DIAGNOSTICS OF THE ERROR FACTOR COVARIANCES

OTTÓ HAJDU[1]

In this paper we explore initial simple factor structure by the means of the so-called '*EPIC*' factor extraction method and the '*orthosim*' orthogonal rotational strategy. Then, the results are tested by confirmative factor analysis based on iteratively reweighted least squares on the one hand and asymptotically distribution free estimation on the other hand. Besides, based on multivariate kurtosis measures, multivariate normality is also investigated to see whether the use of the IWLS method is appropriate or a robust ADF estimator with relatively larger standard error is preferred. Finally, the paper draws attention that confidence intervals for the non-centrality based goodness of fit measures are available.

KEYWORDS: Latent variables; Covariance structural equations; Heteroscedasticity; Goodness of fit.

A model that relates measured variables to latent factors in covariance structure analysis is called a measurement model. These models are mostly factor analysis models and it is standard to distinguish between *confirmatory* and *exploratory* approach. In an exploratory factor analysis, we may not know how many factors are needed to explain the inter-correlations among the indicators. In addition, even if we are sure about the existence of a particular factor, we may not know which variables are the best indicators of the factor. Exploratory factor analysis will give us results: the number of factors, the factor loadings, and possibly the factor correlations.

In contrast, if we anticipate these results, we can do a confirmatory factor analysis. In this type of factor analysis, we presumably have a *hypothesis* about the number of factors, which measured variables are supposedly good indicators of each of the factors, which variables are unrelated to a factor and, how strongly or weakly the factors correlate to each other. In confirmatory models, variables are presumed to be factorially simple. That is, a given indicator is usually expected to be influenced by very few factors, typically only one. In addition, the covariance structure of the error factors can be arbitrary if it is reasonably justified and the model identification permits it. This means that the error (unique) factors are not necessarily uncorrelated but their variances may be equal by the homogeneity hypothesis.[2] Of course, hypothesis may be incorrect hence it must be tested

[1] Associate professor of the Budapest University of Technology and Economic Sciences.
[2] This error covariance structure is analogous to that used in the econometrics literature.

by sample information. Nevertheless, the hypothetical simple factor structure could be explored by orthogonal or oblique rotation of factor loadings carried out on an initial loading matrix produced by some factor extraction method.

The aim of this paper is twofold. In an explorative step we draw attention to a factor extraction method named EPIC (Equal Prior Instant Communalities) and an orthogonal rotation technique called *orthosim*. The EPIC method is used in our analysis as a compromise between two methods: the principal components method, which is computationally simple and the maximum likelihood factor analysis, which frequently leads to convergence problems. The 'orthosim' solution (proposed by *Bentler* [1977]) does not optimize some simplicity criterion within the loading matrix instead it maximizes a generalized variance type factor simplicity index corresponding to the loading matrix as a whole.

The core problem is that the initial unrotated EPIC factor solution (named by *Kaiser* [1990]) is based on a method in which the variances of the *uncorrelated* error factors are initially taken as *equal*, permitting the computations to be done explicitly and untroubled by linear dependencies among the variables (*Anderson* [1984] p. 21.). Nevertheless, the homogeneity assumption of the equal unique variances, as well as the factor pattern itself is merely a hypothesis. Therefore, it must be tested by using a confirmatory factor analysis step. There are two main approaches available to estimate the parameters of a confirmatory factor model. The first is based on some normality assumptions. However, if the normality assumptions are violated, asymptotically distribution free (ADF) approach must be used. This article gives a review of multivariate kurtosis measures to help decision whether the use of ADF method (with relatively larger standard errors) is necessary or not. In addition, the paper draws attention to those goodness of fit measures for which confidence intervals are available.

Finally, the paper illustrates the problems investigated based on microeconomic balance-sheet data. The computations are performed using the statistical programs 'Statistica 6.0' and 'EQS'.

THE ROLE OF UNCORRELATEDNESS IN THE FACTOR MODEL

In factor analysis, one assumes that certain observable variables (indicators) correlate because there are one or several underlying latent factors that generate the observed **x** data. The parametric form of the factor analysis model is given by

$$\mathbf{x}_{(p,1)} = \mathbf{\Lambda}_{(p,m)}\mathbf{f}_{(m,1)} + \mathbf{u}_{(p,1)} \qquad /1/$$

where vector $\mathbf{x}=[x_1, x_2,..., x_p]^T$ consists of p indicators, vector $\mathbf{f}=[f_1,f_2,...,f_m]^T$ consists of m common (latent) factors and $\mathbf{u}=[u_1, u_2,..., u_p]^T$ represents the *error* factors, *unique* to that indicator.[3] The so-called '*pattern matrix*' $\mathbf{\Lambda}$ of order (p, m) consists of λ_{jk} factor loadings. The higher the value of a loading in absolute magnitude the more important the

[3] The 'error factor' and 'unique factor' terminologies are used synonymously in this paper.

factor is. Using /1/ we can express the **C** covariance matrix of order (p,p) among the observed indicators based on covariances as follows:

$$\mathbf{C} = \mathbf{\Lambda C}_{ff}\mathbf{\Lambda}^T + \mathbf{C}_{uu} + \mathbf{\Lambda C}_{fu} + \mathbf{C}_{uf}\mathbf{\Lambda}^T . \qquad /2/$$

It is apparent, that **C** has $p(p+1)/2$ distinct elements (including the variances on the main diagonal as well) but the total number of the unknown parameters in /2/ is

$$pm + \frac{m(m+1)}{2} + \frac{p(p+1)}{2} + mp . \qquad /3/$$

The factor model is identified when the number of parameters to be estimated q is less than the number of the distinct observed covariances that is:

$$df = \frac{p(p+1)}{2} - q > 0 \qquad /4/$$

where *df* is the degree of freedom. Hence, it is necessary to reduce substantially the number of parameters to be estimated relative to the number of indicators. This can be achieved by imposing hypothetical restrictions on the parameters and by increasing the number of indicators.

Such straightforward assumption is that the unique factors are uncorrelated with the common factors, i.e. equation $\mathbf{C}_{fu} = \mathbf{C}_{uf} = \mathbf{0}$ holds in /2/. This restriction yields a decomposition of the observed covariance matrix in the following form:

$$\mathbf{C} = \mathbf{\Lambda C}_{ff}\mathbf{\Lambda}^T + \mathbf{C}_{uu} . \qquad /5/$$

A further reasonable restriction that can be imposed is that the unique factors are uncorrelated with each other as well. This means that the covariance matrix $\mathbf{C}_{uu}$ is diagonal. Based on this additional restriction the decomposition of the observed covariance matrix takes the form:

$$\mathbf{C} = \mathbf{\Lambda C}_{ff}\mathbf{\Lambda}^T + \mathbf{\Psi}^2 \qquad /6/$$

where $\mathbf{\Psi}^2$ is our standard notation for the *diagonal* covariance matrix of the unique factors. In addition, if the unique variances are homogeneous, i.e. all of them are equal to a constant σ^2, the covariance decomposition is as follows:

$$\mathbf{C} = \mathbf{\Lambda C}_{ff}\mathbf{\Lambda}^T + \sigma^2\mathbf{I} . \qquad /7/$$

Using now the conventional notation of $\mathbf{C}_{ff} = \mathbf{\Phi}$ and concerning orthogonal factors, $\mathbf{\Phi}$ is diagonal and, further, assuming standardized factors, $\mathbf{\Phi}$ equals the identity. A non-diagonal $\mathbf{\Phi}$ indicates ‘oblique’ (i.e. correlated) factors.

Explorative factor analysis is basically aimed at estimating the $\left(\mathbf{\Lambda},\mathbf{\Phi},\mathbf{\Psi}^2\right)$ parameters in /6/ without any presumed knowledge about them except, that the common factors are standardized (i.e. $\mathbf{\Phi}$ is a correlation matrix). In contrast, in a confirmative analysis interpretable parameters are selected to be estimated rather than accepting any computationally convenient assumptions. Our focus in this paper is mainly on testing the hypothesized structure of $\mathbf{\Psi}^2$.

Once a solution is obtained, with any $\mathbf{T}_m$ non-singular matrix of order m equation /1/ is still satisfied in the following form:

$$\mathbf{x} = \left(\mathbf{\Lambda}\mathbf{T}^{-1}\right)\left(\mathbf{T}\mathbf{f}\right) + \mathbf{u}\,. \qquad /8/$$

Replacing $\mathbf{f}$ by $\mathbf{f}^{*}=\mathbf{T}\mathbf{f}$ and $\mathbf{\Lambda}$ by $\mathbf{\Lambda}^{*} = \mathbf{\Lambda}\mathbf{T}^{-1}$ we perform an *oblique rotation* which results in the covariance matrix $\mathbf{T}\mathbf{\Phi}\mathbf{T}^{T} = \mathbf{\Phi}^{*}$ of the rotated factors and preserves the reproduced covariance matrix being unchanged:

$$\mathbf{\Lambda}^{*}\mathbf{\Phi}^{*}\mathbf{\Lambda}^{*T} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^{T}\,.$$

Specially, an *orthogonal rotation* is performed when the factors are uncorrelated and $\mathbf{T}$ is orthonormal: $\mathbf{T}^{T} = \mathbf{T}^{-1}$.

