M – 17

Analysis Propensity Score with Structural Equation Model Partial Least Square

Setia Ningsih[1], B. W. Otok[2], Agus Suharsono[2], Reni Hiola[3],
[1]Dept. of Statistics, Institut Teknologi Sepuluh Nopember
[2]Dept. of Statistics, Institut Teknologi Sepuluh Nopember
[3]Dept. of Public Health, Universitas Negeri Gorontalo
thya.setianingsih@gmail.com

Abstract— In research of epidemiology, structural equation modeling (SEM) has been become very popular, especially for latent variables. In SEM there are assumptions that must include the data should be normally distributed multivariate and a used large of data. For overcome these problems required the alternative approach of SEM based variance or partial least square (PLS). SEM-PLS does not require an assumption that a lot. In health sector randomization is not possible, because it concerns the lives of humans. So that assumptions independent can't be achieved. This can lead to imbalances covariates and selection bias. Therefore, to overcome these problems applied propensity score (PS). This method is a statistical analysis that can be used to analyze study design Non-Experimental where can't do randomization to treatment groups. Furthermore, as suggested new methods for handling selection bias is a marginal meanweighting through stratification (MMW-S). The analysis result obtained when using MMW-S is powerful because MMWS show strong reduction in of selection bias. The author uses an innovative method by using empirical data HIV/AIDS. Briefly using MMW-S with a predisposition, clinical manifestations, and opportunistic infection. And adherence to antiretroviral (ARV) as a confounding variable. The results showed that the method of MMW-S can removed bias more than 93.5%.

Keywords: *SEM-PLS, Propensity score, MMW-S, HIV/AIDS*

I. INTRODUCTION

Health problems are one of the factors that have an important role in creating quality human resources. In health, SEM has become a very popular method mainly used to examine the Latent variables. Non-Experimental studies or observational studies are empirical investigations of the effects caused by the treatment as randomized experiments Randomized Controlled Trials (RCT) is impossibble [1]. In general, RCT is very required in the research to the independence assumption so that the bias selection can be minimized. However, in the field of health research involving human, RCT is not always practicable. One method is suggested to be used for such problems is propensity score. Once the propensity score has been estimated in a given dataset, a data preprocessing procedure is performed to create comparability between study groups, it is referred to as pre-processing because it is performed before the final treatment effect is estimated, thus replicating the RCT by separating the study design stage from the outcomes analysis [2].

In general, this first entails stratifying the analytic sample into quantiles of the propensity score, and then generating a weight for each individual based on their corresponding stratum and treatment assignment, the stratification reduces bias in the observed covariates used to create the propensity score, and the weighting standardizes each treatment group to the target population [3]. This approach namely marginal mean weighting through stratification (MMWS), can handle a broad array of experimental conditions that researchers will likely encounter in evaluating health care interventions Once generated, the MMWS can then be used within the appropriate outcome model to estimate unbiased treatment effects [3].

II. LITERATURE REVIEW

A. Structural Equation Modeling Partial Least Square (SEM-PLS)

SEM-based variance or based components called as partial least square (PLS) is a method of analysis that is powerful and often referred to as soft modeling because it does not require assumptions such as

data should not normally distributed multivariate, can be used with data of nominal, ordinal, interval and ratio, in addition sample should not be large [4]. SEM-PLS consists of three components are outer model is specifies the relationship between variables latent and indicators or manifest variables (measurement model), inner model is specifies the relationship between the latent variables (structural model), and the weight relation.

PLS is a powerful modeling methods due to not assume the data must be with particular scale of measurement, samples should not be large, not require extremely assumptions. Types of indicators on the PLS are two as follows:

- Reflective indicators tend to be influenced by the latent variables (indicators is a reflection of the latent variables).
- Formative indicators tend to affect the latent variables (indicators are descriptors of latent variables).

Algorithm SEM-PLS as explained by [5], can be illustrated by Figure 1.

