Estudios de Economía Aplicada (1996)
No. 6, pág. 183-210.

A strategy for the assessment of overall model fit in structural equation modeling for business economics

MANUEL SÁNCHEZ PÉREZ
Universidad de Almería

Esta nueva versión incluye todas las correcciones sugeridas por el censor, las cuales me han parecido oportunas y por las que les quedo muy agradecido.

ABSTRACT

In this paper we propose a strategy for the assessment in economics of the overall model fit obtained using structural equation models. Due to goodness-of-fit of a model is more a relative process than one with absolute criteria, evaluation requieres multiple criteria. The overal evaluation strategy proposed is composed of two main parts: basic assumptions and examination of the solution, and the assessment of the overall model in itself. In both cases, detailed formula and measures are given.

Keywords: Econometrics, economics models, structural equation, model fit assessment

RESUMEN

En el presente trabajo se propone una estrategia de valoración a nivel global de los modelos obtenidos en economía mediante los modelos de ecuaciones estructurales simultáneas. Se proponen múltiples criterios debido a que la evaluación de la bondad de un modelo es más un proceso relativo que una etapa con criterios determinados de forma absoluta. La estrategia planteada consta de dos partes principales: el cumplimiento de los supuestos básicos y el análisis de la solución, de una parte, y la valoración propiamente dicha del modelo a nivel global. En ambos casos, se detallan las fórmulas y medidas especificadas.

Palabras clave: Econometría, Modelos económicos, Ecuaciones estructurales, Evaluación ajuste

Código UNESCO: 530204

Artículo recibido en diciembre de 1995. Revisado en marzo de 1996.

ÍNDICE

1. Introduction

Structural equation modeling (SEM) -also called causal modeling or covariances analysis- constitutes one of the recent developments in multivariate statistics for conducting research in economics and business.

There are three main reasons why the method is so attractive. First, it provides a straighforward and statistically efficient method of dealing with multiple relationships simultaneously. Second, it has the ability to perform exploratory and confirmatory analyses. And third, this method analyses an integrative function, making other techniques such as multiple regression, path analysis, factor analysis, canonical correlation analysis, ANOVA, MANOVA, ANCOVA, MANCOVA, simultaneous equation systems, the analysis of multitrait-multimethod matrices (MTMM), and others, special cases of this more general approach.

Many research work has been done using this tecnique, but not always the researchers evaluate properly the model that has been built. Research in social sciences presents special problems such sampling procedures, indirect measurement, or errors in model specifications that lead to perform a coherent asseement agenda.

In this paper, a strategy is proposed for the assessment of the overall model fit using this method. Then, after exposing the basics of this multivariate tecnique, we will analyse deeply different measures for evaluate the *overall model fit* using this method[1].

2. Models foundations and relations

Let Z_{px1} denote a vector of observable variables that is structured as a linear function of more basic variables, some of which may be latent (unobservable). Let Λ be a matrix of parameters and ξ and ε vectors of respectively common and specific (unobservable) factors, which are assumed to be uncorrelated with each other. Then, we can consider the next *factor analysis model* (also called measurement model)[2]:

$$z = \Lambda\xi + \varepsilon \quad [1]$$

Another typical model is the *structural equation model*, or *'path analysis' model*. This model says that the vector z of observable variables partitions in endogenous and exogenous variables, vector y and x respectively, in such a way that the following linear equations holds:

$$y = By + \Gamma x + u \quad [2]$$

where u is a vector of (unobservable) equation 'disturbances', and B and Γ are parameter matrices. Uncorrelation (non-necessarily independence) between x and u is to be assumed.

A very general multivariate relation is the *structural equation model with latent variables*, so-called *LISREL model*, that combines equations of types [1] and [2] into the structural form of the following three equations:

- the structural equation model: $\eta = B\eta + \Gamma\xi + \zeta$
- the measurement model for y: $y = \Lambda^y\eta + \varepsilon$
- the measurement model for x: $x = \Lambda^x\xi + \delta$

where:
η is a $mx1$ vector of latent endogenous variables;
ξ is a $nx1$ vector of latent exogenous variables;
ζ is a $mx1$ vector of latent errors in equations;
B is a mxm coeficient matrix for the latent endogenous variables;
Γ is the mxn coefficient matrix for the latent exogenous variables;
y is a $px1$ vector of observed indicators of η;
x is a $qx1$ vector of observed indicators of ξ;
ε is a $px1$ vector of measurement errors for y;
δ is $qx1$ vector of measurement errors for x;
Λ^y is a pxm matrix of coefficients relating y to η;
Λ^x is a qxn matrix of coefficients relating x to ξ.

Assumptions:
1) $(I-B)$ is nonsingular, so that $(I-B)^{-1}$ exists.;
2) $E(\eta) = E(\xi) = E(\zeta) = E(\varepsilon) = E(\delta) = 0$;
3) ζ is homoscedastic and non-autocorrelated;
4) ε is uncorrelated witn η and ξ;
5) δ is uncorrelated with ξ and η.

The linear relations of the model imply a specific structure for the covariance matrix Σ of the observed variables; that is, the model expressed by the linear equations gives rise to a covariance structure (Σ) for the observable variables i.e., to a $(p+q)x(p+q)$-dimensional matrix-valued function $\Sigma=\Sigma(\theta)$ of θ. This function is generally non-linear, but smooth enough to be twice continuosly differentiable. This is the fundamental hypothesis for structural equation models: the covariance matrix of the observed variables is a function of a set of parameters. Σ represents

ÍNDICE

the population covariance matrix of observed variables, θ is a vector that contains the model parameters, $\Sigma(\theta)$ is the covariance matrix written as a function of θ.

It can be shown (Bollen, 1989) that the covariance matrix for the observed *y* and *x* variables as a function of the model parameters (θ) is:

$$\Sigma = \Sigma(\theta) = \begin{bmatrix} \Sigma_{yy}(\theta) & \Sigma_{yx}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{xx}(\theta) \end{bmatrix} =$$

$$\begin{bmatrix} \Lambda^{y}(I-B)^{-1}(\Gamma\Phi\Gamma'+\Psi)(I-B)^{-1}\Lambda'^{y}+\Theta^{\varepsilon} & \Lambda^{y}(I-B)^{-1}\Gamma\Phi\Gamma'^{x} \\ \Lambda^{x}\Phi\Gamma'(I-B')^{-1}\Lambda'^{y} & \Lambda^{x}\Phi\Lambda'^{x}+\Theta^{\delta} \end{bmatrix}$$

where:
Φ is a *nxn* covariance matrix of ξ;
Ψ is a *mxm* covariance matrix of ζ;
Θ^{ε} is a *pxp* covariance matrix of ε;
Θ^{δ} is a *qxq* covariance matrix of δ.

The problem is that given the sample covariance matrix of the observed variables, S, we have to find a vector θ of parameter so that $\Sigma(\theta)$ is equal to S.