Our final goal is to give a pattern of loadings as clear as possible that is factors that are clearly marked by high loadings for some variables and low loadings for others. This general pattern is referred to as „simple structure'. This can be achieved by a two-step approach. First, in the explorative step we estimate the orthogonal loadings and subsequently rotate them. Then, in the confirmative step we fix some parameters at some (typically zero or equal) hypothetical value, reestimate the *free* parameters and test the goodness of fit. Specially, the adequacy (goodness of fit) of a specific orthogonal or oblique factor solution can directly be tested by confirmative factor analysis.

There are various rotational strategies that have been proposed in the field to explore a clear pattern of loadings. The most widely used orthogonal rotational strategy is the so-called varimax method (*Kaiser* [1958], *Ten Berge* [1995]). Despite the popularity of varimax, we shall use another method based on a different approach named *orthosim* (*Bentler* [1977]).

The so-called '*orthosim*' orthogonal rotation is based on a factorial *simplicity* index. Let us start with a known loading matrix $\mathbf{A}$ and transform it with an orthonormal $\mathbf{T}$ into $\mathbf{B} = \mathbf{A}\mathbf{T}$, such that $\mathbf{D} = diag\left(\left(\mathbf{B}*\mathbf{B}\right)^{T}\left(\mathbf{B}*\mathbf{B}\right)\right)$ is a diagonal matrix and * denotes the element-wise (Hadamard) product. Then we seek the rotated pattern matrix which *maximizes* the index of factorial simplicity defined as the *generalized variance* as follows:

$$GV = \det\left(\mathbf{D}^{-1/2}\left(\left(\mathbf{B}*\mathbf{B}\right)^{T}\left(\mathbf{B}*\mathbf{B}\right)\right)\mathbf{D}^{-1/2}\right) \rightarrow \max\,.$$

This determinant (based on a symmetric, nonnegative definite matrix with unit diagonal elements) ranges between zero and one. It equals zero when there are linear depend-

encies among the columns of $(\mathbf{B} * \mathbf{B})$. Such a case when some columns of **B** are proportional or identical except for sign. The maximum value of one occurs if the matrix in *GV* is the identity. This means that the factor pattern is factorially simple. It must be emphasized that provided a diagonal scaling matrix *GV* is invariant with respect to the column rescaling of **B**.

The concept of oblique rotations can be used in order to obtain more interpretable simple structure that best represents the 'clusters' of variables, without the constraint of orthogonal factors. One of the recommended widely used methods is the *direct quartimin method* (*Jennrich–Sampson* [1966]).

THE 'EQUAL PRIOR INSTANT COMMUNALITIES' FACTOR EXTRACTION METHOD

Under the hypothesis of this orthogonal factor model, the unique variances of the covariance (correlation) matrix are presumed to be equal. According to the standard orthogonality and the Kaiser-normalization requirements the matrix

$$\mathbf{\Lambda}^T \mathbf{\Psi}^{-2} \mathbf{\Lambda} = \mathbf{D} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m \end{bmatrix}$$

is diagonal and the maximum likelihood (ML) equations /9/ and /10/ must hold (*Lawley–Maxwell* [1971] p. 27. EQ 4.9; p. 30. EQ 4.19):

$$\mathbf{\Psi}^2 = diag(\mathbf{C} - \mathbf{\Lambda}\mathbf{\Lambda}^{T)}, \qquad /9/$$

$$\mathbf{C}\mathbf{\Psi}^{-2}\mathbf{\Lambda} = \mathbf{\Lambda}(\mathbf{I}_m + \mathbf{D}), \qquad /10/$$

where **C** is the covariance matrix of the observed indicators. Alternatively, equation /10/ can be written as:

$$(\mathbf{C} - \mathbf{\Psi}^2)\mathbf{\Psi}^{-2}\mathbf{\Lambda} = \mathbf{\Lambda}\mathbf{D}.$$

It is apparent that the columns of $\mathbf{\Lambda}$ are the eigenvectors corresponding to the largest *m* eigenvalues of matrices $\mathbf{C}\mathbf{\Psi}^{-2}$, or $(\mathbf{C} - \mathbf{\Psi}^2)\mathbf{\Psi}^{-2}$.

Let us suppose, that the uncorrelated unique factors are *homogeneous* i.e. $\mathbf{\Psi}^2 = \sigma^2 \mathbf{I}_p$, and consider the standard spectral decomposition of the covariance matrix **C** of the indicators. Taking only the first *m* eigenvalues on the main diagonal of the diagonal matrix $\mathbf{U}_m$ then:

$$\mathbf{C}\mathbf{W}_m = \mathbf{W}_m \mathbf{U}_m, \quad \mathbf{W}_m^T \mathbf{W}_m = \mathbf{I}_m, \qquad /11/$$

where the columns of $\mathbf{W}_m$ are the corresponding eigenvectors. After some manipulations we can write /11/ equivalently as:

$$\mathbf{C}\left(\sigma^2\mathbf{I}_p\right)^{-1}\left(\sigma\mathbf{W}_m\left(\frac{1}{\sigma^2}\mathbf{U}_m-\mathbf{I}_m\right)^{\frac{1}{2}}\right)=\sigma\mathbf{W}_m\left(\frac{1}{\sigma^2}\mathbf{U}_m-\mathbf{I}_m\right)^{\frac{1}{2}}\frac{1}{\sigma^2}\mathbf{U}_m\,.$$

Clearly, making the

$$\mathbf{\Lambda}=\sigma\mathbf{W}_m\left(\frac{1}{\sigma^2}\mathbf{U}_m-\mathbf{I}_m\right)^{\frac{1}{2}}, \qquad /12/$$

$$\mathbf{D}=\frac{1}{\sigma^2}\mathbf{U}_m-\mathbf{I}_m\,, \qquad /13/$$

$$\mathbf{\Psi}^2=\sigma^2\mathbf{I}_p \qquad /14/$$

substitutions – provided homogeneous unique factors – our all initial ML requirements are met.

The estimation of variance σ^2 happens in the following manner. Given that $tr\left(\mathbf{\Lambda}_m\mathbf{\Lambda}_m^T\right)$ is the sum of the *m* communalities.

$$p\sigma^2=\sum_{j=1}^{p}\mathrm{var}\left(x_j\right)-tr\left(\mathbf{L}_m\mathbf{L}_m^T\right)$$

and based on equations /12/, /13/ and $\mathbf{W}_m^T\mathbf{W}_m=\mathbf{I}_m$, we obtain

$$p\sigma^2=\sum_{j=1}^{p}u_j-tr\left(\mathbf{W}_m\sigma^2\mathbf{D}\mathbf{W}_m^T\right)=\sum_{j=1}^{p}u_j-tr\left(\left(\mathbf{U}_m-\sigma^2\right)\mathbf{W}_m^T\mathbf{W}_m\right)=$$

$$=\sum_{j=1}^{p}u_j-\left(\sum_{j=1}^{m}u_j-m\sigma^2\right),$$

from which

$$\left(p-m\right)\sigma^2=\sum_{j=m+1}^{p}u_j\,. \qquad /15/$$

We draw attention that $\mathbf{WU}^{1/2}$ gives the standard principal components loading matrix.

There are two reasons why the $\mathbf{\Psi}^2 = \sigma^2 \mathbf{I}_p$ assumption is not as restrictive as it seems: first, the unique variances for the covariance matrix are not used in the computations and are not presumed to be equal when an explorative factor extraction is carried out on the correlation matrix as it is the usual case. Rather, the ratio of common factor variance to unique variance is hypothesized as equal for all variables under the model. Second, the estimated communalities for the correlation matrix, obtained from the solution, can vary substantially in practice.

AN EXPLORATIVE STUDY BASED ON MICROECONOMIC FINANCIAL INDICATORS

Based on balance-sheet data of 2117 Hungarian economic units from the branch with NACE code '5011' in 1999, the following indicators have been investigated by EPIC factor analysis followed by orthosim and direct quartimin rotations:

Profit after taxation / Liabilities: 'ATPLIAB'
Cash-Flow / Liabilities: 'CFLIAB'
Current ratio = Current assets / Short term liabilities: 'CURRENT'
Adjusted Current ratio = (Current assets-Inventories) / Short term liabilities: 'ACURRENT'
Long term liabilities / (Long term liabilities + Owner's equity): 'DEBT'
Owner's equity / (Inventories + Invested assets): 'EQUITYR'

The cases with observed value smaller than –10 and those with larger than 10 are excluded from the analysis. The covariance and correlation matrices of these 6 variables are given in Table 1.

