

Asymptotic Relative Efficiency Comparison for some Fit Indices in Structural Equation Modeling

İlkay Doğan^{1,a,*}, İsmet Doğan^{2,b}, Nurhan Doğan^{2,c}

¹ Department of Biostatistics, Medical Faculty, Gaziantep University, Gaziantep, Türkiye.

² Department of Biostatistics, Medical Faculty, Afyonkarahisar Health Science University, Afyonkarahisar, Türkiye.

*Corresponding author

Research Article

History

Received: 18/07/2023

Accepted: 12/03/2024



This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ABSTRACT

There are many fit statistics used in the structural equation modeling, and new ones are consistently being developed. Because of the variety of fit statistics, it is very important to be able to decide which fit statistics are appropriate to use in studies. When comparing any two statistics, the asymptotic relative efficiency (ARE) between them is used. The ARE can use as a power of the fit indices is one of the familiar optimal criteria. It is frequently more convenient, and also more suggestive, to use a measure of relative merit called the relative efficiency. This study aimed to compare of fit indices using Fraser's asymptotic relative efficiency. The data sets were derived from the multivariate normal distribution using the mean vector and covariance matrix. It was determined that the most efficient fit indices in terms of asymptotic relative efficiency were Z-Test of Wilson & Hilferty (W&H), Root Mean Square Error of Approximation (RMSEA), and Chi-Square indices, respectively.

Keywords: Asymptotic relative efficiency, Structural equation model, Fit indices.

^a ilkay_dgn58@hotmail.com

^b <https://orcid.org/0000-0001-7552-6478>

^c nurhandogan@hotmail.com

^b <https://orcid.org/0000-0001-7224-6091>

^b ismet.dogan@afsu.edu.tr

^b <https://orcid.org/0000-0001-9251-3564>

Introduction

Forecasting is a collaborative effort across a wide range of disciplines, and as a result, the question of how forecasts can best be evaluated is of fundamental importance to much of the scientific community [1]. A good forecast is very important for scientific, economic and administrative purposes. Therefore, it is necessary to know whether the forecaster is skilled enough to predict the future. Due to the increasing dependence on forecasting in various disciplines, forecasting skill indices have been proposed. It is very important to clarify that forecasting skill is not the same as forecasting accuracy. A highly skilled forecaster generally tends to have a high rate of forecast accuracy, but the opposite may not be true [2].

Statisticians, when dealing with specific test problems, often try to take or improve the statistical tests that are most efficient in a certain sense. Therefore, since such tests are not usually available in finite sample cases, the main focus is on efficiency concepts that enable comparison of competing procedures through their specific asymptotic properties [3]. Making a verified selection of the most effective statistical test among the few tests at the statistician's disposal is considered one of Statistics' fundamental problems. This problem became particularly important in the mid-twentieth century when computationally simple but "inefficient" rank tests emerged. Asymptotic relative efficiency (ARE) is a concept that enables the quantitative comparison of two different tests (for example T_1 and T_2) used to test the same statistical hypothesis to be applied in large samples. The

concept of asymptotic efficiency of tests is more complex than the asymptotic efficiency of estimates [4]. The ARE of T_2 relative to T_1 is defined as;

$$ARE_{21} = \lim_{n \rightarrow \infty} \left(\frac{n_1}{n_2} \right) \quad (1)$$

where n_1 and n_2 are sample sizes such that T_1 and T_2 have the same power. The oldest known efficiency is the Pitman efficiency [5]. The most familiar and the classical efficiencies are concepts in this respect are Pitman efficiency, Chernoff efficiency, Hodges-Lehmann efficiency and Bahadur efficiency [5-8]. The four basic types of efficiency mentioned are not easy to calculate. Each type of efficiency statistic has its own advantages and disadvantages. Current studies on ARE mainly focus on two categories. First category, consists of method that compare the efficiency of estimators of the same parameter. The other consists of method that compare test statistics of the same hypothesis. Subsequent studies may assume that the test statistics are asymptotically normal. In the circumstances, the ARE's can be easily calculated. There are also some methods to compare ARE of different test statistics, where the same hypothesis may have different asymptotic distributions. These methods (Pitman efficiency, Hodges-Lehmann efficiency and Bahadur efficiency) suggested different ways to calculate ARE, and are difficult to calculate. Although test statistics have the same asymptotic distribution, ARE can be easily

calculated [9]. A simple statement for the relative efficiency under moderate assumptions has been obtained by Fraser [10]. Under the asymptotic normality assumption of the test statistic the efficacy is given by [10-12].

