STATISTICA, anno LXII, n. 2, 2002

FIT ASSESSMENT AND SELECTION BETWEEN COMPETITIVE MODELS IN SEM

Roberto Di Natale

1. INTRODUCTION

Structural equation models (SEM) with latent variables (Bollen, 1989a; Maruyama, 1998) are very often employed by researchers that use nonexperimental and quasi-experimental data. They are a synthesis of procedures developed in econometrics, sociometrics, and psychometrics. The SEM allows the researcher to study causal relationships among latent and observed variables. One or more latent unobserved variables are linked to one or more observed variables to show how the latent variables are measured (the measurement model). The theory suggests the links between latent variables. The relationships between variables have the form of some linear equations and they are simultaneously estimated by the procedures. The interpretation of relations among the variables is the same as in path analysis (Wright, 1934; Duncan, 1966), in which the direct and indirect relationships between variables involved in a causal model are estimated. An important quality of SEM is the generality of the approach. Well-known statistical models like the multiple regression, the path analysis, the measurement models, the confirmatory factor analysis, the "classical" econometric models and others are all contained in the general framework of SEM as special cases. Some assumptions of the "classic" SEM can be relaxed and several extensions are available (Bollen, 1989a, cap. 9). Theory and methodology to study three-level data (Yau *et al.*, 1993), ARMA time series (van Buuren, 1997), the state space model (Oud *et al.*, 1990) have been proposed.

Covariance structure models, path models, dynamic simultaneous equations models, reticular action models, latent variable models and LISREL models, refer to the same class of models. The SEM are implemented in most commercial computer packages (*e.g.* Amos, EQS, LISREL, SAS PROC-CALIS, LISCOMP) and they have become easy to use because they can be run by creating a path diagram of the model.

The first object of this article is to introduce some problems in model fit assessment and in the selection between competitive models in SEM. The methodological consequences are also considered. The second object is to examine the behaviour of the fit indices in the framework of selection between alternative models. To our knowledge, only the recent paper of La Du and Tanaka (1995) extensively examines the problem of misspecification in a Monte Carlo study and their results are not encouraging. In this paper we considered a simple problem in a ideal situation of sampling distributions and sample size. We also include the parameter estimates in the rationale of model selection.

The assessment of fit in a model is a typical point of controversy in statistical analysis. The R^2 in regression analysis is an example of this issue in a older method of analysis than the Structural Equation. In this paper we relied on the point of consensus in testing models resulting from Bollen's (1993a) book. The selection among alternative (or competitive) models theoretically supported by a substantive theory is a usual problem in applied research.

2. THE GENERAL MODEL

The SEM, in LISREL formulation, is defined by the following three equations:

$$\underline{\boldsymbol{\eta}} = \underline{\mathbf{B}}\,\underline{\boldsymbol{\eta}} + \underline{\boldsymbol{\Gamma}}\,\underline{\boldsymbol{\xi}} + \underline{\boldsymbol{\zeta}} \tag{1}$$

$$\underline{\mathbf{y}} = \underline{\boldsymbol{\Lambda}}_y\,\underline{\boldsymbol{\eta}} + \underline{\boldsymbol{\varepsilon}} \tag{2}$$

$$\underline{\mathbf{x}} = \underline{\boldsymbol{\Lambda}}_x\,\underline{\boldsymbol{\xi}} + \underline{\boldsymbol{\delta}} \tag{3}$$

where $\underline{\boldsymbol{\xi}}$ is an n x 1 random vector of latent independent (exogenous) variables, $\underline{\boldsymbol{\eta}}$ is an m x 1 random vector of latent dependent (endogenous) variables, $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$ are vectors of q x 1 and p x 1 observed variables respectively. $\underline{\boldsymbol{\delta}}$ and $\underline{\boldsymbol{\varepsilon}}$ are q x 1 and p x 1 sets of errors of measurement in $\underline{\mathbf{x}}$ and $\underline{\mathbf{y}}$, respectively and $\underline{\boldsymbol{\zeta}}$ is a m x 1 vector of equation errors in the relationships among $\underline{\boldsymbol{\eta}}$ and $\underline{\boldsymbol{\xi}}$. Structural relationship between the latent variables are specified by the $\underline{\mathbf{B}}$ (m x m) and the $\underline{\boldsymbol{\Gamma}}$ (m x n) matrices of coefficients. $\underline{\boldsymbol{\Lambda}}_y$ and $\underline{\boldsymbol{\Lambda}}_x$ are p x m and q x n regression matrices of $\underline{\mathbf{y}}$ on $\underline{\boldsymbol{\eta}}$ and of $\underline{\mathbf{x}}$ on $\underline{\boldsymbol{\xi}}$. We assume that all the variables are deviation of the mean and that $n \leq q$; $m \leq p$. The errors $\underline{\boldsymbol{\zeta}}$, $\underline{\boldsymbol{\varepsilon}}$, and $\underline{\boldsymbol{\delta}}$ are mutually uncorrelated and the errors of measurement are assumed to be uncorrelated with the latent variables measured. The matrix ($\underline{\mathbf{I}}$ - $\underline{\mathbf{B}}$) is assumed non singular and the following covariance matrices are specified: $\underline{\boldsymbol{\Phi}}= E(\underline{\boldsymbol{\xi}}\,\underline{\boldsymbol{\xi}}')$, $\underline{\boldsymbol{\Psi}}=E(\underline{\boldsymbol{\zeta}}\,\underline{\boldsymbol{\zeta}}')$, $\underline{\boldsymbol{\Theta}}^{\varepsilon} =E(\underline{\boldsymbol{\varepsilon}}\,\underline{\boldsymbol{\varepsilon}}')$, $\underline{\boldsymbol{\Theta}}^{\delta} =E(\underline{\boldsymbol{\delta}}\,\underline{\boldsymbol{\delta}}')$.

