InPrime: Indonesian Journal of Pure and Applied Mathematics
Vol. 6, No. 1 (2024), pp. 63 – 75, doi: 10.15408/inprime.v6i1.38914
p-ISSN 2686-5335, e-ISSN: 2716-2478

N-Level Structural Equation Models (nSEM): The Effect of Sample Size on the Parameter Estimation in Latent Random-Intercept Model

Viarti Eminita[1], Asep Saefuddin[2]*, Kusman Sadik[2], and Utami Dyah Syafitri[2]
[1]Faculty of Educational Science, University of Muhammadiyah Jakarta, Indonesia
[2]Department of Statistics and Data Science, IPB University, Bogor, Indonesi
Email: *asaefuddin@apps.ipb.ac.id

Abstract

Multilevel Structural Equation Modeling (MSEM) is claimed to address hierarchical data structures and latent response variables, but it becomes unstable with an increasing number of levels. N-Level SEM (nSEM) is an SEM framework designed to handle a growing number of levels in the model. The nSEM framework uses the Maximum Likelihood Estimation (MLE) method for parameter estimation, which requires a large sample size and correct model specification. Therefore, it is essential to consider the necessary minimal sample size to ensure accurate and efficient parameter estimation in the nSEM model. This study examined how sample size affects the performance of parameter estimators in nSEM models. We propose a method to evaluate the effect of many environments to estimate the results of factor loadings and environmental variance produced by the model. In addition, we also assess the impact of environment size on the estimation results of factor loadings and individual variance. The results were then applied to actual data on student mathematics learning motivation in Depok. The findings show that neither the number of environments nor the size of the environment affects the performance of fixed parameter estimation in the nSEM model. nSEM indicates excellent performance in estimating environmental variance at level 2 when the number of environments increases. Conversely, increasing the size of the environment worsens the performance of estimating individual variance parameters. Overall, the nSEM framework for the latent random-intercept (LatenRI) model performs well with increasing sample sizes. The application data on LatenRI models show almost similar estimation results.
Keywords: hierarchical data; latent random intercept model; multilevel structural equation modeling; n-level structural equation modeling.

Abstrak

Multilevel Structural Equation Modeling (MSEM) *diklaim dapat mengatasi struktur data hierarki dan variabel respons laten, namun menjadi tidak stabil dengan bertambahnya jumlah level.* N-Level SEM *(nSEM) adalah kerangka kerja SEM yang dirancang untuk menangani semakin banyak level dalam model. Kerangka kerja nSEM menggunakan metode* Maximum Likelihood Estimation (MLE) *untuk estimasi parameter, yang memerlukan ukuran sampel yang besar dan spesifikasi model yang benar. Oleh karena itu, penting untuk mempertimbangkan ukuran sampel minimal yang diperlukan untuk memastikan estimasi parameter yang akurat dan efisien dalam model nSEM. Studi ini menguji bagaimana ukuran sampel mempengaruhi kinerja penduga parameter dalam model nSEM. Kami mengusulkan metode untuk mengevaluasi pengaruh banyak lingkungan dalam memperkirakan hasil* factor loadings *dan varians lingkungan yang dihasilkan oleh model. Selain itu, kami juga menilai dampak ukuran lingkungan terhadap hasil estimasi* factor loadings *dan varians individu. Hasilnya kemudian diterapkan pada data aktual motivasi belajar matematika siswa di Depok. Hasil menunjukkan bahwa baik jumlah lingkungan maupun ukuran lingkungan tidak mempengaruhi kinerja estimasi parameter tetap pada model nSEM. nSEM menunjukkan kinerja yang sangat baik dalam memperkirakan varians lingkungan pada level 2 ketika jumlah lingkungan meningkat. Sebaliknya, peningkatan ukuran lingkungan akan memperburuk kinerja pendugaan parameter varians individu. Secara keseluruhan, kerangka*

* Corresponding author
Submitted May 20th, 2024, Revised May 27th, 2024,
Accepted for publication May 29th, 2024, Published Online May 31st, 2024

nSEM untuk model intersepsi acak laten (LatenRI) bekerja dengan baik dengan meningkatnya ukuran sampel. Data penerapan model LatenRI menunjukkan hasil estimasi yang hampir serupa.
Kata Kunci: *data hirarki; model intersep acak laten; model persamaan structural multilevel; model persamaan structural n-level.*

2020MSC: 62D99

1. INTRODUCTION

Researchers often examine the relationship between a response variable and several explanatory variables. The most straightforward way to see the relationship between the two is through a multiple regression analysis with a linear model approach, with only one continuous response variable. However, this analysis cannot handle the complex structure of the response variable [1]. For example, in educational studies, the response variable is latent or cannot be measured directly and has a hierarchical structure. If the data has a hierarchical structure, the random effect on the response variable can be overcome with a linear mixed model. However, if the response variable cannot be observed directly, the linear mixed model and Structural Equation Model (SEM) cannot yet handle both response variable structures. If one of these two response variable structures is ignored, then not only are the parameter estimates and standard errors biased, but essential information about the observed phenomenon can also be lost. In addition, [2] also revealed that severe inferential errors can occur due to complex data analysis if it is assumed that the data was obtained based on a simple random sampling scheme.

