Variance
Journal of Statistics and Its Applications
P-ISSN: 2685-8738 E-ISSN: 2685-872X

Variance: Journal of Statistics and Its Applications
Volume 7 Issue 2 | October 2025
10.30598/variancevol7iss2page167-174

STRUCTURAL EQUATION MODELING-GENERALIZED STRUCTURED COMPONENT ANALYSIS TO ANALYZING THE STRUCTURE OF POVERTY IN INDONESIA IN 2022

Nur Amalia Marukai[1], Djihad Wungguli[2], La Ode Nashar[3], Salmun K. Nasib[4], Asriadi[5], Siti Nurmardia Abdussamad[6]

[1, 2, 3, 4, 5, 6] Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Gorontalo
Jl. Prof. Dr. Ing. BJ Habibie, Tilongkabila, Bone Bolango, 96119, Gorontalo, Indonesia,

Corresponding Author's Email: *amalia.marukai15@gmail.com*

Abstract: *Structural Equation Modeling - Generalized Structured Component Analysis (SEM-GSCA) is a component-based method suitable for limited sample sizes. GSCA is appropriate for structural models that include variables with reflective and formative indicators. This study utilizes the Alternating Least Square (ALS) parameter estimation. Iterations in ALS are used to achieve minimal residuals. Additionally, this study employs jackknife resampling to obtain standard error estimates. This study aims to identify the poverty model structure in Indonesia and examine the relationships among poverty, human resources, economic, and health variables. The results of the structural model of poverty in Indonesia are explained as follows: the influence of human resources and economic variables on poverty is insignificant, while the health variable significantly negatively influences poverty. Furthermore, the health variable significantly influences human resources, and both human resources and health significantly influence the economy.*

Keywords**:** *ALS, jackknife, poverty, SEM-GSCA*

1. INTRODUCTION

Poverty is an economic problem faced by many countries in the world, including Indonesia. This problem is due to the inability of individual or households to fulfill their basic needs, such as food, clothing, shelter, and other social needs to provide a decent life [1].

As of March 2022, the percentage of poverty in Indonesia reached 9,54%, and increased by 0,03% to 9.57% as of September 2022. This number is later decreased to 9,36% by March 2023 [2]. Although decreased, these numbers still show the high level of Poverty in Indonesia and need to be a priority for the government to countermeasure it. To overcome poverty in Indonesia, it is necessary to understand its causes so that efforts to address the high poverty rate can run smoothly. Structural Equation Modeling (SEM) is one method used to understand the relationships among the various dimensions that contribute to poverty.

SEM is a part of multivariate analysis that integrates factor analysis, path analysis, and regression analysis. SEM aims to estimate the correlations between latent variables and their indicators. SEM is divided into two main models: the measurement model and the structural model [3]. SEM can be grouped into two categories, namely Covariance-Based SEM (CBSEM) and Variance-Based SEM (VBSEM). Limitations of CBSEM include the assumption of a large sample size, the requirement that indicators be reflective, the need for multivariate normality in the data, the need for a theoretically grounded model, and the possibility of model inaccuracy (indeterminacy). To mitigate the limitations of CBSEM, VBSEM was developed, including the Partial Least Squares (PLS) and Generalized Structured Component Analysis (GSCA) methods. GSCA, as a component-based method, holds a significant value in calculating scores and can be used even with a limited number of samples. Moreover, GSCA is well-suited for a structural model that includes both reflective and formative indicators. The use of GSCA on a limited number of samples is considered helpful to researchers, as it helps overcome the minimum sample requirement of CBSEM [4], [5].

This introduction presents the aims of this study: to identify the structural model of poverty in Indonesia and the influence of human resources, economic, and health variables on poverty in Indonesia using the GSCA method.