Equations (2) and (3) show how the latent variables $\underline{\boldsymbol{\eta}}$ and $\underline{\boldsymbol{\xi}}$ are measured by the observed variables $\underline{\mathbf{y}}$ and $\underline{\mathbf{x}}$ respectively. The equation (1) has the structure of the classical econometric model but $\underline{\boldsymbol{\eta}}$ and $\underline{\boldsymbol{\xi}}$ are allowed the non-observable variables.

The basic fundamental relation of the SEM is

$$\underline{\boldsymbol{\Sigma}} = \underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}}) \tag{4}$$

where $\underline{\boldsymbol{\Sigma}}$ is the population covariance matrix of $\underline{\mathbf{y}}$ and $\underline{\mathbf{x}}$; $\underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$ is the implied covariance matrix written as function of the t free model parameters in $\underline{\boldsymbol{\theta}} = (\underline{\boldsymbol{\theta}}_1, \underline{\boldsymbol{\theta}}_2,$

$\ldots,\underline{\boldsymbol{\theta}}_t$). The parameter vector $\underline{\boldsymbol{\theta}}$ varies in the parameter space $\underline{\boldsymbol{\Theta}} \subset R^t$. The implied covariance matrix $\underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$ is derived from the specified model (see Bollen, 1989a, 323-326). Identification, estimation and assessments of model fit is based on this relation of $\underline{\boldsymbol{\Sigma}}$ to $\underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$. If the structural equation model is correctly specified and the population parameters are known, then the equation (4) is hold exactly.

Let $\underline{S}$ be the sample unbiased estimate of the population covariance matrix and let n be the sample size. We need to form sample estimates $\hat{\underline{\theta}}$ of the unknown parameters $\underline{\boldsymbol{\theta}}$ based on sample estimate $\underline{S}$ of the covariance matrix $\underline{\boldsymbol{\Sigma}}$. The estimation procedure typically minimise a fitting function F($\underline{S}$, $\underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$) with the following properties: F($\underline{S}$, $\underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$) is a scalar; F($\underline{S}$, $\underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$) ≥ 0; F($\underline{S}$, $\underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$) = 0 if and only if $\underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$ = $\underline{S}$ and F($\underline{S}$, $\underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$) is continuous in $\underline{S}$ and $\underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$. Minimising fitting functions with respect to the free parameters in $\underline{\boldsymbol{\theta}}$ that satisfy these conditions leads to consistent estimators of $\underline{\boldsymbol{\theta}}$ (Browne, 1984). If the model is correct and if we know the parameters, the population covariance matrix would be exactly reproduced. To date the most widely used fitting functions are: the normal theory maximum likelihood F_{ML} function (Jöreskog, 1969), the normal theory Generalised Least Squares F_{GLS} function (Jöreskog and Goldberger, 1972), and the Asymptotic Distribution Free F_{ADF} function which does not depend on distributional assumptions (Browne, 1984).

3. THE χ^2 TEST STATISTIC

Let be

$$T_{ML} = N \cdot F_{ML}(\underline{S}, \underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})) \tag{5}$$

$$T_{GLS} = N \cdot F_{GLS}(\underline{S}, \underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})) \tag{6}$$

$$T_{ADF} = N \cdot F_{ADF}(\underline{S}, \underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})) \tag{7}$$

where $N = n - 1$ and

$$H_o: \underline{\boldsymbol{\Sigma}} = \underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}}) \tag{8}$$

is the null hypothesis, that is the model hold. Theoretical results show that given the null hypothesis is true, the function T_{ML} (Jöreskog, 1969), the function T_{GLS} (Jöreskog, 1972), and the function T_{ADF} (Browne, 1984) are asymptotically chi-squared distributed with $df = (1/2) \cdot (p + q) \cdot (p + q - 1) - t$ degrees of freedom.

The asymptotic robustness of normal-theory methods has been extensively studied and several different conditions were found under which models with non-normally distributed variables can still be studied by use of methods based on normal theory (Satorra, 1990). Therefore, in theory we are well equipped to test the null hypothesis (8). In practice, a series of simulation studies (Muthen and Kaplan, 1985, 1992; Hu *et al.* 1992, Rigdon and Ferguson, 1991; Chou *et al.*,

1991) show that these fitting functions performed poorly and the test cannot be trusted. Normal-theory T statistics are not robust to violation of asymptotic robustness assumptions (Hu *et al.*, 1992). Nonetheless, the Likelihood Ratio (LR) test based on the T_{ML} function remain by far the most widely used methodologies in practice (Yuan and Bentler, 1997). The T_{ADF} function is asymptotically χ^2 distributed only for very large samples. Basically, sample size requirements increase as models become larger and data become more nonnormal (Hu *et al.*, 1992; Muthen and Kaplan, 1985, 1992). A bootstrap correction of additive bias on the ADF test statistic has been proposed (Young and Bentler, 1994), it yields a better tail behaviour as the sample size reaches 500 for a 15-variable 3-factor confirmatory factor-analytic model.

Several corrected T statistics (based on T_{ML} or T_{GLS} or T_{ADF}) have been proposed (Yuan and Bentler, 1997) but little is know on the performance of these new statistics.

