EMPIRICAL LIKELIHOOD APPROACH ESTIMATION OF STRUCTURAL EQUATION MODELS

A Thesis
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree

Master of Arts

by YIYONG YUAN

Dr. Stanislav Kolenikov, Thesis Supervisor

DECEMBER 2007

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

EMPIRICAL LIKELIHOOD APPROACH ESTIMATION OF STRUCTURAL EQUATION MODELS

presented by Yiyong Yuan,

a candidate for the degree of Master of Arts,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Stanislav Kolenikov

Professor Paul Speckman

Professor Douglas J. Miller

Acknowledgments

I am indebted to Dr. Stanislav Kolenikov for his insightful choice of research topic, stimulating and patient guidance in the research. I enjoy the whole process of knowing him, attending his lecture and working as his research assistant.

YIYONG YUAN

University of Missouri at Columbia

Dec 2007

Empirical Likelihood Approach Estimation of Structural Equation Models

Yiyong Yuan

University of Missouri at Columbia, 2007

Supervisor: Stanislav Kolenikov

ABSTRACT

This thesis provides a preliminary investigation of empirical likelihood approach estimation of structural equation models. An auxiliary variable approach built on general estimating equation methods in the EL settings is followed. An auxiliary variable is proposed and estimation/inference based upon it is developed. Testing of model covariance structure for over-identified model is suggested. Asymptotic efficiency connection with Unweighted Least Squares estimator and multi-normal MLE is established. Estimation example of non-elliptical distribution data is provided.

Contents

List of Tables

Chapter 1

Introduction

Structural equation models are statistical techniques analyzing ”invariant” relation among random variables. It has multiple origins. Biometricians, sociologists, psychologists, and econometricians each make their share of contribution.

Biometrician Sewall Wright (1918) first introduced path diagram which actually functions as a system of covariance structure equations, and performed a path analysis that is known today a factor analysis. However, his pioneering accomplishments remained long unnoticed by others. In 1960’s, sociologists Blalock, Boudon and Duncan ”rediscovered” Wright’s path analysis by noting its potential usage in sociology research. Their works laid foundation for the synthesis of latent variable model and measurement models. Later in 1970’s, Jöreskog , Keesing and Wiley developed the most general structural equation model that is known as JKW or LISREL model and is widely used in current SEM notation. There are also representations proposed by Bentler and Weeks (1980) , McArdle and MCDonal (1984) that are essentially of similar representation capacity.

Estimation and inference procedures of early stage SEMs are introduced in the form of examples. The systematic estimation methods is credited to development of theory in multiple areas. SEMs with observable variables in econometrics and factor analysis in psychometrics made important contributions. Lawley (1940) , Anderson and Rubin (1956) , Jöreskog (1969) proposed hypothesis test in factor analysis. Bock and Bargmann (1966) proposed analysis of covariance. Jöreskog (1973) proposed normal theory maximum likelihood estimation of general SEMs. Jöreskog and Goldberg (1972) and Browne (1974, 1982, 1984) provided generalized least squares estimations. Particularly, Browne proposed estimator under elliptical distribution and Arbitrary Distribution Function (ADF) estimator what is also known for Weighted Least Squares (WLS) for distribution-free model. Bentler (1983) suggested models using higher-order moments to identify model that is otherwise not identified by covariance based estimation approach. Muthén (1984, 1987) applied SEMs to categorical and censored data. Bollen (1996) investigated a limited information, Two Stage Least Squares (2SLS) estimator for latent variable SEMs. For an excellent survey of general structural equation model, refer to Bollen (1989). It was written early but weathered the time well.

Since Jöreskog and Sorbom's LISREL, due to SEMs' increasing popularity in social sciences, psychology, and business, there has been many new computer packages arising up for SEM analysis: EQS, AMOS, CALIS, LISCOMP, Mplus, Mx and many others. For a review of LISREL, AMOS and EQS, refer to Kline (1998) and the webpage of Edward E. Rigdon [1] for other choices.

Although there has been tremendous development in SEM estimation technique, given the limitation of ADF in practical estimation, estimation of SEMs still faces the difficulty of non-normal/non elliptical data. According to Chou, Bentler and Satorra (1991), Muthén and Kaplan (1985, 1992), Hu, Bentler and Kano (1992), the ADF estimation performs poorly with small sample size or large model degree of freedom. Alternative approach of estimation might prove helpful here.

Empirical likelihood (EL) theory, as a non-parametric inference method, drew attention of statisticians with the establishment of the Empirical Likelihood Theorem (ELT) by Art. B. Owen in 1980's. Since then it has been developed by many researchers into a systematic theory and found huge following in econometrics. For a sketch of the contributors, refer to Owen (2001). Among

[1] http://www2.gsu.edu/ ˜ mkteer/

the theoretic contribution of the researchers, there are several contributors of incorporating ”side information” into the EL theory. Qin and Lawless (1994) established a way to incorporate ”side information” in the form of general estimating equations into empirical likelihood theory. Their result is capable of bringing non-parametric EL theory to parametric inferences. In fact Owen (1991) has introduced a EL estimation of linear regression problem. Qin and Lawless' result make possible application of EL theory to more complicated models.

In the following part, Chapter 2 reviews the EL theory, especially those part related to this thesis; Chapter 3 provides a more detailed review of SEMs to prepare and motivate Chapter 4's development of alternative estimation method with EL approach; Chapter 5 reports application examples and preliminary performance results of the new estimation methods on simulated data; Chapter 6 concludes the thesis.

Chapter 2

Empirical likelihood theory

2.1 Nonparametric likelihood (NPL) and NPL ratio

Parametric MLE is defined as

$$\hat{\theta_n} = \arg\max_{\theta \in \Theta} l(\theta|x)$$

where

$$l(\theta|x) = \log f(x|\theta)$$

and $f(x|\theta)$ is the pdf or pmf. Under appropriate regularity conditions, by Wald (1948) , or by Huber (1967) under a group of less restrictive conditions, MLE is strongly consistent, asymptotically normal and asymptotically efficient:

$$\hat{\theta_n} \xrightarrow{a.s} \theta_0$$

$$\sqrt{n}(\hat{\theta_n} - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$$

where $I(\theta)$ is the information matrix. MLE is invariant under 1-1 transformation, that is,

$$\widehat{f(\theta)} = f(\hat{\theta}),$$

for 1-1 function f.

