İstanbul Üniversitesi İşletme Fakültesi Dergisi
Istanbul University Journal of the School of Business Administration
Cilt/Vol:40, Sayı/No:1, 2011, 86-123
ISSN: 1303-1732 – www.ifdergisi.org © 2011

Dedicated to the 40th Anniversary of the Journal

Structural equation modeling (SEM) of categorical and mixed-data using the novel Gifi transformations and information complexity (ICOMP) criterion

Eylem Deniz Howe [1]
Department of Statistics
Mimar Sinan Fine Arts University, Istanbul, Turkey

Hamparsum Bozdogan [2]
Department of Statistics, Operations and Management Science
The University of Tennessee, Knoxville, TN, U.S.A

Suman Katragadda [3]
Advanced Analytics
Express-Scripts Company
St. Louis, MO, U.S.A.

Abstract

This paper introduces and develops a novel and computationally feasible alternative approach to the analysis of categorical, dichotomous, and mixed data sets in structural equation models (SEMs) to overcome currently existing problems. Our approach is based on the Gifi system. The Gifi system uses the optimal scaling methodology to quantify the observed categorical variables. In the quantification process, information in the observed variable is retained in the quantified variable. That is, the Gifi system transforms categorical data to continuous data without destroying the scale properties of the categorical variables. The scaling is thus preserved in the transformed nonlinear continuous Gifi data space. Hence the transformation is invertible. This is one of the unique characteristics of the Gifi system which avoids the arbitrary thresholding specification that is currently practiced and used in the literature. After the Gifi transformation, we analyze the transformed data set using SEM based on the multinormal distributional assumption. Such an approach legitimizes the distributional assumption of multivariate normality in SEM.

Information-theoretic model selection criteria such as *Akaike's* [1] *AIC*, *Bozdogan's* [2] *Consistent AIC*, called *CAIC*, and the *information-theoretic measure of complexity ICOMP* criterion of Bozdogan [3-7] are introduced and develop as measures of fit in SEMs. The model with the minimum values of the criteria is selected as the best fitting model among a portfolio of candidate models.

We provide a real benchmark numerical example using SEM on a categorical data set which measures the quality of life (QOL) to illustrate the versatility and flexibility of our approach using the Gifi transformations on this data set and fit five alternative SEM models by scoring the model selection criteria.

Keywords: *Structural Equation Models (SEMs), the Gifi System and Transformation, Akaike's AIC, Consistent AIC, Information-theoretic Measure of Complexity ICOMP Criterion*

[1] *eylemdeniz@gmail.com (E. Deniz Howe)*

[2] *bozdogan@utk.edu (H. Bozdogan)*

[3] *katragaddasuman11@gmail.com (S. Katragadda)*

E. Deniz Howe, H. Bozdogan, S. Katragadda / İstanbul Üniversitesi İşletme Fakültesi Dergisi 40, 1, (2011) 86-123 © 2011

Kategorik ve karma veri setlerinin yapısal eşitlik modellemesinde (YEM) Gifi yaklaşımı kullanımı ve bilgi karmaşıklığı kriteri (ICOMP)

Özet

Bu çalışmada Yapısal Eşitlik Modelleri'nde (YEM) kategorik, ikili veya karma veri setlerinin analizine ilişkin var olan problemleri çözmek için özgün bir alternatif yaklaşım olarak Gifi yöntemi önerilmiştir. Gifi yönteminde, kategorik değişkenleri nicel hale dönüştürmek için optimal ölçekleme yöntemi kullanılır. Nicelleştirme sürecinde gözlenen değişkendeki bilgi, dönüştürülmüş değişkende aynen korunur. Yani Gifi yöntemi, kategorik değişkenlerin ölçek özelliklerini bozmadan kategorik veriyi sürekli veriye dönüştürür ve bu dönüştürme işleminde herhangi bir bilgi kaybı söz konusu olmaz. Ölçek özellikleri, dönüştürülmüş doğrusal olmayan sürekli Gifi veri uzayında saklanır. Bu nedenle dönüştürme işleminden geriye dönüş mümkündür. Bu işlem, literatürde halen uygulanmakta olan rasgele belirlenmiş başlangıç değerlerini göz ardı eden Gifi sisteme özgün bir özelliktir.

Gifi dönüşümünden sonra, çoklu normal dağılım varsayımına dayalı YEM kullanılarak dönüştürülmüş veri seti analiz edilmiştir. Böyle bir yaklaşım YEM'de, kategorik veriler için göz ardı edilen çok değişkenli normal dağılım varsayımını sağlamaktadır.

Akaike'nin [1] Akaike Bilgi Kriteri (AIC), Bozdogan'ın [2] Tutarlı Akaike Bilgi Kriteri (CAIC) ve Bozdogan'ın [3-7] Bilgi Karmaşıklığı Kriteri (ICOMP) gibi bilgiye dayalı model seçim kriterleri YEM'de uyumun bir ölçümü olarak uygulanmaktadır. Minimum kriter değerini veren model, rakip modeller arasında veriye en iyi uyumlu model olarak seçilir.

Bu çalışmada yaşam kalitesinin ölçüldüğü gerçek bir kategorik veri seti kullanılmıştır. Bu veri setine Gifi dönüşüm uygulayarak önerilen yaklaşımın çok yönlülüğü ve esnekliği gösterilmiştir. Ayrıca dönüştürülmüş veri seti üzerinden farklı YEM için model seçim kriter değerleri elde edilmiş ve minimum kriter değerini veren en iyi model belirlenmiştir.

Anahtar Sözcükler: *Yapısal Eşitlik Modelleri (YEM), Gifi Yöntemi ve Dönüşümü, Akaike Bilgi Kriteri (AIC), Tutarlı Akaike Bilgi Kriteri (CAIC), Bilgi Karmaşıklığı Kriteri (ICOMP)*

1. Introduction and Purpose

In this paper, we shall introduce a new and novel data-analytic approach to modeling structural equation models (SEMs) with continuous, categorical and mixed data sets as an alternative to the currently practiced threshold modeling technique where the discrete data are treated as coming from a hidden continuous normal distribution with a specified fixed threshold.

In general, factor analysis (FA), structural equation modeling (SEM), and other statistical modeling techniques have been developed under the assumption that the observed variables have continuous multivariate normal (Gaussian) distributions. These techniques are very much dependent upon the quality and type of the data set at the disposal of a researcher. But much of the real world data sets are of three types- *continuous*, *categorical* and *mixed*. A *continuous data set* is one in which all the variables in it are in continuous form. That is, all the variables are continuous. If a variable can take on any value between its minimum value and its maximum value, we call it a continuous variable; otherwise, it is called a discrete variable. A *categorical data set* is one in which all the variables in it are either ordinal (ordering of the categories exists) or nominal (no specific ordering of the categories exists). A particular form of categorical data set is a *binary* data set in which all the variables take values 0 and 1's. A *mixed data set* is one which contains some of the variables in continuous form and the rest of the variables in categorical form. In other words, a mixed data set is a combination of both *continuous* and *categorical variables*.

In reality, much of the data sets obtained in behavioral, economic, medical, and social sciences typically involve a relatively small number of *continuous variables*. By and large, the measured variables are *categorical, binary*, or *mixed data* types. In such data structures, we cannot any longer assume multivariate normal distribution to model the data.

Consequently, the analysis of SEMs with ordered categorical and mixed data sets is quite difficult and challenging. A major difficulty is that presence of categorical and binary (or dichotomous) variables violate the assumptions of continuity and multivariate normality that are needed in SEM. Also, a more serious consequence of the presence of discrete variables, in the form of categorical variables, is the violation of the covariance structure hypothesis [8]. Furthermore, we need to compute the multiple integrals associated with the cell probabilities that are induced by the ordered categorical variables [9].

As is well known, structural equation models (SEMs) consist of two components, a measurement model and a structural model. When we have categorical, dichotomous or mixed data sets, these have implications concerning the measurement component of SEMs. The structural component of SEM remains intack and it is the same as in the continuous data case (Skrondal and Rabe-Hesketh [10, 11]).

In reviewing the literature, we note that the early work in the analysis of factor analysis (FA) model and SEMs with categorical, dichotomous, and mixed data sets appear to be due to Christoffersson [12], Muthén [13-18], Muthén and Christoffersson [19], Muthén and Kaplan [20], Bartholomew [21], and Bartholomew [22], to mention a few. For a good source of a review article on the recent developments of the factor analysis (FA) of categorical variables, we refer the readers to Mislevy [23]. In most of these previous works by and large research focus has been on using the threshold modeling where the discrete data are treated as coming from a hidden continuous normal distribution with a specified fixed threshold. Based on this, several multistage estimation techniques such as the *weighted least-squares* (*WLS*), *generalized least-squares* (*GLS*), and *full* and *limited information* techniques have been proposed and developed to reduce the computational complexity in SEMs.

Most recently Lee [9], treating discrete data as observations coming from a hidden continuous distribution with a threshold specification, introduced a Bayesian approach for analyzing SEMs with categorical data sets. Although, this is an interesting approach, it still does not resolve the existing problems in SEMs. For example, in the Bayesian approach, we still need to evaluate the posterior distribution of the model which is rather complicated and, moreover, the analysis is computationally intensive using the Gibbs sampler method [24].

With these existing problems in mind, the primary objective of this paper is to introduce and develop a novel and computationally feasible alternative approach to analyze categorical, dichotomous, and mixed data sets in SEMs to resolve the existing problems. Our approach is based on the Gifi [25] system or transformation. The Gifi system uses the optimal scaling methodology to quantify the categorical variables. In the quantification process, information in the observed variable is retained in the quantified variable. In other words, the Gifi system transforms categorical data to continuous data without destroying the scale properties of the categorical variables and the transformation has the *"one-to-one and onto"* feature. In this manner, the scaling is preserved in the transformed nonlinear continuous Gifi data space. Hence, the transformation is invertible. This is one of the unique characteristics of the Gifi system, which avoids the arbitrary thresholding specification that is currently practiced and used in the literature. After the Gifi transformation, we can now analyze the transformed data set using SEM based on the multinormal distributional assumption. Such an approach legitimizes the distributional assumption of multivariate normality or other multivariate distributions in SEM, since this can be easily tested using the existing multivariate tests in the literature.

In addition to the introduction of the Gifi system or transformation, our objective in this paper is also to introduce and develop information-theoretic model selection criteria such as *Akaike's* [1] *classic information criterion* (*AIC*), *Bozdogan's* [2] *Consistent AIC*, called *CAIC*, and the *information-theoretic measure of complexity ICOMP* criterion of Bozdogan [3-7] as measures of fit in SEMs. The model with the minimum values of the criteria is selected as the best fitting model among a candidate of portfolio of models.

The paper is organized as follows. In Section 2, we set up the general structural equation models (SEMs) with latent variables and measurement error. In Section 3, we briefly discuss SEMs with categorical, dichotomous, and mixed data sets and present the idea of threshold modeling. In Section 4, we introduce and discuss the Gifi [25] system. That is, we present homogeneity analysis and mapping the data to Gifi space. Section 5 presents several information-theoretic model selection criteria. For space considerations, we restrict the detailed proofs and dervations of these criteria where appropriate. For more details on information criteria, we will refer the readers to Bozdogan [3-7, 26, 27], Bozdogan and Haughton [28], Bozdogan and Bearse [29] and Bozdogan and Ueno [30]. In Section 6, we provide the derived forms of the model selection criteria in SEMs. In Section 7, we provide a real benchmark numerical example using SEM on a categorical data set which measures the quality of life (QOL) to illustrate the versatility and flexibility of our approach using the Gifi transformations on this data set and fit alternative SEM models by scoring the model selection criteria. We show that, the transformed categorical data using SEM is the best fitting model. Later, after we select the best model, we also choose the best parsimonious SEM using the information criteria. We compare our results with those obtained by Lee's [9] using the Bayesian approach. Section 8 concludes the paper with some discussion and future work.

2. The General Structural Equation Models (SEMs) with Latent Variables and Measurement Error

During the past two decades, structural equation models (SEMs) have become a popular data-analytic tool in social and behavioral sciences for the analysis of the causal modeling of complex multivariate data sets. SEMs have been very useful in solving many substantive problems in several cross-disciplinary areas. These include, but are not limited to, engineering, management and information sciences, market research, genomic research involving DNA data mining. Because of their importance in data analysis, SEMs are well established and known under several other names, such as *LISREL models*, *covariance structure models*, and *latent variables and measurement error models*. An excellent review on SEMs in terms of statistical practice, theory, and directions can be found in Bentler [31] and the references therein.

More formally, structural equestion models (SEMs) with latent variables and measurement error (or LISREL models) are those complex models involving specified causal structures among the unobserved latent variables or hypothetical constructs induced from the observed multivariate X-Y data.

Given the observed data:

$$y' = \left[y_1, y_2, \ldots, y_p \right] = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix} = Y(n \times p)$$

and

$$x' = \left[x_1, x_2, \ldots, x_p \right] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nq} \end{bmatrix} = X(n \times q),$$

there are two aspects in the construction and the use of structural equations models (SEMs) in studying dependence. Hereafter, we will use the acronym SEM or SEMs for brevity.

1) Structural dependence of the latent variables obtained from the measurement models, and
2) The dependence among the p *observed endogenous* y-variables and the q *exogenous* x-variables which are studied by the measurement models.

In general, these two aspects are characterized within the full SEM given in Jöreskog and Sörbom [32] in matrix notation as follows.

Structural equation model:

$$\underset{(r\times1)}{\eta} = \underset{(r\times r)(r\times1)}{B\eta} + \underset{(r\times s)(s\times1)}{\Gamma\xi} + \underset{(r\times1)}{\zeta} \qquad (1)$$

Measurement model for y:

$$\underset{(p\times1)}{y} = \underset{(p\times r)}{\Lambda_y} \underset{(r\times1)}{\eta} + \underset{(p\times1)}{\varepsilon} \qquad (2)$$

Measurement model for x:

$$\underset{(q\times1)}{x} = \underset{(q\times s)}{\Lambda_x} \underset{(s\times1)}{\xi} + \underset{(q\times1)}{\delta} \qquad (3)$$

where η (eta) is a $(r \times 1)$ vector of latent endogenous (or dependent) variables, ξ (xi) is $(s \times 1)$ vector of latent exogenous (or independent) variables, ζ (zeta) is $(r \times 1)$ vector of latent errors in equations, B (beta) is a $(r \times r)$ coefficient matrix for the latent endogenous variables, and Γ is a $(r \times s)$ coefficient matrix for the latent exogenous variables.

The structural model specifies the causal relationships among the latent endogenous variables in B, between the exogenous and endogenous variables in Γ, and describes unexplained residuals of the latent factors in ζ.