The multilevel model (MLM) or Hierarchical Linear Modeling (HLM) is proposed to analyze hierarchical data, where observations at the lowest level (e.g., students) are nested within units at another level (e.g., classes). MLM addresses aggregation bias, standard error estimation errors, and heterogeneity in least squares regression [3]. MLM allows researchers to separate individual and environmental influences [4], [5], understand intergroup diversity [6], [7], and consider hierarchical data structures [8]. The development of multilevel models continues to follow developments in methodological work that result in complex data structures. Multilevel SEM (MSEM) integrated the multilevel model with SEM, which was developed by Muthen [6], [9]. They proposed hierarchical constraints that link the parameters of the SEM model at the lowest and highest levels. The MSEM can become very complex, especially if it involves many hierarchy levels, variables, or relationships between variables. Some software for MSEM analysis has limitations on the number of levels and how to connect those levels. [10] overcame the limitations of the MSEM model with the TEM framework. This framework is claimed to be flexible with a large number of levels and complex relationships between variables [11], [12], [13]. The Latent Variable Random-Intercept Model (LatenRI) is one of the MSEM models in the nSEM framework. This model includes the influence of environmental random intercepts at level 2 that affect individual latent factors at level 1.

LatenRI is a model with a reasonable complexity that affects the determination of the minimum sample size required. LatenRI in the nSEM framework uses the Maximum Likelihood Estimation (MLE) method in parameter estimation, so it needs a large sample size and the correct model specification [14], [15]. This study examined how sample size affects the performance of parameter estimators in nSEM models. The model built in this study has a split plot variance structure with a complexity that is still simple; namely, the model only includes random intercepts. The sample size at the lowest and highest level provides different performance. [16] revealed that the consideration in choosing the sample size is the type and complexity of the model. The determination of the sample size of the multilevel model needs to pay attention to the total sample size for each level [17], [18]. The results of this study provide recommendations for determining the sample size that optimizes the performance of the nSEM model.

2. METHODS

2.1. NSEM: Latent Variable Random-Intercept (LatentRI)

NSEM is an alternative framework combining the SEM approach for cluster and longitudinal data [10]. TEM is also a general approach to formulate and estimate complex data relationships, such as model specifications that include multiple levels, cross-classified [12], or multiple memberships [19], coupled with complex nested structures such as partially nested and multivariate outcomes [20]. The equations used in nSEM to represent the model are pretty complex, so modeling uses superscripts and subscripts [10]. Superscripts indicate each variable and parameter level, while subscripts are, as usual, to reflect the variable with provisions. The modeling framework allows observed variables and latent variables at any level to be influenced by observed variables and latent variables at that level and at higher levels.

The LatenRI model is a 2-level hierarchical model that only includes random intercepts at level 2. The multilevel model's random intercepts are considered latent variables within the SEM framework, namely latent random intercept variables or single latent variables at level 2. This model addresses the problem of non-independence in individual observations, as individuals can be influenced by their environment or group. Figure 1 shows the path diagram of the latent variable random-intercept model (LatenRI). This model includes individual factors at level 1, measured by p observed variables, and influenced by environmental factors at level 2.

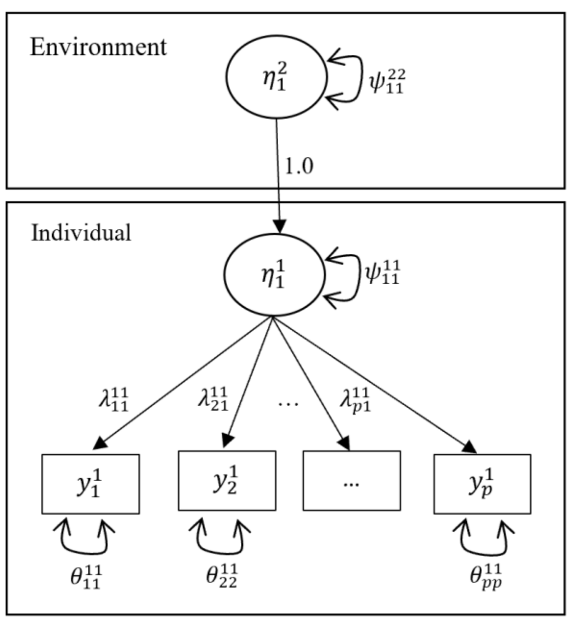

Figure 1. Path diagram of LatentRI model

The scalar model from Figure 1 is expressed in equation (1). Consider y_{pij}^{1} is *p*-th observed variable ($p = 1,2, \dots, P$) measured at level 1 for i-th individual ($i = 1,2, \dots, n$), nested within j-th environment at level 2 ($i = 1,2, \dots, m$):

Level 1 : $y_{pij}^{1} = v_{p}^{1} + \lambda_{p1}^{1,1}\eta_{1ij}^{1} + \varepsilon_{pij}^{1}$

Level 2 → Level 1 : $\eta_{1ij}^{1} = \beta_{11}^{1,2}\eta_{1j}^{2} + \xi_{1ij}^{1} = 1 \cdot \eta_{1j}^{2} + \xi_{1ij}^{1}.$

