Sensitivity analysis of structural equation model with non-normal observed variables

Fahmida Homayra *
Department of Mathematics and Natural Sciences
BRAC University
66, Mohakhali, Dhaka, 1212
Bangladesh

Syed Shahadat Hossain
Azmeri Khan
Institute of Statistical Research and Training
University of Dhaka
Bangladesh

Abstract

Structural equation modeling (SEM) relies on normal theory methods – such as maximum likelihood method (ML) and generalized least squares (GLS) when estimating model and testing model goodness of fit. This limits its applicability since most of the observed variables in social sciences are almost never normally distributed. In this paper, we relax this restriction and conduct a simulation experiment with the simplest observed variable SEM model to see the extent and nature of effects of nonnormality on SEM. We incorporate that the distribution of error of the model belongs to g-*and*-k family of distributions.

Keywords and phrases : *Structural equation modeling, implied covariance matrix, non-normality, generalized least square method.*

1. Introduction

Structural equation modeling (SEM) is a multivariate data analysis technique. It is a combination of multiple regression and factor analysis. As SEM has a number of strength it has become an increasingly popular data analytic option for few decades.

* *E-mail*: fhomayra@bracuniversity.ac.bd

Journal of Interdisciplinary Mathematics
Vol. 12 (2009), No. 4, pp. 499–513

One well-known feature is the ability to specify latent variable models that provide separate estimates of relations among latent constructs and their manifest indicators and of the relations among constructs. Another commonly acknowledged strength is the availability of measures of global fit that can provide a summary evaluation of even complex models that involve a large number of linear equations.

SEM presents a particularly challenging topic for review because it is mathematically complex and constantly evolving. The fundamental hypothesis for the structural equation procedures is that the covariance matrix of the observed variables is a function of a set of parameters. If the model were correct and if we know the parameters, the population covariance matrix would be exactly reproduced.

The equation that formalized this fundamental hypothesis is

$$\Sigma = \Sigma(\theta). \tag{1}$$

In (1) Σ is the population covariance matrix of observed variables, θ is a vector that contains the model parameters, and $\Sigma(\theta)$ is the covariance matrix written as a function of θ.

SEM has been successfully applied to observational and experimental studies in social sciences. The term structural stands for the assumption that the parameters are not just descriptive measures of association but rather that they reveal an invariant causal relation. The links between variables are summarized in the structural parameters.

In typical situations for fitting models SEM uses first and second-order moment structures that are derived from structural equations that prescribe causal connection between variables. Higher order moments have almost never been utilized. This is partly due to the fact that SEM was originally developed based on the normal assumption. But it should be noted that observed variables in social sciences are almost never normally distributed (Micceri, 1989). Higher-order moment structures could be informative for SEM.

A remarkable criticism of structural equation model is the belief that the estimators of structural equation model have no value if the observed variables do not have a multinormal distribution. The tests of statistical significance are dependent on more restrictive conditions. So robustness of structural equation model has to be reviewed. The word robust is loaded with many connotations; in the narrow sense robustness signifies insensitivity to small deviations from the assumptions and as an alternative robust procedure has been proposed. For the statisticians such a theory of robustness, i.e., working with deviation from idealized assumption is valuable and for the applied statisticians it provides both new dimensions of thinking and the reevaluation of existing procedures and development of new procedures.

Most widely used methods of estimation for SEM are maximum likelihood (ML) method, generalized least squares (GLS) and unweighted least square (ULS) method (Bollen, 1989). When the sample size is fairly large, the covariance matrix is analyzed, the observed variables have no excess kurtosis, and $H_0 : \Sigma = \Sigma(\theta)$ holds exactly, $(N-1)F_{ML}$ or $(N-1)F_{GLS}$ provides chi-square estimators to test $H_0 : \Sigma = \Sigma(\theta)$. Brown (1982, 1984) provides the justification for using $(N-1)F_{GLS}$ or $(N-1)F_{ML}$ as chi-square estimator based on generalized least square principles. Brown (1982) shows that $(N-1)F_{GLS}$ and $(N-1)F_{ML}$ are chi-square estimators when observed variables have no excess kurtosis. The best-known distribution with no kurtosis is the multinormal. In practice, non-normal observed variables do occur. Browne (1984) suggests that leptokurtic ('more peaked' than normal) distributions result in too many rejections of H_0, and he speculates that platykurtic distributions will lead to too low of estimates of chisquare. So for effective and valid statistical analysis it is better to avoid relying on such assumptions.