The fitting of $\Sigma(\theta)$ to S can be framed into a least squares approach using next methods of estimation:

- Unweighted Least Squares (ULS)
- Generalized Least Squares (GLS)
- Maximum Likelihood (ML)
- Generally Weighted Least Squares (WLS)
- Diagonally Weighted Least Squares (DWLS)

Next, we expose different concepts and steps in model assessment, which are summarized in table 1.

Table 1. Steps for the assessment of overall model fit in structural equation modeling

PRIOR EVALUATION OF THE MODEL
1) Assessment of the assumptions of structural equation modeling
2) Examination of the solution: • Offending estimates • Standard errors • Correlations of parameter estimates • Squared multiple correlations and coefficients of determination
OVERALL GOODNESS-OF-FIT MEASURES
1) Absolute fit measures • Chi-square test • Power of the chi-square-measure • Goodness of Fit index (GFI) • Root Mean Squared Residual (RMSR) • Critical N (CN)
2) Incremental Fit Measures • Bollen Index (ρ) • Tucker-Lewis Index (TLI), or Non-Normed Fit Index (NNFI) • Normed Fit Index (NFI), or Bentler and Bonett Index • Comparative Fit Index (CFI) • Incremental Fit Index (IFI) • Relative Fit Index (RFI)
3) Parsimonious Fit Measures • Adjusted Goodness of Fit Index (AGFI) • Normed Chi-square • Parsimonious Normed Fit Index (PNFI) • Akaike Information Criteria (AIC) • Other parsimonious fit measures (AIC y ECVI)

3. Prior evaluation of the model: basic assumptions and examination of the solution

There is no single measure or set of measures that have been agreed upon as the only measures needed. The researcher is encouraged to employ one or more measures from each class. Assessing the goodness-of-fit of a model is more a

relative process than one with absolute criteria. The application of multiple measures will enable the researcher to gain a consensus across types of measures as to the acceptability of the proposed model.

In this section we wil analize the fulfillment of the basics assumptions and the coherence of the solution.

3.1. Assessment of the assumptions of structural equation modeling

SEM shares three assumptions with other multivariate methods (multiple regression, factor analysis, multivariate analysis of variance, discriminant analysis): independent observations, random sampling of respondents, and the linearity of all relationships. In addition, a key issue in SEM is the distributional characteristics of the data, particularly the departure from multivariate normality. However, the normality assumption can be relaxed using asymptotic distribution-free methods[3].

The ML and GLS fitting functions are justified when the distribution of the observed variables has a multinormal distribution or when the distribution of the observed variables has no excess kurtosis. Under either of these conditions and with a valid model, θ from ML or GLS is a consistent and asymptotic efficient estimator and (N-1) times the minimum value of the fit function for the specified model (χ^2) has an asymptotic chi-square distribution suitable to test the overall model fit. ULS produces consistent estimators, although these may not be the best in terms of asymptotic efficiency.

An approach to solve the nonnormality or excessive kurtosis is to employ WLS method of estimation. Under WLS, $\hat{\theta}$ is a consistent estimator of θ and may provide asymptotically efficient estimators and an asymptotic chi-square distribution suitable for tests of the overall model fit.

3.2. Examination of the solution

Next step in assesing the overall goodness of a model is to examine the results of an analysis. One should pay careful attention to the following measures:

1) Offending estimates
2) Standard errors
3) Correlations of parameter estimates
4) Squared multiple correlations and coefficients of determination

3.2.1. Offending estimates

An "offending estimate" is any value of a parameter that exceeds its theoretical or acceptable limits. The more common occurrences are: (1) negative error variances or nonsignificant error variances for any latent variable, (2) standardized coefficients exceeding or very close to 1.0, (3) very large standard errors associated with any estimated coefficient, (4) correlations greater than one, or (5) covariance or correlation matrices which are not positive definite. If encountered, the researcher must first resolve the problematic occurrence before evaluating any specific results of the model, as changes in one part of the model can have significant effects on other results.

If identification problems are corrected and the problems still exist, severall other remedies are available. In the case of negative error variances ("Heywood case"). The "offending estimates" issue is due probably to multicollinearity. One possibility suggested by Bentler and Chou (1987) is to fix the offending error variances to a very small positive value (0.005). This procedure provides a solution of the estimation process. However, this remedy only masks the underlying problem and must be considered when interpreting the results due to the fact that this output can not be used to test hypotheses.

Classic approach are not completely satisfactory to deal with multicollinearity. An interesting approach is to use a "ridge regression" in model estimation (a constant times the diagonal is added to the diagonal before estimations begins). This procedure has been used successfully by Jagpal (1982) and it is included in recent versions of software such as LISREL. However, remedies against multicollinearity are valid when it is imperfect. If multicollinearity is perfect, model especification must be verified trying to get more information about the variable that presents an "offending estimate". If, anyway, it is no possible to solve the problem, the obstacle is not the technique, but the data. When correlations in the standardized solution exceed 1.0, or two estimates are correlated highly, then the researcher should consider elimination of one of the latent variables or ensure that true discriminant validity has been established among the constructs. In many instances, such situations are the result of atheoretical models, established without substantive theoretical justification or modified solely on the basis of empirical considerations.

In case of correlations larger than one or a matrix that is not positive definite, it is possible that there are errors in data or the model defined is not the one intended. In the latter case, the researcher should redefine the model again.

ÍNDICE

3.2.2. Standard errors

Standard errors are estimates of the precision of each parameter estimate. The main diagonal of the covariance matrix contains the asymptotic variances of the parameter estimators. When the sample estimates are substituted for the unknown parameters, the square root of the main diagonal of the asymptotic covariance matrix provides estimated asymptotic standard errors. Small standard errors correspond to good precision and large standard errors to poor precision.

The standard errors for ML and GLS are correct under multivariate normality of the observed variables and for WLS if the correct weight matrix is used. Standard errors for ML and GLS are robust against moderate departures from normality. Standard errors for ULS and DWLS are only approximate. A frequent situation where the standard errors are not accurate is when the correlation rather than the covariance matrix is analysed.

3.2.3. Correlations of parameter estimates

An estimate of the asymptotic covariance matrix of the parameter estimates is obtained from the information matrix. The correlations of parameter estimates are obtained by scaling the asymptotic covariance matrix to an asymptotic correlation matrix. This transformation draws on the definition of the correlation as the covariance of two variables divided by the product of their standard deviations:

$$\operatorname{asym}\hat{\rho}_{\hat{\theta}_i\hat{\theta}_j} = \frac{\operatorname{a\,cov}(\hat{\theta}_i, \hat{\theta}_j)}{\sqrt{\operatorname{a\,var}(\hat{\theta}_i)\operatorname{a\,var}(\hat{\theta}_j)}} \qquad [3]$$

where:

$\hat{\theta}_i$ = the *i*th estimated parameter

$\hat{\theta}_j$ = the *j*th estimated parameter

Correlations that are extremely large indicate that the estimates of the two-parameter estimators are very closely associated. High correlations sometimes are a symptom of severe collinearity. Also, the model may be non-identified and some of these parameters cannot be determined from the data.

The asymptotic correlation matrix is accurate for the analysis of covariance matrices but not for the analysis of correlation matrices, because such an analysis leads to unprecise estimates of the variances and covariances of parameter estimates.

