



Comparison of weighted least squares and robust estimation in structural equation modeling of ordinal categorical data with larger sample sizes

Cengiz GAZELOĞLU¹ , Zerrin AŞAN GREENACRE²

¹ Department of Statistics, Suleyman Demirel University, Isparta, Turkey

² Department of Statistics, Eskisehir Technique University, Eskisehir, Turkey

Abstract

The effect of different sample sizes on estimation methods such as weighted least squares and robust weighted least squares that are used in structural equation modeling was studied and compared using information criteria such as Akaike Information Criteria in this study. The simulations were repeated 1000 times with two estimation methods and the average values of criteria were calculated with different sample sizes. The study includes a construct of four factors, with four questions of each that are measured on a five-point Likert scale. Different sample sizes, ranging from 300 to 5000 were selected. According to the simulations results, it is concluded that the robust estimation method provides more effective results at lower sample size. In addition, it was found that as the sample size increases, the efficiency difference between two methods gradually decreases. Moreover, it was detected that there is almost no difference between the two methods for sample sizes over 3000.

Article info

History:

Received:18.11.2019

Accepted:13.01.2020

Keywords:

Structural equation modeling,
Categorical data,
Information criteria,
Simulation

1. Introduction

In the analysis carried out with the variables defined as "Categorical Variable" in the interdisciplinary researches, the implementation of statistical methods within the scope of "Categorical Data Analysis" is necessary. Generally speaking, most statistical techniques are developed assuming interval or ratio levels of measurement. They applied to continuous variables are not appropriate for categorical data. Therefore, it increases the importance of methods within the scope of "Categorical Data Analysis" [1-4]. The most commonly used analytical methods in categorical literature are as follows: Regression Models: Probit / Logistic Regression, Generalized Linear Models, Loglinear Models, Explanatory Factor Analysis, Item Response Theory (IRT), Structural Equation Modeling with Categorical Data [5].

Most of the statistical methods have performed analyses with the assumption that the observed variables are reliable and valid. However, it has been adopted that the analyses done under the assumption that the variables observed in most social sciences such as sociology and psychology are measured perfectly. For this reason, Structural Equation Modeling (SEM), which allows measurement errors of variables, has

been developed. In most disciplines, the main reasons for the increase in SEM's popularity are:

- The ability to analyze simultaneous multiple relationships while providing statistical efficiency
- The ability to carry out versatile assessment of relationships and
- The ability to test causal relationships between observed and hidden variables [6,7].

Because the data is categorical and cannot meet the normality assumption, along with using of various model estimation methods (such as maximum likelihood) and covariance matrices, many researchers do not recommend it. Instead of this, using alternative matrices and estimation methods will provide more consistent and more accurate results in SEM with the researcher.

The most important aspect that reveals the originality of this work is that it provides comparison of methods of estimating both the simulation and the actual results of the application on the data together with the information criteria such as AIC, CAIC and weighted least squares (WLS) and robust weighted least squares (WLSMV), which are used in SEM. In addition, it was aimed to show which estimation method is effective in different sample sizes. For this reason, it was planned to observe how and to what extent the goodness of fit indexes is affected by sample size for Structural

*Corresponding author. Email address: cegazel@gmail.com

Equation Modeling with Categorical Data depending on different sample sizes. In this study, there is no missing data in both the simulation data and the actual data.

2. Structural Equation Models For Categorical Data

In behavioral science researches, response variables are usually measured in an intermittent and sequential or classifying scale. Likert type scales (see Likert, 1932) [8] are examples of such types [9]. In the analysis using these variables in structural equation modeling, there are differences in approach. Depending on the type of the variable, there are many different estimation methods that can be used to predict the parameters included in the structural model. When the variables are continuous, the most commonly used parameter estimation method is the maximum likelihood (ML) method. However, in the case of sequential variables, the most commonly used method is WLS. While the parameter estimation methods vary according to the type of variables, each parameter estimation method has its own assumptions. The method of parameter estimation method to be used depends on whether these assumptions can be met or not. The first step that needs to be done before using the structural equation model is to determine the measurement level of the variables and then to select the appropriate parameter estimation method [10].

In studies where the data are not normally distributed or categorical data is available, using the ML estimation method will not yield accurate results.

There are other estimation methods besides this estimation method. When such data sets are encountered, Jöreskog and Sörbom (1993) [11] recommended using WLS and WLSMV in their work. Moreover, it is recommended that no such normal covariance matrix or Pearson's correlation matrix be used in such data [12].

The various correlation matrices (Tetrachoric, polyserial... etc.) to be used in the analysis should be calculated. In addition to these correlation coefficients, the asymptotic covariance matrix (ACM) generated from the corresponding correlation matrices must also be used [13].

When the desired data to be resolved does not provide the normality assumption or when categorical data is used, normality assumption is distorted so that the correlation coefficient to be calculated also varies. Along with the change in the correlation coefficient, estimation methods will also change. In this case, using classical assumption methods will give non-objective results. As a result, the measurement level of the data has a great importance in the type of analysis to be employed.

Although the Pearson correlation coefficient has a fundamental influence in the field of statistics, the other correlation coefficients vary depending on the measurement level of the variable, too. The correlation coefficient that is calculated according to the measurement levels of the variables (sorter, ordinal, interval and proportional) also varies [14]. The correlation coefficient types to be calculated according to measurement types are shown in Table 1.

Table 1. Type of correlation coefficients.

Correlation Coefficient	Measurement Level	
	Variable I	Variable II
Pearson (Product moment)	Interval	Interval
Spearman, Kendal Tau	Ordinal	Ordinal
Phi	Nominal	Nominal
Point Biserial	Interval	Nominal
Gamma	Ordinal	Nominal
Biserial	Interval	Artificial Variable
Polyserial	Interval	Sorter with a basic variable
Tetrachoric	Two-Level (sorter Artificial)	Two-Level (sorter Artificial)
Polychoric	Ordinal with a basic-continuous variable	

Source. [15]

The most important assumptions in the correlation analysis between variables are the normality and linear relationship assumption. However, when the variables have a categorical structure, the assumption of

normality is distorted in particular. There are a number of correlation coefficients developed to obtain correlation values between categorical variables [16]. In SEM analysis, Pearson's correlation coefficient,

tetrachoric (or polychoric for various types of ordinal variables) correlations and binary series (or polyserial correlation coefficients for continuous and ordinal variables) are generally used [14].

2.1. Theory of structural equation modeling in categorical data

When y^* is approached as a $p \times 1$ dimensional vector of continuous indices of η , measurement model is shown as follows: $y^* = \Lambda_y \eta + \varepsilon$

Here $E(\varepsilon) = 0$ and ε , is unrelated to η . y^* is not the observable variable. However, it is assumed that the μ_{y^*} average of each y ordinal variable is the $\sigma_{y^*}^2$ variance and normal distribution with variance y^* basic continuous variable [17].

For example, while y_1^* has normal distribution, y_1 can be a sequential variable, which has four categories. If $y \neq y^*$

$$y \neq \Lambda_y \eta + \varepsilon \tag{1}$$

When the major mass covariance matrix of y^* and x^* is Σ^* , according to classic SEM, it will be $\Sigma^* = \Sigma(\theta)$. Generally, Σ , which is the population covariance matrix of y and x , is not equal to Σ^* and it becomes $\Sigma \neq \Sigma(\theta)$.

The covariance structure hypothesis contains latent continuous indicators and there is no need for sequential observable indicators. If the assumption that S is a coherent estimator of Σ is adopted so S^* will be a consistent predictor of Σ^* . The parameter estimators are based on S and any concordance function could be an inconsistent estimator of the actual parameter factor.

For example, all variables are standardized in a simple regression equation. Consistent predictors of regression coefficients are the sample correlation between y^* and x^* (r^*).

If y and x are the standardized ordered adaptations of y^* and x^* , the regression coefficient will be the sample correlation between y and x (r). Instead of latent continuous indicators in SEM, when sequential indicators are used, it becomes

$$y \neq \Lambda_y \eta + \varepsilon \tag{2}$$

$$x \neq \Lambda_x \xi + \delta \tag{3}$$

$$\Sigma \neq \Sigma(\theta) \tag{4}$$

$$ACOV(s_{ij}, s_{gh}) \neq ACOV(s_{ij}^*, s_{gh}^*) \tag{5}$$

First, when the equation (3) is considered, the linear relationship between $y(x)$ and $\eta(\xi)$ is not appropriate for making relevant analyses. According to the latent indicators here (y^* and x^*), it is seen that the observed sequential variables are related, and it needs a non-linear function. In analyzing latent variable models where ordinal variables are used, there are two main approaches. These are Item Response Theory- IRT and PRELIS-LISREL approaches (PL) [18].

In SEM, each y ordinal variable μ_{y^*} average and $\sigma_{y^*}^2$ variance and normally distributed y^* is defined as the basic continuous variable. As it is written as $y = i$, it is displayed that y belongs to the sorted category i .

Actual score values for the data may be randomly selected and insignificant. The relationship between ordinal variable and basic y^* variable is as follows;

$$y = i \iff \tau_{i-1} < y^* \leq \tau_i, \quad i = 1, 2, \dots, m$$

Here ;

$$\tau_0 = -\infty, \quad \tau_1 < \tau_2 < \dots < \tau_{m-1}, \quad \tau_m = \infty$$

parameters are called as threshold values.

