International Journal of Water & Hydro Constructions– IJWHC
Volume 3 : Issue 1 [ISSN 2374-1600]
Publication Date : 31 August, 2016

Structural Equation Modelling as Instrument for Water Pollution Factor Analysis

Study Case on Surabaya River, Indonesia

Adam Rus Nugroho, Ali Masduqi

***Abstract—Structural Equation Modelling* (SEM) as a manifestation of multivariate analysis studied to be used as an instrument for decision making in factor analysis of water pollution. Variance based SEM could help with variables of water pollution which consists of qualitative and quantitative variables. Linear structural equation model obtained as a result of the study shows the contribution level among pollution factors. T-statistic values obtained by doing bootstrap model resample show the significances of pollution factors. Study resulted on conclusion that variance based *Structural Equation Modelling* (SEM) using Partial Least Square method (PLS-SEM) is sufficient enough to help decision making related to water pollution factors.**

Keywords— factor analysis, structural equation modelling, water pollution

I. Introduction

Statistical instrument was being overly used in management decision making research, but scarcely found in engineering or science-related research. However, few research related with statistical model found in engineering like risk factor analysis on civil engineering project using Structural Equation Model [4]. For time being, environmental science has a rising number on research related on modeling of observable variable of environment, i.e. organic, suspended solids and nutrient parameter distribution [8]. There's only few researches existed on environmental decision-making modelling. It happens because unobservable variables frequently become necessary on decision-making. This study trying to make a breakthrough for fusing quantitative and qualitative variable on a pollution factors analysis in order to show the applicability of Structural Equation Modeling as instrument to help decision-making analysis.

Adam Rus Nugroho
Department of Environmental Engineering
Universitas Islam Indonesia
Indonesia

Ali Masduqi
Department of Environmental Engineering
Institut Teknologi Sepuluh Nopember
Indonesia

River on area of study, Surabaya River, has been being studied many times and the most of the results conclude that the river is having too much pollution load [1]. But there's one question people still can't be sure of. It is about the scientific reason to call upon the most significant and most contributing pollution source. Confirmatory factor analysis can help with the answer by using statistical multivariate analysis. Multivariate analysis of Structural Equation Modelling (SEM) has ability to illustrate simultaneous linear interconnection among variables by constructing linear equation.

For having complete factor analysis of water pollution, usually it has *unobserved variables* (i.e. Wastewater Treatment Plant Conditions, Regulation, Social Perception) which should be explained by some parameters. These parameters called *observed variables* in SEM. The term to call upon variables relation in SEM is *construct*. Formative construct is variables relation where the *observed variables* are causing *unobserved variables*. Reflective construct is variables relation where the *observed variables* caused by *unobserved variable*.

Structural Equation Model has 2 categories: the covariance-based and the variance-based. The variance-based SEM is recommended to be used to make a structural model of formative construct [6]. Variance-based SEM also recommended when the study has some tentative theories or either the measurement of its latent variables still relatively new [9].

The study's objective is to analyze the application of variance-based Structural Equation Modelling in obtaining the contribution value and significances of each river pollution factors. The study also aims to gain perspective on how strong model obtained. Those evaluation of model is done by using free statistical software: SmartPLS 2.0.

II. Methods

A. *Study Area and Data Description*

Water pollution in Surabaya River in city of Surabaya in East Java, Indonesia was being assessed on this study. Surabaya is the capital city of East Java. The city is known for its diversity of culture because most citizen come as urban immigrant from around East Java. The city location being on the north edge of East Java made it as estuary of main river of East Java. The main river (Brantas River) goes down into

Surabaya as one of Surabaya main river, named Surabaya River.

The study area took up along 2,4 kilometers in river segment before Surabaya River divided into two rivers. Downstream chosen as study area in order to find the impact of industrial sewage from upstream. Thus this one will be upstream and industrial wastewater factor. There's also two input of drainage basin which always take in domestic wastewater accumulation from common drainage basin around north of segment into Surabaya River. The two drainage input will be domestic wastewater factor.

Qualitative factors of river pollution would be come from citizen perception and government regulation policy. Unfortunately, while citizen perception can be obtained from questionnaire, regulation policy is not easy as citizen perception to be obtained. It will need some methodology exploration to obtain and mix it up with other variables. Citizen perception of environment awareness hypothetically will affect the water pollution. The questionnaire will give environment awareness data from citizen in settlement around area of study.