The idea of MLE can be applied to distribution function as well. Since real world sample is observed in discrete form, MLE is actually maximizing the point probability mass of an observed sample, i.e. $Pr(X = x) = F(x) - F(x-)$. Naturally, if empirical distribution function is defined as

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i},$$

then the Nonparametric Likelihood (NPL) can be defined as

$$L(F) = \prod_{i=1}^{n}\{F(X_i) - F(X_i-)\}.$$

Again, similar to MLE maximizing the parametric likelihood, for any DF $F \neq F_n$,

$$L(F) < L(F_n).$$

Thus F_n is the non-parametric MLE of F. Further, analogous to the parametric likelihood ratio test, define nonparametric likelihood ratio

$$R(F) = \frac{L(F)}{L(F_n)}$$

and profile likelihood function:

$$R(\theta) = \sup\{R(F)|T(F) = \theta, F \in \mathbb{F}\}$$

where $T(F)$ is a functional of the distribution F. For example, it might be

$$T(F) = \int_{-\infty}^{\infty} xF(dx) = \mathbb{E}(x)$$

With all these settings, we can test:

$$H_0 : T(F_0) = \theta_0$$

with the confidence set of the form

$$\{\theta | R(\theta) \geq r_0\}$$

where r_0 can be chosen to deliver a given coverage as implied by the Empirical Likelihood Theorem (ELT).

Usually, there might be repetitive data in the sample, however, by some algebraic result, we can use the same definition of profile likelihood function as if there are no ties (Owen 1989, 2001). Now it comes to the problem making inference, it depends on a remarkable result similar to Wilks' theorem, the Empirical Likelihood theorem (ELT) .

2.2 Empirical Likelihood Theorem

Theorem 2.2.1 *Let $x_1, ..., x_n$ be independent random vectors in $\Re^p$ with common distribution F_0 which has mean μ_0 and finite covariance matrix V_0 of rank $q > 0$. Then*

$$C_{r,n} = \{\sum_{i=1}^{n} w_i x_i | \prod_{i=1}^{n} n w_i \geq r, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}$$

is a convex set and

$$-2 \log R(\mu_0) \xrightarrow{d} \chi^2(q)$$

as $n \to \infty$, where

$$R(\mu_0) = \max\{\prod_{i=1}^{n} n w_i | \sum_{i=1}^{n} w_i x_i = \mu_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}$$

An example: Applying ELT to Estimating Equations

By defining

$$R(\theta) = \max\{\prod_{i=1}^{n} nw_i | \sum_{i=1}^{n} w_i m(x_i, \theta) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\} \tag{2.1}$$

ELT can be immediately applied as follows.

Theorem 2.2.2 *Let $x_1, ..., x_n$ be independent random vectors in $\Re^d$ with common distribution F_0. For $\theta \in \Theta \subseteq \Re^p$, and $X \in \Re^d$, let $m(X, \theta) \in \Re^s$. Let $\theta_0 \in \Theta$ such that $\mathbb{V}(m(X_i, \theta_0))$ is finite and has rank $q \geq 0$. If θ_0 satisfies*

$$\mathbb{E}(m(X, \theta_0)) = 0,$$

then

$$-2 \log R(\theta_0) \xrightarrow{d} \chi^2(q)$$

Finally, the empirical likelihood confidence region has the same rate of coverage accuracy as parametric likelihood, jacknife and bootstrap methods. The coverage error decreases to 0 at the rate of $\frac{1}{n}$ as $n \to \infty$. The empirical likelihood confidence region also has competitive power efficiency compared to other inference procedures (Owen, 2001).

2.3 Maximum Empirical Likelihood Estimator (MELE)

Similar to MLE, define maximum empirical likelihood estimator (MELE)

$$\tilde{\theta} = \arg\max_{\theta} R(\theta)$$

where $R(\theta)$ is defined as in equation (2.1) By Qin and Lawless (1994) , we have following results, which is largely that summarized by Owen (2001) in Theorem 3.6.

Theorem 2.3.1 *Let $X_i \in \Re^d$ be iid random vectors, and suppose that $\theta_0 \in \Re^p$ is uniquely de-*

termined by $\mathbb{E}(m(X,\theta)) = 0$*, where* $m(X,\theta)$ *takes values in* $\Re^{p+q}$*, for* $q \geq 0$*. If there is a neighborhood* Θ *of* θ_0 *and a function* $M(x)$ *with* $\mathbb{E}(M(X)) \leq \infty$*, for which:*

- $\mathbb{E}[\partial m(X,\theta_0)/\partial\theta]$ *has rank* p*,*
- $\mathbb{E}[m(X,\theta_0)m(X,\theta_0)']$ *is positive definite,*
- $\partial m(x,\theta)/\partial\theta$ *is continuous for* $\theta \in \Theta$*,*
- $\partial^2 m(x,\theta)/\partial\theta\partial\theta'$ *is continuous in* θ*, for* $\theta \in \Theta$*,*
- $\|m(x,\theta)\|^3 \leq M(x)$ *for* $\theta \in \Theta$*,*
- $\|\partial m(x,\theta)/\partial\theta\| \leq M(x)$*, for* $\theta \in \Theta$*,*
- $\|\partial^2 m(x,\theta)/\partial\theta\partial\theta'\| \leq M(x)$*, for* $\theta \in \Theta$*,*

then MELE is consistent and

$$\lim_{n\to\infty} n\mathbb{V}(\tilde{\theta}) = \left\{\mathbb{E}\big(\frac{\partial m}{\partial\theta}\big)'[\mathbb{E}(mm')]^{-1}\mathbb{E}\big(\frac{\partial m}{\partial\theta}\big)\right\}^{-1}$$

and this asymptotic variance is at least as small as that of any estimator of the form $\mathbb{E}A(\theta)m(X,\theta) = 0$*, where* $A(\theta)$ *is a* $p \times (p+q)$ *matrix with rank* p *for all values of* θ*. Furthermore,*

$$-2\log(R(\theta_0)/R(\tilde{\theta})) \xrightarrow{d} \chi^2_p$$

and

$$-2\log(R(\tilde{\theta})) \xrightarrow{d} \chi^2_q$$

as $n \to \infty$

It might be interesting to examine how restrictive these smoothness conditions concerning estimating equations are compared with those regularity conditions concerning likelihood function and distribution for the asymptotic properties of MLE to hold. It should be observed that these conditions are are not very restrictive compared to Wald (1948) though they do seem to be more restrictive than those required by Huber (1967). Huber's result needs no second and higher order derivative of likelihood function.