 Submitted: 2024-10-06 *Accepted: 2025-11-14*

2. METHODOLOGY

2.1. Materials and Data

This study uses secondary data on factors that cause poverty. These data are obtained from Publikasi Badan Pusat Statistik Indonesia 2023 [6], Statistik Kesejahteraan Rakyat 2023, and Indonesia dalam Angka 2023. The variables used comprise four latent variables and 10 indicators, based on the studies by Novitasari [7] and Nuryanti [8]. The table below shows the list of variables in this study:

Table 1. Variables of this Study

Latent Variables	Observation Variables	Scale
Poverty	Z1: Percentage of Poor Population (P_0)	Ratio
	Z2: Poverty Depth Index (P_1)	Ratio
	Z3: Poverty Severity Index (P_2)	Ratio
Human Resources	Z4: Average Years of Schooling	Ratio
	Z5: Literacy Rate	Ratio
	Z6: Percentage of Population with Elementary School Education or Above	Ratio
Economy	Z7: Percentage of Population Working in the Formal Sector	Ratio
	Z8: Percentage of Non-Food Consumption Per Capita Per Month	Ratio
Health	Z9: Percentage of Population Using Their Own Toilet	Ratio
	Z10: Life Expectancy	Ratio

According to the theory above, there are relations between latent variables and also relations between latent variables and their indicators, as seen in the path diagram conceptual model below:

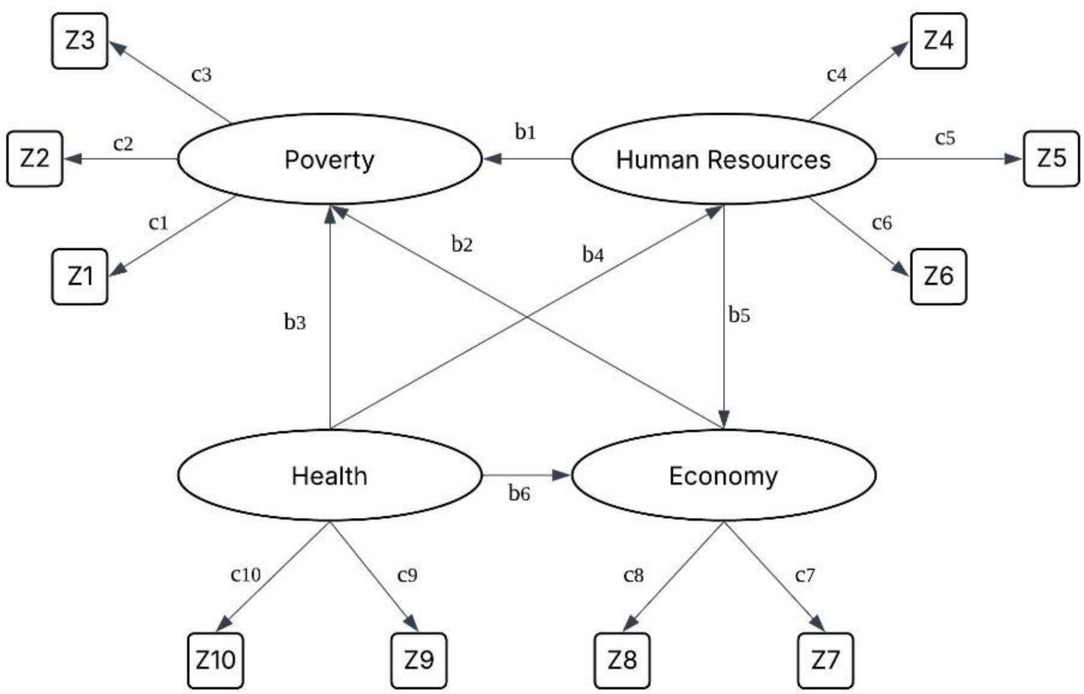

Figure 1. Path Diagram

According to Figure 1, we can put forward the following hypothesis:

H_1: Human resources affect poverty

H_2: The economy affects poverty

H_3: Health affects poverty

H_4: Health affects human resources

H_5: Human resources affect the economy

H_6: Health affects the economy

2.2. Research Methods

According to Leka [9], GSCA includes three models: the measurement model, the structural model, and the weighting model. The measurement model establishes the relationships between latent variables and their indicators. Generally written as follows:

$$z_i = C\gamma_i + \varepsilon_i$$

where z are indicator variables, γ are latent variables, C are loading values between latent variables and their indicators, and ε are residuals of indicator variables. Structural models depict relationships between each latent variable. Generally written as follows:

$$\gamma_i = B\gamma_i + \zeta_i$$

where y are latent variables and B are path coefficients that connect each latent variable, and ζ are the latent variables' residue. Weighting models are used to determine correlations between indicators and latent variables. Generally, written as follows:

$$\gamma_i = Wz_i$$

where W is the weight component from the indicator variables.

Parameter estimation in GSCA was performed using least-squares methods. In GSCA, the structural and measurement models are combined into a single integrated model. The parameter estimation process focuses on minimizing the residuals from this integrated model. The method used to minimize the residuals of the integrated model is Alternating Least Squares (ALS).

The calculation process in ALS is more complex than in Ordinary Least Squares (OLS). Therefore, to achieve minimum residuals, this process involves iterations. Iteration stops when the convergence conditions are met, such as when the difference between the current and previous estimates is $\leq$ 0,001 [4].

Hypothesis testing in GSCA uses jackknife resampling. Resampling involves drawing new samples from the original dataset to form additional samples. These new samples are the same size as the original and are selected randomly, either with or without replacement. Resampling is an alternative when there are not enough observations for the study, which can lead to less accurate parameter estimates. Furthermore, this method enables assessing data validity without assuming normality [10]. Jackknife is a resampling method first introduced by Quenouille (1949) to estimate bias, later expanded by Tukey [11] to estimate standard deviation. This method works by eliminating one data point from the sample and repeating the process as many times as the available samples [12].

The steps of this study are as follows:

1) Determining latent variables and indicator variables for model specification.
2) Creating a path diagram and analyzing the interpretation of the measurement and structural models from the equation.
3) Estimating parameters.
4) Evaluating the measurement model using the estimated loading value, $\sqrt{AVE}$ value, and alpha value.
5) Determining parameter path coefficient and R-squared value.
6) Evaluating the overall goodness of fit of the model to measure how well the model fits the study conducted.
7) Testing the hypothesis for the measurement model and structural model using jackknife resampling by observing the standard error and statistic CR value to see the significance of the variables in the model.

3. RESULTS AND DISCUSSION

Below is the result of GSCA analysis in R, which created an ALS algorithm that met the convergence criteria at 0.001 after four iterations.

3.1. Measurement Model Evaluation

1) Convergent Validity

Latent variables have good convergent validity if their loadings are > 0.5 [13].

Table 2. Loading Estimate of Measurement Model

		Estimate
Poverty	Z1	0.967
	Z2	0.999
	Z3	0.979
Human Resources	Z4	0.896
	Z5	0.887
	Z6	0.972
Economy	Z7	0.949
	Z8	0.820
Health	Z9	0.939
	Z10	0.905

Table 2 shows that the value of loading estimates is already > 0.5, therefore the convergent validity is valid.

2) Discriminant Validity

Validity of reflective indicators are evaluated using AVE value, by comparing $\sqrt{AVE}$ of each latent variable with the correlation between other latent variables inside the model. If $\sqrt{AVE}$ of one latent variable is greater than the other, then the discriminant validity is considered good [13].

Table 3. Correlation Value of $\sqrt{AVE}$ of Measurement Model

	Poverty	Human Resource	Economy	Health
Poverty	0.965			
Human Resource	0.336	0.845		
Economy	0.380	0.454	0.787	
Health	0.559	0.261	0.459	0.850

Table 3 shows that the $\sqrt{AVE}$ of poverty, human resources, economy and health variables are 0.965, 0.845, 0.787 and 0.850 respectively. These numbers are greater than the other correlations between latent variables, meaning this model has good discriminant validity.

3) Internal Consistency Reliability

A good Internal Consistency Reliability have a value of alpha ≥ 0.6 [4]. The alpha values provided in Table 4 shows a value of greater than 0.6 therefore all variables are considered good.