Another condition that restricts the application of structural equation model is that for $(N-1)F_{GLS}$ or $(N-1)F_{ML}$ to approximate a chi-square variate the sample size should be sufficiently large. Here again, much work remains to be done to determine the minimum necessary sample size. Most of the researchers suggested that the chi-square estimator $(N-1)F_{ML}$ is not accurate for sample size smaller than 50 and recommends 100 or more cases. Another guideline for sample size is that the greater the number of free parameters (parameters to be estimated) in a model, the greater the N should be.

In this paper we attempt to study the nature and extent of the effect of violating the assumptions of SEM concerning the sample size and the distribution of the observed variables as well.

2. Structural equation modeling

2.1 *Linear Structural Equation Model* (*LISREL*)

Let X be a q-dimensional observed vector, Y be a p-dimensional observed vector, and ζ an error vector of appropriate dimension. Let η

and ζ are the vectors consisting endogenous and exogenous variables, respectively, and, ordinarily, are not directly observed. They are sometimes called latent variables. The well-known structural equation model (Bollen, 1989) is written as

$$\eta = \beta\eta + \Gamma\xi + \zeta\,, \tag{2}$$

$$Y = \Lambda_y \eta + \varepsilon\,, \tag{3}$$

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

where B and Γ are matrices of path coefficients, B has zeros on the diagonal, $E(\zeta) = 0$, ξ and ζ are independent, ζ, ε and δ are mutually uncorrelated. $I - B$ is nonsingular.

The equation (2) is called the *structural equation*, and shows the relationship between the latent variables. Equations (3) are called the *measurement equations*, and link the latent variables to the observed variables.

2.2 *SEM with observed variables*

Regression based models are common in the social sciences. They may consist of a single equation oriented toward explaining one endogenous variable or multi-equation model with a number of endogenous variables and reciprocal relationships.

The following equation is a general representation of structural equations with observed variables (classical econometric model):

$$y = By + \Gamma x + \zeta\,. \tag{4}$$

Since the disturbances ζ represent random errors in the relationship between the y s and x s these models are sometimes referred as errors in the equation models. The standard assumption is that the errors are uncorrelated with x.

The implicit measurement model for structural equation with observed variables is

$$y = \eta\,, \tag{5}$$

$$x = \xi\,.$$

2.3 *Implied Covariance matrix*

The basic hypothesis of the general SEM model is

$$\Sigma = \Sigma(\theta). \tag{6}$$

We derive the specialization of (6) for observed variables models in this section. The relation of Σ to $\Sigma(\theta)$ is basic to an understanding of identification, estimation, and assessment of model fit.

The implied covariance matrix of y and x is

$$\begin{aligned} \Sigma(\theta) &= \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \\ &= \begin{bmatrix} (I-B)^{-1}(\Gamma \in \Gamma' + \Psi)(I-B)^{-1'} & (I-B)^{-1}\Gamma\Phi \\ \Phi\Gamma'(I-B)^{-1} & \Phi \end{bmatrix}. \end{aligned} \tag{7}$$

2.4 *Tests of overall model fit*

The estimated coefficients and the strength of association require close examination. The overall fit measures help to asses whether the covariance structure hypothesis is valid, and if not they help to measure the departure of Σ from $\Sigma(\theta)$. Being population parameters, Σ and $\Sigma(\theta)$ are unavailable, so researchers examine their sample counterparts S and $\Sigma(\hat{\theta})$. The S is usual sample covariance matrix, and $\Sigma(\hat{\theta})$ is the implied covariance matrix evaluated at the estimate of θ which minimizes either F_{ML}, F_{GLS}, or F_{ULS}, where F_{ML}, F_{GLS}, F_{ULS}, are the fitting functions for the estimation methods maximum likelihood (ML), generalized least square (GLS) and unweighted least square (ULS), respectively (Bollen, 1989).

In our research we use GLS estimator for fitting the model. A general form of the GLS fitting function is

$$F_{GLS} = \frac{1}{2} tr(\{[S - \Sigma(\theta)]W^{-1}\}^2), \tag{8}$$

where W^{-1} is a weight matrix for the residual matrix. The weight matrix W^{-1} is either a random matrix that converges in probability to positive definite matrix as $N \to \infty$, or it is a positive definite matrix of constants. Usually S^{-1} is used as the weight matrix in GLS method.