3.2.4. Squared multiple correlations and coefficients of determination

Squared multiple correlations for each observed variable separately and also for each structural equation can be obtained. The squared multiple correlation is defined as follows:

$$1 - \frac{\hat{\text{var}}(\delta_i)}{\hat{\sigma}_{ii}} \quad [4]$$

where:

$\hat{\text{var}}(\delta_i)$ is the estimated error variance of the *i*th variable, and $\hat{\sigma}_{ii}$ is the fitted variance of the *i*th variable.

The coefficient of determination is:

$$1 - \frac{\left\|\hat{\Theta}\right\|}{\left\|\hat{\Sigma}\right\|} \quad [5]$$

where:

$\left\|\hat{\Theta}\right\|$ is determinant of $\hat{\Theta}$ and $\left\|\hat{\Sigma}\right\|$ is the determinant of the fitted covariance matrix $\hat{\Sigma}$ of the observed variables. These measures show how well the observed variables serve, separately or jointly, as measurement instruments for the latent variables. The measure should be between zero and one, large values being associated with good models.

The squared multiple correlation (R^2_{ii}) for the *i*th structural equation is defined as:

$$1 - \frac{\hat{\text{var}}(\zeta_i)}{\hat{\text{var}}(\eta_i)} \quad [6]$$

where:

$\hat{\text{var}}(\zeta_i)$ and $\hat{\text{var}}(\eta_i)$ are the estimated variances of ζ_i and η_i, respectively.

The total coefficient of determination (R^2) for all structural equations jointly is defined as:

$$1 - \frac{\left\|\hat{\Psi}\right\|}{\left\|\hat{cov}(\eta)\right\|} \quad [7]$$

where:

$\left\|\hat{\Psi}\right\|$ is the determinant of $\hat{\psi}$ and $\left\|\hat{cov}(\eta)\right\|$ is the determinant of the estimated covariance matrix of η. However, this total coefficient has not defined limits. So, we could use the coefficient of alienation (ρ_A) suggested by Dhrymes (1974). This coefficient is defined as:

$$\rho_A = \frac{\begin{vmatrix} \hat{cov}(\eta) & \hat{cov}(\eta, \xi) \\ \hat{cov}(\xi, \eta) & \hat{cov}(\xi) \end{vmatrix}}{\left\|\hat{cov}(\eta)\right\| \left\|\hat{cov}(\xi)\right\|} \quad [8]$$

It assumes the value zero when the covariance between the variables in question do not exist, and it takes the value unity when a perfect linear relation exists between the two variables.

4. Assessment of the overall model fit

The next step is to assess the overall model fit with several goodness-of-fit measures. Goodness-of-fit is a measure of the correspondence of the actual or observed input (covariance or correlation) matrix with that predicted from the proposed model. If the proposed model has an acceptable fit by whatever criteria used, it has not been proved that the proposed model is correct, but only confirmed that it is one of several possible acceptable models. Indeed, several different models might have equally acceptable model fit.

In SEM, there must be a certain ratio between the size of the covariance or correlation matrix and the number of estimated coefficients. The difference between the number of coefficients for a perfectly identified model (one coefficient for each covariance/correlation) and the actual number of coefficients in the proposed model (t) is termed the *degrees of freedom* (df), that is, $df = \frac{1}{2}[(p+q)(p+q+1)]-t$. In doing so, the model achieves *parsimony*, that is, the degree to which a model achieves model fit for each estimated coefficient. The

objective is not to minimize the number of coefficients or maximize the fit, but instead to maximize the amount of fit per estimated coefficient. The objective is to avoid "overfitting" the model with additional coefficients in an attempt to make small gains in model fit.

The goodness of fit measure for the whole model fall into one of three types:

1) Absolute fit measures
2) Incremental fit measures
3) Parsimonious fit measures

4.1. Absolute fit measures

Absolute fit measures determine the degree to which the overall model (structural and measurement model) predicts the observed covariance or correlation matrix and does not express the quality of the model as judged by any other internal or external criteria. Furthermore, if any of the overall measures indicates that the model does not fit the data well, it does not tell what is wrong with the model or which part of the model is wrong.

4.1.1. Chi-square (χ^2) test

For ML, GLS and WLS, the χ^2-measure is (N-1) times the minimum value of the fit function for the specified model:

$$(N-1)F(\hat{\theta}) \quad [9]$$

The χ^2-measure is distributed asymptotically as a chi-square distribution under certain conditions. For ULS and DWLS, certain adjustments are made so as to make the χ^2-measure also asymptotically correct for these methods. The χ^2-measure is correct for ULS, ML, GLS and DWLS under multinormality of the observed variables if a covariance matrix is analyzed. With WLS, the χ^2-measure is correct if a correct weight matrix is used. If a correlation matrix is analyzed with ML, χ^2 is correct only if the model is scale-invariant and $\text{diag}(\hat{\Sigma}) = \text{diag}(S)$.

If the model is correct and the sample size is sufficiently large, the χ^2-measure may be used as a test statistic for testing the model against the alternative that Σ is unconstrained. The null hypothesis of the chi-square test is $H_0 : \Sigma = \Sigma(\theta)$. This implies that the restrictions for the model are correct. Rejection of H_0 suggests that at least one restriction is wrong so that $\Sigma \neq \Sigma(\theta)$. Since H_0 is equivalent to the hypothesis that $\Sigma - \Sigma(\theta) = 0$, the χ^2-test is a simultaneous test that all residuals in $\Sigma - \Sigma(\theta)$ are zero. The chi-square measure is associated to a probability value (p). The level of p has been set at greater than 0.05 or 0.10 to accept the model.

ÍNDICE

Also, the chi-square measure can help us to test the statistical significance of the differences in chi-square estimators between two nested model[4].

4.1.2. Power of the chi-square-measure

Although the χ^2-measure may be viewed theoretically as a test statistic for testing the hypothesis that Σ is of the form implied by the model against the alternative that Σ is unconstrained, it should be emphasized that such a use of χ^2 is not valid in most applications. In most empirical work, the model is only tentative and is regarded as an approximation to reality. From this point of view the statistical problem is not one of testing a given hypothesis (which a priori may be considered false), but rather one of fitting the model to the data and to decide whether the fit is adequate or not.

Instead of regarding χ^2 as a test statistic, one should regard it as a goodness (or badness)-of-fit measure in the sense that large χ^2-values correspond to bad fit and small χ^2-values to good fit. The degrees of freedom serve as a standard by which to judge whether χ^2 is large or small. The χ^2-measure is sensitive to sample size and very sensitive to departures from multivariate normality of the observed variables. Large sample sizes and departures from multivariate normality tend to increase χ^2 over and above what can be expected due to specification error in the model. One reasonable way to use χ^2-measures in comparative model fitting is to use χ^2-differences in the following way. If a value of χ^2 is obtained, which is large compared to the number of degrees of freedom, the fit may be examined and assessed by an inspection of the fitted residuals, the standardized residuals, and the modification indices.