Table 4. Value of Alpha

	Alpha
Poverty	0.983
Human Resource	0.908
Economy	0.749
Health	0.826

3.2. Structural Model Evaluation

1) Path Coefficient

Structural model evaluation aims to measure the accuracy of the currently developed model. The influence of the relationship between latent variables is evaluated accordingly to the estimated path coefficient. The impact of the correlation between latent variables is evaluated using the calculated path coefficient[14]. The structural model is assessed by examining the path coefficients.

Table 5. Structural Model of Path Coefficient

	Estimate
Human Resources→Poverty	-0.242
Economy→Poverty	-0.056
Health→Poverty	-0.586
Health→Human Resources	0.511
Human Resources→Economy	0.443
Health→Economy	0.451

The variable refers to a study by Novitasari [7] that analysed poverty in East Java using SEM to determine the correlations among poverty, the economy, human resources, and health.

Table 5 shows that human resources, economic, and health variables have negative effects on poverty, meaning that the lower they are, the greater the poverty. Health variables have a positive impact on human resources, meaning that higher levels of health are associated with better human resources. Human resources and health variables positively affect the economy, meaning increases in both will also boost the economy.

2) R-squared

A structural model with an R-squared of 0.613 indicates that the structural equations explain 61.3% of the variation in the poverty variables. In other words, economic, human resource, and health variables can explain 61.3% of the fluctuations in poverty in Indonesia, while other variables outside the model explain 38.7%. The R-squared value of 0.261 for the human resource variables indicates that the structural equation explains 26.1% of the variance in these variables. The economy variables' R-squared of 0.604 indicates that 60.4% of their variation is explained by the structural equation.

3.3. Overall Goodness of FIT Evaluation

Overall goodness of FIT evaluation is shown in Table 5 below:

Table 6. Overall FIT

Indicator	Value
FIT	0.727
AFIT	0.705
GFI	0.997
SRMR	0.096

Table 5 shows the FIT value of 0.727, indicating that the model explains approximately 72.7% of the data variation. AFIT value of 0.705 shows that the model explains approximately 70.5%. GFI and SRMR values of $\geq$ 0.9 and $\leq$ 0.1, respectively, indicate that the model is quite suitable, meaning the structural model is unidimensional and includes factors that influence poverty in Indonesia.

3.4. Hypothesis Testing

1) Measurement Model Hypothesis Testing

Below are tests that can be done:

$H_0 : c_{ij} = 0$, loading indicator is not reliable

$H_1 : c_{ij} \neq 0$, loading indicator is reliable

Test Statistics:

$$CRl = \frac{c_{ij}}{SE(c_{ij})}$$

H_0 is denied if the value of |CRl| $\geq$ 2,0 has an error rate of α, which means that the loading measurement of each indicator is reliable. The following is a table of the results of the calculations above:

Table 7. Measurement Model Loading Estimates

		Estimate	SE	CRl
Poverty	Z1	0.967	0.010	120.875
	Z2	0.999	0.001	1000
	Z3	0.979	0.008	122.375
Human Resources	Z4	0.896	0.033	44.8
	Z5	0.887	0.105	10.435
	Z6	0.972	0.034	34.714
Economy	Z7	0.949	0.015	63.267
	Z8	0.820	0.066	12.424
Health	Z9	0.939	0.036	33.536
	Z10	0.905	0.031	33.519

Based on the calculation of the CRl value in Table 7, it is obtained that each indicator value in the table is greater than 2, so that the loading value of each indicator is believed to be valid and reliable. The measurement model from Table 7 is as follows:

$Z_1 = 0.967\gamma_1 + \varepsilon_1$

$Z_2 = 0.999\gamma_1 + \varepsilon_2$

$Z_3 = 0.979\gamma_1 + \varepsilon_3$

$Z_4 = 0.896\gamma_2 + \varepsilon_4$

$Z_5 = 0.887\gamma_2 + \varepsilon_5$

$Z_6 = 0.972\gamma_2 + \varepsilon_6$

$Z_7 = 0.949\gamma_3 + \varepsilon_7$

$Z_8 = 0.820\gamma_3 + \varepsilon_8$

$Z_9 = 0.939\gamma_4 + \varepsilon_9$

$Z_{10} = 0.905\gamma_4 + \varepsilon_{10}$

which are equivalent with the matrix below.

$$\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \\ Z_7 \\ Z_8 \\ Z_9 \\ Z_{10} \end{bmatrix} = \begin{bmatrix} 0.967 & 0 & 0 & 0 \\ 0.999 & 0 & 0 & 0 \\ 0.979 & 0 & 0 & 0 \\ 0 & 0.896 & 0 & 0 \\ 0 & 0.887 & 0 & 0 \\ 0 & 0.972 & 0 & 0 \\ 0 & 0 & 0.949 & 0 \\ 0 & 0 & 0.820 & 0 \\ 0 & 0 & 0 & 0.939 \\ 0 & 0 & 0 & 0.905 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \end{bmatrix}$$

2) Structural Model Hypothesis Test

The parameter coefficient is considered significant if the value of $|CRb| \geq 2$ [15]. The following are the tests carried out:

$H_0 : b_i = 0$, path coefficient of parameter γ_i has no significant effect

$H_1 : b_i \neq 0$, path coefficient of parameter γ_i has significant effect

Test Statistics:

$$CRb = \frac{b_i}{SE(b_i)}$$

H_0 is denied if the value of $|CRb| \geq 2{,}0$, showing that the coefficient for said parameter is highly statistically significant. The result of the calculation above is inside the following table:

Table 8. Structural Model Path Coefficient

	Estimate	SE	CRb
Human Resources→Poverty	-0.242	0.359	-0.674
Economy→Poverty	-0.056	0.243	-0.230
Health→Poverty	-0.586	0.197	-2.975
Health→Human Resources	0.511	0.253	2.019
Human Resources→Economy	0.443	0.110	4.027
Health→Economy	0.451	0.118	3.822

Based on Table 8 with the CRb value, it can be concluded that:

a. HR has a significant statistical effect on poverty

b. Economy has a significant statistical effect on poverty

c. Health has an insignificant statistical effect on poverty

d. Health has an insignificant statistical effect on human resources

e. HR has an insignificant statistical effect on the economy

f. Health has an insignificant statistical effect on the economy

The structural model based on Table 8 is as follows:

$\gamma_1 = -0.242y_2 - 0.056y_3 - 0.586y_4 + \zeta_1$

$\gamma_2 = 0.511y_4 + \zeta_2$

$\gamma_3 = 0.443y_2 + 0.451y_4 + \zeta_3$

Equivalent with the following matrix:

$$\begin{bmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\end{bmatrix} = \begin{bmatrix}0 -0.242 -0.056 -0.586\\ 0 \quad 0 \quad\quad 0 \quad\quad 0.511\\ 0 \;\; 0.443 \quad\; 0 \quad\quad 0.451\end{bmatrix}\begin{bmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\\ \gamma_4\end{bmatrix} + \begin{bmatrix}\zeta_1\\ \zeta_2\\ \zeta_3\end{bmatrix}$$

4. CONCLUSION

According to the discussion in the previous sections, it can be concluded that:

1) Regression analysis shows that health, human resources, and the economy have a negative effect on poverty, meaning that the lower their levels, the higher poverty is in Indonesia. Furthermore, there is a positive effect between variables of health and human resources, and both human resources and health have a positive effect on the economy.
2) Based on the research, human resources and economic variables have no statistically significant effect on poverty variables, whereas health variables have a significant negative effect on poverty in Indonesia. Health has a significant effect on human resources, human resources have a significant effect on the economy, and health has a significant effect on the economy.