Under the conditions stated in Section 1, $(N-1)F_{GLS}$ provide chi-square estimators to test the covariance structure hypothesis. Since H_0 is equivalent to test the hypothesis that $\Sigma - \Sigma(\theta) = 0$, the chi-square test is simultaneous test that all residuals in $\Sigma - \Sigma(\theta)$ are zero. Brown (1982, 1984) provides the justification for using $(N-1)F_{GLS}$ as chi-square estimator based on generalized least square principles. When the sample size is fairly large, the covariance matrix is analyzed, the observed variables have no kurtosis, and H_0 is true, then $(N-1)F_{GLS}$ evaluated at $\hat{\theta}$ is approximately distributed as a chi-square variate with degrees of freedom $\frac{1}{2}(p+q)(p+q+1)-t$, where p is the number of indicators of latent endogenous variable $\eta's, q$ is the number of indicators of latent exogenous variable $\xi's$, and t is the number of free parameters in θ.

In this paper we study the performance of generalized least square method under different levels of non-normality. As this is the most widely used estimation method for SEM and the fitting function F_{GLS} is known to be less sensitive than the other fitting functions, and also it requires less simulation time we chose this method for our study.

3. The *g-and-k* distribution

The g-*and*-k distribution families have shown considerable ability to fit to data, approximate standard distributions, and easy to use in simulation studies (Khan and Rayner, 2003).

In assessment of the robustness to non-normality in ranking and selection procedure, Haynes *et al.* utilize the g-and-k distribution for the above reasons as well as for flexibility in the coverage of distributions with both long and short tails.

Rayner and MacGilivray examine the effect of non-normality on the distribution of (numerical) maximum likelihood estimators. The g-*and*-k distributions (MacGilivray and Cannon) are defined in terms of their quantile functions as

$$Q_x(u|A,B,g,k) = A + Bz_u\left(1 + c\frac{1-e^{-gu_u}}{1+e^{gz_u}}\right)(1+z_u^2)^k, \tag{9}$$

here, A and $B > 0$ are location parameters, g measures skewness in the distribution, and $k > -\frac{1}{2}$ measures the kurtosis of the distributions, $z_u = \Phi^{-1}(u)$ is the uth standard normal quantile and c is a constant chosen to help produce proper distributions.

The sign of the skewness parameter indicates the direction of skewness; $g < 0$ indicates skewness to the left, and $g > 0$ indicates skew-ness to the right. Increasing/decreasing the unsigned value of g increases/decreases the skewness in the indicated direction. When $g = 0$ the distribution is symmetric.

The kurtosis parameter k, for the g-*and*-k distribution, behaves similarly. Increasing k increases the level of kurtosis and vice versa. The value of $k = 0$ corresponds to no extra kurtosis added to the standard normal base distribution. However, this distribution can represent less kurtosis than the normal distribution, as $k > -\frac{1}{2}$ can assume negative values. If curves with even more negative kurtosis are required than base distribution with less than standardized normal can be used.

For these distributions c is the value of the *overall asymmetry*. For an arbitrary distribution, theoretically the overall asymmetry can be as large as one, so it would appear that for $c > 1$, data or distributions could occur with skewness that cannot be matched by these distributions. However for $g \neq 0$, the larger the value chosen for c, the more restrictions on k are required to produce a completely proper distribution. Real data seldom produce overall asymmetry values greater than 0.8 (MacGilivray and Cannon). We have used $c = 0.8$ through this paper. To examine the extent of the effect of different level of non-normality on the test of conventional structural equation model we consider that the random error belong to the g*-and-*k distribution families.

4. Simple regression model as a special case of observed variable model

Let us consider the regression model

$$y = \gamma x + \zeta\,. \tag{10}$$

To examine the sensitivity of chi-square test discussed in Section 5 we consider the model (10) because it is the simplest structural equation model with observed variables. The implied covariance matrix is then

$$\Sigma(\theta) = \begin{bmatrix} \gamma^2\phi & \gamma\phi \\ \phi\gamma & \phi \end{bmatrix}. \tag{11}$$

To reduce the amount of calculation and to avoid mathematical complexity we impose two restrictions:

(1) $v(x) = \phi = 1$, and
(2) $v(\zeta) = \psi = 1$.

