© 2014,
Science-Line Publication
www.science-line.com

ISSN: 2322-4770
Journal of Educational and Management Studies
J. Educ. Manage. Stud.,4 (1):22-38, 2014

JEMS

The Comparison of Two Methods of Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) in Testing Construct Validity of Achievement Goals

Mostafa Baghdarnia[1*], Razieh Firozy Soreh[2] and Raheleh Gorji[3]

[1.] Department of Educational Administration, Allameh Tabatabaei University, Tehran, Iran.
[2.] Department of Sociology, Islamic Azad University, Ashtian Branch, Ashtian, Iran
[3.] Department of Educational psychology, Tehran University, Tehran, Iran

* Corresponding author's Email: Baghdarnia@yahoo.com

ABSTRACT: The present essay was intended to compare between two methods of estimation, namely, Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) in determination of construct validity of achievement goals. The statistical population of this investigation includes all high school third grade students in the fields of empirical sciences and math- physics from Isfahan City during academic year 2005-6, that they were totally 7278. 400 students were chosen by means of stratified sampling technique with proportional assignment. To measure achievement goals, two subscales of mastery goals and performance goals were used that have been exclusively prepared to utilize in the field of mathematics by Middleton and Midgley. The Cronbach alpha coefficients for variables of mastery goals were obtained as 0.76 and 0.82 respectively. The results showed that if the observed variables are ordinal and their categories are lesser than 5, then it is important to select DWLS method as the best choice for estimation before estimating parameters and review of data distribution with respect to type of data. If researcher does not checkup assumptions and adapts default estimation method (ML) so the results may be biased and this may lead to making the erroneous decisions.

Keywords: Estimation Methods, Maximum Likelihood, Diagonally Weighted Least Square (DWLS), Construct Validity, Achievement Goals

ORIGINAL ARTICLE
Received 25 Dec. 2013
Accepted 05 Feb. 2014

INTRODUCTION

Structured Equation Model (SEM) has contributed researchers to formulate and test the new theories by examining simultaneously the correlation between several complex structures during recent years (Anderssen and Lorentzen, 2006). Analysis of structured equations includes five main steps. These steps are the determination, recognition, estimation, testing, and adjustment of model. At first step, researcher may formulate his/ her conceptual model based on the existing researches and theories. At the second phase, subject of identification of model is purposed. The model should be designed in such a way that number of its known parameter to be more than the unknown parameters. At third step, it is dealt with estimation of unknown parameters of the model. And at fourth step, the rate data fitness with the given model is examined. At fifth step, if goodness of fit index is not provided for the model then the model may be adjusted and then the new adjusted model can be evaluated (Schumacker and Lomax, 2004).

LISREL software is often used for confirmatory factor analysis and it has 7 estimation techniques. These techniques comprise of Instrumental Variables (IV), Two- Stage Least Squares (TSLS), Unweight Least Squares (ULS), Generalized Least Squares (GLS), Maximum Likelihood (ML), Weighted Least Squares (WLS), and Diagonally Weighted Least Squares (DWLS). These methods have several goals and they are various infrastructural assumptions (Jöreskog , 2001).

Estimation process includes application of a fitness function in order to minimize variance of sample covariance matrix and implied matrix. There are several fitness functions or estimation techniques. Some of these methods are Unweighted (or Ordinal) Least Squares (ULS or OLS), Generalized Least Squares (GLS), and Maximum Likelihood (ML). ULS estimations are compatible and they have not distributive assumptions or statistical correlation test so they are scale- dependent. Among all estimation techniques, only ULS method of estimation is scale- dependent. ML and GLS methods are scale- free (Schumacker and Lomax, 2004).