The simplified of those both models is:

$$y_{pij}^{1} = v_{p}^{1} + \lambda_{p1}^{1,1}\eta_{1j}^{2} + \lambda_{p1}^{1,1}\xi_{1ij}^{1} + \varepsilon_{pij}^{1}, \tag{1}$$

where v_{p}^{1} is intercept at level 1, and $\lambda_{p1}^{1,1}$ is the loading factor to connected y_{pij}^{1} and latent variable η_{1ij}^{1} for the i-th individual nested within jth environment η_{1ij}^{1}. η_{1j}^{2} represents the random intercept for the j-th environment at level 2 with $\eta_{1j}^{2} \sim N\left(0, \psi_{1,1}^{2,2}\right)$. The level 1 structural error ξ_{1ij}^{1} for the individual latent factor is the deviation between individuals in each environment with the assumption $\xi_{1ij}^{1} \sim N\left(0, \psi_{1,1}^{1,1}\right)$. ε_{pij}^{1} is the measurement error at level 1 ($\varepsilon_{pij}^{1} \sim N(0, \theta^{1,1})$). Equation (1) can also be expressed in matrix notation for all i, j and p, namely:

$$\mathbf{y}^{1} = \mathbf{1}_{mn} \otimes \boldsymbol{\nu}^{1} + (\mathbf{I}_{m} \otimes \mathbf{1}_{n} \otimes \boldsymbol{\Lambda}^{1,1})\boldsymbol{\eta}_{1}^{2} + (\mathbf{I}_{mn} \otimes \boldsymbol{\Lambda}^{1,1})\boldsymbol{\xi}^{1} + \boldsymbol{\varepsilon}^{1}, \tag{2}$$

where $\boldsymbol{\Lambda}^{1,1}$ is a vector of the loading factor, and $\boldsymbol{\eta}_{1}^{2}$ is a vector of a latent variable at level 2. $\boldsymbol{\xi}^{1}$ is a vector of level 1 structural error, and $\boldsymbol{\varepsilon}^{1}$ is a vector of level 1 measurement error. $\gamma_{1,1}^{1,2}$ is a fixed regression coefficient with a value of 1. The Interclass Class Correlation (ICC) in the LatenIA model indicated the proportion of variance in the individual latent factor at level 1 that is explained by the random intercept of the single latent factor at level 2, namely ([21]):

$$ICC = \frac{\psi_{1,1}^{2,2}}{\psi_{1,1}^{2,2} + \psi_{1,1}^{1,1}}. \tag{3}$$

Equation (3) also indicates that individual variance in the underlying latent factor causes variance in each observation.

Model fit indices are crucial in model selection, especially when determining the appropriate variance structure. The choice of the variance-covariance structure is vital in psycholinguistics, genetics, and medical research. Various studies highlight the importance of choosing the most suitable variance structure for the model under analysis [22], [23], [24]. Deviance [25], Akaike's Information Criterion, and Schwartz's Bayesian Information Criterion (BIC) are commonly used indices. Deviance is expressed as -2LL or -2 Log Likelihood, while AIC is [26]:

$$AIC = -2LL + 2p, \tag{4}$$

where p is the number of parameters estimated by Maximum Likelihood Estimation (MLE). Schwartz's BIC is calculated using [23]:

$$BIC = -2LL + p\ln(N), \tag{5}$$

where N is the sample size at level 1. AIC and BIC give a penalty on the number of covariance parameters estimated. These three criteria are used to select a model with a better fit, the model with the smallest value, even close to zero.

2.2. Simulation Data

In this study, we generate simulation data from the LatenRI model, which is a model that only includes the random effect of the latent intercept at level 2. A 2-level model was used to examine how stable the model is against the sample size and the correlation between the observed variables. The response variables in this model were six observed variables. The parameter values are determined based on actual data, considering the correlation between the observed variables generated. The following is the simulation data-generating procedure that was carried out:

1. Specify the nSEM model on the data to be generated, namely with the *p*-th observed response variable ($p = 1,2, \dots, 6$) on the i-th individual ($i = 1,2, \dots, n$) for each *j*-th environment ($j = 1,2, \dots, m$). The nSEM model specification is expressed in graphical and scalar forms in equations (1) or (2).
2. Determine the values of the parameters to be estimated, namely:
   A. Level 1 Single factor: Level 1 factor loadings and intercepts

$$\boldsymbol{\Lambda}^{1,1} = \begin{bmatrix} \lambda_{1,1}^{1,1} \\ \lambda_{2,1}^{1,1} \\ \lambda_{3,1}^{1,1} \\ \lambda_{4,1}^{1,1} \\ \lambda_{5,1}^{1,1} \\ \lambda_{6,1}^{1,1} \end{bmatrix} = \begin{bmatrix} 1 \\ 1,5 \\ 1,0 \\ 1,0 \\ 0,8 \\ 0,7 \end{bmatrix} \text{ and } \mathbf{v}^{1} = \begin{bmatrix} 50 \\ 50 \\ 50 \\ 50 \\ 50 \\ 50 \end{bmatrix},$$

   B. The structural error variance of the i(j)-th latent factor at level 1: $\boldsymbol{\xi}_{i(j)}^{1} \sim N(0, \psi_{1,1}^{1,1})$, with $\boldsymbol{\Psi}^{1,1} = \psi_{1,1}^{1,1} = 25$,

   C. The variance-covariance matrix of the observed residuals at level 1: $\boldsymbol{\varepsilon}^{1} \sim N(\mathbf{0}, \boldsymbol{\Theta}^{1,1})$

