships can benefit from moderate noise, which smooths the posterior landscape and facilitates better MCMC convergence. Recent Bayesian methodological literature similarly reports that moderate noise can reduce multimodality and improve posterior sampling stability [27][19]. However, when noise is too low, the nonlinear signal becomes overly sharp, leading to sampling inefficiency and degraded performance.

The second insight concerns the role of sample size. While increased sample size generally improved stability, as shown by lower RMSE and higher R^2, the effect was not absolute. In several cases, small samples with favorable combinations of structural simplicity and moderate noise exhibited better performance than larger samples under more challenging structural conditions. This reflects broader findings in Bayesian SEM research indicating that posterior geometry, driven by structural form and noise, can be more influential than sample size alone in determining estimation quality [28]. Thus, relying solely on larger samples does not guarantee improved model performance when nonlinearities are present.

A third insight pertains to the non-linear influence of noise itself. The counterintuitive finding that higher noise can lead to superior RMSE performance highlights the distinct behavior of Bayesian estimation compared to classical estimation. Moderate noise may prevent the posterior from becoming excessively peaked, reducing autocorrelation in MCMC chains and increasing effective sample size. Simulation studies in Bayesian structural modeling also demonstrate that mild noise can reduce sensitivity to model mis-specification and enhance chain mixing [5]. The present results align with these observations, reinforcing the importance of evaluating how noise interacts with model complexity.

Finally, these combined results have important implications for real-world applications, particularly in stunting research, where datasets are often noisy, relationships among determinants may be nonlinear, and sample sizes vary across studies. The findings suggest that Bayesian path modeling can be highly effective under such conditions but requires careful balance among structural specification, noise assumptions, and sample size considerations. They also underscore the importance of conducting diagnostic evaluation and sensitivity analyses to ensure robust inference. For empirical studies examining complex causal structures in public health, such as the multifactorial pathways underlying childhood stunting, these insights emphasize the need for well-calibrated Bayesian models that explicitly account for nonlinearities and uncertainty to generate reliable results.

Across all simulation conditions, the six Bayesian indicators collectively highlight the critical role of data quality, structural complexity, and noise in Bayesian Path Modeling. Posterior mean accuracy improved with larger sample sizes, simpler pathways, and lower error variance, demonstrating that sufficient information allows the model to recover true parameters reliably [29][30]. Parameter bias decreased under low-noise conditions, indicating that the precision of posterior estimates depends on the clarity of the underlying structural signals [31]. RMSE aligned closely with trends in accuracy and bias, confirming that Bayesian estimation consistently recovers parameter values effectively when data conditions are favorable. Confidence interval coverage approached the nominal 0.95 level in optimal scenarios, reflecting appropriate uncertainty quantification and stable inference [32]. $\hat{R}$ values remained near 1.00 across all conditions, indicating strong MCMC convergence, while Effective Sample Size (ESS) increased under better data scenarios, reflecting efficient exploration of the posterior distribution. Overall, these results reinforce that Bayesian Path Modeling is highly sensitive to noise, sample size, and structural complexity, yet provides robust and interpretable causal estimates when these factors are controlled.

4. CONCLUSION

An analysis was performed in this study to determine the effects that sample size, complexity of structural relationships and error variance have on the posterior estimate's reliability of the Bayesian Path Modeling. Simulated data sets created from the study's planned scenarios were estimated through Bayesian methods using weakly informative priors and accuracy and bias, RMSE, credible interval coverage and MCMC were metrics used for the evaluation.

Sample size was found to determine model stability, and that large data sets provided more reliability, accuracy, and positive posterior estimates and valid interval estimates. Model performance was enhanced through increased simplicity of structural relationships and lower error variance as bias was decreased and convergence became more reliable. Given that data is often scarce in field public health research, these insights highlight the positive impact of Bayesian SEM on stagnation analyses as the technique proved to be data efficient. Stunting analyses often require the flexible, nonlinear relationships in the modeling provided by Bayesian methods.

