American Journal of Humanities and Social Sciences Research (AJHSSR)
e-ISSN: 2378-703X
Volume-5, Issue-7, pp-11-19
www.ajhssr.com
Research Paper Open Access

STRUCTURAL EQUATION MODEL (SEM)

[1]AJAYI, Lawrence Boboye Phd , [2]ADEBAYO, Adeyinka Taoheed
[1]Department of Finance Ekiti State University, Ado-Ekiti, Ekiti State.
[2]Department of FinanceEkiti State University, Ado -Ejiti, Ekiti State.

ABSTRACT : This paper critically examined a broad view of Structural Equation Model (SEM) with a view of pointing out direction on how researchers can employ this model to future researches, with specific focus on several traditional multivariate procedures like factor analysis, discriminant analysis, path analysis. This study employed a descriptive survey and historical research design. Data was computed viaDescriptive Statistics, Correlation Coefficient, Reliability. The study concluded that Novice researchers must take care of assumptions and concepts of Structure Equation Modeling, while building a model to check the proposed hypothesis. SEM is more or less an evolving technique in the research, which is expanding to new fields. Moreover, it is providing new insights to researchers for conducting longitudinal investigations.
.

KEYWORDS: *factor analysis, discriminant analysis, path analysis, descriptive statistics, and Structural Equation Model (SEM).*

I. INTRODUCTION

Structural Equation Modeling, or SEM, is a very general statistical modeling technique, which is widely used in the behavioral sciences. The most important reason of the spread of this statistical technique is that the direct and indirect relationships among causal variables can be measured with a single model (Meydan&Şen, 2011). Another reason for the widespread adoption of this method is the ability of taking in to the account of the measurement errors and the relationships between errors in the observed variables (Civelek, 2018). This goes a long way in minimizing measurement errors.

Furthermore, it provides a very general and convenient framework for statistical analysis that includes several traditional multivariate procedures like factor analysis, regression analysis, discriminant analysis, path analysis, to mention but a few.The key in the regression analysis is to determine how much of the change in the dependent variable is explained by the independent variable or variables. Differently from the regression, structural equation modeling, allows to test research hypotheses in a single process by modeling complex relationships among many observed and latent variables. In traditional regression analysis, only direct effects can be detected, but in the method of structural equation modeling, direct and indirect effects are put together (Kline, 2011).

Structural equation modeling is used to test the relationships between observed and latent variables. Observed variables are the measured variables in the data collection process and latent variables are the variables measured by connecting to the observed variables because they cannot be directly measured. According to Civelek (2018), most of the statistical methods other than structural equation modeling try to discover relationships through the data set.However, structural equation modeling confirms the correspondence of the data of the relations in the theoretical model. For this reason, it can be said that structural equation modeling is more suitable for testing the hypothesis than other methods (Karagöz, 2016).

II. ASSUMPTIONS OF STRUCTURAL EQUATION MODELING

Similar to regression analysis, structural equation modeling has its assumptions. But in structural equality models, many regression equations work together, whether in the structural model part or in the measurement model part. Therefore, the assumptions that apply to the regression models are valid for the structural equation models. These assumptions can be summarized as follows (Bayram, 2013);

- **Observed variables have multivariate normality**

The multivariate normal distribution is the most important assumption of the maximum likelihood estimation method used in structural equation modeling. This rule is often violated when ordinal and discrete scales are used. The skewness and kurtosis values are examined to determine whether the variables in the data set are normally distributed. These values are calculated on the basis of moments. In general, the packaged softwarecalculates these values to be 0 as base value. In this case values between -2 and +2 are considered normal. In addition, Kolmogorov-Smirnov and Shapiro-Wilk tests can be conducted to test whether the data set is normally distributed (Sarstedt&Mooi, 2014).

- **Latent variables have multivariate normal distribution**

It refers to the endogenous latent variables have normal distribution. In practice, it is a violated assumption.

- **Linearity**

Linearity, which is the most important assumption of regression analysis, also applies to structural equation modeling. In the structural equation model, it is assumed that there are linear relationships between latent variables and also between observed and latent variables.

- **Absence of outliers**

The outlier affects the significance of the existence model negatively.