In particular, the power of the χ^2 to detect discrepancies between Σ and $\Sigma(\theta)$ partially depends on sample size. The estimate of the χ^2 increases in direct proportion to (N-1) and the power of the test increases as N increases. A large sample may increase our confidence that the residual matrix $[\Sigma-\Sigma(\theta)]$ is not zero, but the substantive significance of the difference may be negligible; that is, the χ^2-measure becomes "too sensitive" and almost any difference is detected. Alternatively, a small sample can reduce the power of the test so that with a small N we cannot detect even large differences between Σ and $\Sigma(\theta)$. The χ^2 often shows acceptable fit even when none of the model relationships are shown to be statistically significant. Various analysis establish 50 as the minimum recommended level of sample size for ML, but a sample this small is not recommended. The minimum sample size to get appropiate χ^2-values using ML is 100. Samples exceeding 400 to 500 make χ^2 to indicate poor fit. So, the

recommended sample size or "critical sample size" is 200, with unreliable χ^2 outside of the range between 100 and 200.

Due to these several drawbacks, it seems reasonable to analyse *the power*[5] *of a chi-square test* to assess its quality. To explain the power of the chi-square test to evaluate the null hypothesis (i. e., H_0: $\Sigma=\Sigma(\theta)$), we can use the *likelihood ratio* (LR) test statistic[6]. However, this test is a function of the chi-square from [9]. Then, the strict interpretation of the LR test requires the assumptions underlying the estimation methods be valid.

Let c the chi-square obtained from [9]. Under the standard assumptions c has an asymptotic chi-square distribution when H_0 is valid. However, when the model does not hold ($\Sigma\neq\Sigma(\theta)$), that is, H_a is correct and H_0 is tested, c is distributed as noncentral chi-square[7] with df=½(p+q)(p+q+1) and non-centrality parameter $\lambda=(N-1)F[\Sigma, \Sigma(\theta)]$, an unknown population quantity that may be estimated as: $\hat{\lambda}=$Max {(c-df), 0}. One can also set up a confidence interval for λ. Let $\hat{\lambda}_L$ and $\hat{\lambda}_U$ be the solutions of $G(c|\lambda_L, df)=0.95$, $G(c|\lambda_U, df)=0.05$, where $G(x|\lambda, df)$ is the distribution function of the noncentral chi-square distribution with non-centrality parameter λ and df degrees of freedom. Then $(\hat{\lambda}_L; \hat{\lambda}_U)$ is a 90% confidence (i.e., power) interval for λ.

Table 2. Values of the noncentral chi-square

	Power to detect difference at 5% significance level				
	0.50	0.60	0.70	0.80	0.90
1 df test	3.84	4.90	6.17	7.85	10.51
2 df test	4.96	6.21	7.70	9.64	12.65

From a more general point of view, If we determine the df and the noncentrality parameter, then we know the distribution of the test statistics when H_a is valid and H_0 is tested. From tabled values of the non-central chi-square, we can estimate the power of a test. Table 2 contains values of the noncentrality parameter required to give certain levels of power for 1 df and 2 df tests at the 5% significance level.

Once the validity of the model has been establish, one can test structural hypotheses about the parameters θ in the model such that:

- certain θ's have particular values (fixed parameters):
- certain θ's are equal (equality constraints);
- certain θ's are specified linear or nonlinear functions of other parameter.

ÍNDICE

Each of these types of hypotheses leads to a model with fewer parameters, u, say, for $u<t$ and with a parameter vector ν of order $(ux1)$ containing a subset of the parameters in θ. The model with parameters ν is called the *null hypothesis* H_0 and the model with parameters θ is called the *alternative hypothesis* H_1. The likelihood ratio test statistic for testing H_0 against H_1 is then

$$LR = ch_0 - ch_1,$$

which is used as χ^2 with t-u degrees of freedom. The degrees of freedom can also be computed as the difference between the degrees of freedom associated with ch_0 and ch_1. For instance, in a statistical test of a single parameter, the df equals one.

To use the test formally, one chooses a significant level α (probability of a type 1 error) and rejects H_0 if LR exceeds the $(1-\alpha)$ percentile of the χ^2 distribution with t-u degrees of freedom.

For the usual chi-square test of overall model fit where H_a represents an exactly identified (though usually not fully described) alternative model, the df equals $\frac{1}{2}(p+q)(p+q+1)$-t, where t is the number of free parameters for H_0.

An illustration is given of a procedure with LR to determine the power of the chi-square test of overall model fit. Let us suppose that the following matrices correspond to the estimated parameters of the original model (H_0), with sample size equal to 173:

$$B = \begin{bmatrix} 0 & 0 \\ -0.285 & 0 \end{bmatrix} \qquad \Gamma = \begin{bmatrix} 0 & -0.087 \\ 0 & 0.058 \end{bmatrix}$$

$$\Psi = \begin{bmatrix} 12.961 & 0 \\ 0 & 8.488 \end{bmatrix} \qquad \Phi = \begin{bmatrix} 1.021 & 7.139 \\ 7.139 & 215.662 \end{bmatrix}$$

Now, we specify the values of θ for the alternative hypothesis. To do that, we can fit an unrestricted model by setting free the parameters γ_{11} and γ_{21}. Let us suppose the estimation for γ_{11} and γ_{21} are 0.378 and 0.328, respectively. Then:

$$B = \begin{bmatrix} 0 & 0 \\ -0.285 & 0 \end{bmatrix} \qquad \Gamma = \begin{bmatrix} 0.378 & -0.087 \\ 0.328 & 0.058 \end{bmatrix}$$

$$\Psi = \begin{bmatrix} 12.961 & 0 \\ 0 & 8.488 \end{bmatrix} \qquad \Phi = \begin{bmatrix} 1.021 & 7.139 \\ 7.139 & 215.662 \end{bmatrix}$$

The second step is to generate the implied covariance matrix ($\Sigma(\theta)$) from these parameter values. One way to do this is to start all parameter matrices at the values listed above, and fix all matrices so that no parameters are estimated. The fitted moment matrix from this run is $\Sigma(\theta)$ at $\theta=c$, where c is a known constant. The chi-square for this alternative model is zero by construction, whereas that for H_0 is typically positive.

Then, H_0 is fitted to $\Sigma(\theta)$ ($\theta=c$), that is, the implied covariance matrix obtained is analysed under H_0 (not the alternative model), while keeping the sample size the same. The chi-square from this run is the estimated of the non-centrality parameter (NCP). The df equals the number of parameters that distinguish H_0 from the alternative model (in this case, 2). The estimate of the non-centrality parameter for this example is 6.21, with df=2, α=0.05, and N=173. Then, table 2 suggests that we would have only a 60% chance of detecting a false H_0 ($\gamma_{11}=\gamma_{21}=0$) when the alternative parameterization is true (free γ_{11} and γ_{21}).

In fact, this procedure uses the "wrong model" (H_0) on the correct implied covariance matrix for the alternative model (H_a). We know that H_0 is incorrect, and we analyze a population covariance matrix $\Sigma(\theta)$ where θ is known. The chi-square for the alternative model is zero by construction, whereas that for H_0 is typically positive (taking into account that we are testing the power of the chi-square for the overal model).

