Structural Equation Modeling of the Factors Affecting the Nutritional Status of Children Under Five in Banyuwangi Region using Recursive (one-way) GSCA

I Made Tirta[1], Nawal Ika Susanti[2*] and Yuliani Setia Dewi[3]
[1,3] Program Studi Magister Matematika, FMIPA, Universitas Jember
[2] MTs. Nurul Huda, Jenisari, Genteng Banyuwangi
[*]email: tirtaimade@gmail.com

ABTRACT

Structural Equation Modeling is one among popular multivariate analysis, especially applied in pschology and marketing. There are two main types of Structural Equation Modeling namely covariance-based or CB-SEM and variance-based or Partial Least Square (PLS)- SEM. Both types have advantages and disadvantage. To overcome its limitation, Generalized Structured Component Analysis (GSCA) was then proposed as an extension of PLS-SEM. In estimating the parameters, GSCA uses Alternating Least Squares (ALS) and in estimating the standard error of the parameter estimates it uses the bootstrap method. In this paper, GSCA is applied to study the causality model of Infant nutritional status, in relation with socio-economic status and infantcare status in Banyuwangi Region. The results show that both socio-economic and infantcare status have significant positive influence on infant nutritional status.

Keywords: Alternating least square, generalized structural component analysis, nutritional status of infants, structural equation modelling

INTRODUCTION

Structural Equation Modeling is one among popular multivariate analysis in the fields of psychology and marketing. In its development, there are two main approaches used to solve SEM, namely covariance-based SEM, known as CB-SEM and variance-based SEM known as Partial Least Square (PLS)-SEM. Dispite its popularity, CB-SEM requires some assumptions that must be met including (i) the data must have multivariate normal distribution (at least must be of continuous scale), (ii) the indicator should be reflective, (iii) the sample must be relatively large. While for the PLS-SEM (i) data can be of continuous or ordinal scale, (ii) indicators can be reflective or formative, (iii) sample can be relatively small (Ghazali, 2008). However, PLS-SEM also has a disadvantage of not having a measure of the global fit. Hwang & Takane (Hwang, & Takane, 2004) proposed Generalized Structured Component Analysis (GSCA) which was developed to complement the existing deficiencies in the Partial Least Square approach, especially in measuring overal goodness of fit. In this study, GSCA is applied to model the structural factors that affect the nutritional status of infant (children under five years) in Banyuwangi region, with relatively small samples.

THEORY

The GSCA

GSCA is a component-based SEM, in which latent variables are defined as components or weighted composites of observed variables with the following equation:

$$\gamma_i = \mathbf{W}z_i \qquad (1)$$

z_i is a vector with observed variables with $t \times 1$ observations to measure the i-th observations, for $i=1...n$. and $\mathbf{W}$ is the weight components matrix of the indicator variable sized $j \times t$, where t is the number of composite variables of the model

GSCA also includes the measurement models that describe the relationship between indicators and constructs, as well as the structural model that describes the relationship among the constructs. Mathematically, measurement model is written as follows:

$$z_i = \mathbf{C}\gamma_i + \varepsilon_i \qquad (2)$$

with $\mathbf{C}$ is the $t \times j$ loading matrix between latent variables the indicators, γ_i is the residual vector ($j \times 1$). When $\mathbf{C} = \mathbf{0}$, it indicates that a formative measurement models are used, in which all latent variables are formed by the observed variables as in equation (2). Moreover, only a factor loading of $\mathbf{C}$ can be equal to zero so that the model includes

both formative and reflective models at the same time.

Structural equation models is expressed as below :

$$\gamma_i = B\gamma_i + \xi_i \tag{3}$$

with **B** is a (*t*x*t*) path coefficient matrix that connect amongs latent variables and ξ is the (t×1) residual vector for γ_i. GSCA integrate (combine) the above three equations (1,2,3) into a single equation as follows :

$$\begin{bmatrix} z_i \\ \gamma_i \end{bmatrix} = \begin{bmatrix} C \\ B \end{bmatrix} \gamma_i + \begin{bmatrix} \varepsilon_i \\ \xi_i \end{bmatrix}$$

$$\begin{bmatrix} I \\ W \end{bmatrix} z_i = \begin{bmatrix} C \\ B \end{bmatrix} Wz_i + \begin{bmatrix} \varepsilon_i \\ \xi_i \end{bmatrix} \tag{4}$$

I is the identity matrix. If we make substitution $V = \begin{bmatrix} I \\ W \end{bmatrix}$, $A = \begin{bmatrix} C \\ B \end{bmatrix}$, $e_i = \begin{bmatrix} \varepsilon_i \\ \xi_i \end{bmatrix}$ then the following equations are found:

$$Vz_i = AWz_i + e_i$$
$$ZV = ZWA + E$$
$$\Psi = \Gamma A + E \tag{5}$$

Equation (5) is commonly refers to as a GSCA model. In the GSCA model, all indicators and constructs are included in **ZV** and interdependence is denoted with matrix **A** (Hwang, & Takane, 2004).

