Proceedings of the Eastern Asia Society for Transportation Studies, Vol.9, 2013

Reliability and Stability on the Numerical Analysis in Structural Equation Modeling

Borith LONG [a], Tatsuya MEGURO [b], Kasem CHOOCHARUKUL [c],
Kunihiro KISHI [d], Takashi NAKATSUJI [e]

[a] *Graduate Student, Graduate School of Engineering, Hokkaido University, Sapporo, 060-8628, Japan*
[a] *E-mail: borith_long@hotmail.com*
[b] *Chiba Operation office, Kanto Regional Head office, East Nippon Expressway Company Limited, Chiba, 263-0001, Japan*
[b] *E-mail: t.meguro.ab@e-nexco.co.jp*
[c] *Associate Professor, Department of Civil Engineering, Chulalongkorn University, Bankgok, 10330, Thailand*
[c] *E-mail: kasem.c@eng.chula.ac.th*
[d] *Associate Professor, Graduate School of Engineering, Hokkaido University, Sapporo, 060-8628, Japan*
[d] *E-mail: kishi@eng.hokudai.ac.jp*
[e] *Professor, Graduate School of Engineering, Hokkaido University, Sapporo, 060-8628, Japan*
[e] *E-mail: naka@eng.hokudai.ac.jp*

Abstract: Structural equation modeling has been widely used in various research studies and has been incorporated into many software programs such as Mplus, CALIS, EQS, LISREL, Amos and so on. At the meantime, the diversity, instability of the analysis results or no solutions has been found in some analysis models. Several problems on the numerical analysis such as the constraints of residual variance, initial values and the different solutions by software in use have also been found in SEM analyses. This paper discusses the reliability and stability on the numerical analysis in structural equation modeling by using four kinds of SEM programs and three kinds of sample data. The study will introduce an application of the optimized calculation of genetic algorithms (GA) in structural equation modeling in order to see elaborately what is going on with these issues and also to examine the goodness-of-fit, validity, stability and reliability of structural model. Furthermore, the empirical analysis is presented to discuss the above issue in the questionnaire survey data about drunken driving behavior in Bangkok.

Keywords: structural equation modeling, genetic algorithm, solver, Amos

1. INTRODUCTION

Structural equation modeling (SEM) is a methodology for representing, estimating, and testing a network of relationships between variables (measured variables and latent variables). SEM is called multivariate analysis with latent variables and also called causal modeling or covariance structure analysis. SEM is a valuable methodological tool that has gained popularity across many disciplines in the past two decades perhaps due to its generality and flexibility (Golob, 2003). Essentially the broad framework that includes many well-known procedures such as multiple linear regressions, factor analysis, path analysis; structural equation modeling allows for analysis of causal patterns among unobserved variables represented by multiple measures. It permits testing of causal hypotheses and theory, examination of psychometric, enhancement of the explanatory power of correlational data, extension of theory and so on.

An SEM has two primary components: the measurement model and the structural model. The measurement model describes the relationships between observed variables and the construct or latent variables are hypothesized to measure. In contrast, the structural model describes interrelationships among constructs. When the measurement model and the structural model are considered together, the model may be called the composite or full structural model (Schreiber, 2008). Figure1 shows a basic example of component in structural equation modeling.

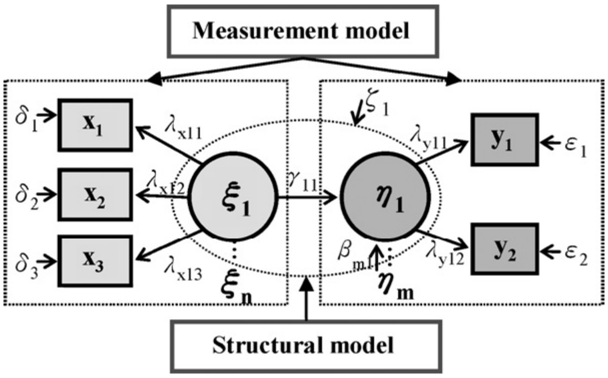


Figure 1. A basic example of SEM component (Lee *et al.* 2008)

Most SEM analyses were conducted using one of the specialized SEM software packages, such as Mplus, CALIS, EQS, LISREL, and Amos and so on. However, there are many options and the choice is not always easy. As an example, Amos was one of the commonly used programs for SEM analysis. Byrne (2001) conducted a study about the comparison of 3 SEM computer programs AMOS, EQS and LISREL. The comparisons focused on key aspects of the programs that bear on the specification and testing of CFA models, preliminary analysis of data, and model specification, estimation, assessment, and misspecification. In Germany, Nachtigall *et al.,* (2003) summarized the efficiency and useful aid of using three SEM programs LISREL, AMOS and EQS. The study proposed the use of LISREL if users have different skill levels and are not sure about the right program. AMOS or EQS are used for an easy way among other SEM program with the risk of understanding the methodological complexity. Lei and Wu (2007) recommended that researchers should consult with software package publishers for more detailed information and current developments before analyzing the model.

From the author knowledge, several curious phenomena have been found in the calculation process of SEM: the residual variance which is a subject on the numerical analysis in the covariance structure analysis could become a negative value or the estimated weight of a path coefficient was heavily fluctuated by constraints and if the positive/negative of an initial value is changed, the mark of the estimated weight of a path coefficient would be reversed. The diversity of solutions results or no solution results are also found with the different software packages. Similarly, Toyoda (1998) and Kojima (2003) pointed out that the value residual variance was negative and the estimated the path coefficient was changed by giving the constraints. Meguro *et al.*, (2012) indicated although the optimized calculation was converged but the solution was still not valid since the estimated coefficient was unacceptable. Toyoda (1998, 2003) also found another problem in numerical analysis that was about initial value. It is known that there are convergence and un-convergence and estimated coefficient changed by an initial value. Although the setting method of the initial value was proposed, initial value problem still could not be solved.

From the literature review, it can be seen that the problems on numerical analysis such as the constraints on residual variance, initial value, software program use, diversity and instability of solution, goodness-of-fit of model are not yet be solved. It is very important to examine and discuss in more detail on these issues.

The objective of this paper is to presents the reliability and stability on the numerical analysis in structural equation modeling and to deal with the problem on the numerical analysis such as the constraint of residual variance, initial value and the unstable and diversity of solutions results by introducing an application of the optimized calculation of genetic algorithms (GA) in structural equation modeling in order to (1) see in more detail what is happening in the neighborhood of the global minimum point, and examine the goodness-of-fit, validity, stability and reliability of structural model. (2) Also present a systematic procedure about what we have to keep in mind when applying SEM. (3) finally, empirical analysis of drunk driver.