This procedure applies to tests of individual parameters or groups of parameters, and not only to tests of overall fit. These power calculations are based on the chi-square difference LR test for comparing models where the parameter values for the least restrictive model usually generate the $\Sigma(\theta)$ that is analysed. The models can differ in a single or in many parameters. The preceding steps are followed except that both models are fitted to the implied covariance matrix ($\Sigma(\theta)$ with $\theta=c$). The non-centrality parameter estimate equals the difference in the non-centrality estimates of the two models, with df equal to the difference in the df's for the models. Generally, there is no need to fit the least restrictive model since its non-centrality parameter estimate is zero by construction.

Power is influenced by many factors. If α is increased, the power of the test goes up. This is one possible means to increase power. Another is to increase the sample size, if this is possible, when collecting new data. In this way, table 2 also indicates that we would have required a non-centrality parameter of 9.64 from

ÍNDICE

the previous analysis to have attained 80% power. We could obtain such a value by increasing the sample size. Now, the chi-square obtained from a test generally increases in proportion to the sample size. So that we can obtain an estimate of the sample size required (SSR) by using the next formula (Dunn, Everitt and Pickles, 1993): SSR = (NCP x SSU) / ECS

where:

SSR = estimate of the sample size required

NCP = obtained non-centrality parameter

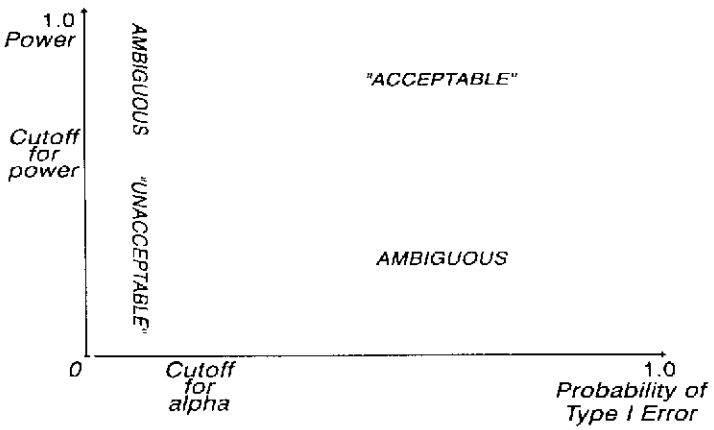
SSU = sample size used for the expected chi-square

ECS = expected chi-square

In our example, the required sample size is 268 (= 9.64 x 173/6.21), at the 5% level.

Figure 1 illustrates several possible outcomes in assessing quality of a statistical test. The horizontal axis shows the probability of a Type I error with the preselected cutoff-level for α. The preselected minimum power level is on the vertical power axis. If the model H_0 has a Type I probability lower than the cutoff α and if the power of the test with respect to H_a is lower than the cutoff for power, this case falls in the lower left-hand quadrant of figure 1 labeled "unacceptable". As the label suggests, H_0 would not be very acceptable since with low power and a significant χ^2-value, it is likely that more than minor specification errors are in the H_0 model. In contrast, when the Type I probability is larger than α and the power of the test is high, we obtain the situation shown in the upper right corner of figure 1. The high power means a high probability of detecting a false H_0. The high Type I probability and high power suggest that the H_0 model is very consistent with the data since, if it were not so, these discrepancies would be detected.

Figure 1. Comparison of Type I Probability and Power of Statistical Test of H_0



Source: Bollen (1989: 175)

The two remaining quadrants of figure 1 represent more ambiguous situations. The lower-right one occurs when the Type I probability for H_0 is greater than α, but the power of the test is low. We cannot be certain whether H_0 is appropriate or whether the low power of the test prevents the detection of large errors. The upper-left quadrant presents a different problem: the Type I probability of H_0 is lower than α, but the power of the test is high. Here we do not know whether the H_0 model has major misspecifications or if the high power of the test reflects minor discrepancies.

As an example of the preceding assessment of test quality, let us consider a model with χ^2 estimate of H_0 is 1.26 with df=3, and the Type I probability (α) is equal to 0.74, indicating a nonsignificant test statistic. Considering the power estimate of 0.44, this is moderate and suggests that even with a nonsignificant χ^2-estimate for the test of H_0, we cannot be sure whether important errors are present in the model. This ambiguous situation corresponds to the lower-right quadrant of figure 1. This demonstrates the need to redesign the study so as to increase the power of the χ^2.

4.1.3. Goodness of Fit Index (GFI)

The Goodness of Fit Index (GFI) is defined as:

$$GFI = 1 - \frac{(s - \hat{\sigma})' W^{-1} (s - \hat{\sigma})}{s' W^{-1} s} \quad [10]$$

The numerator is the minimum of the fit function after the model has been fitted; the denominator is the fit function before any model has been fitted. It compares the squared residuals between prediction and actual data, but is not adjusted for the degrees of freedom. It is a nonstatistical measure ranging in value from 0 (poor fit) to 1.0 (perfect fit), although it is theoretically possible to become a negative value but this would mean that the model fits worse than any model at all. No absolute threshold levels for acceptability have been established.

Although GFI is not explicitly a function of sample size, its distribution depends on sample size, increasing as sample size increases. It also decreases as the number of indicators per latent variable, or the number of latent variables, increase, especially for smaller sample sizes.

ÍNDICE

4.1.4. Root Mean Squared Residual (RMSR)

The Root Mean Squared Residual is defined as:

$$RMSR = \sqrt{\frac{2\sum_{i=1}^{p}\sum_{j=1}^{i}(s_{ij} - \hat{\sigma}_{ij})^2}{(p+q)(p+q+1)}} \quad [11]$$

RMSR is a measure of the average of the fitted residuals and can only be interpreted in relation to the sizes of the observed variances and covariances in S. If covariances are used, it is the average residual covariance. If a correlation matrix is used, then RMSR is in terms of an average residual correlation. However, this measure works best if all observed variables are standardized.

4.1.5. Critical N (CN)

It is defined as: $CN = (\chi^2_{1-\alpha}/F) + 1$ [12]

where the critical χ^2 is the critical value of the chi-square distribution with df equal to the maintained model's df and at a selected alpha level (e.g., 0.05). The F is the value of the fitting function at S and $\hat{\Sigma}$. CN gives the sample size at which the F value would lead to the rejection of H_0 (i.e., $\Sigma=\Sigma(\theta)$) at a chosen alpha level. A tentative cutoff of CN $\geq$ 200 has been suggested. Since N does not enter the formula for CN, its calculated value is the same for a given critical χ^2 and F value for all sample sizes. However, when H_0 is valid, F goes to zero as N grows larger and CN increases with N. So, CN may lead to an overly pessimistic assessment of fit for small samples.