Table 1

Covariances and correlations of the financial microeconomic indicators (N=2117)

Variable	ATPLIAB	CFLIAB	CURRENT	ACURRENT	DEBT	EQUITYR
			Covariance matrix			
ATPLIAB	0.513	0.501	0.118	0.155	–0.086	0.193
CFLIAB	0.501	0.571	0.155	0.181	–0.110	0.223
CURRENT	0.118	0.155	0.837	0.842	–0.189	0.571
ACURRENT	0.155	0.181	0.842	1.566	–0.289	0.671
DEBT	–0.086	–0.110	–0.189	–0.289	0.596	–0.934
EQUITYR	0.193	0.223	0.571	0.671	–0.934	2.721
			Correlation matrix			
ATPLIAB	1.000	0.927	0.180	0.173	–0.155	0.163
CFLIAB	0.927	1.000	0.225	0.191	–0.188	0.179
CURRENT	0.180	0.225	1.000	0.735	–0.268	0.378
ACURRENT	0.173	0.191	0.735	1.000	–0.299	0.325
DEBT	–0.155	–0.188	–0.268	–0.299	1.000	–0.733
EQUITYR	0.163	0.179	0.378	0.325	–0.733	1.000

The eigenvalues of the *correlation* matrix constitute the main diagonal of the diagonal matrix **U**=< 2.709, 1.589, 1.101, 0.306, 0.222, 0.072 > where the first three largest roots account for a variance explained of 90 percentage. In order to extract 3 unrotated factors named F1, F2, F3, the EPIC factor model is used. The estimated variance of the unique factors (see equation /15/) is the average omitted eigenvalue:

$$\hat{\sigma}^2 = \frac{0.306+0.222+0.072}{3} = 0.2\,.$$

The EPIC loadings based on equation /12/ are shown in Table 2. They are computed from the $\mathbf{WU}^{1/2}$ principal components loadings (see also Table 2) according to the following manner:

$$\Lambda_{11}^{EPIC} = 0.601 = \hat{\sigma}\mathbf{W}_{11}\left(\frac{1}{\hat{\sigma}^2}\mathbf{U}_{11} - \mathbf{I}_{11}\right)^{\frac{1}{2}} = \hat{\sigma}\frac{\Lambda_{11}^{PCA}}{\sqrt{u_1}}\left(u_1\frac{1}{\hat{\sigma}^2} - 1\right)^{\frac{1}{2}} =$$

$$= \sqrt{0.2}\,\frac{0.6245}{\sqrt{2.709}}\sqrt{2.709\frac{1}{0.2} - 1}$$

and

$$\Lambda_{63}^{EPIC} = -0.4271 = \sqrt{\frac{0.2}{1.101}}\left(-0.4722\right)\sqrt{\frac{1.101}{0.2} - 1}\,.$$

Table 2

Initial, unrotated and rotated EPIC factor loadings

Variable	PCA factor loadings $\mathbf{WU}^{1/2}$			EPIC factor loadings			Orthosim solution			Direct quartimin solution		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
ATPLIAB	*0.624*	0.756	–0.036	*0.601*	0.707	–0.033	0.088	**0.921**	0.079	–0.013	***0.933***	0.008
CFLIAB	0.653	0.731	–0.029	0.629	0.684	–0.026	0.118	**0.916**	0.100	0.015	***0.924***	–0.009
CURRENT	0.705	–0.309	0.526	0.679	–0.289	0.475	**0.851**	0.107	0.185	**0.872**	0.012	–0.006
ACURRENT	0.688	–0.324	0.536	0.662	–0.303	0.485	**0.853**	0.085	0.175	**0.878**	–0.009	0.004
DEBT	–0.659	0.358	0.559	–0.634	0.335	0.505	–0.139	–0.088	**–0.862**	0.047	–0.008	***0.893***
EQUITYR	0.698	–0.391	*–0.472*	0.672	–0.366	*–0.427*	0.228	0.079	**0.842**	0.056	–0.008	***–0.854***

Since the first three eigenvalues of the correlation matrix account for a large portion of the total variance, it is clear, that the principal components and the EPIC loadings differ just to a slight extent. On the other hand, when some of the subsequent eigenvalues tend to be more important this tendency is not necessary.

The solutions from the orthogonal *orthosim* and oblique *direct quartimin* rotations are given in Table 2 and are almost identical.

According to the rotated loadings the following factors have been explored:

– F1: liability-based ‘profitability’,
– F2: current ratio-based ‘liquidity’,
– F3: long run ‘debtness’.

After oblique rotation the inter-factor correlations are not negligible because: Corr(F1,F2) = –0.233, Corr(F1,F3) = –0.408 and Corr(F2,F3) = –0.21, respectively. As a consequence, in the confirmative analysis step these correlations need to be estimated, increasing hence the number of free parameters.

The question arises at this stage is whether the hypothetical restriction $\sigma^2\mathbf{I}$ imposed on the covariance matrix of the error factors is acceptable or not. Decision on the model will be based on ‘goodness of fit’ measures, evaluated first retaining and then relaxing the restrictions. Such a method that provides tools for inference via the maximum likelihood theory is the *generalized weighted least squares*. However, when the sample does not come from a multivariate normal distribution, the asymptotically distribution free estimator is still available. A detailed overview of it is as follows.

ASYMPTOTICALLY DISTRIBUTION FREE ESTIMATORS

Based on a sample of size N let $\mathbf{S}$ denote the usual unbiased estimator of the population covariance matrix $\boldsymbol{\Sigma}_{(p,p)}$ whose elements are functions of a parameter vector $\boldsymbol{\theta}$:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) .$$

The weighted least squares (WLS) quadratic form discrepancy function measures the discrepancy between the sample covariance matrix $\mathbf{S}$ and the reproduced covariance matrix $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ evaluated at an estimator (*Browne* [1974]):

$$F\left(\mathbf{s},\boldsymbol{\sigma}\left(\boldsymbol{\theta}\right)\right) = \left(\mathbf{s}-\boldsymbol{\sigma}\left(\boldsymbol{\theta}\right)\right)^T \mathbf{W}^{-1}\left(\mathbf{s}-\boldsymbol{\sigma}\left(\boldsymbol{\theta}\right)\right) \to \min ,$$

where $\mathbf{s}$ and $\boldsymbol{\sigma}(\boldsymbol{\theta})$ are column vectors, formed from the $p^* = p\left(p+1\right)/2$ non-duplicative elements of $\mathbf{S}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, respectively and $\mathbf{W}$ is a positive definite weight matrix of order $\left(p^*, p^*\right)$. It is optimal to choose the weight matrix based on the covariance matrix of the sample covariances with typical element:

$$w_{jk,lt} = (N-1)\operatorname{cov}\left(s_{jk}, s_{lt}\right) = (N-1)\sigma_{jk,lt} = \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl} + \frac{N-1}{N}\kappa_{jklt} , \qquad /16/$$

where $\sigma_{jl} = \left[\boldsymbol{\Sigma}\right]_{jl}$ and

$$\kappa_{jklt} = \sigma_{jklt} - \left(\sigma_{jk}\sigma_{lt} + \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl}\right)$$

is a fourth-order cumulant with the fourth-order moment

$$\sigma_{jklt} = E\left(x_j - \mu_j\right)\left(x_k - \mu_k\right)\left(x_l - \mu_l\right)\left(x_t - \mu_t\right).$$

Equation /16/ gives the weight matrix for Browne's Asymptotically Distribution Free (ADF) estimator (*Browne* [1984]). Letting N tend to infinity the ADF weight takes the form without specifying any particular distribution:

$$w_{jk,lt} = \sigma_{jklt} - \sigma_{jk}\sigma_{lt} \qquad /17/$$

with consistent (but not unbiased) estimators

$$m_{jklt} = \frac{1}{N}\sum_{i=1}^{N}\left(x_j - \bar{x}_j\right)\left(x_k - \bar{x}_k\right)\left(x_l - \bar{x}_l\right)\left(x_t - \bar{x}_t\right),$$

$$m_{jk} = \frac{1}{N}\sum_{i=1}^{N}\left(x_j - \bar{x}_j\right)\left(x_k - \bar{x}_k\right).$$

Let us consider the heterogeneous kurtosis theory (*Kano–Berkane–Bentler* [1990]) which defines a general class of multivariate distributions that allows marginal distributions to have heterogeneous kurtosis parameters. Let $\kappa_j^2 = \sigma_{jjjj} / 3\sigma_{jj}^2$ represent a measure of excess kurtosis of the jth indicator. Then the fourth-order moments have the structure

$$\sigma_{jklt} = \frac{\kappa_j + \kappa_k}{2}\frac{\kappa_l + \kappa_t}{2}\sigma_{jk}\sigma_{lt} + \frac{\kappa_j + \kappa_l}{2}\frac{\kappa_k + \kappa_t}{2}\sigma_{jl}\sigma_{kt} + \frac{\kappa_j + \kappa_t}{2}\frac{\kappa_k + \kappa_l}{2}\sigma_{jt}\sigma_{kl}.$$

Under the assumption that all marginal distribution of a multivariate distribution are symmetric and have the same relative kurtosis, the elliptical (homogeneous kurtosis) theory estimators and test statistics can be obtained. The common kurtosis parameter of a distribution from the elliptical class of distributions with multivariate density[4]

$$c\,|\mathbf{V}|^{-\frac{1}{2}}\,h\left[\left(\mathbf{x}-\boldsymbol{\mu}\right)^T\mathbf{V}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right]$$

is

$$\kappa = \frac{\sigma_{jjjj}}{3\sigma_{jj}^2} - 1.$$

Then, the fourth-order moments are

$$\sigma_{jklt} = \left(\kappa + 1\right)\left(\sigma_{jk}\sigma_{lt} + \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl}\right).$$

[4] Here c is a constant, h is a non-negative function and $\mathbf{V}$ is a positive definite matrix.