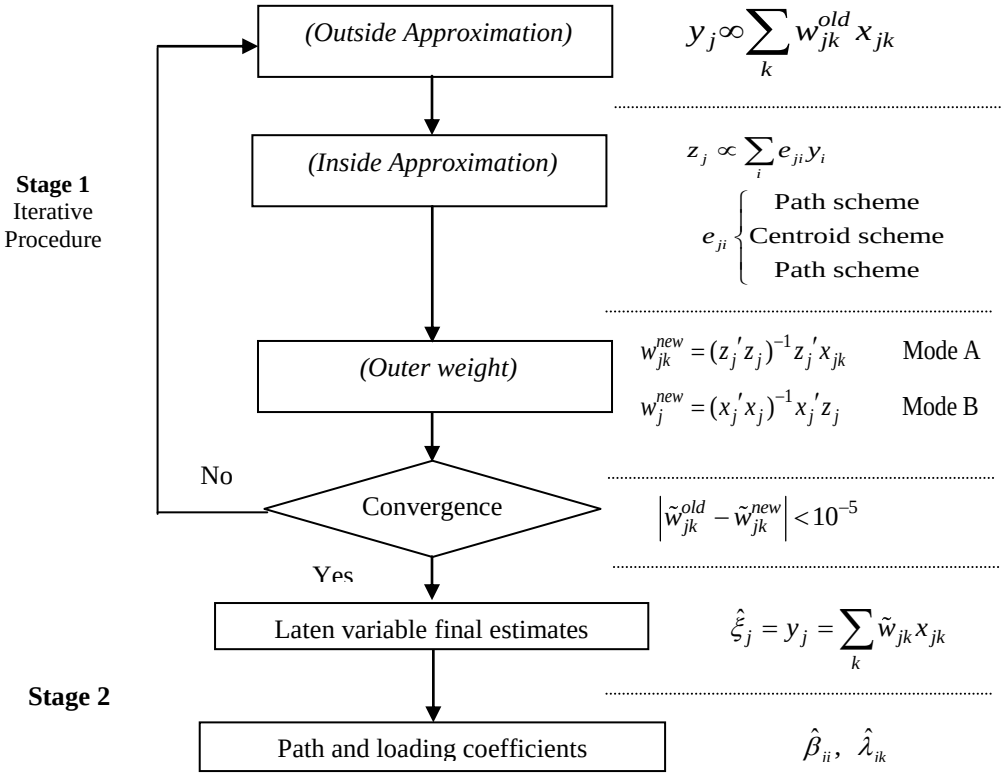

FIGURE 1. PLS ALGORITHM

Evaluation of PLS models is testing the validity and reliability. Validity testing performed to see the value of loading factor produced more than 0.5, if there are indicators has value loading factor below 0.5, then the are excluded from the analysis. Reliability testing show the composite value of each latent variable [6].

B. Propensity Score

The advantage of propensity score in comparison to multivariable adjustment is the separation of confounding factors adjustment and analysis of the treatment effect [7]. If the vector has many covariates were presented in many dimensions, then the propensity score can reduce all the dimensions into one dimension scores [2]. Rosenbaum and Rubin defined the propensity score for $i = 1, 2, \cdots, n$ as a conditional probability of being treated. Indicator of treatment group $(Z_i = 1)$ and control unit $(Z_i = 0)$ based on observed covariates vector (ξ_i). Propensity score can be written mathematically as follows:

$$e(\xi_i) = \mathrm{P}(Z_i = 1 | \xi_i) \quad (1)$$

The goal of propensity scoring is to balance the treated and untreated groups on the confounding factors that affect both the treatment assignment and the outcome, thus it is important to verify that treated and untreated patients with similar propensity score values are balanced on the factors included in the propensity score. Demonstrating that the propensity score achieves balance is more important than showing that the propensity score model has good discrimination [8].

C. Marginal Mean Weigthing through Stratification (MMWS)

Marginal mean weighting through stratification (MMW-S) was introduced as a flexible approach, combining propensity score weighting and propensity score stratification to remove imbalances of pre-intervention characteristics between two or more groups [3]. MMW-S produces more robust analysis than the methods of propensity score matching, propensity score stratification, and propensity score weighting [9].

$$MMWS = \frac{n_q \times \Pr(Z = z)}{n_{z=z,q}} \quad (2)$$

Where

n_q = Number of individuals in each stratum

$\Pr(Z=z)$ = Probability of the treatment group

$n_{z=z,q}$ = Number of individuals in each stratum is treated as treated / non-treated

