Generalized Structured Component Analysis to Analyze Measurement Models: Utilization of Health Insurance

D V Ferezagia¹, K A Safitri¹, N F Dewi¹, D Anggara²

¹ Vocational Education Program, Universitas Indonesia, Depok, 16426, Indonesia
² Directorate of Price Statistics, Statistics Indonesia, Central Jakarta, 10440, Indonesia

*E-mail: debrinaferezagiaa@gmail.com

Abstract. This study aims to analyze measurement models of the Utilization of Health Insurance. A survey is conducted annually by Statistics Indonesia in measuring the Utilization of Health Insurance through National Socio-Economic Survey (SUSENAS). Code 81301- 81309 on SUSENAS Core is blocking code XIII with a description of Health Insurance Utilization. Y₁: Do you have health insurance? Y₂: have you ever used health insurance? Y₃: What are the reasons you have never used health insurance? Y₄: in the last year, have you ever been refused a health check? Y₅: what are the reasons you experience refusal to check your health?, Y₆: in the last year, have you ever used health insurance for hospitalization?, Y₇: What are the reasons you never used health insurance for hospitalization?, Y₈: in the last year, have you ever been refused hospitalization?, and Y₉: what are the reasons you experience refusal to be hospitalized? Generalized Structured Component Analysis (GSCA) is part of SEM-based on components that have global least-square optimization criteria, which can be consistently minimizing sum squares residuals to obtain model parameter estimates. GSCA is a powerful analysis method. This is because it is not based on many assumptions such as variables that do not have to have multivariate normal distributions (indicators with category, ordinal, interval to ratio can be used on the same model), the amount of data does not have to be large. The results showed that some indicator variables had no significant effect on the construct variables. The loading factor> 0.6 are latent construct indicators that provide good convergent validity.

1. Introduction
Health insurance in Indonesia is still an important topic for researchers to study. This is because the health insurance literacy index among the community is still low. The government continues to strive to improve health insurance literacy and community participation as participants in this insurance. The government in 2014 through Law Number 40 of 2004 and Law Number 24 of 2011 obliged all people to use health insurance called National Health Insurance.

A survey is conducted regularly by the Central Bureau of Statistics through the National Socio-Economic Survey which is usually called SUSENAS. National Health Insurance which is abbreviated as health insurance cannot be measured directly so it requires a manifest variable to measure the use of health insurance among the public. The measurement model is often called the Outer model, showing how the manifest/observer variable (indicator) variables represent latent constructs to be measured, namely by testing the validity and reliability of these latent constructs.

The construct variable becomes an important part to determine how the representative variable manifest represents the latent variable. Research on how many manifest variables are required in a
variant construct has been proposed by Hair et al. [5]. He stated that the number of manifest variables was not determined in a specific number. A construct can have one manifest variable even though this one variable is of course not representative enough. Even if there is only one manifest variant it can cause problems on the reliability prediction. Hair said that it is better to have a multi-indicator approach in each of its constructs. Resinger & Mayondo [9] states that 3 manifest variables are good enough. Meanwhile, Hair et al [5] suggest that the number of manifest variables is 5 to 7.

In addition to the number of manifest variables that determine the representative of a constructed variable, the sample size is also an important thing in parameter estimation. This is closely related to sampling error. As research conducted by Reisinger and Mavondo [9], the sample size is a very important thing that can affect the size of the model fit. Reisinger and Mavondo [9], Schreiber et al. [10], and Hair et al. [5] argue that there is no standard sample size but the absolute thing is that the sample size must be greater than the number of corrections in the input data matrix. Recommendations from other researchers such as Reisinger and Mavondo [9], Schreiber et al. [10], and Hair et al. [5] are ten respondents per parameter.

The sample size will increase with the complexity of the model [10]. In other words, if the model built is very complex, the small sample size will not meet the adequacy size in estimating the model. The procedure used in parameter estimation is Maximum Likelihood Estimation (MLE). In this method of aging, a sample size above 400 will increase the sensitivity in detecting differences between data. So that the goodness of fit will increase too. Meanwhile, Hair et al [5] recommend a minimum sample size of 177. Reisinger and Mavondo [9] argue that the basic assumptions that need to be met to get accurate conclusions to include independent observation, random sampling (except for longitudinal data), linearity, multivariate normality, validity, no outliers, and no skew or kurtosis. in the data [5,9]. Large variations in normality can invalidate all statistical tests, and small sample sizes can lead to poor parameter predictions, a negative variance which is often called Heywood cased [1].