2. ANALYSIS METHODS

2.1 Analysis Programs

In order to improve the reliability of the analysis, four kinds of the SEM software programs are used as the analytical methods in this study.

(1) Amos 18.0 is denoted as P1.

Amos (analysis of moment structure) version 18 by Arbuckle (2011) is distributed with SPSS. It has two components: Amos Graphics and Amos Basic. Amos Graphics permits the specification of models by diagram drawing whereas Amos Basic allows the specification from equation statements. An alternative full-information maximum likelihood estimation method for missing data is also available in Amos (Lei and Wu, 2007).

(2) SEM software developed by Kojima (2003) is denoted as P2.

This SEM software is called K-Solver. Its numerical computation is calculated by using "Solver" function which is one of the add-in functions in Microsoft Excel, 2003.

(3) SEM program developed by author is denoted as P3.

We have developed the SEM program in Visual Basic Application (VBA) by using the Solver function in Microsoft Excel, 2010. This program is more flexible; it is able to set any initial value and any constraints of residual variance and estimation weight.

(4) Optimization program by Genetic Algorithms (GA) which is denoted as P4.

Extending from SEM program P3, we have developed a new SEM program by introducing the application of optimization by GA based on GENECOP III concept (Michalewicz, 1992) in order to improve the solution at the global minimum.

2.2 Data Use

Three samples data are used in this study, the first and the second sample data are getting from the book of covariance structure analysis-structural equation modeling written by Toyoda (1998). The first sample data is an exercise of covariance structure analysis which investigated the image of respondents' hobby which is affected by themselves and their family. Also the second data is an exercise which investigated the influence of the income, education on the social status and also the influence of social status to the social environment. And the third data is the empirical data, questionnaire survey data which was getting from the questionnaire survey of the drunk driving behavior in Bangkok.

2.3 Analytical Methods

In the four SEM analyses program (P1, P2, P3 and P4); the numerical calculation is calculated by the Maximum likelihood method and the least-squares method. To be more concrete, when assessing population-level data- model fit for models with full and reduce sets of parameters, one starts with fit function associated with the desired method of estimation. The maximum likelihood (ML) approach will estimate by minimizing the fit function $f(\boldsymbol{\theta})$. All the solutions in this paper are based on Maximum likelihood method.

$$f(\boldsymbol{\theta}) = (N-1)\left[tr(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{S}) - \ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{S}| - n_x\right]$$

Where, N: number of sample data, S: the sample covariance matrix for the observed data, n_x: number observed variable, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$: population covariance matrix implied by the model with parameters $\boldsymbol{\theta}$.

Genetic Algorithms (GA) is also applied in our analysis which is included in numerical calculation of the SEM program P4. The genetic algorithms based optimization method is used to calibrate the model parameters so that the model produces minimum error in the estimation of the variable. GA is an adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetic. The basic concept of GA is designed to simulate process in natural system necessary for evolution. The idea with GA is to use this power of evolution to solve optimization problems. GA has been widely studied, experimented and applied in many fields in engineering. GA provides an alternative method to solving problem, to finding optimal parameter which might prove difficult for other methods. Its usefulness and gracefulness of solving problems has made it the more favorite choice among the other methods, namely gradient search, random search and others. The concept of GENECOP which stands for genetic algorithm for numerical optimization for constrained problem was also used. This concept provides a way of handling constraints that is both general and problem independent. Based on the concept of GENECOP III, a percentile error function is used as the objective function to be minimized (Michaelwicz, 1992).

Regarding to the fit of the model, many criteria have been developed for assessing overall goodness-of-fit of an SEM and measuring how well one model does versus another model. Most of the evaluation criteria are based on the chi-square statistic given by the product of the optimized fitting function and the sample size (Golob, 2003). One rule of thumb for good fit is that the chi-square should be less than two or three times its degrees of freedom (Golob, 2003). Goodness-of-fit measures for a single model based on chi-square values include root mean square error of approximation (RMSEA) which measures the discrepancy per degree of freedom. It is generally accepted that the value of RMSEA for a good model should be less than 0.05 (Browne and Cudeck, 1992). MacCallum *et al.* (1996) recommends that the entire 90% confidence interval for RMSEA should be less than 0.05. But Byrne (2009) accepted that RMSEA, the obtained value less than 0.05 indicate good fit; those ranging from 0.08 to 0.10 indicate mediocre fit and those greater than 0.10 indicate poor fit. For several goodness-of-fit indices, baseline comparison such as normed fit index (NFI), comparative fit index (CFI), a rule of thumb for most of the indices is that a good model should exhibit a value greater than 0.90 (Bentler, 1990; McDonald and Marsh, 1990). But Byrne (2009) accepted that the recommended acceptance of a good fit to a model requires the obtained NFI, CFI value should be in range from zero to one.

To improve the fit of model, the optimized calculation such as nonlinear optimization of genetic algorithms (GA) was examined and discussed in this paper.

3. PROBLEMS ON THE NUMERICAL ANALYIS IN COVARIANCE STRUCTURE ANALYSIS

As from the literature review (Toyoda, 1998 and 2003; Kojima, 2003) and the author knowledge, several problems such as the constraints of the residual, initial value and diversity of solution have been found in SEM numerical analysis. These issues will be proved and analyzed more detail in this article. At the meantime, the genetic algorithm (GA) which is the optimized calculation is also applied in SEM analysis in order to analyze more deeply about the diversity and uniqueness of the solution and to improve goodness-of-fit of models.

3.1 Constraints of Residual Variance

From the analysis result using the first sample data by running in the Amos (P1) and K-Solver (P2) without applying any constraints, it is found that the solutions of the estimated coefficient are almost the same and the residual variance have a negative value. From the figure 2, it can be seen that the residual variance (e2 and e6) of the observed variable V2 and V6 are found to be negative. It can be concluded from this analysis result that this solution is not valid as the value of residual covariance is negative. In this case, setting the constraints of residual variances is required. However, the Amos and K-Solver program are not able to deal with the constraints of the residual variance; therefore SEM program (P3) will be used in the analysis.