The usual assumptions for the structural model are that: $E(\eta) = 0,\ E(\xi) = 0,\ E(\zeta) = 0$, ζ is uncorrelated with ξ, and $(I - B)$ is nonsingular. The covariance matrices are:

$$E(\xi\xi') = \Phi\,(phi) \qquad (4)$$

is an $(s \times s)$ covariance matrix of the latent exogeneous variables, and

$$E(\zeta\zeta') = \Psi\,(psi) \qquad (5)$$

is a $(r \times r)$ covariance matrix of the latent errors in equations.

The measurement model specifies how the observed variables, x and y, are determined through Λ_y and Λ_x by the latent variables, ξ and η, respectively. The ε and δ terms represent the residuals in x and y unexplained by ξ and η. The usual assumptions are: $E(\eta)=0,\ E(\xi),\ E(\varepsilon)=0,$ and $E(\delta)=0$. Furthermore, ε is uncorrelated with $\eta, \xi,$ and δ. Likewise, δ is uncorrelated with $\xi, \eta,$ and ε. The covariances matrices in the case of the measurement model are:

$$E(\varepsilon\varepsilon') = \Theta_\varepsilon = \Sigma(\varepsilon) \tag{6}$$

a $(p \times p)$ covariance matrix of ε, and

$$E(\delta\delta') = \Theta_\delta = \Sigma(\delta) \tag{7}$$

a $(q \times q)$ covariance matrix of δ.

For more on the above set up of SEM, we refer the readers to Jöreskog and Sörbom [32], Bollen [8], and others.

2.1. Implied Model Covariance Matrix

Let Σ be the population covariance matrix of y and x, and let θ be tha parameter vector of the model. Then Σ is a function of the free parameters in θ, denoted by $\Sigma(\theta)$ and is given by the following partitioned matrix of dimension $(p+q)\times(p+q)$:

$$\begin{aligned}\Sigma(\theta) &= \begin{bmatrix} \Sigma_{yy}(\theta) & \Sigma_{yx}(\theta) \\ \Sigma_{xy}(\theta) & \Sigma_{xx}(\theta) \end{bmatrix} \\ &= \begin{bmatrix} \Lambda_y (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I-B')^{-1}\Lambda_y + \Theta_\varepsilon & \Lambda_y (I-B)^{-1}\Gamma\Phi\Lambda'_x \\ \Lambda_x \Phi\Gamma'(I-B')^{-1}\Lambda'_y & \Lambda_x\Phi\Lambda'_x + \Theta_\delta \end{bmatrix}.\end{aligned} \tag{8}$$

Hence, if the structural equation model (SEM) is correct, then the basic fundamental hypothesis of the general SEM is

$$\Sigma = \Sigma(\theta). \tag{9}$$

We note that the elements of the implied model covariance matrix $\Sigma(\theta)$ is a function of the elements of Λ_y, Λ_x, B, Γ, Φ, Ψ, Θ_δ, and Θ_ε matrices.

To be able to use the SEMs, we need to specify the pattern of the elements of these eight matrices. There are three kinds of specifications that are given in Jöreskog and Sörbom [32]. These are:

1) *Fixed parameters*- that have been assigned given values,
2) *Constrained parameters*- that are unknown but equal to one or more other parameters, and
3) *Free parameters*- that are unknown and not constrained to be equal to any other parameter.

2.2. Estimation of Model Parameters

To estimate the unknown parameters of the SEMs, with the assumption that the observed variables have a multivariate normal distribution, and providing that the model and parameter identifiability hold, we use one of the most popular methods. Namely, we use the *maximum likelihood* (*ML*) estimation.

In SEM, there are usually more unobserved variables than observed variables in the model and identifiability depends on the choice of model and the specification of fixed, constrained, and free parameters. In SEM, there are $1/2(p+q)(p+q+1)$ equation in t unknown parameters $\theta_1, \theta_2, \ldots, \theta_t$, t being the *total number of free parameters* in the model. Hence, a necessary condition for identification of all parameters is that

$$t \leq \frac{1}{2}(p+q)(p+q+1). \tag{10}$$

If a parameter θ can be determined from these equations, we say that this parameter is identified. Otherwise, it is not. Another way to check the parameter identifiability is to check the rank of the estimated *Fisher information matrix* of the model. Indeed, the existing canned *"black-box"* SEM software checks the identification of the model by checking the *positive definiteness of the information matrix* (*FIM*). For more on these, we refer the readers to Jöreskog and Sörbom [32].

Under the assumption that the observed variables are continuous and have interval scales, we consider that

$$z = (y', x')' \sim N_{(p+q)}\left(0, \Sigma(\theta)\right) \ (\textit{multivariate normal}) \tag{11}$$

with the probability density function

$$f(z;\Sigma) = (2\pi)^{-(p+q)/2} \left|\Sigma(\theta)\right|^{-1/2} \exp\left\{-\frac{1}{2} z'\Sigma(\theta) z\right\}. \tag{12}$$

Suppose we have a random sample of n observations $z_1, z_2, \ldots, z_n$. We define the likelihood of all the samples for the SEM by

$$L(\theta) = (2\pi)^{-n(p+q)/2} \left|\Sigma(\theta)\right|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} z_i'\Sigma(\theta) z_i\right\}. \tag{13}$$

The log likelihood function is

$$\begin{aligned} l(\theta) \equiv \log L(\theta) &= -\frac{n}{2}(p+q)\log(2\pi) - \frac{n}{2}\log\left|\Sigma(\theta)\right| - \frac{1}{2}\sum_{i=1}^{n} z_i'\Sigma(\theta) z_i \\ &= -\frac{n}{2}(p+q)\log(2\pi) - \frac{n}{2}\log\left|\Sigma(\theta)\right| - \frac{n}{2} tr\left(S\Sigma^{-1}(\theta)\right), \end{aligned} \tag{14}$$

where S is the sample ML estimator of the covariance matrix of the data which uses n as the divisor. That is, $S = A/n$, and where $A = \sum_{i=1}^{n} z_i' z_i$ is the sums-of-squares-and-cross-product (SSCP) matrix.

The negative two times the log likelihood function becomes

$$-2l(\theta) \equiv -2\log L(\theta) = n(p+q)\log(2\pi) + n\log\left|\Sigma(\theta)\right| + ntr\left(S\Sigma^{-1}(\theta)\right). \tag{15}$$

Instead of maximizing the log likelihood function, we equivalently minimize the following fit (or discrepancy) function to obtain the MLE's.

$$F_{ML} = \log\left|\Sigma(\theta)\right| + tr\left(S\Sigma^{-1}(\theta)\right) - \log\left|S\right| - (p+q). \tag{16}$$

Similarly, other specific fit functions are based on different types of parameter estimates. The most commonly used ones are: *unweigted least squares* (*ULS*), *weighted least squares* (*WLS*), and the *generalized least squares* (*GLS*).

There are many small scale practical examples where explicit solutions for the structural parameters which minimize F_{ML} exist. However, in general, in SEMs, the implied model covariance matrix $\Sigma(\theta)$, even in the simplest hybrid models is a nonlinear function of the parameters, and, consequently, of the parameter estimates. This makes F_{ML} a complicated nonlinear function of the structural parameters, and explicit solutions are not always found (Bollen [8, p. 108]). Therefore, in the literature of SEM, there are several iterative algorithms used to minimize F_{ML} to find the maximum likelihood (ML) estimators of the structural parameters. These include, *Fletcher-Powell* [33] (*FP*), *Fletcher-Reeves* [34] (*FR*), *Gauss-Newton* (*GN*), *Newton-Raphson* (*NR*), and *Fisher Scoring* (*FS*) algorithms. For the analysis of SEM, these algorithms are implemented in various existing standard canned computer programs such as LISREL (Jöreskog and Sörbom [32]), EQS (Bentler [35]), and other computer packages.

The minimization procedure, for example, for the *Fletcher-Powell* (*FP*) algorithm uses the first-order derivatives and approximations to the second-order derivatives of F_{ML}, and the convergence of the algorithm is quite rapid to obtain the MLEs. At the convergence, let $\hat{\Sigma}(\hat{\theta})$ denote the maximum likelihood estimator (MLE) of the implied covariance matrix $\Sigma(\theta)$. Then, the estimated implied model covariance matrix is

$$\hat{\Sigma}(\hat{\theta}) = \begin{bmatrix} \hat{\Lambda}_y (I-\hat{B})^{-1}(\hat{\Gamma}\hat{\Phi}\hat{\Gamma}' + \hat{\Psi})(I-\hat{B}')^{-1}\hat{\Lambda}_y + \hat{\Theta}_\varepsilon & \hat{\Lambda}_y (I-\hat{B})^{-1}\hat{\Gamma}\hat{\Phi}\hat{\Lambda}'_x \\ \hat{\Lambda}_x \hat{\Phi}\hat{\Gamma}'(I-\hat{B}')^{-1}\hat{\Lambda}'_y & \hat{\Lambda}_x\hat{\Phi}\hat{\Lambda}'_x + \hat{\Theta}_\delta \end{bmatrix}. \tag{17}$$

For a comprehensive review of the iterative algorithms for SEM, we refer the readers to Lee and Jennrich [36] and Jamshidian and Bentler [37].

2.3. Fisher Information Matrix (FIM) and Its Inverse (IFIM)

Let $\hat{\theta}$ denote a vector of maximimum likelihood (ML) estimates of the model structural parameters that maximizes $\log L(\theta)$ in (14), or equivalently, minimizes F_{ML} and that both functions contain the sample covariance matrix S of the observed variables.

Let us denote

$$\mathcal{H}(\theta) = \frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'} \tag{18}$$

to be the Hessian matrix of the log likelihood function. Then the expected Fisher information, or in short *Fisher information matrix* (*FIM*) is defined as

$$\begin{aligned}\mathcal{F}(\theta) &= -E\left(\mathcal{H}(\theta)\right) \\ &= -E\left(\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'}\right) = E\left(\left(\frac{n}{2}\right)\left(\frac{\partial^2 F_{ML}}{\partial\theta\partial\theta'}\right)\right).\end{aligned} \tag{19}$$

If we are able to take the expectation in (19), then the estimated covariance matrix $\hat{C}ov(\hat{\theta})$ of the parameter vector of the model, or quivalently, the *estimated inverse Fisher information matrix* (*IFIM*) $\hat{\mathcal{F}}^{-1}$ is given by

$$\hat{C}ov(\hat{\theta}) = \hat{\mathcal{F}}^{-1} = \left\{ -E\left(\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'}\right)\Bigg|_{\theta=\hat{\theta}} \right\}^{-1}. \tag{20}$$

If we cannot take the expectation in (19), then the *estimated inverse of the observed Fisher information matrix* $\hat{\mathcal{F}}^{-1}_{Obs}$ is given by

$$\hat{\mathcal{F}}^{-1}_{Obs} = \left\{ \left(-\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'}\right)\Bigg|_{\theta=\hat{\theta}} \right\}^{-1}. \tag{21}$$

The primary advantage of using $\hat{\mathcal{F}}^{-1}$ over $\hat{\mathcal{F}}^{-1}_{Obs}$ as an estimator of $\mathcal{F}^{-1}$ is that $\hat{\mathcal{F}}^{-1}$ is a maximum likelihood estimator (MLE) of $\mathcal{F}^{-1}$ which follows from the invariancy property of the MLE's. In many cases, taking expectations is difficult especially when the log likelihood function is a nonlinear function of the parameters. In such cases, working with $\hat{\mathcal{F}}^{-1}_{Obs}$ is a reasonable choice. Indeed in SEM software such as EQS, LISREL, Mplus, SAS (Calis), the standard error of the parameter for the jth parameter in θ is computed as

$$SE(\hat{\theta}_j) = \sqrt{\hat{\mathcal{F}}_{jj}^{-1}} \tag{22}$$

where $\hat{\mathcal{F}}_{jj}^{-1}$ is the $(j, j)th$ entry of $\hat{\mathcal{F}}^{-1}$.

Since we assumed that the observed variables are continuous and have interval scales, and since $z = (y', x')' \sim N_{(p+q)}\left(0, \Sigma(\theta)\right)$ multivariate normal, modifying the results in Bozdogan [38], Williams, Bozdogan, Aiman-Smith [39], and following the derivations in Bozdogan [7], the estimated inverse-Fisher information matrix (IFIM) for the general SEM is given by

$$\begin{aligned}\hat{C}ov(\hat{\theta}) = \hat{\mathcal{F}}^{-1} &= \begin{bmatrix} \hat{\mathcal{F}}^{-1}_{11} & 0 \\ 0 & \hat{\mathcal{F}}^{-1}_{22} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n}\hat{\Sigma}(\hat{\theta}) & 0 \\ 0 & \frac{2}{n} D^{+}_{(p+q)}\left(\hat{\Sigma}(\hat{\theta}) \otimes \hat{\Sigma}(\hat{\theta})\right) D^{+\prime}_{(p+q)} \end{bmatrix},\end{aligned} \tag{23}$$

where $\hat{\Sigma}(\hat{\theta})$ is given in (17). In (23) the matrix $D^{+}_{(p+q)}$ is the Moore-Penrose inverse of the duplication matrix $D_{(p+q)}$. The duplication matrix $D_{(p+q)}$ is a unique $(p+q)^2 \times \frac{1}{2}(p+q)(p+q+1)$ matrix, and so its Moore-Penrose inverse is

$$D^{+}_{(p+q)} = \left(D'_{(p+q)} D_{(p+q)}\right)^{-1} D'_{(p+q)} \qquad (24)$$

which is a $\frac{1}{2}(p+q)(p+q+1) \times (p+q)^2$ matrix. Further, note that

$$D_{(p+q)} vech\left(\hat{\Sigma}(\hat{\theta})\right) = vec\left(\hat{\Sigma}(\hat{\theta})\right), \qquad (25)$$

where $vech(\bullet)$ denotes the half-vec operator. For any $(m \times m)$ matrix A the vector $vech(A)$ denotes the $1/2m(m+1) \times 1$ vector that is obtained from $vec(A)$ by eliminating all supradiagonal elements of A. For example, for $m = 2$,

$$vec(A) = \left(a_{11}, a_{21}, a_{12}, a_{22}\right)' \quad and \quad vech(A) = \left(a_{11}, a_{21}, a_{22}\right)'.$$

where the supradiagonal element a_{12} has been removed. Thus, for symmetric A, $vech(A)$ only contains the distinct elements of A. Now, if A is symmetric, the elements of $vec(A)$ are those of $vech(A)$ with some repetitions. For more on these, see, e.g., Magnus [40], and Magnus and Neudecker [41].

We shall use these results in computing the information complexity (ICOMP) criterion later for the fitted structural equation models (SEMs).

LISREL computer program checks the parameter identifiability by computing the *Fisher information matrix* (*FIM*) and the *standard errors* (*SEs*) of the parameters. If *FIM* is positive definite, then the model and the structural parameters are identified. If *FIM* is singular, the model and the structural parameters are not identified.