Letting again N tend to infinity, substitution into /17/ yields the weight

$$w_{jk,lt} = \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl} + \kappa\left(\sigma_{jk}\sigma_{lt} + \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl}\right) = \left(\kappa+1\right)\left(\sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl}\right) + \kappa\sigma_{jk}\sigma_{lt}$$

Obviously, if $\kappa = 0$, then, multivariate normal distributions are considered and the typical element of the weight matrix takes the form

$$w_{jk,lt} = \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl} .$$

Because the size of **W** in practice can be very large it is reasonable to perform computations based on an equivalent form of the discrepancy function. Namely, assuming elliptical distributions, the quadratic form discrepancy function takes the form:

$$F_E = \frac{1}{2(\kappa+1)} tr\left[\left((\mathbf{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\mathbf{V}^{-1}\right)^2\right] - \frac{\kappa}{4(\kappa+1)^2 + 2p\kappa(\kappa+1)} tr^2\left((\mathbf{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\mathbf{V}^{-1}\right)$$

which reduces to the normal theory discrepancy function when $\kappa = 0$, i.e. the distributions have no kurtosis:

$$F_N = \frac{1}{2} tr\left[\left((\mathbf{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))\mathbf{V}^{-1}\right)^2\right]. \qquad /18/$$

When **V=I**, one obtains the unweighted least squares estimator F_{ULS}, the substitution **V=S** yields the generalized least squares estimator F_{GLS} and an iteratively reweighted solution F_{IWLS} is obtained when $\mathbf{V} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ is the reproduced covariance matrix generated by $\hat{\boldsymbol{\theta}}$ in each iterative step. Finally, asymptotically, F_{IWLS} leads to maximum likelihood estimate F_{ML} for exponential families of distributions.[5]

If $[\mathbf{V}]_{jk}$ is a consistent estimator of $\left[\boldsymbol{\Sigma}\right]_{jk} = \sigma_{jk}$ then $\hat{w}_{jk,lt}$ will be a consistent estimator of cov(**s,s**). Further, the *unbiased* estimator of $\hat{w}_{jk,lt}$ is

$$\hat{w}_{jk,lt} = \frac{N}{(N-2)(N-3)} \times$$

$$\times\left[(N-1)\left(m_{jklt} - m_{jk}m_{lt}\right) - \left(m_{jl}m_{kt} + m_{jt}m_{lk} - \frac{2}{N-1}m_{jk}m_{lt}\right)\right].$$

[5] The statistical distribution of the elements of a covariance matrix is not the same as that of a correlation matrix. This is obvious if you consider the diagonal elements of a covariance matrix, which are the variances of the variables. These are random variables – they vary from sample to sample. On the other hand, the diagonal elements of a correlation matrix are not random variables – they are always 1. The sampling distribution theory employed for the case of a covariance matrix is not applicable to a correlation matrix, except in special circumstances. It must be emphasized that it is possible (indeed likely) to get some incorrect results if we analyze a correlation matrix as if it were a covariance matrix. This has been described in the literature (see, for example, *Cudeck* [1989]). In order to analyse of the correlation matrix of the input data correctly, computations are based on the constrained estimation theory developed by *Browne* [1982]. As a result, we give the correct standard errors, estimates, and test statistics when a correlation matrix is analyzed directly.

If **W** consists of these unbiased elements, it may not be positive definite, but it would be unlikely the case when N is substantially larger than p*.

As it is apparent, the measures of multivariate kurtosis play a key role from the multivariate normality point of view.

THE MEASURES OF KURTOSIS

The statistics described subsequently allow us to examine whether the assumptions of *multivariate* normality have been violated. The consistent estimator of the common relative multivariate kurtosis parameter κ is the *rescaled* Mardia's sample measure:

$$\left(\hat{\kappa}+1\right)=\sum_{i=1}^{N}\frac{\left[\left(\mathbf{x}_i-\overline{\mathbf{x}}\right)^T\mathbf{S}^{-1}\left(\mathbf{x}_i-\overline{\mathbf{x}}\right)\right]^2}{Np\left(p+2\right)}.$$

This measure should be close to 1 if the distribution is multivariate normal.

If the sample comes from a multivariate normal distribution, the Mardia-coefficient of multivariate kurtosis defined as

$$MK=\hat{\kappa}p\left(p+2\right)$$

should be close to zero.[6] Further, the normalized multivariate kurtosis

$$\hat{\kappa}_0=\frac{MK}{\sqrt{8p\left(p+2\right)/N}}$$

has a distribution that is approximately standard normal at large samples.

The elliptical distribution family includes the multivariate normal distribution as a special case. As mentioned in this distribution family all variables have a common kurtosis parameter κ. This parameter can be used to rescale the *Chi*-square statistic if the assumption of an elliptical distribution is valid. The Mardia-based *kappa*

$$\hat{\kappa}_1=\frac{MK}{p\left(p+2\right)}$$

is an estimate of *kappa* obtained by rescaling the Mardia's coefficient of multivariate kurtosis. This number should be close to zero if the population distribution is multivariate normal.

Distribution theory provides a lower bound for *kappa*. It must never be less than –6/(p+2), where p is the number of variables. The adjusted mean scaled univariate kurto-

[6] The expected value and variance of $\left(\hat{\kappa}+1\right)p\left(p+2\right)$ are $\left(N-1\right)p\left(p+2\right)/\left(N+1\right)$ and $8p\left(p+2\right)/N$ respectively.

sis is an alternate estimate of *kappa*, which takes into account this requirement and is obtained simply as the average univariate kurtosis:

$$\hat{\kappa}_2 = \frac{1}{3p}\sum_{j=1}^{p} \max\left\{\left(\frac{m_{jjjj}}{m_{jj}^2} - 3\right), \frac{-6}{p+2}\right\},$$

where

$$\frac{m_{jjjj}}{m_{jj}^2} - 3$$

is the rescaled (i.e. uncorrected), biased estimate of *univariate* kurtosis for variable x_j. The asymptotic variance of this univariate measure is 24/*N*, which is used to standardize the uncorrected kurtosis in order to produce the '*normalized*' kurtosis.

The $\hat{\kappa}_2$ estimate averages the scaled univariate kurtosis, but adjusts each one that falls below the bound to be at the lower bound point. This coefficient should be close to zero if the distribution is multivariate normal.

Table 3

Measures of Multivariate Kurtosis

Measure	Value
Mardia Coefficient of Multivariate Kurtosis	403.085
Normalized Multivariate Kurtosis	946.437
Mardia-Based Kappa	8.398
Mean Scaled Univariate Kurtosis	9.580
Adjusted Mean Scaled Univariate Kurtosis	9.580
Relative Multivariate Kurtosis	9.398

Table 4

Univariate measures of skewness and kurtosis

Variable	Measures of skewness			Measures of kurtosis		
	Skewness	Corrected	Normalized	Kurtosis	Corrected	Normalized
ATPLIAB	0.490	0.490	9.201	46.798	46.912	439.524
CFLIAB	2.176	2.178	40.876	50.111	50.232	470.636
CURRENT	4.056	4.059	76.188	22.257	22.313	209.040
ACURRENT	3.311	3.314	62.203	13.932	13.968	130.850
DEBT	4.256	4.259	79.940	27.112	27.179	254.632
EQUITYR	–1.387	–1.388	–26.051	11.809	11.840	110.913

Considering our 6 financial microeconomic indicators, the computed values of the measures discussed are presented in Table 3 and Table 4. Results show that the requirement of zero kurtosis is violated. Nevertheless, the homogeneous kurtosis hypothesis

about a common non-zero kurtosis parameter could still be valid. But as we can see from Table 4 the univariate uncorrected kurtosis measures do not justify accepting the case of a common kurtosis parameter.