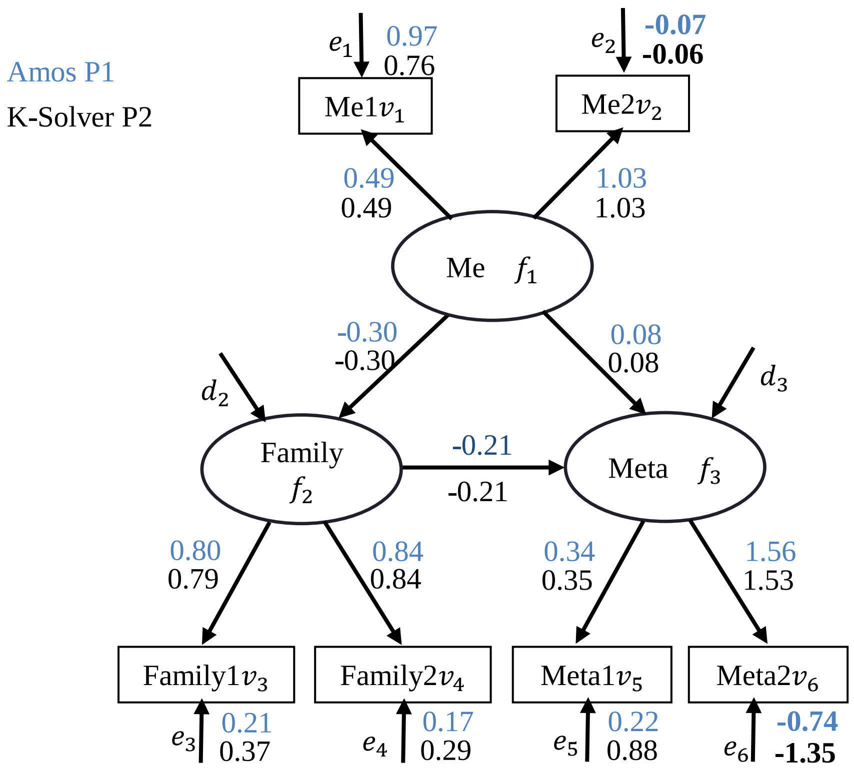


Figure 2. Comparison of the analysis results between SEM program P1 and P2

By setting the constraint of residual into the previous problem, given in advance the path coefficient and the residual variance of initial value are 0.5 and path coefficient is a standardized estimation and residual variance is a non-standardized estimation. It is shown that the difference of solution is found not only on the residual variance but also on the coefficient of path matrix. It can be observed that the estimated coefficient of the path matrix getting from SEM program (P3) is also changed comparing to P2. The estimated coefficients of the path matrix from both solutions of Amos without constraints (P1) and SEM program (P3) with constraints are not much differences and the solution of residual variance using P3 program is found to be positive. From these solutions, it can be said that the solution using P3 program is valid given the constraints setting and the residual variance is positive. It can be deducted that the restriction of the constraints should be considered to get more reliable solution. The results in figure 2 and 3 suggested that we could have unexpected solutions if we applied Amos (P1) and K-solver (P2) without much attention to the constraints of residual.

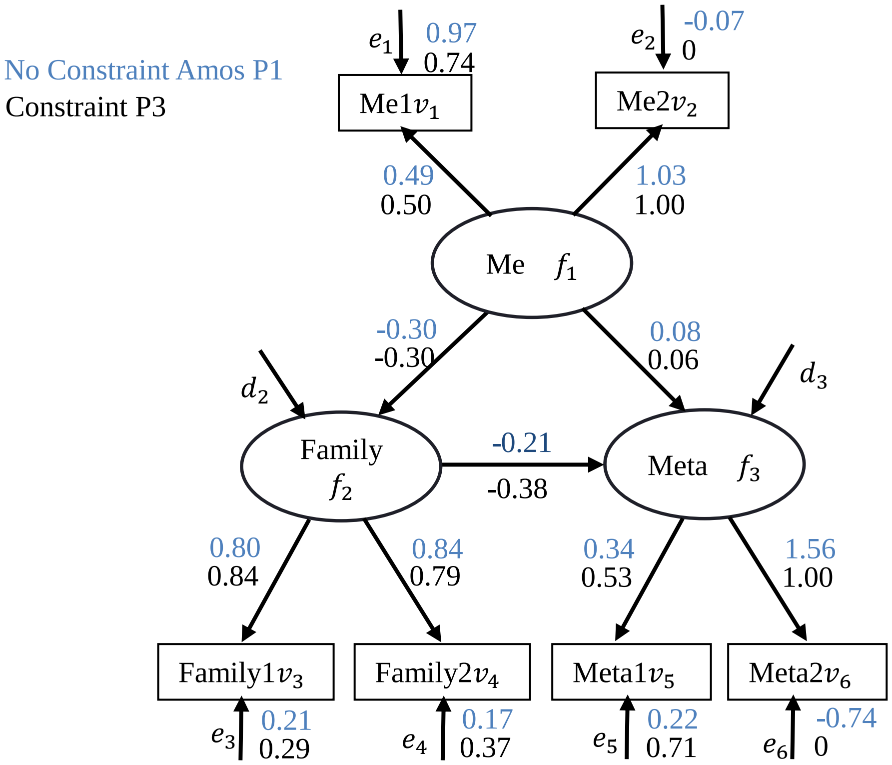


Figure 3. Comparison of the analysis results between SEM program P1 and P3

3.2 Initial Value

The initial value is also found as one of problem in SEM numerical analysis which often occurred. This analysis also uses the first sample data and SEM program P3 with two different input of initial value (shown in Table 1). The initial value A is given as the constant number with the value of 0.5 in both path coefficient and residual variance. Whereas the initial value B, the initial value of path coefficient is given as a random number between -1 to 1 and the initial value of residual variance is also a random number in the range of 0 to 1. It can be assumed that these initial values are the range of value which may be drawn to a solution with appropriate initial value.

Table 1. Value of the initial value A and B

	Path coefficient	Residual variance
Initial Value A	0.5	0.5
Initial Value B	Random -1~1	Random 0~1

It is observed that the estimated coefficient of the path matrix is always reversed, when the positive or negative of the initial value is changed. From the analysis result in the figure 4, it is found that the absolute value of each estimated coefficients are equal but the sign of some estimated coefficients are reversed, given the difference initial value (constant and random). As from other solution results of other structural models, it is also found that the solutions are reversed given the different initial value; moreover although initial value A is change into -0.5, the same phenomenon is occurred. It can be said that the sign of the estimated coefficient is reversed when the positive or negative of initial value is given. It can be concluded that when the positive solution is assumed, the positive initial value should be given; in contrast if the negative solution in is assumed the negative initial value should be applied. The result in figure 4 suggested that we must pay much attention to the initial value.