3. SEMs with categorical, dichotomous, and mixed data sets: The current state

As we note from the discussion of in Section 2 above, and in general, the theory of factor analysis (FA) and structural equation models (SEMs) have been developed under the assumption that the observed variables have continuous multivariate normal (Gaussian) distributions. These techniques are very much dependent on the quality and type of the data sets at hand.

In reality, much of the data sets obtained in behavioral, economic, medical, and social sciences usually involve a relatively small number of *continuous variables*. By and large, they contain many *categorical, binary*, or *mixed* variables. In such data structures, we cannot any longer assume multivariate normal distributions to model the data. Consequently, the analysis of SEMs with ordered categorical and mixed data sets is quite difficult and challenging. This is due to the fact that, presence of categorical and binary (or dichotomous) variables violate the assumptions of continuity and multivariate normality that are needed in SEM. Also, a more serious consequence of the presence of these discrete variables is the violation of the covariance structure hypothesis [8, p. 434]. Furthermore, we need to compute the multiple integrals associated with the cell probabilities that are induced by the ordered categorical variables [9, p. 141].

In the literature of SEM, currently, the state-of-the affairs is that when the categorical variable is ordinal it is thought of as the discretized version of unobserved continuous variable and where the discrete data are treated as coming from a *hidden continuous normal distribution with a specified fixed threshold*. It is because of this that the name *"threshold model"* is used.

Threshold model:

Let y_j^* denote the latent continuous variable and let y_j be the corresponding observed ordered categorical variables with c_j response categories. Each observed categorical response y_j is related to a latent continuous response y_j^* via a threshold model. For ordinal observed responses it is assumed that

$$y_j = \begin{cases} 1, & if\ y_j^* \le \tau_{j,1} \\ 2, & if\ \tau_{j,1} \le y_j^* \le \tau_{j,2} \\ \vdots & \\ c_j - 1, & if\ \tau_{j,c_j-2} \le y_j^* \le \tau_{j,c_j-1} \\ c_j, & if\ \tau_{j,c_j-1} < y_j^* \end{cases} \tag{26}$$

where $\tau_{j,1}, \tau_{j,2}, \cdots, \tau_{j,c_j-1}$ are thresholds. For example, if we have four odered response categories, Figure 1 shows the histogoram of four odered response categories.

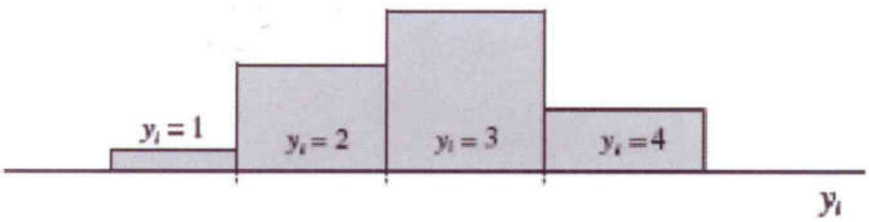

Figure 1 Histogoram of Four Odered Response Categories [42]

Figure 2 shows the underlying normal distribution with a threshold specification.

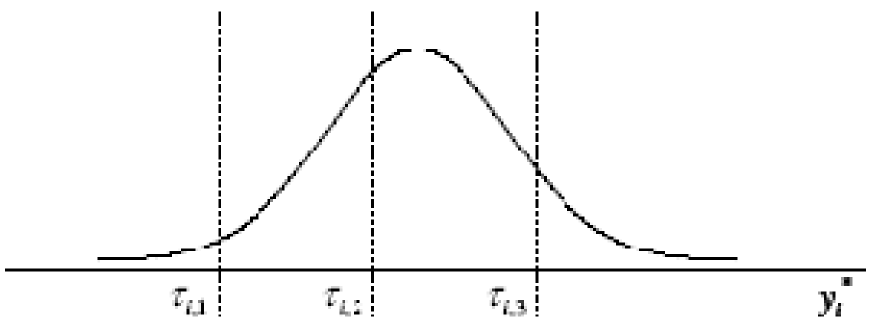

Figure 2 The Underlying Normal Distribution with a Threshold Specification [42]

The ordered categorical observations that give the histogram in Figure 1 can be captured by the standard normal distribution $N(0,1)$ with appropriate thresholds as in Figure 2.

For a random vector $y = (y_1, y_2, \ldots, y_p)$ of ordinal categorical items, the distribution of the underlying continuous random vector $y^* = (y_1^*, y_2^*, \ldots, y_p^*)$ is multivariate normal with a correlated structure. As the probability density function of y involves a complicated integral of high dimension, the statistical analysis is nontrivial and challenging. In the literature, current research focus has been on using the threshold modeling where the discrete data are treated as coming from a hidden continuous normal distribution with a specified fixed threshold. Based on this, several multistage estimation techniques such as the *weighted least-squares* (*WLS*), *generalized least-squares* (*GLS*), and *full* and *limited information* techniques have been proposed and developed to reduce the computational complexity in SEMs.

Most recently Lee [9], treating that the discrete data as observations coming from a hidden continuous distribution with a threshold specification, introduced a Bayesian approach for analyzing SEMs with categorical, dichotomous, and mixed data sets.

Although, Bayesian approach is an interesting approach, it still does not resolve the currently existing problems in SEMs. For example, in the Bayesian approach, we still need to evaluate the posterior distribution of the model which is rather complicated and, moreover, the analysis is computationally intensive using the Gibbs sampler method [24].

With these existing problems in mind, and rather than assuming that the discrete data are treated as coming from a *hidden continuous normal distribution with a specified fixed threshold*, in the next section we introduce a rather novel and computationally feasible alternative approach to analyze categorical, dichotomous, and mixed data sets in SEMs.

This approach is called the Gifi system or transformation.

4. Homogeneity Analysis and Mapping the Data to Gifi Space

The Gifi [25] system was originally developed by Albert Gifi at the Department of Data Theory at Leiden University dating back to 1968. This novel procedure was later popularized in the Netherlands and nore widely in Europe by the emergence of the Gifi [25] book by his academic followers. For this see, e.g., Heiser and Meulman [43], Meulman [44-49], Meulman and der Kooij [50], Meulman et al. [51], and Michailidis and de Leeuw [52] who reviewed the concepts of Gifi transformation applied on a pure categorical data set. It was shown in detail the application of several classical multivariate techniques on the transformed scale to identify patterns in the categorical set.

In the United States, not much work has been done using this procedure from model selection point of view, except the work of Katragada [53], Katragada and Bozdogan [54].

4.1. What is Optimal Scaling?

The idea behind optimal scaling is to assign numerical quantifications to the categories of each variable, thus allowing standard procedures to be used to obtain a solution on the quantified variables. The optimal scale values are assigned to categories of each variable based on the optimizing criterion of the procedure in use. Unlike the original labels of the nominal or ordinal variables in the analysis, these scale values have metric properties.

Two algorithms are used to analyze the data in Gifi space. One algorithm is *Optimal Scaling Method* (*OSM*) and other one is *Linear Combination Method* (*LCM*). In *OSM*, the p-dimensional categorical variables are transformed to p-dimensional continuous variables. The *LCM*, on the other hand, does linear combinations of the categories of the p-dimensional categorical variables and transforms them into 1-dimensional continuous space. *LCM* is useful when the dimension of the categorical variables is very large.

Let $\left(k_1,\ldots,k_j,\ldots,k_m\right)$ be the m-vector containing the number of categories of each variable, and let p denote the dimensionality of the analysis that one needs to choose. Let each variable $v_j\,(j=1,\ldots,m)$ be coded into an $(n\times k_j)$ indicator matrix G_j. An indicator matrix indicates which categories are scored by which objects. Rows of an indicator matrix usually refer to objects and columns to categories. Its elements consist of zeros (not scored) and ones (scored).

Homogeneity analysis determines quantifications of the categories of each of the variables such that homogeneity is maximized. Let X be the $(n\times p)$ matrix (usually $p\le m$) containing the object scores. If y_j, a k_j-vector, is the quantification of the categories of variable v_j, then $G_j y_j$ represents a single quantification or transformation of the n objects for variable v_j. Without additional conditions on the y_j, objects in the same categories get the same quantification. In homogeneity analysis, simultaneous quantifications for each of the variables are collected in the $k_j\times p$ matrices Y_j, called multiple nominal quantifications. Thus matrices G_jY_j induce p multiple quantifications of the objects for variable v_j. For example,

$$v_j=\begin{bmatrix}a\\b\\a\\c\\c\\a\end{bmatrix}\Rightarrow\begin{bmatrix}y_1\\y_2\\y_1\\y_3\\y_3\\y_1\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\1&0&0\\0&0&1\\0&0&1\\1&0&0\end{bmatrix}\times\begin{bmatrix}y_{j1}\\y_{j2}\\y_{j3}\end{bmatrix}=G_jY_j, \qquad (27)$$

where $G_j y_j$ represents a single transformation of the n objects induced by variable j.

Perfect homogeneity is defined if all multiple quantifications of the objects are the same for all variables which means that $X=G_1Y_1=\cdots=G_mY_m$. Homogeneity analysis thus amounts to minimizing

$$\begin{aligned}\sigma\left(X;Y_1\ldots\ldots Y_m\right)&=\frac{1}{m}\sum_{j=1}^{m}SSQ\left(X-G_jY_j\right)\\&=\frac{1}{m}trace[\left(X-G_jY_j\right)'\left(X-G_jY_j\right)]\end{aligned} \qquad (28)$$

over object scores X and multiple nominal quantifications Y_j under appropriate normalization conditions. It should be emphasized that the choice of normalization of X is crucial. In equation (28), $SSQ(\bullet)$ denotes the sum of squares. The loss function in equation (28) is at the heart of the Gifi [25] system.

We minimize the above loss function simultaneously over X and Y_j by employing an iterative method called *Alternating Least Squares* (ALS) algorithm. After the current quantifications are used to find a solution, the quantifications are updated using that solution. The updated quantifications are then used to find a new solution, which is used to update the quantifications, and so on, until some criterion is reached that signals the process to stop.

4.2. Dimensionality

The dimensionality p is important in this study, because choosing a different dimensionality will lead to different transformations of the data as we can no longer assume the solutions to be nested when we require category points to be on a line. Multiple nominal variables have different quantifications in p dimensions. In general, the categories of non-multiple-nominal variables fit on a straight line. If there are no dependencies in the data, the maximum number of dimensions, when the first α variables are multiple nominal, is

$$\sum_{j=1}^{\alpha}(k_j - 1) + (m - \alpha) \tag{29}$$

In the case of non-multiple-nominal variables only, people use the first p dimensions for which the eigenvalues $\lambda_j \,(j = 1,\ldots,p)$ of the correlation matrix between the quantified variables is larger than $1/m$. If a dimension has an eigenvalue smaller than $1/m$, it explains less variance than an individual variable; such a dimension has little or no generalizability. If there are multiple nominal variables, there is no easy rule. It remains true that if an eigenvalue is smaller than the reciprocal of the maximum number of dimension, the corresponding dimension has little generalizability and could better be discarded.

4.3. Optimally-Scaled Analysis

We suppose our observations are measured on a set of nominal, ordinal or numerical variables or any mixture of these variables collected in a data matrix X $(n \times p)$ in which the rows correspond to n objects (or individuals) measured on m columns (variables). Suppose for the $(m-\alpha)$ non-multiple nominal variables that $q_j = G_j y_j$ and $q_j : I \to R$ is a function that assigns quantifications to the categories of variable v_j, such that the transformed categories $q_j = q_j v_j$ are optimally scaled. This produces a p-dimensional matrix X of scores for all objects where X is the ortho-normalized (on Ω) version of

$$\tilde{X} = \frac{1}{m}\left(\sum_{j=1}^{\alpha} G_j Y_j + \sum_{j=\alpha+1}^{m-\alpha} q_j a_j^t\right). \tag{30}$$

We define scores in X with means zero, and uncorrelated dimensions for the n observations. Then we propose to use nonparametric discriminant procedures to estimate the unknown distribution of the object scores X which are the (standardized) averages over quantified variables $q_j a_j^t$.

After transforming the mixed-data into a Euclidean space, the new data preserves the information as the one before the transformation. The only new characteristic of the transformed data is to be homogeneous, continuous, and free of any distribution. Once the nonlinear variables are transformed to a linear scale (Euclidean space), we apply SEM on the transformed continuous data to carry out our analysis under a multivariate distributional assumption. Of course, any distributional assumption requires mostly multivariate normality assumption at the outset as the starting point. But one should bear in mind that this may not be tenable always. What is unique and advantageous about our proposed approach is that, we do not need to assume that the observations are coming from a hidden continuous distribution with a threshold specification to analyze SEMs with categorical, dichotomous, and mixed data sets.

In the next section, Section 5, we discuss the information theoretic model selection criteria we use in this paper and give their derived forms for SEMs.

5. Information Theoretic Model Selection Criteria and Complexity Measure

As is well known model selection is arguably one the most fundamental problems in scientific research. One of the major difficulties in statistical modeling of data is the choice of an appropriate model that fits the data. The main purpose of model evaluation is to *"understand"* the observed data. Researchers and practitioners alike seek to learn the model and study the quality of the model by a process which is called statistical model identification or selection. In recent years, in the literature, the necessity of introducing the concept of model selection has been recognized and the problem is posed in how to choose the *"best approximating"* model among a class of competing models with different numbers of parameters by a suitable model selection criterion given a data set. Also, there is presently a great deal of interest in simple criteria represented by parsimony of parameters for choosing one of a set of competing models to describe a given data set. The general principle for parsimony is that a simpler model is preferable to a more complex one, known as Occam's razor, named after the English Franciscan friar William of Ockham who is known in the scientific community by his famous razor. Occam's razor states that *"entities should not be multiplied beyond necessity."* It advocates the simplest possible explanation for the data that we have observed, but no simpler explanation than that (see, e.g., Bozdogan [2, 7], Paquet [55], and others.).

Therefore, to operationalize these concepts in statistical modeling and model evaluation problems, the concept of information theoretic or entropic underpinning and a measure of complexity plays an important role. At the philosophical level, complexity involves not just the number of parameters but also the notion of connectivity patterns and the interdependencies of the estimated parameters of the model and its model components. Without a measure of *"overall"* model complexity, prediction of model behavior and assessing model quality is difficult. This requires detailed statistical analysis and computation to choose the best fitting model among a portfolio of competing models for a given finite sample Bozdogan [2].

Based on Akaike's [1] classic AIC, in the literature many model selection procedures that take the form of a *penalized likelihood* (a *negative log likelihood* plus a *penalty term*) have been proposed.