With these limitations, the component or variety-based measurement model approach is a good method of analysis. Component-based measurement model approaches include Partial Least Square (PLS) and Generalized Structured Component Analysis (GSCA). Both methods can avoid the problem of inadmissible solutions and factor indeterminacy [3]. PLS has limitations in estimating parameters, because it does not have a global optimum, so it does not have a single criterion consistently in determining model estimates. As a result, there is no guarantee that PLS will provide an optimal solution and it is difficult to determine the overall model fit test [6]. GSCA is a method proposed by Hwang H & Takane [6] to overcome weaknesses and maintain PLS strengths. GSCA has the optimum global least-square criteria, which consistently minimizes the number of squared errors to obtain a model parameter estimator. The GSCA is also equipped with an overall model to fit size. Based on this background, the study of the health insurance latent variable measurement model used GSCA.

2. Data and Methodology

2.1. Data
The data used consisted of latent variables, namely the use of health insurance with 9 manifest variables. A survey is conducted annually by Statistics Indonesia in measuring the Utilization of Health Insurance through National Socio-Economic Survey (SUSENAS). The data were taken randomly from as many as 399 respondents. The variables used in this study are

$Y_1$: National Health Insurance ownership
$Y_2$: The use of National Health Insurance
$Y_3$: The reasons never used National Health Insurance
$Y_4$: Health checkin in the last year
$Y_5$: The reasons for being refused for a health checking
$Y_6$: The use of national health insurance for hospitalization in the last year
$Y_7$: The reasons never used insurance for hospitalization
$Y_8$: Refusal of hospitalization, in the last year
Yv: The reasons for refusal to be hospitalized

2.2. Methodology

2.2.1. Validity and reliability
The validity test of the Pearson product-moment correlation uses the principle of correlating item scores with the total scores of respondents' answers. The correlation results are compared with the critical point. If the statistical test value \( r > \) the critical point \( r \), it is declared valid. Reliability measures the accuracy of the latent variables used. According to Ringle (1986), reliability is the term used to describe one of the most significant properties of a test value in a consistent manner. Reliability testing with Cronbach Alfa. If the Cronbach Alpha value is \( > 0.6 \), it can be said that the instrument is reliable.

2.2.2. Multivariate Normal
This normality test can be seen in the value of the Critical Ratio (CR) value of its skewness and kurtosis. If the CR value ranges from -2.58 to 2.58 (2.58) at a significance level of 1% (0.01), it can be concluded that the data is normally distributed both univariate and multivariate.

2.2.3. Estimated Parameters
If \( Z \) represents the normalized indicator variable matrix (size \( n \times j \)), where \( n \) represents the number of observations and \( j \) the number of indicator variables. GSCA can be viewed as component-based SEM where the latent variable is defined as a component or weighted composite of the indicator with the equation [6]:

\[
\gamma = ZW
\]

where \( \gamma \) is a matrix of latent variables \( (n \times t) \) sized and \( W \) is the component weight matrix of the indicator variable \( (j \times t) \) sized, where \( t \) is the number of latent variables.

GSCA also includes a measurement model that describes the relationship between indicators and constructs as well as a structural model that connects between constructs. The measurement model can be systematically written as follows [6]:

\[
Z = \gamma C + \epsilon
\]

Where \( C \) is the loading matrix between the latent variables and their indicators \( (t \times j) \) sized, \( \epsilon \) is the error matrix \( (n \times j) \) sized for the measurement model [6].

The parameters of GSCA are \( V, W, \) and \( A \) which need to be estimated so that the sum of the squares of all residuals \( E = ZV - ZWA = \Psi - \tau A \) as small as possible. This means minimizing the optimum criteria Least Square [6]:

\[
f = \text{trace}((ZV-ZWA)'(ZV-ZWA))
\]

\[
f = \text{trace}((\Psi - \tau A)'(\Psi - \tau A))
\]

with \( V, W \) dan \( A \). Components in \( \Psi \) and \( \tau \) normalized for identification purposes, eg \( \gamma'\gamma_i = 1 \).

Alternating Least Square (ALS) is used to minimize this equation [7]. The ALS method is a general approach to parameter estimation that involves grouping a parameter into several subsets and then obtaining the least squares for one of the parameter subsets assuming that all remaining parameters are constant. The ALS algorithm used in the GSCA consists of 2 stages, namely \( A \) is predicted with \( V \), and \( W \) is fixed and \( V \) and \( W \) are predicted with \( A \) is fixed.