2.4 Shape of Empirical likelihood Confidence Regions

Empirical likelihood confidence region of the mean with the form $C_\mu = \{\mu | R(\mu) \geq r_0\}$ is convex. To show its convexity, it suffices to show

$$R(\mu_0) = \max\{\prod_{i=1}^{n} nw_i | \sum_{i=1}^{n} w_i x_i = \mu_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}$$

is a concave function. It is easy to see that for $R(\mu_1)$ with maximizing w_{i1} and $R(\mu_2)$ with maximizing w_{i2}, if we take $w_i = tw_{i1} + (1-t)w_{i2}$ for $0 \leq t \leq 1$, then w_i satisfy $w_i \geq 0, \sum_{i=1}^{n} w_i = 1$ and $\sum_{i=1}^{n} w_i x_i = t\mu_1 + (1-t)\mu_2$, by definition of $R(\mu)$, $R(\mu) \geq \prod_{i=1}^{n} nw_i$, but

$$\log\{\prod_{i=1}^{n} nw_i\} = \sum_{i=1}^{n} \log(nw_i)$$

$$= \sum_{i=1}^{n} \log[tnw_{i1} + (1-t)nw_{i2}]$$

$$\geq \sum_{i=1}^{n} \{t \log[nw_{i1}] + (1-t) \log[nw_{i2}]\}$$

$$= t \sum_{i=1}^{n} \log[nw_{i1}] + (1-t) \sum_{i=1}^{n} \log[nw_{i2}]$$

this is actually showing that

$$R(\mu_0) = \max\{\log\{\prod_{i=1}^{n} nw_i\}|\sum_{i=1}^{n} w_i x_i = \mu_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}$$

is a concave function. It has to be noted that this result says nothing about the convexity of confidence region in the form $C_\theta = \{\theta | r_1 \geq R(\theta) \geq r_0\}$.

When it comes to the estimating equations case, confidence regions of aforementioned form can be constructed either by direct extension of ELT of the mean or the result by Qin and Lawless. The confidence region is defined as either

$$C_\theta = \{\theta | R(\theta) \geq r_0\}$$

or

$$C_\theta = \{\theta | \frac{R(\theta)}{R(\tilde{\theta})} \geq r_0\} \tag{2.2}$$

where $R(\theta)$ is defined as

$$R(\theta) = \max\{\log\{\prod_{i=1}^{n} nw_i\}|\sum_{i=1}^{n} w_i m(x_i, \theta) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\} \tag{2.3}$$

and $\tilde{\theta}$ is defined as

$$\tilde{\theta} = \arg\max_{\theta} R(\theta)$$

For either case, it also suffices to check if $R(\theta)$ is still a concave function. Considering $R(\theta_1)$ with maximizing w_{i1} and $R(\theta_2)$ with maximizing w_{i2}, similarly, if we take $w_i = tw_{i1} + (1-t)w_{i2}$ for $0 \leq t \leq 1$, then w_i satisfy $w_i \geq 0, \sum_{i=1}^{n} w_i = 1$ and

$$\log\{\prod_{i=1}^{n} nw_i\} \geq tR(\theta_1) + (1-t)R(\theta_2),$$

but it may not be true that

$$\sum_{i=1}^{n} w_i m(x_i, \theta) = \sum_{i=1}^{n} w_i m(x_i, t\theta_1 + (1-t)\theta_2) = 0,$$

therefore $R(\theta) \geq \prod_{i=1}^{n} nw_i$ may not hold. Hence, generally, the confidence region for parameter obtained through EL estimating equation may not be convex.

Chapter 3

Structural equation models

Structural equation models includes factor analysis, regression and simultaneous equations as special cases. Among the estimation methods of structural equation model, Maximum Likelihood Estimation and Minimum Distance Estimation (MDE) are two popular ones to make statistical inference.

A general structural equation model consists of a latent model relating the latent variables:

$$\eta = \alpha_\eta + \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta \tag{3.1}$$

and a measurement model linking the latent and the observed variables:

$$\mathbf{y} = \alpha_y + \mathbf{\Lambda}_y\eta + \varepsilon \tag{3.2}$$

$$\mathbf{x} = \alpha_x + \mathbf{\Lambda}_x\xi + \delta \tag{3.3}$$

where ξ are exogenous latent variables, η are endogenous latent variables, x and y are indicators of the exogenous and endogenous latent variables, and error terms $\zeta, \varepsilon, \delta$ are uncorrelated with one another and ξ. The set of model parameters includes the intercepts α, the effect of exogenous variables on the endogenous ones $\mathbf{\Gamma}$, the effect of endogenous variables on one another $\mathbf{B}$, the loadings $\mathbf{\Lambda}$, and variances and covariances of unrestricted elements of $\mathbb{V}[(\xi', \zeta', \varepsilon', \delta')']$. Equations (3.1) – (3.3) imply a certain covariance structure referred to as $\Sigma(\theta) = \mathbb{V}\Big[\begin{pmatrix} x \\ y \end{pmatrix}\Big]$.

In SEMs settings there is an identification issue. The idea is about the existence of unique solution to equations $\Sigma(\theta) = \Sigma$. A model is said to be identified if solving $\Sigma(\theta) = \Sigma$ leads to only one solution. If the conditions is just enough to identify all the elements of θ, the model is just or exactly identified. If there is at least one θ that has more than one condition to determine from, the model is termed over-identified. For a detailed discussion on identification, see Bollen (1989).

Existing estimation methods of SEMs include Unweighted Least Squares, Maximum Likelihood and Generalized Least Squares.

Unweighted Least Squares estimator is defined as

$$\hat{\theta}_{ULS} = \arg\min_{\theta}\{\frac{1}{2}tr[(S - \Sigma(\theta))^2]\}$$

The idea is to minimizing the "distance" between the sample variance matrix and implied variance matrix by define the distance as the unweighted average of the squared difference between the elements of two variance matrices. Browne (1982) provides ways to make statistical inference.