Parameter Estimation and Bootstrap

In GSCA, the unknown parameters (**V**, **W** and **A**) are estimated using *Alternating Least Square method* (ALS) so that the value of all the residual sum squares (**E**) are as small as possible for all observations. It is the same as the least squares optimization to minimize the following criterion

$$f = SS(ZV - ZWA) = SS(\Psi - \Gamma A$$

(10)

with respect to **V**, **W** and **A**, where SS (**X**) = trace($X^T X$). Components in Ψ or τ are normalized for identification purposes, for example $\gamma_1^T \gamma_i = 1$, so that equation (10) can be expressed as follows

$$f = \text{trace}(ZV - ZWA)'(ZV - ZWA) \tag{11}$$

ALS is a common approach to estimate the parameters that involve grouping parameters into several subsets and then get the least squares for one subset of parameters by assuming that all the remaining parameters are constant. The GSCA method consists of two subsets of parameters, **A** and (**V,W**). The algorithm are described as follows:

1. Updating **A**, assuming (**V,W**) constant
a. Initialize **V** and **W**

b. Compute $I \otimes \Gamma$
c. Form matrix Ω which is a matrix formed by deleting column component related to nol element in vec(A).
d. Updates **A** using *least square* estimate:
$$\hat{A} = (\Omega^T \Omega)^{-1} \Omega' \text{ vec } (\Psi)$$
Assuming that $\Omega^T \Omega$ is nonsingular, otherwise *Moore-Penrose inverse* will be applied [5].

2. Updating (**V,W**), assuming **A** constant
a. Initialize **A** by using previously updated **A**
b. Compute **S** containing weighting parameters to be estimated.
c. Define each column of the matrix **S** (as many as k columns) which is derived from any column of the matrix **W** (as many columns q) and **V** (as many as p column).
d. Defined $\Lambda = WA$.
e. Define β^T and Δ, by $\beta^T = e_p^T - a_q^T$ and $\Delta = \Lambda_{(-q)} - V_{(-p)}$
f. Define $\beta \otimes Z$

Model Detail and Evaluation

In this study, the data are the determining factor of poor nutritional status in infants (Cheah, et. al, 2010). In Banyuwangi region. Refering to the theory and relevant research results, then framework embodied in the structural model, is designed as seen in Figure 2 which is a unidimensional structural equation. The variables in the framework are symbolized according to the type of latent variables. There are 3 latent constructs and total of 10 observed variables, which are summarized in Table 1.

Table 1. Constructs and indicators involved in the research.

Construct	Indicator	Types of Data	Type of Indicator
Nutrition (η)	Weight for age ($z1=y1$)	Ordial	Reflective
	Height for age ($z2=y2$)	Ordinal	Reflective
	Weight for height ($z3=y3$)	Ordinal	Reflective
Socio-Economic (ξ_1)	Income ($z4=x1$)		Formatiive
	Mother's Education ($z5=x2$)	Ordinal	Formatiive
	Number of children ($z6=x3$)	Ordinal	Formatiive
	Mother's Job status ($z7=x4$)	Ordinal	Formatiive
Infantcare (ξ_2)	Breast milk ($z8=x5$)	Ordinal	Reflective
	Immunization ($z9=x6$)	Ordinal	Reflective
	Eating pattern ($z10=x7$)	Ordinal	Reflective

Socioeconomic status, Infantcare, are exogenous latent variables (independent), whereas nutritional status is an endogenous latent variable treated as the dependent variable. Considering the nature of the indicators, it is reasonable to treat indicators of socio-economics as formatives. The detail of model can be described in the path diagram (as Figure 1). The composite model of the path diagram with three components is as follows:
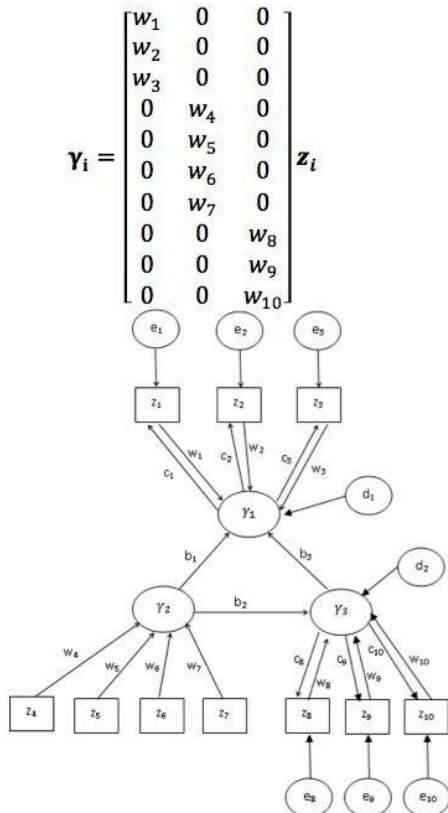
$$\gamma_i = Wz_i$$