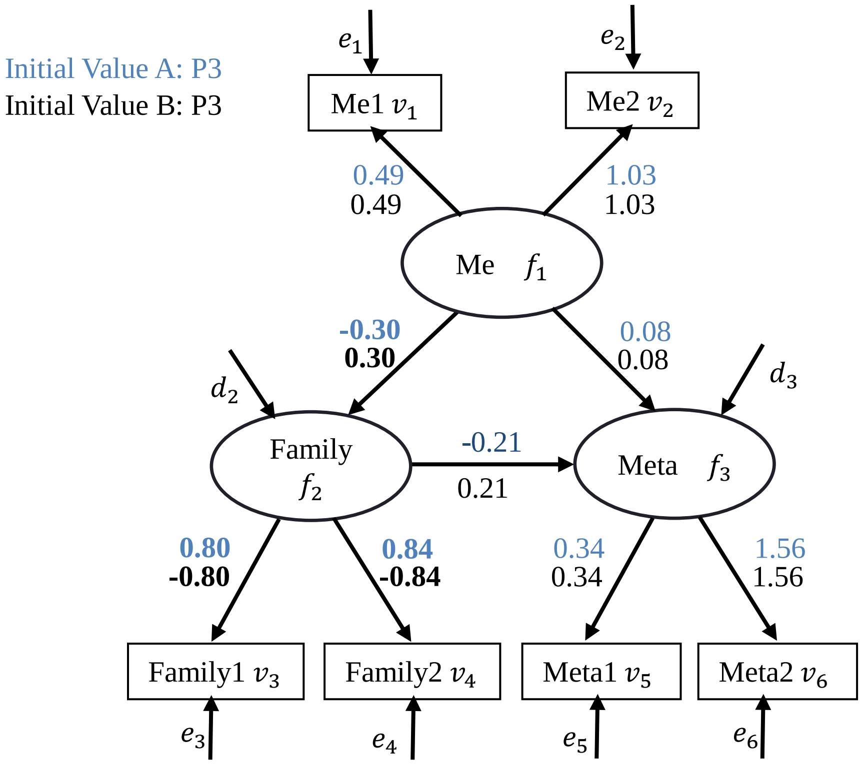


Figure 4. Comparison of the analysis results between initial value A and B

3.3 Optimization Analysis by Genetic Algorithm (GA)

In order to consider the diversity and uniqueness of solution results more deeply, the genetic algorithm (GA) has been applied into SEM in this study. Some operators' value which are used in the process of GA analysis, are set as following: number of uniform mutation 20, boundary mutation 20, non-uniform mutation 40, simple crossover 20, arithmetical crossover 20, heuristic crossover 20 and other parameters are using with the default value of GA. The SEM program P4 which is the developed program associated with GA will be used in the analysis with the second sample data. The further analysis was conducted under the various setting of conditions using in the model as show in figure 5. The analysis results obtaining from P3 and P4 will be compared and discussed. It was found that even though having constraints (P3) or no constraints (P1 or P2), the value of objective functions are still wider comparing to the number of observed variables hence the solutions are found unreliable; in other words the solver solution had stopped at a local minimum. Therefore, the SEM program (P4) incorporated with GA and estimated by maximum likelihood method was proposed. The population parameter was set to 100 and the number of repetition was 10,000 times and the constraints and initial value were set as shown in table 2 above.

Table 2. Value of the constraints and initial value

	Path coefficient	Residual variance diagonal	Residual variance non-diagonal
Constraints	-1~1	0~1	-1~1
Initial Value	0.5	0.5	0.5