MLE under multi-normality is defined as

$$\arg\min_{\theta} F_{ML} = \arg\min\{\log|\Sigma(\theta)| + tr[S\Sigma^{-1}(\theta)] - \log|S| - dim(x'y')'\}$$

It can be obtained either by the multiple normality of observed variables of by the Wishart distribution of the unbiased variance matrix. MLE has well known good properties: it's scale free, consistent and asymptotically efficient.

Defining the "distance" between two matrices has multiple choices. Contrasted to the uniform weights put on each element of variance difference matrix, Generalized Least Squares (GLS) is defined as

$$\arg\min_{\theta}\{\frac{1}{2}tr[[(S-\Sigma(\theta))V^{-1}]^2]\}$$

where V^{-1} is a weight matrix which might be a positive definite constant one, like identity matrix, leading back to ULS, or might be a random matrix converging in probability to a positive definite matrix. When observed variable follows distributions without excessive kurtosis including normal so that elements of S has multinormal asymptotic distribution, if V^{-1} is chosen to be $V^{-1} \xrightarrow{p} \Sigma^{-1}$, then the GLS will have the desirable asymptotic properties of MLE. A popular choice of V^{-1} is S^{-1}. According to Lee and Jennrich (1969) , choosing $V^{-1} = \Sigma^{-1}(\hat{\theta})$, where $\hat{\theta}$ is MLE, will lead back to MLE. GLS is also scale invariant.

MLE needs the assumption of multinormality of the observed variables, while GLS expands the range of possible distributions, when the distribution of the data is nonnormal or heavy tailed, using MLE or GLS lead to the loss of asymptotic efficiency.

The Weighted Least Squares Estimator is defined as

$$\arg\min_{\theta}\{[S^* - \Sigma(\theta)^*]'W^{-1}[S^* - \Sigma(\theta)^*]\}$$

where $S^* = Vech(S)$, and $\Sigma(\theta)^* = Vech[\Sigma(\theta)]$. W^{-1} is a positive definite weight matrix relating to V: $W = diag[\mathbb{V}(S^*)]$.

By Browne (1982, 1984) , if W is chosen to be asymptotic sample covariance matrix of S^* or a consistent estimator of it, the WLS is consistent and asymptotically efficient as long as the 8-th order moments of the observed variables are finite. This is the so-called Arbitrary Distribution Function (ADF) estimation method.

In fact, by Browne (1974) , the fitting function of WLS can be written as

$$\frac{1}{n}tr\{[S-\Sigma(\theta)]'V^{-1}\}^2$$

by which the connection between WLS and it specialized version GLS, MLE and ULS can be established by different choice of V^{-1}.

To incorporate the nonnormality in the form of variable kurtosis, Browne (1984) introduced another specialized version of WLS with the fitting function

$$\frac{1}{2}(K+1)^{-1}tr[(S-\Sigma(\theta))V^{-1}]^2 - C_1[tr(S-\Sigma(\theta))V^{-1}]^2$$

where $c_1 = \frac{K}{4(K+1)^2+2(p+q)K(K+1)}$ and K is the kurtosis parameter.

It is interesting to note the similarity between WLS and GMM(Hansen, 1982, Hall 2005). Both of them define a ”distance” between moments, or estimating equations.

The asymptotic efficiency result about the WLS estimator is strong in the sense that it assumes nothing about the distribution. However, as noted in the introduction, its performance under small sample size or large model degree of freedom is poor.

Since empirical likelihood theory and the estimating equation result built upon it offer a distribution free technique to make statistical inference, it is very interesting to investigate the application of these theory to SEMs for inference on heavy tailed, and especially non-elliptical data. This motivates why we need non-elliptical data for test of estimation feasibility. This will be discussed later in Chapter 5

Chapter 4

Application of empirical likelihood theory to SEM

Owen (1991) has introduced the approach of auxiliary variable to apply EL theory to linear regression model. The idea is as follows. Rewrite linear regression Least Squares procedure:

$$(X'X)\beta = X'Y \Rightarrow \frac{1}{n}\sum_{i=1}^{n}(x_i x_i'\beta - x_i y_i')$$

$$\Rightarrow \mathbb{E}(x_i x_i'\beta - x_i y_i') = 0$$

where y_i' is 1 by m, β is k by m. when $m = 1$, it's the single equation case.
Define an auxiliary matrix with dimension k by m

$$Z = x_i x_i'\beta - x_i y_i'$$

In fact, this works for either single-equation model or multivariate regression model as long as there is no identification problem.

For general structural equation models with latent variable, recall the theorem by Qin and Lawless (1994), it seems straightforward to define appropriate auxiliary variable. However, the auxiliary variable can be derived based on different choices of existing fitting functions. Establishing a connection between them and making selection is a more important task. The first section derived auxiliary variables based on the ULS fitting function. the second section derived auxiliary variable based on multinormal ML fitting function, found the common part in the two auxiliary variable, and arrived at a better auxiliary variable.

4.1 Applying ELT to Unweighted Least Squares estimation of general SEM

Suppose θ is of dimension u, covariance matrix of observed variable Σ is $v \times v$. By Unweighted Least Squares,

$$\hat{\theta}_{ULS} = \arg\min_{\theta}\{\frac{1}{2}tr[(\Sigma - \Sigma(\theta))^2]\},$$

where Σ is the covariance matrix of $Z = (\mathbf{y}', \mathbf{x}')'$, the observed variable vector. The population counterpart of S, Σ, rather than S is used here, since it offers a formality convenience to write an estimating equation form such as $\mathbb{E}[m(Z,\theta)] = 0$ that is meant to represent the sample version $\frac{1}{n}\sum_{i=1}^{n} m(Z_i,\theta) = 0$. Taking derivatives with respect to θ leads to

$$\nabla_{\theta}[\Sigma - \Sigma(\theta)]'[\Sigma - \Sigma(\theta)] = 0,$$

where $\nabla_{\theta}[\Sigma - \Sigma(\theta)]'$ is a Cubix of dimension $u \times v \times v$, by chain rule and derivative of matrix trace. For a reference to matrix calculus notation used here, refer to matrix calculus appendix of