- **Multiple measurements**

In the structural equation model, three or more observed variables must be used to measure each latent variable.

- **Absence ofmulti-co-linearity**

It is assumed that there is no relation between the independent variables in the structural equation model.

- **Sample size**

In the structural equation modeling, many of the fit indices are influenced by sample size. In some sources, a minimum sample size of 150 is recommended for structural equation models (Bentler& Chou, 1987). The minimum sample size that should be used in the structural equation modeling method is at least 10 times the number of parameters that can be estimated in the model. (Jayaram, Kannan,&Tan, 2004). According to some researchers, the sample size required for structural equation modeling should be at least 200 and 200 – 500 (Çelik&Yılmaz, 2013).

- **No correlation between error terms**

It is assumed that there is no correlation between error terms in the structural equation modeling method. However, if it is explicitly stated by the researcher in the conceptual model, a correlation can be made between the error terms (Doğan, 2015).

III. VALIDITY AND RELIABILITY ANALYSIS

Reliability means that a scale always measures the same value under the same conditions consistently.Scale is the method used to find the numerical values of the dimensions that constitute a concept. Since concepts cannot be directly measured in social sciences, questionnaires are formed to define these concepts.For example, a questionnaire form is reliable if the same group is given the same result when applied two different times. So if we ask the same questions about the same people, if the conditions are not changed, they are expected to give the same answers. Otherwise, this means that the persons in the sample either they did not understand the questions on the questionnaire or they did not read them.
Validity is a measure of what we really want to measure. For example, if a questionnaire actually measures a different concept than the dimension we want to measure, it is not valid. If the questions we ask about the concept A are confused with the questions about the concept B, then it means that the concepts we consider to measure are not perceived or perceived as different from those in the sample. In this case, the scale we use is not a valid measurement tool for this sample. For this reason, it is necessary to test the validity and reliability of the scale before any analysis is started. As a result of these tests, verification of unidimensionality is generally provided. Unidimensionality means that the observed variables used to measure each dimension must measure only one dimension (Avcılar&Varinli, 2013).
Construct validity and reliability must be determined in order to confirm unidimensionality.The construct validity indicates that the observed variables do not measure any latent variable other than they connected in the conceptual model. But in this case it would not be correct to say that the validity of the construct is fully realized without confirming the reliability of the scale (Gerbing& Anderson, 1988).

1. Determination of Convergent Validity

Convergent validity indicates that the correlations between questions constituting a construct are high. In structural equation modeling method, it is necessary to look at the results of confirmatory factor analysis to determine the convergent validity of the scales used to measure the dimensions constituting the conceptual model of the research. The measurement model part of structural equation models correspond to confirmatory factor analysis (CFA). Therefore, if the measurement model fit indices are low, there is no need to test the structural model.
The t test results of all the coefficients in the measurement model should indicate that the coefficient values are different from zero. The standard value of each coefficient in the measurement model is the factor loadings of the confirmatory factor analysis. Each factor load should be higher than 0.50.Otherwise, the fit indices of the

general model will beadversely affected. The fact that the factor loads are above 0.5 is evidence of convergent validity. If the critical rate value of a question in CFA results is greater than 2 as an absolute value this means that this item is loaded to the factor it is connected.

Before applying confirmatory factor analysis (CFA), it is first necessary to look at the results of explanatory factor analysis (EFA) in practice. Even though scales generally accepted in the literature are used, to see if the survey fillers correctly perceive the questions principle component analysis should be conducted in SPSS before set up CFA model in AMOS. And how many different dimensions the questions are perceived by those who solve the questionnaire should be clarified. At this stage, the necessary questions should be eliminated. This step is also called the purification stage. Principle component analysis is a type of analysis that assigns the variables in the data set into groups so that the relationship between the variables in the group is maximized. Main purpose of this analysis is to obtain the least number of factors to represent the relationship among items at the highest level.