$$\boldsymbol{\Theta}^{1,1} = \begin{bmatrix} \theta_{1,1}^{1,1} & \theta_{1,2}^{1,1} & \theta_{1,3}^{1,1} & \theta_{1,4}^{1,1} & \theta_{1,5}^{1,1} & \theta_{1,6}^{1,1} \\ \theta_{2,1}^{1,1} & \theta_{2,2}^{1,1} & \theta_{2,3}^{1,1} & \theta_{2,4}^{1,1} & \theta_{2,5}^{1,1} & \theta_{2,6}^{1,1} \\ \theta_{3,1}^{1,1} & \theta_{3,2}^{1,1} & \theta_{3,3}^{1,1} & \theta_{3,4}^{1,1} & \theta_{3,5}^{1,1} & \theta_{3,6}^{1,1} \\ \theta_{4,1}^{1,1} & \theta_{4,2}^{1,1} & \theta_{4,3}^{1,1} & \theta_{4,4}^{1,1} & \theta_{4,5}^{1,1} & \theta_{4,6}^{1,1} \\ \theta_{5,1}^{1,1} & \theta_{5,2}^{1,1} & \theta_{5,3}^{1,1} & \theta_{5,4}^{1,1} & \theta_{5,5}^{1,1} & \theta_{5,6}^{1,1} \\ \theta_{6,1}^{1,1} & \theta_{6,2}^{1,1} & \theta_{6,3}^{1,1} & \theta_{6,4}^{1,1} & \theta_{6,5}^{1,1} & \theta_{6,6}^{1,1} \end{bmatrix} = \begin{bmatrix} 20 & 5 & 5 & 5 & 5 & 5 \\ 5 & 20 & 5 & 5 & 5 & 5 \\ 5 & 5 & 20 & 5 & 5 & 5 \\ 5 & 5 & 5 & 20 & 5 & 5 \\ 5 & 5 & 5 & 5 & 20 & 5 \\ 5 & 5 & 5 & 5 & 5 & 20 \end{bmatrix},$$

   D. Across-Level Model (level 2 → level 1): Latent factor regression coefficients

$$\boldsymbol{\Gamma}^{1,2} = \left[\gamma_{1,1}^{1,2}\right] = [1],$$

E. The variance of the latent environmental factors at level 2: $\boldsymbol{\Psi}^{2,2} = \psi_{1,1}^{2,2} = 30$,

3. Generate m times data of a single latent factor at level 2: $\boldsymbol{\eta}_1^2 \sim N(0, \psi_{1,1}^{2,2})$,
4. Generate $m \times n$ times data with a latent factor error $\boldsymbol{\xi}^1 \sim N(0, \psi_{1,1}^{1,1})$ at level 1,
5. Generate $m \times n$ times data with measurement errors at level 1, $\boldsymbol{\varepsilon}^1 \sim \mathbf{N}(0, \boldsymbol{\Theta}^{1,1})$ for each p variables, $p = 1,2, \dots, 6$,
6. Substitute the result from steps 2 to 5 into the observed response variable y_{pij}^1 in equation (1).

2.3. Simulation Design

Simulation design to study the LatenRI Model on Complex Data. The simulation design aimed to investigate the LatenRI model on complex data, where the model includes random components from both individuals and environments. The specific objectives of this simulation were:

1. To determine whether the number of environments (m) affects the estimation results of factor loadings and environmental variance produced by the model.
2. To determine whether the environment size (n) affects the estimation results of factor loadings and individual variance.

We propose a method as follows:

1. Generate sample data, namely steps 3-6 (sub-chapter 2.2) on both models with m environments and n sample sizes per environment for a 300-data set.
2. Generate simulation data using several variations to investigate the effect of increasing sample size. At this stage, there are 4 combinations of $m = (10, 25)$ and $n = (30, 100)$.
3. Perform an nSEM analysis that includes random components of each simulated data set, resulting in 300 sets of parameter estimates.
4. Evaluate the performance of the nSEM model to determine the goodness of the model's performance. We evaluate model using the bias and Mean Square Error (MSE) as follows:

$$Bias = E(\theta - \hat{\theta}) = \frac{1}{S}\sum_{s=1}^{300}(\theta - \hat{\theta}_s), \quad (4)$$

and

$$MSE = E(\theta - \hat{\theta})^2 = \frac{1}{S}\sum_{s=1}^{300}(\theta - \hat{\theta}_s)^2. \quad (5)$$

2.4. The Actual Data

We use the actual data to illustrate the LatentRI model, explicitly focusing on the motivation of mathematics students in Depok. The data had previously been analyzed by [27] to investigate the random effects of teachers on students' mathematics learning motivation. The data analysis employed the LatentRI model, which included the influence of random intercepts only, and LatentRI, which encompassed both random intercepts and coefficients. The study involved three key variables: the teachers' ability factor (a single endogenous variable), teacher competence (an exogenous variable at the teacher level), and student motivation (another endogenous variable). Two questionnaires were utilized as research instruments to evaluate teachers' competence and students' learning motivation. The study employed a stratified random sampling technique in three specific districts in Depok, namely Sawangan, Bojong Sari, and Limo, encompassing 11, 7, and 6 schools, respectively. The research utilized this sampling method to ensure that the selected sample was representative of the population. Out of the 24 schools selected, 32 math teachers and 768 students participated as respondents.

3. RESULTS

Model identification in nSEM is generally similar to SEM; that is, it is based on applicable theory. nSEM modeling also requires initial values that affect the convergence of the resulting model. If the initial values we input are far from the actual parameters, the resulting parameter estimates may not converge. Thus, the determination of initial values is essential in this case. In addition, the sample size must also be considered to determine the initial value. So, it is essential to study the appropriate sample size to estimate the parameters of the nSEM model.

Sample size can affect the goodness of the resulting model. Table 1 presents the goodness of fit for the nSEM model with various sample sizes. Based on this table, the deviance increases as the sample size increases. This is because the deviation calculation depends on the sample size.