More clearly, for ordinal y_1 indicator, equation 6 applies.

$$y_1 = \begin{cases} 1, & y_1^* \leq \tau \\ 2, & \tau_1 \leq y_1^* \leq \tau_2 \\ \vdots & \vdots \\ m-1, & \tau_{m-2} \leq y_1^* \leq \tau_{m-1} \\ m, & \tau_{m-1} \leq y_1^* \end{cases} \tag{6}$$

The number of the categories here for m , y_1 , is τ_i ($i = 1, 2, \dots, m-1$) category thresholds and y_1^* constant latent variable. For y variable that has m category, there are $m-1$ threshold parameters [17].

2.2. Model estimation methods by structural equation modeling in categorical data

The most commonly used estimation methods in the structural equation model, which is one of the analytical methods of categorical data are the ML method, 2-step least Squares (2SLS), generalized least squares (GLS), WLS, WLSMV, tool variables (TV) and cross weighted least squares (CWLS) methods. Among these mentioned estimation methods, estimation methods vary according to the measurement level of the data. Since this study was carried out using

categorical data, weighted least squares estimation and robust weighted least squares estimation were used.

Table 2. The differences between WLS and WLSMV estimators.

	Chi-Square Estimated Value	Parameter Estimation	Standard Errors	When is it applied?
WLS	All weight and reverse matrices are used.	All weight matrices are used.	All weight and reverse matrices are used.	When there are categorical variables or continuous internal variables
WLSMV	All weight matrices are used but are not reverse.	Diagonal weight matrices are used.	All weight matrices are used but are not reverse.	When there is at least one categorical internal variable

When working with continuous variables that do not provide the normal distribution assumption, in this case, asymptotic distribution free (ADF) estimation method is recommended. Although simulation studies have shown that maximum likelihood estimation (MLE) is better than ADF in normal conditions, it isn't recommended to use MLE. The ADF function is described as arbitrary generalized least squares (AGLS) in EQS package and weighted least squares (WLS) in LISREL [11]. In contrast to MLE, raw data are needed to analyze the data in ADF. This method can be used if some of the variables are sequential and others are continuous if the distribution of continuous variables shows deviations from the normal distribution and model contains two-level variables [18]. The fit function that minimizes MLS is shown in Equation 7.

$$F_{WLS} = [s - \sigma(\theta)]'W'[s - \sigma(\theta)] \tag{7}$$

s is the vector of nonredundant elements in the empirical covariance matrix, θ is the (tx1) vector of parameters, W^{-1} is a (kxk) positive definite weight matrix with $k = p(p + 1)/2$ and p=number of observed variables in Equation 7 [19].

The Robust weighted least squares method is one of the most commonly used estimation methods for estimating parameters when categorical data is used. As a result of these studies, in some sources, this method is known as the estimation method of Diagonally weighted least squares estimation (DWLS).

The DWLS method aims to estimate WLS parameters using the diagonal weight matrix (W) and robust standard errors, mean and variance- adjusted. It is based on polychoric correlations alone when there are no constraints in the thresholds. Assuming that $\rho(\theta)$ is represented all of the elements of polychoric

correlation matrix, the least squares function is shown as in Equation 8 [19].

$$F_{WLSMV} = (\hat{\rho} - \rho(\theta))' \hat{W} (\hat{\rho} - \rho(\theta)) \tag{8}$$

Let $\hat{\Gamma}$ be an estimate of the asymptotic covariance matrix ($\hat{\rho}$) of polychoric correlations predicted in Equation 8. When $\hat{W} = \hat{\Gamma}$, this method is called the Weighted least squares (WLS). When $\hat{W} = (diag(\hat{\Gamma}))^{-1/2}$, this method is called the Diagonal weighted least squares (DWLS). In the second case, only variances of the polychoric correlations estimated in the weight matrix are used [20].

2.3. Information criteria in the structural equation model with categorical data

Amongst the most popular information criteria are Akaike Information Criteria [21,22,23], Consistent Akaike Information Criteria [24], Expected Cross-Validity Indexes [25], Consistent Akaike Information Criteria Based on Fisher Information [26], Information Complexity Criteria Based on Inverse Fisher Information Matrix [27], Schwarz Information Criteria [28], Mallow's Cp Criteria [29]. In this study, Akaike Information Criteria was used. In most of the studies, only the AIC information criteria is used [22]. Cheung and Rensvold (2002) [30] studied for evaluating for goodness of fit index. Also, it is be seen articles about Comparison of Akaike information criterion for model selection like Bozdoğan (1987) [24], Anderson and White (1998) [31].

3. Simulation Study

3.1. Literature simulation study

In the literature, simulation studies have searched the properties of different estimation methods at different

sample sizes for continuous and categorical data. It can be seen many papers about model fit with ordinal data.

For example, it can be seen related to working with Muthen (1983) [32], Muthen (1984) [19], Muthen and Kaplan (1985) [33], Jöreskog (1991) [34] and Muthen and Satorra (1995) [35]. It was given the theory for ordinal data in SEM by Muthens (1984) [19] and Muthen and Satorra (1995) [35].

In addition to, Hipp and Bollen (2003) [36] worked model fit in SEM with censored, ordinal and dichotomous variables. Another study belongs to Flora and Curran (2004) [37]. It was shown that the general performance of WLS was low that of WLSMV in CFA for ordinal data at almost every condition by Flora and Curran. It was researched WLSMV and robust unweighted least squares with two or five category ordinal variables under one- or three-factor correctly specified CFA models by Forero et al. (2009) [38]. It was obtained less biased factor loading estimates by Beauducél and Herzberg (2006) [39] and DiStefano (2002) [40]. It was investigated that the parameter estimates obtained by WLS were substantially biased, whereas those obtained by WLSMV and ML were essentially unbiased, regardless of the number of categories and the shape of the observed distributions by Yang-Wallentin et al. (2010) [41].

Savalei and Rhemtulla's (2013) [42] paper reported on a simulation study that evaluated the performance of five structural equation model test statistics appropriate for categorical data. Different model sizes, sample sizes, numbers of categories, and threshold distributions were considered in their article. Statistics associated with both the diagonally weighted least squares estimator and with the unweighted least squares estimator were studied. It showed a comparison of robust continuous and categorical SEM estimation methods by Rhemtulla et al. (2012) [43]. Diagonal weighted least squares robust estimation techniques were compared for ordinal data in another study by Stefano and Morgan (2014). They used two different statistics programs (LISREL and Mplus) for comparison with sample sizes ranging from 200 to 800. It was found these estimators generally produced a suitable model-data fit except for a few categories were used with non-normal data at smaller sample sizes in this study. Their studies results were resounding the findings of Forero et al. (2009) [38, 44]. Li (2016) compared robust maximum likelihood and diagonally weighted least squares estimators in CFA for ordinal data. This simulation does provide conclusive evidence that WLSMV performs uniformly better than MLR in factor loading estimates regardless

of the number of categories, sample size, and the degree of latent [45].

3.2. Present Simulation Study

Apart from the studies mentioned in the literature, it is compared estimation methods WLS and WLSMV for SEM with ordinal data (a five point Likert scale) according to ranging sample size from 300 to 5000 in this study. When the literature review is carried out about the methods used in the analysis of categorical data, it will be seen that they cover a very large area. Within this area, analyses related to categorical data in SEM are also widely available. However, both SEM and categorical data analysis are supported with very few studies. With this work, great contributions will be made in the name of the transition of these two fields.

In this study, it was aimed to demonstrate how SEM makes analysis with categorical data and how the WLS estimation method and the WLSMV estimation methods perform at different sample sizes. Besides, it was intended to find out how the information criteria such as AIC, behaves according to sample size. For this purpose, Mplus package program was used to demonstrate the performance of the related estimation methods and information criteria, and simulation studies were carried out at different sample sizes. The most important reason for choosing the M-plus package program in work is that it owns a very good simulation module. When data are both categorical and non-normally distributed, this package provides for the use of both the WLSM and WLSMV estimators [46]. Moreover, WLSMV has been specifically proposed to deal with ordinal data in Mplus, because it makes no distributional assumptions about the observed variables [45]. When it comes to an SEM model with ordinal data, applied researchers tend to choose one or another estimator to perform data analysis in Mplus. The disadvantage of the program is that it takes up a lot of space for data generation and simulation results occupy time depending on the characteristics of the computer used. The reason why a simulation was carried out in the study is that collecting actual data at 13 different sample sizes cause problems to researchers just like time, cost, and qualified worker ... etc. Because a 5-point Likert scale is used in the simulation study, $5-1 = 4$ will be the threshold value. (See Equation 6) Next, these mentioned threshold values were calculated according to the normal distribution threshold values and data were generated at the corresponding sample sizes. These analyses have been implemented through WLS and WLSMV estimation methods. Analyses were repeated 1000 times and averages were taken [47].