$$\gamma_i = \begin{bmatrix} w_1 & 0 & 0 \\ w_2 & 0 & 0 \\ w_3 & 0 & 0 \\ 0 & w_4 & 0 \\ 0 & w_5 & 0 \\ 0 & w_6 & 0 \\ 0 & w_7 & 0 \\ 0 & 0 & w_8 \\ 0 & 0 & w_9 \\ 0 & 0 & w_{10} \end{bmatrix} z_i \quad (6)$$

Figure 1. Path diagram of the model

Measurement model for Figure 1 can be written as follows for exogenous variable,

$$z_i = C\gamma_i + \varepsilon_i$$
$$z_i = C_i\gamma_1 + \varepsilon_i \text{ for } i \le 3$$
$$z_i = C_i\gamma_3 + \varepsilon_i \text{ for } 7 < i \le 10$$

For endogenous variable,

$$\gamma_2 = w_4 z_4 + w_5 z_5 + w_6 z_6 + w_7 z_7$$

in matrix form, these equations can be expressed as equation (7):

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \\ z_8 \\ z_9 \\ z_{10} \end{bmatrix} = \begin{bmatrix} w_1 & 0 & 0 \\ w_2 & 0 & 0 \\ w_3 & 0 & 0 \\ 0 & w_4 & 0 \\ 0 & w_5 & 0 \\ 0 & w_6 & 0 \\ 0 & w_7 & 0 \\ 0 & 0 & w_8 \\ 0 & 0 & w_9 \\ 0 & 0 & w_{10} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ 0 \\ 0 \\ 0 \\ 0 \\ e_8 \\ e_9 \\ e_{10} \end{bmatrix}$$
(7)

Meanwhile the structural model for Figure 1 is as follows,

$$\gamma_1 = b_1\gamma_2 + b_2\gamma_3 + d_1$$
$$\gamma_3 = b_3\gamma_2 + d_2,$$

which are equivalent with the equations in the following matrix form:

$$\gamma_i = B\gamma_i + \xi_i$$
$$\begin{bmatrix} \gamma_1 \\ \gamma_3 \end{bmatrix} = \begin{bmatrix} 0 & b_1 & b_2 \\ 0 & b_3 & 0 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

(8)

Equations (6), (7), and (8) can then be combined into one equation as equation (A1) in the Appendix A. The relationship can be expressed as equation (9)
That is

$$ZV = ZWA + E$$
$$\Psi = \Gamma A + E$$
(9)

With

$$\Psi = [z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, z_{10}],$$
$$\Gamma = [\gamma_1, \gamma_2, \gamma_3] \text{ and } A = [C, B],$$

$$C =$$
$$\begin{bmatrix} c_1 & c_2 & c_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_4 & c_5 & c_6 & c_7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_8 & c_9 & c_{10} \end{bmatrix}$$
and

$$B = \begin{bmatrix} 0 & 0 \\ b_1 & b_3 \\ b_{21} & 0 \end{bmatrix}$$

In evaluating the model, GSCA conducts three stages, the first stage is the evaluation of the measurement model to see convergent validity based on the value of the factor loading of each indicator of latent constructs. There are some

criteria to be used. A latent construct is said to have a good convergent validity if the value of loading factor is more than 0.70 and significant (Chin, 1998), discriminant validity is good if the value of the square root of AVE of each construct is greater than the value of the correlation between the other constructs in the model (Fornell & Larcker, 1981). The recommended composite reliability value is greater than or equal to 0.70. Meanwhile, the average variance extracted (AVE) recommended value is greater than or equal to 0.50.

Three hypotheses can be proposed are as follows:

H_1: socioeconomic status has a positive effect on the nutritional status of infant

H_2: the Infantcare has o positive effect on nutritional status of infant

H_3: socioeconomic status has a positive effect on infant care

Location, Population and Sample
1. The location of this study are health centers in the district Genteng, in the villages: Genteng topWetan, Genteng Kulon, Kaligondo, Setail and Kembiritan;
2. Population in this study are 200 infants who visit the clinic in Genteng - Banyuwangi (mostly from the lower middle socio-economic class);
3. Sample in this study are 50 infants who visit the health centers in Genteng, who were taken randomly from total of 200 infants in Genteng region.