2. Determination of Discriminant Validity

Discriminant validity is the measure of the level at which a structure in a measurement model differs from other structures. It is an indicator of a low correlation between the questions that form a construct and other questions that form other construct. To find the discriminant validity for each dimension, we first need to calculate the Average Variance Extracted (AVE) value for each dimension. The acceptable AVE value must be greater than 0.50 or 0.50. However, this value confirms convergent validity when examined alone (Fornell&Larcker, 1981). In order to determine discriminant validity, it is also desirable that the values of the AVE for each construct in the data set are larger than the correlation coefficients of that construct with the other constructs. In this case, it can be determined that the scales used have discriminant validity for each dimension. AVE value alone does not indicate discriminant validity but the square root of the AVE value of each construct is larger than the inter-dimensional correlation value it can be said that there is discriminate validity (Fornell&Larcker, 1981). The AVE value is not calculated by the AMOS package program. However, it is easy to find ready-made excel files that provide this value calculation on the internet.

3. Determination of Reliability

After determination of the validity of the scales by means of CFA reliability analysis must be conducted for each construct. First of all, Cronbach's α value is calculated for each dimension separately. Values greater than 0.7 threshold indicate that the internal reliability of the scale used is sufficient. Cronbach's α is a measure based on correlations between items in a construct. It is obtained by dividing the sum of the variances of the items constituting a scale by the general variance. It takes a value between 0 and 1. Values beyond 0.7 threshold indicate that the scale is reliable. If it is below 0.6, the reliability of the scale is low (Karagöz, 2016).

Another value that is used to calculate the reliability of the scale for each dimension is the composite reliability value. The composite reliability value is calculated from the factor loads found in the confirmatory factor analysis. After CR values beyond 0.7 threshold or equals to 0.7 it can be said that there is composite reliability (Raykov, 1997).

The Table below shows a sample table showing Cronbach's α, AVE and CR values calculated for each construct and the correlation values between constructs. Cronbach's α value can be calculated from the scale reliability menu in the SPSS program. The AVE and CR values are found by placing the results of the CFA factor loadings in to the formulas. There are readymade calculation tools on the Internet.

Descriptive Statistics, Correlation Coefficient, Reliability Results and Discriminant Validity

	Avr.	Std. Dev	1	2	3	4	5	6	7	8	9
1.Construct	3,25	0,81	(0,842)								
2.Construct	3,28	0,71	,216*	(0,711)							
3.Construct	3,63	0,72	,427*	,383*	(0,840)						
4.Construct	3,72	0,68	,228*	,533*	,457*	(0,718)					
5.Construct	3,62	0,70	,449*	,192*	,378*	,298*	(0,769)				

6.Construct	3,76	0,68	,430*	,394*	,551*	,450*	,499*	(0,734)			
7.Construct	3,23	0,87	,585*	,166*	,452*	,174*	,479*	(0,800)	,449*		
8.Construct	3,68	0,67	,394*	,496*	,672*	,508*	,350*	,508*	,358*	(0,722)	
9.Construct	3,02	0,77	,340*	,374*	,353*	,335*	,209*	,302*	,219*	,410*	(0,754)
Cronbach Alpha Reliability Coefficient			0,927	0,861	0,901	0,851	0,781	0,771	0,828	0,808	0,721
Composite ReliabilityCoefficient (CR)			0,924	0,854	0,905	0,841	0,791	0,777	0,840	0,813	0,725
Avrerage Variance Extracted (AVE)			0,710	0,506	0,706	0,516	0,592	0,539	0,640	0,522	0,570

* $P<0,05$, Note: the values written in brackets indicate the square root of the AVE values.

There are statistically significant relationships among the constructs in the sample in Table above. Correlation is the coefficient that indicates the power of linear relationship between variables. This coefficient must be statistically significantnordertobeabletosaythatthereisarelationship between variables. The correlation coefficient takes a value between-1and+1(Sipahi,Yurtkoru,&Çinko,2010).