Jon Dattorro (2005). Since Σ is constant and $\Sigma(\theta)$ symmetric, above equation simplifies to

$$-\nabla_\theta[\Sigma(\theta)][\Sigma - \Sigma(\theta)] = 0,$$

This is equivalent to

$$-\nabla_\theta[vec[\Sigma(\theta)]]vec[\Sigma - \Sigma(\theta)] = 0.$$

Writing Σ explicitly and rearranging leads to

$$\nabla_\theta[vec[\Sigma(\theta)]]vec[\mathbb{E}(ZZ') - \Sigma(\theta)] = \nabla_\theta[vec[\Sigma(\theta)]]vec[\mathbb{E}(Z)\mathbb{E}(Z)']$$

Either by model assumption of SEM or considering the centered case, $\mathbb{E}(Z) = 0$, therefore,

$$\nabla_\theta[vec[\Sigma(\theta)]]vec[\mathbb{E}(ZZ') - \Sigma(\theta)] = 0$$

and

$$\mathbb{E}\{\nabla_\theta[vec[\Sigma(\theta)]]vec[ZZ' - \Sigma(\theta)]\} = 0, \tag{4.1}$$

which is finally equivalent to

$$\mathbb{E}\{\nabla_\theta[H * vech[\Sigma(\theta)]]vech[ZZ' - \Sigma(\theta)]\} = 0, \tag{4.2}$$

where

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

is of dimension $\frac{v(v+1)}{2} \times \frac{v(v+1)}{2}$ to assign appropriate coefficients to elements of $vech[\Sigma(\theta)]$ formed by diagonal (assigned "1") and off-diagonal elements (assigned "2") of the matrix $\Sigma(\theta)$ since it is symmetric.

Define an auxiliary variable vector

$$U = \nabla_\theta [H * vech[\Sigma(\theta)]] vech[ZZ' - \Sigma(\theta)]$$

with dimension equal to number of parameters in the model. By theorem 2.2.2 If $\mathbb{V}(U)$ is finite and has rank $r > 0$, then

$$-2 \log R(\theta) \xrightarrow{d} \chi^2_r$$

ELT can be applied to this estimating equations situation. However, it turns out there is better choice of auxiliary variable. This definition of auxiliary variable is not further pursued.

4.2 Alternative estimating equations and asymptotic efficiency

Consider the matrix estimation equaion

$$\mathbb{E}\{\Sigma - \Sigma(\theta)\} = 0$$

or equivalently

$$\mathbb{E}\{[ZZ' - \Sigma(\theta)]\} = 0 \tag{4.3}$$

Notice that equation (4.2) satisfies

$$\mathbb{E}[A(\theta)m(x,\theta)] = 0$$

while equation equation (4.3) satisfy

$$\mathbb{E}[m(x,\theta)] = 0$$

where $A(\theta) = \nabla_\theta[H * vech(\Sigma(\theta))]$, of dimension $u \times \frac{v(v+1)}{2}$.

Note that, the rank of $\nabla_\theta[H * vech\Sigma(\theta)]$ being u for all values of θ is only stating that the SEM is globally identified: H is of full rank, $\nabla_\theta[H * vech\Sigma(\theta)] = H'\nabla_\theta[vech\Sigma(\theta)], rank\{\nabla_\theta[vech\Sigma(\theta)]\} \geq u$ will lead to $rank\{\nabla_\theta[vech\Sigma(\theta)]\} = u$. So if all the required smoothness conditions by Theorem 2.3.1 are met, asymptotic variance of MELE $\tilde{\theta}$ is at least as small as that of estimators with the form (4.2) .

As a popular estimation method of SEM, multinormal MLE $\hat{\theta} = argmin_\theta F_{ML}$, and

$$F_{ML} = log|\Sigma(\theta)| + tr[\Sigma\Sigma^{-1}(\theta)] - log|\Sigma| - dim(x'y')'$$

Taking derivative of F_{ML} with respect to θ leads to

$$\nabla_\theta F_{ML} = -\nabla_\theta log|\Sigma^{-1}(\theta)| + \nabla_\theta tr\{\mathbb{E}(ZZ')\Sigma^{-1}(\theta)\}$$

$$= -\nabla_\theta[\Sigma^{-1}(\theta)]\Sigma(\theta) + \nabla_\theta[\Sigma^{-1}(\theta)]\mathbb{E}(ZZ')$$

$$= \nabla_\theta[\Sigma^{-1}(\theta)][\mathbb{E}(ZZ') - \Sigma(\theta)]$$

therefore the ML estimator is defined as

$$\nabla_\theta[\Sigma^{-1}(\theta)][\mathbb{E}(ZZ') - \Sigma(\theta)] = 0$$

where $\nabla_\theta \Sigma^{-1}(\theta)$ is also a cubix of dimension $u \times v \times v$. It is equivalent to

$$\nabla_\theta\{vec[(\Sigma^{-1}(\theta))']\}vec[\mathbb{E}(ZZ') - \Sigma(\theta)] = 0$$

and

$$\mathbb{E}\{\nabla_\theta\{vec[(\Sigma^{-1}(\theta))']\}vec[ZZ' - \Sigma(\theta)]\} = 0, \tag{4.4}$$

which is finally equivalent to

$$\mathbb{E}\{\nabla_\theta\{H * vech[(\Sigma^{-1}(\theta))']\}vech[ZZ' - \Sigma(\theta)]\} = 0, \tag{4.5}$$

where H is defined as before since $\Sigma^{-1}(\theta)$ also symmetric, and $\nabla_\theta\{H * vech[(\Sigma^{-1}(\theta))']\}$ is $u \times \frac{v(v+1)}{2}$. Take

$$A(\theta) = \nabla_\theta\{H * vech[(\Sigma^{-1}(\theta))']\}$$

of dimension of $u \times \frac{v(v+1)}{2}$ and note that, the rank of $\nabla_\theta\{H * vech[(\Sigma^{-1}(\theta))']\}$ being u for all values of θ is also a restatement that the SEM is globally identified. This is because H is of full rank,$\Sigma(\theta) = \Sigma$ is equivalent to $\Sigma^{-1}(\theta) = \Sigma^{-1}$ in terms of determining θ.