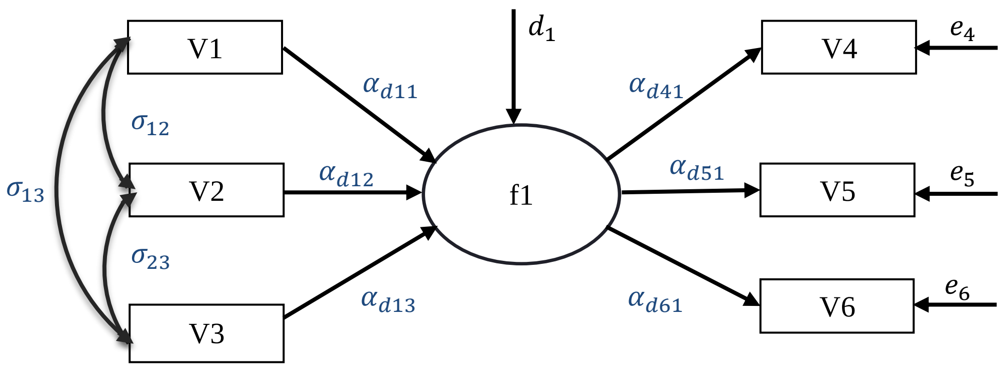


Figure 5. Structural model of second sample data

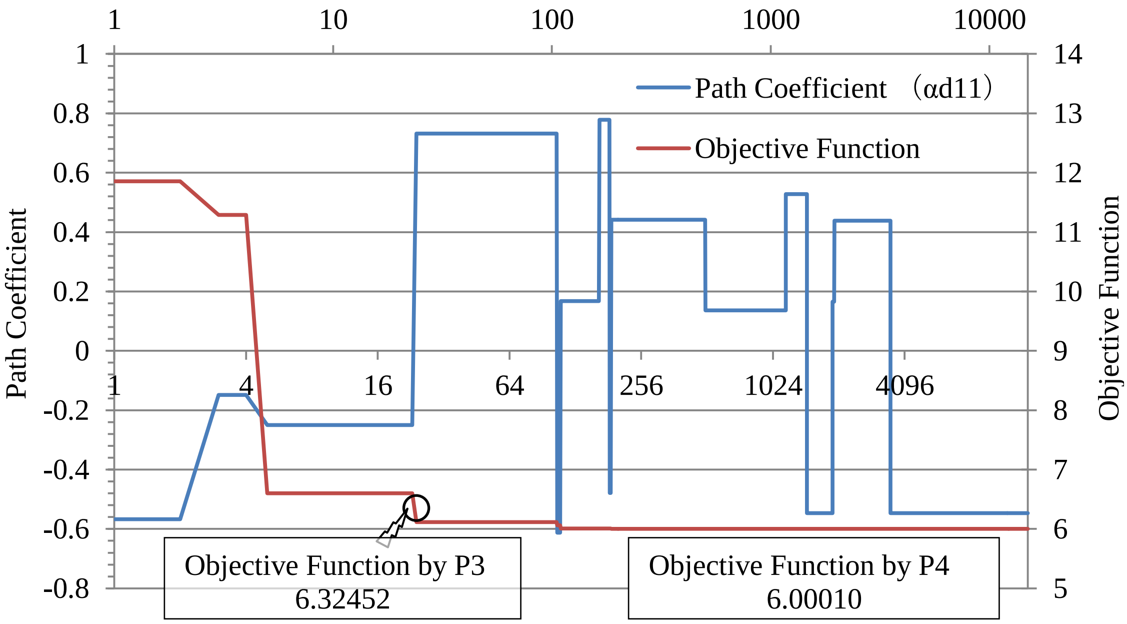


Figure 6. Graph of objective function and path coefficient

Figure 6 represents the graph showing how the value of path coefficient α_{d11} and the objective function changed with the number of iteration. The horizontal axis presents the number of time of calculation with a logarithm scale and the vertical axis on the left hand side shows the path coefficient of α_{d11} and on the right side shows the value of the objective function of the analysis model.

It can be seen from the graph that the objective function is converged after 100 times of iteration. The value of the objective function by the time of the ending calculation of P3 is 6.32452 and 6.00010 for P4. The objective function of P4 is close to the convergence yet the P3 is still far; in other words the objective function of P4 (GA) reached the global minimum but P3 (solver solution) stopped at the local minimum. The value of the path coefficient is changed rapidly hence fluctuation of path coefficient and residual variance are found. Figure 7 shows the comparison of the analysis results between P3 and P4. It can be conclude from the results that solver solution stopped at local minimum and sharply changed the value of path coefficient or few residual variances.

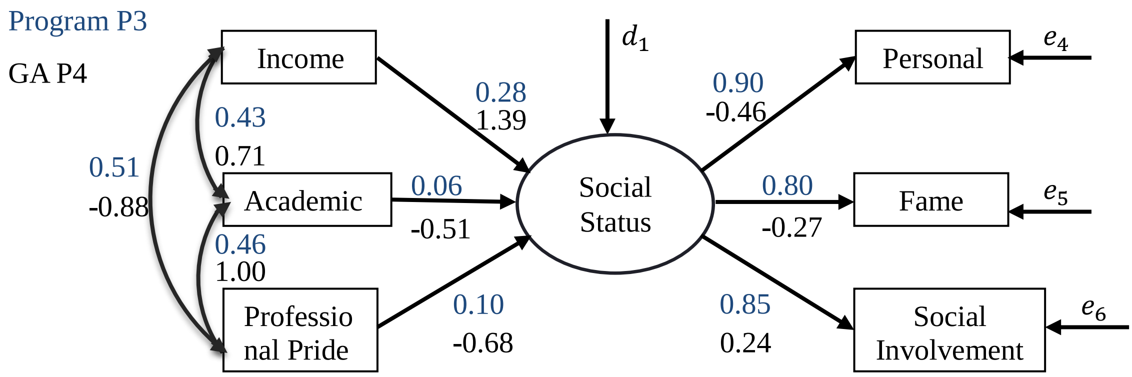


Figure 7. Comparison of the analysis results between P3 and GA P4

3.4 Lessons of the Numerical Analyses

The problem on numerical analysis and the reference of the diversity of solution in SEM were analyzed by using four kinds of SEM programs. It can be noticed that the solution of SEM is affected by the initial value, constraints and the software use. In order to get a valid and reliable analysis result, the key points below should be considered in the numerical analysis.

(1) Initial value

When a positive solution is assumed, the positive value of initial value is required; similarly if the negative solution is assumed, the negative of initial value is given. The random values of initial value either negative or positive are found ineffectiveness (should not be used).

(2) Constraints

The positive value of the constraints (0~1) is given to the diagonal of residual variance. The restriction of the constraints should be considered to get more reliable solution.

(3) GA optimization

GA program is used when the solution is far different from the assumption and also when the goodness-of-fit is found to be not good. When various solutions exist, it should be reconsidered of the sample data and model structure.

3.5 Systematic Approach

Finally, the calculation procedure is summarized into the flow chart as shown in the figure 8 below. For empirical data analysis, at first we have to check the internal consistency (Cronbach's alpha index) to check the homogeneity of questionnaire items.