Briefly, the χ^2 test statistic cannot be trusted in most of practical situations, since the variables are not normally distributed and the ADF theory request very large samples that are rarely available in applied research. It does not provide information regarding the degree of fit because high fit can be obtained simply with a diminution of the degrees of freedom. In the extreme case in which $df = 0$ then $\underline{\mathbf{S}} = \underline{\boldsymbol{\Sigma}}(\hat{\underline{\boldsymbol{\theta}}})$ and $F(\underline{\mathbf{S}}, \underline{\boldsymbol{\Sigma}}(\hat{\underline{\boldsymbol{\theta}}})) = 0$. The T statistic indicate a perfect fit and it does not depend on the proposed model. Furthermore, if the fitting function $F(\underline{\mathbf{S}}, \underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})) = k$, $k > 0$ (F_{ML} or F_{GLS} or F_{ADF}) is constant, as the sample size increases, each model will be rejected by the asymptotic chi-square test statistic T at any fixed level of significance. This sample size issue is linked in part to the statistical power available to test the null hypothesis. In applied research, the H_o: $\underline{\boldsymbol{\Sigma}} = \underline{\boldsymbol{\Sigma}}(\underline{\boldsymbol{\theta}})$ does not hold exactly, thus the chi-square should be compared with a non-central rather than a central chi-square distribution (see Browne, 1984). To overcome these problems in fit assessment by χ^2 statistic, many fit indices has been proposed.

4. FIT INDICES

We now examine the problem of the overall fit assessment and the problem of selection between competitive models using some important indices proposed in literature. To date, many fit indices has been proposed but none of which has been endorsed as the "best index" by the majority of researchers. As an illustration of the proliferation of indices, the computer programs LISREL8 (Jöreskog and Sörbom, 1996b) and PROC CALIS (SAS Institute, 1989) prints the values of almost 20 such fit indices. The problem of testing model fit is the subject of considerable discussion and Bollen's (1993a) book concentrate on this issue, the first chapter reports the major points of consensus. Following these points of consensus, we prefer fit indices that take account of the degrees of freedom (parsimony) of a model and indices whose means of their sampling distribution are not or are only weakly related to the sample size.

Following McDonald and Marsh (1990) the indices examined could be considered of two sorts: absolute measures and relative measures. Absolute measures are typically made up on discrepancy between sample covariance matrix $\underline{\mathbf{S}}$ and the estimated population matrix $\underline{\mathbf{\Sigma}}(\underline{\boldsymbol{\theta}})$. The absolute measures to be considered in this paper are "badness" of fit measures, in the sense that small values correspond to good fit and large values correspond to bad fit. In addition, the relative fit indices require a third matrix that is used as a reference point in assessing fit and they range from 0 to 1 (indicating the best fit). The model used as a reference point is often the null model. Usually it is a more restricted nested model in which all observed variables are uncorrelated. The null model is a statistical baseline of comparison for the evaluation of fit.

The already examined χ^2 statistic is considered between absolute measures of fit, the use of χ^2 statistic as test of the model is rarely appropriate in real data applications. The LR test is also considered for the selection between competitive models.

We consider a set of g competing models, models $M_1, M_2, \ldots, M_g$. Each involves a s x s matrix valued function $\underline{\mathbf{\Sigma}}_k(\underline{\boldsymbol{\theta}}_k)$ of a parameter vector $\underline{\boldsymbol{\theta}}_k$ with t_k elements, k = 1,2, ..., g, where $s = p + q$ is the number of observed variables. We think of the models M_k as the k-th in a series of increasing complexity. A model $\underline{\mathbf{\Sigma}}_l(\underline{\boldsymbol{\theta}}_l)$ is a submodel of $\underline{\mathbf{\Sigma}}_k$, nested with it, if $\underline{\mathbf{\Sigma}}_l$ can be obtained by placing further restrictions on $\underline{\boldsymbol{\theta}}_k$ in $\underline{\mathbf{\Sigma}}_k$ (usually by setting some of the parameters in $\underline{\boldsymbol{\theta}}_k$ to zero). In brief, the nested model is a special case of the less restrictive model.

The LR test is not the golden rule in selection between nested competitive models because empirical results show its inadequacy for nonnormal data (Hu *et al.* 1992). The LR test has the following form:

$$LR = -2 \cdot [\log L(\hat{\underline{\theta}}_l) - \log L(\hat{\underline{\theta}}_k)]$$

where $\hat{\underline{\theta}}_l$ and $\hat{\underline{\theta}}_k$ are the ML estimator for the restrictive, nested model and the less restricted model respectively. LR has a limiting χ^2 distribution with degrees of freedom $df = df_l - df_k$ when the restrictive model is valid and the assumptions underlying the ML estimators are valid. The LR test may be written as

$$LR = T_{ML}^{\ l} - T_{ML}^{\ k}$$

where $T_{ML}^{\ l}$ and $T_{ML}^{\ k}$ are the chi-square statistics in equation (5) for the nested model and the model without restriction respectively.

Three absolute fit indices are examined in this study: the chi-square statistic, the Akaike information criteria (AIC: Akaike, 1987;) and the single sample cross-validation index ECVI (Browne and Cudeck, 1989).

The chi-square statistic T is a fundamental value in the derivation of different fit indices.

The RMSEA_k index (Steigher, 1990) for the k-th model is:

$$\mathrm{RMSEA}_k = [\max((F_k / df_k - 1/N), 0)]^{1/2}$$

where df_k and F_k are the minimum value of the fit function and the degrees of freedom for the k-th model respectively. RMSEA is a measure of the discrepancy per degree of freedom for the model.

The $ECVI_k$ index (Browne and Cudeck, 1989) for the *k*-th model is:

$$ECVI_k = (T_k / N) + 2 \cdot (t_k / N)$$

where $N = (n - 1)$. The $ECVI_k$ is intended to provide guidance for the selection of a model appropriate for a specified sample size. Underlying the use of this index is the assumption that none of the models under consideration (excluding the saturated model) is correct (Browne and Cudeck, 1989).