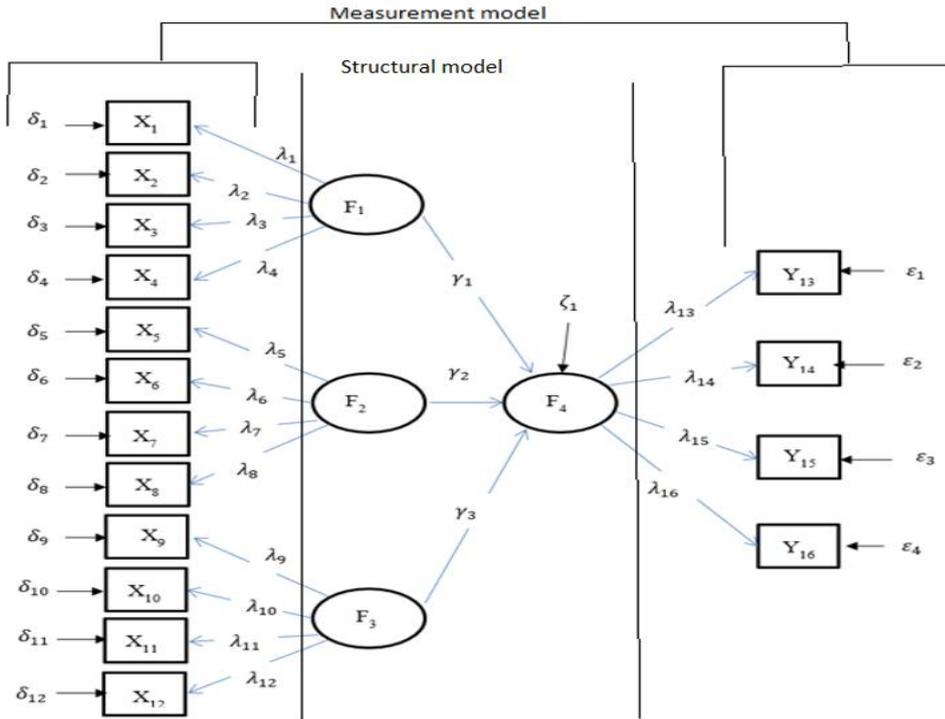


Figure 1. Structural equation modeling with 4 factor 16 ordinal variables

The simulation study consists of 4 factors as shown in Figure 1. In addition, each of these 4 factors was composed of 5 independent Likert and 4 independent variables. A total of 16 independent variables were included in the study. Each stage of the simulation study performed with the M-Plus package program was averaged 1000 times. The results obtained indicate the general average. In Figure 1, you can see the structural equation model that is made up of 4 latent variables and 16 observed variables. Furthermore, each latent variable is explained by four observed variables.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{13} \\ \lambda_{14} \\ \lambda_{15} \\ \lambda_{16} \end{bmatrix} \cdot F_4 + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

As the sample size in the simulation study, 300, 350, 400, 500, 750, 1000, 1250, 1500, 1750, 2000, 2500, 3000 and 5000 were used. The main reason why a minimum of 300 was taken in the sample size is that the covariance matrix used in the calculations is either negative or it cannot be calculated by the used program. The main reason for taking 5000 as the maximum value is that the changes in the goodness of fit indexes after the 3000 sample size in the simulation results are too low to be tested. There are not any changes in some index values. In addition, RMSEA, CFI, TLI, AIC, CAIC, χ^2 and χ^2/sd were considered as model evaluation indices.

General Matrix Representation

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \\ X_9 \\ X_{10} \\ X_{11} \\ X_{12} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \\ \lambda_9 \\ \lambda_{10} \\ \lambda_{11} \\ \lambda_{12} \end{bmatrix} \cdot F_1 + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \end{bmatrix}$$

Table 3. Structural equation model indexes established in different sample sizes and calculated by WLSMV method.

Sample Size n	χ^2	RMSEA	CFI	TLI	AIC	χ^2/sd
n = 300	101.882	0.011	0.991	0.994	273.882	1.039
n = 350	101.742	0.010	0.993	0.995	273.742	1.038
n = 400	100.766	0.009	0.994	0.997	272.766	1.028
n = 500	99.993	0.008	0.996	0.998	271.993	1.020
n = 750	99.895	0.006	0.997	0.999	271.895	1.019
n = 1000	99.570	0.005	0.998	0.999	271.570	1.016
n = 1250	98.383	0.004	0.998	1.000	270.383	1.003
n = 1500	98.202	0.004	0.999	1.000	270.102	1.002
n = 1750	97.660	0.003	0.999	1.000	269.660	0.996
n = 2000	98.499	0.003	0.999	1.000	270.499	1.005
n = 2500	98.178	0.003	0.999	1.000	270.178	1.001
n = 3000	98.531	0.003	0.999	1.000	270.531	1.005
n = 5000	98.351	0.002	1.000	1.000	270.351	1.003

Table 3 shows the most commonly used goodness of fit indexes for model evaluation such as RMSEA, CFI, TLI, AIC and χ^2/sd . There are different sampling sizes in the leftmost column of the table. The table summarizes how goodness of fit indexes vary according to sample size in general.

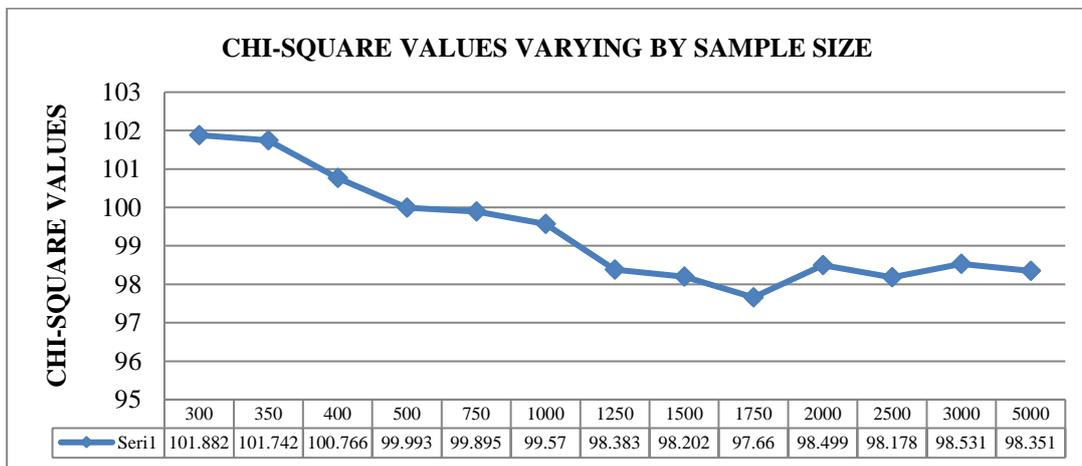


Figure 2. Chi-Square values varying by sample size through estimated model by WLSMV method.

The graph in Figure 2 shows the Chi-Square statistic, which varies according to sample volumes. When the figure is examined, it is seen that when n equals 300, Chi-Square value is calculated as 101.882. At n = 350 sample volume, the Chi-Square value is found to be 101.742. Chi-square values were then calculated for increasing sample volumes as 100.766, 99.993, 99.895, 99.570, 98.383, 98.202, 97.660, 98.499, 98.178, 98.531 and 98.351, respectively. The Chi-square value in SEM is a statistic used to test the meaning of the theoretical model. It is desirable state that the Chi-square value statistic is close to zero. If the Chi-square value takes a zero value, it can be seen that there is no difference between the covariance matrices for the theoretical model derived from the sample covariance matrix. It can be said that the theoretical model approximates to the real model with the increase of sample volume in the simulation study. It can be said that the theoretical model gets closer to the real model

by the increase of sample size within this simulation study.

This situation is shown in Figure 2. In other words, the increase in sample size and the decrease in Chi-Square value is the indicator of reduction of the difference between sample covariance matrix and covariance matrix of the theoretical model. However, Chi-square is an influential statistic in the number of observations. The sample size has certain limitations according to the work done. If it stays below or above these limit values, the calculated Chi-square value varies. This study includes a structural equation model with 4 factors and 16 variables and with increases in sample size above 2000, there will be an increase or decrease in the Chi-Square value. It can be said that this is an indicator that it has gone out of the limitations. In such a case, when working with categorical data in a model as in Figure 1, it is suggested that the sample size should not exceed 2000. In addition to this situation, it is seen that sample

size should not be under 300 and the covariance matrix is negative or incalculable in the simulation study.

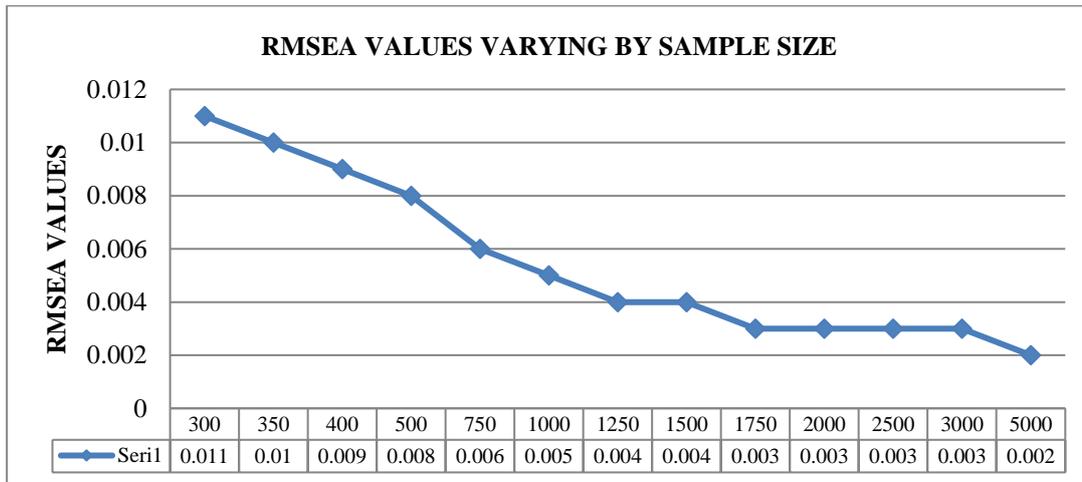


Figure 3. RMSEA values varying by sample size through WLSMV method.