Confirmatory factor analysis is often done with ordinal data in social sciences since most of measurement tools include questions in LIKERT spectrum. Moreover, on most of occasions, data have no normal multivariate distribution. Data often violates from assumptions in some estimation

methods and this point should be considered upon estimation of model parameters. If the given assumptions are not noticed fitting parameters of the model may be biased so this leads to potentially improper decisions concerning to the theory. For example, values of the inflated Chi-2 and standard deviations increase the likelihood of error (type I). Likewise, under optimistic situation, over- estimation of parameters and fitness parameters may increase likelihood of error (type II) (Mindrila, 2004). Thus, when data are not distributed normally, using ML method may lead to weak results. Under such circumstances, application of alternative estimations like WLS and DWLS will provide better results. Nevertheless, with respect to data attributes like intensity of abnormality and sample size, the alternate estimations may not act better than ML technique (Kortmann, 2011). Hence, with respect to importance of estimation technique in analysis of structured equations, this article is intended to compare two common methods of maximum likelihood (ML) and Diagonally Weighted Least Squares (DWLS).

Theoretical Framework

Data are obtained usually by questionnaire in social sciences in which respondents are asked for selection of great number of questions in one of the limited classes. Choices of “very low” and “very high” in LIKERT scale often lead to distribution with positive skewness while inferential statistics is based on this assumption that data are normally and continually distributed. Multivariate normal distribution is the generalization of single variable normal distribution to higher orders. A matrix has multivariate normal distribution if all its linear elements have normal distribution. Abnormality may be shown by means of skewness and kurtosis. Skewness occurs when responses have higher frequency in some part of measurement scale and this factor effects on variance-covariance among variables. Kurtosis indicates widening of data distribution. The convex data are protruded more than normal distribution while the wide data are more leveled and dispersed on x-coordinate and they have lower frequency on y-coordinate. Both convex data and outliers effect on accuracy of statistical methods (Schumacker and Lomax, 2004). There is no consensus in the reasonable level of abnormality but on most of occasions, the cutoff values are selected as 2 for single variable skewness, 7 for single variable kurtosis, and 3 for multivariate kurtosis (Mindrila, 2004). What it to be noticed is to coming across data in which continuity assumptions of measurement scale and or skewness and kurtosis of normality do not apply to them. This point is noticed in model estimation.

ML is the most well- known method of estimation and it is more likely due to this fact that default option is in LISREL software (Schumacker and Lomax, 2004). This technique is an estimator for normal theory and it is assumed that sample size is adequate and observations are independent (randomly selected); the model has been properly formulated and data have multivariate normal distribution and are continuous. Like other estimators of normal theory (GLS), in ML method iterative estimation process is employed that minimizes variance among the observed covariance matrix in sample and implied covariance matrix of model. Model parameters, which are obtained by means of this technique, maximize likelihood of the existing data. This method has been suggested more than other estimators in normal theory since if a model has been formulated inappropriately then its results are less biased (Mindrila, 2004). Therefore, in general the assumptions for using ML estimation technique are as follows: 1) Sample is too big (asymptotic), 2) The observed variable scale is continuous, 3) The observed normal variable distribution is multivariate, and 4) The assumed model is valid (West et al., 1995; after Byrne, 1998). This model is not suitable for some data. As it mentioned above, variables are at ordered categorical scales (i.e. discrete variables with only a few possible values) in applied researches and researchers are usually interested in behaviors with abnormal distribution. In this type of studies, ML approach is not proportional to Pearson’s correlation and it has negative consequences in evaluation of fitness of model (Nye and Drosgove, 2011).

The important advantage of ML is in that it purposes statistical testing of model general fitness for the over- identified models. The other privilege of ML is that its estimators are scale invariant and scale free (Schermelleh- Engel et al., 2003). As a result, values of fitness function do not depend on this point that whether correlation matrix is used or covariance matrix or original data are employed or the transformed data (Schermelleh- Engel et al., 2003).

If data are continuous but abnormal then Asymptotically Distribution Free (ADF) method is the most common suggested estimation technique. ADF technique exists in LISREL software under title of Weighted Least Squares (WLS) and in EQS under title of Arbitrary Generalized Least Squares (AGLS).

Unlike ML method, raw data need for data analysis. Similarly, when some observed variables have been distributed orderly and some others continually, and distribution of variables is noticeably biased from normality and variables with dual values exist in the model, this method can be used (Schermelleh- Engel et al., 2003).

