EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 5, No. 3, 2012, 282-301
ISSN 1307-5543 – www.ejpam.com

Performance of Information Complexity Criteria in Structural Equation Models with Applications

Eylem Deniz Howe[1,*], Hamparsum Bozdogan[2], and Gülay Kıroğlu[1]

[1] *Mimar Sinan Fine Arts University, Istanbul Turkey*
[2] *Department of Statistics, Operations, and Management Science, The University of Tennessee, Knoxville, TN, 37996, USA*

Abstract. A common problem in structural equation modeling is that of model selection. Many researchers have addressed this problem, but many methods have provided mixed benefits until recently. Akaike's well-known criteria, AIC, has been applied in the context of structural equation modeling, but the effectiveness of many other information criteria have not been studied in a convincing manner. In this paper, we compare the SEM model selection prowess of several AIC-type and $ICOMP$-type criteria. We also introduce two new large sample consistent forms of Bozdogan's $ICOMP$ criteria - one of which is robust to model misspecification.
To study the empirical performance of the information criteria, we use a well-known SEM simulation protocol, and demonstrate that most of the information-theoretic criteria select the "pseudo true" model with very high frequencies. We also demonstrate, however, that the performance of AIC is inversely related to the sample size. Finally, we apply the new criteria to select an analytical model for a real dataset from a retail marketing study of consumer behavior. Our results show the versatility of the new proposed method where both the goodness-of fit and the complexity of the model is taken into account in one criterion function.

2010 Mathematics Subject Classifications: 62H12,62H15,62H20,62H25,62P15,62F35

Key Words and Phrases: Structural Equation Model, Information Criteria, AIC-type Criteria, $ICOMP$-type Criteria

1. Introduction

Structural Equation Modeling (SEM) has been a popular tool in social and behavioral sciences since the early 1990's for the causal modeling of complex, multivariate data sets. Presently, it has penetrated engineering, management and information sciences, and genomic research, to mention a few. It has changed the perspective of researchers on how to do good statistical modeling and model selection. It has also fostered the curiosity of social

*Corresponding author.

Email addresses: eylemdeniz@gmail.com (E. Howe), bozdogan@utk.edu (H. Bozdogan), gkiroglu@msu.edu.tr (G. Kıroğlu)

scientists about the underlying constructs or factors in many applications. According to [11], the SEM approach has been regarded, perhaps, as the most significant and effective statistical revolution in social sciences.

In SEM one of the fundamental problems is how to evaluate different models and how to select among the rival models [18]. In the literature of SEM, there are many alternative fit indices available to the researchers. An excellent recent book by [22], "Linear Causal Modeling with Structural Equations" devotes a fair amount of discussion on many goodness-of-fit indices. Mulaik [22] further discusses information theoretic measures of model discrepancy such as AIC, see, e.g., [1, 2, 27, 4]. He devotes a section on the information theoretic measure of complexity (ICOMP) developed in [6, 7]. However, [22] does not provide numerical examples or evaluate the empirical performance of these model selection criteria in the SEM framework.

The first author in [14], evaluated the performance of several information criteria for measurement models and SEMs. In this paper, our objective is to to extend that research and to present some new results on ICOMP which is consistent with respect to sample size and robustness against model misspecification in SEM and in other high dimensional model fitting.

Therefore, in this paper we revisit and discuss information criteria used to gauge model fit in structural equation modeling (SEM). This is based on the objective that statistical methods of fit are related to sample size, any statistical test is increasingly likely to imply rejection of a model as sample size increases, even if model misfit is trivial in magnitude. Various practical fit indices have been proposed over the years - and information criteria are among the most widely used. We review prior information criteria, and propose a couple of new information criteria based on information measure of complexity of a covariance matrix called $ICOMP$-type criteria.

We demonstrate the empirical performance of $ICOMP$-type criteria along with AIC, Consistent Akaike Information Criterion ($CAIC$), Consistent Akaike Information Criterion with Fisher Information ($CAICF$) both due to [4], and a Bayesian Model Selection (BMS) criterion developed by [10].

The rest of the paper is organized as follows. Section 2 provides the requisite background on SEM, which is then followed by Section 3 regarding AIC-type and $ICOMP$-type criteria for model selection. Here, we define $ICOMP$ under the correct and misspecified models. The basic idea is that one can use the difference between ICOMP(misspecified model) and ICOMP(correctly specified model) as an indication of possible departures from the distributional form of the model. This brings out the most important weakness of Akaike-type criteria for model selection, that AIC depends crucially on the assumption that the specified family of models includes the pseudo true model. In general, this may not be the case. Finally, results with both simulated and real datasets are shown in Section 5. Our simulation results are based on three models based on a slight modification of Fan's simulation protocol [15] to achieve convergence in LISREL.

2. Structural Equation Model with Latent Variables

General SEMs consist of two models referred to as measurement and structural equation (or latent variable) models. The structural equation model defines relationship between latent variables. This model is a path model adapted to latent variables. The measurement model is a confirmatory factor (CF) model which defines relationship between observed and latent variables. Latent variables in the measurement model are the factors in the CF model. Following [19] in matrix notation, we define the general SEM model with these two models by three equations:

Structural Equation model:	$\underset{(r\times 1)}{\eta} = \underset{(r\times r)(r\times 1)}{B\eta} + \underset{(r\times s)(s\times 1)}{\Gamma\xi} + \underset{(r\times 1)}{\zeta}$
Measurement model for y:	$\underset{(p\times 1)}{y} = \underset{(p\times r)(r\times 1)}{\Lambda_y \eta} + \underset{(p\times 1)}{\varepsilon}$
Measurement model for x:	$\underset{(q\times 1)}{x} = \underset{(q\times s)(s\times 1)}{\Lambda_x \xi} + \underset{(q\times 1)}{\delta}$

where η is a $(r\times 1)$ vector of latent endogenous (or dependent) variables, ξ is $(s\times 1)$ vector of latent exogenous (or independent) variables, ζ is $(r\times 1)$ vector of latent errors in equations, β is an $(r\times r)$ 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 β, 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:
$E(\eta)=E(\xi)=E(\zeta)=0$, ζ is uncorrelated with ξ, and $(I-\beta)$ is nonsingular. The covariance matrix,

$$E(\xi\xi')=\Phi, \tag{1}$$

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

$$E(\zeta\zeta')=\Psi \tag{2}$$

is an $(r\times r)$ covariance matrix of the latent errors in equations. The measurement model specifies how the observed variables, x and y, are determined through Λ_x and Λ_y 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 for the measurement model are:
$E(\eta)=E(\xi)=E(\varepsilon)=E(\delta)=0$. Furthermore, ε is uncorrelated with ξ, η and δ. Likewise, δ is uncorrelated with ξ, η and ε. The covariances matrices in the case of the measurement model are:

$$E(\varepsilon\varepsilon')=\Theta_\varepsilon=\Sigma(\varepsilon), \tag{3}$$

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

$$E(\delta\delta')=\Theta_\delta=\Sigma(\delta), \tag{4}$$

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

According to the measurement and structural equation models, the implied full covariance matrix, $\Sigma(\theta)$, for the general SEM is given by

$$\Sigma(\theta)=\begin{bmatrix} \Lambda_y(I-B)^{-1}(\Gamma\Phi\Gamma'+\Psi)[(I-B)^{-1}]'\Lambda_y'+\Theta_\epsilon & \Lambda_y(I-B)^{-1}\Gamma\Phi\Lambda_x' \\ (\Lambda_y(I-B)^{-1}\Gamma\Phi\Lambda_x')' & \Lambda_x\Phi\Lambda_x'+\Theta_\delta \end{bmatrix}. \quad (5)$$

The elements of the general implied covariance matrix are functions of B, Γ, Λ_y, Λ_x, Φ, Ψ, Θ_ϵ and Θ_δ. To be able to use SEM, we need to specify the pattern of the elements of these eight matrices. There are three kinds of specifications. These are:

- Fixed parameters- that have been assigned specific values,
- Constrained parameters- that are unknown but equal to one or more other parameters, and
- Free parameters- that are unknown and not constrained to be equal to any other parameter.

For more on these, see [3, 19], and others.

The fundamental hypothesis for these structural equations procedures is that the covariance matrix of the observed variables is a function of a set of parameters. If the model is correct and if we know the parameters, then the population covariance matrix can be exactly reproduced. The hypothesis in this case is:

$$\Sigma=\Sigma(\theta). \quad (6)$$

In (6), Σ is the population covariance matrix of the observed variables, θ is a vector containing the free parameters of the model, and $\Sigma(\theta)$ is the covariance matrix written as a function of the covariance matrix implied by a specific model [3].

3. Information Theoretic Model Selection Criteria

One of the fundamental difficulties in statistical analysis and data mining is the choice of an appropriate model, estimating and determining the order or dimension of a model. In general statistical modeling and model evaluation problems, the concept of information theoretic measure of complexity plays an important role. At the philosophical level, *complexity* involves notions such as connectivity patterns and the *interactions* of 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 [4].

Based on Akaike's AIC, many model-selection procedures that take the form of a *penalized likelihood* plus a *penalty term*) have been proposed.

For a general multivariate linear or nonlinear model defined by

$$Statistical\ model=Signal+Noise \quad (7)$$

a summary diagram for AIC and ICOMP in terms of a *loss function* is give by

$$\text{Loss} = \left.\begin{array}{l} \left.\begin{array}{l}\text{Lack of Fit}\\ +\text{Lack of Parsimony}\end{array}\right\} \longrightarrow \boxed{AIC} \\ +\text{Profusion of Complexity}\end{array}\right\} \longrightarrow \boxed{ICOMP}$$

3.1. Akaike-Type Information Criteria

Since Akaike (1973), entropic information criteria has had a substantial impact in statistical model evaluation problems. Its introduction furthered the recognition of the importance of good modeling and model fitting in the statistical science. As a result, many important statistical modeling techniques have been developed in various fields of statistics such as in biostatistics, control theory, econometrics, engineering, medical data mining, psychometrics, and many others.

Akaike Information Criterion (AIC)

AIC, which is one of the first information criterion developed by Akaike, can be thought of as a generalized form of the maximum likelihood. AIC can be formulated as maximizing generalized entropy, or equivalently minimizing *Kullback-Leibler (KL) information*. See, e.g. [21], and [20]. It is obtained from asymptotic unbiased estimation of the logarithm of the average expected likelihood of a model and it is defined by

$$AIC = -2\log L(\widehat{\theta}_k) + 2k, \tag{8}$$

where θ is a k-dimensional unknown parameter vector, $\widehat{\theta}$, denotes the maximum likelihood estimator of θ, $L(\widehat{\theta})$, is the maximized likelihood function of the model, and k is the number of unknown parameters estimated in the model.

Consistent Akaike Information Criterion (CAIC)

Without violating Akaike's principles and using the results in mathematical statistics, Bozdogan [4] extended AIC analytically in two ways. These extensions make AIC asymptotically consistent, and penalize overparameterization more stringently, so as to pick the simplest of the true models whenever there is nothing to be lost in doing so. $CAIC$ is defined by

$$CAIC = -2\log L(\widehat{\theta}_k) + k(\log(n) + 1). \tag{9}$$

Using a correction factor based on the sample size (n) $CAIC$ is an attempt to overcome the tendency of the AIC in overestimating the complexity (i.e., number of parameters) of the underlying model [17].

Consistent Akaike Information Criterion with Fisher Information (CAICF)

Also in [4], a different estimator for negative twice the expected entropy was proposed. Similar to $CAIC$, this approach extends AIC analytically to make it consistent without deviating from Akaike's original premise. In this manner [4] penalizes overparameterization more strongly, in particular, in large samples. $CAICF$ is defined by

$$CAICF = -2\log L(\widehat{\theta}) + k(\log(n) + 2) + \log|\hat{\mathscr{F}}|, \tag{10}$$

where $\hat{\mathscr{F}}$ is the estimated Fisher information matrix (FIM) of the model.

Bayesian Model Selection Criterion (BMS)

Bozdogan and Ueno [10] provided yet a further extension of AIC called Bayesian Model Selection Criterion (BMS), given by

$$BMS = -2\log L(\widehat{\theta}) + k\log(n) + 2(\frac{nk}{n-k-2}) + \log|\hat{\mathscr{F}}|. \tag{11}$$

As we note, these extended forms AIC all attempt to repair the inconsistency problem of the AIC.

3.2. The General Form of Information Complexity: ICOMP

Motivated from considerations similar to those in AIC, here we give details a *new entropic or information-theoretic measure of complexity* called $ICOMP$ of [6, 7, 9] as a decision rule for model selection in SEM.

The development and construction of $ICOMP$ is based on a generalization of the *covariance complexity index* originally introduced in [29]. Instead of penalizing the number of free parameters directly, $ICOMP$ penalizes the covariance complexity of the model. It is defined by

$$ICOMP = -2\log L(\hat{\theta}_k) + 2C(\widehat{Cov}(\hat{\theta}_k)), \tag{12}$$

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 with k unknown parameter, and C represents a *real-valued complexity measure*.