The equation (11) becomes

$$\Sigma(\theta) = \begin{bmatrix} 1+\gamma^2 & \gamma \\ \gamma & 1 \end{bmatrix}. \tag{12}$$

According to t rule and null B rule the model (10) is identified (Bollen, 1989). To quantify the extent and nature of deviations of the powers and

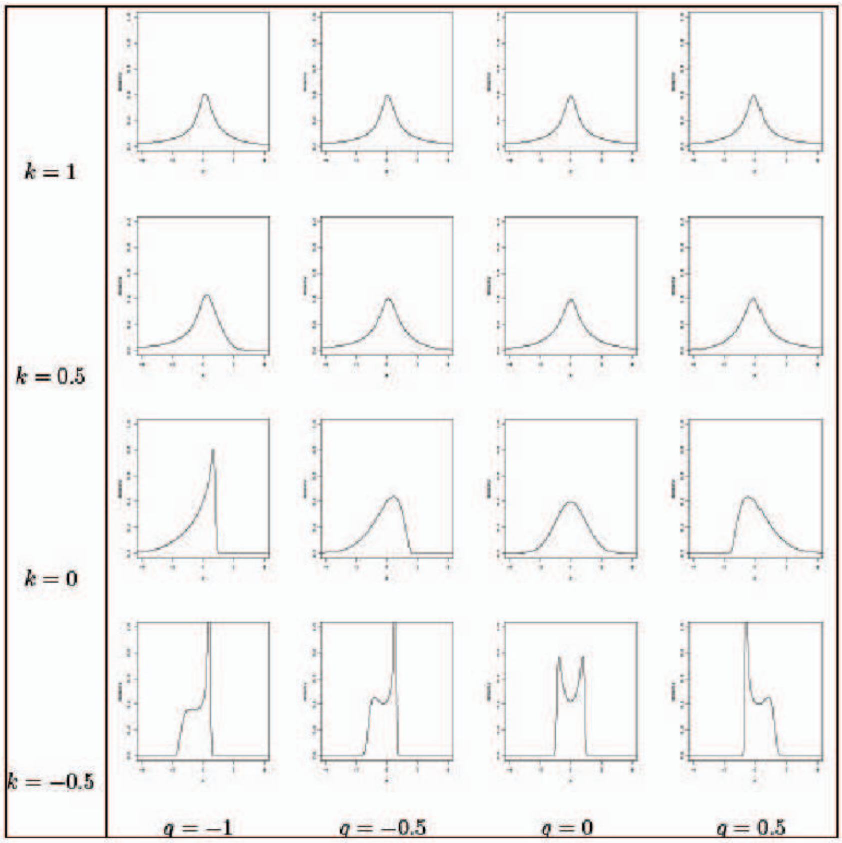

Figure 1

g-and-k **distribution densities corresponding to various combinations of** (g, k)**. Here the location and scale parameters are** $A = 0$ **and** $B = 1$

size of chi-square test under non-normality model (10) is considered and GLS fitting function (8) is used to estimate the parameter and to calculate the chi-square statistic.

5. Simulation study

In Section 1 the assumptions of chi-square tests are given. But what will be the consequences of violation of these assumptions? To answer this question we conduct a simulation experiment with model (10).

We generate the observed exogenous variable x and random error variable ζ from g-*and*-k distribution.

To see the performance of chi-square estimator in non-normal situation we simulate the size and power of the test with different levels of skewness and kurtosis and compare them with the situation when observed variables x and y are from multinormal distribution. The use of g-*and*-k distribution facilitates us to see the effect of specific level of skewness through the parameter g and specific level of kurtosis through the parameter k on the chi-square estimator.

Since we restrict that $v(x) = \phi = 1$ and $= v(\zeta) = \psi = 1$, there is only one free parameter in the model (10), which is the regression coefficient γ.

The size and power of chi-square test are examined by generating data from sample of size $n = 5, 10, 100, 350$ and 460 with different (g, k) combinations.