Finally, if the univariate kurtosis and skewness measures separately reject the assumption of univariate normality, the hypothesis about the multivariate normality must also be rejected as a consequence. Therefore, the *corrected univariate kurtosis* and *skewness* measures are also useful unbiased estimates for investigation of the assumption of normality.[7] They are, respectively:

Corrected univariate kurtosis

$$b^*_{2(j)} = \frac{(N+1)N^2}{(N-1)(N-2)(N-3)}\left[\frac{m_{jjjj}}{(m_{jj})^2}\right] - \frac{3(N-1)^2}{(N-2)(N-3)}.$$

Corrected univariate skewness

$$b^*_{1(j)} = \frac{N^2}{(N-1)(N-2)}\left[\frac{m_{jjj}}{(m_{jj})^{3/2}}\right].$$

The asymptotic variance of this latter measure is $6/N$, which is used to standardize the uncorrected skewness to produce the '*normalized*' skewness.

As a consequence of the kurtosis and skewness measures, we prefer the ADF estimator under arbitrary distribution as long as it produces interpretable results.

Estimates of free parameters and their inference statistics (standard error, T-value) are given in Table 5 based on both IWLS and ADF estimators considering both homogeneous and heterogeneous error-variance models. The corresponding converged values of the discrepancy function are also included. (Each of the four model-estimation converged within 10 iteration steps.) The type of the free parameters is indicated by the following scheme in the first column of the table: (.) contains latent variable, [.] includes measured indicator, the numbered -#-> arrow represents directed relationship and the numbered -#- wire represents undirected relationship (i.e. variance, covariance). Finally, the numbered name of an error factor is DELTA#.

As we can see, only the 'ADF, Homogeneous' T-values for parameters #11 and #13 are not significant with P-values 0.336, 0.454, respectively. All other P-values are practically zeros. Obviously, in the case of the ADF estimator (because of the distributional knowledge omitted) the estimated standard errors are higher than those computed by IWLS.

Based on the discrepancy function the results from the ADF method seem to be preferred. Contrary, based on the Root Mean Square (RMS) standardized residual[8], the IWLS results exhibit a better fit. The former results are based on the assumption of multivariate normality, while the latter is not, producing greater standard errors. Nevertheless our main purpose is to compare the homogeneous model with the heterogeneous one.

[7] One can find the uncorrected counterparts closed in the [.] bracket.
[8] Residual is divided by its standard error.

Table 5

Model characteristics from IWLS and ADF estimators

Free parameters	Iteratively reweighted least squares estimator						Asymptotically distribution free estimator					
	Heterogeneous variances			Homogeneous variances			Heterogeneous variances			Homogeneous variances		
	Estimate	St.Error	T	Estimate	St.Error	T	Estimate	St.Error	T	Estimate	St.Error	T
(F1)-1->[ATPLIAB]	0.664	0.012	56.444	0.634	0.014	44.706	0.465	0.060	7.773	0.569	0.064	8.885
(F1)-2->[CFLIAB]	0.756	0.012	65.054	0.673	0.015	46.140	0.530	0.066	8.066	0.585	0.066	8.849
(F2)-3->[CURRENT]	0.837	0.022	37.855	0.765	0.015	49.839	0.556	0.045	12.334	0.859	0.049	17.515
(F2)-4->[ACURRENT]	1.005	0.030	33.997	1.159	0.020	56.991	0.990	0.068	14.671	0.875	0.050	17.528
(F3)-5->[DEBT]	-0.590	0.019	-31.016	-0.600	0.013	-44.7	-0.763	0.044	-17.28	-0.298	0.020	-15.27
(F3)-6->[EQUITYR]	1.581	0.043	36.631	1.589	0.026	60.476	1.203	0.072	16.782	1.297	0.071	18.362
(DELTA1)-7-(DELTA1)	0.072	0.002	32.527	***0.189***	0.003	56.338	0.031	0.010	**3.191**	***0.018***	0.005	3.667
(DELTA2)-8-(DELTA2)	0.000	0.000	-	***0.189***	0.003	56.338	0.000	0.000	-	***0.018***	0.005	3.667
(DELTA3)-9-(DELTA3)	0.137	0.027	5.013	***0.189***	0.003	56.338	0.201	0.033	6.031	***0.018***	0.005	3.667
(DELTA4)-10-(DELTA4)	0.555	0.042	13.081	***0.189***	0.003	56.338	0.273	0.106	**2.579**	***0.018***	0.005	3.667
(DELTA5)-11-(DELTA5)	0.248	0.017	14.668	***0.189***	0.003	56.338	0.030	0.031	**0.962**	***0.018***	0.005	3.667
(DELTA6)-12-(DELTA6)	0.220	0.108	2.028	***0.189***	0.003	56.338	0.878	0.128	6.882	***0.018***	0.005	3.667
(F2)-13-(F1)	0.244	0.022	11.028	0.238	0.024	9.927	0.047	0.063	**0.749**	0.212	0.072	2.944
(F3)-14-(F1)	0.193	0.022	8.778	0.207	0.024	8.651	0.220	0.035	6.300	0.250	0.053	4.701
(F3)-15-(F2)	0.427	0.022	19.541	0.399	0.020	19.842	0.321	0.026	12.540	0.431	0.034	12.807
Discrepancy Function	0.0441			0.973			0.0301			0.0897		
degree of freedom	6			11			6			11		
RMS Stand. Residual	0.0158			0.0625			0.243			0.285		
Chi-Square Statistic	93.3284			2057.94			63.7372			189.81		
Goodness of fit indices	*Confidence intervals at 90 percent level*											
Noncentrality based indices	LB	PE	UB	LB	PE	UB	LB	PE	UB	LB	PE	UB
Population Noncentrality Index	0.026	0.038	0.055	0.500	0.551	0.606	0.017	0.027	0.041	0.065	0.085	0.107
Steiger-Lind RMSEA Index	0.066	0.080	0.095	0.213	0.224	0.235	0.053	0.067	0.083	0.077	0.088	0.099
McDonald Noncentrality Index	0.973	0.981	0.987	0.738	0.759	0.779	0.980	0.986	0.992	0.948	0.959	0.968
Population Gamma Index	0.982	0.987	0.991	0.832	0.845	0.857						
Adjusted Population Gamma Index	0.938	0.956	0.970	0.679	0.704	0.727						
Other fit indices												
Joreskog GFI		0.986			0.844			0.832			0.499	
Joreskog AGFI		0.952			0.701			0.412			0.044	
Akaike Information Criterion		0.058			0.982			0.044			0.099	
Schwarz's Bayesian Criterion		0.098			1.009			0.084			0.126	
Browne-Cudeck Cross Validation		0.058			0.982			0.044			0.099	
Null Model Chi-Square		7989.2			7989.2			291.8			291.8	
Null Model *df*		15			15			15			15	
Bentler-Bonett Normed Fit Index		0.988			0.742							
Bentler-Bonett Non-Normed Fit Index		0.973			0.650							
Bentler Comparative Fit Index		0.989			0.743							
James-Mulaik-Brett Parsimonious Fit Index		0.395			0.544							
Bollen's Rho		0.971			0.649							
Bollen's Delta		0.989			0.743							

Note: Where a baseline model is involved, it is assumed to be the null model, defined as a model without any common factors.

var(F1)=var(F2)=var(F3)=1 and the error factors (Delta1-Delta6) are uncorrelated.

Once a minimized converged value of the discrepancy function has been reached and selected as the best one, the subsequent evaluation of its goodness of fit is necessary. For this purpose a wide selection of fit-indices is available. Some of them are hypothesis theory-based, others are heuristic. On the other hand, we can distinguish noncentrality-based goodness of fit indices and other indices including incremental type indices as well. In the following section we discuss those employed in this paper.

NONCENTRALITY-BASED GOODNESS-OF-FIT INDICES

Let us consider the null hypothesis that the restricted model $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ holds for the population covariance matrix $\boldsymbol{\Sigma}$, against the alternative that it does not hold:

$$H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}),$$

$$H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}(\boldsymbol{\theta}).$$

In other words, the H_1 hypothesis states that a significant improvement is expected in the discrepancy between the restricted and the unrestricted models due to a simple switch from $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ to $\boldsymbol{\Sigma}$. Then, the discrepancy between the *true* and the hypothesized model is

$$F(\boldsymbol{\sigma}, \boldsymbol{\sigma}(\boldsymbol{\theta})) = (\boldsymbol{\sigma} - \boldsymbol{\sigma}(\boldsymbol{\theta}))^T \mathbf{W}^{-1} (\boldsymbol{\sigma} - \boldsymbol{\sigma}(\boldsymbol{\theta})) \rightarrow \min$$

which could be minimized with respect to the parameter vector $\boldsymbol{\theta}$. Let $F(\boldsymbol{\sigma}, \boldsymbol{\sigma}(\boldsymbol{\theta}^*))$ denote the minimized value at some $\boldsymbol{\theta}^*$. Then, asymptotically, $\chi^2 = (N-1) F(\mathbf{s}, \boldsymbol{\sigma}(\boldsymbol{\theta}))$ is distributed as a *noncentral* Chi-square with

$$df = \frac{p(p+1)}{2} - q$$

degrees of freedom and *noncentrality parameter*

$$\tau = (N-1) F(\boldsymbol{\sigma}, \boldsymbol{\sigma}(\boldsymbol{\theta}^*))$$

or

$$\frac{\tau}{N-1} = F(\boldsymbol{\sigma}, \boldsymbol{\sigma}(\boldsymbol{\theta}^*))$$

rescaled noncentrality parameter, where q is the number of parameters to be estimated for the model. Obviously, when the model holds, $\tau = 0$ and χ^2 is distributed as a *central* Chi-square with *df* degrees of freedom.