The ECVI and the AIC (Akaike, 1987) are similar, but AIC is derived from statistical information theory whereas ECVI is a measure of the discrepancy between the fitted covariance matrix in the analysed sample and the expected covariance matrix that would be obtained in another sample of the same size.

The following three relative fit indices are examined: the Comparative Fit Index (CFI; Bentler, 1990), the Incremental Fit Index (IFI; Bollen, 1989b), the Adjusted Goodness of Fit Index (AGFI; Jöreskog and Sörbom, 1989). The Relative Noncentrality Index (RNI: McDonald and Marsh, 1990) was excluded from this study because CFI and RNI are algebraically equivalent in most applications (Goffin, 1993).

Let F_i and F_k be the minimum value of the fit function for the null model and the k-th model respectively. Let df_i and df_k be the corresponding degrees of freedom.

The IFI_k (Bollen, 1989b) for the *k*-th model is

$$IFI_k = (N \cdot F_i - N \cdot F_k) / (N \cdot F_i - df_k)$$

The CFI_k (Bentler, 1990), has the following form:

$$CFI_k = 1 - \tau_k / \tau_i$$

where $\tau_k = \max(N \cdot F_k - df_k, 0)$, and $\tau_i = \max(N \cdot F_i - df_i, N \cdot F_k - df_k, 0)$.

The $AGFI_k$ (Jöreskog and Sörbom, 1989) index is

$$AGFI_k = 1 - [s \cdot (s + 1) / (2 \cdot df_k)] \cdot (1 - GFI_k)$$

where s is the number of observed variables and the Goodness of Fit Index GFI_k is

$$GFI_k = 1 - [F_k(\underline{\mathbf{S}}, \underline{\mathbf{\Sigma}}(\hat{\underline{\theta}})) / F(\underline{\mathbf{S}}, \underline{\mathbf{\Sigma}}(\underline{\mathbf{0}}))]$$

The numerator is the minimum of the fit function after the model has been fitted; the denominator is the fit function when all free parameters are set to zero. The GFI assess the discrepancy between the predicted covariance $\underline{\mathbf{\Sigma}}(\hat{\underline{\theta}})$ and the observed covariance $\underline{\mathbf{S}}$. The AGFI adjusts the GFI for the degrees of freedom of a model relative to the number of variables.

The LR test can not be used out of a nested sequence of models. The IFI and CFI are based on the null model logic (Bentler and Bonnet, 1980), they are appropriate for nested sequence of models (see La Du and Tanaka, 1995).

5. THE PROBLEM OF MODEL SELECTION

In this work we are specifically interested in the selection of a model between competitive models on the basis of the analysis of a single set of data. This is a well know problem in applied research (Di Natale and Saba, 1997; Saba and Di Natale, 1998).

We agree with the points of consensus identified in Bollen's (1993a) book on testing model fit. These points and the recent results of simulation studies (Muthen and Kaplan, 1985, 1992; Hu *et al.* 1992, Rigdon and Ferguson, 1991; Chou *et al.*, 1991; La Du and Tanaka, 1995) are used in subsequent discussion.

It is assumed that each of the identified k alternative (or competitive) models are theoretically supported by a substantive theory. The models give reasonable results of signs and magnitudes of coefficient estimates and R^2 of equations. If these assumptions did not occur a model can be eliminated from further consideration. Even under these conditions the problem of model selection would not be reduced to the selection of the model with the better fit. The strategy of the addition of parameters primarily to improve model fit may be inappropriate. Using this strategy we probably will select a very good model for our sample but this does not mean that this model has a reasonable correspondence to reality.

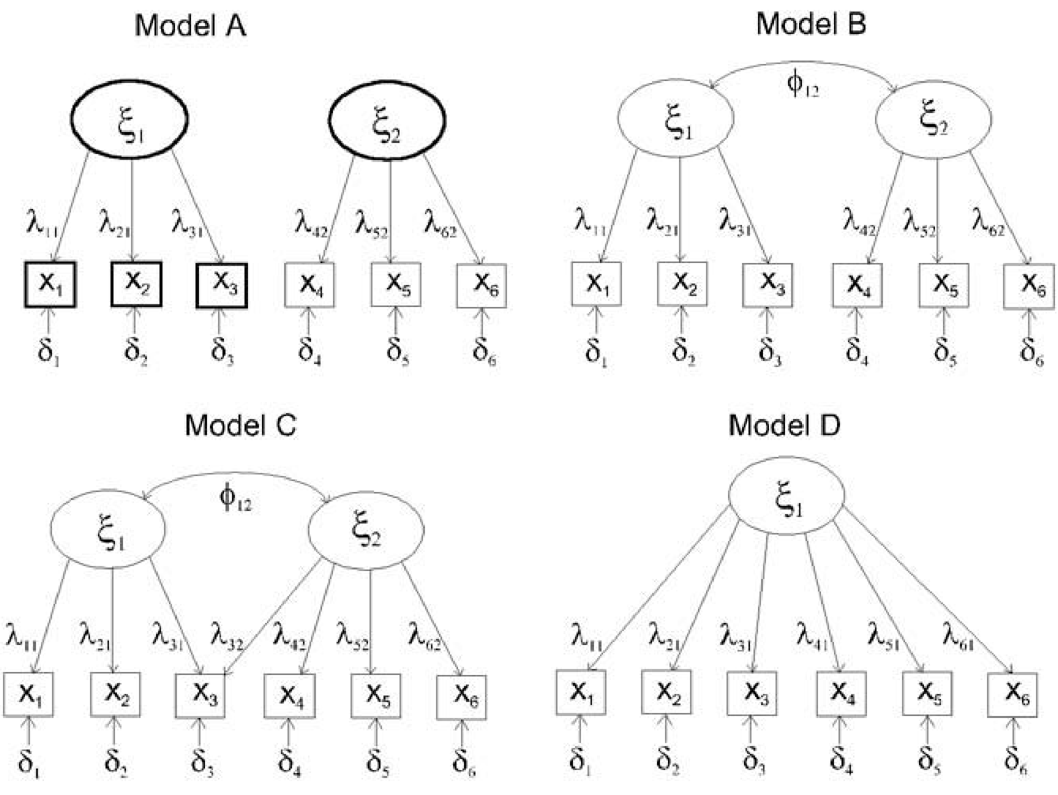

Figure 1 – Path diagram of the four models.