Table 1. Deviance of LatenRI model for each sample size

Sample Size	Deviance
300	10803.335
750	27041.469
1000	36006.145
2500	90049.843

3.1. Fixed Effect Parameter

The bias and MSE for the nSEM fixed parameter estimator are presented in Figure 6. Both bias and MSE for all fixed parameters were small or close to 0. Most of the biases produce negative values (underestimates). The range of bias in the parameter estimates produced was quite large for small numbers of observations (nj = 30) in both numbers of environments. However, the differences between the four sample sizes were insignificant and only ranged from -0.0056 to 0.0008. In addition, there was a decrease in MSE as the sample size increased for each parameter estimator. At each environment size, it showed good performance as the number of environments increased. This finding indicated that nSEM was very good at estimating the fixed parameters of the model.

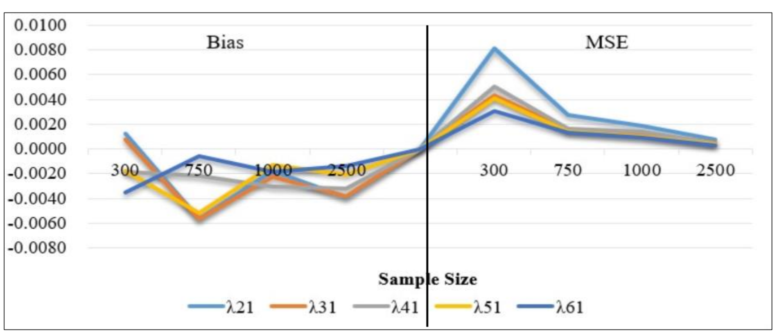

Figure 2. Graph of bias and MSE for the nSEM fixed estimator

3.2. Random Effect Parameter

The performance of random parameter estimation in the nSEM model was studied based on bias and MSE values. Both values are presented in Figure 3. Twenty-one random measurement error parameters and two random structural error parameters were studied in this section. The comparison of bias and MSE results for each random parameter estimate showed different results than the previous fixed parameter estimates. The magnitude of the bias produced for all measurement error parameters was negative at environment size $n = 100$, while for $n = 30$, the bias was positive. The trend pattern of bias and MSE is illustrated in Figure 3(a) - (d).

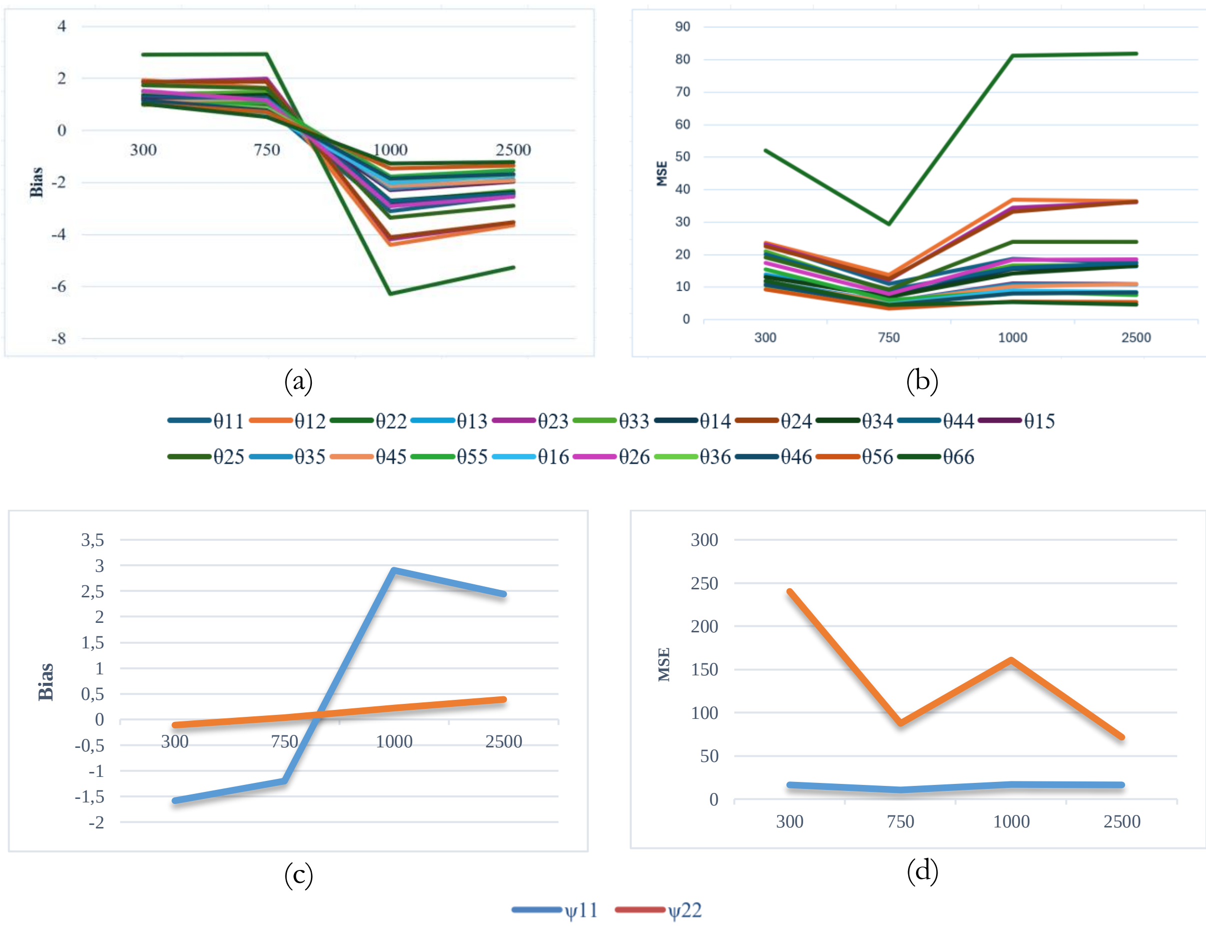

Figure 3. Bias and MSE for the nSEM fixed estimator (a) Bias of measurement error, (b) MSE of measurement error, (c) Bias of structural error, and (d) MSE of structural error