Hence, the size of τ can be considered as a *population* measure of model *misspecification*, with larger values of τ indicating greater misspecification. As it follows from the probability theory, the expected value of the noncentral Chi-square statistic is

$$E\left(\chi^2\right) = E\left\{(N-1)F\left(\mathbf{s},\boldsymbol{\sigma}\left(\boldsymbol{\theta}\right)\right)\right\} = df + \tau\,.$$

Hence, based on only one observation for χ^2, the estimated value of the noncentrality parameter is

$$\hat{\tau} = NCP = \chi^2 - df\ ,$$

where

$$\chi^2 = (N-1)F\left(\mathbf{s},\boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}\right)\right)$$

is the estimated measure of distance between the *currently* investigated model which is the *target* of our hypothesis and the saturated model with $p(p+1)/2$ free parameters, say, **s**. Therefore, the discrepancy function (named also fitting function) is calculated as

$$F = \frac{\chi^2}{N-1}\,.$$

Note, that *NCP* can be negative when the estimated Chi-square is less than the *df*. Dividing the noncentrality parameter by (N–1) yields the population noncentrality index *PNI* which is a measure of population *badness-of-fit* and depends only on the model, and the method of estimation:

$$PNI = \max\left\{\frac{\chi^2 - df}{N-1}, 0\right\}.$$

The population noncentrality index *PNI* is an unbiased estimate of the rescaled noncentrality parameter and is relatively unaffected by the sample size. However, *PNI* fails to compensate for model complexity. In general, for a given **S**, the more complex the model the better its fit. A method for assessing population fit which fails to compensate for this will inevitably lead to choosing the most complex models, even when simpler models fit the data nearly as well. Because *PNI* fails to compensate for the size or complexity of a model, it has limited utility as a device for comparing models.

The adjusted root mean square error index, first proposed by *Steiger* and *Lind* [1980], takes a relatively simplistic approach to solving these problems. Since model

complexity is reflected directly in the number of free parameters, and inversely in the number of degrees of freedom, the *PNI* is divided by degrees of freedom, then the square root is taken to return the index to the same metric as the original standardized parameters.

$$RMSEA = \sqrt{\frac{PNI}{df}} .$$

The RMSEA index can be thought of roughly as a root mean square standardized residual. Values above .10 indicate an inadequate fit, values below .05 a very good fit. Point estimates below .01 indicate an outstanding fit. The rule of thumb is that, for 'close fit', RMSEA should be less than $c = .05$ yields a rule that

$$\frac{\chi^2}{df} < 1 + (N-1)c^2 = 1 + \frac{N-1}{400} .$$

With this criterion, if $N = 401$, the ratio of the *Chi-square* to its degrees of freedom should be less than 2. Note that this rule implies a *less stringent* criterion for the ratio χ^2/df as sample size increases.

Rules of thumb that cite a single value for a critical ratio of χ^2/df ignore the point that the *Chi-square* statistic has an expected value that is a function of degrees of freedom, population badness of fit, and *N*. Hence, for a fixed level of population badness of fit, the expected value of the *Chi-square* statistic will increase as sample size increases.

McDonald [1989] proposed an index of noncentrality that represents one approach to transforming the population noncentrality index *PNI* into the range from 0 to 1. The index does not compensate for model parsimony, and the rationale for the exponential transformation it uses is primarily pragmatic. The index may be expressed as

$$MDNI = e^{-\frac{1}{2}PNI} .$$

Good fit is indicated by values above 0.95. Similarly, the *scaled likelihood ratio criterion* is

$$LHR = e^{-\frac{1}{2}F} .$$

Further, the weighted population coefficient of determination can also be defined as

$$\Gamma = 1 - \frac{\left(\boldsymbol{\sigma} - \boldsymbol{\sigma}(\boldsymbol{\theta})\right)^T \mathbf{W}^{-1} \left(\boldsymbol{\sigma} - \boldsymbol{\sigma}(\boldsymbol{\theta})\right)}{\boldsymbol{\sigma}^T \mathbf{W}^{-1} \boldsymbol{\sigma}} ,$$

where **W** is a positive definite weight matrix. Under arbitrary weighted least squares estimation, the *population gamma index* of *Tanaka* and *Huba* [1985] is given as a general

form for the *sample* fit index for covariance structure models. It assumes the covariance structure model has been fit by minimizing the WLS discrepancy function. Then, the index is

$$\gamma = 1 - \frac{\left(\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\right)^T \mathbf{W}^{-1} \left(\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\right)}{\mathbf{s}^T \mathbf{W}^{-1} \mathbf{s}}.$$

When the distributions have no kurtosis ($\kappa = 0$) based on /18/ we can write γ as the parametric form of the *Jöreskog–Sörbom* [1984] index of fit:

$$JSI_V = 1 - \frac{\frac{1}{2} tr\left(\left(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\right)\mathbf{V}^{-1}\right)^2}{\frac{1}{2} tr\left(\mathbf{S}\mathbf{V}^{-1}\right)^2}.$$

If **V=I**, or **V=S**, one obtains the Jöreskog–Sörbom (JS) index for the *ULS* and *GLS* estimators, respectively. Specially, using *IWLS*, i.e. $\mathbf{V} = \hat{\boldsymbol{\Sigma}}$, gives asymptotically the *JSI* index for the maximum likelihood (ML) estimation with

$$F_{IWLS} = \frac{1}{2} tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1} - \mathbf{I}\right)^2 = \frac{1}{2}\left(tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right)^2 - 2tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right) + p \right).$$

Hence,

$$\gamma_{IWLS} = 1 - \frac{tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1} - \mathbf{I}\right)^2}{tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right)^2} = \frac{2tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right) - p}{2F_{IWLS} + 2tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right) - p}.$$

In addition, when the

$$tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right) = p$$

equation holds, the $\gamma_{IWLS,ML}$ index reduces to the classic JS goodness of fit index:

$$GFI = \frac{p}{tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right)^2} = \frac{p}{2F_{IWLS} + p}.$$

As a consequence, *GFI* can be thought of as the sample equivalent of the index defined in the population as

$$\Gamma_1 = \frac{p}{tr\left(\boldsymbol{\Sigma}\left[\boldsymbol{\Sigma}(\boldsymbol{\theta})\right]^{-1}\right)^2} = \frac{p}{2F\left(\boldsymbol{\sigma}, \boldsymbol{\sigma}(\boldsymbol{\theta})\right) + p} = \frac{p}{2\frac{\tau}{N-1} + p}.$$

Any consistent estimate of τ will give a consistent estimate for Γ_1. This index like *PNI*, fails to compensate for the effect of model complexity. Consider a sequence of *nested* models, where the models with more degrees of freedom are special cases of those with less degrees of freedom. For such a nested sequence of models, the more complex models (i.e. those with more free parameters and less degrees of freedom) will always have Γ_1 coefficients as low or lower than those which are less complex.

The adjusted population *gamma* index Γ_2 attempts to compensate for this tendency:

$$\Gamma_2 = 1 - \frac{p(p+1)}{2 \cdot df}(1 - \Gamma_1)$$

and its sample counterpart is

$$AGFI = 1 - \frac{p(p+1)}{2 \cdot df}(1 - GFI) .$$

Values of the Joreskog GFI above .95 indicate good fit. This index is a *negatively biased* estimate of the population GFI, so it tends to produce a slightly pessimistic view of the quality of population fit. We give this index primarily because of its historical popularity.

The Population Gamma index is a superior realization of the same rationale. The values of the Joreskog AGFI above .95 also indicate good fit. This index is, like the GFI, a negatively biased estimate of its population equivalent. As with the GFI, the Adjusted Population Gamma Index is a superior realization of the same rationale.

At this stage we have arrived at an important conclusion that the lower and upper bounds of an α level confidence interval of the Chi-square statistic can be inserted into any goodness of fit measure that involves the Chi-square statistic. Consistent estimates and confidence intervals for Γ_1 may thus be converted into corresponding quantities for Γ_2.

OTHER INDICES OF FIT

Rescaled Akaike Information Criterion

In a number of situations the user must decide among a number of competing *nested* models of different dimensions. This criterion is useful primarily for deciding which of several nested models provides the best approximation to the data. The most typical example is the choice of the number of factors in common factor analysis. *Akaike* ([1973], [1974], [1983]) proposed a criterion for selecting the dimension of a model. *Steiger* and *Lind* [1980] presented an extensive Monte Carlo study of the performance of the Akaike criterion. Here the criterion is rescaled (without affecting the decisions it indicates) so that it remained more stable across differing sample sizes. The rescaled Akaike criterion (modified by *Cudeck* and *Brown* [1983]) is as follows.