As explained in [7, 8, 9], there are several forms and theoretical justifications of $ICOMP$. For brevity, here, we will recapitulate and use one of the most general forms of $ICOMP$ referred to as $ICOMP(IFIM)$. $ICOMP(IFIM)$ exploits the well-known asymptotic optimality properties of the maximum likelihood estimators (MLEs), and uses the information-based complexity of the estimated *inverse-Fisher information matrix* ($IFIM$ or $\hat{\mathscr{F}}^{-1}$) of a model. This is known as the *Cramér-Rao lower bound (CRLB) matrix*. See, e.g., [12] and [23, 24, 25]. As such, $ICOMP(IFIM)$ is an approximation to the sum of two KL distances. This approach provides us an achievable accuracy of the parameter estimates of the model by considering the entire parameter space.

For a multivariate normal linear or nonlinear structural model we define the general form of $ICOMP(IFIM)$ as

$$ICOMP(\hat{\mathscr{F}}^{-1}) = -2\log L(\hat{\theta}_k) + 2C_1(\hat{\mathscr{F}}^{-1}), \tag{13}$$

where C_1 denotes the maximal informational complexity of $\hat{\mathscr{F}}^{-1}$ and is given by

$$C_1(\hat{\mathscr{F}}^{-1}) = \frac{s}{2}\log[\frac{tr(\hat{\mathscr{F}}^{-1})}{s}] - \frac{1}{2}\log|\hat{\mathscr{F}}^{-1}|, \tag{14}$$

and where $s = dim(\hat{\mathscr{F}}^{-1}) = rank(\hat{\mathscr{F}}^{-1})$.

The first component of $ICOMP(IFIM)$ measures the lack of fit of the model, and the second component measures the complexity of the accuracy of the estimated parameters and implicitly adjusts for the number of free parameters included in the model. It is a measure of the state of disorder of a model for a given data set. For more on this, and for some immediate physical motivation, we refer the readers to the interesting book by [16], entitled *"Physics from Fisher Information."*

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

$$ICOMP(\hat{\mathscr{F}}^{-1}) = -2\log L(\hat{\theta}_k) + s\log\left(\overline{\lambda}_a/\overline{\lambda}_g\right). \tag{15}$$

We note that $ICOMP(IFIM)$ now looks in appearance like the $CAIC$, MDL [26], and the Bayesian criterion SBC [28], except for using $\log\left(\overline{\lambda}_a/\overline{\lambda}_g\right)$ instead of using $\log(n)$, where $\log(n)$ denotes the natural logarithm of the sample size n. In this sense, $ICOMP(IFIM)$ resembles a penalized likelihood method similar to AIC and AIC-type criteria, except that the penalty depends on the curvature of the log likelihood function via the scalar C_1 complexity value of the estimated $IFIM$.

3.3. Consistent and Misspecification Forms of ICOMP

Bozdogan [9] suggested several other different approaches to generalize and derive the $ICOMP$ criteria by maximizing the *posterior expected utility* (PEU) when the models are misspecified (which is the case in practice). Here, we give derived forms of these generalized $ICOMP$ criteria as follows.

$$ICOMP\left(\hat{\mathscr{F}}^{-1}\right)_{PEU\_Miss} = -2\log L\left(\hat{\theta}\right) + k + 2\left[tr\left(\hat{\mathscr{F}}^{-1}\hat{\mathscr{R}}\right) + C_1\left(\hat{\mathscr{F}}^{-1}\right)\right] \tag{16}$$

In (16), whereas $\hat{\mathscr{F}}^{-1}$ is $IFIM$ in inner-product form, $\hat{\mathscr{R}}$ is $IFIM$ in outer-product form. If the model is correctly specified, these two forms of Fisher information matrices would be equal to one another. That is, if $\hat{\mathscr{F}}^{-1} = \hat{\mathscr{R}}$, then $tr(\hat{\mathscr{F}}^{-1}\hat{\mathscr{R}}) = tr\left(I_k\right) = k$. In this case, we have

$$ICOMP\left(\hat{\mathscr{F}}^{-1}\right)_{PEU} = -2\log L(\hat{\theta}) + 3k + 2C_1(\hat{\mathscr{F}}^{-1})$$

$$= AIC_3 + 2C_1(\hat{\mathscr{F}}^{-1}) \tag{17}$$

Bozdogan [9] approximates the term $tr(\hat{\mathscr{F}}^{-1}\hat{\mathscr{R}})$ in equation (16) by

$$tr\left(\hat{\mathscr{F}}^{-1}\hat{\mathscr{R}}\right) = \frac{nk}{n-k-2}$$

which corrects the bias for small as well as large sample sizes. Thus, equation (16) reduces to

$$ICOMP\left(\hat{\mathscr{F}}^{-1}\right)_{PEU\_Miss} = -2\log L(\hat{\theta}) + k + 2(\frac{nk}{n-k-2}) + 2C_1(\hat{\mathscr{F}}^{-1}). \tag{18}$$

This form of $ICOMP$ is useful in cases where we cannot determine the closed form expression of $\hat{\mathscr{R}}$, outer-product form of $IFIM$ in many modeling situations and SEM is a case in point example.

Consistent $ICOMP$ that maximizes the *posterior expected utility* (PEU)is given by

$$\begin{aligned} CICOMP(\hat{\mathscr{F}}^{-1})_{PEU} &= -2\log L(\hat{\theta}|X) + k + 2k\log(n) + 2C_1(\hat{\mathscr{F}}^{-1}) \\ &= CAIC + 2C_1(\hat{\mathscr{F}}^{-1}). \end{aligned} \tag{19}$$

Further, consistent $ICOMP$ that maximizes the *posterior expected utility* (PEU) and also guards us against misspecification is given by

$$CICOMP(\hat{\mathscr{F}}^{-1})_{PEU\_Miss} = -2\log L(\hat{\theta}|X) + k + 2\log(n)\frac{nk}{n-k-2} + 2C_1(\hat{\mathscr{F}}^{-1}). \tag{20}$$

A model with minimum ICOMP-type and AIC-type criteria are chosen to be the best model among all possible competing alternative models.

With $ICOMP$-type criteria, 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, $ICOMP$-type criteria provide a more judicious penalty term than AIC, MDL, SBC, or $CAIC$. The lack of parsimony is automatically adjusted by $C_1(\hat{\mathscr{F}}^{-1})$ across the competing alternative portfolio of models as the parameter spaces of these models are constrained in the model selection process. For more on $ICOMP$-type criteria, we refer the readers to [9] and [13].