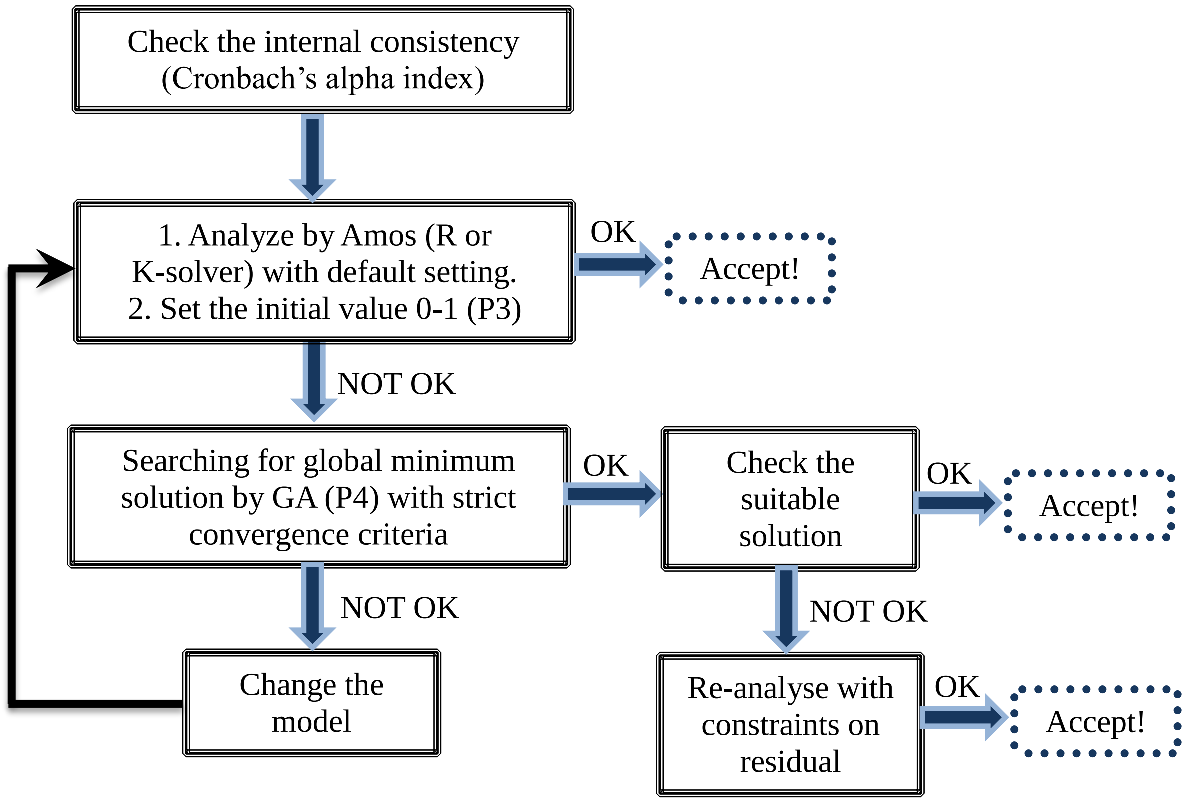


Figure 8. Flow chart of analysis

4. EMPIRICAL ANALYSIS ON DRUNK DRIVING BEHAVIOR

Learning from the problems on numerical analysis above, the empirical analysis will be conducted and the improvement of solution result is highly expected. In this section, the drunk driving behavior which is considered as a serious cause of the traffic accident in developing countries will be analyzed. The report from International Center for Alcohol Policies (ICAP, 2012) shows that among the Southeast Asia countries, Thailand were found to have the highest alcohol consumption country which is in average 7.0 liters per person following by Laos 6.7 liters, Philippine 6.3 liters, and Cambodia 4.7 liters and other countries. Drunk driving is one of the main causes for road accidents in Thailand. In addition, the numbers of road accident due to drunk driving seem to be increasing from 2 percent to 6 percent from 2001 to 2009 (ICAP, 2012). Thailand government and private sectors have tried to implement the campaigns in order to reduce road accidents due to drunk drivers and also have plenty of strong laws and penalties related to drunk driving. However, the number of road accidents due to the drunk driver is still increasing. ICAP also reviewed a study showing that, not only a strong laws but also law enforcement and public awareness, which will results in reduction of victims due to the drunk drivers. The study also mentioned that Japan was one of the great examples of the successful country, which the number of victim due to drunk drivers per year can be reduced as much as 5,000 victims in 2008; although drinking is one of the Japanese's daily life activities (ICAP, 2012). Various factors such as lack of law consciousness, drinking level, lack of safety consciousness and lack of knowledge about drunk driving of drivers will be considered as a psychological factor which related to drunk driving behavior. The factors lead to the high risk of drunk driving of Thai drivers will be revealed in this study.

The empirical data collecting from the drivers about drunk driving behavior in Bangkok will be used. The questionnaire survey about drunk driving was designed on the assumption of SEM analysis and distribute to road users.

4.1 Respondents

Respondents are voluntarily recruited from the passengers, students, sellers, workers who stopped at airport, university, commercial building and working office. The survey was conducted during November 2012 for a total of one week. A total of 106 of questionnaire surveys were distributed and 96 respondents were useable for data analysis.

From our survey, 47.3 percent of the total respondents are female and 52.7 percent are male. About 55.2 percent of respondents are less than 20 years old, 39.6 percent are in range of 21 years old to 40 years old and 5.2 percent are more than 50 years old. It is found that most of respondents are car user which is about 66 percent of the total respondents following by motorcycle 12.8 percent, bicycle 11.7 percent and walking 9.6 percent.

4.2 Measurements

The survey questionnaire consists of two sections. The first section asks the respondents about their socioeconomic including age, gender, income, education and travel characteristic such as travel mode, trip purpose and so on. The second section consists of the question related to drunk driving consciousness such as drinking level, lack of safety consciousness, lack of skill consciousness, lack of knowledge about drunk driving and lack of law consciousness. To avoid errors in measurement, the questionnaire survey is done in double translation, which is from Japanese to Thai and Thai to Japanese. Items used in the questionnaires are measured base on a five-point Likert scale with "strongly disagree" and "strongly agree" at each end point and also the questionnaire with answers.

Drinking level is measured by asking the respondents to choose the answers which is preparing in five levels. Drinking level is measure by asking 3 statements quoted from AUDIT (Thomas *et al.*, 2001): "How often do you have a drink containing alcohol?" (Q01); "How many drinks containing alcohol do you have on a typical day when you are drinking?" (Q02); "How often do you have six or more drinks on one occasion?" (Q03)

Lack of safety consciousness is measured by asking respondents to rate three statements: "I think that driving after little drinking of alcohol is ok" (Q04); "I think that driving for short distance after drinking alcohol is ok" (Q05); "I think that drunk driving is bad and dangerous" (Q06)

As for lack of skill consciousness, three statements were asked. "Despite driving under the influence of alcohol it would not cause accident" (Q07); "I think that the judgment will be lower if driving under the influence of alcohol" (Q08); "I think that the reaction behavior will be lower if driving under the influence of alcohol" (Q09)

Lack of knowledge about drunk driving is measured by three statements: "I think that education about danger of drunk driving is very important" (Q10); "How often you read or watch the report of traffic accident caused by drunk driving in newspaper or television?" (Q11); "Have you ever joined the seminar about drunk driving?" (Q12)

Lack of law consciousness is measured by asking respondents to rate two statements: "I think that the law of drunk driving should be made more severe" (Q13) and "I think that the control of drunk driving should be increased" (Q14)

As for age in the model analysis, the dummy variable was used where 1= 30 years old or more and 0 = 20 years old or less. For education background, 1 is for junior high school, 3 for high school and 5 for colleague or higher.

The homogeneity of the items such as drinking level, lack of safety consciousness, lack of skill consciousness, lack of knowledge about drunk driving and law is evaluated by means of the Cronbach's alpha coefficients (See Table 3). The Cronbach's alpha indicates the overall reliability of a questionnaire and the acceptable value of Cronbach's alpha range from 0.7 to 0.8 whereas the value substantially lower indicates unreliable scale (Wright, 2005). Due to the homogeneity of the items (Cronbach's alpha) is lower than acceptable limit, some items was deleted. As for lack of safety consciousness, lack of skill of consciousness and lack of knowledge about drunk driving, only two statements will be used in the analysis model.

