

Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) Estimation Procedures: A Comparison of Estimation Bias with Ordinal and Multivariate Non-Normal Data

Diana Mîndrilă
University of South Carolina, USA

Abstract

This paper aims to identify the effect of using the maximum likelihood (ML) parameter estimation method when data do not meet the assumption of multivariate normality and are not continuous. Both ML and the diagonally weighted least squares (DWLS) procedure were applied to simulated sets of data, which have different distributions and include variables that can take different numbers of possible values. Results were also compared to the ideal situation of a data set consisting of continuous, normally distributed variables. Outcomes indicate that ML provides accurate results when data are continuous and uniformly distributed, but is not as precise with ordinal data that is not treated as continuous, especially when variables have a small number of categories and data do not meet the assumption of multivariate normality. In contrast, DWLS provides more accurate parameter estimates, and a model fit that is more robust to variable type and non-normality.

1. Introduction

LISREL 8 software is frequently used for confirmatory factor analysis and provides a choice of seven estimation methods of parameter estimation: instrumental variables method (IV), two stage least squares (TSLS), unweighted least squares (ULS), generalized least squares (GLS), maximum likelihood (ML), weighted least squares (WLS), and diagonal weighted least squares (DWLS) [4]. These methods have different purposes, as well as different underlying assumptions.

In social sciences, confirmatory factor analysis is frequently conducted with ordinal data, because

measurement instruments often consist of Likert scale items. Furthermore, in numerous situations data do not have a multivariate normal distribution. These aspects violate the assumptions of some estimation methods, and should be taken into account when estimating model parameters. When assumptions are not taken into consideration, model fit indicators may be biased, and lead to potentially incorrect decisions about the hypothesized theory. For instance, inflated chi-square values and standard errors may increase the probability of Type I error, because in such situations it is more likely to reject a correctly specified model than expected by chance. Similarly, overoptimistic parameter estimates and fit indices may increase the probability of Type II error, because researchers are more likely to accept a potentially misspecified model.

2. Theoretical Framework

Inferential statistics rely on the assumption that data are normally distributed. The multivariate normal distribution is a generalization of the univariate normal distribution to higher dimensions. A matrix has a multivariate normal distribution when all of its linear components are normally distributed.

Non-normality is indicated by the degree of skewness and kurtosis. Skewness occurs when responses are more frequent at one part of the measurement scale, and affects the variance-covariance among variables. Kurtosis reflects the flatness in data distribution. Leptokurtic data are more peaked than normal, whereas platykurtic data are flatter and more dispersed on the X-axis, with low frequencies on the Y-axis. Both leptokurtic and platykurtic data impact the accuracy of statistical procedures [9]. There is no consensus regarding an acceptable degree of non-normality, but most often

cut-off values are 2 for univariate skewness, 7 for univariate kurtosis, and 3 for multivariate kurtosis [5].

ML is the most popular estimation procedure, most likely because it is the default option in LISREL [9]. It is a normal theory estimator, and assumes that the sample has an adequate size, observations are independent (randomly selected), the model is correctly specified, and data are multivariate normal and continuous. Like other normal theory estimators (GLS), ML employs an iterative estimation process, which minimizes the difference between the observed sample covariance matrix and the model implied covariance matrix. The model parameters obtained with this method maximize the likelihood of observing the available data if one were to collect data from the same population again. This estimation procedure is recommended over other normal theory estimators, because its results are less biased when the model is misspecified [5].

With non-normal continuous data, ML produces relatively accurate parameter estimates, but the bias in chi-square and standard errors increases with non-normality [2]. Even when the model is correctly specified, the use of ML in conditions of multivariate non-normality results in inflated chi-squares, particularly when the data have a leptokurtic distribution [3]. Consequently, fit indices such as the Tucker-Lewis Index (TLI), the root-mean square error approximation (RMSEA), and the comparative fit index (CFI), which are functions of chi-square, are also biased. Although ML produces accurate parameter estimates with non-normal continuous data, the standard errors are underestimated, especially when data are leptokurtic. [6]