An important measurement problem emerges if the measured scale is ordinal. Under such circumstances, creation of ploy-choric correlation (correlation between two ordinal variables) and application of Weighted Least Squares (WLS) are commonly used in asymptotic covariance matrix (Jöreskog, 2001). Nevertheless, sample size is the major problem in this approach. The minimum needed sample size is unknown in this approach since its value varies with number of categories in each marker, quantity of observation in each question, dimension, and number of zero cells in agreement tables. Stimulation studies have indicated that WLS act poorly without a big sample size (more than 1500). This technique acts weakly with sample sizes less than 1000 for a model with 10 markers (Coursey and Pandey, 2007). Jöreskog (2001) suggests minimum sample size of 400 for only 6 markers and 1 dimension.

Generalized Least Squares (GLS) and Unweighted Least Squares (ULS) are some special cases of WLS estimation. Since GLS technique is based on ML assumptions so this estimation method is employed under the same conditions. But whereas it has weaker results with small sample sizes it is better to use ML method with small sample sizes. ULS has been derived from GLS technique. This method is benefitted from this advantage that compared to ML and DWLS, it leads to compatible estimator but in contrast to ML, it does not need to distributive assumptions. ULS constraints are due this fact that it may not present the most efficient estimations and it does not contain invariant scale and scale free (Schermelleh- Engelet al., 2003). Additionally, some of software programs do not present chi-2 statistics and standard errors when ULS is employed. Also some of software presents standard errors and ch-2 only multivariate normal assumption for ULS. Thus, ULS results should be cautiously interpreted (Schermelleh- Engel et al., 2003).

WLS estimations have some limitations. In addition to necessity for high sample size (e.g. according to views of Jöreskog (2001) in above sample size, the minimum size of sample for WLS models should be 1.5(p+10) for (p>12) where p is number of observed variables). The second constraint corresponds to weighted matrix used in WLS. Due to size of this matrix (W is matrix), WLS estimation contains complex computations. For example, when p = 50, W is a 210×210 matrix with 22155 separate arrays. As a result, Jöreskog (2001) have purposed DWLS estimator. In this approach, only diameter of weighted matrix (asymptotic variances) comprises of expression is used. Therefore, this method often reduces the amount of needed calculations to acquire a solution (Nye and Drosgove, 2011).

ML method purposes estimation of relatively accurate parameter with abnormal continuous data but due to abnormality, bias is increased in chi-2 and standard errors (Mindrila, 2004). Even when this model has been properly formulated, application of ML under abnormal multivariate conditions leads to inflation of chi square values, especially if data have distribution with positive kurtosis (leptokurtic). Consequently, fitness parameters are biased like Tucker- Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Comparative Fitness Index (CFI) that are types of chi square functions as well. Although ML offers accurate estimation parameter with abnormal continuous data but standard errors are under- estimated, particularly when data possess kurtosis (leptokurtic) 1 (Mindrila 2004).

The studies have shown that when you work on the ordered categorical observed variables it is better to employ DWLS method. WLS technique that is similar to DWLS that needs to great data and it produces biased testing statistics with N<500 (Kortmann, 2011). Wand and Kingham (Kortmann, 2011) found that polykurtic correlation matrix is used with DWLS method the model is often fitted with goodness of data. Due to the discrete nature of categorical data, some of authors considered it as abnormal intuitively (Merinda, 2004). But at the same time, when the ordered data have many categories and are almost normal so ML technique does not create extremely biased results. As the number of categories of responses reduces, bias and abnormality of multivariate scale increases. Since ML calculation methods are based on Pearson's momentum correlation methods (PPM) so if number of categories of responses is low, fitness indices, parameters estimations, and standard errors may be biased. When data are with the same order and abnormal, by application of ML, chi square and residue of mean square root (MRM) is inflated and normative normalized Fitness Index (NNFI) and Goodness of Fitness Index (GFI) are under- estimated. Furthermore, when data have skewness or kurtosis provided that responses category is low or sample size is small or if there is weak relations among factors and indices then bias increases in estimations of parameters and standard errors (Mindrila, 2004).

Similarly, using Pearson's correlation with the ordered data may under- estimated the power of relationships among variables (Cho et al., 2009). Application of polykurtic correlation coefficients may estimate non- biased parameters in confirmatory and exploratory factor analyses (Cho et al., 2009).