The ALS algorithm is repeated at both stages until it converges. GSCA takes the bootstrap method developed by Efron [2] to estimate standard error parameters which will be used to test the significance.

2.2.4. The goodness of Fit in the model
The goodness of this model aims at the validity and reliability of the research instrument. Evaluation of the measurement model by looking at convergent validity, discriminant validity, and average variance extracted (AVE) [4]. Convergent validity is assessed based on the loading factor value of each latent construct forming indicator. A latent construct is said to have a good convergent validity of the loading factor value is more than 0.6 and is significant. Discriminant validity is assessed by comparing the root
value of the AVE of each latent construct with the correlation between the construct in question and other constructs in the model. If the root value of the AVE of each construct is greater than the correlation value between constructs and other constructs, then it is said to have good discriminant validity [4]. There are several measures of model fit, namely FIT, AFIT, GFI, and SRMR. FIT measures how much diversity the data can be explained by the model and the value ranges from 0 to 1 [6].

3. Result and Discussion

3.1. Data Exploration

According to Tabel 1, it reaches 96.74% of 399 respondents own national health insurance which the government target to oblige society to have health insurance. The ownership of this health insurance can be through the respondents’ office or they paid for its premium independently.

| Table 1. The percentage of insurance ownership |
|----------------------------------------------|
| Answers | Percentage |
| Yes     | 96.74%     |
| No      | 3.26%      |

Table 2 shows that 57.64% of 399 respondents used their health insurance that would help the availability of all doctor’s expenses, medicines, hospitalization, to surgery. If at any time they fall ill, the insurance company will cover the medical expenses, according to the contract or agreement. 42.36% of 399 respondents said they never used the health insurance.

| Table 2. The Percentage of ever using national health insurance |
|---------------------------------------------------------------|
| Answers | Percentage |
| Yes     | 57.64%     |
| No      | 42.36%     |

The biggest percentage of total respondents reasoned that they did not experience health issues so they never use their health insurance even they already have it based on Tabel 3. The 2nd position of the reason why never use health insurance is they cure illness they selves. It means they could lookout for the traditional medication considering some Indonesians still believe in it whose raw materials are made of plants, rocks, and in the form of animals, although health testing or research has not been carried out or they would see doctors with their expenses. The 8.1% of total respondents said that this national health insurance that the Indonesian Government provides to protect citizens has difficult procedures and requirements. This finding can be a critic to be fixed and monitored regularly.

| Table 3. The percentage of why never use health insurance |
|----------------------------------------------------------|
| Answer | Percentage |
| Do not experience health complaints | 33.57% |
| Cure the illness their selves | 21.64% |
| Long queue | 14.36% |
| Procedures and Requirements that are difficult to fulfill | 8.1% |
| Inactive insurance card | 5.3% |

Tabel 4 explained that 79.2% of total respondents refusing a health check and 20.8% said they don’t refuse to do a health check. Table 5 showed the reason why they are being refused for a health check is that they did not fulfill the procedure 31.82%.

| Table 4. Percentage of respondents refusing a health check |
|------------------------------------------------------------|
| Answers | Percentage |
| Yes | 20.8% |
| No | 79.20% |
Table 5. The reason for being refused to do a health check

| Answer                                      | Percentage |
|---------------------------------------------|------------|
| did not fulfill the procedure               | 31.82%     |
| not according to the service schedule       | 27.57%     |
| no medical personnel required as service providers | 6.1%     |
| does not have a supporting examine         | 3.3%       |

National health insurance cards or policyholders get inpatient benefits along with optional benefits in the form of outpatient care, childbirth, dental care, and optional additional benefits and almost 92% of total respondents used insurance for hospitalization according to Table 6.

Table 6. Percentage of using insurance for hospitalization

| Answers                  | Percentage |
|--------------------------|------------|
| Yes                      | 91.22%     |
| No                       | 8%         |

But apparently, some respondents did not use insurance for hospitalization. Table 7 shows their reasons for doing so, we see that some of them had other private insurances to cover their expense in health that each 10.2%.

Table 7. The reasons never used health insurance for hospitalization

| Answers                                      | Percentage |
|----------------------------------------------|------------|
| Do not experience health complaints          | 85.71%     |
| Use another private insurance                | 10.2%      |
| Procedures and Requirements that are difficult to fulfill | 2.6%      |

Table 8 showed that 90.2% of 399 respondents were never refused when they need to be hospitalized. While table 9 explained most of the reason they experience refusal of hospitalization is no available vacancy in that hospital.