4. Derived Forms of Information Criteria in Structural Equation Models

4.1. AIC-type Criteria in SEM

Under the assumption that the observed variables are continuous and have interval scales, and multivariate normal, that is,

$$z = (y', x')' \sim N_{(p+q)}(0, \Sigma(\theta)) \tag{21}$$

and using the maximum likelihood estimators, we obtain AIC, $CAIC$, $CAICF$ and BMS for the SEM. These are given in [13] as follows.

$$AIC = n(p+q)\log(2\pi) + n\log|\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}S) + 2k, \tag{22}$$

$$CAIC = n(p+q)\log(2\pi) + n\log|\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}S) + k\left[\log(n)+1\right], \quad (23)$$

$$CAICF = n(p+q)\log(2\pi) + n\log|\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}S) + k\left[\log(n)+2\right] + \log|\hat{\mathscr{F}}|, \quad (24)$$

$$BMS = n(p+q)\log(2\pi) + n\log|\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}S) + k\log(n) + (\frac{nk}{n-k-2}) + \log|\hat{\mathscr{F}}|, \quad (25)$$

where k is the number of parameters in SEM.

4.2. ICOMP-type Criteria in SEM

Let $\hat{\Sigma}(\hat{\theta})$ denote the maximum likelihood estimator (MLE) of the implied covariance matrix $\Sigma(\theta)$ given by

$$\underset{(p+q)\times(p+q)}{\hat{\Sigma}(\hat{\theta})} = \begin{bmatrix} \hat{\Lambda}_y(I-\hat{B})^{-1}(\hat{\Gamma}\hat{\Phi}\hat{\Gamma}'+\hat{\Psi})(I-\hat{B}')^{-1}\hat{\Lambda}_y+\hat{\Theta}_\varepsilon & \hat{\Lambda}_y(I-\hat{B})^{-1}\hat{\Gamma}\hat{\Phi}\hat{\Lambda}_x' \\ \hat{\Lambda}_x\hat{\Phi}\hat{\Gamma}'(I-\hat{B}')^{-1}\hat{\Lambda}_y' & \hat{\Lambda}_x\hat{\Phi}\hat{\Lambda}_x'+\hat{\Theta}_\delta \end{bmatrix}. \quad (26)$$

The first $ICOMP$ criterion in SEM is given by

$$ICOMP = n(p+q)\log(2\pi) + n\log|\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}S) + 2C_1(\hat{\Sigma}(\hat{\theta})). \quad (27)$$

We can also use the estimated $IFIM$, the covariance of the model parameters, for the general SEM as given by

$$\hat{\mathscr{F}}^{-1} = \begin{bmatrix} \frac{1}{n}\hat{\Sigma}(\hat{\theta}) & 0 \\ 0' & \frac{2}{n}\underset{(p+q)}{D}^{+}[\hat{\Sigma}(\hat{\theta})\otimes\hat{\Sigma}(\hat{\theta})]\underset{(p+q)}{D}^{+'} \end{bmatrix} \quad (28)$$

to obtain several other new forms of $ICOMP$ based on the work of [5].

Note that in one dimension equation in (28) reduces to the $IFIM$ of the normal distribution $N(\mu,\sigma^2)$ given by

$$\hat{\mathscr{F}}^{-1} = \begin{bmatrix} \frac{1}{n}\hat{\sigma}^2 & 0 \\ 0' & \frac{2}{n}\hat{\sigma}^4 \end{bmatrix} \quad (29)$$

which checks and shows the correctness of the formula in (28) and that $\hat{\theta}$ has the consistency property.

In (28), the matrix D_p^+ is the Moore-Penrose inverse of the duplication matrix D_p. The duplication matrix D_p is a unique $p^2 \times \frac{1}{2}p\left(p+1\right)$ matrix, and so its Moore-Penrose inverse is

$$D_p^+ = \left(D_p'D_p\right)^{-1}D_p'$$

which is a $\frac{1}{2}p\left(p+1\right)\times p^2$ matrix. Further, note that the duplication matrix D_p is implicitly defined by

$$D_p vech(\hat{\Sigma}(\hat{\theta})) = vec(\hat{\Sigma}(\hat{\theta}))$$

where $vech(\hat{\Sigma})$ vectorizes the distinct elements of $\hat{\Sigma}$ by vertically stacking those on and below the principal diagonal. Consequently

$$vech(\hat{\Sigma}(\hat{\theta})) = D_p^+ vec(\hat{\Sigma}(\hat{\theta})).$$

Based on the above results, we establish $ICOMP(IFIM)$ for the SEM with latent variables given by

$$ICOMP(\hat{\mathscr{F}}^{-1}) = n(p+q)\log(2\pi) + n\log|\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}S) + 2C_1(\hat{\mathscr{F}}^{-1}). \tag{30}$$

We further note that with $C_1(\hat{\mathscr{F}}^{-1})$, we do not actually need to construct $\hat{\mathscr{F}}^{-1}$, since it is a scalar measure of complexity. As derived in [13], using the definition of complexity, we get a computationally convenient open form of the expression for $C_1(\hat{\mathscr{F}}^{-1})$ — in the sense that all the required inputs are available as a part of the standard output of most SEM packages — is

$$\begin{aligned} C_1(\hat{\mathscr{F}}^{-1}) \quad = \quad & \frac{s}{2}\log\left[\frac{\frac{1}{n}tr(\hat{\Sigma}) + \frac{1}{2n}\left[tr(\hat{\Sigma}^2) + (tr(\hat{\Sigma}))^2 + 2\sum_{j=1}^{p+q}\left(\hat{\sigma}_{jj}\right)^2\right]}{s}\right] \\ & -\frac{1}{2}(p+q+2)\log\left|\hat{\Sigma}\right| + \frac{1}{2}\left[(p+q) + \frac{1}{2}\left(p+q\right)\left(p+q+1\right)\right]\log(n) \\ & +\frac{1}{4}\left(p+q\right)\left(p+q-1\right)\log(2), \end{aligned} \tag{31}$$

where

$$s = dim(\hat{\mathscr{F}}^{-1}) = rank(\hat{\mathscr{F}}^{-1}) = \frac{1}{2}(p+q)(p+q+3). \tag{32}$$

Therefore, $C_1(\hat{\mathscr{F}}^{-1})$ avoids the construction of $IFIM$ which is very attractive. This shows the scalability properties of the $ICOMP$ criterion.