In model selection an underlying principle of parsimony is often used (see Bollen, 1989a, pag. 71). Briefly, the parsimony rule suggests to select the simplest model between proposed competing models with nearly the same fit to data.

Three simulation study was used to highlight the problems in model selection. We evaluated the ability in the above indices to distinguish the correct model. A 3 x 4 factorial Monte Carlo approach was used in confirmatory factor analysis models. The treatment variables were sample size and model specification. Four different models (fig.1) were analysed in each of the three studies. A study differ from the others for the know population model. Model A, B, C differ for the setting of one or two parameters to zero, they form a sequence of nested models. They are only slightly underparametrized and overparametrized compared with themselves. We considered it a slightly misspecification. Model D is really different from the others and it introduce a misspecification. The dependent variables in the study were parameter estimates and fit statistics. For each know population model, 100 replications are performed drawing sample size of n=100, 200, 500. The classical ML estimator was used under multivariate normality. Under this assumption, ML and GLS have the same asymptotic properties (Browne, 1974). For the purposes of this paper only results for fit statistics are showed. The results of simulation on parameter estimates are examined in subsequent discussion when the results for fit indices are unsatisfactory. However, in agreement with other results (Anderson and Gerbing, 1984; Rigdon and Ferguson, 1991), the mean of AGFI was affected by sample size for the four model under consideration.

The Schrage (1979) uniform random numbers generator was used to produce the independent normal observations. The misspecified model D (fig.1) is not nested with the other models, thus the LR test can not be used, the IFI and the CFI indices are not appropriate. The analyses were carried out using the statistical packages PRELIS2 and LISREL8.

5.1 *Results*

In the study 1, A is the know population model. The mean, the standard deviation and the extreme values (only for n=200) of the fit indices obtained in the simulation study 1 for n=100, 200, 500 are listed in table 1. The LR statistic chi-square estimates are not statistically significant for models B and C in every one of the situations considered. These additional parameters did not significantly improve model fit and we select the right model A. In this study only ECVI select always the real model. The others fit indices select a slightly misspecified model or are neutral in the problem of model selection. The variability of the RMSEA is considerably high and in many cases, it decrease to zero that is the best value for wrong models. All the indices excluded the model misspecified D. For models A, B and C the changes in the indices are small and it is not difficult to fall into error in model selection. In this case, the values of parameter estimates can help us in the selection of the correct model. For model B (n=200), the mean of parameter estimates of ϕ_{12} is 0.003, this value is not significantly different from zero. This result suggests that we should set the parameter ϕ_{12} of model B to zero. The parameters estimate are ϕ_{12} = -0.006 and λ_{32} = 0.001 for model C, these improperly added parameters are not significantly different from zero. The parameters estimate suggests that we should set ϕ_{12} and λ_{32} to zero. Therefore, we are led to the correct model by the parameters estimate of the models B and C.

TABLE 1

Study 1. Descriptive statistics of fit indices (A=correct model)

	n=100		*n*=200				*n*=500	
chi2	Mean	Std.dev.	Mean	Std.dev.	Min.	Max.	Mean	Std.dev.
Mod.A (9df)	9,181	4,121	8,461	3,460	1,836	20,169	7,889	3,576
Mod.B (8df)	8,188	3,846	7,368	3,364	1,327	20,042	6,878	3,245
Mod.C (7df)	6,795	3,488	6,334	2,872	1,000	17,386	5,985	3,128
Mod.D (9df)	137,46	20,82	268,86	31,96	182,45	347,50	667,03	56,21
RMSEA								
Mod.A (9df)	0,026	0,034	0,014	0,021	0,000	0,079	0,008	0,012
Mod.B (8df)	0,028	0,035	0,014	0,021	0,000	0,087	0,008	0,013
Mod.C (7df)	0,026	0,034	0,013	0,021	0,000	0,086	0,008	0,014
Mod.D (9df)	0,378	0,031	0,380	0,023	0,311	0,435	0,382	0,016
ECVI								
Mod.A (9df)	0,335	0,042	0,163	0,017	0,130	0,222	0,064	0,007
Mod.B (8df)	0,345	0,039	0,168	0,017	0,137	0,231	0,066	0,007
Mod.C (7df)	0,351	0,035	0,173	0,014	0,146	0,228	0,068	0,006
Mod.D (9df)	1,631	0,210	1,472	0,161	1,038	1,867	1,385	0,113
AGFI								
Mod.A (9df)	0,933	0,028	0,968	0,013	0,926	0,993	0,988	0,005
Mod.B (8df)	0,932	0,030	0,969	0,014	0,917	0,994	0,988	0,006
Mod.C (7df)	0,935	0,032	0,969	0,014	0,917	0,995	0,988	0,006
Mod.D (9df)	0,271	0,064	0,283	0,048	0,178	0,429	0,287	0,033
CFI								
Mod.A (9df)	0,994	0,011	0,998	0,004	0,979	1,000	0,999	0,001
Mod.B (8df)	0,994	0,010	0,998	0,004	0,978	1,000	0,999	0,001
Mod.C (7df)	0,995	0,009	0,998	0,003	0,981	1,000	0,999	0,001
Mod.D (9df)	0,504	0,059	0,501	0,042	0,390	0,601	0,504	0,030
IFI								
Mod.A (9df)	1,000	0,015	1,001	0,007	0,979	1,015	1,001	0,003
Mod.B (8df)	1,000	0,014	1,001	0,006	0,978	1,014	1,001	0,002
Mod.C (7df)	1,001	0,013	1,001	0,005	0,981	1,013	1,001	0,002
Mod.D (9df)	0,516	0,057	0,507	0,041	0,398	0,606	0,506	0,030