Figure 3 shows how the value of RMSEA varies according to the sample size. If the RMSEA value is between zero and 0.05, it indicates that it is in good fit [20, 48]. In addition, if the RMSEA is between 0.05 and 0.08, it means that it is an acceptable fit. On the other hand, if the RMSEA is between 0.08 and 0.10, it indicates that it is a mediocre fit. If the RMSEA value is greater than 0.10, it is an indication that this model cannot be accepted. When Figure 3 is examined, it can be seen that when the sample size is 300, the RMSEA value is 0.011. Through increasing sample size, the

RMSEA values are gradually decreasing, approaching zero. Because of this assessment, the increase in sample size is as follows: That is, as for the model acceptance evaluation criteria, the RMSEA value should decrease and the sampling size should be at the sufficient size. Especially when the categorical data are concerned, the importance of size of the sample increases as well. When evaluated in this sense, there is a need for a larger sample size for categorical data compared to the continuous variables.

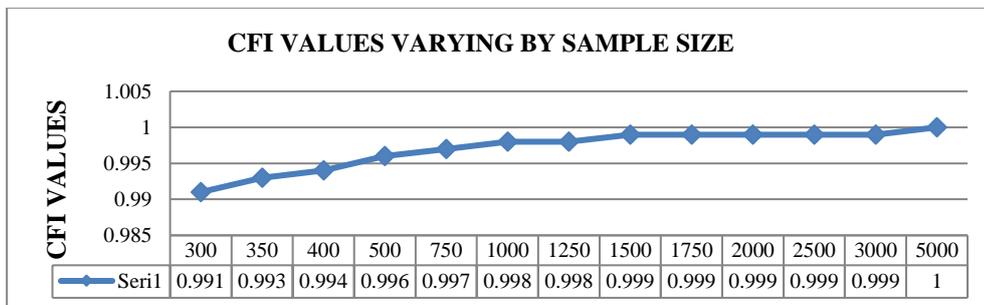


Figure 4. CFI values varying by sample size through WLSMV method.

Figure 4 shows how the CFI fit index changes according to the sample size [20, 48]. When the sample volume is 300, the CFI value is calculated as 0.991. According to the increasing sample size in Figure 4, the CFI values are 0.993, 0.994, 0.996, 0.997, 0.998, 0.999, 0.999, 0.999, 0.999, 0.999 and 1, respectively. When the CFI fit index takes a value between 0.95 and 0.97, it is in an acceptable fit. In addition, when the CFI fit index has a value between 0.97 and 1.00, it is

regarded as good fit. If the index takes a value below 0.95, it indicates that the established model cannot be accepted. It was observed in this study that as the sample size increases, so does the CFI value. It is seen in Fig. 4 that after 1500 sample size, the fit index is very close to 1. This can be explained as follows: When making a structural equation model in categorical data, it can be said that 1,500 samples will suffice as it is evaluated in terms of CFI index.

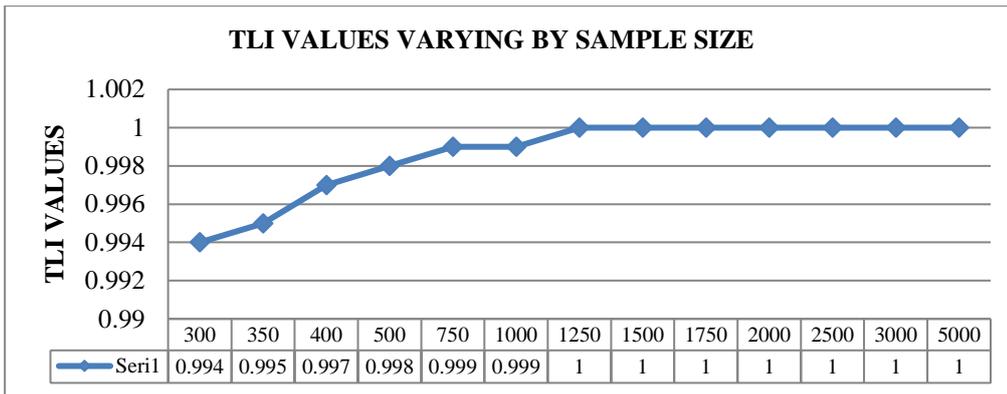


Figure 5. TLI values varying by sample size through WLSMV method.

Figure 5 shows the change in TLI value according to sample size [20, 48]. The TLI index is one of the model evaluation indices that are less affected by sample size. When the TLI index takes a value between 0.97 and 1.00, it is considered a good fit. Also, if the value is between 0.95 and 0.97, the TLI index is in an acceptable fit. If the value falls below 0.95, it indicates

that the model is not compatible. As can be seen in Figure 5, as the sample volume increases, it is observed that the TLI index increases, too. After 1250 sample size, the TLI value is constantly 1.00. It can be interpreted as that a maximum of 1250 samples will be sufficient when an analysis is to be done with categorical data in SEM.

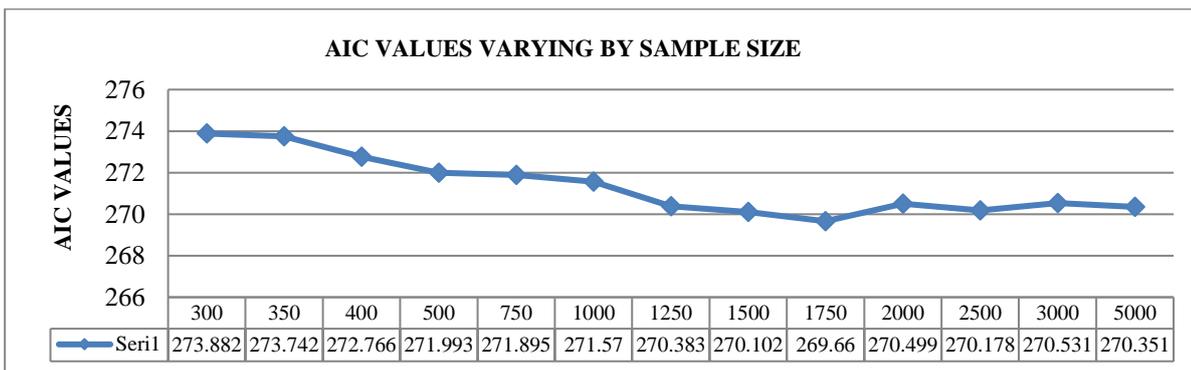


Figure 6. AIC values varying by sample size through WLSMV method.

The AIC value, a predictive fit index, is an information criterion used in model comparison. The AIC index is used both for comparing different models and for testing model consistency. The model with a lower AIC value is interpreted as a better and more consistent model. When Figure 6 is examined, it is perceived that when the sample size is 300, the AIC value is calculated as 273.88. According to Figure 6, as the sample size increases, the corresponding information criteria fall to 1750 sample size. However, after 1750

sample volumes, it appears that the increasing AIC values fluctuate with increases or decreases in sampling sizes. The main reason for this fluctuation is that the optimum sample volume was calculated as 1750 in the model structure established depending on the data structure generated in the simulation. The next sample increments, when n equals to 1750, are not lower than this size.

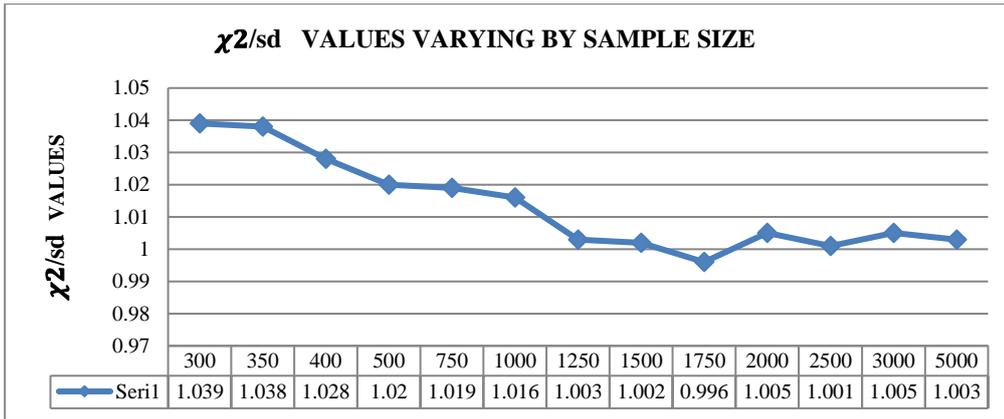


Figure 7. χ^2/sd values varying by sample size through WLSMV method.

Another index used in evaluating the compatibility of the established theoretical model is χ^2/sd . If this value takes a value between 0 and 2, this is the indicator of a good fit. Also, if it has a value between 2 and 3, it is in an acceptable fit. If the sampling volume is 300, it is calculated as $\chi^2/sd = 1.039$. As the sampling size increases, χ^2/sd value also decreases depending on the

decrease in Chi-Square value. However, in sample volumes after $n = 1750$, the χ^2/sd value shows some fluctuations by including increases or decreases. These fluctuations are due to the limitations of the Chi-Square value. A detailed description of this situation was given in Figure 7

Table 4. Structural equation model evaluation indexes established in different sample sizes and calculated by WLSMV method.