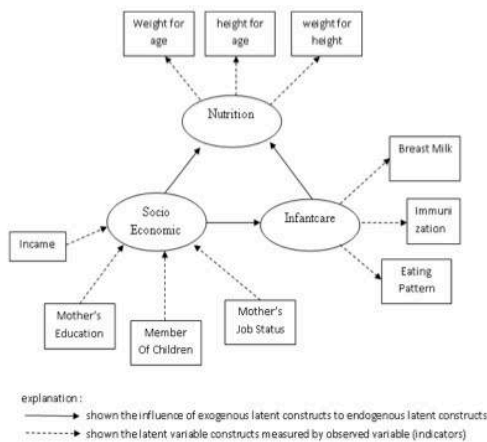The proposed structural model of the research is given in Figure 2.

explanation :
→ shown the influence of exogenous latent constructs to endogenous latent constructs
-----→ shown the latent variable constructs measured by observed variable (indicators)

Figure 2. The Propose structural model of research

RESULT AND DISCUSSION

The result (output) from GESCA (webbased or online program for GSCA) are given in Table 2. We find that loading value of all reflextive indicators of nutritional status and health status are well above 0.560 and they are statistically significant. Similarly AVE values are above 0.600 indicating that the average variance of indicators can be explained by the latent variable by more than 60%

Table 2. Estimates of measurement models

Variable	Loading			Weight			SMC		
	Estimate	SE	CR	Estimate	SE	CR	Estimate	SE	CR
Nutrition al	AVE = 0.609, Alpha =0.653								
y1	0.980	0.007	137.42*	0.694	0.062	11.19*	0.960	0.014	68.81*
y2	0.567	0.125	4.53*	0.254	0.063	4.05*	0.322	0.131	2.45*
y3	0.739	0.083	8.92*	0.238	0.060	3.95*	0.546	0.118	4.62*
Socio-economic	AVE = 0.000, Alpha =0.792								
x1	0	0	0	0.479	0.333	1.44	0	0	0
x2	0	0	0	0.236	0.346	0.68	0	0	0
x3	0	0	0	0.353	0.145	2.43*	0	0	0
x4	0	0	0	0.186	0.212	0.88	0	0	0
infantcar e	AVE = 0.714, Alpha =0.708								
x5	0.798	0.089	9.0*	0.344	0.069	5.01*	0.637	0.132	4.82*
x6	0.871	0.048	18.15*	0.446	0.062	7.22*	0.759	0.081	9.31*
x7	0.863	0.050	17.13*	0.390	0.064	6.09*	0.745	0.085	8.8*

*) significant at .05 level

Also evident that the alpha (Cronbach) value is above 0.4 indicating that reliability of the measurement model for reflective indicators are quite reliable. While the root of the AVE are respectively 0.780 and 0.882 which shows that the root of AVE values higher than the correlation between latent constructs. It can be concluded that all reflective indicators are valid and reliable measurement for latent variables.

For the nutritional status, there are three indicators that describe the variable . Based on From Table 1 it is seen that the BW/U indicator has the greatest influence in determining the nutritional status of children and become a dominant factor of 0.980 . For socioeconomic status, While for the Infantcare status, immunization has a great influence in determining the status and becoming a dominant factor of 0.871. The estimates of the parameters for the structural model are given in Table 2.

Table 2. The Estimate of the structural model

a. Path coefficients

Path Coefficients			
	Estimate	SE	CR
Socio-economic->Nutrition	0.272	0.213	1.28
Socio-economic->Infantcare	0.677	0.064	10.67*
Infantcare ->Nutrition	0.571	0.168	3.39*

CR* = significant at .05 level

b. R squares

R square of Latent Variable	
Nutritional	0.611
Infantcare	0.459

c. Mean Scores

Means Scores of Latent Variables	
Nutritional	2.267
Socio-economic	2.703
Infantcare	1.998

d. Correlation

Correlations of Latent Variables (SE)			
	Nutritional	Socio-economic	Infantcare
Nutritional	1	0.659 (0.100)*	0.756 (0.066)*
Socio-economic	0.659 (0.100)*	1	0.677 (0.064)*
Infantcare	0.756 (0.066)*	0.677 (0.064)*	1

* significant at .05 level

Based on Table 2 it can be said that the structural model are statistically significant. Path coefficient of family socio-economic status to the nutritional status is 0.272 (but not significant at 95% level) which means that socioeconomic status has no significant direct effect on nutritional status. Path coefficient of family socioeconomic status to infantcare status is significant, 0.677, which means that socioeconomic status has a significat positive effect on infantcare and the path coefficient of the infantcare to the nutritional status is 0.571 which means stastus infantcare has a significat positive effect on nutritional status.