Figure 3(a)-(b) showed that the variance-covariance matrix of measurement errors ($\boldsymbol{\Theta}^{1,1}$, $\theta pq = \theta_{p,q}^{1,1}$, $p, q = 1,2, \dots, 6$ for the observed variable produces a larger bias and MSE compared to other variance-covariance matrices of measurement errors. In addition, the MSE for 750 sample size with $n = 25$ and $m = 30$ had the smallest value among other sample sizes, while the bias was relatively the same. More negligible bias and MSE indicated better estimation performance. Therefore, this sample size could be said to be quite optimal in estimating the variance-covariance matrix of measurement errors. In addition, the larger the sample, the better nSEM estimates measurement errors.

The structural error variance of the second-level latent variable was also essential to examine. Figure 3(c) showed that the bias of these two parameters $\left(\psi 11 = \psi_{1,1}^{1,1}, \psi 22 = \psi_{1,1}^{2,2}\right)$ was relatively small, ranging from -1.5837 to 2.9066 for all sample sizes. The structural variance parameter at level 2 showed a more stable bias pattern than the structural variance at level 1. Conversely, the structural variance parameter at level 1 showed a more stable MSE graph than the structural variance at level 2 (Figure 3(d)). The structural variance at level 2 was quite large, ranging from 16.5418 to 240.2847. A sizeable individual sample size ($n = 100$) produced a superior MSE compared to $n = 30$.

3.3. Analysis of the Actual Data

We apply the LatenRI model to the actual data about student mathematics learning motivation, and the analysis is shown in Table 2. This table show the random effect estimator for LatenRI model to actual data with sample size of 768 and $m = 32$ (unbalanced environment sizes). Based on the simulation results, the sample size for actual data was still entirely accurate in estimating both fixed and random parameters. Both models produced relatively similar and significant analysis results for fixed and random effect parameters, except for the random coefficient estimator of teacher competence (95% confidence interval includes 0) [27]. Furthermore, the LatenRI model produced a smaller goodness-of-fit measure than the LatenRI with random coefficients. An ICC of 4.77% was quite good in measuring Education [10].

Table 2. Random effect estimator for LatenRI model

Parameter	**LatentRI Model**		**LatenRI Model with random coefficient**	
	Estimate	**Confidence interval (CI) 95%**	**Estimate**	**Confidence interval (CI) 95%**
Fixed Effects				
Student				
Relevance $\left(\lambda_{2;1}^{1;1}\right)$	1.360*	[1.218; 1.523]	1.358*	[1.216; 1.521]
Confidence $\left(\lambda_{3;1}^{1;1}\right)$	1.666*	[1.483; 1.875]	1.669*	[1.486; 1.879]
Satisfaction $\left(\lambda_{4;1}^{1;1}\right)$	0.743*	[0.588; 0.908]	0.744*	[0.589; 0.909]
Teacher				
Personality $\left(\lambda_{2;1}^{2;2}\right)$	-	-	0.477*	[0.120; 0.861]
Social $\left(\lambda_{2;1}^{2;2}\right)$	-	-	1.003*	[0.649; 1.417]
Professional $\left(\lambda_{2;1}^{2;2}\right)$	-	-	1.132*	[0.8165; 1.5146]
Competence (β_{12}^{22})	-	-	-0.070	[-0.0697; 0.0296]
Random Effects				
Motivation (ψ_{55}^{11})	0.0598	[0.049; 0.072]	00597	[0.049; 0.072]
Teacher (ψ_{11}^{22})	0.0030	[0.001; 0.007]	0.0027	[0.001; 0.007]
Teacher comp. (ψ_{22}^{22})	-	-	0.0943	[0.052; 0.170]

Table 2. (continued)

Goodness of fit		
Latent-ICC	4.77%	
Deviance	2985.744	3017.808
AIC	3011.744	3069.808
BIC	3090.135	3227.652

4. DISCUSSION

The SEM research focuses on fixed effect parameters (factor loadings) to see how observed variables contribute to building latent factors. The contribution of observed variables also reflects the measurement method used to generate observed data [28]. The chosen sample size does not explicitly affect the performance of nSEM factor loadings. This meant that the fixed parameter estimators produced by nSEM were accurate and efficient for various sample size variations. However, observed variables with factor loadings greater than 1.0 tend to perform poorly in estimating the variance-covariance matrix of measurement errors at the individual level.

The bias performance was relatively small, as shown in the estimation of random parameters of the model at both the lowest and highest levels. However, MSE was quite strict in assessing the accuracy of the model. This was seen in the poor performance of nSEM under conditions of small sample sizes at both levels. The TEM parameter estimation process becomes more sensitive if the sample size is larger. The larger the number of environments, the greater the computing power required. Setting the initial value in nSEM, further away from the actual parameter, prevents the model parameter estimation from converging. Therefore, nSEM computing still needs to be developed, one of which is determining the initial value in its modeling.