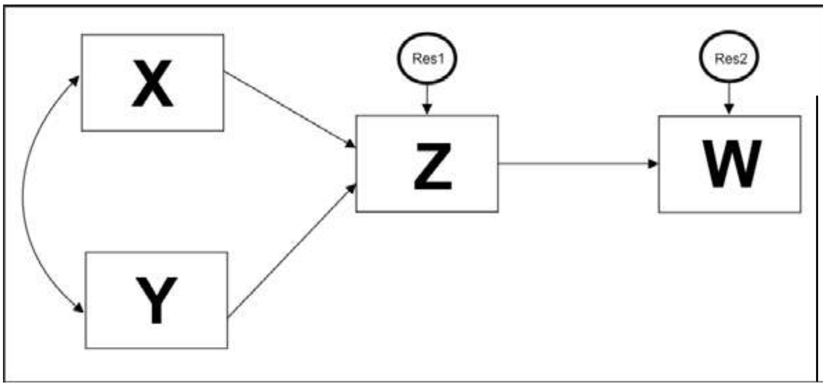
IV. TYPES OF STRUCTURAL EQUATION MODELS

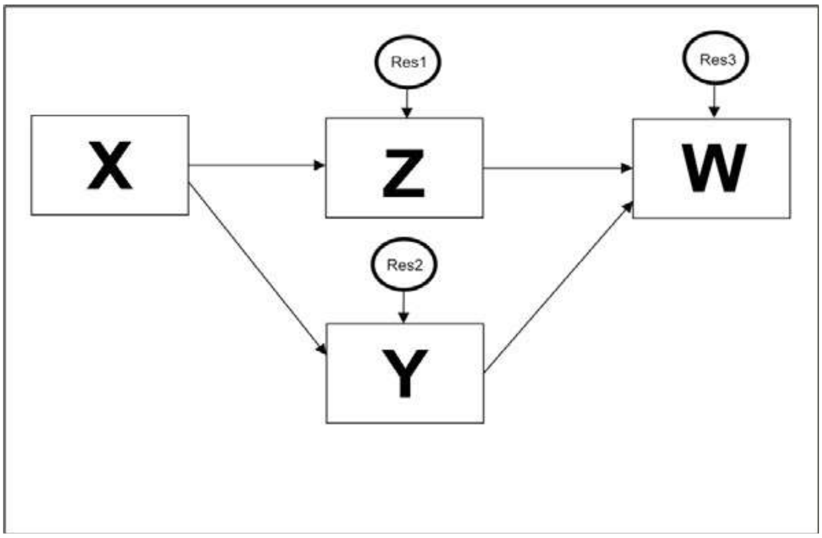
There are four basic types of structural equation models. These are explained below:

- **Path Analysis Models**

In the method of structural equation modeling, the models established with only observed variables are called path analysis models. The basis of the structural equation modeling depends upon path analysis. This model was developed by biologist Sewall Wright (Taşkın&Akat, 2010) and wasfirst implemented in the 1920s.The path analysis is similar to multiple regression as it is done with observed variables. However, it is superior to multiple regression, because there is one dependent variable in the multiple regression. Although, there may be more than one dependent variable in the path analysis, and a variable can be both a dependent variable and an independent variable, more than one regression model can be analyzed at the same time, and indirect and direct effects can be measured at the same time. Direct effect is the effect of one variable on another variable without any mediation. However, the indirect effect arises from the intervention of a variable which is playing mediator role between independent and dependent variables. This variable is named as the mediator variable. The sum of the direct effect and the indirect effect of a variable on another variable is called the total effect (Raykov&Marcoulides, 2006). Path analysis do not contain latent variables, they cannot be saved from measurement errors (Meydan&Şen, 2011). For this reason, structural regression models generated by latent variables give more accurate results.