4.2. Incremental fit measures

This class of measures compares the proposed model to some baseline model, most often referred to as the *null model*. The null model is hypothesized to be the most simple model that can be theoretically justified. It should be some realistic model that all other models should be expected to exceed. In most cases, the null model is a single-construct model with all indicators perfectly measuring the construct. There is, however, some disagreement over exactly how to specify the null model in many situations.

4.2.1. Bollen Index (ρ)

This index compares the fit of the proposed model per degrees of freedom for the null and proposed model, and is defined as:

$$\rho = \frac{(F_{null} / df_{null}) - (F_{proposed} / df_{proposed})}{(F_{null} / df_{null})} = \frac{(\chi^2_{null} / df_{null}) - (\chi^2_{proposed} / df_{proposed})}{(\chi^2_{null} / df_{null})} \quad [13]$$

Since introducing additional parameters will lower the number of degrees of freedom, it is possible for ρ to stay the same or to decrease for more complex specifications. The lowest possible value is zero and the maximum is one (although it may be less than zero in theory). The mean of the sampling distribution of ρ increases with N.

4.2.2. Tucker-Lewis Index (TLI), or Non-Normed Fit Index (NNFI)

First, this measure was proposed as a means of evaluating factor analysis but it has been extended to SEM. It is better than the Bollen index (ρ) due to lessens the dependency of the mean of ρ on N. It combines a measure of parsimony into a relative index between the proposed and null models, resulting in values ranging from 0 to 1.0. It is expressed as:

$$TLI = \frac{(\chi^2_{null} / df_{null}) - (\chi^2_{proposed} / df_{proposed})}{(\chi^2_{null} / df_{null}) - 1} \quad [14]$$

In the numerator, TLI compares the worst fit to the sample fit. The denominator contrasts the worst fit (χ^2_{null}/df_{null}) to the best fit (1).

A recommended value for TLI is 0.90 or greater. This measure may be valued less than zero or greater than one. The latter case means that the model is overfitting. The case of TLI less than zero is unlikely in practice. This measure is called also Non-Normed Fit Index.

4.2.3. Normed Fit Index (NFI), or Bentler and Bonett Index

The NFI, or Δ_1, was proposed by Bentler and Bonett and it is defined as:

$$NFI = \frac{F_{null} - F_{proposed}}{F_{null}} = \frac{\chi^2_{null} - \chi^2_{proposed}}{\chi^2_{null}} \quad [15]$$

ÍNDICE

F_{null} is the fitting function value of a null model and $F_{proposed}$ is the value of the fitting function for the maintained or proposed model. The alternative formula in χ^2's is sometimes computationally more convenient[8].

One way to obtain the null model is to specify a model such that $q=n$, $x=\xi$, $\Theta^{\delta}=0$, $\Lambda^{x}=I$ and Φ is a diagonal, free matrix. Of course, other baseline models may be more appropriate.

The NFI measure ranges from zero (no fit at all) to 1.0 (perfect fit). A limitation of Δ_1 is that it does not control for degrees of freedom. As a result, the NFI-value can be increased adding parameters. Thus a complex model may have a higher Δ_1 even though it has fewer degrees of freedom and may be "overfitting" the data.

The measure is also influenced by sample size. The mean of the sampling distribution of NFI is larger for bigger samples than it is for smaller ones. Thus comparing Δ_1 for two samples of different sizes can give the impression that the large sample has a better fit than the smaller one, even if the identical model holds for both samples.

A proposed modification on Δ_1 that lessens the dependence of its mean on N and takes account of $df_{proposed}$ is:

$$\Delta_2 = \frac{F_{null} - F_{proposed}}{F_{null} - \left[df_{proposed} / (N-1)\right]} = \frac{\chi^2_{null} - \chi^2_{proposed}}{\chi^2_{null} - df_{proposed}} \quad [16]$$

4.2.4. Other incremental fit measures

The CFI (Comparative Fit Index), IFI (Incremental Fit Index) and RFI (Relative Fit Index) measures range from 0 to 1, but values outside this interval can occur. Higher values indicate a better model. Let F be the minimum value of the fit function for the estimated or proposed model, let F_i be the minimum value of the fit function for the null or baseline model, and let d and d_i be the corresponding degrees of freedom. Furthermore, let $f = (N-1)F/d$, $f_i = (N-1)F_i/d_i$, $\varpi = \max\{((N-1)F-d), 0\}$, and $\varpi_i = \max\{((N-1)F_i-d_i), ((N-1)F-d), 0\}$. Then,

$CFI = 1 - \varpi/\varpi_i$ [17]

$IFI = \{(N-1)F_i - (N-1)F\}/\{(N-1)F_i - d\}$ [18]

$RFI = (f_i - f)/f_i$ [19]

4.3. Parsimonious fit measures

These measures relate the goodness-of-fit of the model to the number of estimated coefficients required to achieve this level of fit. The basic objective is to

diagnose whether model fit has been achieved by "overfitting" the data with too many coefficients.

4.3.1. Adjusted Goodness of Fit Index (AGFI)

The Goodness of Fit Index adjusted for the degrees of freedom, or the adjusted GFI, AGFI, is defined as:

$$AGFI = 1 - \frac{(p+q)(p+q+1)}{2d}(1 - GFI) \quad [20]$$

where d is the degrees of freedom of the model. This corresponds to using mean squares instead of total sums of squares in the numerator and denominator of 1-GFI. This measure ranges from zero (not fit at all) to 1.0 (perfect fit), though it is possible for it to be negative. The maximum is reached when $S = \hat{\Sigma}$.The recommended acceptance level is a value greater than or equal to 0.90.

4.3.2. Normed Chi-Square

This measure consists in adjusting the χ^2 by the degrees of freedom to assess model fit for various models.

This measure provides a way to assess inappropiate models in two ways: (1) a model that may be overfitted by capitalizing on chance, which is typified by values less than 1.0, and (2) models that are not truly representative of the observed data and thus need improvement, having values greater than an upper threshold (either 2.0 or 3.0, or the most liberal limit of 5.0). However, since the χ^2 value is the major component of this measure, it is subject to the sample size effects discussed earlier to the χ^2-measure itself.

4.3.3. Parsimonious Normed Fit Index (PNFI)

This measure is a modification of the NFI. The PFI takes into account the number of degrees of freedom used to achieve a level of fit. Parsimony means achieving higher measure relative to the number of degrees of freedom used. Thus more parsimony is desirable. The PNFI is defined as:

$$PNFI = [\text{degrees of freedom}_{proposed} / \text{degrees of freedom}_{null}] \times NFI \quad [21]$$

ÍNDICE

The principal use of PNFI is for the comparison of models with differing degrees of freedom. Higher values are better but there are no recommended levels of acceptable fit. Anyway, differences of 0.6 to 0.9 are proposed to be indicative of substantial model differences.

4.3.4. Akaike Information Criteria (AIC)

This is another comparative measure between models with differing numbers of constructs. The AIC is calculated as:

$$AIC = \tfrac{1}{2}(-\chi^2) - \text{Number of estimated coefficients} \quad [22]$$

This measure is always negative but values closer to zero indicate better fit and greater parsimony. A small AIC generally occurs when small χ^2 values are achieved with fewer estimated coefficients. This shows not only a good fit of observed versus predicted covariances/correlations, but a model not prone to "overfitting" as well.