The other problem of working with bi- value or ordinal data is in that in factor analysis within some

software like SAS and SPSS it is assumed that there is linear relation between questions and factors. Given that such problems, many studies have suggested using polykurtic correlation as input data in confirmatory factor analysis of ordered or bi- value data (Cho et al., 2009). These data have shown that ordered data ranged at spectrum (2-7) and compared to Pearson's correlation, application of polykurtic correlation may lead to more accurate estimation of factor coefficients.

Jöreskog (2001) studied on this point that if the observed variables in all SEM analyses have ordered scale or a composition of ordered or relative scale then they consider categorical nature of these variables. In particular, they have implied that analyses should not be assumed based on Pearson's correlation in them and all variables should have continuous scale; however, analyses should be based on polykurtic or multi- categorical correlation and they should employ Weighted Least Squares (WLS). Polykurtic correlation coefficient is adapted for study on correlation among two ordered variables. If a relationship is considered among an ordered variable and continuous variable multi- categorical correlation coefficient is used and in some cases the ordered variable is of bi- value type, bi- categorical correlation coefficient is used.

Application of DWLS method has been prevalent as a method of estimation for the ordered factor analysis during recent years. One of the reasons for this point is that this method will be an invariant scale estimator if continuous markers are employed. Namely, if the fitted model is an invariant scale and its markers are continuous then DWLS technique for minimum functions may produce the identical fitness and parameters are converted into linear form when data become linear.

Due to the assumption of multivariate normal distribution, it is suggested generally to use ML technique only when bias from multivariate normal scale is dispensable. Besides, ML method may be also used for ordered data only when variables have at least 5 different values so they are treated like continuous values. DWLS method may estimate more accurate parameters when covariance or correlation matrix is computed and on the occasions the multivariate normal assumption is extremely biased and/ or data are ordinal. DWLS technique is resistant and based on polykurtic correlation matrix of the entered variables in this study. Few studies have been conducted regarding advantages and disadvantages of using this estimation method. Compared to WLS, this method may be employed with size of small data in great models, ordered data and with skewness. This method uses asymptotic variance of asymptotic covariance matrix presented by PRELIS (Merinda, 2004).

With study on SEM in journal of marketing, journal of marketing research, journal of consumer research, and journal of academy of marketing science during 1995-1999, they only used ML method in 16 cases, GLS technique in 5 times, WLS in 3 cases, and Elliptical Reweighted Least Squares (ERLS) techniques in 4 cases for this purpose while in many studies (46 researches) it has not been characterized that which method has been employed. Likewise and Breckler that has used 72 essays in journal of personality and social psychology which had employed structured equation modeling examined it and concluded that 19% of assumptions for theory of normality were reviewed and less than 10% have noticed to this point that whether assumptions of normality theory has been violated or not. The fact that in GLS and ML estimators it assumed that the observed variables have multivariate normal distribution is important for estimation techniques in SEM. Violation from these assumption may manipulate standard error of route coefficient between latent variables and test statistics. As a result, with respect to extreme kurtosis and skewness, the observed variables are examined in such a way that asymptotic efficiency is not usually considered and this reduces the accuracy of asymptotic covariance matrix, chi square, and t-tests. For example, while most of the related sizes to agreement are skewed negatively (Anderssen et al., 2006) and quantitative issues are reported about kurtosis, extreme bias from this assumption that the observed variables with multivariate normal distribution may create some problems in validity of statistical conclusion (Anderssen et al., 2006).

MATERIALS AND METHODS

The studied samples

The statistical population in this study comprised of all high school third graders who have studied in the fields of experimental science and math- physics during academic year 2005-6 that were totally 7278. Although there is no general agreement over the optimal sample size for such researches, according to Jöreskog and Sörbom regarding choosing of maximally 30 respondents versus any observed variable (and in order to increase accuracy in estimation, 400 students (200 females and 200 males) were elected from five educational areas by means of classified sampling with proportional assignment.