Examples of Path Analysis

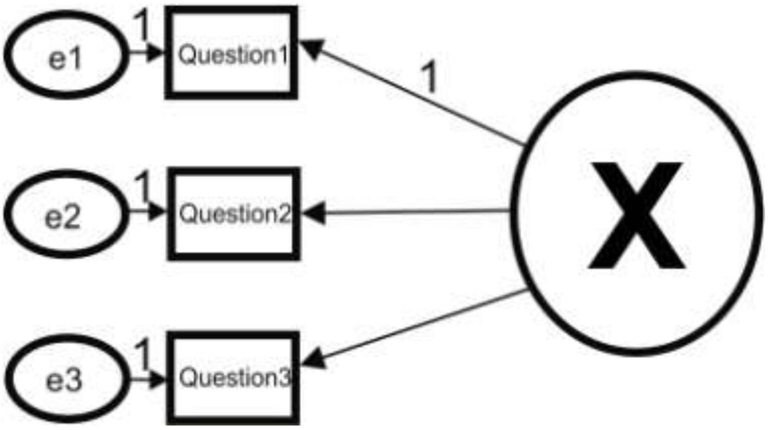




- **Confirmatory Factor Analysis**

This is divided into exploratory and confirmatory. In explanatory factor analysis, factors are revealed from relations among variables. In explanatory factor analysis, the observed variables can be loaded on any factor or on multiple factors. However, in the confirmatory factor analysis, the theoretically predetermined factor structure is confirmed by the current data. In other words, in the confirmatory factor analysis, which factor will be loaded on an observed variable is predetermined. By means of the explanatory factor analysis, the latent variables are revealed from the observed variables. However, in the confirmatory factor analysis, previously discovered scales are confirmed again with the collected data.

Single Factor CFA Model



Results of Confirmatory Factor Analysis

The Table below shows the way in which confirmatory factor analysis results are given. What is important here is that the standard factor loads of the questions under each conceptual variable are over 0.50. By looking at this table, questions with a standard factor load of less than 0.50 are discarded.

Items	Conceptual Variable	Standardized Factor Loads	Unstandardized Factor Loads	Standard Error	t-Value (Critical Ratio)
Qestion1		0,818	1		
Qestion2	X	0,906	1,104	0,049	22,523
Qestion3		0,907	1,111	0,049	22,570
Qestion4		0,825	1		

Qestion5	Y	0,732	0,882	0,057	15,549
Qestion6		0,718	0,885	0,058	15,187
Qestion7	Z	0,757	1		
Qestion8		0,835	1,102	0,062	17,785
Qestion9		0,939	1,255	0,062	20,176
Qestion10	W	0,676	1		
Qestion11		0,799	1,131	0,083	13,555
Qestion12		0,785	1,158	0,087	13,379

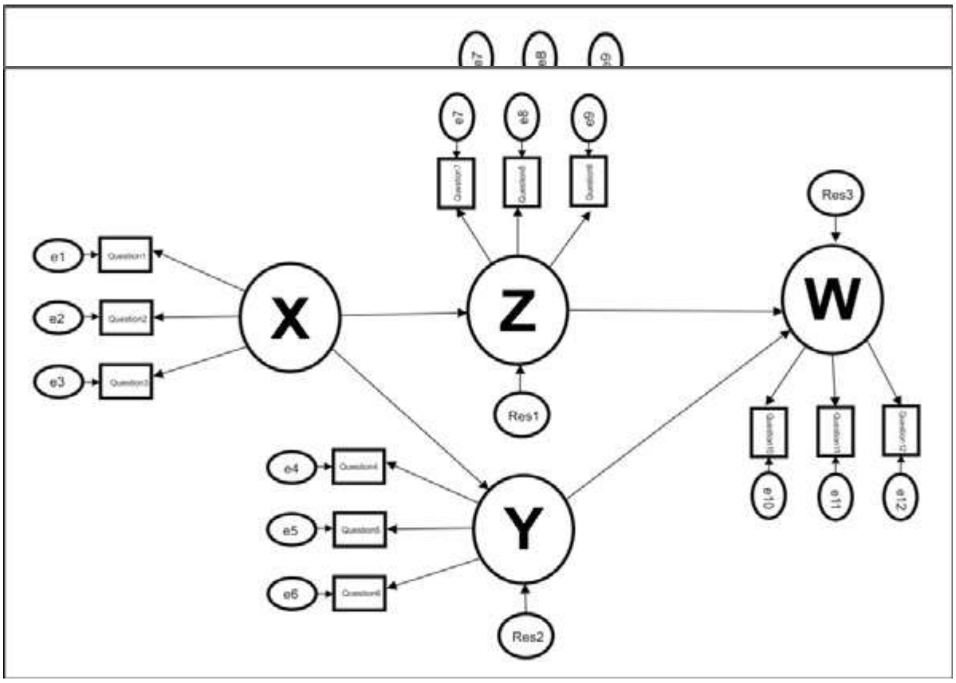
Note: For all values P<0.01

- **Structural Regression Models**

This is formed between latent variables in structural equation models. It consists of a combination of measurement model and structural model. Incorporating the measurement model and the structural model allows the inclusion of measurement errors so that more accurate results can be obtained. In other words, confirmatory factor analysis and multiple regression analysis coexist.