The other new forms of $ICOMP$ criteria for SEM are:

$$ICOMP\left(\hat{\mathscr{F}}^{-1}\right)_{PEU} = n(p+q)\log(2\pi) + n\log|\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}S) + 3k + 2C_1(\hat{\mathscr{F}}^{-1}), \tag{33}$$

and

$$\begin{aligned} ICOMP\left(\hat{\mathscr{F}}^{-1}\right)_{PEU\_Miss} =& n(p+q)\log(2\pi) + n\log|\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}S) + k \\ & + 2(\frac{nk}{n-k-2}) + 2C_1(\hat{\mathscr{F}}^{-1}). \end{aligned} \tag{34}$$

The consistent and misspecification and robust forms of $ICOMP$-type criteria are obtained in a similar fashion which are given by

$$CICOMP(\hat{\mathscr{F}}^{-1})_{PEU} = n(p+q)\log(2\pi) + n\log|\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}S) + k(1+2\log(n)) + 2C_1(\hat{\mathscr{F}}^{-1}), \tag{35}$$

and

$$\begin{aligned} CICOMP(\hat{\mathscr{F}}^{-1})_{PEU\_Miss} =& n(p+q)\log(2\pi) + n\log|\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}S) + k \\ & + 2\log\left(\frac{nk}{n-k-2}\right) + 2C_1(\hat{\mathscr{F}}^{-1}). \end{aligned} \tag{36}$$

Note that these two above forms of $ICOMP$ criteria are both consistent and misspecification resistant.

Comparing AIC, $CAIC$, $CAICF$, BMS, and the different forms of $ICOMP$-type criteria, we see that the difference between these criteria are in their penalty terms. For convenience, we summarize these model selection criteria in Table 1, which we will use in our numerical examples. In the next section we give two numerical examples to demonstrate the performance

Table 1: Summary of Information Criteria and Their Penalties.

Information Criteria	Penalty term
AIC	$2k$
$CAIC$	$k\left[\log(n)+1\right]$
$CAICF$	$k\left[\log(n)+2\right]+\log\lvert\hat{\mathscr{F}}\rvert$
BMS	$k\log(n)+(\frac{nk}{n-k-2})+\log\lvert\hat{\mathscr{F}}\rvert$
$ICOMP$	$2C_1(\hat{\Sigma}(\hat{\theta}))$
$ICOMP(IFIM)$	$2C_1(\hat{\mathscr{F}}^{-1})$
$CICOMP(IFIM)_{PEU}$	$k(1+2\log(n))+2C_1(\hat{\mathscr{F}}^{-1})$
$CICOMP(IFIM)_{PEU\_Miss}$	$k+2\log\left(\frac{nk}{n-k-2}\right)+2C_1(\hat{\mathscr{F}}^{-1})$

of these information criteria in SEM.

5. Numerical Results

5.1. A Large Scale Monte Carlo Simulation Study

In this simulation study, we used the widely known general SEM simulation protocol given in [15]. This model is formed from 1 latent exogenous variable, 2 latent endogenous variables, 2 observed exogenous variables, and 4 observed endogenous variables. We slightly modify the structure of [15] in θ_ϵ matrix to achieve convergence in LISREL. We generated 500 sample covariance matrices for four sample sizes (100, 400, 1000, and 4000). Each sample covariance matrix was analyzed for each of the three analytic models given below using LISREL.

Figure 1 shows the population parameters of the model and the path diagram as expressed in LISREL notation, with parameters shown here.

$$\Lambda_x=\begin{bmatrix}1.00\\0.50\end{bmatrix},\Lambda_y=\begin{bmatrix}1.00&0\\0.95&0\\0&1.00\\0&0.90\end{bmatrix},\Gamma=\begin{bmatrix}-0.60\\-0.25\end{bmatrix},\Phi=\begin{bmatrix}7\end{bmatrix},B=\begin{bmatrix}0&0\\0.60&0\end{bmatrix},$$

$$\Psi=\begin{bmatrix}5.00&0\\0&4.00\end{bmatrix},\theta_\delta=\begin{bmatrix}3.00&0\\0&2.50\end{bmatrix},\theta_\epsilon=\begin{bmatrix}3.00&0&0&0\\0&3.00&0&0\\0&0&4.00&0\\0&0&0&4.00\end{bmatrix}$$

Once the model parameters are specified, the implied population covariance matrix is ob-

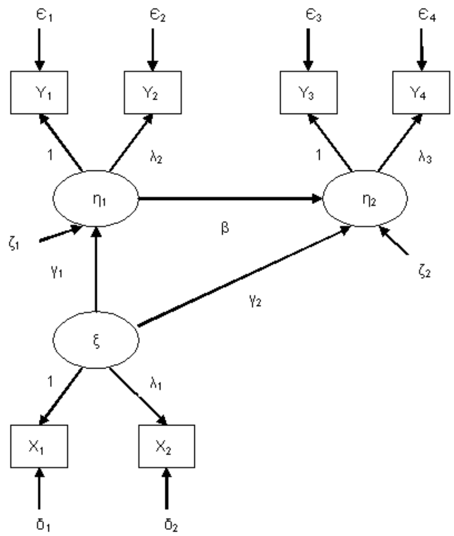

Figure 1: True model with population parameters and model specification conditions.

tained through Equation (5). For the analysis of this covariance matrix, we used three different analytic models, given below.

- Overfitting model: AM1 (15 free parameters)
  $\Lambda_x(1,1) = \Lambda_y(1,1) = \Lambda_x(3,2) = 1$
  $\Gamma, \Phi, B, \Psi, \theta_\delta, \theta_\epsilon$ are estimated freely

$$\Lambda_x = \begin{bmatrix} 1.00 \\ \lambda_1 \end{bmatrix}, \Lambda_y = \begin{bmatrix} 1.00 & 0 \\ \lambda_2 & 0 \\ 0 & 1.00 \\ 0 & \lambda_3 \end{bmatrix}, \Gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \Phi = \begin{bmatrix} \phi \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ \beta & 0 \end{bmatrix},$$

$$\Psi = \begin{bmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{bmatrix}, \Theta_\delta = \begin{bmatrix} \theta_{\delta_1} & 0 \\ 0 & \theta_{\delta_2} \end{bmatrix}, \Theta_\epsilon = \begin{bmatrix} \theta_{\epsilon_1} & 0 & 0 & 0 \\ 0 & \theta_{\epsilon_2} & 0 & 0 \\ 0 & 0 & \theta_{\epsilon_3} & 0 \\ 0 & 0 & 0 & \theta_{\epsilon_4} \end{bmatrix}$$