For a general multivariate linear or nonlinear model defined by

Statistical Model=Signal+Noise,

a summary diagram for AIC and $ICOMP$ in terms of a loss function is given by

$$\left.\begin{array}{r} \left.\begin{array}{r} Lack\,of\;\;fit \\ Loss = +Lack\,of\;\;Parsimony \end{array}\right\} \Rightarrow AIC \\ +Profusion\,of\;\;Complexity \end{array}\right\} \Rightarrow ICOMP$$

5.1. Akaike's Information Criterion (AIC)

AIC is considered as the grandfather of all the information criteria. It was first developed and introduced by Akaike [1]. Since its introduction, AIC has had a fundamental impact on statistical model evaluation in scientific research. The introduction of AIC transformed the recognition of good modeling in statistics away from conventional hypothesis testing type procedures. As a result, many important statistical modeling techniques have been developed in various cross-disciplinary fields.

AIC as an extension of the maximum likelihood principle is a measure of the *"goodness of fit"* of a model to the data. Data with greater uncertainty will exhibit less information. Akaike's [1] *information criterion* (*AIC*) is defined by

$$AIC(k) = -2\log L(\hat{\theta}_k) + 2m(k), \tag{31}$$

where $L(\hat{\theta}_k)$ is the maximized likelihood function, $\hat{\theta}_k$ is the maximum likelihood estimate of the parameter vector θ_k under the model M_k, and $m(k)$ is the *number of free parameters* estimated within the model (the *penalty component*) which is a measure of complexity that compensates for the *bias* in the lack of fit when the *maximum likelihood estimators* (*MLEs*) are used. Therefore, in *AIC*, the compromise takes place between the maximized log likelihood, i.e., $-2\log L(\hat{\theta}_k)$ (the *lack of fit component*) and $m(k)$, the *number of free parameters* estimated within the model (the *penalty component*).

For an untrained eye, the AIC can be easily explained as a punishment fitness function, because it punishes the model for any variation that the model fails to explain, and at the same time, for any additional number of parameters that the model employs. The minimum of AIC is chosen to indicate the best fitting model.

5.2. Consistent Akaike's Information Criterion (CAIC)

Since it is well known that AIC has several disadvantages, i.e., there is an overfitting tendency and there is no model selection consistency, to penalize the overparametrization more strongly, Bozdogan [2] improved and extended AIC analytically in several ways. These extensions make AIC asymptotically consistent, and that overparameterization is penalized more stringently to pick the simplest of the true models whenever there is nothing to be lost in doing so. Here, we only give one of the forms of these extensions.

The *Consistent* AIC is defined by

$$CAIC(k) = -2\log L(\hat{\theta}_k) + m(k)[\log(n) + 1], \tag{32}$$

where $\log(n)$ denotes the *"natural logarithm"* of the sample size n, and $m(k)$ is the *number of free parameters* estimated within the model (the *penalty component*) when M_k is the model.

Similar to AIC, the minimum of $CAIC$ is chosen to be the best fitting model among a portfolio of competing alternative models. Note that $CAIC$ punishes the model more stringently than does AIC.

5.3. Information Complexity (ICOMP) Criteria

To measure how the parameter estimates are correlated with one another in the model fitting process, that is, to take into account the *"Profusion of Complexity"*, and to guard against model misspecification of structural relationships, Bozdogan [4, 6, 7] introduced several forms of a new class of information-theoretic measure of complexity criterion called *ICOMP* as a decision rule for model selection in statistical modeling to help provide new approaches relevant to statistical inference. In *ICOMP*, *I* is for *information* and *COMP* for *complexity* to distinguish it from other non-information theoretic complexity measures.

The development of *ICOMP* has been motivated in part by Akaike's [1] classic AIC. However, in contrast to AIC, *ICOMP* is based on the *structural complexity* of an element or set of random vectors via a generalization of the *information-based covariance complexity index* of van Emden [56].

5.3.1. ICOMP as an Approximation to the Sum of Two Kullback-Leibler Distances

Instead of penalizing the number of free parameters directly, *ICOMP* penalizes the covariance complexity of the model. It is defined by

$$ICOMP = -2\log L(\hat{\theta}_k) + 2C(\hat{\Sigma}_{Model}), \tag{33}$$

where $L(\hat{\theta}_k)$ is the maximized likelihood function, $\hat{\theta}_k$ is the maximum likelihood estimate of the parameter vector θ_k under the model M_k, and C represents a real-valued complexity measure and $\hat{\Sigma}_{Model} = \hat{C}ov(\hat{\theta}_k)$ represents the estimated covariance matrix of the parameter vector of the model. This covariance matrix in *ICOMP* is estimated several ways. One of the ways to estimate this covariance matrix is to use the celebrated *Cramer-Rao lower bound (CRLB)* matrix through its inverse. That is, the *estimated inverse Fisher information matrix* (*IFIM*) $\hat{\mathcal{F}}^{-1}$ of the model given by

$$\hat{\mathcal{F}}^{-1} = \left\{ -E\left(\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'} \right)_{\hat{\theta}} \right\}^{-1}, \tag{34}$$

where the expression in bracket is the $(s \times s)$ matrix of second partial derivatives of the log-likelihood function of the fitted model evaluated at the maximum likelihood estimators $\hat{\theta}$. For this, see, e.g., Cramér [57] and Rao [58-60].

The estimated *IFIM* provides us an achievable accuracy of the parameter estimates by considering the entire parameter space of the model. *IFIM* is a measure of the best precision with which a parameter can be estimated from statistical data. It measures the quantum of information and measures the curvature of the log likelihood function of the model. The diagonal elements of *IFIM* contain the *estimated variances* or *squared standard errors* of the estimated parameters, while the off-diagonals of the matrix contain their covariances.

In its general form, for univariate and multivariate models (linear *and/or* nonlinear) *ICOMP* is defined by

$$ICOMP = -2\log L(\hat{\theta}_k) + 2C_1(\hat{\mathcal{F}}^{-1}), \tag{35}$$

where

$$C_1(\hat{\mathcal{F}}^{-1}) = \frac{s}{2}\log\left[\frac{tr\hat{\mathcal{F}}^{-1}}{s}\right] - \frac{1}{2}\log\left|\hat{\mathcal{F}}^{-1}\right| \tag{36}$$

is the maximal information complexity of the estimated *inverse Fisher information matrix* (*IFIM*) of the model, and where $s = \dim(\hat{\mathcal{F}}^{-1}) = rank(\hat{\mathcal{F}}^{-1})$.

The use of $C_1(\hat{\mathcal{F}}^{-1})$ in the information-theoretic model evaluation criteria takes into account the fact that as we increase the number of free parameters in a model, the accuracy of the parameter estimates decreases. As preferred according to *the principle of parsimony*, $ICOMP(IFIM)$ chooses simpler models that provide more accurate and efficient parameter estimates over more complex, overspecified models.

We note that, the trace of $IFIM$ in the complexity measure involves only the diagonal elements analogous to *variances* while the determinant involves also the off-diagonal elements analogous to *covariances*. Therefore, $ICOMP(IFIM)$ contrasts the *trace* and the *determinant* of $IFIM$, and this amounts to a comparison of the *geometric* and *arithmetic means* of the eigenvalues of $IFIM$ given by

$$ICOMP(IFIM) = -2\log L\left(\hat{\theta}_M\right) + s\log\left(\overline{\lambda}_a / \overline{\lambda}_g\right), \tag{37}$$

where $s = \dim\hat{\mathcal{F}}^{-1}(\hat{\theta}) = rank\hat{\mathcal{F}}^{-1}(\hat{\theta})$, and where $\overline{\lambda}_a$ is the *arithmetic mean* and $\overline{\lambda}_g$ is the *geometric mean* of the eigenvalues of $\hat{\mathcal{F}}^{-1}$.

We note that $ICOMP(IFIM)$ now looks in appearance like the $CAIC$ of Bozdogan [2], Rissanen's [61] MDL, and Schwarz's [62] Bayesian criterion SBC, except for using $\log\left(\overline{\lambda}_a / \overline{\lambda}_g\right)$ instead of using $\log(n)$ denotes the natural logarithm of the sample size n.

A model with minimum $ICOMP(IFIM)$ is chosen to be the best among all possible competing alternative models.

With $ICOMP(IFIM)$, complexity is viewed not as the number of parameters in the model, but as the *degree of interdependence* (i.e. the *correlational structure among the parameter estimates*). By defining complexity in this way, $ICOMP(IFIM)$ provides a more judicious penalty term than AIC, MDL, SBC, or $CAIC$. The lack of parsimony and the profusion of complexity are automatically adjusted by $C_1(\hat{\mathcal{F}}^{-1})$ across the competing alternative portfolio of models as the parameter spaces of these models are constrained in the model selection process.

5.3.2. ICOMP as an Estimate of Posterior Expected Utility: $ICOMP_{PEU}$

In the literature, the idea of using two utility functions U_1 and U_2 that are multiplied to define a utility U whose *posterior expectation* is (approximately) maximized to select a model was considered notably by Poskitt [63], and others. If we relate utility U_1 to the *lack of fit component* of the model and U_2 to the *complexity of the parameter space of the model,* i.e., *the dimension of the model*, we introduce a new *ICOMP* class of criteria as a Bayesian criterion in maximizing a *posterior expected utility(PEU)* following the results from Bozdogan and Haughton [28].

ICOMP as a Bayesian criterion in maximizing a posterior expected utility (PEU) is given by

$$ICOMP\left(IFIM\right)_{PEU} = -2\log L\left(\hat{\theta}_M\right) + k + 2C_1\left(\hat{\mathcal{F}}^{-1}\left(\hat{\theta}_M\right)\right). \tag{38}$$

The decision rule is to choose the minimum of $ICOMP\left(IFIM\right)$ over the class of models $M_k,\ k = 1, 2, \ldots, K$ that is the best fitting model.

5.3.3. ICOMP for Misspecified Models: $ICOMP_{PEU\_Miss}$

Suppose that the fitted model is the wrong or misspecified model.

Different choices of utility U_2 may depend on other characteristics that a researcher can consider on the parameter vector θ_M if the model M is under consideration. Therefore, the full specification of the form of the utility function U_2 is important. By defining different forms of the utility U_2 we can, therefore, obtain other forms of $ICOMP\left(IFIM\right)_{PEU}$ that give us many useful class of model selection criteria one of which is the form of $ICOMP$ in the case of a misspecified model. Therefore, the choice of the utility

$$U_2 = \exp\left[-tr\left(\hat{\mathcal{F}}^{-1}\hat{\mathcal{R}}\right) - C_1\left(\hat{\mathcal{F}}^{-1}\right)\right] \tag{39}$$

would lead to

$$\begin{aligned} ICOMP\left(IFIM\right)_{PEU\_Miss} &= -2\log L\left(\hat{\theta}_M\right) + k + 2tr\left(\hat{\mathcal{F}}^{-1}\hat{\mathcal{R}}\right) + 2C_1\left(\hat{\mathcal{F}}^{-1}\left(\hat{\theta}_M\right)\right) \\ &= -2\log L\left(\hat{\theta}_M\right) + k + 2\left[tr\left(\hat{\mathcal{F}}^{-1}\hat{\mathcal{R}}\right) + C_1\left(\hat{\mathcal{F}}^{-1}\left(\hat{\theta}_M\right)\right)\right]. \end{aligned} \tag{40}$$

Note that $tr\left(\hat{\mathcal{F}}^{-1}\hat{\mathcal{R}}\right)$ is the well-known Lagrange-multiplier test statistic. See, for example, Takeuchi [64], Hosking [65], and Shibata [66].

We can approximate $tr\left(\hat{\mathcal{F}}^{-1}\hat{\mathcal{R}}\right)$ by

$$tr\left(\hat{\mathcal{F}}^{-1}\hat{\mathcal{R}}\right) \cong \frac{nk}{n-k-2}. \tag{41}$$

This corrects the bias for small as well as large sample sizes if the model is misspecified. Hence, (40) becomes

$$\begin{aligned} ICOMP\left(IFIM\right)_{PEU\_Miss} &= -2\log L\left(\hat{\theta}_M\right) + k + 2tr\left(\hat{\mathcal{F}}^{-1}\hat{\mathcal{R}}\right) + 2C_1\left(\hat{\mathcal{F}}^{-1}\left(\hat{\theta}_M\right)\right) \\ &= -2\log L\left(\hat{\theta}_M\right) + k + 2\left(\frac{nk}{n-k-2}\right) + 2C_1\left(\hat{\mathcal{F}}^{-1}\left(\hat{\theta}_M\right)\right). \end{aligned} \quad (42)$$

If the model is correctly specified, then

$$tr\left(\hat{\mathcal{F}}^{-1}\hat{\mathcal{R}}\right) = tr(I_k) = k. \quad (43)$$

Therefore, $ICOMP\left(IFIM\right)_{PEU\_Miss}$ reduces to

$$\begin{aligned} ICOMP\left(IFIM\right)_{PEU\_AIC_3} &= -2\log L\left(\hat{\theta}_M\right) + k + 2k + 2C_1\left(\hat{\mathcal{F}}^{-1}\left(\hat{\theta}_M\right)\right) \\ &= -2\log L\left(\hat{\theta}_M\right) + 3k + 2C_1\left(\hat{\mathcal{F}}^{-1}\left(\hat{\theta}_M\right)\right) \\ &= AIC_3 + 2C_1\left(\hat{\mathcal{F}}^{-1}\left(\hat{\theta}_M\right)\right). \end{aligned} \quad (44)$$

For more details on these, we refer the readers to Bozdogan and Haughton [28], and Bozdogan [6, 7].

6. Derived Computational Forms of Information Criteria in SEMs

In this section we show the derived forms of the information theoretic model selection criteria we discussed in Section 5 to trade off between parsimony and fit in structural equation modeling (SEM) and how this trade off affects the scoring of model selection to rank the models according to the minimum of the model selection criteria. It is well known that the number of free model parameters influences complexity. In SEM even if the models may have the same number of parameters but different structures, such models can posses different complexities. With this in mind, our goal is to score the models with model selection criteria and study the models in terms of their generalizability.

Under the assumption that the observed variables are continuous and have interval scales, and multivariate Gaussianity, i.e.,

$$z = (y', x')' \sim N_{(p+q)}\left(0, \Sigma(\theta)\right) \ (multivariate\ normal) \quad (45)$$

and using the maximum likelihood estimators, we have

- *Akaike's* [1] *information criterion* (*AIC*) for the *structural equation model* (*SEM*) given by

$$AIC(SEM) = n(p+q)\log(2\pi) + n\log\left|\hat{\Sigma}(\hat{\theta})\right| + ntr\left(S\hat{\Sigma}^{-1}(\hat{\theta})\right) + 2k, \quad (46)$$

where k is the number of parameters.