While the quantitative factor has parameter of wastewater quality to be analyzed, qualitative factor of water pollution need to be quantified first. Questionnaire data will be converted to quantitative data with scoring method before being taken in multivariate analysis of Structural Equation Modelling.

Water quality data used in the case study of Surabaya River take account of BOD, COD, TSS. Citizen perception divided into environment awareness, the tendency of citizen to make sewage discharge and the tendency to make garbage discharge into water body. The variable of water pollution level in downstream observed by STORET method and Pollutant Index method corresponding the method used in Indonesia Ministry of Environment regulation.

B. Variables and Number of Samples

Variables determination on confirmatory factor analysis based on references and study hypotheses. Structural Equation Models has two latent variables: endogenous variables and exogenous variables. Endogenous variables are variable which its value affected by exogenous variables in structural model. Value of exogenous variables determined outside of the model. Endogenous variables have symbol of Y, while exogenous variables have symbol of X. The variables on this study summarized on Table I.

Minimum number of sample in Partial Least Square (PLS) analysis should be at least 10 times of the total indicator in the most complex construct [2]. It means minimum sample number for PLS-SEM analysis should be 10 times from total number of indicator on a latent variable which has the most indicator number. The case study has one latent variable which has most number of indicator. It is the variable of *Upstream Condition* that has four indicators. Thus, the minimum number of samples on the case is 4 times 10, resulting 40 samples.

C. Model Concept

The statistical model in PLS-SEM has concept which allow the latent variable measured by reflective or formative indicators. Reflective indicator is the result of latent variable (on path diagram, the arrow directed from latent variable to indicators), whereas formative indicator is the cause of latent variable (on path diagram, arrow directed from indicator to latent variable).

TABLE I. VARIABLES OF CASE STUDY

Latent Variables	Indicator Variables	
	Symbol	***Indicator Name***
Upstream Condition (X_1) *HULU*	$X_{1.1}$	BOD of upstream
	$X_{1.2}$	COD of upstream
	$X_{1.3}$	TSS of upstream
	$X_{1.4}$	River flow of upstream
Citizen Perception (X_2) *PM*	$X_{2.1}$	Environmental awareness
	$X_{2.2}$	Tendency to discharge untreated wastewater
	$X_{2.3}$	Tendency to dispose solid waste into water body
Domestic Wastewater Quality (Y_1) *ALD*	$Y_{1.1}$	BOD of domestic wastewater
	$Y_{1.2}$	COD of domestic wastewater
	$Y_{1.3}$	TSS of domestic wastewater
Water Pollution Level (Y_2) *TP*	$Y_{2.2}$	STORET method
	$Y_{2.2}$	Water Pollution Index method

The latent variable interaction with its indicators referred as outer model, while the interaction among latent variables themselves referred as inner model. Outer model also referred as measurement model, while inner model also called as structural model. The equation of measurement model and structural model for the case study given below. The path diagram which display the interaction among variables and indicators given in Fig.1.

1) Measurement Model Equation

Formative Measurement Model

$X_1 = \pi_1 X_{1.1} + \pi_2 X_{1.2} + \pi_3 X_{1.3} + \pi_4 X_{1.4}$

$Y_1 = \pi_5 Y_{1.1} + \pi_6 Y_{2.2} + \pi_7 Y_{2.3}$

Reflective Measurement Model

$X_{2.1} = \lambda_1 \xi_3 + \delta_1$

$X_{2.2} = \lambda_2 \xi_3 + \delta_2$

$X_{2.3} = \lambda_3 \xi_3 + \delta_3$

$Y_{2.1} = \lambda_4 Y_2 + \varepsilon_1$

$Y_{2.2} = \lambda_5 Y_2 + \varepsilon_2$

Symbols Description:

$X_{1.1} - X_{2.3}$ = indicators of exogenous variabels

$Y_{1.1} - Y_{2.2}$ = indicators of endogenous variabels

$\pi_1 - \pi_7$ = weights loading of formative indicator

$\lambda_1 - \lambda_5$ = loading factor of reflective indikator

$\zeta_1 - \zeta_2$ = error of formative indicator
$\delta_1 - \delta_6$ = error of reflective exogenous latent
$\varepsilon_1 - \varepsilon_2$ = error of reflective endogenous latent

2) **Structural Model Equation**

$Y_1 = \gamma_{12}X_2 + \zeta_1$

$Y_2 = \gamma_{21}X_1 + \gamma_{22}X_2 + \beta_{21}Y_1 + \zeta_2$

Symbols Description:

$\gamma_{12}, \gamma_{21}, \gamma_{22}$ = regression weight exogen to endogen
Y_1 = endogenous *domestic wastewater quality*
Y_2 = endogenous *water pollution level*
X_1 = exogenous *upstream condition*
X_2 = exogenous *citizen perception*
ζ_1, ζ_2 = error of structural model

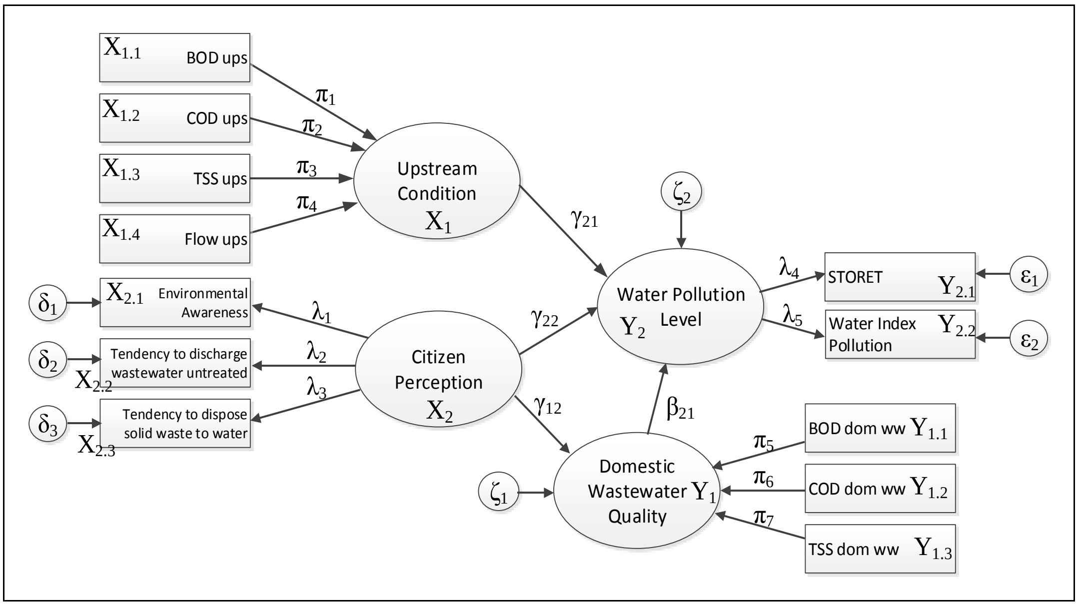

Figure 1. SEM Path Diagram Concept

III. Results and Discussion

A. *Outer Model Evaluation*

Outer model evaluation is a test of measurement model for each latent variable's construct. On this evaluation, reflective and formative indicator are tested for validity and reliability (Table III). Formative indicator should be tested for the *weight loading* and also the presence of multicollinearity (Table II).

Multicollinearity is a phenomenon in which two or more predictor in a multiple regression model are highly correlated. Linear regression method would fail or have poor model in the presence of multicollinearity. Following problems are caused by multicollinearity: serious statistical problems in estimation of parameter, low significance level, wrong trajectory on model selection process, and statistical inference is not reliable [5]. It has been a common rule of thumb that variance inflation factor (VIF) becomes indicator of multicollinearity. VIF value of more than 10 indicates that there will be a huge multicollinearity problem [7].

TABLE II. FORMATIVE INDICATOR REFERENCES FOR OUTER MODEL EVALUATION [3]

Parameter	Reference Value
Weight Significances	• t-value > 1,65 → α = 0,1 • t-value > 1,96 → α = 0,05 • t-value > 2,58 → α = 0,01
Multicollinearity	• VIF < 10 • Tolerance > 0,20

Variables which have formative indicator on the case study are X_1 and Y_1. Their indicator are parameters of water quality, which means the indicator data's validity and reliability has been taken in laboratory where the data comes from. Thus, on this outer model evaluation, only reflective indicator which should be evaluated for their validity and reliability. On this case, only X_2 and Y_2 outer model which will be evaluated.