- Pseudo true model: AM2 (13 free parameters)
  $\Lambda_x(1,1) = \Lambda_y(1,1) = \Lambda_x(3,2) = 1$
  $\theta_\epsilon(1,1) = \theta_\epsilon(2,2), \theta_\epsilon(3,3) = \theta_\epsilon(4,4)$
  $\Gamma, \Phi, B, \Psi, \theta_\delta$ are estimated freely

$$\Lambda_x = \begin{bmatrix} 1.00 \\ \lambda_1 \end{bmatrix}, \Lambda_y = \begin{bmatrix} 1.00 & 0 \\ \lambda_2 & 0 \\ 0 & 1.00 \\ 0 & \lambda_3 \end{bmatrix}, \Gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \Phi = \begin{bmatrix} \phi \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ \beta & 0 \end{bmatrix},$$

$$\Psi = \begin{bmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{bmatrix}, \Theta_\delta = \begin{bmatrix} \theta_{\delta_1} & 0 \\ 0 & \theta_{\delta_2} \end{bmatrix}, \Theta_\epsilon = \begin{bmatrix} \theta_{\epsilon_1} & 0 & 0 & 0 \\ 0 & \theta_{\epsilon_1} & 0 & 0 \\ 0 & 0 & \theta_{\epsilon_3} & 0 \\ 0 & 0 & 0 & \theta_{\epsilon_3} \end{bmatrix}$$

- Underfitting model: AM3 (11 free parameters)
  $\Lambda_x(1,1) = \Lambda_y(1,1) = \Lambda_x(3,2) = 1$
  $\Lambda_x(2,1) = \Lambda_y(2,1)$
  $\Gamma(2,1) = 0$
  $\theta_\epsilon(1,1) = \theta_\epsilon(2,2), \theta_\epsilon(3,3) = \theta_\epsilon(4,4)$
  $\Phi, B, \Psi, \theta_\delta$ are estimated freely

$$\Lambda_x = \begin{bmatrix} 1.00 \\ \lambda_1 \end{bmatrix}, \Lambda_y = \begin{bmatrix} 1.00 & 0 \\ \lambda_1 & 0 \\ 0 & 1.00 \\ 0 & \lambda_3 \end{bmatrix}, \Gamma = \begin{bmatrix} \gamma_1 \\ 0 \end{bmatrix}, \Phi = \begin{bmatrix} \phi \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ \beta & 0 \end{bmatrix},$$

$$\Psi = \begin{bmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{bmatrix}, \Theta_\delta = \begin{bmatrix} \theta_{\delta_1} & 0 \\ 0 & \theta_{\delta_2} \end{bmatrix}, \Theta_\epsilon = \begin{bmatrix} \theta_{\epsilon_1} & 0 & 0 & 0 \\ 0 & \theta_{\epsilon_1} & 0 & 0 \\ 0 & 0 & \theta_{\epsilon_3} & 0 \\ 0 & 0 & 0 & \theta_{\epsilon_3} \end{bmatrix}$$

Based on the above AM structures above, we carried out a large scale Monte Carlo simulation study. Since we have three AMs and four different sample sizes $n = 100, 400, 1000,$ and 4000, we have in total 12 different simulation experiments. For each simulation experiment, we performed 500 replications and scored each of the information criteria to study their empirical performance based on their hit ratios.

One of the important characteristics of AM3 is that it is a misspecified model according to [15]; that is misspecified as it relates to underfitting . We note that his misspecification does not indicate to distributional misspecification since there are many other ways we can misspecify a model.

The steps of our Monte Carlo simulation are as follows,

1. The covariance matrices are generated with population parameters given in Figure 1, using PRELIS.

2. Each generated sample covariance matrix is analyzed by means of the three analytic models using LISREL. Goodness of fit indices and implied covariance matrices from LISREL outputs are saved.

3. LISREL outputs are passed to program are read into MATLAB, which computes both AIC-type and $ICOMP$-type criteria scores, and AMs selection frequencies. Our MATLAB modules are available from the authors upon request which marries LISREL results with MATLAB.

5.1.1. Results of Monte Carlo Simulation Experiments

Table 2: Correct model selection frequency by criteria (in %).

Information Criteria	n	$AM1$	$AM2$	$AM3$
AIC	100	16	84	0
	400	23.4	76.6	0
	1000	41.6	58.6	0
	4000	85.6	14.4	0
$CAIC$	100	0.4	98.8	0.8
	400	0.2	99.8	0
	1000	1.2	98.8	0
	4000	5.8	94.2	0
$CAICF$	100	0	98.4	1.6
	400	0.2	99.8	0
	1000	0.6	99.4	0
	4000	3.6	96.4	0
BMS	100	0	98.6	1.4
	400	0.2	99.8	0
	1000	1.0	99.0	0
	4000	5.6	94.4	0
$ICOMP$	100	98.6	1.4	0
	400	99.8	0.2	0
	1000	100	0	0
	4000	100	0	0
$ICOMP(IFIM)$	100	37	63	0
	400	48.8	51.2	0
	1000	65.4	34.6	0
	4000	95.4	4.6	0
$ICOMP(IFIM)_{PEU}$	100	13.8	86.2	0
	400	22.8	77.2	0
	1000	41.2	58.8	0
	4000	85.4	14.6	0
$ICOMP(IFIM)_{PEU\_Miss}$	100	9.2	90.8	0
	400	20.8	79.2	0
	1000	40.2	59.8	0
	4000	85.4	14.6	0
$CICOMP(IFIM)_{PEU}$	100	0	80.6	19.4
	400	0	100	0
	1000	0	100	0
	4000	0.2	99.8	0
$CICOMP(IFIM)_{PEU\_Miss}$	100	0	52.4	47.6
	400	0	100	0
	1000	0	100	0
	4000	0.2	99.8	0

In our simulation study, we intentionally did not include the nine traditional goodness of fit indices which were scored in [15], as they do not have the provision of taking into account different type of model misspecifications. Therefore, it is misleading even to score and report their results under such circumstances.

In Table 2, we summarize our result of percent hit ratios of the three AMs, AIC $CAIC$, $CAICF$, BMS, $ICOMP$, $ICOMP(IFIM)$, $ICOMP(IFIM)_{PEU}$, $ICOMP(IFIM)_{PEU\_Miss}$, $CICOMP(IFIM)_{PEU}$, and $CICOMP(IFIM)_{PEU\_Miss}$ criteria, for each sample size. We note that $CAIC$, $CAICF$, BMS, $ICOMP$, $ICOMP(IFIM)$, $ICOMP(IFIM)_{PEU}$, $ICOMP(IFIM)_{PEU\_Miss}$, $CICOMP(IFIM)_{PEU}$, and $CICOMP(IFIM)_{PEU\_Miss}$ all hit the pseudo true $AM2$ with very high frequencies. Specifically the performances of $CAIC$, $CAICF$, BMS, $CICOMP(IFIM)_{PEU}$, and $CICOMP(IFIM)_{PEU\_Miss}$ are outstanding and all above 90%.