Sample Size n	χ^2	RMSEA	CFI	TLI	AIC	χ^2/sd
n = 300	181.313	0.052	0.957	0.947	377.313	1.850
n = 350	164.569	0.043	0.964	0.956	360.569	1.679
n = 400	154.164	0.037	0.970	0.963	350.164	1.573
n = 500	138.900	0.027	0.979	0.974	334.900	1.417
n = 750	123.890	0.017	0.989	0.987	319.890	1.264
n = 1000	116.270	0.012	0.993	0.992	312.270	1.186
n = 1250	112.377	0.010	0.995	0.995	308.377	1.146
n = 1500	109.698	0.008	0.996	0.996	305.698	1.119
n = 1750	108.170	0.007	0.997	0.997	304.170	1.103
n = 2500	104.023	0.005	0.998	0.999	300.023	1.061
n = 3000	103.917	0.004	0.999	0.999	299.917	1.060
n = 5000	101.714	0.003	0.999	1.000	297.714	1.037
n = 2500	104.023	0.005	0.998	0.999	300.023	1.061
n = 3000	103.917	0.004	0.999	0.999	299.917	1.060
n = 5000	101.714	0.003	0.999	1.000	297.714	1.037

Similar to Table 3, Table 4 shows the most commonly used goodness of fit indexes and sample sizes in different volumes. However, the point to be noted here is that Table 4. contains the results predicted by the WLS method.

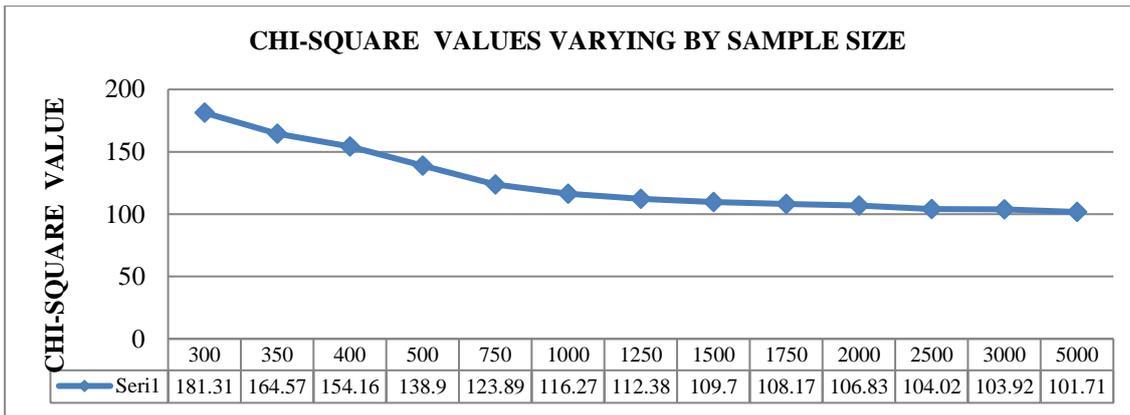


Figure 8. Chi-Square values varying by sample size through WLS method.

Figure 8 shows the Chi-Square values estimated by the WLS method and varying according to different sample sizes. When the sample size is 300, it is seen that the Chi-Square value is 181.31 and as the sample volume is 350, the Chi-Square value is 164.57. That is, it is seen that the Chi-Square value decreases when the

sample volume increases. It requires that the Chi-Square value is close to zero. Because if the value is zero, it means that there is no difference between the covariance matrices for the theoretical model derived from the sample covariance matrix.

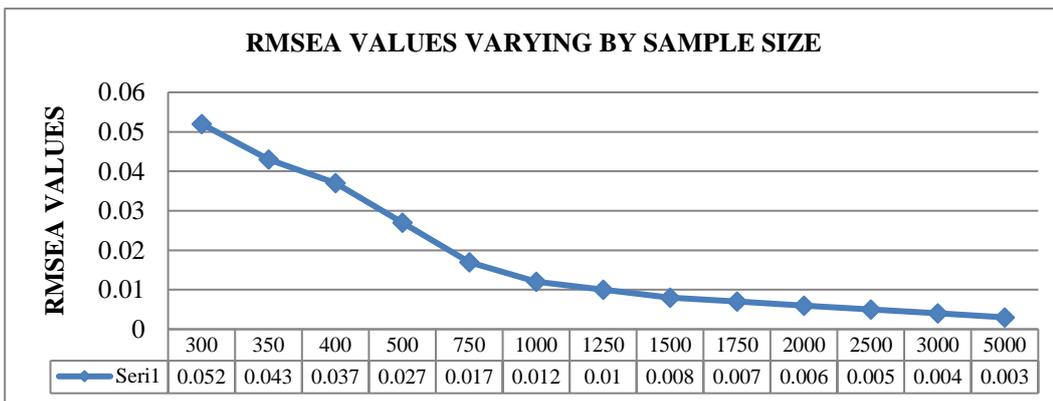


Figure 9. RMSEA Values varying by sample size through WLS method.

According to the sample size in Figure 9, the change in RMSEA value is seen. If the RMSEA value is between 0 and 0.5, it indicates a good fit. If the RMSEA value is greater than 0.1, it is an unacceptable fit. When Figure 9 is examined, in the smallest sample size, that is n=300, it is seen that the RMSEA value was calculated to be 0.052. It was observed that as the

sample volume increases, the RMSEA value decreases. However, after 1250 sample size it appears that the declines are very small. When it is assessed in this sense, it can be said that when working with categorical data, taking at least 1250 samples size will provide more consistent results with the researchers.

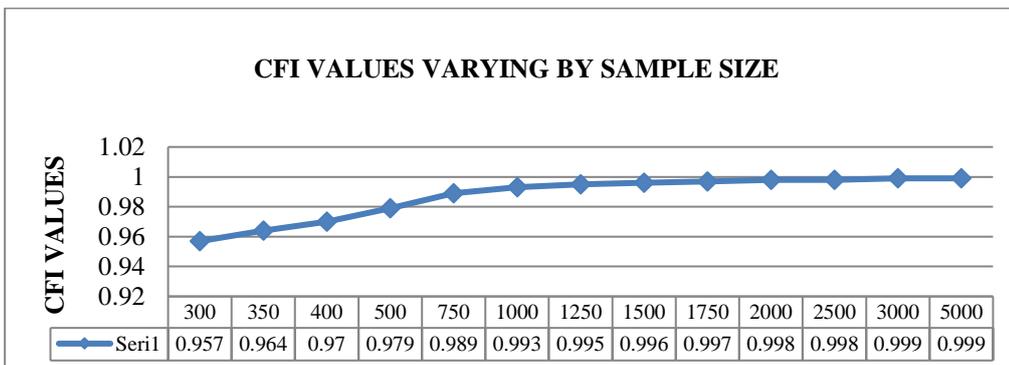


Figure 10. CFI values varying by sample size through WLS method.

Figure 10 shows the change in CFI values according to the sample size, which was analyzed according to the WLS method. As the sample volume increases, the CFI good fit index appears to increase. However, after the sample size of 3000, there is no change observed in

CFI value. When the CFI value is thought to be theoretically equal to maximum 1, it can be said that researchers need to work with a data set that has 3000 samples when the WLS method is preferred.

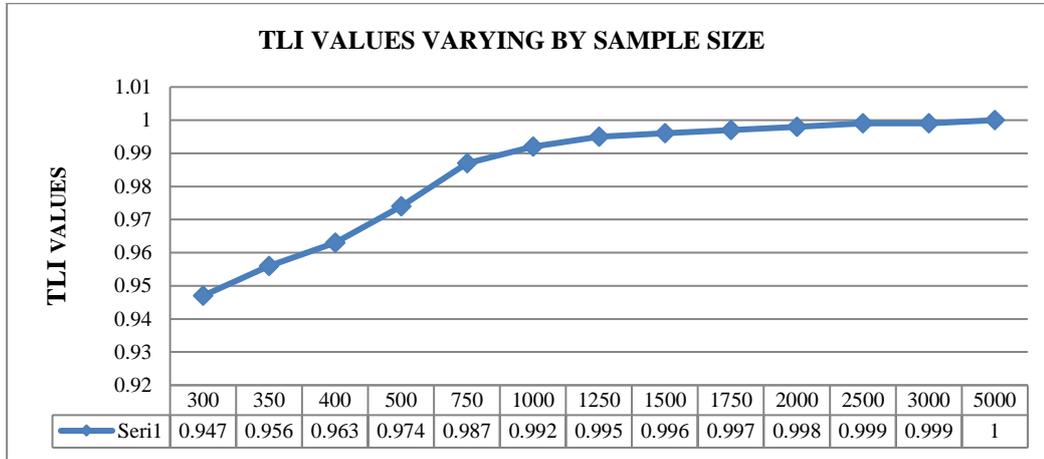


Figure 11. TLI values varying by sample size through WLS method.

The change of TLI goodness of fit index with the sample size can be seen in Figure 11. When the TLI index is between 0.97 and 1.00, it is considered a good fit. Furthermore, when the TLI index takes a value between 0.95 and 0.97, it is in an acceptable fit. If it

falls below 0.95, it is stated that the model is not compatible. In the TLI index estimated according to the WLS method, in the case of increasing sample size, it is seen that the mentioned index value also increases. The index takes its maximum value in 5000 samples.