Let $F_{ML,k}$ be the maximum likelihood discrepancy function and q_k be the number of free parameters for the model M_k. Let N be the sample size.

When trying to decide between several nested models, choose the one with the smallest Akaike criterion:

$$AC_k = F_{ML,k} + \frac{2q_k}{N-1}.$$

Schwarz's Bayesian Criterion

This criterion (*Schwarz* [1978] also modified by *Cudeck* and *Brown* [1983]) is similar in use to Akaike's index, selecting, in a sequence of nested models, the model for which

$$SC_k = F_{ML,k} + \frac{q_k \ln(N)}{N-1}$$

is a minimum.

Browne–Cudeck Single Sample Cross-Validation Index

Browne and *Cudeck* [1990] proposed a single sample cross-validation index as a follow-up to their earlier (*Cudeck–Browne* [1983]) paper on cross-validation. Cudeck and Browne had proposed a cross-validation index which, for model M_t in a set of competing models is of the form $F_{ML}(\mathbf{S}_{cv},\boldsymbol{\Sigma}_t(\boldsymbol{\theta}))$. In this case, F is the maximum likelihood discrepancy function, $\mathbf{S}_{cv}$ is the covariance matrix calculated on a cross-validation sample, and $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ the reproduced covariance matrix obtained by fitting model M_t to the original calibration sample. In general, better models will have smaller cross-validation indices.

The drawback of the original procedure is that it requires two samples, i.e. the calibration sample for fitting the models, and the cross-validation sample. The new measure estimates the original cross-validation index from a single sample.

The measure is

$$C_t = F_{ML}(\mathbf{S}_{cv}, \boldsymbol{\Sigma}_t(\boldsymbol{\theta})) + \frac{2q_t}{N-p-2}.$$

Null Model Chi-square and df

This is the *Chi-square* goodness-of-fit statistic, and the associated degrees of freedom, for the hypothesis that the population covariances are all zero. Under the assumption of multivariate normality, this hypothesis can only be true if the variables are all independent. The 'Independence Model' is used as the 'Null Model' in several comparative fit indices.

Bentler–Bonett Type Fit Indices

One of the most historically important and original fit indices, the *Bentler–Bonett* [1980] *normed index* measures the *relative decrease* in the discrepancy function caused by switching from a 'Baseline Model' (typically the null model) to a more complex model. It is defined as:

$$0 \le NFI_{t/b} = \frac{F_b - F_t}{F_b} = \frac{\chi_b^2 - \chi_t^2}{\chi_b^2} \le 1,$$

where F_b is the discrepancy function for the 'baseline model', F_t is the discrepancy function for the *target* (typically the current) model. This index approaches 1 in value as fit becomes perfect. However, it does not compensate for model parsimony.

The comparative *Bentler–Bonett* [1980] non-normed fit index takes into account model parsimony. It is defined as

$$NNFI_{t/b} = \frac{\chi_b^2 - \frac{df_b}{df_t}\chi_t^2}{\chi_b^2 - df_b} = \frac{\frac{\chi_b^2}{df_b} - \frac{\chi_t^2}{df_t}}{\frac{\chi_b^2}{df_b} - 1} = \frac{\frac{F_b}{df_b} - \frac{F_t}{df_t}}{\frac{F_b}{df_b} - \frac{1}{N-1}}$$

or it can be written as

$$NNFI_{t/b} = \frac{\chi_b^2 - df_b - \left(\frac{df_b}{df_t}\chi_t^2 - df_b\right)}{\chi_b^2 - df_b} = 1 - \frac{df_b}{df_t}\frac{NCP_t}{NCP_b} = 1 - p_{b/t}\frac{PNI_t}{PNI_b},$$

where $p_{b/t}$ is the so-called parsimony coefficient.

Bentler Comparative Fit Index

The comparative index (*Bentler* [1990]) estimates the relative decrease in the population noncentrality obtained by changing from the 'Baseline Model' to the *t* model. The index may be computed as:

$$BCFI_{t/b} = 1 - \frac{NCP_t}{NCP_b},$$

where NCP_t is the estimated non-centrality parameter for the target model and NCP_b is that for the base line model.

James–Mulaik–Brett Parsimonious Fit Index

This index was one of the earliest (along with the Steiger–Lind index) to compensate for model parsimony. Basically, it operates by rescaling the Bentler–Bonett Normed fit index to compensate for model parsimony. The formula for the index is:

$$PI = \frac{df_t}{df_b} NFI_t,$$

where *NFI* denotes the Bentler–Bonett normed fit index.

Bollen's Rho

This comparative fit index computes the relative reduction in the discrepancy function per degree of freedom when moving from the 'Baseline Model' to the *t* model. It is computed as

$$\rho_{t/b} = \frac{\frac{F_b}{df_b} - \frac{F_t}{df_t}}{\frac{F_b}{df_b}} = 1 - \frac{df_b}{df_t}\frac{F_t}{F_b}.$$

Comparing with *NNFI*, we see that, for even moderate *N,* there is virtually no difference between Bollen's *Rho* and the Bentler–Bonett Non-normed fit index.

Bollen's Delta

This index is also similar in form to the Bentler–Bonett index, but rewards simpler models (those with higher degrees of freedom). It is computed as:

$$\Delta_{t/b} = \frac{F_b - F_t}{F_b - \frac{df_t}{N-1}}.$$

EVALUATION OF MODEL FIT

Based on the results given in Table 5 and discussion of the goodness of fit measures presented earlier the following statements can be established:

Both the IWLS and the ADF estimates exhibit an outstanding goodness of fit. The model with more parameters, of course, performs a better fit. Except the *population gamma* and *Jöreskog–Sörbom* indices, the ADF estimator seems to be preferred against the IWLS estimator.

As a brief summary measure the pseudo R-square defined as

$$R^2 = 1 - \frac{\chi^2}{\chi^2_{null\text{-}model}}$$

are 98.83 and 74.24 percent for the IWLS homogeneous and heterogeneous models, respectively. Hence, switching from one model to the other seems to cause a considerable difference. These measures for the ADF estimators are 78.16 and 34.95 percent, respectively.

The null-model goodness of fit Chi-square value (the distance from the saturated model with $p(p+1)/2$ parameters) is substantially smaller in the ADF case. This null model Chi-square means the moving range given for the Chi-square to get closer to the saturated model (to the sample points). The corresponding R-square values thus must be interpreted on this shorter range of improvement for the ADF case. The null model chi-square estimated by the ADF method is 291.781, maybe underestimated to a great extent. Therefore additional goodness of fit measures based on this distance are not published in Table 5.

Only the James-Mulaik-Brett Parsimonious Index prefers the improvement in the degree of freedom versus worsen in the discrepancy function.

Considering any model of our interest, because of the large sample size the model chi-square statistic is relatively large as compared with the small degree of freedom resulted in from the (6,6) order of the sample covariance matrix. As a consequence, in spite of the goodness of fit measures, the chi-square test suggests to reject each of our models at any significance level.

Even a moderately large sample size is given, as it is the present case, it is not possible to choose between the competing homogeneous and heterogeneous models based on chi-square-difference test statistic. Namely, the difference between these chi-square statistics 2057.94–93.33=1964.61 and 189.81–63.74=126.07 are still significant at (11–6=5) degrees of freedom no matter whether the IWLS or the ADF results are considered.

Nevertheless, hypothesis testing can be avoided if we use some so-called incremental goodness of fit index such as the Bentler-type indices. Normed indices that fall into the interval of (0,1) are preferred because of their easy interpretation. In our investigation the Bentler-type incremental indices are as follows

$$NFI_{\text{homogeneous/heterogeneous}} \mid IWLS = \frac{\chi_o^2 - \chi_e^2}{\chi_o^2} = \frac{2057.94 - 93.3284}{2057.94} = 0.9546\,,$$

$$NNFI_{\text{homogeneous/heterogeneous}} \mid IWLS = \frac{\frac{\chi_o^2}{df_o} - \frac{\chi_e^2}{df_e}}{\frac{\chi_o^2}{df_o} - 1} = \frac{\frac{2057.94}{11} - \frac{93.3284}{6}}{\frac{2057.94}{11} - 1} = 0.9218\,,$$

$$BCFI_{\text{homogeneous/heterogeneous}} \mid IWLS = 1 - \frac{NCP_o}{NCP_e} = 1 - \frac{93.3284 - 6}{2057.94 - 11} = 0.9573\,,$$

and in the case of the ADF estimation they take the values as follows

$$NFI_{\text{homogeneous/heterogeneous}} \mid ADF = \frac{\chi_o^2 - \chi_e^2}{\chi_o^2} = \frac{189.81 - 63.7372}{189.81} = 0.6642,$$

$$NNFI_{\text{homogeneous/heterogeneous}} \mid ADF = \frac{\frac{\chi_o^2}{df_o} - \frac{\chi_e^2}{df_e}}{\frac{\chi_o^2}{df_o} - 1} = \frac{\frac{189.81}{11} - \frac{63.7372}{6}}{\frac{189.81}{11} - 1} = 0.408,$$

$$BCFI_{\text{homogeneous/heterogeneous}} \mid ADF = 1 - \frac{NCP_o}{NCP_e} = 1 - \frac{63.7372 - 6}{189.81 - 11} = 0.6771.$$

The smaller the value of an incremental index, the closer the models of interest are to one another. Hence, the results from IWLS may suggest that the assumption of the homogeneous variances is acceptable but, in contrary, based on the ADF method we may conclude that the error factor variances are heterogeneous. Recall here that the use of IWLS is questionable because of rejecting the normality and the zero kurtosis assumption.