Table 3. Value of Cronbach's alpha

Constructs	Cronbach's alpha
Lack of knowledge about drunk driving	0.811
Drinking level	0.900
Lack of safety consciousness	0.700
Lack of safety consciousness	0.500
Lack of law consciousness	0.840

4.3 Statistical Analysis

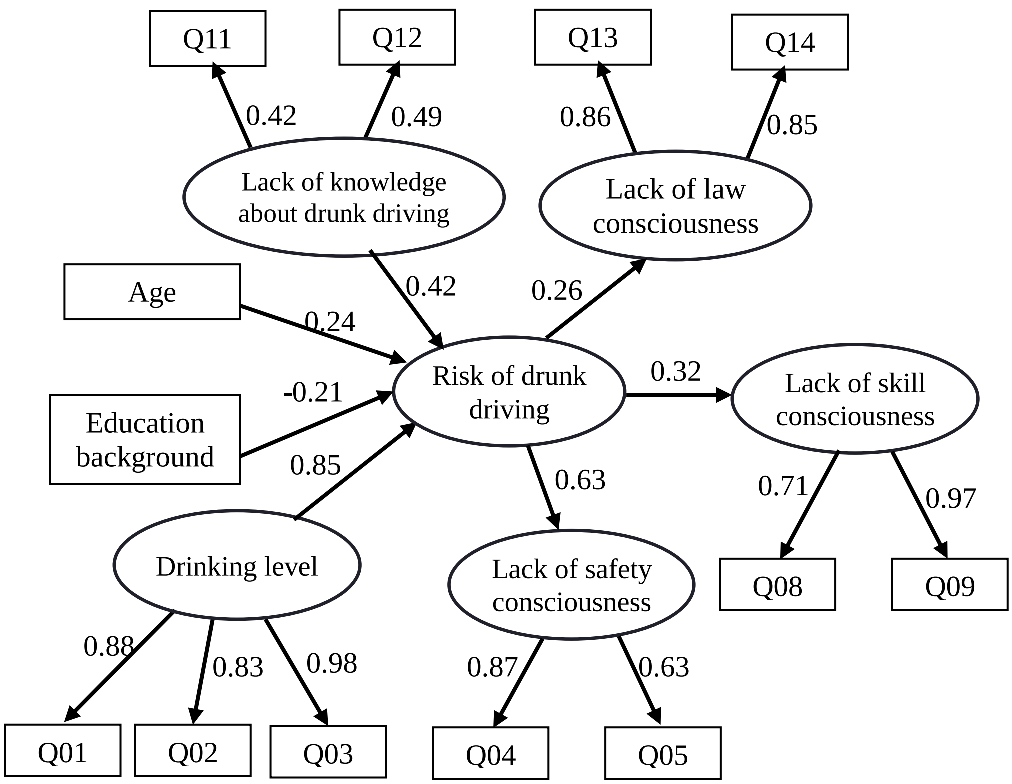
This research estimates the consciousness factors which result in drunk driving by covariance structure analysis (SEM) with the consideration of initial value, constraint and reliability condition that we have raised up in the numerical analysis above. SEM program P2 (K-solver) and P4 (GRG, Evolutionary GA) were used with the estimation of maximum likelihood method. By using the Microsoft Excel 2010, the Generalized Reduced Gradient (GRG) and Evolutionary which is an algorithm for optimizing nonlinear problems are applied. GRG increases the ability of solver more powerful by changing the given initial value/constraints into the unknown. GRG Solver will take longer time to analyze since it will start with the iteration of initial value. It normally stops when the first of three tests is satisfied. The Evolutionary method is like other genetic or evolutionary algorithms which will be able to find a good solution to a reasonably well-scaled model. Because the evolutionary method does not rely on derivative or gradient information, it cannot determine whether a given solution is optimal; so it never really knows when to stop. It knows only that a new candidate solution is "better" than other solutions found earlier. In this study, the positive solution is assumed; therefore the initial value of path coefficient and residual are given as constant value and the constraints was given in range value.

As for the procedure of the calculation, at first the structural model was analyzed by using Amos; but no solution was found. Then re-analysis using the K-solver (P2), after that the verification for some unsuitable solution was found. Next the calculation by P4 (GRG) was applied, after that the goodness-of-fit of model was checked if the solution and goodness-of-fit still not valid, the final analysis using P4 with evolutionary (GA) was conducted through verification of goodness of fit. From this, it is required to be careful and pay more attention on the numerical parameters in the analysis of SEM.

For the fit of the model, the chi-square statistics, degree of freedom, normed fit index (NFI), comparative fit index (CFI), and root mean square error of approximation (RMSEA) were examined. The recommended acceptance of a good fit to a model requires that the obtained NFI, CFI value should be in range from zero to 1 (Byrne, 2009) and for the RMSEA, the obtained values less than 0.05 indicate good fit; those ranging from 0.08 to 0.10 indicate mediocre fit and those greater than 0.10 indicate poor fit (Byrne, 2009).

4.4 Results

It was hypothesized that the drinking level, lack of knowledge about drunk driving and the education background of drivers will have positive influence on the risk of drunk driving and the risk of drunk driving will also have positive influence on driver who have lack of safety consciousness, lack of skill consciousness, and lack of law consciousness. Figure 8 shows the results of the structural model with standardized path coefficients. Overall, this model gives a χ^2value of 80.672 with 60 degrees of freedom. The standardized direct effect on risk of drunk driving are 0.42 for lack of knowledge about drunk driving, 0.24 for age, -0.21 for education background and 0.85 for drinking level.



$\chi 2$=80.672, d.f.= 60, GFI = 0.890, NFI = 0.840, CFI = 0.953, RMSEA = 0.060
Note: Age is a dummy variable, where 1= 30 years old or more and 0 = 20 years old or less.

Figure 9. SEM analysis results of drunk driving behavior

The risk of drunk driving has influence on lack of safety consciousness 0.63, lack of skill consciousness 0.32 and lack of law consciousness 0.20. The goodness-of-fit indicates that this model fits the data well. Specifically, the RMSEA value of 0.060 is lower than the upper limit 0.10 and GFI value of 0.890, NFI value of 0.840 and CFI value of 0.953 are better in range of the cutoff value of 0 to 1. As hypothesized, the drinking level, lack of knowledge about drunk driving and the education background of drivers have a significantly positive influence on the risk of drunk driving and the risk of drunk driving also significantly influent on driver who have lack of safety consciousness, lack of skill consciousness, and lack of law consciousness.