StructuralRegressionModelExample

These following are examples of structural regression models. Although the models in the first and second figure are basedon the same measurement model, the path analysis created is different. In the first figure, there are more than one exogenous variable and therefore covariance is placed between them. In the second figure there is only one exogenous variable. In both modelsresidualtermsarelinkedtotheendogenousvariables.Careful attention should be paid to these rules when constructing structural models. Otherwise the model will not work.



Hypothesis Test Results Table Example

The Table below shows an example of the hypothesis test results. The values in this table are in the estimates section of the output screen of the AMOS program. The notation *** in AMOS output means that P is equal to zero.

Relations	Standard Coefficients	Unstandardized

			Coefficients	
X → Z	0.533*	0.594*		
Y → Z		0.437*		0.638*
Z → W		0.493*	0.377*	

$^*p < 0.05$

- **Latent Change Models**

They are also named as “latent growth curve models” or “latent curve analysis”. They are models that describe longitudinal variation in time series (Raykov&Marcoulides, 2006). They are also used to explain the growth and decay of an event over time, similarities or differences within and between units (Doğan, 2015). Structural equation modeling is a very useful method for analyzing changes in time.In the figure below, two factorial growth models are observed fortwotimepoints(T1,T2). Repeated measurements over time are needed to use the latent change models. Such data are called longitudinal data (vertical crosssection data).These models are the models used to explain the growth and decay of an event over time, similarities or differences within and between units. (Doğan, 2015).

According to Baltes and Nesselroade (1979), thismodel can be used for the following purposes:

(1) Describeobservedandunobservedverticalsectiondata.

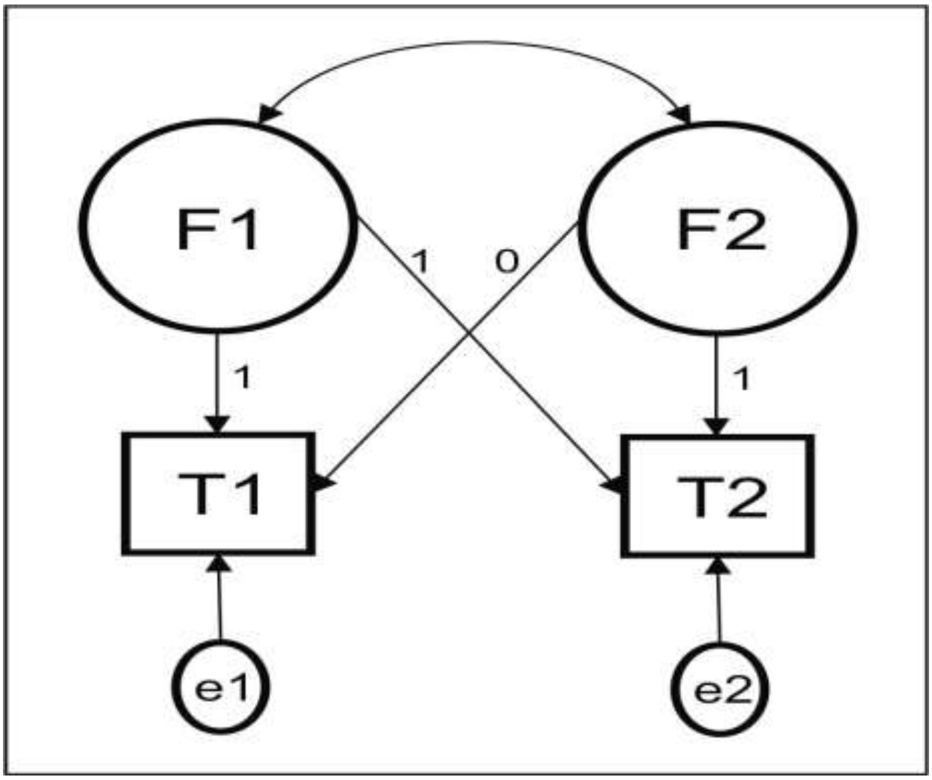
(2) Characterizethedevelopmentofindividualsandgroups.