Due to the discrete nature of categorical data, some authors consider it to be inherently non-normal [8]. However, when ordinal data have a large number of categories and are approximately normal, ML does not produce severely biased results. Bias tends to increase as the number of response categories decreases, and multivariate non-normality increases. Because ML computational procedures are based on Pearson product-moment (PPM) correlational techniques, when the number of response categories is small, the fit indices, parameter estimates, and standard errors can be biased. [5]

When data are both ordinal and non-normal, using ML inflates the chi-square and the root mean square residual (RMR), and underestimates the non-

normed fit index (NNFI), and the goodness of fit index (GFI). Furthermore, the bias in parameter estimates and standard errors increases when data are skewed or kurtotic, when there are few response categories, the sample is small, or the relationships between factors and indicators are weak [1].

Because of the assumption of multivariate normal distribution, it is generally recommended to use ML only when the violations of multivariate normality are only slight. Additionally, ML can be used with ordinal data only if variables can take at least 5 different values, and they are treated as continuous when computing the correlation or covariance matrix [10].

In situations in which the assumption of multivariate normality is severely violated and/or data are ordinal, the diagonally weighted least squares (DWLS) method provides more accurate parameter estimates. The DWLS is the robust WLS method, and is based on the polychoric correlation matrix of the variables included in the analysis. It uses only the diagonal of weights in inversion, and all weights in estimation of fit and standard error. There is little research on the advantages and disadvantages of using this estimation procedure. As opposed to WLS, this method can be used with small sample sizes, large models, as well as skewed and ordinal data. It uses the asymptotic variance from the asymptotic covariance matrix, which is produced by PRELIS [10].

3. Data Sources

The five data sets used in the study were artificially generated and have a sample size of 500 cases, which meets the requirement of 5-20 cases per parameter estimate. One data set represents the ideal situation of having a perfectly normal distribution and continuous variables. The other four data sets consist of ordinal variables. Two of them have three category data, whereas the other two have seven category data. In both cases, one data set is multivariate normal, while the other is non-normally distributed (Table 1). All data sets were screened using Jöreskog and Sörbom's PRELIS software. By default, PRELIS recognizes categorical data and treats it accordingly, by using polychoric correlation to compute covariance matrices. Polychoric correlation allows the estimation of the correlation between theorized normally distributed continuous

latent variables, by using the corresponding observed ordinal variables.

Additionally, for each data set PRELIS was also used to compute asymptotic covariance matrices for categorical variables. These matrices are not required to apply the ML method, but must be included when using DWLS. They are not used in iterations, and therefore, do not affect parameter estimates, and do not have to be inverted. However, they should be used to compute robust standard errors and Chi-Squares, which are corrected for non-normality.

In addition to these procedures, ordinal variables were also treated as continuous, and covariance

matrices were computed using a simple Pearson product moment correlation.

4. Confirmatory Factor Analysis

All data sets were used to replicate the true model illustrated in Figure 1, which has 20 variables and five factors (four variables per factor). All the variables in the true model have factor loadings of .7, and all factor correlations are .3.

As indicated in Table 2, the T rule [4] shows that the model is overidentified: it provides sufficient information to estimate the requested parameters and enough degrees of freedom are left to be able to compute and compare fit indices.

Table 1. Data sets.

Continuous	Ordinal			
	3 Category Data		7 Category Data	
	Normal	Non-Normal	Normal	Non-Normal
500	500	500	500	500

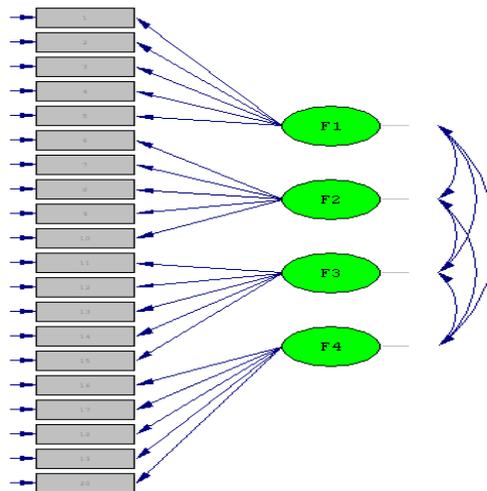


Figure 1. Estimated model.