$$eff = \left(\frac{\mu}{\sigma}\right)^4 \quad (2)$$

Here μ and σ are the mean and standard deviation of the limiting normal distribution. Definition of the ARE of two statistics, T_1 with respect to T_2 is as follows [10, 11];

$$ARE(T_1, T_2) = \frac{eff(T_1)}{eff(T_2)} \quad (3)$$

The value of $ARE(T_1, T_2) > 1$ indicates that T_2 is efficient, $ARE(T_1, T_2) < 1$ indicates that T_1 is efficient. In structural equation modeling (SEM), the fit indices establish whether, overall, the model is acceptable. The model fit evaluating issue in SEM analysis has been at the main topic of theoretical and empirical research for years [13-17]. The simplest model that describes the data well enough and makes the best prediction is the best fit model. Therefore, one of the important goals of scientific theory should be to establish criteria that will enable determining models that produce accurate predictions [18]. Research in this area has focused on various aspects of model fit indices, such as which model fit indices should be used in which situations and how they should be interpreted. Most studies have examined the changing of fit indices under different data conditions, such as estimation methods, sample size, and model misspecification [19].

In this article, we focus our study to the asymptotic relative efficiency via Fraser [10] of the fit indices. We propose an analytic comparison of the some fit indices. To this aim, we introduce Fraser [10] asymptotic relative efficiency, and compare the fit indices' asymptotic relative efficiencies, in the Fraser asymptotic relative efficiency sense.

Material and Methods

Simulation Study

In the study, a model occurred from four latent variables was used. Also each latent variable were explained by four variables. The mean vector and covariance matrix created for this model were obtained from the study of Doğan and Özdamar [15]. The data sets were derived under the assumption of multivariate normal distribution. While the sample sizes were determined as 100, 150, 250, 500, 1000 and 5000 units and the replication number was determined as 1000. The maximum likelihood (ML) was used for parameter estimates. This study was conducted to compare the model fit measures by using Fraser efficiency with different sample sizes. For this purpose, Chi-Square, Goodness of Fit Index (GFI), Root Mean Square Residuals

(RMR), Standardized Root Mean Square Residuals (SRMR), Z-Test of Wilson & Hilferty (W&H), Comparative Fit Index (CFI), Normed Fit Index (NFI), Non-normed Fit Index (NNFI), Incremental Fit Index (IFI), Adjusted Goodness of Fit Index (AGFI), Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Expected Cross Validation Index (ECVI), McDonald's Fit Index (MFI), Root Mean Square Error of Approximation (RMSEA), and Schwarz Bayesian Criterion (SBC) goodness of fit statistics were evaluated. The simulation study was performed in the SAS 9.3 program using the PROC IML and PROC WORK procedures.

There are two aspects to measuring model performance: discriminative capacity and reliability. However, it is generally stated that discrimination capacity is more important than reliability. Discriminative capacity measures a model's ability to distinguish between the presence and absence of the phenomenon in question. Reliability refers to the agreement between predicted values and observed values. Reliability is an important characteristic of the quality of probabilistic forecast models. When the modeling result is continuous, both aspects of model performance (discrimination capacity and reliability) can be evaluated. When the modeling result is binary, only discriminatory capacity can be evaluated. Various indices are used to assess discriminatory capacity and/or reliability. Some of these can only be applied to binary results or to continuous results converted to a binary solution using a specific cut-off value called a threshold. These indices are called threshold-dependent indices. Indices that can be directly applied to continuous cases are called threshold-independent indices [20]. The indices used in this study are non-threshold-dependent measures.

Data Generation

The steps of the simulation study were performed as suggested by Doğan and Özdamar [15] and Fan Xia and Fan Xia [21]. First of all, a 5000-unit data set was generated from the multivariate normal distribution by using the mean vector and covariance matrix of the specified model [15]. Secondly, the covariance matrix of the obtained data set in the first step was calculated to avoid the singular covariance matrix structure. Thirdly, another data set containing 5000 units is generated from the multivariate normal distribution with the help of the covariance matrix calculated in second step. Finally, parameter estimations of the model specified by Doğan and Özdamar [15] were performed with the help of the data set obtained in the third step. The fit statistics of the specified model and the correlation matrix (R) were examined. The correlation coefficients for the variables of each latent variable change $0.30 \leq r \leq 0.90$ intervals.

Results

The comparative summarized table of fit indices based on ARE is given in Table 1. As a result, it was determined that the most efficient fit indices in terms of asymptotic

relative efficiency were W&H, RMSEA, and Chi-Square indices, respectively. The results of the simulation study are same for all sample sizes (n=100, 150, 250, 500, 1000 and 5000). When all fit indices considered in the study are

sorted in terms of efficiency, W&H, RMSEA, Chi-Square, RMR, SRMR, AIC / MFI, ECVI, SBC, CAIC, ACFI, NNFI, IFI, NFI, CFI, and GFI are obtained.