It can be observed again by Theorem 2.3.1, when $\mathbb{E}[m(X, \theta)] = 0$ is taken as

$$\mathbb{E}\{vech[ZZ' - \Sigma(\theta)]\} = 0$$

in SEM settings, the MELE $\tilde{\theta}$ via the auxiliary variable defined by (4.3) is also at least as asymptotically efficient as multinormal MLE. When we have the normal data, we lose nothing in asymptotic efficiency if we choose to use MELE by such defined auxiliary variable.

4.3 A test of model fit

In SEM analysis, it is interesting to test $H_0 : \Sigma = \Sigma(\theta)$. Among many available tests, is the popular likelihood ratio test with test statistic equal to $n-1$ times $F_{ML}(\hat{\theta})$, the MLE fitting function evaluated at MLE in parametric likelihood setting.

In the EL framework, we can also develop a test of overall model fit as follows. Consider

$$R(\theta) = \max\{\prod_{i=1}^{n} nw_i | \sum_{i=1}^{n} w_i vech[Z_i Z_i' - \Sigma(\theta)] = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\},$$

Looking at the constraint $\sum_{i=1}^{n} w_i vech[Z_i Z_i' - \Sigma(\theta)] = 0$, if $w_i = \frac{1}{n}$, i.e. we have uniform weights across all observation of the auxiliary variable, this constraint becomes $S = \Sigma(\theta)$. With two other remaining constraints automatically satisfied, this will lead to the supremum of $\prod_{i=1}^{n} nw_i$ that equals 1. All other values of $\prod_{i=1}^{n} nw_i$ due to a differently specified covariance structure $\Sigma(\theta)$ with the constraints satisfied is not bigger than 1, therefore the ratio $R(\theta_{MELE})$ can serve as an indicator of model fit. Fortunately, its distribution is known. It is by second asymptotic distribution result of Theorem 2.3.1 that

$$-2\log(R(\theta_{MELE})) \xrightarrow{d} \chi_q^2,$$

where $q = \frac{v(v+1)}{2} - u$ since here $\frac{v(v+1)}{2} = p + q$ and $u = p$. Note that, though this result parallels those of MLE approach very well, it is actually the ratio between two empirical likelihood ratios.

4.4 Computation and inference miscellaneous

To estimate SEM following a EL approach with the auxiliary variable suggested by last section, we need to compute MELE and empirical likelihood ratio to make inference. The computation of MELE and EL ratio can be computed via ready numerical methods in constrained programming. Following the definition of MELE in previous text, θ_{MELE} is defined as

$$\tilde{\theta} = \arg\max_{\theta}\{\prod_{i=1}^{n} nw_i | \sum_{i=1}^{n} w_i vech[Z_i Z_i' - \Sigma(\theta)] = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}$$

In fact, the only difference between the programming problems of computing MELE and computing EL ratios of the mean (with an appropriately defined auxiliary variable) is whether θ is known or not. There are multiple choices of computation. Solvers capable of dealing with large number of variables perform better since the variable in the problem increase with then number of observation. A interior point optimizing package "Knitro" is used in the later simulation section. Alternative optimizer such as NPSOL is much faster but often return slightly violated constraints conditions which make the value of χ^2 statistic inaccurate. With MELE and and EL ratio at hand, an overall model fit test can be performed directly with EL ratio value, and by (2.2) – (2.3), the confidence set can be explicitly written as

$$C_{\theta,\alpha} = \{\theta | 2\log\frac{R(\theta)}{R(\tilde{\theta})} \geq \chi^2_{\alpha}(u), \sum_{i=1}^{n} w_i vech[Z_i Z_i' - \Sigma(\theta)] = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}$$

For individual confidence interval of θ_k, we can solve the following constrained programming problem:

$$\max \mathit{or} \min\{\theta_k | 2\log\frac{R(\theta)}{R(\tilde{\theta})} \geq \chi^2_\alpha(u), \sum_{i=1}^n w_i vech[(Z_i Z_i' - \Sigma(\theta))] = 0, w_i \geq 0, \sum_{i=1}^n w_i = 1\}$$

to compute the maximum and minimum values allowed for specified significance level. In fact, define

$$R^*(\theta_k) = \{\prod_{i=1}^n n w_i | \sum_{i=1}^n w_i vech[(Z_i Z_i' - \Sigma(\theta))] = 0, w_i \geq 0, \sum_{i=1}^n w_i = 1\},$$

in which treating all components of θ other than θ_k as unknown, by a search around MELE $\tilde{\theta_k}$ gives confidence interval of θ_k. A contour of a pair of θ components can also be computed in a similar way.

For a test of $H_0 : \theta = \theta_1$, it is also straightforward to apply

$$-2\log(R(\theta)/R(\tilde{\theta})) \xrightarrow{d} \chi^2_u$$

For a composite test on the restriction of parameter θ such as $H_0 : g(\theta) = 0$ or $H_0 : g(\theta) \geq 0$, when we can not solve the restriction for a solution we may need to take $\theta = \arg\min_\theta R(\theta)$ subject to $g(\theta) = 0$ (or ≥ 0) to compute $R(\theta)$ and the test statistic $-2\log(R(\theta)/R(\tilde{\theta}))$ for safe accepting of H_0, and take $\theta = \arg\max_\theta R(\theta)$ subject to $g(\theta) = 0$ (or $\geq$ 0) (obviously, such θ can be directly expressed as $\arg\max_\theta\{\prod_{i=1}^n n w_i | \sum_{i=1}^n w_i vech[(Z_i Z_i' - \Sigma(\theta))] = 0, w_i \geq 0, \sum_{i=1}^n w_i = 1, g(\theta) = 0 (\geq 0)\}$) for a safe rejection of H_0, since $\arg\min_\theta R(\theta)$ subject to $g(\theta) = 0$ (or ≥ 0) produces largest test statistic value and vice versa.

Chapter 5

Estimation, inference and simulation examples

According to Hu, Bentler and Kano (1992) , the ADF estimation performs poorly with small sample size, we want to simulate non-elliptical data and make small to moderate sample model estimation to test the viability of the approach proposed above. Small sample size is meaningful because typical sample size in practice is seldom large enough.