- *Consistent AIC* (*CAIC*) (Bozdogan[2]) for *SEM* is:

$$CAIC(SEM) = n(p+q)\log(2\pi) + n\log\left|\hat{\Sigma}(\hat{\theta})\right| + ntr\left(S\hat{\Sigma}^{-1}(\hat{\theta})\right) + k[\log(n)+1]. \tag{47}$$

To derive $ICOMP\left(IFIM\right)_{PEU}$ and $ICOMP\left(IFIM\right)_{PEU\_Miss}$ we need to compute the *trace* and the *determinant* of *IFIM* given in (23), so that we can score the complexity measure $C_1(\hat{\mathcal{F}}^{-1})$. From Magnus [40] and Magnus and Neudecker [41], and following the results in Bozdogan and Bearse [29], we have

$$\begin{aligned} tr\left(\hat{\mathcal{F}}^{-1}\right) &= \frac{1}{n}tr(\hat{\Sigma}) + tr\left[\frac{2}{n}D^{+}{}_{(p+q)}(\hat{\Sigma}\otimes\hat{\Sigma})D^{+\prime}{}_{(p+q)}\right] \\ &= \frac{1}{n}tr(\hat{\Sigma}) + \frac{2}{n}\left[\frac{1}{4}tr(\hat{\Sigma}^2) + \frac{1}{4}\left(tr(\hat{\Sigma})\right)^2 + \frac{1}{2}\sum_{j=1}^{p+q}(\hat{\sigma}_{jj})^2\right] \\ &= \frac{1}{n}tr(\hat{\Sigma}) + \frac{1}{2n}\left[tr(\hat{\Sigma}^2) + \left(tr(\hat{\Sigma})\right)^2 + 2\sum_{j=1}^{p+q}(\hat{\sigma}_{jj})^2\right], \end{aligned} \tag{48}$$

and

$$\begin{aligned} \left|\hat{\mathcal{F}}^{-1}\right| &= \left|\frac{1}{n}\hat{\Sigma}\right| \bullet \left|\frac{2}{n}D^{+}{}_{(p+q)}(\hat{\Sigma}\otimes\hat{\Sigma})D^{+\prime}{}_{(p+q)}\right| \\ &= \left(\frac{1}{n}\right)^{p+q}\left|\hat{\Sigma}\right| \bullet \left|\frac{2}{n}D^{+}{}_{(p+q)}(\hat{\Sigma}\otimes\hat{\Sigma})D^{+\prime}{}_{(p+q)}\right| \\ &= \left(\frac{1}{n}\right)^{p+q}\left|\hat{\Sigma}\right| \bullet \left(\frac{1}{n}\right)^{\frac{1}{2}(p+q)(p+q+1)}\left|2D^{+}{}_{(p+q)}(\hat{\Sigma}\otimes\hat{\Sigma})D^{+\prime}{}_{(p+q)}\right|. \end{aligned} \tag{49}$$

Since

$$\left|2D^{+}{}_{(p+q)}(\hat{\Sigma}\otimes\hat{\Sigma})D^{+\prime}{}_{(p+q)}\right| = 2^{-1/2(p+q)(p+q-1)}\left|\hat{\Sigma}\right|^{p+q+1}, \tag{50}$$

then

$$\left|\hat{\mathcal{F}}^{-1}\right| = \left(\frac{1}{n}\right)^{p+q}\left|\hat{\Sigma}\right| \bullet \left(\frac{1}{n}\right)^{\frac{1}{2}(p+q)(p+q+1)} \bullet 2^{-1/2(p+q)(p+q-1)}\left|\hat{\Sigma}\right|^{p+q+1}. \tag{51}$$

Simplifying (51) further, we have

$$\left|\hat{\mathcal{F}}^{-1}\right| = \left|\hat{\Sigma}\right|^{p+q+2}\left(\frac{1}{n}\right)^{(p+q)+\frac{1}{2}(p+q)(p+q+1)} 2^{-1/2(p+q)(p+q-1)}. \tag{52}$$

Now taking the natural log of (52), we obtain

$$\begin{aligned}\log\left|\hat{\mathcal{F}}^{-1}\right| &= (p+q+2)\log\left|\hat{\Sigma}\right| + \left[(p+q)+\frac{1}{2}(p+q)(p+q+1)\right]\log\left(\frac{1}{n}\right) \\ &\quad -\frac{1}{2}(p+q)(p+q-1)\log(2) \\ &= (p+q+2)\log\left|\hat{\Sigma}\right| - \left[(p+q)+\frac{1}{2}(p+q)(p+q+1)\right]\log(n) \\ &\quad -\frac{1}{2}(p+q)(p+q-1)\log(2).\end{aligned} \tag{53}$$

Using the definition of complexity in (36), we obtain a computationally convenient form of the expression for $C_1(\hat{\mathcal{F}}^{-1})$ given by

$$\begin{aligned}C_1\left(\hat{\mathcal{F}}^{-1}\right) &= \frac{s}{2}\log\left[\frac{\frac{1}{n}tr(\hat{\Sigma})+\frac{1}{2n}\left[tr(\hat{\Sigma}^2)+\left(tr(\hat{\Sigma})\right)^2+2\sum_{j=1}^{p+q}(\hat{\sigma}_{jj})^2\right]}{s}\right] \\ &\quad -\frac{1}{2}(p+q+2)\log\left|\hat{\Sigma}\right| + \frac{1}{2}\left[(p+q)+\frac{1}{2}(p+q)(p+q+1)\right]\log(n) \\ &\quad +\frac{1}{4}(p+q)(p+q-1)\log(2)\end{aligned} \tag{54}$$

which requires only the computation of traces and determinants and that $C_1(\hat{\mathcal{F}}^{-1})$ avoids the construction of the full *IFIM.* Note that all the required inputs to (54) are readily available as part of the standard output of most *SEM* packages or can be programmed in *MATLAB* or *R* macro language. This saves computational time when the *IFIM* is a very large matrix when we have high-dimensional multivariate X-Y data which is very attractive. Further, it shows the scalability property of the complexity measure.

Hence, $ICOMP\left(IFIM\right)_{PEU}$ for the *SEM* is given by

$$\begin{aligned}ICOMP\left(IFIM\right)_{PEU} &= n(p+q)\log(2\pi)+n\log\left|\hat{\Sigma}(\hat{\theta})\right|+ntr\left(S\hat{\Sigma}^{-1}(\hat{\theta})\right) \\ &\quad +k+2C_1\left(\hat{\mathcal{F}}^{-1}\right)\end{aligned} \tag{55}$$

with $C_1(\hat{\mathcal{F}}^{-1})$ given in (54) in its scalar open form.

Similarly, $ICOMP\left(IFIM\right)_{PEU\_Miss}$ for *SEM* is given by

$$\begin{aligned}ICOMP\left(IFIM\right)_{PEU\_Miss} &= n(p+q)\log(2\pi)+n\log\left|\hat{\Sigma}(\hat{\theta})\right|+ntr\left(S\hat{\Sigma}^{-1}(\hat{\theta})\right) \\ &\quad +k+2\left(\frac{ns}{n-s-2}\right)+2C_1\left(\hat{\mathcal{F}}^{-1}\right).\end{aligned} \tag{56}$$

Comparing *AIC*, *CAIC*, and the two forms of *ICOMP(IFIM)* criteria, we see that the difference between these criteria are in the crutial penalty term.

Next, we show a real numerical example using our approach on a benchmark data set.

7. A Real Numerical Example: Structural Equation Modeling and Analysis of Quality of Life (QOL) Data

According to Wikipedia, the term *quality of life* (*QOL*) is used to evaluate the general well-being of individuals and societies. This term has been used in a wide range of contexts and it should not be confused with the concept of standard of living. Here our aim is to model the quality of life (QOL) in healthcare.

The Wolrd Health Organization Quality of Life (WHOQOL) project was initiated in 1991. The aim was to develop an international cross-culturally comparable quality of life assessment instrument.

Quality of life (QOL) is a broad multidimensional concept that usually includes subjective evaluations of both positive and negative aspects of life. The concept of health related quality of life (HRQOL) and its determinants have evolved since the 1980s to cover those overall quality of life that can be clearly shown to affect health. It has become an important concept for health care. Focusing on HRQOL as a national standard can bridge boundaries between disciplines and between social, mental, and medical services. Therefore, HRQOL has great value for clinical work and the planning and evaluation of health care as well as for medical research.

QOL has generally been accepted as a multidimentional concept that is best modeled by a number of latent constructs.

Items in a QOL instrument comprises 26 items of questions Q1 to Q26 measure the broad domains such as *physical health*, *psychological health*, *social relationships*, and *environment*. All of the items are measured on an ordinal categorical scale with 1 to 5 points.

Our data set for structural equation modeling (SEM) of the QOL is taken from Lee [9] and it consists of n=338 observations on 26 categorical variables (questions). Figure 3 shows frequency distribution of the questions in the WHOQOL data set.

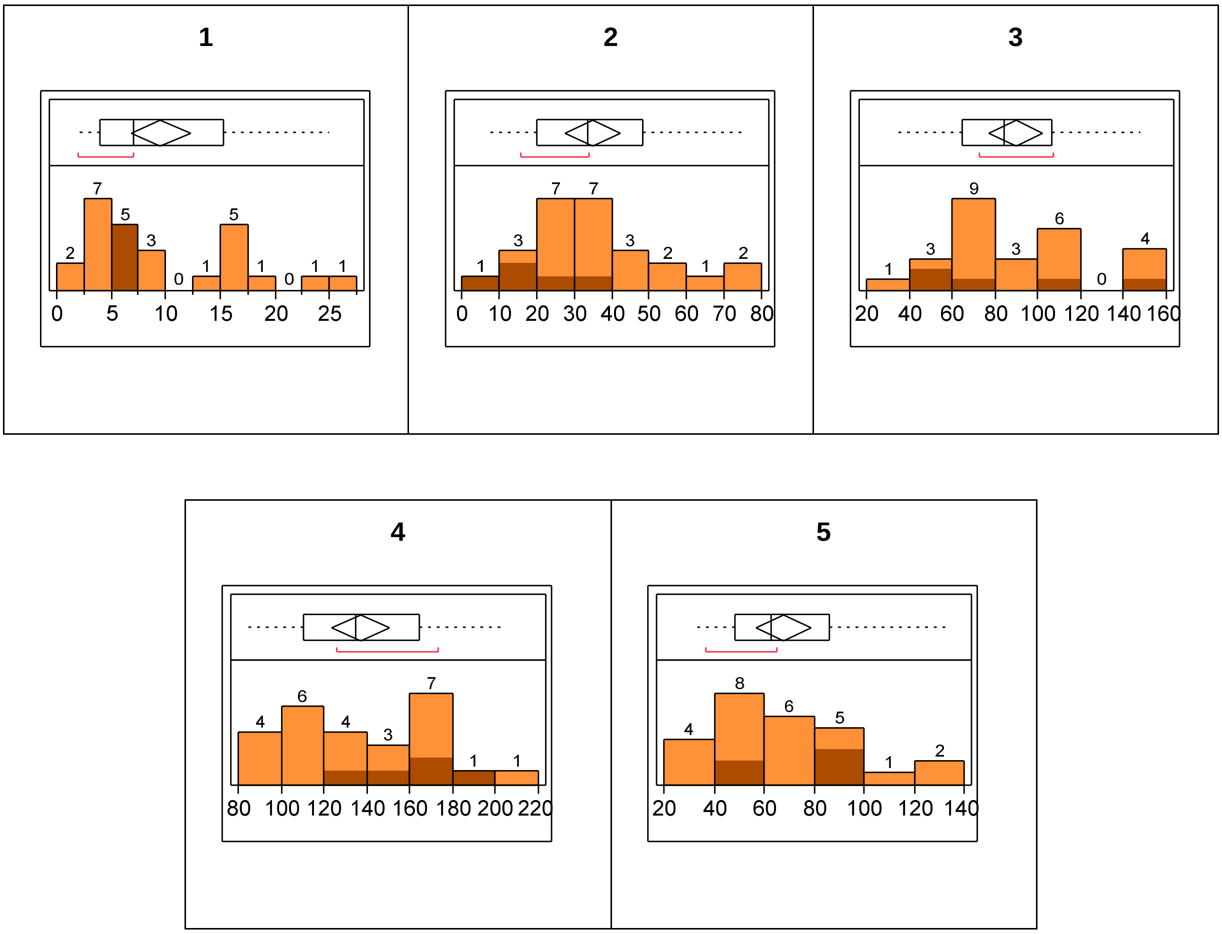

Figure 3 Box Plots and Frequency Distribution of the Questions in the WHOQOL Data Set

Figure 4 shows the frequency distributions of the first 12 questions Q1-Q12 out of 26 questions.

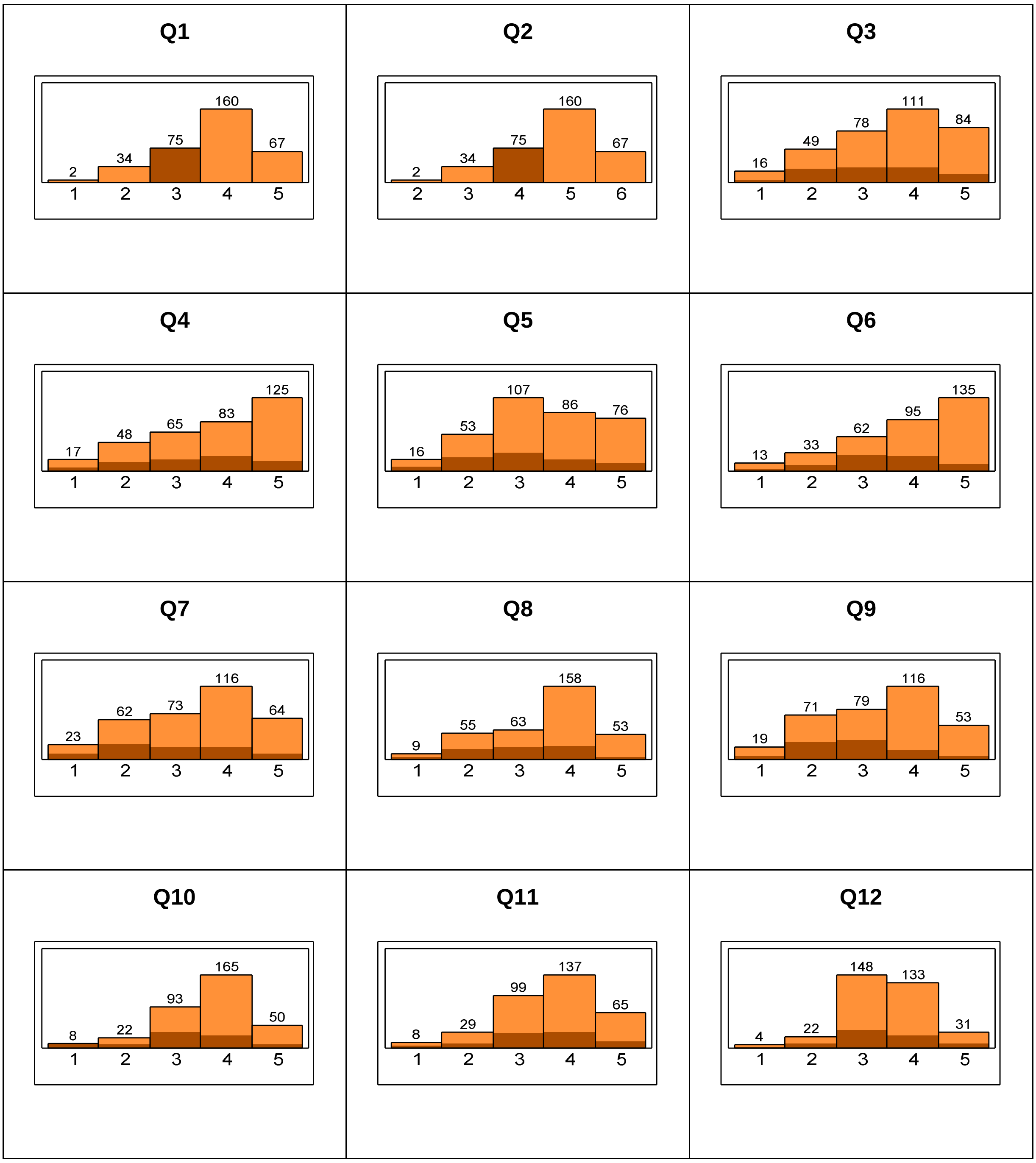

Figure 4 Frequency Distributions of the First 12 Questions Q1 to Q12.