Table 1. The comparisons of the fit indices

	GFI	RMR	SRMR	W&H	CFI	NFI	NNFI	IFI	AGFI	AIC	CAIC	ECVI	MFI	RMSEA	SBC	
Chi-Square	Chi-Square	Chi-Square	Chi-Square	W&H	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	RMSEA	Chi-Square	
GFI	-	RMR	SRMR	W&H	CFI	NFI	NNFI	IFI	AGFI	AIC	CAIC	ECVI	MFI	RMSEA	SBC	
RMR		-	RMR	W&H	RMR	RMR	RMR	RMR	RMR	RMR	RMR	RMR	RMR	RMSEA	RMR	
SRMR			-	W&H	SRMR	SRMR	SRMR	SRMR	SRMR	SRMR	SRMR	SRMR	SRMR	RMSEA	SRMR	
W&H				-	W&H	W&H	W&H	W&H	W&H	W&H	W&H	W&H	W&H	W&H	W&H	
CFI					-	NFI	NNFI	IFI	AGFI	AIC	CAIC	ECVI	MFI	RMSEA	SBC	
NFI						-	NNFI	IFI	AGFI	AIC	CAIC	ECVI	MFI	RMSEA	SBC	
NNFI							-	NNFI	AGFI	AIC	CAIC	ECVI	MFI	RMSEA	SBC	
IFI								-	AGFI	AIC	CAIC	ECVI	MFI	RMSEA	SBC	
AGFI									-	AIC	CAIC	ECVI	MFI	RMSEA	SBC	
AIC										-	AIC	AIC	AIC/MFI	RMSEA	AIC	
CAIC											-	ECVI	CAIC/MFI	RMSEA	SBC	
ECVI												-	ECVI/MFI	RMSEA	ECVI	
MFI														-	SBC/MFI	
RMSEA															-	RMSEA

Chi-Square, Goodness of Fit Index (GFI), Root Mean Square Residuals (RMR), Standardized Root Mean Square Residuals (SRMR), Z-Test of Wilson & Hilferty (W&H), Comparative Fit Index (CFI), Normed Fit Index (NFI), Non-normed Fit Index (NNFI), Incremental Fit Index (IFI), Adjusted Goodness of Fit Index (AGFI), Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Expected Cross Validation Index (ECVI), McDonald's Fit Index (MFI), Root Mean Square Error of Approximation (RMSEA) and Schwarz Bayesian Criterion (SBC)

Discussion and Conclusion

Structural equation modeling is a very powerful multivariate analysis technique that makes it possible to evaluate hidden structures, while asymptotic relative efficiency is an analysis and testing tool due to its unique properties compared to other techniques. After parameter estimation is made with the appropriate method for the specified model, the model needs to be tested. In testing the model, the compatibility of the data with the specified model is determined. In other words, an answer is sought to the question "to what extent is the theoretical (specified, proposed, and established) model compatible with the sample data" [22]. Compliance is called the ability of a model to reproduce the data, that is, the variance-covariance matrix [23]. There are many fit criteria to evaluate model fit in SEM. Most of these fit measures are based on comparing the covariance matrix of the theoretically proposed model with the sample covariance matrix. The fact that these two matrices are not similar to each other, that is, the difference between the matrices is large, indicates that the data does not fit

the theoretical model, and the fact that the difference between these two matrices is very small indicates that the data fits well with the theoretical model [22, 24, 25]. Evaluating model performance, that is, comparing the predictions produced by the model with observed values, is a fundamental step in model development and use. Once models are obtained, it is necessary to validate some aspects of them. This validation process usually involves a definition of criteria based on mathematical measurements that indicate how well the model's predictions are produced by simulating observed values [26]. The validity of the structural equation model is tested by calculating fit indices based on the collected data.

The results suggest that W&H is the best goodness of fit statistic for all specified assumptions. In the study, the multivariate normality assumption was taken into consideration and the fit indices were compared in terms of their asymptotic relative efficiency for the case where this assumption is valid. Although it is emphasized in the literature [15, 19, 25, 27] that the Chi-Square indices is affected by the sample size and should not be used, its effectiveness was better than the other indices after W&H and RMSEA fit indexes. In future studies, it is recommended to make comparisons for cases where this assumption is violated. Since the W&H fit indices is not calculated in some statistical packages, it is recommended to use the RMSEA fit indices in future studies. Doğan and Doğan [18] emphasize that model selection criteria should be considered as a model comparison or evaluation tool because the term selection includes the idea that something more certain is achieved. In addition, the criteria used in model selection should not be duplicated beyond necessity [18]. As a result, it is more important to decide which one gives better results rather than

duplicating model fit indices. This study shows that the RMSEA fit indices gives better results than others.

Conflicts of interest

There are no conflicts of interest in this work.