Table 2. Using the T rule to determine whether the estimated model is overidentified.

Available Information	Information to Estimate	Degrees of Freedom
210	46	164
	<i>Errors</i> 20	
	<i>Loadings</i> 20	
	<i>Relationships between Factors</i> 6	

With each one of the four ordinal data sets, confirmatory factor analysis was conducted using three different procedures. In the first procedure, ordinal variables were treated as continuous, and parameters were estimated using the default LISREL method (ML). In the second procedure variables were treated as ordinal (default option), and parameters were estimated using the DWLS method, for which the asymptotic covariance matrix was also computed. The last procedure treated variables as ordinal, and used ML the estimation method without including the asymptotic covariance matrix. In summary, the first method uses the default estimation procedure, the second method uses the default variable type, and the third method uses default options for both variable type and estimation procedure. For comparison, the model was also estimated using the normally distributed data with continuous variables, using the ML procedure. Parameter estimates and goodness of fit indices resulted from all methods were compared to determine whether or not they might lead to different results.

5. Results

As indicated in Table 3, with the three category and the seven category data, the DWLS method estimates factor loadings, standard errors and factor correlations most precisely (closest to the true model), with both multivariate normal and non-normal distributions.

Although factor loadings are closer to .7 when non-normal three category data is treated as ordinal and ML is used, the standard errors are very large and the factor correlations are poorly estimated. Results depend more on the correct identification of

variable type, rather than addressing the variable distribution. Even when the use of ML is justified by normal distribution and variables are treated as continuous, results are biased because this method is not adequate for categorical data. DWLS has less restrictive assumptions, but the statistical quality of the estimates remains to be determined. For the three category data, this method produces a relatively large number of significant modification indices, suggesting that some items should load to different factors (this does not coincide with the true model).

Maximum likelihood is the most precise method, when the data are continuous and normally distributed. However, this is rarely the case in social science research. With ordinal data, maximum likelihood was more sensitive to variable type than normal distribution. Nevertheless, the impact of ordinal data is attenuated when the number of categories is larger.

In addition to factor loadings, standard errors, factor correlation and modification indices, goodness of fit indices for each solution were compared to determine which of the estimating procedures produces the model that best fits the data.

As shown in Table 4, chi-square values associated with each estimating procedure are significant. However, the chi-square test is sensitive to both sample and model size, and can lead to the inappropriate rejection of plausible models. Therefore chi-square divided by the degrees of freedom was used as an index of model fit. Generally, values lower than 3 indicate a good model fit. The lowest Chi-square/df ratios occur when ML is used and data is treated as continuous, (or has a larger number of categories). However, ML artificially inflates the model fit, whereas DWLS computes robust Chi-Squares and subsequent indices, by correcting for non-normality.

The RMSEA index estimates how well the proposed model approximates reality [7]. Values between .05 and .08 indicate a fair model fit, whereas values smaller than .05 show excellent fit. In most cases, DWLS method has the lowest RMSEA index (close to 0), indicating that this method has the lowest average error left unaccounted for, and the model fits the data almost perfectly.

Table 3. Range of parameters estimates and number of significant modification indices for each data set and estimation procedure.

	Loadings	Standard Errors	Factor Correlations	Significant Modification Indices
<i>Norm Cont ML</i>	(.65-.71) all significant	(.48-.58) all significant	(.23-.39) all significant	0
3NCont ML	(.53-.71) all significant	(.53-.72) all significant	(.26-.31) all significant	0
3Ncateg DWLS	(.60-.77) all significant	(.40-.64) all significant	(.27-.33) all significant	2
3Ncateg ML	(.59-.78) all significant	(.39-.66) all significant	(.26-.32) all significant	7
3NNCont ML	(.43-.57) all significant	(.82-.62) all significant	(.06-.28) all significant	0
3NNcateg DWLS	(.58-.77) all significant	(.41-.66) all significant	(.09-.35) all significant	10
3NNcateg ML	(.62-.74) all significant	(.48-.66) all significant	(.08-.36) all significant	0
7NCont ML	(.56-.71) all significant	(.50-.69) all significant	(.31-.37) all significant	2
7Ncateg DWLS	(.65-.72) all significant	(.48-.67) all significant	(.31-.37) all significant	0
7Ncateg ML	(.57-.72) all significant	(.48-.68) all significant	(.31-.38) all significant	3
7NNCont ML	(.47-.67) all significant	(.55-.76) all significant	(.19-.36) all significant	2
7NNcateg DWLS	(.63-.80) all significant	(.36-.67) all significant	(.19-.43) all significant	2
7NNcateg ML	(.57-.78) all significant	(.39-.67) all significant	(.18-.42) all significant	29