Mathematically structural model of the nutritional status of children as follows,

$$\gamma_1 = 0,272\gamma_2 + 0,571\gamma_3$$

$$\gamma_3 = 0,677\gamma_2$$

That is,
Nutrition = 0,272* Socio econonmic +
            + 0,571* infantcare
Infantcare = 0,667* socio-economic

Because both of these pathways are significant then the next step is testing the fit of model. The summary of the model fit are given in Table 3.

Table 3. Overall model fit measures

Model Fit	
FIT	0.560
AFIT	0.543
NPAR	19

It is seen in Table 3 that the FIT value ≥ 0.5, AFIT value ≥ 0.5. We conclude that unidimensional structural model of the factors that influence nutritional status in Genteng district is a good model and also an acceptable model. The overall structural model composed of nutritional status of children in Genteng-Banyuwangi is as given in Figure 3.
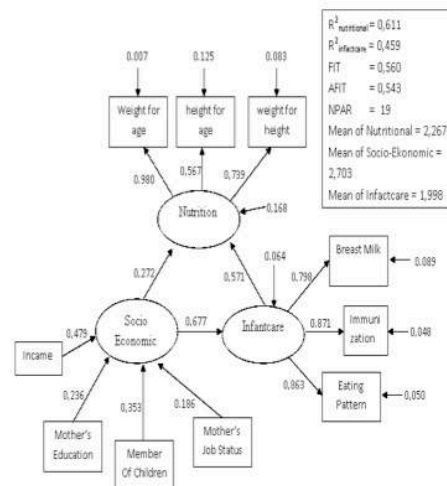

Figure 3. The Emperical structural model of nutritional status

CONCLUSION

Based on the results of the discussion in the previous section it can be concluded as follows:
1. There are three significant paths leading to the nutritional status of children. The first path directly from socioeconomic status to nutritional status, the second path is from infantcare status to nutritional status and third lines of socioeconomic status to the status

infantcare and then to the nutritional status. Of the three paths, the greatest influence on the nutritional status is good through the infantcare status of the second and third lines rather than socioeconomic status.

2. From the measurement model and the structural model, it is indicated that low mother's education may lead to little or no understanding of the importance of immunization and also the important a good diet and vice versa. On the other hand poor immunization and poor diet can lead to underweight of infants which is considered as malnutrition or poor nutrition.

REFFERENCES

Cheah, WL, Manan, A, Husein, Z., *A Structural Equation Model of the Determinants of Malnutrition among Children in Rural Kelantan Malaysia,* Original Research Rural and Remote Health,10,1248 (2010)

Chin,W.W.,.*Issues and opinions on structural equation modeling,* MIS Quarterly, 22 (1),pp.287-298 (1998)

Fornell & Larcker, *Evaluating Structural Equation Models with Unobservable Variables and Measurement Error,* Journal of Marketing Research, 18 (February), 39-50 (1981)

Ghazali, Imam, *Generalized Structured Component Analysis (GSCA): Model Persamaan Struktural Berbasis Komponen,* Semarang: Badan Penerbit Universitas Diponegoro (2008)

Hwang, H & Takane, Y., *Generalized Structured Component Analysis,* Psychometrika.Vol 69. No.1, 81-99(2004)

APPENDIX

$$\mathbf{Z}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & w_1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & w_2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & w_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & w_8 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & w_9 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & w_{10} \end{bmatrix} = \mathbf{Z}\begin{bmatrix} w_1 & 0 & 0 \\ w_2 & 0 & 0 \\ w_3 & 0 & 0 \\ 0 & w_4 & 0 \\ 0 & w_5 & 0 \\ 0 & w_6 & 0 \\ 0 & w_7 & 0 \\ 0 & 0 & w_8 \\ 0 & 0 & w_9 \\ 0 & 0 & w_{10} \end{bmatrix}\begin{bmatrix} c_1 & c_2 & c_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & b_2 & b_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_8 & c_9 & c_{10} & 0 & 0 \end{bmatrix} + \mathbf{E}$$

$$\text{(A1)}$$

For

$$\mathbf{Z} = [z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8, z_9, z_{10}], \text{ and}$$
$$\mathbf{E} = [e_1, e_2, e_3, 0,0,0,0, e_8, e_9, e_{10}].$$