(3) To predict individual and group differences in developmentalforms.

(4) To examine the dynamic determinants among variablesin time.

(5) To reveal the group differences of the dynamic determinants between variables intime.

Latent Change Model Example



Source:Raykov,T.,&MarcOULIDES,G.(2006).AFIRStCourSEinStructuralEquation Modeling.Mahwah:LawrenceErlbaumASSOCIATES.

V. METHODS TO BE APPLIED IN CASE OF DATA INADEQUACY

Sometimes, there may be cases where the assumptions of the estimation methods used are not met by the existing data set. In this case, there are methods that can be applied if it is necessary to be satisfied with the dataset available. Leading methods among them are bootstrap partial least square.

1. Bootstrap Method

The bootstrap technique is applied when one of the assumptions of normal distribution or being continuous variable is not met. This method was developed by B. Efron in 1979 (Efron, 1979). In many studies in the literature the condition of normal distribution obligation is neglected. It is also seen in many studies in the literature that X2 value is derived by maximum likelihood and generalized least squares methods. Estimation methods which are frequently used in the structural equation model are these two methods. In particular, with the non-normal distribution, the number of observations is also low cause X2 value to increase. At the same time, irreversible and inadequate modifications made during the analysis of such data are not scientifically acceptable and result in inconsistent estimations about the population. In the bootstrap method, a different data set is obtained from the existing observations (Sacchi, 1998). This method is basically the derivation of the sample from the sample.

There are advantages and limitations of the bootstrap process. The main advantage of the bootstrap technique is the ability to evaluate the accuracy of the predicted parameters. The idea underlying the bootstrap technique is to create sub-samples of the current data and look at the distribution of the parameters computed from each sub-sample.

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

It is also called covariance-based structural equation modeling since the structural equation model that has been examined in the previous sections is based on the covariance matrix. However partial least square structural equation modeling is based on variance. For this reason, it is also called as the variance-based structural equation modeling. Partial least square structural equation modeling (PLS-SEM) is an advantageous method when the assumptions of least squares are not met. It is an alternative of covariance-based structural equation modeling (CB-SEM). It is a second generation multivariate analysis method that enables measurement model and structural model to be analyzed together like covariancebased structural equation modeling.

According to Civelek (2018), covariance-based structural equation modeling is a more powerful and reliable method. For this reason, the partial least square structural equation modeling method is generally preferred in cases where the conditions listed below are found:

1. If the sample is small.
2. If the data do not distribute normally.
3. If the number of indicators connected to the latent variable is less than three.
4. If there is a multicollinearity.
5. There is missing value.
6. If the number of observations is less than the number of explanatory variables.

If the above listed conditions are found, method PLS-SEM method is far superior to method CB-SEM, because, in these cases, it reduces the unexplained variance to the lowest level. As the model is complex, such as in the CB-SEM method, no larger sampling is required in PLS-SEM. However, some researchers who have done research on the sampling sensitivity of the PLS-SEM method have raised the ten-fold rule. According to this rule, there is a necessity to have 10 times observation of the number of indicators used to measure a construct in the measurement model and 10 times observation the number of the path in a structural model (Barclay, Higgins, & Thompson, 1995).

However, the PLS-SEM method is a non-parametric method because it does not have any distributional assumption (Hair, Hult, Ringle, &Sarstedt, 2017). It is also an explanatory approach, which is why it is preferred in exploratory research. In other words, when the theory is underdeveloped, it can be said that researchers prefer to use partial least squares structural equation modeling. This judgment is partially correct in cases where the structure need to be predicted and relations need to be explained (Rigdon, 2012).