References

- [1] Wheatcroft E., Interpreting the skill score form of forecast performance metrics. *Int. J. Forecast.*, 35 (2) (2019) 573-579.
- [2] Sitthiyot T., Holasut K., On the evaluation of skill in binary forecast. *Thailand and The World Economy*, 40 (3) (2022) 33-54.
- [3] Baringhaus L., Gaigall D., On an asymptotic relative efficiency concept based on expected volumes of confidence regions, *Statistics*, 53 (6) (2019) 1396–1436.
- [4] Nikitin Y., Asymptotic Relative Efficiency in Testing, University of Saint-Petersburg, (2010).
- [5] Pitman E.J.G., Mimeographed Lecture Notes on Nonparametric Statistics, Columbia University, New York, (1948) 52-59.
- [6] Chernoff H.A., Measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Stat.*, 23 (1952) 493–507.
- [7] Hodges J.L., Lehmann E.L., The efficiency of some nonparametric competitors of the t –test, *Ann. Math. Stat.*, 27 (1956) 324–335.
- [8] Bahadur R.R., Stochastic comparison of tests. *Ann. Math. Stat.*, 31 (1960) 276–295.
- [9] Yuan A., Fan R., Xu J., Xue Y., Li Q., Asymptotic relative efficiencies of the score and robust tests in genetic association studies, *The Open Statistics & Probability*, 9 (2018) 26-41.
- [10] Fraser D.A.S., Nonparametric Methods in Statistics, John Wiley & Sons Inc, (1957) 270-276.
- [11] Jammalamadaka S.R., Wells M.T., A test of goodness of fit based on extreme spacings with some efficiency comparisons, *Metrika*, 35 (1988) 223-232.
- [12] Puri M.L., Rao J.S., Yoon Y., A simple test for goodness of fit based on spacings with some efficiency comparisons (Report No.441). University of Wisconsin, (1976) 11-13.
- [13] Anderson J.C., Gerbing D.W., The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis, *Psychometrika*, 49 (1984) 155–173.
- [14] Curran P.J., West, S.G., Finch J.F., The robustness of test statistics to non-normality and specification error in confirmatory factor analysis, *Psychol. Methods*, 1 (1) (1996) 16–29.
- [15] Doğan İ., Özdamar K., The effect of different data structures, sample sizes on model fit measures, *Commun. Stat-Simul. C.*, 46 (9) (2017) 7525-7533.
- [16] Doğan İ., Farklı veri yapısı ve örneklem büyüklüklerinde yapısal eşitlik modellerinin geçerliği ve güvenilirliğinin değerlendirilmesi, PhD Thesis, Eskişehir Osmangazi University, Institute of Health Sciences, (2015).
- [17] Fan X., Thompson B., Wang L., Effects of sample size, estimation methods and model specification on structural equation modeling fit indexes, *Struct. Equ. Modeling*, 6 (1) (1999) 56–83.
- [18] Doğan İ., Doğan, N., Model Performans Kriterlerinin Kronolojisine ve Metodolojik Yönlerine Genel Bir Bakış: Bir Gözden Geçirme, *Türkiye Klinikleri J. Biostat.*, 12 (1) (2020) 114-125.
- [19] Fan X., Sivo S.A., Sensitivity of fit indices to model misspecification and model types, *Multivar. Behav. Res.*, 42 (3) (2007) 509–529.
- [20] Liu C., White M., Newell G., Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography*, 34 (2) (2011) 232-243.
- [21] Fan X., Fan X., TEACHER'S CORNER: Using SAS for Monte Carlo Simulation Research in SEM, *Struct. Equ. Modeling*, 12 (2) (2005) 299-333.
- [22] Schumacker R.E., Lomax R.G., A beginner's guide to structural equation modeling (2nd ed.). Mahlah, New Jersey,London: Lawrence Erlbaum Associates. (2004) 79-84.
- [23] Erkorkmaz Ü., Etikan İ., Demir O., Özdamar K., Sanisoğlu S.Y., Doğrulayıcı faktör analizi ve uyum indeksleri, *Türk. Klin. J. Med. Sci.*, 33(1) (2013) 210-223.
- [24] Bollen, K.A. Structural equations with latent variables, New York: John Wiley & Sons, Inc., (1989) 256-289.
- [25] Kline B.R., Principles and practice of structural modeling. New York-London: The Guilford Press, (2011) 189-230.
- [26] Pereira H.R., Meschiatti M.C., Pires R.C.M., Blain G.C., On the performance of three indices of agreement: an easy-to-use r-code for calculating the Willmott indices, *Bragantia*, 77(2) (2018) 394-403.
- [27] Weng L.J., Cheng C.P., Why might relative fit indices differ between estimators?, *Struct. Equ. Modeling*, 4 (2) (1997) 121-128.