We further observe that AIC's performance is not satisfactory as the sample size increases. For sample sizes $n = 100$, AIC picks $AM2$ 84% of the time, and as n gets large, AIC's hit percentage diminishes, and deteriorates. We know that AIC is not a consistent criterion. Specifically, AIC leans toward the $AM1$ which is an overfitting model which is the behaviour that is often demonstrated in the literature about AIC.

Of course one should note that, the large scale Monte Carlo simulation experiment we performed is only using one model set here. This can be easily extended to a large dimensional other model settings to further study the performances of these criteria in SEM. This requires high speed computation and computational capability on a super computer and a stand alone SEM software to carry out the task that is limited at this point.

5.2. A Real Data Example in Market Research

In this example, we apply the information criteria to a real data set from a soft drink company in Turkey. For propriety reasons we cannot disclose the name of the company. The data set in this study consists of a sample of $n = 135$ marketing surveys to study the product quality and to build a predictive operating model for the soft drink company. The goal of this survey was to enhance the market positioning and determine the influence of the investment of the company on their marketing campaign. For this data set we have seven characteristics which are measured to establish the companies objectives. These seven characteristics are: *refreshing*, *tastes great*, *good with meal*, *exhilarating*, *feels good*, *worth the money*, and *brands having products of good quality*.

Since we do not know a priori the generating model for this real data set, first we applied an Exploratory Factor Analysis (EFA) to learn which factors were related to the original variables. According to EFA results, three latent variables were named as *"taste"*, *"feeling"*, and *"quality"*, based on pre-determined factors. In the EFA model, variables which were loaded with factor 1 were: *"refreshing"*, *"tastes great"*, and *"good with meal"*. These were taken as exogenous variables and related to the latent exogenous variable *"taste"*.

Variables *"exhilarating"*, and *"feels good"* were loaded with factor 2. These were taken as exogenous variables and related to the latent exogenous variable; *"feeling."* Finally, variables *"worth money"* and *"quality products"* were loaded with factor 3. These were taken as endogenous variables and related to the latent endogenous variable *"quality"*.

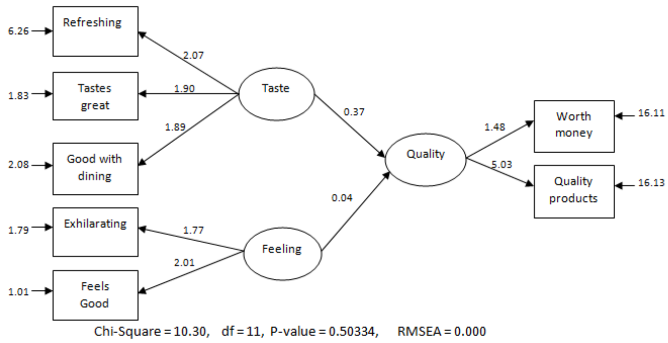

Figure 2: Full SEM with 17 free parameters (Model 1).

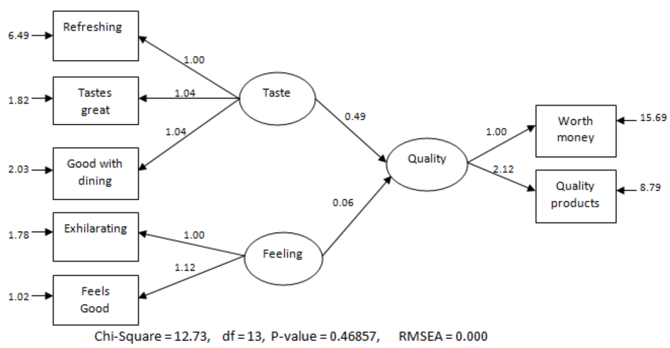

Figure 3: Full SEM with 15 free parameters (Model 2).

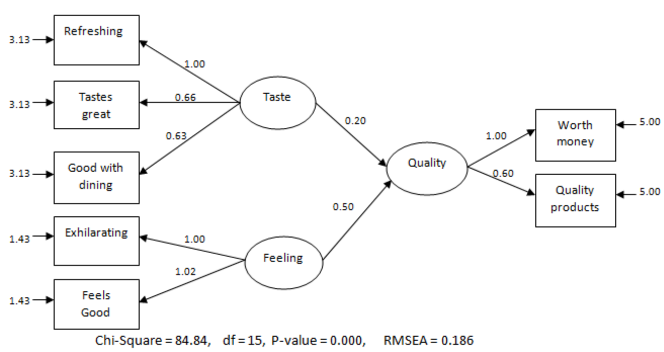

Figure 4: Full SEM with 13 free parameters (Model 3).

Since the information criteria used in the study were not applied to the just identified models, over-identified models were chosen. In accordance with the t-rule [19], one of the rules of identification must have free parameters under 28 to be over identified. Accordingly, three different models with free parameters, 17, 15, and 13 were compared to assess the information criteria. The path diagram of each of these three models, obtained from LISREL, are given in Figures 2, 3, and 4. Model 1 was fit to the data with 17 free parameters in Figure 2 with $P-value = 0.29 > 0.05$. In this model, the diagonal elements of the covariance of latent exogenous variables and a single path coefficient between latent and observed endogenous were fixed. The other parameters were estimated freely. Model 2 was fit to the data with 15 free parameters shown in Figure 3 with $P - value = 0.42 > 0.05$. It was restricted by setting to 1 a single path coefficient in each latent variable, and equating diagonal elements of the covariance matrix of latent variables. In the case of Model 3, the model was fit to the data with 13 free parameters shown in Figure 4 with $P - value = 0.00009 < 0.05$. It also was restricted by setting to 1 as single path coefficient in each latent variable, and equating the measurement errors of the observed variables belonging to each latent variable.

All information criteria scores for each of these three SEMs, obtained from LISREL and our MATLAB module, are reported in Table 3. According to the minimum of the information

Table 3: The values of criteria for three General SEMs.

Model Selection Criteria	**Model 1**	**Model 2**	**Model 3**
AIC	4218.1	**4216.6**	4298.8
$CAIC$	4284.5	**4275.2**	4349.5
$CAICF$	4333.8	**4322.4**	4389.0
BMS	4319.6	**4309.5**	4377.6
$ICOMP$	**4187.9**	4190.5	4275.9
$ICOMP(IFIM)$	**4245.7**	4248.6	4328.3
$ICOMP(IFIM)_{PEU}$	4279.7	**4278.6**	4354.3
$ICOMP(IFIM)_{PEU\_Miss}$	4285.2	**4282.9**	4357.5
$CICOMP(IFIM)_{PEU}$	4429.4	**4410.8**	4468.8
$CICOMP(IFIM)_{PEU\_Miss}$	4456.8	**4432.0**	4484.8

criteria, we choose the "best" model among those compared, as the model that achieves the best balance of fit with respect to parameter cardinality. In this case, all criteria are minimized at Model 2. These are indicated by the boldfaced scores in Table 3. We note that for this data set all criteria agreed. In general, this would not typically be the case with other real data sets.