Table 4. Fit indices.

	Chi-Square	df	Chi-Square/ df	RMSEA (CI 90%)	SRMR	ECVI (CI 90%)	AIC	NNFI	AGFI
Norm Cont ML	165.46	164	1.01	.0042 (.00-.02)	.029	.52 (.51-.58)	256.46	1.0	.96
3NContML	129.67	164	0.79	0.0 (.0-.0)	.027	.51(.51-.51)	221.67	1.01	.97
3NcategDWLS	260.29	164	1.59	0.0 (.0-.0)	.032	.51(.51-.51)	224.31	1.01	.99
3NcategML	259.86	164	1.58	.034 (.03--.04)	.033	.71(.63-.80)	351.86	.98	.94
3NNContML	165.61	164	1.01	.004 (.00-.07)	.037	.52(.51-.59)	257.61	1.0	.96
3NNcategDWLS	1116.84	164	6.81	0.0 (.0-.0)	.076	.51(.51-.51)	6062.50	1.01	.98
3NNcategML	1090.85	164	6.65	.11 (.10-.11)	.077	2.37(.22-.59)	6062.5	.77	.77
7NContML	182.73	164	1.11	.015 (.0-.07)	.033	.55(.51-.62)	274.73	1.00	.95
7NcategDWLS	214.46	164	1.31	.014 (.0-.025)	.034	.55(.51-.62)	272.03	1.00	.99
7NcategML	213.65	164	1.03	.025 (.01-.03)	.035	.61(.54-.70)	305.65	.99	.95
7NNContML	188.89	164	1.51	.017 (.0-.03)	.038	.56(.51-.64)	280.89	.99	.95
7NNcategDWLS	549.19	164	3.34	0.0 (.0-.01)	.054	.51(.51-.53)	233.14	1.0	.98
7NNcategML	540.48	164	3.29	.068 (.062-.074)	.055	1.27(1.13-1.42)	632.48	.92	.87
Benchmark			< 3 acceptable fit	0=perfect fit <.05 good fit <.08 fair fit >.10 poor fit		Lowest value shows better fit.		>.9 acceptable fit >.95 good fit	

SRMR reflects the size of the fitted residuals, with small values indicating a better fit [7]. Regardless of the number of data categories and variable distribution, the ML estimation method produces the lowest SRMR values, if the data is treated as continuous.

The ECVI value shows how well the proposed model is expected to cross-validate, and low values indicate good model-fit [7]. Results obtained with the different estimation methods show that in most cases the model obtained with DWLS is most likely to cross validate. The AIC index is used to compare models with different numbers of latent variables, taking into account the model complexity and fit [7]. AIC values that are closest to 0 show the most parsimonious model. Regardless of data distribution, the maximum likelihood method produces a lower AIC index when the data has only three categories, and is treated as continuous. When the data has seven categories and the DWLS method is used, AIC indicates a more parsimonious model.

NNFI indicates how much better a model fits the data relative to a model with no structure, assuming that sampling error explains covariation among measured variables [7]. The closer its values are to one, the better is the model fit. When the data are normally distributed, both ML (when variables are treated as continuous) and DWLS provide high NNFI values. Nevertheless, when the assumption of multivariate normality is not met, NNFI indicates that the DWLS method produces a model that better fits the data.