Tools

- Achievement goals: To measure achievement goals, two subscales of proficiency goals and performance goals were used, which have been prepared exclusively in mathematical field by

Middleton and Midgley (1997). Proficiency goals scale denotes student's focus on growth, skill, and proficiency in assignments and internal value of learning and it has 5 questions while performance goals scale also suggests student's focus on his/ her performance in comparison with others and being seemed as top student where 4 questions have been tested in this scale. The rate of alpha coefficient which reported by Middleton and Midgley for these two scale was 0.84 while in the present study, alpha coefficient values have been derived as 0.76% for proficiency goals scale and 0.82 for performance goals scale. All questions have been prepared within LIKERT spectrum ranging from "completely agree" to "completely disagree" and students should express their comment about each of questions with respect to lesson of math in current year.

RESULTS

At first, descriptive statistics (mean, standard deviation, skewness, and kurtosis) were reported for total sample group of (400) students and analyzed and presented in Table (1). Mean value and the computed standard deviation signify this point that scores are appropriately dispersed. Also two statistics of skewness and kurtosis suggest this fact that data dispersion in each variable is normally distributed.

Table 1: Descriptive indices of questions in achievement goals scale

Variables	Mean	Standard deviation	Skewness	Kurtosis
Proficiency goals	19.54	3.60	-0.93	1.23
Performance goals	14.82	3.70	0.57	-0.14

Table 2: GFI indices of bi- factor structure for achievement goals in two method of fitting

Indices	ML	DWLS
GFI	0.94	0.98
AGFI	0.90	0.97
RMSEA	0.09	0.06
X^2	115.92	65.29
d.f	26	26

In the following, in order to compare two methods of fitness, initially confirmatory factor analysis was conducted with the maximum likelihood to estimate the model and then DWLS was examined. Some of indices like Chi- square ( ), ratio of Chi Square to degree of freedom ( ), Comparative Fitness Index (CFI), Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), and Root Mean Square Error of Approximation (RMSEA) for fitting of both model may be observed in the following table.

In general, there are several fitness indices for evaluation of factor analysis models so here some of them have been utilized including indices of chi-2, Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), and Root Mean Square Error of Approximation (RMSEA).

If sample size is 75-200, Chi Square is an appropriate index for fitting but for any model with greater n, Chi Square is almost statistically significant all the times. Under such conditions, with respect to a series of reasoning, degree of freedom is purposed as a value based on which magnitude of can be measured. Although this index lacks a fixed criterion for a reasonable model, the values less than 3 may be considered as appropriate parameter for goodness of fit.

Value of Goodness of Fit Index (GFI) and Adjusted Goodness of Fit Index (AGFI) range from zero to one so the more this index approaches to one the greater goodness of fit index it has. But Root Mean Square Error of Approximation (RMSEA) should be minimized as possibly while the values lesser than 0.05 indicate perfectly appropriate fitness and this value is appropriate up to 0.08 and the fitness will be weak up to 0.1 or higher. As it shown in table above, fitness indices for the second model i.e. fitting by DWLS technique are better than by ML method in terms of fitness. Standardized parameters indicate orientation of goal in both fitting methods.

Table 3: Standardized coefficients, measurement error, t- value and its significance level for any question

Factor	Question	Route standardized coefficients		Standard Error Measurement (SEM)		t- value	
		DWLS	ML	DWLS	ML	DWLS	ML
Proficiency goals	1. I intend to learn deeply mathematical concepts in math class	0.67	0.64	0.10	0.10	13.20	12.64
	2. I solve additional exercises rather than textbook exercises in order to become proficient over mathematical concepts	0.41	0.39	0.11	0.10	6.76	7.10
	3. I like math assignments that make me to think	0.68	0.69	0.12	0.12	14.13	13.61
	4. One of my goals in math class is to learn more new knowledge as I could	0.71	0.70	0.10	0.09	12.93	13.93
	5. I like that mathematical exercises to become difficult to the extent that I could learn something from them	0.62	0.64	0.07	0.08	12.52	12.50
Performance goals	1. I like to show my math teacher that I am topic students than other	0.64	0.60	0.14	0.14	13.42	12.08
	2. It is important for me that my classmates think I am proficient in lesson of math	0.71	0.73	0.08	0.09	17.65	15.52
	3. It is important for me to become more proficient in lesson of math than other students	0.82	0.83	0.08	0.10	22.25	18.29
	4. It is important for me to be seemed proficient in lesson of math compared to other my classmate	0.75	0.76	0.12	0.13	16.76	16.50

Significance of questions factorial load is considered with respect to t-tests and the existing coefficients above the route between measurement indices and factors as its relevance of each question. According to above table, standardized coefficients ranged from 0.39 to 0.83 and all routes are significant.