The description of the 26 questions for the QOL data is given in Table 1.

In his analysis, Lee [9], treating that the QOL data as observations coming from a hidden continuous normal distribution with a fixed threshold specification, introduced a Bayesian approach to fit SEM to QOL data set despite the fact that non-rigorous treatments of the ordinal items as continuous can be subjected to criticism and can lead to erroneous conclusions and errors. Also, it is not clear how the discrete nature of the data is taken into account to analyze this ordered categorical data set within the Bayesian framework.

Table 1 The Description of the Questions Q1-Q26 for the Quality of Life (QOL) Data

Q1=Overall Quality of Life (QOL)	Q14=Self-esteem
Q2=Overall health	Q15=Negative feeling
Q3=Pain and discomfort	Q16=Personal relationship
Q4=Medical treatment dependence	Q17=Sexual activity
Q5=Energy and fatigue	Q18=Social support
Q6=Mobility	Q19=Phsical safety and security
Q7=Sleep and rest	Q20=Physical environment
Q8=Daily activities	Q21=Financial resources
Q9=Work capacity	Q22=Daily life information
Q10=Positive feeling	Q23=Participation in leisure activity
Q11=Sprituality/personal beliefs	Q24=Living condition
Q12=Memory and concentration	Q25=Health accessibility and quality
Q13=Bodily image and appearance	Q26=Transportation

In this paper, we analyzed the same data set regarding the quality of life (QOL) with our novel Gifi transformation approach without assuming that the ordinal categorical data as observations coming from a hidden continuous distribution.

We also considered four latent variables of the instrument WHOQOL-100 (Power, Bullingen and Hazper [67]). The first item (Q1) is an ***observed endogenous variable*** intended to address one ***latent endogenous variable named QOL***. The seven items (Q3 to Q9) are intended to address ***physical health***, the next six items (Q10 to Q15) are intended to address ***psychological health***, the next three items (Q16 to Q18) are intended to address the ***social relationship***, and the last eight items (Q19 to Q26) are intended to address the ***environment***.

To be able to model and analyze this data set rigorously using SEM, first, we transformed the categorical variables using the Gifi transformation to continuous variables by our newly developed Gifi MATLAB module. See, e.g., Katragada [53], Katragada and Bozdogan [54]. After the Gifi transformations, we then used LISREL to fit SEMs to both the original categorical data and the continuous Gifi transformed data. We scored four of the above information criteria on the fitted models. Namely, we scored *AIC*, *CAIC*, $ICOMP\left(IFIM\right)_{PEU}$, and $ICOMP\left(IFIM\right)_{PEU\_Miss}$ on Model on Continuous Data and Model on DiscreteData. The description of these two models is given as follows.

Model on Continuous Data: Full structural equation model with 1 observed endogenous variable, 24 observed exogenous variables, 1 latent endogenous variable, 4 latent exogenous variables fit to categorical data with 59 free parameters.

Model on DiscreteData: Full structural equation model with 1 observed endogenous variable, 24 observed exogenous variables, 1 latent endogenous variable, 4 latent exogenous variables fit to discrete data with 59 free parameters.

Our results are given Table 2.

Table 2 Information Criteria Scores for Gifi Transformed and Discrete Data Sets

Models	*AIC*	*CAIC*	$ICOMP(IFIM)_{PEU}$	$ICOMP(IFIM)_{PEU\_Miss}$
Continuous Data	**1160**	**11444**	**14672**	**14816**
Discrete Data	20621	20905	22967	23111

According to the results in Table 2, we see that the model on continuous Gifi transformed data set is better than the model using the original data discrete data set based on the minimum of the information criteria.

Next, we specify different SEM sub models models to fit to the Gifi transformed data set and compare them with the full saturated SEM. Model 1 is the full SEM which includes all observed endogenous and exogenous variables.

Model 1: Full structural equation model consist of 2 observed endogenous variables, 24 observed exogenous variables, 1 latent endogenous variable, 4 latent exogenous variables with 50 free parameters. The structure of Model 1 is given by

$$\begin{aligned} x_{(24\times1)} &= \Lambda_{x(24\times4)}\xi_{(4\times1)} + \delta_{(24\times1)} \\ y_{(1\times1)} &= \Lambda_{y(1\times1)}\eta_{(1\times1)} + \varepsilon_{(1\times1)} \\ \eta_{(1\times1)} &= \Gamma_{(1\times4)}\xi_{(4\times1)} + \zeta_{(1\times1)} \end{aligned} \quad (57)$$

The estimated structural relationship between latent exogenous variables (ξ_1, ξ_2, ξ_3, and ξ_4) and the latent endogenous variable (η) is

$$\eta = \underset{(3.13)}{2.27}\xi_1 + \underset{(0.19)}{0.09}\xi_2 + \underset{(1.08)}{0.92}\xi_3 - \underset{(-0.26)}{0.14}\xi_4 + \zeta \quad (58)$$

In (58), the numbers in the parenthesis indicate the Wald statistics. According to these values, with $\alpha = 0.05$, only physical health (ξ_1) has a significant effect on QOL.

Psychological health (ξ_2), social relationship (ξ_3) and environment (ξ_4) are not significant at $\alpha = 0.05$ level of significance.

Model 2: Sub structural equation model with 2 observed endogenous variables, 21 observed exogenous variables, 1 latent endogenous variable, 3 latent exogenous variables with 49 free parameters is given by

$$\begin{aligned} x_{(21\times1)} &= \Lambda_{x(21\times3)}\xi_{(3\times1)} + \delta_{(21\times1)} \\ y_{(1\times1)} &= \Lambda_{y(1\times1)}\eta_{(1\times1)} + \varepsilon_{(1\times1)} \\ \eta_{(1\times1)} &= \Gamma_{(1\times3)}\xi_{(3\times1)} + \zeta_{(1\times1)} \end{aligned} \quad (59)$$

For **Model 2**, we excluded latent exogenous variable social relationship, because if we excluded psychological health or environment, the other latent exogenous variables were not significant to study QOL.

The estimated structural equation for Model 2 is

$$\eta = \underset{(5.28)}{1.66}\xi_1 + \underset{(2.37)}{0.53}\xi_2 + \underset{(1.98)}{0.38}\xi_4 + \zeta. \qquad (60)$$

According to Model 2, physical health, psychological health and environment are significant to study QOL.

Model 3: Is a measurement model with 24 observed exogenous variables and 4 latent exogenous variables with 54 free parameters given by

$$x_{(24\times1)} = \Lambda_{x(24\times4)}\xi_{(4\times1)} + \delta_{(24\times1)} \qquad (61)$$

In Model 3, we analyzed the measurement model only with observed exogenous variables. Again, we used four latent exogenous variables in this model.

The results from fitting these three SEMs to select the best fitting model and the scores of four information criteria are summarized in Table 3.

Table 3 Scores of the Information Criteria for Models 1, 2, and 3

Models	*No. of parameters*	*No. of free parameters*	*AIC*	*CAIC*	$ICOMP(IFIM)_{PEU}$	$ICOMP(IFIM)_{PEU\_Miss}$
Model 1	325	59	1160	11444	14672	14816
Model 2	**253**	**49**	**10402**	**10639**	**12793**	**12908**
Model 3	200	54	10598	10858	13277	13407

Looking at Table 3, we see that all the criteria are minimized at Model 2. We would therefore select Model 2 as the best fitting structural equation model (SEM). Note that, the information criteria do not choose Model 1, which is the full saturated SEM, indicating that it is an inferior model to address the relations of QOL with latent constructs.

8. Conclusions and Discussion

Analysis of categorical, dichotomous, and mixed data sets is one of the most difficult and challenging problems in structural equation modeling (SEM). This is due to the fact that, presence of categorical and dichotomous variables violate the assumptions of continuity and multivariate normality that are needed in SEM.

In this paper we introduced and developed a novel and computationally feasible approach to analyze categorical, dichotomous, and mixed data sets in SEMs to resolve the current existing problems using the novel Gifi [25] system or transformations to analyze categorical, binary, and mixed-data sets in SEM research. Gifi system uses optimal scaling of the categorical variables to analyze data that are difficult or impossible for the usual standard statistical techniques to handle.

In addition to the introduction of the Gifi system or transformation, in this paper we also introduced and derived information-theoretic model selection criteria such as *Akaike's* [1] *classic information criterion* (*AIC*), *Bozdogan's* [2] *Consistent AIC*, called *CAIC*, and the *information-theoretic measure of complexity ICOMP* criterion of Bozdogan [3-7] to choose the best fitting SEMs. The model with the minimum values of the criteria is selected as the best fitting model among a candidate of portfolio of models.

We applied our proposed approach to model and analyze the quality of life (QOL) data in heathcare by showing the generalizability and flexibility of our approach over the currently used threshold modeling techniques where the categorical data are treated as coming from a hidden continuous normal distribution with a specified arbitrary threshold.

Based on our numerical results for the QOL data the results obtained with real categorical variables, the model which is generated from continuous Gifi transformed data is better than the model which is generated from the original categorical data based on the minimum of the information criteria. Furthermore, the advantage of our approach is that, we did not analyze the categorical data set by treating the data as coming from a univariate standard normal distribution. Also we do not have an arbitrary threshold specification. And, we did not assume that the normal distribution was hidden. Our approach in general takes into account the dimensionality of the data among the categories in the continuous Gifi space in the transformed data set. This allows us to use multivariate Gaussian, or other more general multivariate distributions in SEM framework.

For the QOL data set, the best SEM selected the Gifi transformed continuous data. It used 1 observed *endogenous variable*, 21 observed *exogenous variables*, 1 latent *endogenous variable*, and 3 *latent exogenous variables*.

We claim that the analysis and modeling based on the Gifi transformed continuous data, is much easier and more reliable than direct analysis of the categorical data set assuming a hidden univariate normal distribution.

As we note, Lee [9] also analyzed the QOL data using WinBUGS (a Bayesian program) through Bayesian treatment of the ordered categorical variables by fixing the thresholds at both ends in order to solve the identification problem. He estimated the other unknown thresholds simultaneously with the structural parameters. Lee [9] indicates that WinBUGS is not straightforward to apply to estimate the unknown thresholds and structural parameters simultaneously. When we compare our best fitting model Model 2 for QOL data set, although we did not follow a Bayesian approach, we see that our Model 2 is much more simple and parsimonious than that of Lee's [9] Bayesian models. In the Bayesian SEM, Lee also assumed the multivariate normality under the fitted models.

Although this is one of the caveats in the analysis of any multidimensional data set, and we are not always guaranteed to satisfy the multivariate Gaussian assumption, we can always use Mardia's [68-70] – test for multivariate normality to test our assumptions. This test consists of two tests – one for kurtosis, and one for skewness. To this end, we used Mardia's test to test the multivariate normality for the QOL data set. Figure 5 shows the kurtosis and skewness tests results for the QOL data set.

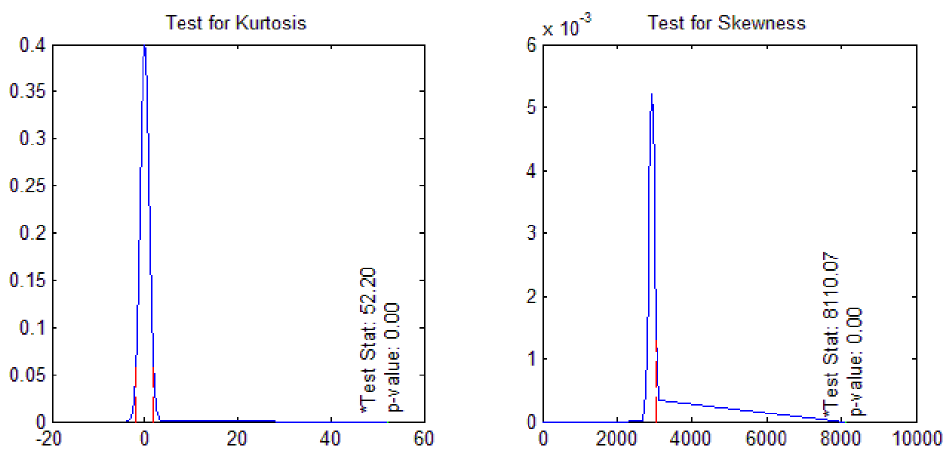

Figure 5 Mardia's [68-70] Multivariate Normality Test for the QOL Data Set

Under the null hypothesis of multivariate normality, the theoretical value for kurtosi is 675. The sample value of 883.63 suggests the data are more highly peaked than expected. The test statistic follows a standard normal distribution under the null hypothesis, and we see the sample test statistic of 52.2 is well outside of the ±1.96 critical range. Thus, we reject the null hypothesis of Gaussian kurtosis. In the left pane of Figure 5, we see the theoretical distribution with the red vertical bars indicating the critical region. The test statistic is indicated by *.

Under the null hypothesis, the skewness should be 0 and Mardia's test statistic follows a chi-square distribution. For this dataset, the sample skewness is 143.97, indicating more skew than expected. The sample test statistic 8110.07 is well out of the [0, 3051.93] critical range, and the p-value is much smaller than 0.0000. Thus, we also reject the null hypothesis of Gaussian skewness. These results are shown graphically on the right pane of Figure 5.