REFERENCE

[1] N. W. R. Suadnyani and I. B. Darsana, "Pengaruh Pertumbuhan Ekonomi, Pengangguran Dan Pendidikan Terhadap Kemiskinan Di Kabupaten Bangli," *E-Jurnal Ekonomi Pembangunan Universitas Udayana*, vol. 7, no. 5, pp. 1022–1049, 2018, [Online]. Available: https://ojs.unud.ac.id/index.php/eep/article/view/38745

[2] Badan Pusat Statistik, "Persentase Kemiskinan 2023." [Online]. Available: bps.go.id

[3] F. Anugrah Pratiwi, N. Nessyana Debataraja, and S. Martha, "Analisis Derajat Kesehatan di Kalimantan Barat dengan Generalized Structured Component Analysis," *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, vol. 10, no. 2, pp. 277–286, 2021, doi: https://doi.org/10.26418/bbimst.v10i2.46541.

[4] Solimun, Nurjannah, L. Amaliana, and A. A. R. Fernandes, *Metode Statistika Multivariat Generalized Structured Component Analysis (GSCA) Pemodelan Persamaan Struktural (SEM)*. Malang: UB Press, 2019.

[5] H. Hwang and Y. Takane, "GSCA Pro," 2004. doi: https://doi.org/10.1007/BF02295841.

[6] Badan Pusat Statistik, *Statistik Indonesia 2023*. 2023. [Online]. Available: https://www.bps.go.id/id/publication/2023/02/28/18018f9896f09f03580a614b/statistik-indonesia-2023.html

[7] D. A. Novitasari, "Analisis Data Kemiskinan Di Jawa Timur Berdasarkan Data Susenas 2017 Dengan Menggunakan Structural Equation Modelling," *Media Mahardhika*, vol. 17, no. 3, p. 414, 2019, doi: 10.29062/mahardika.v17i3.97.

[8] R. R. Nuryanti and T. Soebagijo, "SEM-PLS Untuk Analisis Struktur Kemiskinan Pada Masa Pandemi Covid-19," *Seminar Nasional Official Statistics*, vol. 2021, no. 1, pp. 195–203, 2021, doi: 10.34123/semnasoffstat.v2021i1.836.

[9] S. S. Leka and T. S. Yanti, "GSCA Model untuk Menentukan Pengaruh Kualitas Layanan dan Fasilitas Terhadap Kepuasan Mahasiswa Program Studi Statistika FMIPA Univeristas Islam Bandung," in *Prosiding Statistika*, 2020, pp. 72–79. doi: http://dx.doi.org/10.29313/.v6i2.22846.

[10] M. F. Papalia, Solimun, and Nurjannah, "COMPARISON OF RESAMPLING EFFICIENCY LEVELS OF JACKKNIFE AND DOUBLE JACKKNIFE IN PATH ANALYSIS," vol. 17, no. 2, pp. 807–818, 2023, doi: https://doi.org/10.30598/barekengvol17iss2pp0807-0818.

[11] J. W. Tukey, "The Problem of Multiple Comparisons," 1953.

[12] R. S. Dwiputra, "Perbandingan Resampling dengan Metode Jackknife dan Bootstrapping pada Model WarpPLS," Undergraduate Thesis, Universitas Brawijaa Malang, 2018.

[13] F. Fitriani, A. Rusgiyono, and T. Widiharih, "PENERAPAN METODE GENERALIZED STRUCTURED COMPONENT ANALYSIS PADA KEPUASAN KONSUMEN (Studi Kasus: Pasien Klinik Q)," *Jurnal Gaussian*, vol. 9, no. 4, pp. 454–463, 2020, doi: 10.14710/j.gauss.v9i4.29416.

[14] B. Efron, "Better Bootstrap Confidence Intervals," *J Am Stat Assoc*, vol. 82, no. 397, pp. 171–185, 2012, doi: http://dx.doi.org/10.1080/01621459.1987.10478410.

[15] H. Hwang, "Fuzzy Clusterwise Generalized Structued Component Analysis," *Psychometrika*, vol. 72, no. 2, pp. 1–20, 2007, doi: https://doi.org/10.1007/s11336-005-1314-x.