There is no absolute value indicating an acceptable level of fit, but the researcher should choose the model with the smallest value.

4.3.5. Other parsimonious fit measures

All interesting parsimonious fit measures are the CAIC, and the single sample cross-validation index ECVI. All of these are simple functions of chi-square and the degrees of freedom:

$$CAIC = \chi^2 + 51 + \ln N)t \quad [23]$$
$$ECVI = (\chi^2/(N-1)) + 2(t/(N-1)) \quad [24]$$

where the mentioned parameters has been defined above. Although ECVI is quite similar to AIC, the rationale for ECVI is quite different from that of AIC and CAIC. Whereas AIC and CAIC are derived from statistical information theory, the ECVI is a measure of the discrepancy between the fitted covariance matrix in the analyzed sample and the expected covariance matrix that would be obtained in another sample of the same size.

To apply these measures to the decision problem, one estimates each model, ranks them according to one of these criteria and chooses the model with the smallest value. One can also take the precision of the estimated value of ECVI into account. For example, a 90% confidence interval for ECVI is:

$$\{(\hat{\lambda}_L + s + t)/(N-1);\ (\hat{\lambda}_U + s + t)/(N-1)\}, \quad [25]$$

where $s = \tfrac{1}{2}(p+q)(p+q+1)$ and the rest parameters as defined above.

For a given data set, N is the same for all models. Therefore, AIC and ECVI will give the same rank order of the models, whereas the rank ordering of AIC and CAIC can differ.

5. Application

To illustrate the application of the proposed assessment strategy, we shall assess a causal model with eight observed variables and three latent or nonobserved variables.

The observed measures in this example are: X_1 to X_6 (exogenous), and Y_1 and Y_2 (endogenous). The latent variables are: η_1 (endogenous), and ξ_1 and ξ_2 (exogenous).

The specification of the LISREL model is as follows:

1) Measurement model:

$X_1 = \lambda^x_{11}\,\xi_1 + \delta_1 \qquad X_2 = \lambda^x_{21}\,\xi_1 + \delta_2$

$X_3 = \lambda^x_{31}\,\xi_1 + \delta_3 \qquad X_4 = \lambda^x_{42}\,\xi_2 + \delta_4$

$X_5 = \lambda^x_{52}\,\xi_2 + \delta_5 \qquad X_6 = \lambda^x_{62}\,\xi_2 + \delta_6$

$Y_1 = \lambda^y_{11}\,\eta_1 + \varepsilon_1 \qquad Y_2 = \lambda^y_{21}\,\eta_1 + \varepsilon_2$

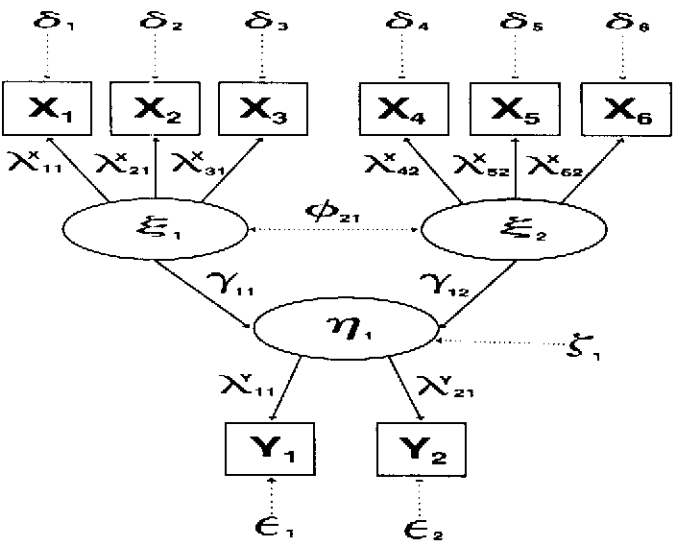
2) Structural model:

$\eta_1 = \gamma_{11}\,\xi_1 + \gamma_{12}\,\xi_2 + \zeta_1$

ϕ_{21}: covariance between ξ_1 and ξ_2.

The path diagram for the specified model is shown in figure 2.

Figure 2. Path diagram for the model



ÍNDICE

The analysis of the specified model was conducted by means of LISREL 8 software, using Maximum Likelihood (ML) estimation method. The estimation process provides estimates for all parameters (table 3).

Table 3. Parameter Estimates

STRUCTURAL MODEL			MEASUREMENT MODEL		
Parameter	Estimation	*t*-value	Parameter	Estimation	*t*-value
γ_{11}	0.49	2.10	λ^{y}_{11}	1.01	5.60
γ_{12}	-0.18	-0.80	λ^{y}_{21}	0.82	5.88
ϕ_{21}	0.63	5.08	λ^{x}_{11}	0.85	6.92
			λ^{x}_{21}	0.88	7.23
			λ^{x}_{31}	0.40	2.87
			λ^{x}_{42}	0.53	3.63
			λ^{x}_{52}	0.59	4.11
			λ^{x}_{62}	0.82	5.66
			$\theta^{\varepsilon}_{11}$	-0.02	-0.07
			$\theta^{\varepsilon}_{22}$	0.33	1.90
			θ^{δ}_{11}	0.28	2.60
			θ^{δ}_{22}	0.23	2.10
			θ^{δ}_{33}	0.84	5.05
			θ^{δ}_{44}	0.72	4.55
			θ^{δ}_{55}	0.65	4.24
			θ^{δ}_{66}	0.33	1.99

Prior Evaluation of the Model

1) Assessment of the assumptions of structural equation modeling

The eight variables must be assessed for their distributional characteristics, particularly normality and kurtosis, in order to use ML method. No variable was found to have significant departure from normality nor pronounced kurtosis.

2) Examination of the solution

A. *Offending estimates*

In table 3 we see that the loading for Y_1 is greater than 1.0. The corresponding error measurement value for the same variable is negative (-0.02). Such estimates are theoretically inappropriate and must be corrected. There is no right solution to this problem; if the error measurement is set to a small positive value (0.005), practical requirements of the estimation process will be met. The theoretical

justification of the model and the evident discriminant validity between latent variables must be examinated.

B. *Standard errors*

Under multivariate normality of the observed variables and using ML method, standard errors are correct.

C. *Squared multiple correlations and coefficients of determination*

The squared multiple correlation for the structural equation (R^2) is 0.16.

The squared multiple correlations for *y*-variables are following:

METHOD	Y_1	Y_2
ML	1.02	0.67

And the squared multiple correlations for *x*-variables are:

METHOD	X_1	X_2	X_3	X_4	X_5	X_6
ML	0.72	0.77	0.16	0.28	0.35	0.67

The squared multiple correlation for the structural equation is low. It means that η_1 is not well explain by the exogenous variables. Results are very high for *y*-variables, but we observed an anomalous value for Y_1. Again, there is an "offending estimate" for value related to Y_1. Error variance is negative and correlation exceeds 1.0. We face with a clear sympton of multicollinearity. Then we should review model especification for η_1 and consider elimination of observed variable Y_1. Y_1 and Y_2 measure the same concept. Linear relationships between *x*-observed variables and the latent variable are strong enough. The results are very good for almost all *x*-variables, except for X_3 and X_4, indicating that only a small proportion of their variance is accounted for by the latent exogenous variables (ξ_1 and ξ_2).