Based the results above, we can use Model 2 as our operating predictive SEM model in the subsequent surveys for market campaign to study the product quality. In passing we note that, although we used the EFA model to determine the SEM structure for this real data set, one can also use an expert pattern analysis method using the genetic algorithm with information complexity to obtain the "best" factor pattern structure to establish a confirmatory factor analysis to derive the SEMs. This new approach is well explained in [30] for further reading. We have not pursued this approach in this paper.

6. Conclusions and Discussion

In this paper, we showed the performance of several information criteria, some old and new ones in general SEMs, under different sample sizes, and different models. Based on our results from this specific large scale Monte Carlo simulation experiment, for general SEM, $CAIC$, $CAICF$, BMS, $CICOMP(IFIM)_{PEU}$ and $CICOMP(IFIM)_{PEU\_Miss}$ criteria show the best performance. On the other hand, the performance of AIC seems to degrade as sample sizes increase - it tends to select overfitting models. This behavior is based on the sensitivity of the log likelihood function to sample size. Because $CAIC$, $CAICF$, BMS, $CICOMP(IFIM)_{PEU}$ and $CICOMP(IFIM)_{PEU\_Miss}$ criteria are consistent with respect to sample size (penalty term includes $\log(n)$), they become more accurate as the sample size gets larger. Our simulation results demonstrates the performance and the versatility of these criteria. In summary, we recommend that the information criteria: $CAICF$, BMS, $CICOMP(IFIM)_{PEU}$ and $CICOMP(IFIM)_{PEU\_Miss}$, be used in SEM. In the real example, we showed how one can build an operating predictive SEM in market research. We are currently studying other forms of misspecification such as the distributional misspecification, presence of high multicollinearity, and error variance heteroscedasticity within the SEM framework using the information criteria along with other applications. Our results will be reported elsewhere.

ACKNOWLEDGEMENTS This research was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) for the first author at the Department of Statistics, Operations, and Management Science at the University of Tennessee as a Visiting Scholar under the supervision of Professor Bozdogan. The first author extends her gratitude and thanks to Professor Bozdogan for the hospitality and conducive research atmosphere provided.

References

[1] H Akaike. Information theory and an extension of the maximum likelihood principle. In B.N. Petrox and F. Csaki, editors, *Second International Symposium on Information Theory.*, pages 267–281, Budapest, 1973. Academiai Kiado.

[2] H Akaike. Factor analysis and AIC. *Psychometrica*, 52(3):317–332, 1987.

[3] K A Bollen. *Structural Equations with Latent Variables.* John Wiley,New York, 1989.

[4] H Bozdogan. Model selection and akaike's information criteria (AIC): the general theory and its analytical extensions. *Psychometrica*, 52:317–332, 1987.

[5] H Bozdogan. A new information theoretic measure of complexity index for model evaluation in general structural equation models with latent variables. Rutgers the State Univarsity, June 13-16 1991. Symposium on Model Selection in Covariance Structures at the Joint Meeting of Psychometric Society and The Classification Society.

[6] H Bozdogan. Choosing the number of clusters, subset selection of variables, and outlier detection in the standard mixture-model cluster analysis. In E. Diday et al., editor, *New*

Approaches in Classification and Data Analysis, pages 169–177. Springer-Verlag, New York, 1994.

[7] H Bozdogan. Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, 44:62–91, 2000.

[8] H Bozdogan. *Statistical Data Mining Knowledge Discovery*, chapter Intelligent Statistical Data Mining with Information Complexity and Genetic Algorithms., pages 15–56. ChapmanHall,CRC, 2004.

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

[10] H Bozdogan and M Ueno. A unified approach to information-theoretic and bayesian model selection criteria. Crete, Greece, 2000. The 6th World Meeting of the International Society for Bayesian Analysis.

[11] N Cliff. Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*, 18:115–126, 1983.

[12] H Cramér. *Mathematical Methods of Statistics.* Princeton University Press, Princeton, New Jersey, 1946.

[13] E Deniz, H Bozdogan, and S Katraggadda. Structural equation modeling (SEM) of categorical and mixed-data using the novel gifi transformations and information complexity (ICOMP) criterion. *Journal of the School of Business Administration*, 40(1):86–123, 2011.

[14] E A Deniz. *Information Criteria in Structural Equation Models.* unpublished ph.d. thesis, Mimar Sinan Fine Arts University, 2007.

[15] X Fan and X Fan. Using SAS for monte carlo simulation research in SEM. *Structural Equation Modeling*, 12(2):299–333, 2005.

[16] B Frieden. *Physics from Fisher Information.* Cambridge University Press, Cambridge, UK, 1998.

[17] N Gheissari and A Bab-Hadiashar. Detecting cylinders in 3D range data using model selection criteria. Proceeding of Fifth International Conference on 3-D Digital Imaging and Modelling, 2005.

[18] K Jöreskog. *Testing Structural Equation Models.*, chapter Testing Structural Equation Models. Sage Publication, London, 1993.

[19] K Jöreskog and D Sörbom. *LISREL 7 A Guide to the Program and Applications 2nd Edition.* SPSS Inc, Chicago, 1989.

[20] S Kullback. *Information Theory and Statistics.* Dover Publications, Dover, New York, 1968.

[21] S Kullback and R Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[22] S A Mulaik. *Linear Causal Modeling with Structural Equations.* Chapman and Hall, 2009.

[23] C R Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81, 1945.

[24] C R Rao. Minimum variance and the estimation of several parameters. In *Proceedings of the Cambridge Philosophical Society.*, volume 43, page 280, 1947.

[25] C R Rao. Sufficient statistics and minimum variance estimates. In *Proceedings of the Cambridge Philosophical Society.*, volume 45, page 213, 1948.

[26] J Rissanen. Modeling by shortest data description. *Automatica*, pages 465–471, 1978.

[27] Y Sakamoto, M Ishiguro, and G Kitagawa. *Akaike Information Criterion Statistics.* Dordrecht, The Netherlands: Reidel, 1986.

[28] G Schwarz. Estimating the dimension of a model. *Annals of Statistics*, pages 461–464, 1978.

[29] M Van Emden. An analysis of complexity. In *Mathematical Centre Tracts.*, volume 35. Mathematisch Centrum, 1971.

[30] H Y Yang and H Bozdogan. Learning factor patterns in exploratory factor analysis using the genetic algorithm and information complexity as the fitness function. *Journal of Pattern Recognition Research*, 2:307–326, 2011.