Finally, the question of our interest is whether the magnitude of improvement in the discrepancy function due to involving *correlated error* (unique) factors is significant or not. Enabling cov(DELTA3, DELTA5) to be freely estimated provided *heterogeneous* error variances by the ADF method, the discrepancy function reduces to 0.0105 and the Chi-Square Statistic becomes 22.3062 with 5 degrees of freedom and P-value of 0.000458. Parameter estimates are given in Table 6. As compared with the corresponding Chi-Square Statistic 63.7372, the difference is (63.7372–22.3062)=41.431 with 1 degree of freedom which is significant at any level.

Table 6

Parameter estimation by ADF with heterogeneous but correlated error factors

Free parameters	Estimation	Standard error	T-value	Prob.
(F1)-1->[ATPLIAB]	0.635	0.058	10.894	0.000
(F1)-2->[CFLIAB]	0.668	0.061	11.001	0.000
(F2)-3->[CURRENT]	0.818	0.052	15.869	0.000
(F2)-4->[ACURRENT]	0.979	0.050	19.647	0.000
(F3)-5->[DEBT]	–0.633	0.040	–15.822	0.000
(F3)-6->[EQUITYR]	1.463	0.077	18.951	0.000
(DELTA1)-7-(DELTA1)	0.018	0.009	1.975	0.048
(DELTA2)-8-(DELTA2)	–0.000	0.000		
(DELTA3)-9-(DELTA3)	0.123	0.028	4.373	0.000
(DELTA4)-10-(DELTA4)	0.557	0.098	5.687	0.000
(DELTA5)-11-(DELTA5)	0.191	0.032	6.008	0.000
(DELTA6)-12-(DELTA6)	0.567	0.141	4.027	0.000
(F2)-13-(F1)	0.284	0.061	4.648	0.000
(F3)-14-(F1)	0.241	0.032	7.557	0.000
(F3)-15-(F2)	0.465	0.030	15.671	0.000
(DELTA3)-16-(DELTA5)	0.061	0.008	7.741	0.000

Table 7

Goodness of fit confidence intervals for IWLS and ADF measures with heterogeneous but correlated error factors

Noncentrality measures	90 percent IWLS confidence interval			90 percent ADF confidence interval		
	Lower	Point	Upper	Lower	Point	Upper
Population Noncentrality Index	0.007	0.015	0.025	0.003	0.008	0.017
Steiger-Lind RMSEA Index	0.038	0.054	0.071	0.024	0.040	0.058
McDonald Noncentrality Index	0.987	0.993	0.996	0.992	0.996	0.999
Population Gamma Index	0.992	0.995	0.998			
Adjusted Population Gamma Index	0.965	0.980	0.990			

Finally, we conclude that the $\sigma^2\mathbf{I}$ restriction imposed in our factor model is strongly questionable.

REFERENCES

AB MOOIJAART – BENTLER, P. M. [1985]: The weight matrix in asymptotic distribution-free methods. *British Journal of Mathematical and Statistical Psychology*. No. 38. p. 190–196.

AKAIKE, H. [1973]: Information theory and an extension of the maximum likelihood principle. In: *Petrov, B. N. – Csaki, F.* (Eds.) *Second International Symposium on Information Theory*. Budapest. Akadémiai Kiadó.

AKAIKE, H. [1974]: A new look at the statistical model identification. IEEE transactions on automatic control. AC-19, p. 716–723.

AKAIKE, H. [1983]: Information measures and model selection. *Bulletin of the International Statistical Institute: Proceedings of the 44th Session*. Vol. 1. p. 277–290.

ANDERSON, T. W. [1984]: Estimating linear statistical relationships. *The Annals of Statistics*. Vol. 12. No. 1. p. 1–45.

BENTLER, P. M. [1977]: Factor simplicity index and transformations. *Psychometrika*. Vol. 42. No. 2. p. 277–295.

BENTLER, P. M. [1990]: Comparative fit Indexes in structural models. *Psychological Bulletin*. Vol. 107. No. 2. p. 238–246.

BENTLER, P. M. [1983]: Some contributions to efficient statistics in structural models: sand estimation of moment structures. *Psychometrica*. Vol. 48. No. 4. p. 493–517.

BENTLER, P. M. – AB MOOIJAART [1989]: Choice of structural model via parsimony: a rationale based on precision. *Psychological Bulletin*. Vol. 106. No. 2. p. 315–317.

BENTLER, P. M.– BONNET, D. G. [1980]: Significance tests and goodness of fit in the analysis of covariance structures.

BOLLEN, K. A. [1989]: Structural equations with latent variables. Wiley. New York.

BOLLEN, K. A. [1986]: Samle size and Bentler and Bonnet's nonnormed fit index. *Psychometrika*. Vol. 51. No. 3. p. 375–377.

BROWNE, M. W. [1982]: Covariance structures. In: *Hawkins, D. M.* (Ed.) *Topics in Applied Multivariate Analysis.* MA: Cambridge University Press. Cambridge.

BROWNE, M. W. [1974]: Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*. No. 8. p. 1–24.

BROWNE, M. W. [1984]: Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*. No. 37. p. 62–83.

CUDECK, R. – BROWNE, M. W. [1983]: Cross-validation of covariance structures. *Multivariate Behavioral Research*. No. 18. p. 147–167.

BROWNE, M. W. – CUDECK, R. [1990]: Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*. No. 24. p. 445–455.

CUDECK, R. – HENLY, S. J. [1991]: Model selection in covariance structures analysis and the „problem' of sample size: a clarification. *Psychological Bulletin*. Vol. 109. No. 3. 512–519.

GOLUB, G. H. – VAN LOAN, C. F. [1996]: Matrix computations (3rd ed.). MD: Johns Hopkins University Press. Baltimore.

JENNRICH, R. I. – SAMPSON, P. F. [1966]: Rotation for simple loadings. *Psychometrika*. Vol. 31. No. 3. p. 313–323.

KAISER, H. F. [1990]: Outline of EPIC, a new method for factoring a reduced correlation matrix. *Paper presented at Society of Multivariate Experimental Psychology*. Providence. RI.

KAISER, H. F. [1958]: The varimax criterion for analytic rotation in factor analysis. *Psychometrika.* No. 23. p. 187–200.

LAWLEY, D. N. – MAXWELL, A. E. [1971]: Factor analysis as a statistical method. American Elsevier Publishing Company. Inc., New York.

KANO,Y. – BERKANE, M. – BENTLER, P. M. [1990]: Covariance structure analysis with heterogenous kurtosis parameters. *Biometrika*. No. 77. p. 575–585.

LI-TZE HU – BENTLER, P. M. [1992]: Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*. Vol. 112. No. 2. p. 351–362.

MCDONALD, R. P. [1989]: An index of goodness-of-fit based on noncentrality. *Journal of Classification*. No. 6. p. 97–103.

MCDONALD, R. P. – MARSH, H. W. [1990]: Choosing a multivariate model: noncentrality and goodness of fit. *Psychological Bulletin*. Vol. 107. No. 2. p. 247–255.
MULAIK, S. A. ET AL. [1989]: An evaluation of goodness of fit indices for structural equation models. *Psychological Bulletin*. Vol. 105. No. 3. p. 430–445.
RAFTERY, A. [1993]: Bayesian model selection in structural equation models. In: *Bollen, K. – Long, J.* (Eds.) *Testing structural equation models.* Newbury Park. California. p. 163–180.
RAFTERY, A. [1995]: Bayesian model selection in social research. In: *Marsden, P.* (Ed.) *Sociological Methodology 1995*. San Francisco. p. 111–163.
SCHWARZ, G. [1978]: Estimating the dimension of a model. *Annals of Statistics*. No. 6. p. 461–464.
STEIGER, J. H. – LIND, J. C. [1980]: Statistically-based tests for the number of common factors. Paper presented at the annual Spring Meeting of the Psychometric Society in Iowa City. May 30.
TANAKA, J. S. – HUBA, G. J. [1985]: A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*. No. 38. p. 197–201.
TUCKER, L. R. – LEWIS, C. [1973]: The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*. No. 38. p. 1–10.