Overall Goodness-of-Fit Measures

1) Absolute fit measures

Next table contains all abolute fit measures:

MEASURE	ML
χ^2 (p) -df-	14.96 (0.60) -17-
Power of χ^2	90%
GFI	0.94
RMSR	0.056
Critical N (α=0.05)	121.60

ÍNDICE

Almost all measures are acceptables. The chi-square shows a very good p-level (0.60). The only measure that is slightly bad is the critical N. This is because this measure is abnormally low for small sample sizes (sample size=55).

2) Incremental fit measures

Next table shows the incremental fit measures:

MEASURE	ML
NFI (Bentler-Bonett)	0.92
NNFI	1.02
CFI	1.00
IFI	1.01
RFI	0.86

Results are very high, indicating that the proposed model is much better than the baseline or null model, that is, a simpler or more realistic model than which has been estimated.

3) Parsimonious fit measures

The parsimonious fit measures are shown in next table:

MEASURE	ML
AGFI	0.87
Normed χ^2	0.8
PNFI	0.56
AIC	52.96^9 (-17)
CAIC	110.10
ECVI	0.98
90% confidence interval for ECVI	(1.02; 1.22)

The normed chi-square has a marginal acceptability, because the model is overfitted.

Almost all measures of overall fit show acceptable values to support the hypothesized latent variables. The main measure (χ^2) is very low and the rest of measures provide good values. Then, the proposed model fits quite well the observed covariances. However, since the model has a data inconsistency (the «Heywood case» for Y_1), the researcher should review the model especification,

gather additional data and considerer elimination of Y_1 (the option of developing a «ridge» regression to solve multicollinearity is automatically determined by the software LISREL). This decision must carried out taking into account both statistical information and theory. The model must be revised a tested on a new covariance matrix.

Notes

1. Also, there are other types of assessment that should be carried aout to evaluate a structural equation model such the measurement model fit, the tructural model fit, the analysis of residuals, or the model modification indices. These types exceed the aims of this paper.

2. Without loss of generality, we can also assume that ξ and ε are centered variables.

3. In particular, the WLS method allows to perform an asymptotic distribution-free analysis.

4. A nested sequence of models, in the parameter-nested sense, is a sequence of similar models having the same parameters but ordered according to increasingly more restricted a priori constraints placed on their parameters. For example, two nested models with five parameters can be:

- b_1 b_2 b_3 b_4 b_5 (completely unrestricted model, i.e., all parameters are free), and

- b_1 b_2 b_3 b_4 0 (restricted model, one parameter, b_5, is restricted to 0)

5. The power of a test equals one minus the probability of a Type II error, that is, the probability of rejecting H_0 when it is incorrect, given that an alternative hypothesis, H_a, is true.

6. Other avalaible test statistics are the Lagrangian multiplier test (LM), and the Wald test (W). For further details see Bollen (1989: 289-303, and 338-349).

7. The familiar (central) chi-square distribution is the special case that arises when the *n* independent N (α, 1) variables have a mean of zero (i. e, α=0).

8. The fitting-function value (F) is computed from the next formula: $F=\frac{\chi^2}{N-1}$.

9. Value according to LISREL formula.

References

AISH, A. M. AND JÖRESKOG, K. G. (1990): «A Panel Model for Political efficacy and Responsiveness: An Application of LISREL 7 with Weighted Least Squares», *Quality & Quantity*, vol. 24: 405-426.

ARNOLD, S. F. (1981): *The Theory of Linear Models and Multivariate Analysis*, New York: John Wiley & Sons.

BAGOZZI, R. P. (1990): «Structural Equation Models in Marketing Research», AMA *Conference Proceedings of First Annual Advanced Research Techniques Forum*, Beaver Creek (CO): American Marketing Association, 335-379.

ÍNDICE

BEARDEN, W. O.; SHARMA, S. AND TEEL, J. E. (1982): "Sample Size Effects on Chi Square and Other Statistics Used in Evaluating Causal Models", *Journal of Marketing Research*, vol. XIX (November): 425-430.

BENTLER, P. M. Y CHOU, C. (1987): «Practical Issues in Structural Modeling», *Sociological Methods and Research*, vol. 16 (Agosto): 78-117.

BOLLEN, K. A. (1989): *Structural Equations with Latent Variables*, New York: John Wiley & Sons.

CUDECK, R. (1989): «Analysis of Correlation Matrices Using Covariance Structure Models», *Psychological Bulletin*, 105 (2): 317-327.

DHRYMES, *PHOEBUS J. (1974): Econometrics: Statistical Foundations and Applications*, Nueva York: Springer-Verlag.

DUNN, G.; EVERITT, B. AND PICKLES, A. (1993): *Modelling Covariances and Latent Variables Using EQS*, London: Chapman & Hall.

FOX, J. (1984): *Linear Statistical Models and Related Methods*, New York: John Wiley & Sons.

GOLDBERGER, A. S. (1991): *A Course in Econometrics*, Cambridge (MA): Harvard University Press.

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L., AND BLACK, W. C. (1992): *Multivariate Data Analysis*, 3rd ed., New York: Macmillan Publishing Company.

JAGPAL, H. S. (1982): «Multicollinearity in Structural Equation Models With Unobservable Variables», *Journal of Marketing Research*, vol. XIX (November): 431-439.

JÖRESKOG, K. G. (1988): «New Developments in LISREL: Analysis of Ordinal Variables using Polychoric Correlations and Weighted Least Squares», *International Conference on Social Science Methodology*, Dubrovnik (Yugoslavia).

JÖRESKOG, K. G., AND SÖRBOM, D. (1989): *LISREL 7 - A Guide to the Program and Applications*, 2nd ed., Chicago: SPSS Publications.

JÖRESKOG, K. G., AND SÖRBOM, D. (1993): *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*, Chicago: Scientific Software International, Inc.

JÖRESKOG, K. G., AND SÖRBOM, D. (1994): *LISREL 8 - User's Reference Guide*, Chicago: Scientific Software International, Inc.

MULAIK, S. A.; JAMES, L. R.; ALSTINE, J.; BENNET, N.; LIND, S., AND STILWELL, C. D. (1989): «Evaluation of Goodness-of-Fit Indices for Structural Equation Models», *Psychological Bulletin*, 105 (3): 430-445.

SÁNCHEZ, M., AND SWINNEN, G. (1995): "Structural Equation Modeling with Latent Variables. A Multivariate Advanced Method for Marketing Research", *Research Paper* BEDR/1995/03, Department of Business Economics, University of Limburgs.

SATORRA, A. (1990): «Robustness Issues in Structural Equation Modeling: A Review of Recent Developments», *Quality & Quantity*, vol. 24: 367-386.