4.5 Discussion and Summary

From the result of the analysis, it is found that the proposed hypothesis is significant. The drivers who have high drinking level, lack of knowledge about drunk driving and the high education background have the high risk of drunk driving and the driver who have high risk of drunk driving result into the lack of safety consciousness, lack of skill consciousness, and lack of law consciousness.

It can be seen that drinking level has the high influence on risk of drunk driving compared to the lack of knowledge about drunk driving, age and education background. This finding shows that drivers who have high drinking level have high risk of drunk driving and also similarly for those drivers who have lack of knowledge about drunk driving. It was found that old drivers have more risk of drunk driving than young drivers. The education background is found to be negatively influence on risk of drunk driving; this mean that the drivers who have high education tend to have more risk than those who have low education background. It is also found that risk of drunk driving has the highest influence on the lack of safety consciousness. This may stem from the respondents' belief that little drinking or a short distance driving after drinking alcohol will not cause any accident; in other words the drivers who have lack of safety consciousness may think driving under the influence of alcohol for a short distance or with little drink are ok.

In term of numerical analysis, it was found that the solution result by SEM program P4 with GRG and Evolutionary (GA) are almost the same. Comparing with K-solver (P2), the solution result by running with GRG and evolutionary (GA) got better objective functions which mean the goodness-of-fit of the model is good.

In conclusion, the present study demonstrates the strategies to reduce the risk of drunk driving behavior of Bangkok driver should be focus on driver's drinking level, drivers' consciousness on knowledge of drunk driving, age, and take some measures to improve the driver's consciousness on driving skill and safety and make law more severe or increase the control of the alcohol checking.

5. CONCLUSION

In this study, some numerical analyses issues associated with SEM were analyzed by using four kinds of SEM software programs. It can be summarized that the consideration of the initial value, constraints, and software program use should be done carefully in order to get a valid and reliable solution. When a positive solution is assumed, the positive value of initial value should be given; similarly if the negative solution is assumed, the negative of initial value is given. The random values of initial value either negative or positive should not be used. The positive value of the constraints (0~1) is given to the diagonal of residual variance. The restriction of the constraints should be considered to get more reliable solution. The optimization calculation method (GRG or evolutionary GA) is used in order improve the goodness-of-fit of the model and to avoid the diversity and instability of solution results. The reliability and stability on the numerical analysis in structural equation modeling is revealed. The key points and method of analyses to get a reliable, valid solution was introduced such as setting constraint of residual and initial value, selecting the appropriate software program, using the optimized calculation GRG nonlinear and evolutionary (GA) method. For the calculation procedure, at first the internal consistency (Cronbach's alpha index), the homogeneity of questionnaire items, have to be checked. After that pre-analysis using Amos (P1) or K-solver (P2) with the default setting was analyzed. If the result was not suitable, the program P3 was used with the setting of initial value and constraints. From this stage, if the solution was still not good, the program P4 with the optimized calculation by GRG or evolutionary (GA) would be used. Furthermore, the empirical analysis from the real data survey of drunk driving in Bangkok was also analyzed. Further study should be considered on the development of nonlinear SEM. It is thought that the human brain, human behavior and their mechanism of thinking are nonlinear. Therefore, the numerical analysis of applying sigmoid function to the latent variable which is considered similar to human brain should be clarified.

REFERENCES

Arbuckle, J.L. (2011) *SPSS Amos 19 user's guide*, Amos Development Corporation.

Bentler, P.M. (1990) Comparative fit indexes in structural models. *Psychometrika 107*, 238–246.

Browne, M.W. and Cudeck, R. (1992) Alternative ways of assessing model fit. *Sociological Methods and Research 21*, 230-258.

Byrne, B. M. (2009) *Structural equation modeling with AMOS: Basic Concepts, Applications, and Programming*, 2nd Edition, Routledge, New York.

Byrne, B. M. (2001) Structural equation modeling with AMOS, EQS, and LISREL: comparative approach to testing for the factorial validity of a measuring instrument, *International journal of testing 1(1)*, 55-86.

Golob, T. F. (2003) Structural equation modeling for travel behavior research, *Transportation Research Part B 13*, 1-25.

Kojima, T. (2003) *Covariance structure analysis: a graphical modeling learning with Excel*, Ohmsha printing, Japan. (in Japanese)

Lee, J.Y.L., Chung, J.H., Son, B. (2008) Analysis of traffic accident size for Korean highway using structural equation models, *Accident Analysis and Prevention 40*, 1955-1963.

Lei, P.W., Wu, Q. (2007) *Introduction to structural equation modeling: issues and practical considerations*, National council on measurement in education.

MacCallum, R.C., Browne. M.W., Sugawara. H.M. (1996) Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods 1*, 130-149.

Meguro, T., Long B., Nakatsuji, T. (2012) Some numerical analysis on structural equation modeling, *Proc. Infrastructure planning*, vol 44.

Michaelwicz, Z. (1992) *Genetic algorithms + data structures=evolution programs*, Spinger, London.

McDonald, R.P. and Marsh, H.W. (1990) Choosing a multivariate model: non-centrality and goodness of fit. *Psychological Bulletin 107*, 247-255.

Nachtigall, C., Kroehne, U., Funke, F., Steyer, R. (2003) Why should we use SEM? Pros and Cons of structural equation modeling, *Methods of psychological research online* *vol 8*, No.2, 1-22.

Scheriber, J.B. (2008) Core reporting practices in structural equation modeling, *Research in social and administrative pharmacy 4*, 83-97.

Thomas, F.B., John, C.H.B., John, B.S., Maristel, G.M. (2001) *The alcohol use disorders identification test*, World health organization, 2nd edition.

Tokitkla, A., Iamoapat, K., Nusit, K. (2012) *Drink driving initiative: situation assessment in Thailand*, International center for alcohol policies (ICAP).

Toyoda, H. (1998) *Covariance structure analysis: structural equation modeling (introductory)*, Asakura printing, Japan. (in Japanese)

Toyoda, H. (2003) *Covariance structure analysis: structural equation modeling*, Asakura printing, Japan. (in Japanese)

Wright., B.D. (2005) *Discovering statistics using SPSS*, SAGE Publication, London.