Since both tests strongly reject the null hypothesis, with p-values of <<0.000, we can safely reject the null hypothesis of multivariate Gaussian distribution. It appears that the QOL data are both more peaked and more skewed than expected.

To remedy such situations, our future goal in this direction will involve the extension of the proposed novel approach here under a more general multivariate distribution, such as the multivariate skewed power exponential (MVSPE), where we can take both skewness and kurtosis of the data into account. Alternatively, we can develop a new SEM under misspecification by relaxing the distributional assumptions on the model and utilize multivariate kernel density estimation (KDE) approach to SEM to robustify the model selection process in the presence of skewness and kurtosis. A first attempt toward this goal has been made in Deniz, Howe, and Bozdogan [71], and further work is in progress to be reported elsewhere.

In conclusion, we note that with the novel Gifi transformations we achieve an important advantage in that the results of the final best fitting model can be reverse-mapped to the original scale from the transformed scale since the Gifi transformation has a one-to-one mapping of the nonlinear values to the linear values. With the currently utilized techniques and the Bayesian methods, we do not have such benefits. Other advantages of our approach include the applicability of the proposed methods to many other multivariate data mining techniques, such as classification and clustering, probabilistic principle component analysis, mixture-model cluster analysis, factor analysis, discriminant analysis to mention a few, when the observed data are mixed, categorical, or purely binary.

Acknowledgements

This research for the first author was supported by the Scientific and Technological Council of Turkey (TUBITAK) as a Visiting Scholar at the University of Tennessee in Knoxville. For the second author, this research was partially supported by the prestigious *Jefferson Faculty Prize Award* at the University of Tennessee. Further, first and third authors extend their gratitude to Professor Bozdogan for his continued encouragement and hospitality while studying under him at the University of Tennessee in Knoxville. All the authors extend their gratitude and thanks to Prof. Dr. Eyüp Çetin, the Editor-in-Chief of the Istanbul University *Journal of the School of Business Administration (JSBA)*, for inviting us to make a contribution to the 40th year anniversary of *JSBA*. Authors extend their thanks to Dr. Kirkor Bozdogan, Senior Researcher at MIT, who read the draft of this paper very carefully and made some valuable editorial comments.

References

[1] H. Akaike, Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csáki (Eds.), *Second International Symposium on Information Theory*, Académiai Kiadó, Budapest, 267-281 (1973).

[2] H. Bozdogan, Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, 52, 3, 345-370 (1987).

[3] H. Bozdogan, Mixture-model cluster analysis using a new informational complexity and model selection criteria. In Multivariate Statistical Modeling, H. Bozdogan (Ed.), Vol. 2, Proceedings of the First US/Japan Conference on the *Frontiers of Statistical Modeling: An Informational Approach*, Kluwer Academic Publishers, the Netherlands, Dordrecht, 69-113 (1994).

[4] H. Bozdogan, Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, 44, 62-91 (2000).

[5] H. Bozdogan, Intelligent Statistical Data Mining with Information Complexity and Genetic Algorithms. In H. Bozdogan (Ed.), *Statistical Data Mining & Knowledge Discovery*, Chapman & Hall/CRC, 15–56 (2004).

[6] H. Bozdogan, A new class of information complexity (ICOMP) criteria with an application to customer profiling and segmentation, *Istanbul University Journal of the School of Business Administration*, 39, 2, 370-398 (2010).

[7] H. Bozdogan, Information Complexity and Multivariate Learning in High Dimensions in Data Mining. A forthcoming book to appear (2011).

[8] K. Bollen, *Structural Equations with Latent Variables*, Wiley, 1989.

[9] S-Y Lee, *Structural Equation Modeling: A Bayesian Approach*, Wiley, 2007.

[10] A. Skrondal, and S. Rabe-Hesketh, *Generalized Latent Variable Modeling*, Chapman & Hall/CRC, 2004.

[11] A. Skrondal, and S. Rabe-Hesketh, Structural equation modeling: categorical variables, *Entry for the Encyclopedia of Statistics in Behavioral Science*, Wiley, 2005.

[12] A. Christoffersson, Factor analysis of dichotomized variables, *Psychometrika*, 40, 5-32 (1975).

[13] B. Muthén, Contributions to factor analysis of dichotomous variables, *Psychometrika*, 43, 551-560 (1978).

[14] B. Muthén, A structural probit model with latent variables. *Journal of American Statistical Association*, 74, 807-811 (1979).

[15] B. Muthén, Factor analysis of dichotomous variables: American attitudes towards abortion. In D.J. Jackson and E. F. Borgetta (Eds.), *Factor Analysis and Measurement in Sociological Research*. Beverly Hills, Sage, 1980.

[16] B. Muthén, Some categorical response models with continuous latent variables. In K. G. Joreskög and H. Wold (Eds.), *Systems under Indirect Observation: Causality, Structure, and Prediction*. Noth Holland Publishing Co., Amsterdam, 1981.

[17] B. Muthén, Latent variable structural equation modeling with categorical data, *Journal of Econometrics*, 22, 43-65 (1983).

[18] B. Muthén, A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators, *Psychometrika*, 49, 115-132 (1984).

[19] B. Muthén, and A. Christoffersson, Simultaneous factor analysis of dichotomous variables in several groups, *Psychometrika*, 46, 407-419 (1981).

[20] B. Muthén, and D. Kaplan, A comparison for some methodologies for the factor analysis of non-normal Likert variables, *British Journal of Mathematical and Statistical Psychology*, 38, 171-189 (1985).

[21] D.J. Bartholomew, Factor analysis for categorical data (with discussion). *Journal of the Royal Statistical Society*, Series B, 42, 293-321 (1980).

[22] D.J. Bartholomew, *Latent Variable Models and Factor Analysis*, Charles Griffin & Company, London, 1987.

[23] R. Mislevy, Recent developments in the factor analysis of categorical variables, *Journal of Educational Statistics*, 11, 3-31 (1986).

[24] S. Geman, and D. Geman, Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741 (1984).

[25] A. Gifi, *Nonlinear Multivariate Analysis*, Wiley, 1990.

[26] H. Bozdogan, ICOMP: A new model-selection criterion. In H. H. Bock (Ed.), *Classification and Related Methods of Data Analysis*, Elsevier Science Publishers, Amsterdam; 599-608 (1988).

[27] H. Bozdogan, On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics, Theory and Methods*, 19, 221-278 (1990).

[28] H. Bozdogan, and D.M.A. Haughton, Informational complexity criteria for regression models. *Computational Statistics and Data Analysis*, 28, 51-76 (1998).

[29] H. Bozdogan, and P.M. Bearse, Subset selection in vector autoregressive models using the genetic algorithm with informational complexity as the fitness function. *Systems Analysis, Modeling, Simulation* (SAMS) (1998).

[30] H. Bozdogan, and M. Ueno, A unified approach to information-theoretic and Bayesian model selection criteria. Invited paper presented in the Technical Session Track C on: Information Theoretic Methods and Bayesian Modeling at the 6th World Meeting of the International Society for Bayesian Analysis (ISBA), May 28-June 1, 2000, Hersonissos-Heraklion, Crete, (2000).

[31] P.M. Bentler, Covariance structure analysis: statistical practice, theory, and directions, *Annu. Rev. Psychol*. 47, 563-592 (1996).

[32] K.G. Jöreskog, and D. Sörbom, *LISREL 7, A guide to the program and applications*, 2nd Edition, SPSS Inc., (1989).

[33] R. Fletcher, and M.J.D. Powell, A rapidly convergent descent method for minimization, *Computer Journal*, 6, 163-168 (1963).

[34] R. Fletcher, and C.M. Reeves, Function minimization by conjugate gradients, *Computer Journal*, 7, 149-154 (1964).

[35] P.M. Bentler, *EQS structural equations program manual*, Encino, CA: Multivariate software, 1995.

[36] S.Y. Lee, and R.I. Jennrich, A study of algorithms for covariance structure analysis with specific comparisons using factor analysis, *Psychometrika*, 44, 1, 99-113 (1979).

[37] M. Jamshidian, and P.M. Bentler, ML estimation of mean and covariance structures with missing data using complete data routines, *Journal of Educational and Behavioral Statistics*, 24, 1, 21-41 (1999).

[38] H. Bozdogan, A new information theoretic measure of complexity index for model evaluation in general structural equation models with latent variables, invited paper presented at the *Symposium on Model Selection in Covariance Structures at the Joint Meeting of Psychometric Society & the Classification Society*, June 13-16, Rutgers University, Newark, NJ. (1991).

[39] J.L. Williams, H. Bozdogan, H., and L. Aiman-Smith, Inference problems with equivalent models, in *Advanced Structural Equation Modeling Issues and* Techniques, Eds. Marcoulides, A.G. & Schumaker, R.E., Lawrence Erlbaum Associates, New Jersey, pp. 279-314, 1995.

[40] J.R. Magnus, *Linear Structures*, Charles Griffin & Company, London and Oxford University Press, New York, 1988.

[41] J.R. Magnus, and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, Chichester/New York, 1999.

[42] S.B. Cho, Decomposing individual and group differences of categorical variables with genetic factor model, Unpushed Masters Thesis, the University of Missouri-Columbia, 2007.

[43] W.J. Heiser, and J.J. Meulman, Homogeneity Analysis: Exploring the distribution of variables and their nonlinear relationships. In M. Greenacre & J. Blasius (Eds.), Correspondence analysis in the social sciences: Recent developments and applications, pages pp. 179–209 (1994).

[44] J.J. Meulman, *Homogeneity analysis of incomplete data*. Leiden, The Netherlands: DSWO Press, 1982.

[45] J.J. Meulman, The integration of multidimensional scaling and multivariate analysis with optimal transformations of the variables. *Psychometrika*, 57:539–565 (1992).

[46] J.J. Meulman, Nonlinear principal coordinates analysis: minimizing the sum of squares of the smallest eigenvalues. *British Journal of Mathematical and Statistical Psychology*, 46:287–300 (1993).

[47] J.J. Meulman, Fitting a distance model to homogeneous subsets of variables: Points of view analysis of categorical data. *Journal of Classification*, 13:249–266 (1996).

[48] J.J. Meulman, *Optimal scaling methods for multivariate categorical data analysis*. SPSS White Paper, 1998.

[49] J.J. Meulman, Prediction and Classification in NonLinear Data Analysis: Something Old, Something New, Something Borrowed, Something Blue. *Physchometrica*, Vol 68 (No 4):493–517 (2003).

[50] J.J. Meulman, and A. J. V. der Kooij, Transformations towards independence through optimal scaling. Paper presented at the International Conference on Measurement and Multivariate Analysis (ICMMA), Banff, Canada (2000).

[51] J.J. Meulman, A.J.V. der Kooji, and W. J. Heiser, Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In D.

Kaplan (Ed.), Handbook of qunatitative methodology for the social sciences, pp. 49-70, 2004.

[52] G. Michailidis, and J. de Leeuw, The Gifi System of Descriptive Multivariate Analysis. *Statistical Science*, 1998, Vol 13 (No. 4):307–336 (1996).

[53] S. Katragada, Multivariate Mixed-Data Mining with Gifi System using Genetic Algorithm and Information Complexity. Ph.D. Thesis under the Supervision of Prof. Bozdogan, Department of Statistics, The University of Tennessee, Knoxville, TN, 2008.

[54] S. Katragada, and H. Bozdogan, *Multivariate Mixed-Data Mining in High-Dimensions with Gifi System using Genetic Algorithm and Information Complexity*. A forthcoming book to appear, 2011.

[55] U. Paquet, Bayesian Inference for Latent Variable Models, Ph.D. thesis, Wolfson College, University of Cambridge, Cambridge, U.K., March, 2007.

[56] M.H. van Emden, An Analysis of Complexity. *Mathematical Centre Tracts*, Amsterdam, 35, (1971).

[57] H. Cramér, Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ, 1946.

[58] C.R. Rao, Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math Soc*., 37, 81 (1945).

[59] C.R. Rao, Minimum variance and the estimation of several parameters. *Proc. Cam. Phil. Soc.*, 43, 280 (1947).

[60] C.R. Rao, Sufficient statistics and minimum variance estimates. *Proc. Cam. Phil. Soc.*, 45, 213 (1948).

[61] J.Rissanen, Modeling by shortest data description. *Automatica*, 14, 465-471 (1978).

[62] G.Schwarz, Estimating the dimension of a model. *Ann. Statist.*, 6, 461-464 (1978).

[63] D.S. Poskitt, Precision, Complexity and Bayesian model determination. *J. Roy. Statist. Soc*. 49, 199-208 (1987).

[64] K. Takeuchi, Distribution of information statistics and a criterion of model fitting. *Suri-Kagaku (Mathematical Sciences)*, 153, 12-18 (1976).

[65] J.R.M. Hosking, Lagrange-multiplier tests of time-series models. *Journal of the Royal Statistics Society*, (Series B), 42, 170-181 (1980).

[66] R. Shibata, Statistical aspects of model selection. In J.C. Willems (Ed.), From the data to modeling, Berlin: Springer-Verlag, pp. 216-240, 1989.

[67] M. Power, M. Bullingen, and A. Hazper, The World Health Organization WHOQOL-100: Tests of universality of quality of life in 15 different cultural groups worldwide. *Health Psychology*, 18, 495-505 (1999).

[68] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519-530 (1970).

[69] K.V. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya*, Ser B 36(2), 115-128 (1974).

[70] K.V. Mardia, Tests of univariate and multivariate normality. In: P.R. Krishnaiah (ed.), Handbook of Statistics, Vol 1, 279-320, North Holland, 1980.

[71] E. Deniz, J.A. Howe, and Bozdogan, Robustifying Structural Equation Modeling Against Nonnormality with Adaptive Kernels, Paper under review in Multivariate Behavioral Research, 2010.

Appendix

Here we provide the input parameters and the the path diagrams of Model 1, Model 2, and Model 3 in our analysis. The path diagrams for the fitted models are shown in Figures A.1 to A.3.

We note that our models are much simpler than that of the Bayesian approach proposed by Lee (2007).