AGFI indicates how much variance the proposed model accounts for [2]. Values above .9 are generally considered acceptable, while an AGFI value of .95 indicates that the proposed model fits the data very well. This index is used to compare the fit of different models with the same data, and is adjusted for the degrees of freedom of a model relative to the number of variables. It is, therefore, an important index to consider in interpreting the results of our analysis. In all cases, regardless of data distribution and number of data categories, the best AGFI values are recorded when the data is treated as ordinal, and the DWLS method is used to estimate model parameters.

Goodness of fit analysis also indicates that in all situations, estimation is less precise when both default LISREL options are used: data is treated as categorical, and parameters are computed based on maximum likelihood. This outcome is not unexpected, because there is a mismatch between the

nature of the data and the estimation method used: covariance matrices are based on polychoric correlations, but ML treats variables as continuous, and the asymptotic covariance matrix is not included in the analysis. Model fit is more drastically affected when the distribution is skewed, because the ML assumption of multivariate normality is not met.

6. Conclusion

Based on the results presented above it can be concluded that the methods of parameter estimation in confirmatory factor analysis should be carefully chosen, by taking into account the extent to which the data sets utilized meet their inherent assumptions. ML is adequately used in a perfect case scenario, when the data are continuous and uniformly distributed. The accuracy of this estimation method is mostly affected when it is used with ordinal data that is not treated as continuous, especially when variables have a small number of categories. Additional estimation error occurs when the data does not meet the assumption of multivariate normality. In contrast, DWLS provides more accurate parameter estimates, and the fit of the model is more robust to variable type and non-normality.

Therefore, it is critical to investigate the distribution of the data before estimating parameters, and to address the variable type by choosing the most appropriate estimation method. When researchers do not check these assumptions and simply employ the default estimation procedure, results may be biased and may lead to erroneous decisions regarding the model tested.

7. Limitations

Although the results presented above support the effectiveness of the DWLS estimation method with ordinal and multivariate non-normal data, they are based on testing only one model. Additionally, all items in this model have the same loading value, and the same distribution. To better simulate real data, further research on this topic should test several models and include items with various loadings and different distributions.

Moreover, further investigation is needed to determine whether the findings can be replicated

with a large number of samples, as well as across different sample sizes.

8. References

- [1] E. Babakus, C. e. Ferguson & K. G. Jöreskog, "The Sensitivity of Confirmatory Maximum Likelihood Factor Analysis to Violations of Measurement Scale and Distributional Assumptions", *Journal of Marketing Research*, 1987, Vol. 24, 2228.
- [2] K. A. Bollen, *Structural Equation Modeling with Latent Variables*, New York: Wiley, 1989.
- [3] M. W. Browne, "Asymptotic Distribution Free Methods in the Analysis of Covariance Structures", *British Journal of Mathematical and Statistical Psychology*, 1984, Vol. 37, 62-83.
- [4] B. M. Byrne, *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*, Lawrence Erlbaum Associates, Mahwah, N.J. 1998.
- [5] S. J. Finney, C. A. DiStefano, *Nonnormal and Categorical Data in Structural Equation Modeling*, Information Age Publishing, Greenwich, Connecticut, 2006, pp. 269-314.
- [6] J. J. Hoogland, A. Boomsma, "Robustness Studies in in Covariance Structure Modeling: An Overview and Meta-Analysis, *Sociological Methods & Research*, 1998, Vol. 26, pp. 329-367.
- [7] K. G. Jöreskog, D. Sörbom, S. H. C. Du Toit, M. Du Toit, *LISREL 8*, Scientific Software International, Uppsala Universitet, 1996.
- [8] B. O. Muthen, D. Kaplan, "A Comparison of some Methodologies for the Factor Analysis of Non-Normal Likert Variables: A Note on the Size of the Model", *British Journal of Mathematical and Statistical Psychology*, 1985, Vol. 45, pp. 19-30.
- [9] R. E. Schumacker, , R. G. Lomax, *A beginner's Guide to Structural Equation Modeling.*, Lawrence Erlbaum Associates, Mahwah, N.J., 2004.
- [10] R. E. Schumacker, S. T. Beyerlein, "Teacher's Corner: Confirmatory Factor Analysis With Different Correlation Types and Estimation Methods", *Structural Equation Modeling: A Multidisciplinary Journal*, 2000, Vol. 7, 4 , pp. 629 – 636.