D. HIV/AIDS

Human Immunodeficiency Virus (HIV) is a virus that reduces the body's immune system so that the people affected by this virus will be susceptible to various infections and then causes *Acquired Immune Deficiency Syndrome* (AIDS) [10]. The HIV is decreases gradually the immune system and leads to death as a direct result of one or more opportunistic infections. Opportunistic infection is an infection caused by immune deficiencies as a result of the HIV. Factors that influence the Opportunistic infection is predisposition and clinical manifestation. Predisposition is the internal factors that exist in individuals, families, communities that make easier individuals to behave. Clinical manifestations is presence indication of a disease that is perceived as complaints from patients and has been examined by a doctor or clinic. Predisposition include of age, level of education, work and marital status. And clinical manifestation include of CD4 and clinical stage.

III. METHODOLOGY

A. Source of Data

The data used in this research is secondary data on the medical records of HIV / AIDS patients in one hospital. The number is 91 patients HIV/AIDS. By using several variables as follows:

1. Exogenous Variables
   a Predisposition : age, level of education, work and marital status,
   b. Clinical manifestation : CD4 and clinical stage
2. Endogenous Variables: Opportunistic infection
3. Confounding variables: Adherence therapy ARV

B. Method of Analysis

Based on the research objectives, analysis methods used in this study is

1. Select confounding variables
2. Determine the propensity score approach to SEM
   a. Develop the conceptual model based on the theory
   b. Construct the path diagram
   c. Convert the path diagram into an equation system
   d. Estimate the parameters of model included
   e. Get the path coefficient value
   f. Determine the e propensity score value
3. Divide sample into Q strata based on propensity score and calculate the marginal mean weight

4. Examine the balancing of the covariates
5. Determine percentage bias reduction (PBR)

IV. ANALYSIS AND DISCUSION

A. *Select confounding variable*

Confounding variable according to the epidemiology is a situation where the size of the effect of distorted risk factors because of the correlation between exposure and other factors that influence the results [11]. The actual relationship between exposure factors and impact /disease factors are disappear or covered by other factors, so the influence of confounding factors can increase or decrease the actual relationship. Chi-square test was used to examine the relationship among variables, the following hypotheses [12]:

H_0: There is no significant relationship among variables

H_1: There is a significant relationship among variables

Significance level: $\alpha = 5\%$

Critical region: reject H_0 if $\chi^2 > \chi^2_{1-\alpha}; df = (i-1)(j-1)$ or p-value < α

TABLE 1. RELATIONSHIP BETWEEN ADHERENCE WITH PREDISPOSISI & CLINICAL MANIFESTASI

Variable	χ^2	*P-value*	Decision
ADH*Predisposition	5.315	0,021	reject H_0
ADH*Clinical manifestation	7.662	0,006	reject H_0

Based on the Table 1, can be seen that the adherence has a relationship with the predisposition and clinical manifestation. Therefore variable adherence is confounding variables. Diagram path after the formed variable interactions from compliance with adherence ARV the relationship among predisposition variables and adherence ARV the relationship among clinical manifestations of the opportunistic infection can be seen at figure 2.