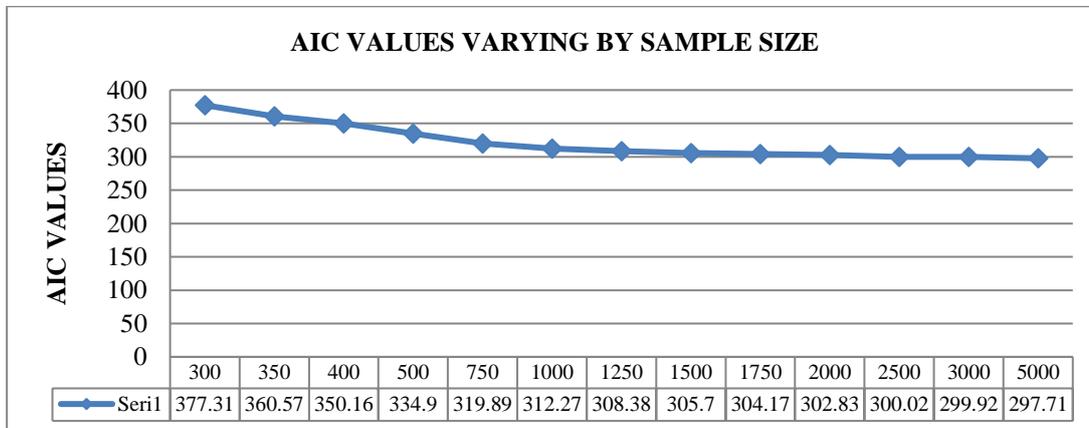


Figure 12. AIC values varying by sample size through WLS method.

In Figure 12, as an information criteria, it is seen how the AIC index changes with respect to the sample size. It is interpreted that the model with lower AIC value is better and more consistent model. It was determined in Figure 12 that if the sample size is 300, the AIC value

is 377.31 while the sample volume is 350, it is found to be 360.57. Along with an increased sample size, when the sample size is 5000, it was found to be 297.71. In this case, it can be said that the optimum sample size is 5000.

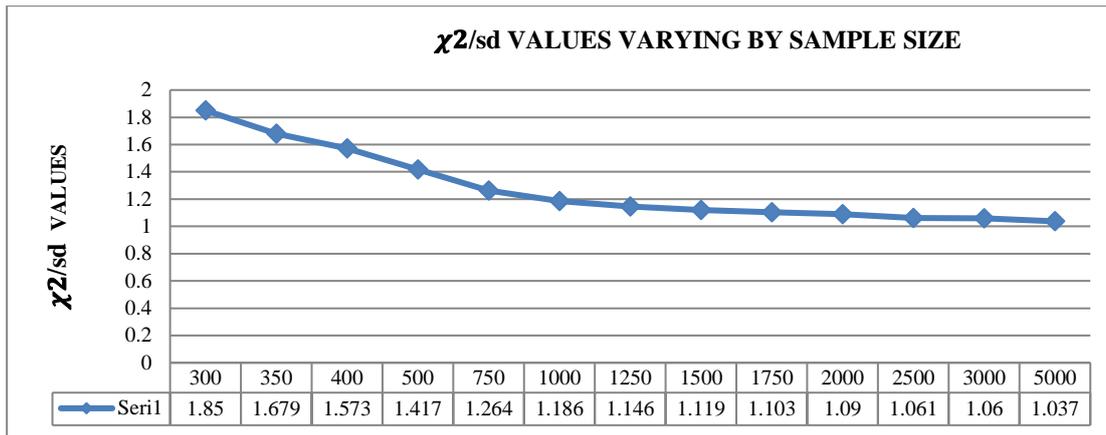


Figure 13. χ^2/sd values varying by sample size through WLS method.

Another index used in evaluating the fit of the theoretical model established is χ^2/sd value. If this value is between 0 and 2, it is an indicator of good fit. Also, if it takes a value between 2 and 3, it is in an acceptable fit. According to Figure 13, when the

sampling size is 300, it is calculated as $\chi^2/sd = 1.85$. As the sample volume increases, depending on the decrease of the Chi-Square value, χ^2/sd decreases. The lowest index value is calculated when n equals to 5000.

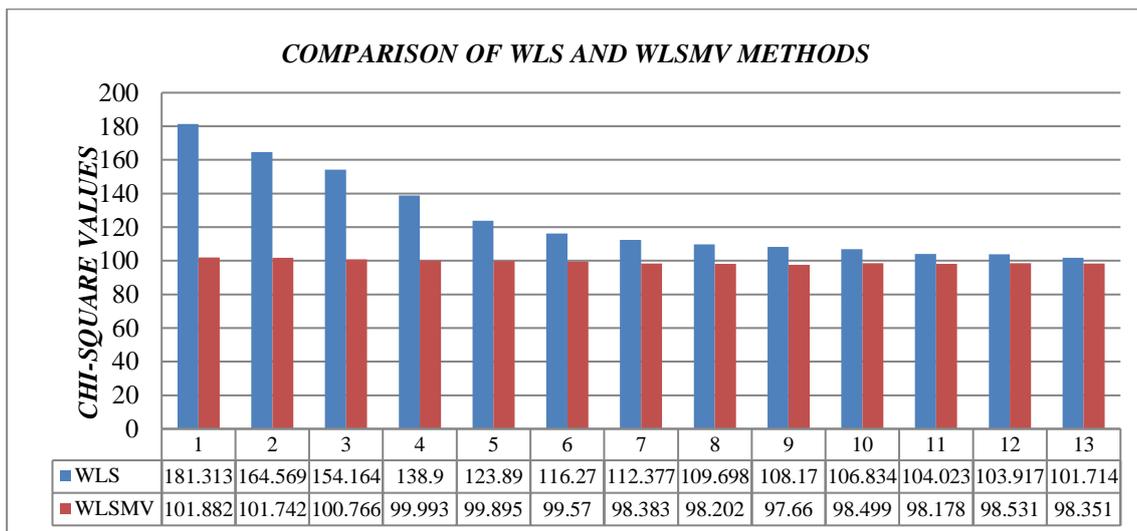


Figure 14. Comparison of WLS and WLSMV methods in terms of chi-square values.

In Figure 14, the comparisons of Chi-Square values in different sample volumes obtained with WLS and WLSMV methods can be seen. As a result, it is seen throughout the graph that the Chi-Square values obtained from the WLS method are higher than the Chi-Square values obtained from the WLSMV method. The difference is gradually decreasing with

increasing sample size. The difference is minimal when the sample size is 5000. The point that researchers should pay attention to in this case is that it is recommended that you select 5000 or more sample size if the WLS method is to be selected. On the other hand, if the WLSMV method is to be chosen, it can be suggested that the sampling size is more than 2000.

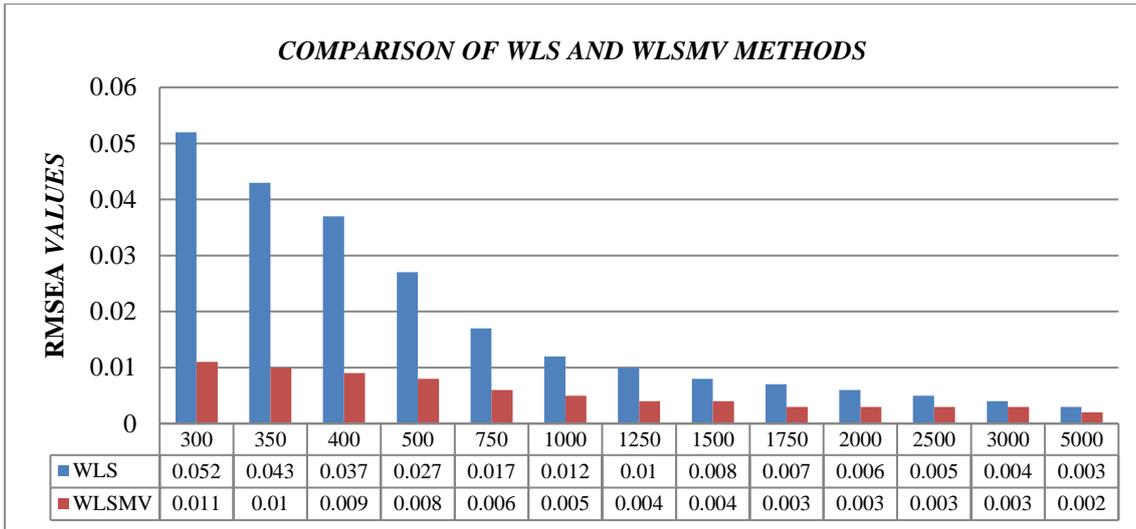


Figure 15. Comparison of WLS and WLSMV methods in terms of RMSEA values.

As RMSEA values estimated by WLS and WLSMV methods were compared, when n equals to 300, it noted that the estimated RMSEA value by the WLS method is higher than estimated by the WLSMV method in Figure 15. However, as the sample volume increases,

the difference between the two methods gradually decreases. For example, when the sample size is 350, the difference is 0.033 while n equals 5000, the difference is 0.001.

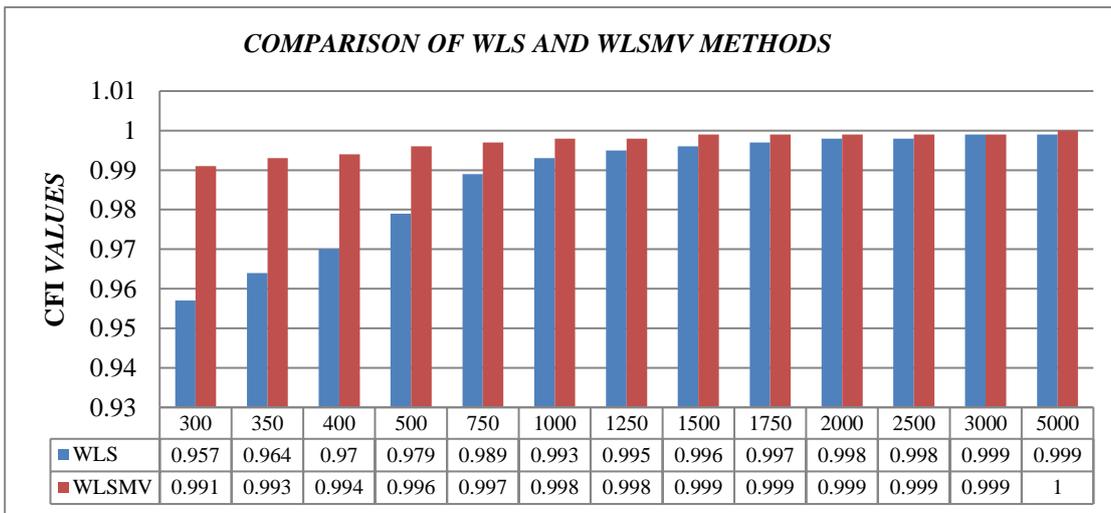


Figure 16. Comparison of WLS and WLSMV methods in terms of CFI values.