MODEL 1 Parameters are:

$$\Lambda_x^{'} = \begin{pmatrix} 1.00 & 0.95 & 1.92 & 1.62 & 0.97 & 2.44 & 1.93 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.00 & 0.59 & 0.44 & 0.48 & 0.93 & 0.82 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.00 & 0.13 & 0.80 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.00 & 0.61 & 0.55 & 0.59 & 0.58 & 0.85 & 0.46 & 0.57 \end{pmatrix}$$

$$\Lambda_y = \begin{pmatrix} 1.00 \\ 1.00 \end{pmatrix} \quad \Gamma = \begin{pmatrix} 1.54 & 0.06 & 0.63 & -0.10 \end{pmatrix} \quad \Phi = \begin{pmatrix} 0.08 & & & \\ 0.11 & 0.28 & & \\ 0.05 & 0.17 & 0.16 & \\ 0.09 & 0.19 & 0.15 & 0.21 \end{pmatrix} \quad \Psi = \begin{pmatrix} 0.27 \end{pmatrix}$$

$$\theta_\delta = diag\begin{pmatrix} 0.16 & 0.17 & 0.21 & 0.17 & 0.20 & 0.14 & 0.22 & 0.22 & 0.20 & 0.17 & 0.18 & 0.23 & 0.22 & 0.16 & 0.06 & 0.13 & 0.23 & 0.18 & 0.15 & 0.20 & 0.25 & 0.19 & 0.12 & 0.16 \end{pmatrix}$$

$$\theta_\in = diag\begin{pmatrix} 0.00 & 0.00 \end{pmatrix}$$

MODEL 2 Parameters are:

$$\Lambda_x^{'} = \begin{pmatrix} 1.00 & 0.96 & 1.91 & 1.62 & 0.98 & 2.45 & 1.94 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.00 & 0.59 & 0.44 & 0.49 & 0.94 & 0.82 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.00 & 0.60 & 0.54 & 0.59 & 0.61 & 0.83 & 0.44 & 0.57 \end{pmatrix}$$

$$\Lambda_y = \begin{pmatrix} 1.00 \\ 1.00 \end{pmatrix} \quad \Gamma = \begin{pmatrix} 1.12 & 0.36 & 0.26 \end{pmatrix} \quad \Phi = \begin{pmatrix} 0.08 & & \\ 0.11 & 0.28 & \\ 0.09 & 0.19 & 0.22 \end{pmatrix} \quad \Psi = \begin{pmatrix} 0.28 \end{pmatrix}$$

$$\theta_\delta = diag\begin{pmatrix} 0.16 & 0.17 & 0.21 & 0.17 & 0.20 & 0.14 & 0.22 & 0.22 & 0.20 & 0.17 & 0.18 & 0.23 & 0.22 & 0.23 & 0.18 & 0.15 & 0.20 & 0.24 & 0.20 & 0.13 & 0.16 \end{pmatrix}$$

$$\theta_\in = diag\begin{pmatrix} 0.00 & 0.00 \end{pmatrix}$$

MODEL 3 Parameters are:

$$\Lambda_x^{'} = \begin{pmatrix} 1.00 & 0.59 & 0.43 & 0.48 & 0.91 & 0.81 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.00 & 0.60 & 0.54 & 0.58 & 0.58 & 0.83 & 0.45 & 0.57 \end{pmatrix}$$

$$\Lambda_y = \begin{pmatrix} 1.00 \\ 1.00 \end{pmatrix} \quad \Gamma = \begin{pmatrix} 0.72 & 0.39 \end{pmatrix} \quad \Phi = \begin{pmatrix} 0.29 & \\ 0.19 & 0.22 \end{pmatrix} \quad \Psi = \begin{pmatrix} 0.33 \end{pmatrix}$$

$$\theta_\delta = diag\begin{pmatrix} 0.21 & 0.19 & 0.18 & 0.18 & 0.23 & 0.22 & 0.23 & 0.18 & 0.15 & 0.20 & 0.24 & 0.20 & 0.12 & 0.16 \end{pmatrix}$$

$$\theta_\in = diag\begin{pmatrix} 0.00 & 0.00 \end{pmatrix}$$

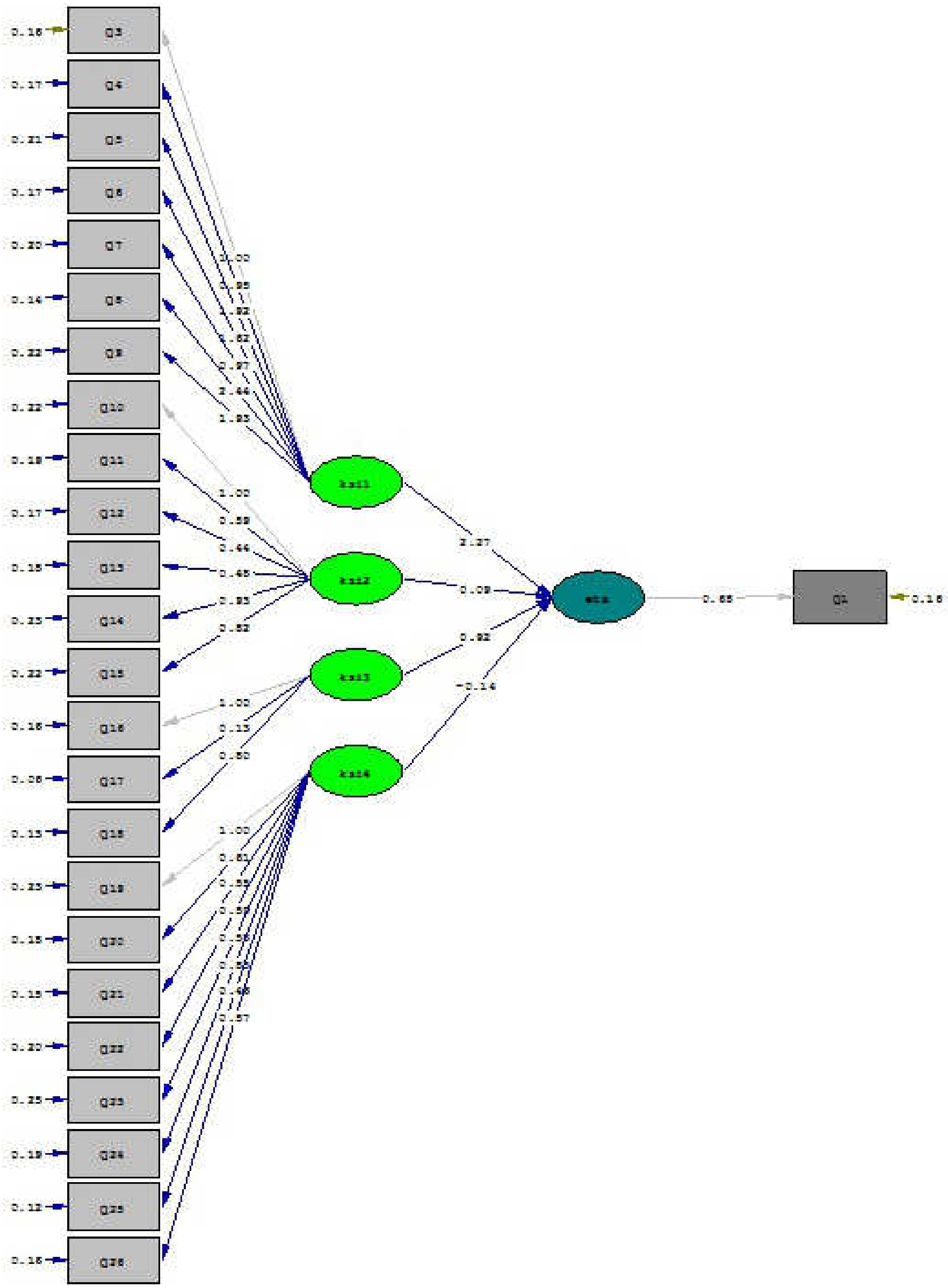

Figure A1 Path Diagram for Model 1

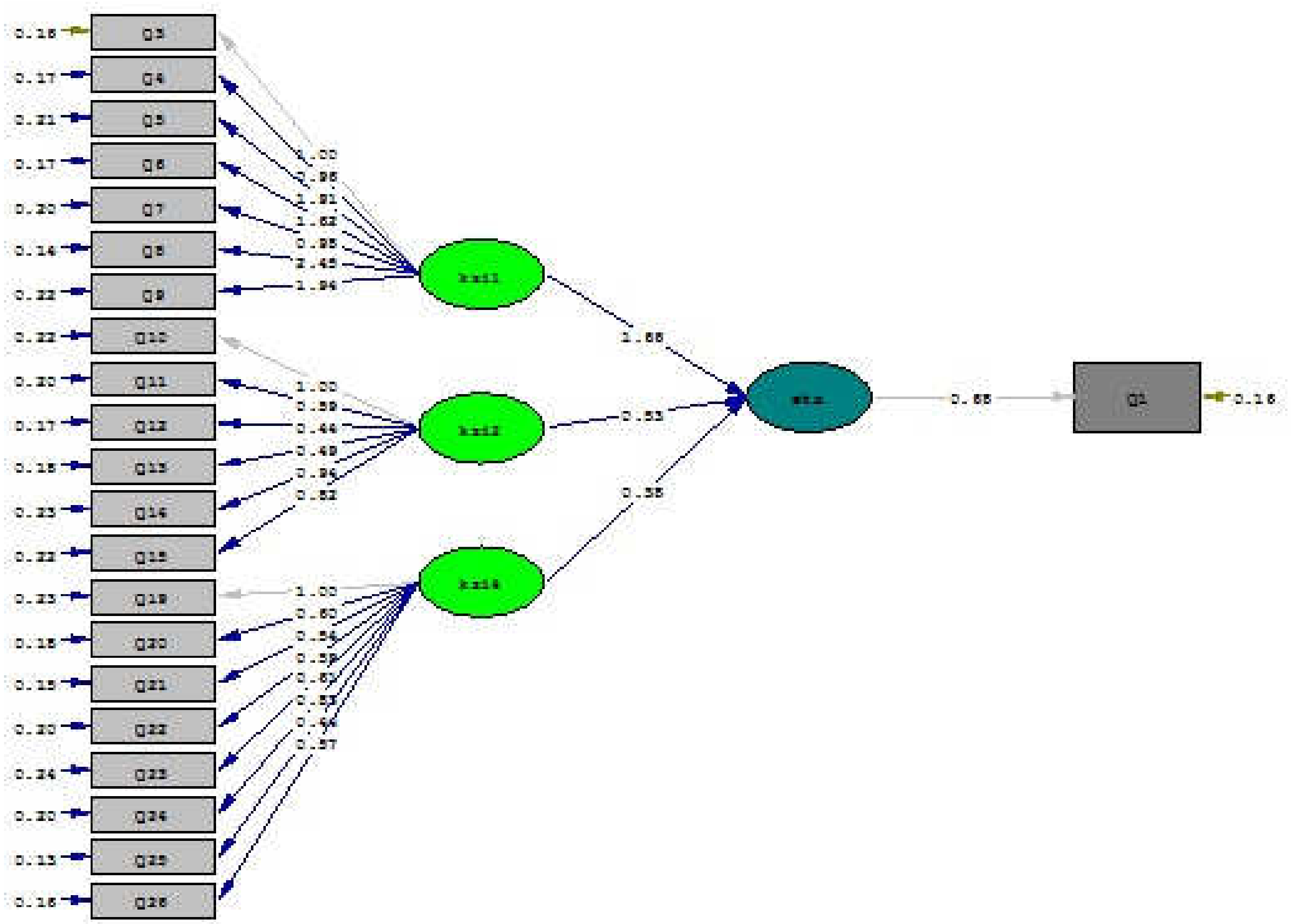

Figure A2 Path Diagram for Model 2

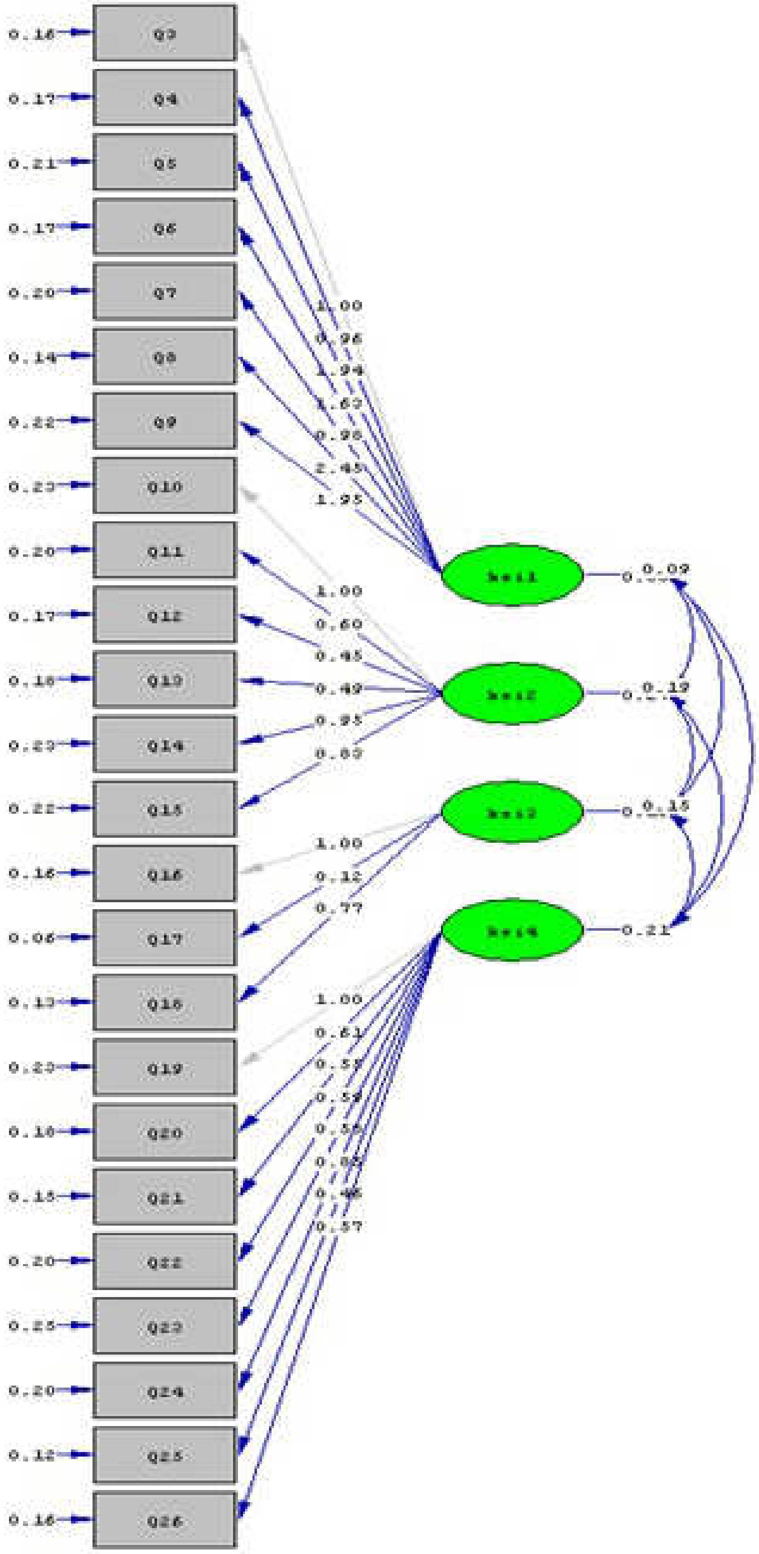

Figure A3 Path Diagram for Model 3