5.1 Data generation

Mattson (1997) proposed a method of generating data with flexible skewness and kurtosis for SEMs. The method realized flexible univariate skewness and kurtosis of observable variables by controlling the univariate skewness and kurtosis of the random latent variables and errors. Compared to their purpose, we need a non-elliptical data on observable variables. It turns out that by assigning different distribution to individual latent variable and error, the computed observable variables can have marginal distributions with different characteristic generator and therefore the distribution of the observable variables will be non-elliptical.

Consider SEM written as:

$$\eta = (I - \beta)^{-1}(\Gamma\xi + \zeta)$$

$$X = \Lambda_x \xi + \delta$$

$$Y = \Lambda_y \eta + \epsilon$$

To generate a non-elliptical data, substitute the first equation into the third equation, rewrite the model as

$$Z = AV,$$

where

$$Z = (X'Y')',$$

$$V = (\xi'\zeta'\delta'\epsilon')',$$

and A is the correspondent coefficient matrix. Since each component of the observable variable Z is the sum of part of components of V, univariate distribution with different characteristic generator can be used to generate data for the components of V and based on V to produce a distribution of Z. This way, the marginal distribution of Z will have different characteristic generator. Letting each of the components have zero conditional expectation given others can ensure the un-correlation of these components in the model. It can be chosen that all but one component of Z are univariate normal. The non-normal one can be the sum of a normal and a heavy tailed student t.

Now consider an simplified example as follows:

$$\eta = \gamma\xi + \zeta$$

$$X = \Lambda_x \xi + \delta$$

$$Y = \Lambda_y \eta + \epsilon$$

where η, ξ, ζ are univariate, X is of dimension 3; Y is of dimension 4; Λ_x and Λ_y are of dimension 3 and 4 respectively; both Θ_δ and Θ_ϵ are also diagonal. Overall this model is identified by Two-Step rule since the latent model is identified by Null B rule and the measurement model is identified by the Three indicator rule following the approach by Bollen (1989) .

Let $\xi \sim N(0, \phi = 1)$; $\zeta \sim N(0, \psi = 1)$; $\delta_i \sim N(0, \sigma^2_{\delta i} = 1)$ for $i = 1, 2, 3$; $\epsilon_i \sim N(0, \sigma^2_{\epsilon i} = 1)$ for $i = 1, 2, 3$; $\epsilon_4 \sim t_4$, i. e. $\sigma^2_{\epsilon 4} = 2)$; $\lambda_x = (1, 1, 1)'$; $\lambda_y = (1, 1, 1, 1)'$ and $\gamma = 1$. There are 17 parameters in this model:

$$\phi, \psi, \gamma, \lambda_{x1}, \lambda_{x2}, \lambda_3, \lambda_{y1}, \lambda_{y2}, \lambda_{y3}, \lambda_{y4}, \sigma^2_{\delta 1}, \sigma^2_{\delta 2}, \sigma^2_{\delta 3}, \sigma^2_{\epsilon 1}, \sigma^2_{\epsilon 2}, \sigma^2_{\epsilon 3}, \sigma^2_{\epsilon 4}$$

Note that for this example in conventional SEM model estimation only 15 parameters are allowed to ensure identifiability. To make comparison of the estimation methods, we set $\lambda_{x1} = 1$, $\lambda_{y1} = 1$ following one of the conventional scaling option. Also note that, this mode is over identified and makes a test of model fit possible and necessary.

Following notations in above subsection, i. e.

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

.

This model can be written in the RAM (reticular action model) form as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ x_1 \\ x_2 \\ x_3 \\ \eta \\ \xi \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{y1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{y2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{y3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{y4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{x1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{x2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{x3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ x_1 \\ x_2 \\ x_3 \\ \eta \\ \xi \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \delta_1 \\ \delta_2 \\ \delta_3 \\ \zeta \\ \xi \end{pmatrix}$$

5.2 Estimation example and Preliminary simulation results

Below is the result of an estimation of this model with a moderate sample size 100. As a tentative examination of EL estimation performance, it is compared with ML estimation. With 100 simulated samples with sample size 100 following the data generation scheme in last example, the variances of the 4 among the 15 parameter estimates is smaller than that of MLE. The variance difference is mostly small, except that the variance of $\sigma^2)\epsilon 4$ of MELE estimate is 0.4202 while it is 1.1258 for MLE. Note that, $\sigma^2_{\epsilon 4}$ is exactly the variance of $\epsilon 4$ with a heavy tailed t distribution.

For information purpose, table 5.1 reported MELE/MLE estimation results: empirical means and standard deviation of parameter estimates and χ^2 statistic. The test statistic can not pass the Kolmogorov-Smirnov test of $\chi^2(13)$ for this simulation result.
Overall, this simulation results point many issued to be investigated later: the accuracy of the algorithm; Bartlett correction.

The S.D in the table is based on sample variance defined as $\mathbb{V}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Table 5.1: Means and Standard Deviations of 16 parameter estimates from 100 estimations.

Parameter/True value	*MELE/MLE Mean*	*MELE/MLE S.D*
ϕ	1.012	0.3314
1	1.032	0.3181
ψ	0. 9855	0.2892
1	1.012	0.2765
γ	1. 026	0.2572
1	1.014	0.2253
λ_{x2}	1. 035	0.2768
1	1.0084	0.2023
λ_{x3}	1. 0425	0. 2428
1	1.017	0.2498
λ_{y2}	1. 0117	0. 1239
1	1.006	0.1113
λ_{y3}	1.0181	0. 1363
1	1.016	0.1189
λ_{y4}	1. 0300	0. 1602
1	1.022	0.1564
$\sigma^2_{\delta 1}$	0. 9454	0. 2089
1	0.9768	0.2006
$\sigma^2_{\delta 2}$	0. 9537	0. 2183
1	0.9893	0.2221
$\sigma^2_{\delta 3}$	0. 9611	0. 2313
1	1.007	0.2039
$\sigma^2_{\epsilon 1}$	0. 9830	0. 2192
1	1.009	0.1989
$\sigma^2_{\epsilon 2}$	0. 9775	0. 2009
1	1.013	0.1864
$\sigma^2_{\epsilon 3}$	0. 9456	0. 2334
1	0.9832	0.2349
$\sigma^2_{\epsilon 4}$	1. 7014	0. 4202
2	2.026	1.1258
test statistic	Mean	S.D
$\chi^2 statistic$	17.87	8.011