In Figure 16, you can find a comparison of the two methods in terms of CFI values. When the sample volume is 300, choosing the WLSMV method will supply with more consistent results. As the sample size increases, it is noticed that CFI values get closer to 1 for both methods. When the sample size is 3000, it was determined that the CFI values were equal for the two methods. It should be noted that the WLSMV method

is perfectly compatible with all sample size while it was figured out that WLS method reaches perfect fit interval after 400 sample size. As for the 300 and 350 sample sizes, it was determined that the WLS method is in the acceptable fit.



Figure 17. Comparison of WLS and WLSMV methods in terms of TLI values.

In Figure 17, you can see the comparison of the estimated TLI values by two methods. TLI values estimated by the WLSMV method reach the maximum value in 1250 sample size. On the other hand, the WLS method has reached a maximum in 5000 sample sizes.

The largest difference is 0.047 in 300 sample size while the difference is zero in 5000 sample size. 1250 sample sizes are sufficient for the WLSMV method in the TLI fit index whereas 2500 samples are sufficient for the WLS method.

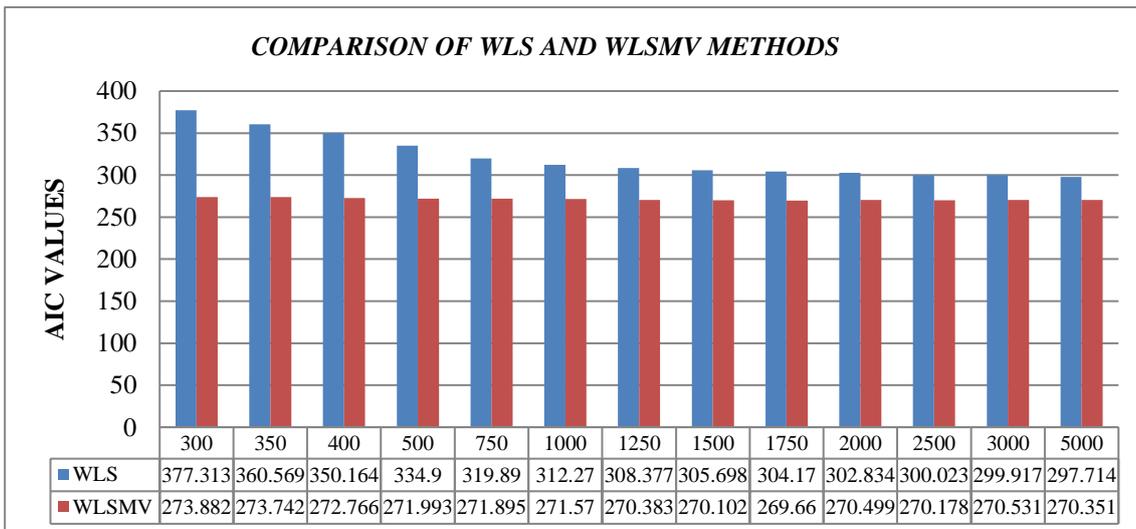


Figure 18. Comparison of WLS and WLSMV methods in terms of AIC values.

In Figure 18, you can find the changes in the two methods according to the sample size in terms of AIC information criteria. Figure 3.18 shows that the difference between the two methods size gradually decreases as the sample size increases. However, an important point here is that the AIC value calculated by the WLSMV method in all sample volumes is always lower than the WLS method. When it is evaluated in this sense, it can be said that the WLSMV method may be preferred to the WLS method.

It is included in the related sources that Chi-Square/sd should be less than 2. In Figure 19, you can find the

changes in these ratios according to sample size by two methods. The desired ratios have been reached in all sample sizes. However, it is necessary to pay attention is that the estimated ratio with the WLSMV method is always lower than the estimated ratio with the WLS method. Since the degree of freedom of the model used is the same, the method with the lower Chi-Square value will be lower than Chi-Square/sd value. As a result, it can be said that choosing the WLSMV prediction method for small sample sizes may give more favorable results.

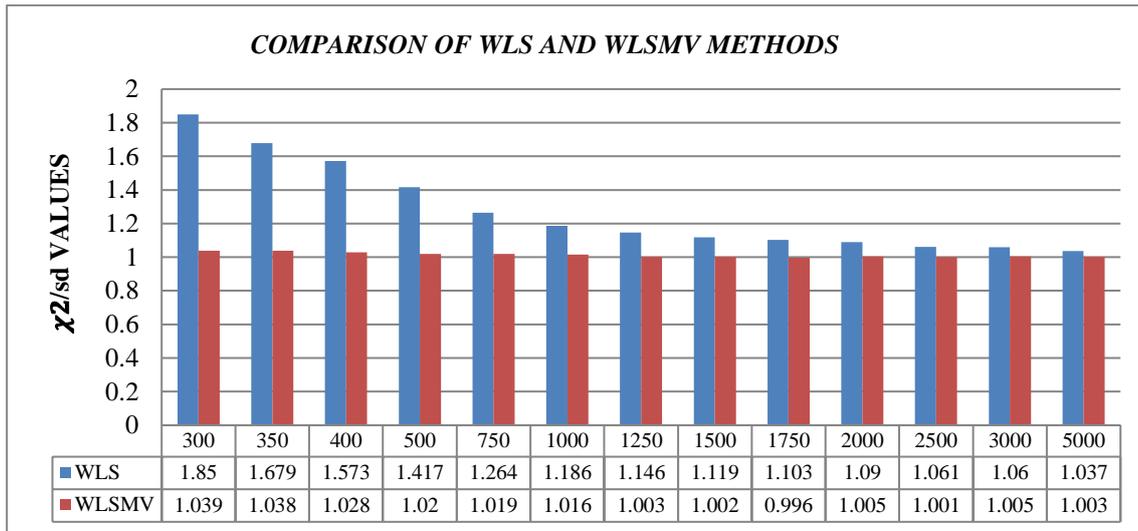


Figure 19. Comparison of WLS and WLSMV methods in terms of χ^2 / sd values.

4. Conclusion and Discussion

In this study, it was aimed to explain how the structural equation model makes analysis in Likert type categorical data. It was also intended in this study to find out how WLS and WLSMV estimation methods show a change between the theoretical model and the established model in different sample sizes in terms of the model fit measures and the goodness of fit indexes. In the literature, when we compared the simulation results with our simulation results, it has been worked sample size up to 1000 as the large sample size in other simulation studies [44,45] but we also showed the results for sample size up to 5000.

According to the WLSMV method, the optimal chi-square statistic should not exceed the sample size of 2000 in the simulation studies performed. On the other side, it was found out in the corresponding simulation results that at least 5000 sample sizes should be taken according to the WLS method.

If the RMSEA goodness of fit index takes a value between 0 and 0.05, it indicates that this index is in a good fit. Also, if the RMSEA is between 0.05 and 0.08, it is in an acceptable fit. In the simulation studies performed, it was determined that the RMSEA index in all sample sizes in the WLSMV estimation method is in good fit whereas, in the WLS estimation method, it was found out that in all sample size after 350 samples are in good fit. If researchers prefer the WLSMV method as a method of predicting when they

work with categorical data, taking 300 samples and more will suffice. However, if the researchers prefer WLS estimation method, 350 sample sizes will be sufficient.

The TLI goodness of fit index is an index that is less affected by sample volume. When this index value is between 0.95 and 0.97, it is in an acceptable fit. It was detected in the simulated studies that the TLI index value is in a good fit in all sample sizes in both WLSMV and WLS estimation methods. Researchers can take advantage of 300 sample sizes in the WLSMV estimation method. However, on the condition that TLI index takes a value of 1 continuously in the samples after 1250 sample sizes, it would be sufficient for researchers to use 1250 samples. This state takes place in the WLS estimation method in a sample size of 5000. That is, the TLI index takes the value of 1 in 5000 sample sizes.

In the simulation studies, the optimum AIC value in the WLSMV estimation method takes in 1750 sample sizes. However, it can be said that the optimum sample volume in the WLS estimation method is 1000. In other words, it is advised to the researchers that they select 1750 sample volumes in the WLSMV estimation method and 1000 sample volumes in the WLS estimation method.

Table 5. Recommended sample sizes for WLS and WLSMV estimation methods

Goodness of fit index	WLS	WLSMV
Chi-square	5000	1750
RMSEA	1750	1750
TLI	5000	1250
AIC	5000	1750
χ^2 / sd	5000	1750
CFI	3000	1500

As a result of the evaluations made for this study, you can see in Table 5 that in which sample sizes WLS and WLSMV parameter estimation methods showed the best performance according to the goodness of fit indexes. In general, according to Table 5, it was advised that the researchers who prefer the WLSMV estimation method should select 1750 sample size while the researchers who prefer the WLS estimation method should select 5000 sample size for chi-square of goodness of fit index.