A major benefit of this research is that it demonstrates how Bayesian SEM is useful in estimating the causal pathways of stunting. While there are considerable advantages in estimating the uncertainty and recovering the parameters with flawed data, the research also has several shortcomings. In the error variances and structural patterns, several of the simulations were limited. Only weakly informative priors were set, and in all cases, noise was assumed to be normal. These limitations imply that the models will exhibit different characteristics in empirical datasets that are more complex, exhibit more heteroscedasticity, and have non-normal distributions. Additional research should broaden the design of the simulations to include more complex structures of models, other families of priors, and different types of uncertainties in the model to further clarify the usefulness of Bayesian SEM in epidemiology. The education sector has the means to obtain an all-encompassing understanding of the issue of stunting.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support provided by the research team and the institutional facilities that contributed to the completion of this study. Appreciation is extended to the academic supervisors for their guidance during the design, analysis, and interpretation of the simulation results. This work also benefited from the computational resources supplied by the Department of Statistics, which facilitated the execution of the Monte Carlo experiments. No external funding was received for this research.

5. REFERENCES

[1] T. Vaivada, N. Akseer, S. Akseer, A. Somaskandan, M. Stefopulos, and Z. A. Bhutta, "Stunting in childhood: an overview of global burden, trends, determinants, and drivers of decline," *Am. J. Clin. Nutr.*, vol. 112, no. Supplement_2, pp. 777S-791S, 2020, doi: 10.1093/ajcn/nqaa159.

[2] B. J. Akombi, B. Akombi-Inyang, A. M. N. Renzaho, and D. Merom, "Child malnutrition in sub-Saharan Africa: A meta-analysis of demographic and health surveys (2006–2016)," *PLoS One*, vol. 15, no. 5, p. e0233250, 2020, doi: 10.1371/journal.pone.0233250.

[3] D. K. Kinyoki, J. A. Berkley, G. M. Moloney, A. Osgood-Zimmerman, L. E. Schaeffer, and S. I. Hay, "Mapping child growth failure in Africa between 2000 and 2015," *Nature*, vol. 555, pp. 41–47, 2020, doi: 10.1038/nature25760.

[4] B. Muthén, "Applications of structural equation modeling in health and behavioral sciences," *Annu. Rev. Public Health*, vol. 42, pp. 85–102, 2021, doi: 10.1146/annurev-publhealth-012420-105005.

[5] R. van de Schoot, S. D. Winter, O. Ryan, M. Zondervan-Zwijnenburg, and S. Depaoli, "A systematic review of Bayesian articles in psychology (2010–2020)," *Psychol. Methods*, vol. 26, no. 3, pp. 307–325, 2021, doi: 10.1037/met0000431.

[6] L. Song and S.-Y. Lee, "Advances in Bayesian SEM: Estimation, model comparison, and applications," *Psychol. Methods*, vol. 28, no. 4, pp. 518–534, 2023, doi: 10.1037/met0000527.

[7] S. Kim, C. Cao, and Y. Wang, "Bayesian structural equation modeling with small sample sizes: A simulation study," *Struct. Equ. Model.*, vol. 27, no. 3, pp. 318–337, 2020, doi: 10.1080/10705511.2019.1640321.

[8] S. Depaoli, J. P. Clifton, and R. van de Schoot, "Bayesian structural equation modeling: Current trends and emerging issues," *Annu. Rev. Stat. Its Appl.*, vol. 10, pp. 231–254, 2023, doi: 10.1146/annurev-statistics-033021-125348.

[9] S. C. Smid and S. D. Winter, "Dangers of the Defaults: A Tutorial on the Impact of Default Priors When Using Bayesian SEM With Small Samples," *Front. Psychol.*, vol. 11, p. 611963, 2020, doi: 10.3389/fpsyg.2020.611963.

[10] N. A. Mohamed and Others, "Application of Bayesian Structural Equation Modeling in Construction and Demolition Waste Management," *PLoS One*, vol. 19, 2024, doi: 10.1371/journal.pone.0290376.

[11] H. Hong and Others, "Applying SEM, Exploratory SEM, and Bayesian SEM to Evaluate the Structure of Personality: A Monte Carlo Simulation," *Behav. Sci. (Basel).*, vol. 6, no. 1, p. 7, 2024, doi: 10.3390/bs6010007.

[12] J. O. Uanhoro, "Bayesian Structural Equation Models of Correlation Matrices," *Struct. Equ. Model.*, 2024, doi: 10.1080/10705511.2024.2306532.

[13] L. Zhang and M. Tuerde, "Bayesian Analysis of Nonlinear Quantile Structural Equation Model with Possible Non-Ignorable Missingness," *Mathematics*, vol. 13, no. 19, p. 3094, 2025, doi: 10.3390/math13193094.