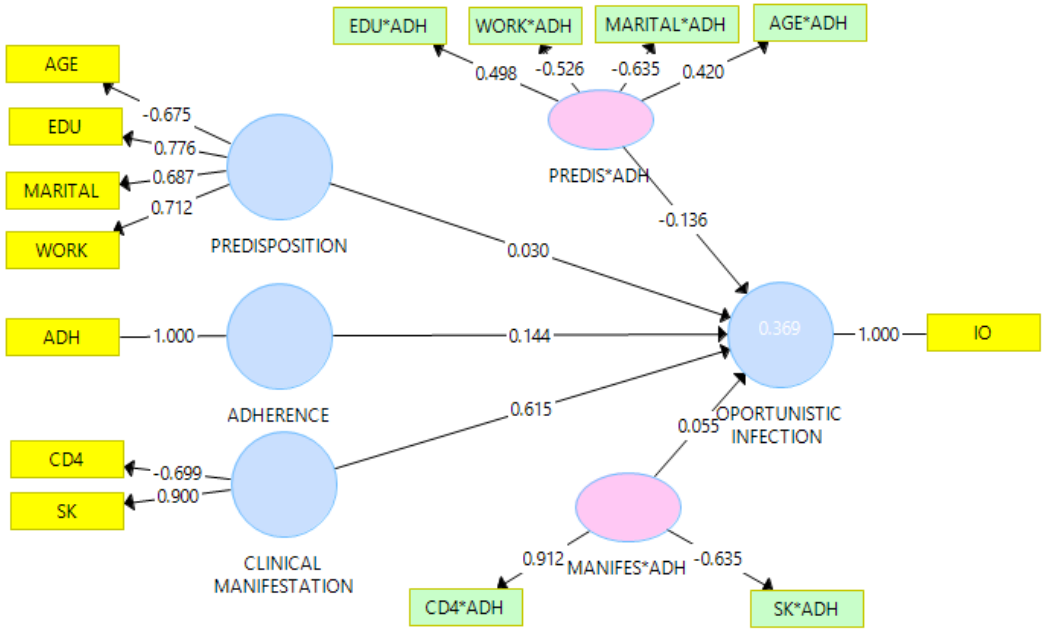

FIGURE 2. DIAGRAM PATH VARIABLE CONFOUNDING

Based on the figure 2 diagram path after putting confounding variables, the structural equation model is:

Opportunistic Infection = 0.030 predisposition + 0.144 clinical manifestation + 0.615 adherence
- 0.136 (Predis*adh) + 0.050 (manifes*adh)

B. Calculating of MMWS

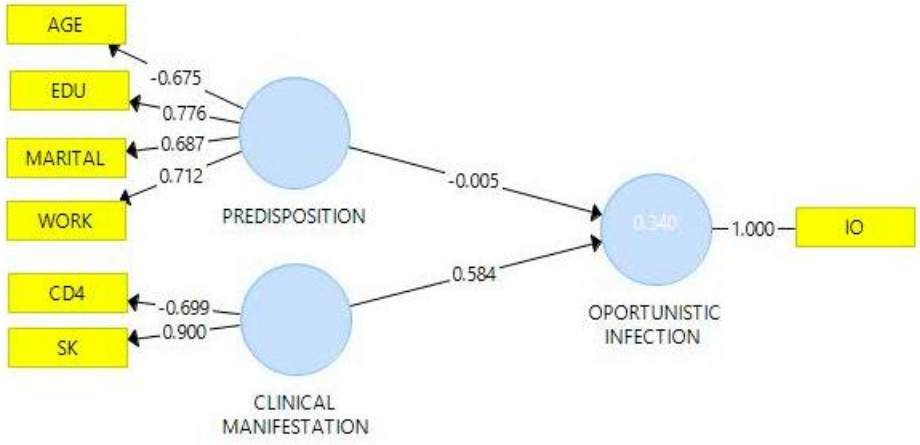

FIGURE 3 : LOADING FACTOR OF EACH LATEN VARIABLE

Based on the figure 2, can be seen that the loading factor of each indicator more than 0.5. Hence can be concluded that a valid indicator to measure the construct predisposition and clinical manifestation. Then calculate the propensity score using SEM-PLS. The propensity score for all respondents are used to divide respondents into five strata. Furthermore, calculating the marginal mean weight to tread groups and untreated groups uses the equation recommended as follows [9].

$$\frac{n_s \times \Pr(Z = z)}{n_{z=z,s}} \tag{3}$$

Where, n_s is the total number of individuals in stratum s, $\Pr(Z = z)$ is assignment probability to treatment groups z. $n_{z=z,s}$ is the total number of individuals in stratum s which is the actual treatment assignment for z.

TABEL 2. THE CALCULATION OF THE MMWS

Stratum	Sample	Unweighted sample		MMWS		Weighted sample	
		Treated	Untreated	z	z'	z	z'
1	19	10	9	0.626	1.42	6	13
2	18	6	12	0.989	1.01	6	12
3	18	4	14	1.484	0.86	6	12
4	18	8	10	0.742	1.21	6	12
5	18	2	16	2.967	0.75	6	12

After calculate MMWS can increase the homogeneity of propensity score between the treatment group and the control group in each stratum. The homogeneity or balance covariates in each stratum using the t-statistic for numerical variables. Insignificant T-values indicate adequate MMWS. The results of cheking balance covariate of each stratum are presented in Table 3.