Table 8. The percentage of being refused for hospitalization

| Answers                      | Percentage |
|------------------------------|------------|
| Yes                          | 9.78%      |
| No                           | 90.22%     |

Patients who will be hospitalized using health insurance will sometimes experience rejection. Of the respondents who had experienced rejection, 28.07% was due to the unavailability of rooms. The old reason, 27.5% did not comply with existing procedures. This is shown in table 9.

Table 9. The reasons for experiencing refusal to be hospitalized

| Answer                                      | Percentage |
|---------------------------------------------|------------|
| No available vacancy                        | 28.07%     |
| Did not comply with the procedure           | 27.57%     |
| Do not have a supporting health examination | 10.36%     |
| No medical personnel required as service providers | 2.4%      |

3.2. Reliability and Validity

The reliability test based on the results of the analysis amounted to 0.615. It can be said that the health insurance construct variable is reliable. Validity using the r value in table 10 Based on the results of the analysis of the statistical value $r >$ the critical point $r$ (0.128) it can be said that the manifest variable is valid.
Table 10. Validity and Reliability Test

| Manifest Variables | r    | Critical Point r |
|--------------------|------|------------------|
| Y1                 | 0.61 | 0.128            |
| Y2                 | 0.70 | 0.128            |
| Y3                 | 0.70 | 0.128            |
| Y4                 | 0.38 | 0.128            |
| Y5                 | 0.13 | 0.128            |
| Y6                 | 0.64 | 0.128            |
| Y7                 | 0.63 | 0.128            |
| Y8                 | 0.27 | 0.128            |
| Y9                 | 0.30 | 0.128            |

3.3. Normal Multivariate

Based on the results of normal multivariate testing, it can be seen that the Critical Ratio (CR) value is outside the normal distribution range +/- 3, so it can be said that the latent variables are not normally distributed. So that parameter estimation using OLS and MLE methods cannot be used. Estimation of the correct parameters using the ALS contained in the GSCA model showed in Table 11.

Table 11. Estimation of the correct parameters using the ALS

| Variable | skew  | c.r.  | kurtosis | c.r.  |
|----------|-------|-------|----------|-------|
| Y9       | 4.730 | 38.568| 20.814   | 84.867|
| Y8       | -2.777| -22.647| 5.875   | 23.955|
| Y7       | 5.122 | 41.768| 27.452   | 111.931|
| Y6       | -2.987| -24.358| 7.098   | 28.942|
| Y5       | 2.992 | 24.399| 7.408   | 30.206|
| Y4       | -1.478| -12.050| .260   | 1.060|
| Y3       | 3.857 | 31.455| 15.821   | 64.507|
| Y2       | .283  | 2.308 | -1.875   | -7.644|
| Y1       | -5.266| -42.939| 25.726  | 104.895|
| Multivariate | 241.252 | 171.236 |

3.4. Measurement Model

Evaluation of the measurement model is done by looking at the loading value on the indicator variable against the latent variable. The t value of each real indicator variable and the factor loading value > 0.6 is a latent construct indicator that provides good convergent validity (Chin 1998). The use of loading criteria is more than 0.6 because the percentage of diversity that can be explained by the indicator variable against the latent variable is quite large, namely 60%. Indicator variables that have a loading factor <0.6 are excluded from the next analysis. Poor indicator variables can be input and further evaluated on indicators as question items that can represent latent variables by interested parties. Estimates of the Measurement Model can be seen in Table 12.

Table 12. Estimates of Measurement Model

| Manifest variables | Estimates | Std.error | 95% CL-LB | 95% CL-UB |
|--------------------|-----------|-----------|-----------|-----------|
| Y1                 | 0.7226    | 0.0443    | 0.6542    | 0.8242    |
| Y2                 | 0.7274    | 0.0555    | 0.6001    | 0.8057    |
| Y3                 | 0.6958    | 0.0713    | 0.5548    | 0.8065    |
A loading value of less than 0.6 can be an evaluation of the manifest variable on the health insurance latent variable. In this study, indicators Y4 and indicator 8 are not representative of the manifest variables, namely whether health insurance has been rejected or not, both outpatient (Y4) and inpatient (Y8).