In the study 2, D is the know population model. The mean, the standard deviation and the extreme values (only for *n*=200) of the fit indices obtained in the simulation study 2 for *n*=100, 200, 500 are listed in table 2. The ECVI select the correct model D for all the sample sizes. The RMSEA select the correct model only for *n*=500 and it is equal to zero for wrong models. The other fit indices are often neutral in the selection between the correct model D and the misspecified models B and C. Only the model A is always excluded by the indices. The parameter estimate of $\phi_{12} \cong 1$ for models B and C, therefore it suggest the real model D.

TABLE 2

Study 2. Descriptive statistics of fit indices (D=correct model)

	n=100		*n*=200				*n*=500	
chi2	Mean	Std.dev.	Mean	Std.dev.	Min.	Max.	Mean	Std.dev.
Mod.D (9df)	10,232	4,693	10,191	4,528	1,754	32,382	8,403	4,438
Mod.A (9df)	140,81	17,68	273,41	23,89	226,09	347,54	664,32	32,37
Mod.B (8df)	9,002	4,383	9,033	4,581	1,279	31,155	7,527	4,026
Mod.C (7df) *	7,46	3,67	7,56	3,40	1,27	17,16	6,60	3,31
RMSEA								
Mod.D (9df)	0,036	0,039	0,025	0,026	0,000	0,114	0,009	0,015
Mod.A (9df)	0,384	0,026	0,384	0,017	0,348	0,435	0,382	0,009
Mod.B (8df)	0,034	0,040	0,025	0,028	0,000	0,121	0,010	0,016
Mod.C (7df) *	0,030	0,038	0,023	0,025	0,000	0,085	0,010	0,015
ECVI								
Mod.D (9df)	0,346	0,047	0,172	0,023	0,129	0,283	0,065	0,009
Mod.A (9df)	1,665	0,179	1,495	0,120	1,257	1,867	1,379	0,065
Mod.B (8df)	0,354	0,044	0,176	0,023	0,137	0,287	0,067	0,008
Mod.C (7df) *	0,358	0,037	0,179	0,017	0,147	0,227	0,069	0,007
AGFI								
Mod.D (9df)	0,925	0,033	0,962	0,017	0,877	0,993	0,987	0,007
Mod.A (9df)	0,516	0,026	0,528	0,016	0,483	0,565	0,538	0,009
Mod.B (8df)	0,925	0,035	0,962	0,019	0,864	0,994	0,987	0,007
Mod.C (7df) *	0,928	0,034	0,963	0,016	0,918	0,994	0,987	0,007
CFI								
Mod.D (9df)	0,994	0,009	0,997	0,004	0,973	1,000	0,999	0,002
Mod.A (9df)	0,667	0,032	0,666	0,022	0,619	0,721	0,667	0,013
Mod.B (8df)	0,995	0,008	0,997	0,004	0,973	1,000	0,999	0,001
Mod.C (7df) *	0,996	0,007	0,998	0,003	0,987	1,007	0,999	0,001
IFI								
Mod.D (9df)	0,997	0,012	0,999	0,006	0,973	1,010	1,000	0,002
Mod.A (9df)	0,672	0,032	0,668	0,022	0,622	0,723	0,668	0,013
Mod.B (8df)	0,998	0,011	0,999	0,006	0,973	1,009	1,000	0,002
Mod.C (7df) *	0,999	0,009	0,999	0,004	0,987	1,007	1,000	0,002

* Values based on the convergement estimates only.

In the study 3, B is the know population model. The mean, the standard deviation and the extreme values (only for *n*=200) of the fit indices obtained in the simulation study 3 for *n*=100, 200, 500 are listed in table 3. Also in this study, ECVI select the correct model for all the sample sizes. The RMSEA is equal to zero for wrong models. All the indices exclude models A and D. The relative measures and RMSEA are inadequate to the choice between model B and C. In this case the parameters estimate of $\lambda_{32} \cong 0$ for model C suggest the correct model B.

TABLE 3

Study 3. Descriptive statistics of fit indices (B=correct model)