RMSEA goodness of fit index showed the best performance in WLS and WLSMV estimation method with 1750 sample size. In WLS method, while 5000 sample size is sufficient, the sample size 1750 and 1250 is sufficient in WLSMV for TLI, AIC and χ^2 / sd of goodness of fit indexes. Finally, while WLS method shows the best performance in 3000 sample size, WLSMV shows the best performance in 1500 sample size for CFI index. More detailed comparisons of the indices are given in the related tables and graphs.

The most important aspects of this study that may contribute to the literature are that if a researcher wants to model a problem by structural equation model using Likert data, he can reach detailed information about which methods should be used. It will also help the researcher to find out in which sample size the estimation methods perform better and to get rid of the errors caused by the unnecessary, too much or misleading small sample sizes. In addition to that, it will help determine in which sample size the goodness of fit index has the optimum value in the chosen estimation method [47].

In terms of directing researchers to a future study, some of the indices such as ECVI, RMSR, SRMR and critical N values of absolute fitness goodness indices, RNR and CI from the decentralized goodness of fit index and PGFI and PNFI indices that fall outside the scope of the simulation in this study may be used. It can be investigated that how these indices perform in different sample volumes by the structural equation model in categorical data. This study may function as

a road map for studies recommended to the researchers.

In our further research will be another simulation study. Although unweighted least square and diagonally weighted least square than less efficient WLS [39,45], we include them in the planned simulation study for large sample size. Also, other robust ML estimators can be included and the other methods in ICS based on the Bayes factors in the estimation for comparisons.

References

- [1] Agresti A., Categorical Data Analysis, Second edition, John Wiley Sons, 2003.
- [2] Edward C., Wirht R.J., Houts C.R. and Xi N., Categorical Data in The Structural Equation Modeling Framework. Hoyle RH. (Ed.) Handbook of Structural Equation Modeling. Guilford Press, 2012;. 195-208.
- [3] Arıcıgil Ç., Sosyal Bilimlerde Kategorik Verilerle İlişki Analizi, (Second Edition), Ankara, Pegem Akademi, 2013.
- [4] Kateri M., Contingency Table Analysis Methods and Implementation Using R, New York, Springer, 2014.
- [5] Agresti A., Booth J.G., Hobert J.P. and Caffo B., Random-Effects modeling of Categorical Response Data. *Sociological Methodology.*, 30 (2000) 27-80.
- [6] Akıncı E.D., Yapısal eşitlik modellerinde bilgi kriterleri, Doctoral Thesis, Fen Bilimleri Enstitüsü, İstanbul, Mimar Sinan Güzel Sanatlar Üniversitesi, 2007.
- [7] Schumacker R.E. and Lomax R.G., A beginner's guide to structural equation modeling, Third edition, Routledge, Taylor and Francis Group, LLC 2010.

- [8] Likert R., A Technique for the Measurement of Attitudes, New York University 1932.
- [9] Lee S.Y., Structural Equation Modeling A Bayesian Approach, England, John Wiley Sons, 2007.
- [10] Arslan M.S.T., Ordinal değişkenli yapısal eşitlik modellerinde kullanılan parametre tahmin yöntemlerinin karşılaştırılması, Master Thesis, Fen Bilimleri Enstitüsü, Eskişehir, Eskişehir Osmangazi Üniversitesi, 2011.
- [11] Jöreskog K.G. and Sörbom D., LISREL 8: Structural Equation Modelling with the Simples Command Language, Scientific Software International, 1993.
- [12] Kline R.B., Principles and Practice of Structural Equation Modelling, London, Guilford Press, 2011.
- [13] Kline R.B., Principles and Practice of Structural Equation Modelling, London, Guilford Press, 2005.
- [14] Schumacker R.E. and Lomax R.G., A beginner's guide to structural equation modeling, Second edition, Mahwah, Lawrence Erlbaum Associates, 2004.
- [15] Çelik H.E., Yapısal eşitlik modellemesi ve bir uygulama: genişletilmiş online alışveriş kabul modeli, Doctoral Thesis, Fen Bilimleri Enstitüsü, Eskişehir Osman Gazi Üniversitesi, Eskişehir, 2009.
- [16] Agresti A., Analysis of Ordinal Categorical Data, John Wiley Sons, 2010.
- [17] Bollen K.A., Structural Equation with Latent Variables, New York, John Wiley Sons, 1989.
- [18] Çelik H.E., Yılmaz V., LISREL 9.1 ile Yapısal Eşitlik Modellemesi, Ankara, Anı Yayıncılık, 2013.
- [19] Muthén B. O., A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika.*, 49 (1984) 115–132.
- [20] Doğan İ., Farklı veri yapısı ve örneklem büyüklüklerinde yapısal eşitlik modellerinin geçerliliği ve güvenilirliğinin değerlendirilmesi. Doktora Tezi. Sağlık Bilimleri Enstitüsü. Eskişehir: Eskişehir Osmangazi Üniversitesi, 2015.
- [21] Akaike H., A new look at the statistical model identification, *IEEE Transactions on Automatic Control.*, 19(6) (1974), 716–723.
- [22] Akaike H., Information theory as an extension of the maximum likelihood principle. In: B.N. Petrov and F. Csaki (Eds.), Proceedings of the Second International Symposium on Information Theory, Akademiai Kiado, Budapest, 1973; pp 267–28.
- [23] Akaike H., Likelihood of a model and information criteria, *Journal of Econometrics.*, 16 (1981) 3-14.
- [24] Bozdoğan H., Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika.*, 52(3) (1987), 345-370.
- [25] Browne M.W. and Cudeck R., Alternative ways of assessing model fit. In: K. A. Bollen & J. S. Long (Eds.), Testing structural equation models, Beverly Hills, CA: Sage Publication, 1993; pp136-162.
- [26] Bozdoğan H., Akaike's Information Criterion and Recent Developments in Information Complexity. *Journal of Mathematical Psychology.*, 44 (2000) 62-91.
- [27] Bozdoğan H., Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-fisher information matrix. Information and Classification, Concepts, Methods and Applications Proceedings of the 16th Annual Conference of the "Gesellschaft für Klassifikation e.V." University of Dortmund, April 1–3, 1992.
- [28] Schwarz G., Estimating the dimension of a model, *The Annals of Statistics*, 6 (1978) 461-464.
- [29] Mallows C.L., Some comment on Cp, *Technometrics.*, 8 (1973) 661-675.
- [30] Cheung G.W. and Rensvold R.B., Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance, *Structural Equation Modeling.*, 9(2) (2002), 233-255.
- [31] Anderson D.R., Burnham K.P. and White G.C., Comparison of akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies. *Journal of Applied Statistics.*, 25(2) (1998), 263-282.

- [32] Muthén B.O., Latent variable structural equation modeling with categorical data. *Journal of Econometrics.*, 22 (1983) 48-65.
- [33] Muthén B.O. and Kaplan D., A comparison of some methodologies for the factor analysis of non- normal Likert variables. *British Journal of Mathematical and Statistical Psychology.*, 38 (1985) 171-189.
- [34] Jöreskog K.G., Latent variable modeling with ordinal variables. Paper presented at the international Workshop on statistical Modeling and Latent Variables in Trento, Italy, 1991.
- [35] Muthén B.O. and Satorra A., Technical aspects of Muthen's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, 60 (1995) 489-503.
- [36] J.R. Hipp J.R. and Bollen K.A., Model fit in structural equation models with censored, ordinal, and dichotomous variables: Testing vanishing tetrads, *Sociological Methodology.*, 33(1) (2003), 267-305.
- [37] Flora D.B. and Curran P.J., An empirical evaluation of Alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psycho Methods.*, 9(4) (2004), 466-491.
- [38] Forero C. G., Maydeu-Olivares M. and Gallardo-Pujol D., Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling.*, 16 (2009) 625–641.
- [39] Beauducél A., and Herzberg P: Y., On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling.*, 13 (2006) 186– 203.
- [40] DiStefano C., The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling.*, 9 (2002) 327–346.
- [41] Yang-Wallentin F., Jöreskog K.G. and H. Luo H., Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling.*, 17 (2010) 392–423.
- [42] Savalei V. and Rhemtull M., The performance of robust test statistics with categorical data, *British Journal of Mathematical and Statistical Psychology.*, 66 (2013) 201-223.
- [43] Rhemtulla M., Brosseau-Liard P.E. and Savalei V., When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM Estimation methods under suboptimal conditions, *Psychological Methods.*, 17(3) (2012) 354-373.
- [44] DiStefano C. and Morgan B.G., A comparison of diagonal weighted least squares robust estimation techniques for ordinal data, *Structural Equation Modeling: A Multidisciplinary Journal.*, 21(3) (2014) 425-438.
- [45] Li C.H., Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavioral Research.*, 48 (2016) 936–949.
- [46] Byrne M.B., Equation modeling Computer Software Snapshots of LISREL, EQS, Amos, and Mplus. Hoyle RH. (Ed.) Handbook of Structural Equation Modeling, Guilford Press, 2012; pp 195-208.
- [47] Gazeloğlu C., Methods for Estimating Weighted and Unweighted Parameters and Information criteria used in structural equation modeling of ordinal categorical data and comparison of results with different sample sizes. Unpublished doctoral thesis, Anadolu University, 2016.
- [48] Doğan İ. and Özdamar K., The effect of different data structures, sample sizes on model fit measures. *Communication in Statistics-Simulation and Computation.*, 46(9) (2017) 7525-7533 .