TABLE 3: *T*-TEST FOR CHECKING BALANCE

Stratum	Predisposition			Clinical Manifestation		
	***T*-value**	$T_{(df,\alpha)}$	**Decision**	***T*-value**	$T_{(df,\alpha)}$	**Decision**
1	0.390	6.314	Not reject H_0	0.911	1.812	Not reject H_0

2	0.298	1.655	Not reject H_0	0.495	1.652	Not reject H_0
3	1.196	6.314	Not reject H_0	0.367	1.703	Not reject H_0
4	1.582	6.314	Not reject H_0	0.747	1.648	Not reject H_0
5	1.116	1.687	Not reject H_0	1.216	1.653	Not reject H_0

Based Table 3 shows that after MMWS there is no difference between the treatment group and control. Furthermore, compute a percentage bias reduction (PBR) on the covariate is another criterion to assess the effectiv of MMWS.

$$PBR = 100 \times \frac{\left(\bar{x}_t - \bar{x}_c\right)_{beforeMMWS} - \left(\bar{x}_t - \bar{x}_c\right)_{AfterMMWS}}{\left(\bar{x}_t - \bar{x}_c\right)_{AfterMMWS}} \quad (4)$$

Based on calculate of percentage bias reduction (PBR) obtained 93.5%. So that MMWS able to reduction bias 93.5%. It this sufficient of the bias reduction based on the examples in Cochran and Rubin PBR value of 80% or higher is satisfactory.

V. CONCLUSION

SEM-PLS can see that loading factor for each indicator on each latent variable is greater than 0.5, so that the indicators (age, level of education, work and marital status) were able to explain the predisposition variable and indicators (CD4, clinical stage) capable explain the clinical manifestation variable. Furthermore, score factor of each of the latent variables used to calculate a propensity score that will be used at this stage of Marginal mean through weighting stratification (MMWS) to reduce bias due to confounding variable. Marginal mean weighting through stratification (MMWS) method is a powerful because MMWS showed a reduction from the selection bias more than 93.5%.

.

REFERENCES

[1] Rosenbaum, P.R., & Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Journal Biometrika, vol.70, No.1, pp 41-55J*.

[2] Guo, S. & Fraser, M. W. (2010). Propensity score analysis: Statistical methods and applications. *Thousand Oaks, CA: Sage Publications*

[3] Linden, Ariel. (2014). Combining propensity score-based stratification and weighting to improve causal inference in the evaluation of health care interventions. *Journal of Evaluation Clinical Practice 20, 1065-1071.*

[4] Ghozali, Imam. (2011). *Structural Equation Modelling Metode Alternatif dengan Partial Least Square*. Badan Penerbit Universitas Diponegoro, Semarang

[5] Trujillo, G.S. (2009). *PATHMOX Approach: Segmentation Trees in Partial Least Squares Path Modeling*. Universitat Politècnica de Catalunya.

[6] Chin, W.W. (1998). *The Partial Least Squares Approach for Structural Equation Modelling. Modern Method for Business Research*. London: Lawrence Erlbaum Associates.

[7] Littnerova Simona, Jarkovsky Jiri, Parenica Jirib, Pavlik Tomas, Spinar Jindric, Dusek Ladislav. (2003) Why to use propensity score in observational studies? Case study based on data from the Czech clinical data base AHEAD 2006–09. *Original Research Article. Coret Vasa 55. pp. e 383 – e 390*

[8] Crowson Cynthia S. Schenck Louis A., Green Abigail B., Atkinson Elizabeth J., Therneau. (2013). *Terry M The Basics of Propensity Scoring and Marginal Structural Models*. Mayo Clinic, Rochester, Minnesota

[9] Hong, G. (2010). Marginal mean weighting through stratification: Adjustment for selection bias in multi-level data. *Journal of Educational and Behavioral Statistics, 35 (5), 499-531*.

[10] Djauzi. S. dan Djoerban, Z. (2003). *Penatalaksanaan Infeksi HIV di Pelayanan Kesehatan Dasar. Edisi II*. Jakarta: Balai Penerbit FK UI; 2003.

[11] Wunsch, Guillaume. (2007). Confounding and control. *Demografi Research Volume 16. Article 4, page 97-120* Published 06 Februari 2007

[12] Daniel, W. W. (1978). *Statistik Nonparametrik Terapan.* Jakarta: Gramedia.