We consider that the value of the regression parameter $\gamma = 6$ to construct the implied covariance matrix under null hypothesis. Then the implied covariance matrix becomes

$$\Sigma(\theta) = \begin{bmatrix} 37 & 6 \\ 6 & 1 \end{bmatrix}.$$

5.1 *Size of chi-square test*

First, we consider the effect of non-normality on size of the chi-square test. For simulating the size of the chi-square test we generate data from g-*and*-k distribution with location and scale parameter $A = 0$ and $B = 1$, respectively. Using the statistical software R we generate data for sample size $n = 5, 50, 100$, and 460, test the $H_0 : \Sigma = \Sigma(\theta)$ against the alternative $H_1 : \Sigma \neq \Sigma(\theta)$.

To determine the size of these tests we generate data under the null hypothesis and repeat the test 5000 times and divide the number of times the hypothesis rejected by 5000; tests are carried out using a nominal size of $\alpha = 0.05$

In Table 1 simulation results are given to see the effect of different level of nonnormality on size of the chi-square test. From the simulation results it can be said that for small samples the size of the test is very high even if the observed variables follows normal distribution and the size equals to the nominal size $\alpha = 0.05$ when the sample size is 450. And for nonnormal observed variable presence of kurtosis affects the size badly than the skewness. For small samples the size of the test is large

Table 1

Size of chi-square test for different combinations of (g, k) with varying sample sizes

n	g	k				
		0.0	0.5	0.8	−0.3	−0.5
5	0.0	0.9048	0.7408	0.5982	0.9616	0.9806
	0.5	0.9162	0.7472	0.6286	0.9682	0.9796
	1.0	0.9184	0.7660	0.6542	0.9590	0.9796
	−0.5	0.9630	0.7468	0.6112	0.9608	0.9818
50	0.0	0.5444	0.0736	0.0036	0.7720	0.8982
	0.5	0.5374	0.0892	0.0076	0.7590	0.8912
	1.0	0.5114	0.0892	0.0110	0.7420	0.8412
	−0.5	0.5310	0.0874	0.0080	0.7640	0.8890
100	0.0	0.3798	0.1114	0.0000	0.6738	0.8432
	0.5	0.3734	0.0160	0.0000	0.6572	0.8346
	1.0	0.3360	0.0110	0.0006	0.6352	0.8030
	−0.5	0.3360	0.0122	0.0004	0.6514	0.8326
460	0.0	0.0500	0.0000	0.0000	0.3486	0.6702
	0.5	0.0480	0.0000	0.0000	0.3408	0.6366
	1.0	0.0370	0.0000	0.0000	0.2870	0.5942
	−0.5	0.0476	0.0000	0.0000	0.3308	0.6310

in presence of excessive kurtosis. For large samples size of the test is zero in presence of positive kurtosis and large in presence of negative kurtosis. And the size approaches to unity in presence of higher negative kurtosis.

So it can be said that chi-square test is not size robust for small samples and for data having excessive kurtosis than those of normal data.

5.2 *Power of chi-square test*

Power of a test is an important and considerable matter. Powerful test gives better conclusion, so it is desired to use powerful test.

To simulate power we test the $H_0 : \Sigma = \Sigma(\theta)$ where $\Sigma(\theta)$ is the implied covariance matrix given earlier, against the alternative H_1, $\Sigma \neq \Sigma(\theta)$. We generate data using $\gamma \in (5.5, 5.55, 5.60, \ldots, 6.5)$ and repeat the test procedure 5000 times. We First determine the number of rejections of the test out of the 5000 times for each value in the mentioned set and dividing the total number of rejections by 5000, with level of significance $\alpha = 0.05$.