The sample size in LatenRI was divided into two categories: the number of environments and the environment size. The larger the number of environments, the more accurate it was in distinguishing between individuals and environments. However, in many environments, most measurement error parameter estimators perform poorly as the size of the environment increases. In addition, the size of the environment gave different results on the model's performance in distinguishing between individuals and environments. The larger the size of the environment, the more accurate it was in determining the environment, but on the contrary, it was less accurate in distinguishing individuals.

LatenRI was used to distinguish groups accurately. The limitation of this study was that the number of environments evaluated was moderate. However, these findings provided enough information that a sufficiently large number of environments, or 25, gives more accurate results in distinguishing groups. This was because the size of the environment has a more significant influence on the performance of the nSEM model. The nSEM model with a large environment size showed better performance. This was also revealed by [6], who showed that the effect of sample size at the individual level was generally greater than the effect of sample size at the environmental level. This meant that increasing the individual sample size had a greater impact on the accuracy of estimation, detection power, and generalization of results than increasing the group sample size. With a larger sample size, nSEM could better detect smaller effects with higher precision. [29] revealed that the main problem with between-group models is producing inaccurate and inefficient parameter estimates, which occurs when the number of environments is small (<50) while the ICC is low. This problem was overcome at least with the number of environments 100. [30] revealed that a relatively simple MSEM, at least 60 environments were needed to detect structural influences at the highest level.

The simulation results, which emphasized the importance of sufficient sample sizes and environment numbers for accurate model performance, are supported by applying the LatenRI model to student mathematics learning motivation data. In this real-world scenario, a sample size of 768 with 32 environments (although unbalanced) proved accurate in estimating both fixed and random parameters. It aligns with the simulation findings that larger sample sizes generally lead to better parameter estimation accuracy. Applying the LatenRI model to actual student data reinforces the findings from the simulation studies and demonstrates its practical utility in educational research. The results highlight the importance of considering both sample size and environmental numbers when applying LatenRI models to real-world data and suggest that simpler models might sometimes be preferable.

5. CONCLUSIONS

The number of environments and the size of the environment did not affect the performance of fixed parameter estimation in the nSEM model because the bias and MSE of the fixed parameter estimator were close to 0. However, if the factor loading was large or > 1.0, the model performance deteriorated in estimating the variance-covariance matrix of measurement errors at the lowest level. In estimating environmental variance, nSEM showed excellent performance when the number of environments grew. Conversely, increasing the size of the environment made the performance of estimating individual variance parameters worse.

The study of MSEM in the nSEM framework still needs to be evaluated for a large number of environments, at least 50, while the environment size was more than 100. The nSEM framework for simple models (LatenRI) performs well when increasing sample sizes. For further research, we suggest analyzing how nSEM performs for simple and complex models.

REFERENCES

[1] M. D. M. Rueda, B. Cobo, and A. Arcos, "Regression Models in Complex Survey Sampling for Sensitive Quantitative Variables," *Mathematics*, vol. 9, no. 6, pp. 1–13, Mar. 2021, doi: 10.3390/math9060609.

[2] D. Holt, A. J. Scott, and P. D. Ewings, "Chi-Squared Tests with Survey Data," *Journal of the Royal Statistical Society Series A*, vol. 143, no. 3, pp. 303–320, 1980.

[3] K. J. Preacher, M. J. Zyphur, and Z. Zhang, "Multilevel Structural Equation Models for Assessing Moderation Within and Across Levels of Analysis," *Psychol Methods*, vol. 21, no. 2, pp. 189–205, 2016, doi: 10.1037/met0000052.supp.

[4] R. D. Goddard, M. Tschannen-Moran, and W. K. Hoy, "A Multilevel Examination of the Distribution and Effects of Teacher Trust in Students and Parents in Urban Elementary Schools," *Elem Sch J*, vol. 102, no. 1, pp. 3–17, 2001.

[5] S. De Iaco and S. Maggio, "Using Multilevel Models to Evaluate the Attitude of Separate Waste Collection in Young People," *Metron*, vol. 80, pp. 77–95, Apr. 2022, doi: 10.1007/s40300-020-00194-2.

[6] B. O. Muthén, "Mean and Covariance Structure Analysis of Hierarchical Data," in *The Psychometric Society Meeting*, in UCLA Statistics. Princeton: UCLA, 1990, pp. 1–57.

[7] I. Tsaousis, G. D. Sideridis, and K. Al-Harbi, "Examining Differences in Within-and Between-Person Simple Structures of an Engineering Qualification Test Using Multilevel MIMIC Structural Equation Modeling," *Front Appl Math Stat*, vol. 4, no. 3, pp. 1–14, 2018, doi: 10.3389/fams.2018.00003.

[8] A. H. Leyland and P. P. Groenewegen, "Multilevel Data Structures," in *Multilevel Modelling for Public Health and Health Services Research*, Springer International Publishing, 2020, pp. 49–68. doi: 10.1007/978-3-030-34801-4_4.

[9] B. O. Muthen, "Multilevel Covariance Structure Analysis," *Sociol Methods Res*, vol. 22, no. 3, pp. 376–398, 1994, doi: 10.1177/0049124194022003006.

[10] P. Mehta, "n-Level Structural Equation Modeling," in *Applied Quantitative Analysis in Education and the Social Sciences*, New York: Routledge, 2013, pp. 329–361.

[11] L. Branum-Martin, "Multilevel Modeling: Practical Examples to Illustrate a Special Case of SEM," in *Applied Quantitative Analysis in Education and the Social Sciences*, 2013, pp. 95–124.