TABLE III. VALIDITY & RELIABILITY REFERENCES FOR OUTER MODEL EVALUATION [3]

Category	Parameter	Reference Value
Convergent Validity	Loading Factor	• > 0,70 for confirmatory research • > 0,60 for exploratory research • > 0,50 for start-up research
	Average Variance Extracted (AVE)	• > 0,50 for confirmatory and exploratory research
Discriminant Validity	Cross Loading	• > 0,70 for every variable
	√AVE and Correlation among Latent Construct	• √AVE > correlation between latent construct
Reliability	Cronbach's Alpha	• > 0,70 for confirmatory research • > 0,60 for exploratory research

	Composite Reliability	• > 0,70 for confirmatory research • 0,60-0,70 for exploratory research

Discriminant validity evaluation needs *cross loading* value and also comparison between square root AVE with corelation among latent variables. The result (Table IV) shows that all of reflective indicators have *cross loading* value more than 0,70 and value of square root AVE are higher than latent variables correlation. These mean reflective indicators tested is valid.

Reliability of reflective indicators tested by the value of *cronbach's alpha* and *composite reliability* (Table II). Value of *cronbach's alpha* and *composite reliability* of this case study indicators are higher than 0,7 (Table V), thus their reliability proved.

TABLE IV. CROSS LOADINGS OF REFLECTIVE INDICATOR

	PM	TP
PM1	0,7767	-0,1146
PM3	0,9606	-0,2607
TP1	-0,2343	0,8291
TP2	-0,1950	0,9218

TABLE V. AVE, COMPOSITE RELIABILITY, CRONBACHS ALPHA, R^2

	AVE	Composite Reliability	Cronbachs Alpha	R^2
HULU	-	-	-	-
ALD	-	-	-	0,0002
PM	0,7630	0,8642	0,7269	-
TP	0,7685	0,8688	0,7076	0,2996

B. *Inner Model Evaluation*

Having outer model evaluated, it continued to inner model evaluation. Inner model evaluation is structural model test. The evaluation's objective is to predict the relation among latent variables. The prediction strength of structural model shown by value of R^2 each of endogenous latent variable. R^2 value of 0,75, 0,50 and 0,25 explain the model strength, respectively are strong, moderate and weak [6]. Weight significances then evaluated by the value of t-statistics. The value of t-statistics higher than 1,96 means the latent variable is significance within α=0,05. However, t-statistics value only will be shown if the model has been through *bootstrap* process. Bootstrap in statistic refer to test that relies on random sampling with replacement. Bootstrapping allows assigning measures of accuracy to sample estimates. On the case study, the bootstrapping done with 100 times resampling. T-statistic values shown in Table VI. The result of path diagram after 100 resample bootstrap shown in Fig. 2.

The R^2 value of endogenous *water pollution level* variable is 0,2996. It can be referred that the structural model of this case study is rather weak. Based on Table VI, the only significance variable is *upstream condition* (confidence interval of 95%), because it was the only variable which has t-statistic value more than 1,96.

TABLE VI. √AVE COMPARED WITH LATENT VARIABLES CORRELATION

	√AVE			ALD	HULU	PM	TP
PM	0,8735	compared	**PM**	0,0156	-0,1924	1,0000	-
TP	0,8766		**TP**	0,2739	0,4955	-0,2389	1,0000

TABLE VII. PATH COEFFICIENTS

	Original Sample	Sample Mean	Standard Error	T Statistics
	O	*M*	*STERR*	*|O/STERR|*
DOM → WPLv	0,185	0,215	0,169	1,099
UPS → WPLv	0,425	0,466	0,190	2,243
CP → DOM	0,016	-0,078	0,286	0,055
CP → WPLv	-0,160	-0,129	0,173	0,924

Table VI is giving the path coefficients of the structural equation model (sample mean). Thus, the structural equation of the model is:

$$Water\ Pollution\ Level = (0{,}215 \times upstream\ condition) + (0{,}466 \times domestic\ wastewater\ quality) - (0{,}129 \times citizen\ perception)$$