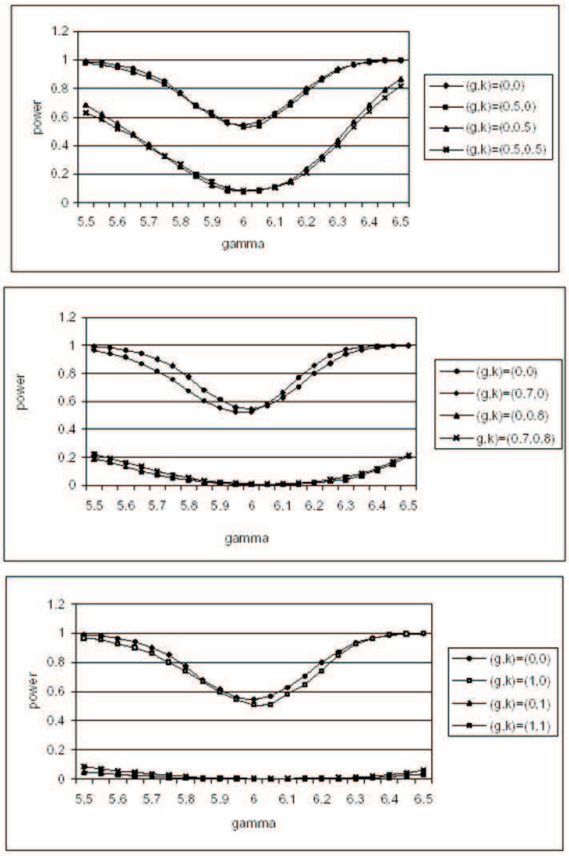

Figure 2

Power curves of generalized least square test under different combinations of (g, k) with sample size 50

The simulation is conducted by using R. We generate power for sample size $n = 50, 100, 350$ with various combinations of the shape parameters g and k to see the effect of skewness and kurtosis on the power of chi-square test.

6. Discussion of results

From the Figure 2 to 5 we can say that chi-square test is neither optimally robust nor size robust. Because both the power and the size of the test are badly affected by sample size and excess kurtosis. The simulation results can be summarized in the following ways:

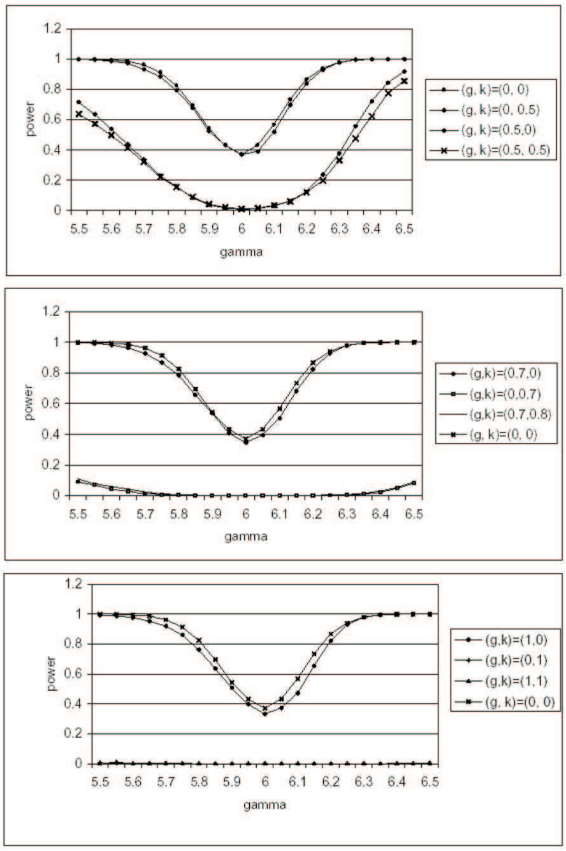

Figure 3

Power curves of generalized least square test for different combinations of (g, k) with sample size 100

(1) Comparing the Figures 2, 3 and 4, we can say that as the sample size increases the loss of power increases when there are positive kurtosis in data.

(2) From the Figures 2 to 4 it is apparent that as the value of kurtosis parameter k increases in positive direction chi-square test gives less power than that with normal data.

(3) From Figure 5 we can say that as the sample size increases the size of the test increases when there are negative kurtosis in data and size increases with the increase of sample size.