[12] O. Kwok, M. H. Lai, F. Tong, and R. Lara-alecio, "Analyzing Complex Longitudinal Data in Educational Research : A Demonstration With Project English Language and Literacy Acquisition (ELLA) Data Using xxM," *Frontiers in Psycology*, vol. 9, no. June, pp. 1–12, 2018, doi: 10.3389/fpsyg.2018.00790.

[13] E. Ryu, "Multiple Group Analysis in Multilevel Structural Equation Model Across Level 1 Groups," *Multivariate Behav Res*, vol. 50, no. 3, pp. 300–315, 2015, doi: 10.1080/00273171.2014.1003769.

[14] I. Devlieger and Y. Rosseel, "Multilevel Factor Score Regression," *Multivariate Behavioral Research*, vol. 55, no. 4. Routledge, pp. 600–624, Jul. 03, 2019. doi: 10.1080/00273171.2019.1661817.

[15] B. Kelcey, K. Cox, and N. Dong, "Croon's Bias-Corrected Factor Score Path Analysis for Small-to Moderate-Sample Multilevel Structural Equation Models," *Organ Res Methods*, vol. 24, no. 1, pp. 55–77, Jan. 2021, doi: 10.1177/1094428119879758.

[16] A. Sagan, "Sample Size in Multilevel Structural Equation Modeling – the Monte Carlo Approach," *Econometrics*, vol. 23, no. 4, pp. 63–79, 2019, doi: 10.15611/eada.2019.4.05.

[17] T. A. B. Snijders, "Power and Sample Size in Multilevel Modeling," *Encyclopedia of Statistics in Behavioral Science*, vol. 3, pp. 1570–1573, 2005, [Online]. Available: https://www.researchgate.net/publication/228079507

[18] M. T. Barendse and Y. Rosseel, "Multilevel SEM with Random Slopes in Discrete Data Using the Pairwise Maximum Likelihood," *British Journal of Mathematical and Statistical Psychology*, vol. 76, no. 2, pp. 327–352, May 2023, doi: 10.1111/bmsp.12294.

[19] E. Ryu, "Model Fit Evaluation in Multilevel Structural Equation Models," *Front Psychol*, vol. 5, no. 81, pp. 1–9, 2014, doi: 10.3389/fpsyg.2014.00081.

[20] Y. Petscher and C. Schatschneider, "Using n-Level Structural Equation Models for Causal Modeling in Fully Nested, Partially Nested, and Cross-Classified Randomized Controlled Trials," *Educ Psychol Meas*, vol. 79, no. 6, pp. 1075–1102, 2019, doi: 10.1177/0013164419840071.

[21] H. Y. Hsu, J. H. Lin, O. M. Kwok, S. Acosta, and V. Willson, "The Impact of Intraclass Correlation on the Effectiveness of Level-Specific Fit Indices in Multilevel Structural Equation

Modeling: A Monte Carlo Study," *Educ Psychol Meas*, vol. 77, no. 1, pp. 5–31, Jan. 2017, doi: 10.1177/0013164416642823.
[22] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal," *J Mem Lang*, vol. 68, no. 3, pp. 255–278, 2013.
[23] H. Matuschek, R. Kliegl, S. Vasishth, H. Baayen, and D. Bates, "Balancing Type I Error and Power in Linear Mixed Models," *J Mem Lang*, vol. 94, pp. 305–315, Jun. 2017, doi: 10.1016/j.jml.2017.01.001.
[24] E. S. W. Ng, K. Diaz-Ordaz, R. Grieve, R. M. Nixon, S. G. Thompson, and J. R. Carpenter, "Multilevel Models for Cost-Effectiveness Analyses that Use Cluster Randomised Trial Data: An Approach to Model Choice," *Stat Methods Med Res*, vol. 25, no. 5, pp. 2036–2052, Oct. 2016, doi: 10.1177/0962280213511719.
[25] J. N. Pritikin, M. D. Hunter, T. von Oertzen, T. R. Brick, and S. M. Boker, "Many-Level Multilevel Structural Equation Modeling: An Efficient Evaluation Strategy," *Structural Equation Modeling*, vol. 24, no. 5, pp. 684–698, Sep. 2017, doi: 10.1080/10705511.2017.1293542.
[26] M. J. Brewer, A. Butler, and S. L. Cooksley, "The Relative Performance of AIC, AICC and BIC in the Presence of Unobserved Heterogeneity," *Methods Ecol Evol*, vol. 7, no. 6, pp. 679–692, Jun. 2016, doi: 10.1111/2041-210X.12541.
[27] V. Eminita, A. Saefuddin, K. Sadik, and U. D. Syafitri, "Analyzing Multilevel Model of Educational Data: Teachers' Ability Effect on Students' Mathematical Learning Motivation," *Journal on Mathematics Education*, vol. 15, no. 2, pp. 431–450, 2024, doi: 10.22342/jme.v13i1.xxxx.
[28] T. D. Little, *Longitudinal Structural Equation Modeling*. New York: The Gilford Press, 2013.
[29] J. J. Hox and C. J. M. Maas, "The Accuracy of Multilevel Structural Equation Modeling with Pseudobalanced Groups and Small Samples," *Structural Equation Modeling*, vol. 8, no. 2, pp. 157–174, 2001.
[30] B. Meuleman and J. Billiet, "A Monte Carlo Sample Size Study: How Many Countries are Needed for Accurate Multilevel SEM?," *Surv Res Methods*, vol. 3, no. 1, pp. 45–58, 2009, [Online]. Available: http://www.surveymethods.org