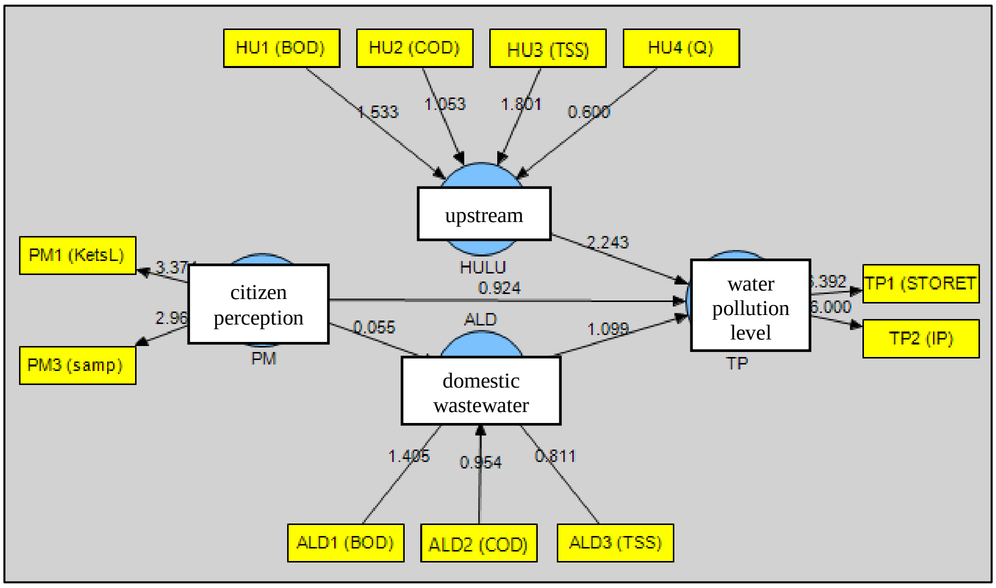

Figure 2. Path Diagram after 100 Resample Bootstrap by SmartPLS

IV. Conclusion

The variance-based SEM can be used well-enough for helping the pollution factor analysis. It can mix qualitative and quantitative variables with good perseverance. The case study resulted in a weak structural model with significance value identified for one variable (upstream condition). In order to strengthen the model, it need more variable and indicators to put up. It is recommended to have many observed indicators when make the model concept in first place. It is also recommended to add latent variable of regulation or policy that leads into pollution rising.

Acknowledgment

The study has been successful thanks to two peoples in Institut Teknologi Sepuluh Nopember (Surabaya, Indonesia): Ali Masduqi, for giving the support on data access and research management; Bambang Widjanarko Otok, for giving insight on statistical approach method. Great gratitude is given also to Universitas Islam Indonesia for support on research publication funding.

References

[1] Binnie & Partners Ltd. "Surabaya River Pollution Control Action Plan Study". Vol. 2, Main Report, 1999, Redhill, UK.

[2] D. Gefen, D. W. Straub, and M. C. Boudreau. "Structural Equation Modeling and Regression: Guidelines for Research Practice," Communications of the Association for Information Systems, Vol. 4, No. 7, 2000.

[3] H. Latan, I. Ghozali. "Partial Least Squares – Concept, Technique and Application of SmartPLS 2.0 M3," 2012, Semarang, Indonesia.

[4] H. P. Chandra. "Structural Equation Model for Investigating Risk Factors Affecting Project Success in Surabaya," Procedia Engineering vol. 125, 2015, pp. 53-59.

[5] H. Yu, S. Jiang, K. C. Land. "Multicollinearity in Hierarchical Linear Models," Social Science Research, Vol. 53, 2015, pp. 118-136.

[6] J. F. Hair, C. M. Ringle, and M. Sarstedt. "PLS-SEM: Indeed a Silver Bullet", Journal of Marketing Theory and Practice, Vol. 19, No. 2, 2011, pp. 139-151.

[7] M.H. Kutner, C.J. Nahctsheim, J. Neter. "Applied Linear Regression Models," 4th ed., 2004, Mc.Graw-Hill

[8] Suntoyo, et al. "Modelling of the COD, TSS, Phosphate and Nitrate Distribution Due to the Sidoardjo Mud Flow into Porong River Estuary," Procedia Earth and Planetary Science vol. 14, 2015, pp. 144-151.

[9] W. W. Chin, P. R. Newsted. "Structural Equation Modeling Analysis with Small Samples Using Partial Least Squares", in Statistical Strategies for Small Sample Research, R. Hoyle (Ed.), 1999, pp. 307-341, Sage Publications, Thousand Oaks (California).

About Author:

Adam Rus Nugroho was born in Indonesia on 1991. He studied Environmental Engineering for Bachelor Degree (2008-2012) and Magister Degree (2012-2014) in Institut Teknologi Sepuluh Nopember (Surabaya, Indonesia). Since 2015 May, he was working as lecturer in Universitas Islam Indonesia (Yogyakarta).