Chapter 6

Concluding remarks

Empirical likelihood approach estimation of structural equation models is feasible based on the general estimating equation methods in the EL settings. Testing of model covariance structure is available. Point estimate is available in the form of Maximum Empirical Likelihood Estimator (MELE) . MELE derived from the auxiliary variable defined in equation (4.3) has at least as small asymptotic variance as Multinormality MLE and estimator by method of Unweighted Least Squares (ULS) . This provides a starting point for EL style alternative distribution-free SEM estimation technique. However, the asymptotic efficiency of the MELE such derived relative to GLS, Elliptical estimator and ADF is still unknown, either theoretically or by simulation. Looking into higher order asymptotic bias might be required. It will be of great interest to investigate these issues in the sense of searching for the "best" and "distribution-free" estimator. One way of establishing a EL approach estimator under elliptical distribution, might be treating MLE as

$$\mathbb{E}(\frac{\partial \log f(x_i;\theta)}{\partial \theta}) = 0$$

directly and include it into equation (2.1) to find MELE with ready asymptotic efficiency.

Bibliography

[1] Anderson, T. W., and Rubin, H., (1956), 'Statistical inference in factor analysis', Proceedings of the Third Berkeley Symposium for Mathematical Statistics Problems, 5: 111-150.

[2] Bentler, P. M., and Weeks D. G., (1980), 'Multivariate analysis with latent varibles', In P. R. Krishnaiah and L. Kanal, eds., Handbook of Statistics. Vol. 2. Amsterdam: North-Holland, pp. 747-771.

[3] Bentler, P. M., (1983), 'Simultaneous equation systems as moment structure models', Journal of Econometrics, 22: 13-42.

[4] Bock, R. D. and Bargmann, R. E., (1966), 'Analysis of covariance sturctures', Psychometrika, 31: 507-534.

[5] Bollen, K. A., (1989), 'Structural equations with latent variables', Wiley series in probability and mathematical statistics, Applied probability and statistics section, Johnson Wiley Sons, Inc.

[6] Bollen, K. A., (1996), 'An Alternative Two Stage Least Squares (2SLS) Estimator for Latent Variable Equations'. Psychometrika 61: 109-21.

[7] Browne, M. W., (1974), 'Generalized Least squares estimators in the analysis of covariances structures', South African Statistical Journal, 8: 1-24.

[8] Browne, M. W., (1982), 'Covariance structures', In D. M. Hawkins, ed. , Topics in Multivariate Analysis. Cambridge: Cambridge University Press, pp. 72-141.

[9] Browne, M. W. , (1984), 'Asymptotic distribution free methods in analysis of covariance structures', British Journal of Mathematical and Statistical Psychology, 37: 62-83.

[10] Chou, C. P., Bentler P.M., and Satorra A., (1991), 'Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: A Monte Carlo study', British Journal of Mathematical and Statistical Psychology, 44: 347-357.

[11] Dattorro J., (2005), Convex Optimization and Euclidean Distance Geometry, Meboo Publishing USA.

[12] Hayashi, F., (2000), 'Econometrics', Preseton University Press.

[13] Hansen, L., (1982), 'Large sample properties of generalized method of moments estimators', Econometrica, 50: 1029-1054.

[14] Hu, L. -T., Bentler, P. M. and Kano, Y., (1992), 'Can test statistics in covariance structural analysis be trusted?', Psychological Bulletin, 112: 352-362.

[15] Jöreskog, K. G., (1969), 'A general approcah to confirmatory maximum likelihood factor analysis', Psychometrika, 34: 183-202.

[16] Jöreskog, K. G. and Goldberg A. S., (1972), 'Factor Analysis by generalized least squares', Psychometrika, 37: 243-260.

[17] Jöreskog, K.J., (1973), 'A general method for estimating a linear structural equation system', In A. S. Goldberger and O. D. Duncan, eds. , Structural Equation Models in the Social Sciences. New York: Academic Press, pp. 85-112.

[18] Kline, R. B., (1998), 'Software Review: Software Programs for Structural Equation Modeling: Amos, EQS, and LISREL, ' Journal of Psychoeducational Assessment, Vol. 16, No. 4, 343-364

[19] Lawley, D. N., (1940), 'The estimation of factor loadings by the method of maximum likelihood', Proceedings of the Royal Society of Edinburgh, 60: 64-82.

[20] Mattson, S. (1997) , 'How to generate non-normal data for simulation of Structural Equation Models', Multivariate Behavioral Research, 32 (4) , 355-373.

[21] McArdle, J. J., and MCDonal, R. P., (1984), 'Some algebric properties of the reticular action model for moment structures', British Journal of Mthematical and Statistical Psychology, 37: 234-251.

[22] Muthén, B., (1984), 'A general structural eaution model with dichotomous, ordered categorical and continuous latent variable indicators', Psychometrika, 49: 115-132.

[23] Muthén. B and Kaplan, D., (1985), 'A comparison of some methodologicals for the factor analysis of non-normal likert variables', British journal of mathematical and statistical psychology, 38: 171-189.

[24] Muthén, B. (1987), LISCOMP:Analysis of linear structural equations with a comprehensive measurement model', Mooresville, IN: Scientific Software, Inc.

[25] Muthén, B. and Kaplan, D., (1992), 'A comparison of some methodologicals for the factor analysis of non-normal likert variables: A note on the size of the model', British journal of mathematical and statistical psychology, 45: 19-30.

[26] Newey, W and D. McFaden., (1994), 'Large sample Estimation and hypothesis testing', Chapter 36 in R. Engle and D. McFaden (eds.) , Handbook of econometrics, Volume IV, New York: North-Holland.

[27] Owen. A., (2001), 'Empicial Likelihood', Monographs on statistics and applied probability, Chapman Hall/CRC.

[28] Owen. A,. (1991), 'Empirical likelihood for linear models', The Annals of Statistics, 19: 1725-1747.

[29] Qin, J and Lawless, J. (1994), 'Empirical likelihood and general estimating equations', The Annals of Statistics 22, 300-325.

[30] Wright, S., (1918), 'On the nature of size factors', Genetics, 3: 367-374.