DISCUSSION

The results were acquired from comparison of two estimation methods in order to examine that when number of observed ordinal variables and their categories is less than 5, DWLS method has Chi square value smaller than in ML. Thus, it is better to select parameter estimation methods properly in confirmatory factor analysis with respect to this point that to what extent data could estimate their assumptions accurately. If data are continuous and uniformly distributed ML method is the best option. This estimation method is effective often when the ordered data are not treated like continuous data especially if there are some quantitative classifications. Over- estimation error occurs when data could not estimate multivariate normal assumption and model fitness is more resistant tan type of variable and its abnormality. Therefore, it is crucially important to review data distribution before parameters estimation and considering type of variable through selection of the best estimation method. If researcher does not examine these assumptions and simply uses software default estimation technique the results may be biased and it may lead to false decisions regarding tested model. The results of this study were in line with results of Mindrila (2004). Similarly, findings of the given research corresponded to this study since the results of both researches showed that Lambda values are the same in both methods. Despite of findings of Mindrila (2004), when data are continuous and normally distributed ML method is the most accurate technique but such a condition is rare in studies of behavioral sciences. With the presence of ordered data, this method is more sensitive to type of variable than normal distribution. Nevertheless, the impact of ordered data is reduced when number of categories becomes greater.

This point means that questions are related to its measurement factor. As it observed, values derived in both methods of estimation are identical and near to each other.

REFERENCES

Anderssen, T.W., Lorentzen, G. & Olsson, U.H. (2006).The impact of no0normality and estimation methods in SEM on satisfaction research in marketing. Quality & Quantity, 40:39–58.

Breckler, S.J. (1990). Application of covariance structure modeling in psychology: Cause for concern? Psychological Bulletin, 107 (2): 260–273.

Byrne, B.M. (1998). Structural equation modeling with LISREL, PERLIS and SIMPLIS: Basic concepts, applications and programming. New Jersey: Lawrence Erlbaum Association, publishers.

Carlos G. Forero, Alberto Maydeu-Olivares, and David Gallardo-Pujol (2009). Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. Structural Equation Modeling, 16:625–641.

Cho, S et al. (2009).Accuracy of the Parallel Analysis Procedure with Polychoric Correlations. Educational and Psychological Measurement, 69 (5):748-759.

Coursey, D.H. & P Andey, S.K. (2007). Public service motivation measurement, testing an abridged version of Perry's proposed scale. Administration & Society Vol. 39 Number 5: PP. 547-568.

Jöreskog, K. (2001). Analysis of ordinal variables 2: Cross-sectional data. Unpublishedmanuscript.http://www.ssicentral.com/lisrel/ordinal.htm

Kortmann, S.C.J.J. (2011). Confirmatory Factor Analysis of the Bandung Family Relation Test: A Simulation Study Comparing ML, DWLS and WLS Estimation.

Middleton, M.J. & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An under explored aspect of goal theory. Journal of Educational psychology, 89 (4), 710-718.

Mindrila, D. (2004). Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) Estimation Procedures: A Comparison of Estimation Bias with Ordinal and Multivariate Non-Normal Data. International Journal of Digital Society (IJDS), 1(1): 1-7.

Nye, C.D. & Drasgow, F. (2011). Assessing goodness of fit: simple rules of thumb simply do not work.Organizational Research Methods 14(3) 548-570.

Olsson, U.H., Troye, S.V. & Foss, T. (2001). The role of fit assessment in structural equation modeling: Can fit be compared across estimation methods? Working paper, Norwegian School of Management.

Schermelleh, K., Moosbruger, H. & Muller, H. (2003).Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods of Psychological Research, 8(2):23-74, available at: http://www.mpr-online.de

Schumacker, R.E. & Lomax, R.G. (2004). A beginner's Guide to Structural Equation Modeling. Lawrence Erlbaum Associates, Mahwah, N.J.