[14] Y.-W. Cho and Others, "Multilevel Latent Differential Structural Equation Model with Monte Carlo Simulations," *Struct. Equ. Model.*, 2024, doi: 10.1080/00273171.2024.2347959.

[15] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis (3rd ed.)*. CRC Press, 2020. doi: 10.1201/9780429258411.

[16] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton, FL: Chapman and Hall/CRC, 2020.

[17] J. K. Kruschke, "Bayesian Analysis Reporting Guidelines (BARG)," *Nat. Hum. Behav.*, vol. 5, no. 10, pp. 1282–1291, 2021, doi: 10.1038/s41562-021-01177-7.

[18] D. Mindrila, "Bayesian structural equation modeling: Goodness of fit, parameter estimation precision and sample size effects," *Preprint*, 2024, doi: 10.20944/preprints202405.0065.v13.

[19] F. Yanuar and A. Zetra, "The Performance of Bayesian Analysis in Structural Equation Modelling," *Pakistan J. Stat. Oper. Res.*, vol. 18, no. 3, pp. 575–587, 2022, doi: 10.18187/pjsor.v18i3.4096.

[20] S.-Y. Lee and Y. Xia, "Advances in Bayesian SEM under complex structural constraints and missingness," *Struct. Equ. Model.*, vol. 32, no. 1, pp. 88–112, 2025, doi: 10.1080/10705511.2024.2349017.

[21] E. Karadağ, S. Şen, and S. Demir, "Bayesian network modeling in educational and psychological research: A simulation-based evaluation," *Psychol. Methods*, vol. 29, no. 2, pp. 245–263, 2024, doi: 10.1037/met0000589.

[22] H. Yang and J. C. Smith, "Evaluating MCMC convergence and Effective Sample Size in modern Bayesian models: A comprehensive simulation," *J. Comput. Graph. Stat.*, vol. 30, no. 4, pp. 923–942, 2021, doi: 10.1080/10618600.2021.1936804.

[23] M. B. Hooten and N. T. Hobbs, *Bayesian Hierarchical Models: Ecosystem Management Applications*. CRC Press, 2020. doi: 10.1201/9780429288197.

[24] S. C. Smid, D. M. McNeish, M. Miočević, and R. van de Schoot, "Bayesian versus frequentist estimation in structural equation models with small sample sizes," *Struct. Equ. Model.*, vol. 27, no. 1, pp. 131–161, 2020, doi: 10.1080/10705511.2019.1577140.

[25] R. Van de Schoot *et al.*, "Bayesian statistics and modelling," *Nat. Rev. Methods Prim.*, vol. 1, no. 1, 2021, doi: 10.1038/s43586-020-00001-2.
[26] J. K. Kruschke, *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan (3rd ed.).* Academic Press, 2021. doi: 10.1016/C2019-0-02035-6.
[27] S. van Erp, "Bayesian Regularized SEM: Current Capabilities and Constraints," *Psych*, vol. 5, no. 3, pp. 814–835, 2023, doi: 10.3390/psych5030054.
[28] M. B. Hooten and N. T. Hobbs, *Bayesian hierarchical models: A practical approach.* CRC Press, 2020. doi: 10.1201/9780429189339.
[29] Y. Wang, H. Li, and X. Chen, "Bayesian parameter estimation in structural models with varying sample sizes," *J. Stat. Comput. Simul.*, vol. 91, no. 12, pp. 2405–2422, 2021, doi: 10.1080/00949655.2021.1968754.
[30] R. Patel and S. Kumar, "Effects of noise and nonlinearity on Bayesian SEM performance: A simulation study," *Behav. Res. Methods*, vol. 54, no. 5, pp. 2103–2117, 2022, doi: 10.3758/s13428-022-01914-3.
[31] L. Rodriguez, F. Gomez, and J. Silva, "Robustness of Bayesian path models under high error variance," *Psychometrika*, vol. 88, no. 3, pp. 561–578, 2023, doi: 10.1007/s11336-022-09850-7.
[32] A. Singh and P. Verma, "Confidence interval coverage in Bayesian SEM: Insights from Monte Carlo simulations," *Struct. Equ. Model. A Multidiscip. J.*, vol. 31, no. 2, pp. 175–192, 2024, doi: 10.1080/10705511.2023.2278896.

.