When the theory needs to be tested and verified, in case of there is cycles in the structural model and if the model needs to be verified in general with fit indices it is more accurate to use CB-SEM method, because the PLS-SEM method cannot explain loop-related relations. In addition, it does not give general fit indices of the model. Partial least squares method can be easily implemented by means of a packet program called SmartPLS. SmartPLS is a packet program that allows the creation of partial least squares based structural equation models. Structural equation modeling programs outside of SmartPLS makes the maximum likelihood estimation method the default choice.

VI. CONCLUSION

Novice researchers must take care of assumptions and concepts of Structure Equation Modeling, while building a model to check the proposed hypothesis. SEM is more or less an evolving technique in the research, which is expanding to new fields. Moreover, it is providing new insights to researchers for conducting longitudinal investigations.

REFERENCES

[1]. Avcılar, M., &Varinli, İ. (2013).*PerakendeMarkaDeğerininÖlçümüveYapısalEşitlikModeliUygulaması* (1 b.). Ankara: DetayYayıncılık.

[2]. Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least sequares approach to causal modeling: Personal computer adoption and use as illustration. *Technology Studies*, 2, 285-309.

[3]. Bayram, N. (2013). *YapısalEşitlikModellemesineGiriş*. Bursa: EzgiKitapevi

[4]. Bentler, P. M., & Chou, C.-P.(1987). Practical Issues in Structural Modeling.*Sociological Methods Research*, 16(1), 78-117.

[5]. Çelik, H. E.&Yılmaz, V. (2013).Lısrel 9.1 *ileYapısalEşitlikModellemesi.* Ankara: AnıYayıncılık.

[6]. Civelek, M. (2018). The Mediator Roles Of Attitude Toward The Web Site And User Satisfaction On The Effect Of System Quality On Net Benefit: A Structural Equation Model On Web Site Success. *European Scientific Journal*, 61-73.

[7]. Doğan, İ. (2015).*FarkliVeriYapisiveÖrneklemBüyüklüklerindeYapisalEşitlikModellerininGeçerliğiveGüvenirliğininDeğerlendirilmesi.*Eskişehir: EskİşehİrOsmangazİÜnİversİtesİDoktoraTezi.

[8]. Efron, B. (1979). Bootstrap method: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.

[9]. Fornell, C., &Larcker, D. (1981).Evaluating Structural Equation Models with Unobservable Variables and Measurement Error.*Journal of Marketing Research*, 18(1), 39-50.

[10]. Gerbing, D., & Anderson, J. (1988).An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its.*Journal of Marketing Research*, 25(2), 186-192.

[11]. Hair, J., Hult, G., Ringle, C., &Sarstedt, M. (2017).*A primer on partial least squares structural equation modelingPLS-SEM.* Los Angeles: SAGE.

[12]. Jayaram, J., Kannan, V., & Tan, K. (2004). Influence of initiators on supply chain value creation. *International Journal of Production Research*, 42(20), 4377-4399.

[13]. Karagöz, Y. (2016). *SPSS ve AMOS 23 UygulamalıİstatistikselAnalizler*. Ankara: Nobel.

[14]. Kline, R. (2011). *Principles and practice of structural equation modeling* (3nd Ed. b.). New York: Guilford.

[15]. Meydan, C. H. &Şen, H. (2011).*YapısalEşitlikModellemesi AMOS Uygulamaları. Ankara*: DetayYayıncılık.

[16]. Raykov, T. (1997).Estimation of composite reliability for congeneric measures.*Applied Psychological Measurement*, 21(2), 173-184.

[17]. Raykov, T., &Marcoulides, G. (2006).*A First Course in Structural Equation Modeling.* Mahwah: Lawrence Erlbaum Associates.

[18]. Rigdon, E. (2012). Rethinking partial least squares path modelling: In praise of simple methods. *Long Range Planning*, 45, 341-358.

[19]. Sacchi, M. (1998).A bootstrap procedure for high-resolution velocity analysis.*Geophysics*, 63(5).

[20]. Sarstedt, M., &Mooi, E. (2014).*A concise guide to market research: The process, data, and methods using* IBM SPSS statistics (2nd b.). Berlin: Springer.

[21]. Taşkın, Ç.&Akat, Ö. (2010).*AraştırmaYöntemlerindeYapısalEşitlikModelleme*. Bursa: EkinBasımYayın