The loading value also shows the manifest variables that have the most influence on the use of health insurance. Based on the results of the analysis, the indicators that most influence the use of health insurance are Y6 and Y9. The y6 indicator refers to the use of health insurance for hospitalization. This study shows that the use of health insurance for hospitalization is the most dominant indicator used by the community in the use of health insurance. Also, the y9 indicator refers to the reasons for refusal by health facilities to be hospitalized for patients. It can be recommended that the most important thing is the use of health insurance for hospitalization.

3.5. Overall Model Evaluation
The overall model evaluation for the modified model can be seen from the suitability test as shown in Table 13. It can be seen that the middle square root value of standardized error (SRMR) is 0.2342, and the GFI value is close to 1 (0.9674). So, it can be said that the model used is quite good. The FIT value measures the total diversity of all variables that can be explained by the model. This means that 38.96% of the diversity of all variables can be explained by the model. The corrected FIT value is 38.59%.

| Manifest variables | Estimates | Std.error | 95% CL-LB | 95%CL-UB |
|--------------------|-----------|-----------|-----------|-----------|
| Y4                 | -0.1533   | 0.1672    | -0.5151   | 0.1222    |
| Y5                 | 0.7206    | 0.1216    | 0.4508    | 0.8702    |
| Y6                 | 0.8632    | 0.0804    | 0.6413    | 0.9322    |
| Y7                 | 0.7861    | 0.0534    | 0.6775    | 0.8741    |
| Y8                 | 0.1529    | 0.1885    | -0.1782   | 0.5639    |
| Y9                 | 0.8632    | 0.0401    | 0.7416    | 0.8843    |

Table 13. The Goodness of Fit Models

| Model Fit          | Measure | Std.Error | 95% CI-LB | 95%CI-UB |
|--------------------|---------|-----------|-----------|-----------|
| FIT                | 0.3896  | 0.0091    | 0.3801    | 0.4137    |
| AFIT               | 0.3859  | 0.0091    | 0.3764    | 0.4101    |
| GFI                | 0.9674  | 0.0923    | 0.6967    | 0.9952    |
| Standardized Root Mean Square (SRMR) | 0.2342 | 0.0235 | 0.1897 | 0.2824 |

4. Conclusion and Recommendation
GSCA is best used on data that does not comply with normal assumptions. The testing phase starts with validity and reliability testing, normality test of parameter estimation, and model evaluation. Only 38.96% of the diversity of all variables can be explained by the model. So that for the manifest variable to be more representative of the latent variable, it is necessary to add another variable following the supporting theory. The most dominant manifest variables are Y6 and Y9. The y6 indicator refers to the use of health insurance for hospitalization. This study shows that the use of health insurance for hospitalization is the most dominant indicator used by the community in the use of health insurance.
Also, the y9 indicator refers to the reasons for refusal by health facilities to be hospitalized for patients. It can be recommended that the most important thing is the use of health insurance for hospitalization.

References
[1] Chou C P and Bentler P M 1985 Estimate and Tests in Structural Equation Modelling In R Hoyle (Md) Structural Equation Modelling Concepts Issues and Application (Newbury Park CA Sage) pp 37-55
[2] Efron B 1982 The Jackknife The Bootstrap and Other Resampling Plans (Philadelphia: SIAM)
[3] Fornell C and Bookstein F 1982 Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory (Journal of Marketing Research) vol 18 pp 440-452
[4] Fornell C and Larcker D 1981 Evaluating Structural Equation Models with Unobservable Variable and Measurement Error (Journal of Marketing Research) vol 19 pp440-452
[5] Hair J F, Anderson R E, Tatham R L, and Black WC 2006 Multivariate Data Analysis (Upper Saddle River: NJ Pearson Education Incorporation)
[6] Hwang H and Takane Y 2004 Generalized Structured Component Analysis (Psychometrica) Vol 69 No 1 pp 81-99
[7] Leeuw D J, Young F W, and Takane Y 1976 Regression with Qualitative and Quantitative Variables: An Alternating Least Squares Method Wit Optimal Scaling Features (Psychometrica) Vol 41 No 4 pp 505-529
[8] Robert L E and Frisbie D A 1991 Essential of Educational Measurement (U.S.A: Prentice-Hall)
[9] Reisinger Y and Mavondo F 2007 Structural Equation Modeling (Journal of Travel & Tourism Marketing) Vol 21 (4) pp 41-71
[10] Schreiber J B, Frances K S, Jamie K, Amaury N, and Elizabeth A 2006 Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review (The Journal of Educational Research) Vol 99(6) pp 323-337