	n=100		*n*=200				*n*=500	
chi2	Mean	Std.dev.	Mean	Std.dev.	Min.	Max.	Mean	Max.
Mod.B (8df)	8,068	3,757	7,313	3,547	1,422	17,067	7,109	18,269
Mod.A (9df)	41,99	10,56	75,24	14,72	38,17	112,32	173,12	238,30
Mod.C (7df)	6,828	3,545	6,331	3,356	1,338	15,306	6,229	18,208
Mod.D (9df)	93,95	20,44	181,92	30,11	104,80	258,92	462,65	587,53
RMSEA								
Mod.B (8df)	0,026	0,035	0,015	0,022	0,000	0,075	0,009	0,051
Mod.A (9df)	0,190	0,031	0,191	0,022	0,128	0,240	0,191	0,226
Mod.C (7df)	0,026	0,036	0,015	0,024	0,000	0,077	0,009	0,057
Mod.D (9df)	0,307	0,037	0,310	0,027	0,231	0,374	0,317	0,359
ECVI								
Mod.B (8df)	0,344	0,038	0,167	0,018	0,138	0,216	0,066	0,089
Mod.A (9df)	0,667	0,107	0,499	0,074	0,312	0,685	0,395	0,526
Mod.C (7df)	0,349	0,050	0,173	0,017	0,147	0,218	0,069	0,093
Mod.D (9df)	1,191	0,206	1,035	0,151	0,647	1,422	0,975	1,225
AGFI								
Mod.B (8df)	0,932	0,031	0,969	0,015	0,931	0,994	0,988	0,998
Mod.A (9df)	0,755	0,045	0,777	0,033	0,704	0,871	0,792	0,844
Mod.C (7df)	0,924	0,099	0,969	0,016	0,928	0,993	0,988	0,998
Mod.D (9df)	0,364	0,088	0,357	0,069	0,135	0,565	0,338	0,480
CFI								
Mod.B (8df)	0,996	0,007	0,998	0,003	0,987	1,000	0,999	1,000
Mod.A (9df)	0,905	0,026	0,905	0,019	0,857	0,956	0,907	0,935
Mod.C (7df)	0,986	0,100	0,998	0,003	0,988	1,000	0,999	1,000
Mod.D (9df)	0,753	0,056	0,751	0,041	0,650	0,852	0,742	0,824
IFI								
Mod.B (8df)	1,000	0,011	1,001	0,005	0,987	1,011	1,001	1,004
Mod.A (9df)	0,907	0,026	0,906	0,018	0,859	0,956	0,907	0,935
Mod.C (7df)	0,990	0,101	1,001	0,005	0,989	1,009	1,000	1,003
Mod.D (9df)	0,757	0,055	0,753	0,040	0,653	0,853	0,743	0,825

In the three simulation studies we have the following common results. In some samples the RMSEA is equal to zero for wrong models. All the means of fit indices depend on sample size except for IFI. Only ECVI select on the average the correct model. All the fit indices examined exclude the severe misspecified model. For small specification errors, the parameters estimate always suggest the correct model.

6. CONCLUSION

A wide variety of fit indices has been proposed and many are available on software packages. The effect of estimation methods, sample size, nonnormal data on fit indices has been the object of many studies (see Gerbing and Anderson, 1993) but little is know about the sensitivity of the fit measures to detect misspecification.

To date there is not an index considered as the best index in fit assessment and there are not threshold values accepted to distinguish between a "correct" and an "incorrect" model. However, there are some points of consensus in overall fit controversy, they are listed in Bollen and Long (1993b). It is important to underline that several different measures of overall fit should be used and the fit of the components of a model should not be ignored.

In this study, the distribution of the data is normal and the replication reduce the effect of sampling variability. Thus, we examine the behaviour of fit indices in a nearly ideal situation of research. Notwithstanding, in the three studies only ECVI select on the average the correct model for all the sample sizes. The others fit indices do not seem sensitive in detection of slightly misspecification and RMSEA is equal to zero in some samples for wrong models. All the fit indices examined exclude the severe misspecified model. For small specification errors, the parameters estimate always suggest the correct model. Further researches are necessary for a better understanding of the sensitivity of the fit measures to detect misspecification. The overall fit indices are only one of the available statistical tools in model selection. In our example, the parameter estimate were the principal tool in the selection of the correct model among slightly misspecified models. Other empirical tools in model selection are the analysis of residuals between fitted and observed covariances, the t-values of parameter estimates, the modification indices, but the study of these empirical tools are not the purpose of this paper.

In applied research the sample variability, the quality of the measurements and the nonnormal distribution of the variables are other potential source of errors in model selection. In SEM, as in many other statistical models, the choice of a model among competitive models with nearly the same fit to the data is often a subjective matter instead of a statistical problem.

Ministero della salute ROBERTO DI NATALE

ACKNOWLEDGMENTS

The Author thanks Ms. Marina Scardecchia for supporting the realization of the paper.

REFERENCES

H. AKAIKE (1987), *Factor analysis and AIC,* "Psychometrika", 52, pp. 317-332.

J.C. ANDERSON, D.W. GERBING (1984), *The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis*, "Psychometrika", 49, pp. 155-173.

P.M. BENTLER (1990), *Comparative fit indices in structural models*, "Psychological Bulletin", 107, pp. 238-246.

P.M. BENTLER, D.G. BONNET (1980), *Significance tests and goodness of fit in the analysis of covariance structures*, "Psychological Bulletin", 88, pp. 588-606.

K.A. BOLLEN (1989a), *Structural equations with latent variables*. Eds. Wiley, New York.

K.A. BOLLEN (1989b), *A new incremental fit index for general structural equation models,* "Sociological Methods and Research", 17, pp. 303-316.

K.A. BOLLEN, J. S. LONG. EDITORS (1993a), *Testing structural equation models*, Newbury Park, CA: Sage.

K.A. BOLLEN, J. S. LONG (1993b), Introduction, In K.A. Bollen and J.S. Long (Eds.), *Testing structural equation models*, Newbury Park, CA: Sage.

M.W. BROWNE (1974), *Generalized least squares estimators in the analysis of covariance structures*, "South African Statistical Journal", 8, pp. 1-24.

M.W. BROWNE (1984), *Asymptotic distribution free methods in analysis of covariance structures*, "British Journal of Mathematical Statistical Psychology", 37, pp. 62-83.

M.W. BROWNE, R. CUDECK, (1989), *Single sample cross-validation indices for covariance structures* "Multivariate Behavioural Research", 24, pp. 445-455.