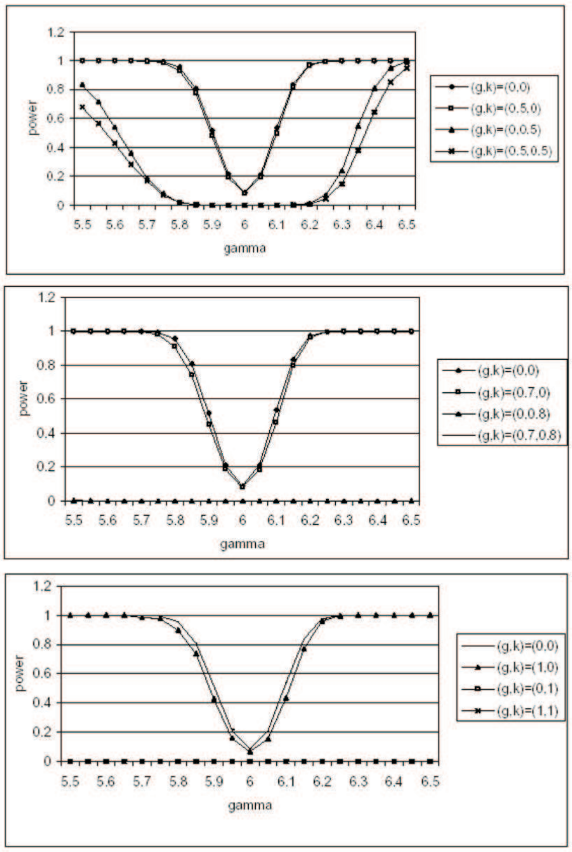

Figure 4

Power curves of generalized least square test with different combinations of (g, k) with sample size 350

7. Conclusion

In this paper, we have used higher-order moment structures to examine the effect of non-normality to conventional SEM. We took simple regression model as an example of the SEM with observed variables. We used a simple bivariate model since it is simple, comprehensive and illustrates a problem that many researches encounter commonly. We conducted a small simulation experiment.

We have quantified the effect of specific level of skewness and kurtosis on the generalized least square method generating data from the g-*and*-k quantile distribution family. From the simulation results we can

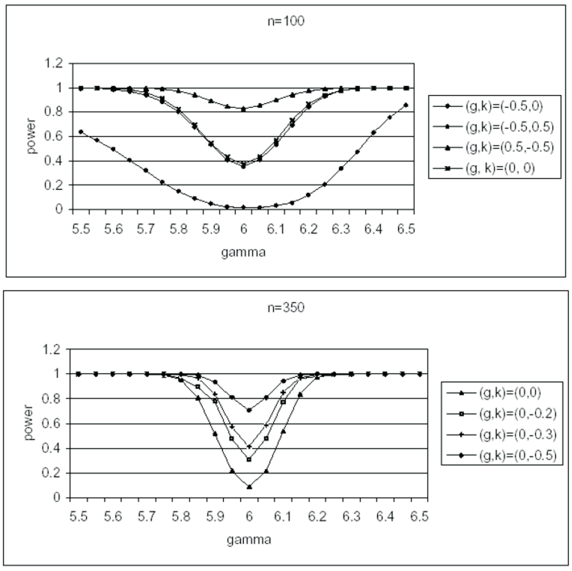

Figure 5

Power curves of generalized least square test under different combinations of (g, k) with various sample sizes

say that GLS method gives poor result in presence of kurtosis in data. And even for normal data when sample size is not sufficiently large it gives poor results. When there is negative kurtosis in data the size of the test is affected badly and when there is positive kurtosis in data the power of the test is affected. And the skewness does not affect the test that much. And for normal data the sample size should be sufficiently large (> 100), but as the sample size increases the test gives worse results for non-normal data comparative to the case of normal data.

References

[1] K. A. Bollen (1989), *Structural Equations with Latent Variables*, Wiley, New York.

[2] M. W. Brown (1982), Covariance Structures, in *Topics in Applied Multivariate Analysis*, D. M. Hawkins (editor), Cambridge University Press, Cambridge, pp. 72–141.

[3] M. W. Brown (1984), Asymptotic distribution-free methods for the analysis of covariance structures, *British Journal of Mathematical and Statistical Psychology*, Vol. 9, pp. 665–672.

[4] M. A. Haynes, H. L. MacGillivray and K. L. Mengersen (1997), Robustness of ranking and selection rules using generalized *g-and-k* distributions, *J. Statist. Plann. Inference*, Vol. 65, pp. 45–66.

[5] A. Khan and G. D. Rayner (2003), Robustness to non-normality of common tests for the many-sample location problem, *Journal of Applied Mathematics and Decision Sciences*, Vol. 7, pp. 187–206.

[6] H. L. MacGilivray and G. D. Rayner (1986), Skewness and asymmetry: measures and orderings, *The Annals of Statistics*, Vol. 14 (3), pp. 994–1011.

[7] T. Micceri (1989), The unicorn, the normal curve, and other improbable creatures, *Psychological Bulletin*, Vol. 105 (1), pp. 156–166.

Received August, 2008