C.-P. CHOU P.M. BENTLER, A. SATORRA (1991), *Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study*, "British Journal of Mathematical Statistical Psychology", 44, pp. 347-357.

R. DI NATALE, A. SABA (1997), *LISREL models to study the determinants of food choice*, "Statistica Applicata: Italian Journal of Applied Statistics", 9, 4, pp. 473-488.

O. D. DUNCAN (1966), *Path analysis: sociological examples*, "American Journal of Sociology", 72, pp. 1-16.

D.W. GERBING, J. C. ANDERSON (1993), *Monte Carlo evaluations of goodness-of-fit indices for structural equation models*. In K.A. Bollen and J.S. Long (Eds.), Testing structural equation models, (pp. 40-65). Newbury Park, CA: Sage.

R. D. GOFFIN (1993), *A comparison of two new indices for the assessment of fit of structural equation models*, "Multivariate Behavioural Research", 28, 2, pp. 205-214.

L. HU P. M. BENTLER, Y. KANO (1992), *Can test statistics in covariance structure analysis be trusted?*, "Psychological Bulletin", 112, 2, pp. 351-362.

K. G. JÖRESKOG (1969), *A general approach to confirmatory maximum likelihood factor analysis*, "Psychometrika", 34, pp. 183-202.

K. G. JÖRESKOG, A.S. GOLDBERGER (1972) *Factor analysis by generalized least squares*, "Psychometrika", 97, pp. 243-260.

T.J. LA DU, J.S. TANAKA (1995), *Incremental fit index changes for nested structural equation models*, "Multivariate Behavioural Research", 30, 3, pp. 467-482.

G.M. MARUYAMA (1998), *Basic of structural equation modeling*, Eds. Sage Publications, Inc., California.

R.P. MCDONALD, H.W. MARSH (1990), *Choosing a multivariate model: noncentrality and goodness-of-fit*, "Psychological Bulletin", 107, pp. 247-255.

B. MUTHEN, D. KAPLAN (1985), *A comparison of some methodologies for the factor analysis of non-normal Likert variables,* "British Journal of Mathematical Statistical Psychology", 38, pp. 171-180.

B. MUTHEN, D. KAPLAN (1992), *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 Statistical Psychology", 45, pp. 19-30.

J. H. OUD J. H. VAN DEN BERCKEN, R. J. ESSERS (1990), *Longitudinal factor score estimation using the Kalman filter*, "Applied Psychological Measurement", 14, pp. 395-418.

E.E. RIDGON, C.E. FERGUSON (1991), *The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data*, "Journal of Marketing Research", 28, pp. 491-497.

A. SABA, R. DI NATALE (1998) *Attitudes, Intention and Habit: their role in Predicting the Actual Consumption of Fats and Oils*, "Journal of Human Nutrition and Dietetics",11, pp. 21-32.

SAS INSTITUTE INC. (1989), SAS/STAT user's guide version6 (4th ed. Vol1.) Cary, NC:Author.

A. SATORRA (1990), *Robustness issues in structural equation modeling: a review of recent developments*, "Quality and Quantity", 24, pp. 367-386.

L. SCHRAGE (1979), *A more portable FORTRAN random number generator*, "ACM Transactions on Mathematical Software", 5, pp. 132-138.

J.H. STEIGHER (1990), *Structural model evaluation and modification: An interval estimation approach*, "Multivariate Behavioural Research", 25, pp. 173-180.

L. H.-Y. YAU S.-Y. LEE, W.-Y. POON (1993), *Covariance structure analysis with three-level data*, "Computational Statistics and Data Analysis", 15, pp. 159-178.

K. H. YUAN, P. M. BENTLER (1997), *Mean and Covariance Structure Analysis: Theoretical and Practical Improvements*, "Journal of the American Statistical Association, Theory and Methods", June 1997, 92, 438, pp. 767-774.

Y.-F. YUNG, P. M. BENTLER (1994), *Bootstrap-corrected ADF test statistics in covariance structure analysis*, "British Journal of Mathematical and Statistical Psychology", 47, pp. 63-84.

S. VAN BUUREN (1997), *Fitting ARMA Time Series by Structural Equation Models*, "Psychometrika", 62, 2, pp. 215-236.

S. WRIGHT (1934), *The method of Path Coefficients*, "Annals of Mathematical Statistics", 5, pp. 161-215.

SUMMARY

Fit assessment and selection between competitive models in SEM

This paper outlines some issues in fit assessment and in selection between competitive models in Structural Equation Models (SEM). The theory on chi-square statistic and the results of some simulation studies are reviewed and the methodological consequences are discussed. Several goodnees-of-fit indices are discussed and their employment in fit assessment and in the selection between competitive models are taken into consideration. A Monte Carlo study was used to point out the difficulties connected with model selection in SEM under misspecification.

Keywords: structural equation models, LISREL, fit assessment, competitive models, misspecification.

RIASSUNTO

Modelli ad equazioni strutturali: la valutazione dell'adattamento e la selezione tra modelli competitivi.

In questo lavoro si evidenziano alcuni problemi nella valutazione dell'adattamento del modello ai dati e nella selezione di modelli competitivi nel contesto dei modelli ad equazioni strutturali formalizzati secondo il modello LISREL. La teoria sulla statistica del chiquadrato è riesaminata sulla base dei risultati di alcune recenti simulazioni e si individuano le conseguenze metodologiche di tali risultati empirici. Si discutono quindi i principali indici di adattamento proposti e il loro utilizzo nell'ambito della selezione tra modelli competitivi. Uno studio di tipo Monte Carlo è utilizzato al fine di sottolineare le difficoltà connesse alla selezione di modelli competitivi con errori di specificazione.