British Journal of Mathematical and Statistical Psychology (1992) **45**, 19–30 *Printed in Great Britain*

A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model

Bengt Muthen†

Graduate School of Education, University of California Los Angeles, Los Angeles, CA 90024, USA

David Kaplan

Department of Educational Studies, University of Delaware

This paper expands on a recent study by Muthen & Kaplan (1985) by examining the impact of non-normal Likert variables on testing and estimation in factor analysis for models of various size. Normal theory GLS and the recently developed ADF estimator are compared for six cases of non-normality, two sample sizes, and four models of increasing size in a Monte Carlo framework with a large number of replications. Results show that GLS and ADF chi-square tests are increasingly sensitive to non-normality when the size of the model increases. No parameter estimate bias was observed for GLS and only slight parameter bias was found for ADF. A downward bias in estimated standard errors was found for GLS which remains constant across model size. For ADF, a downward bias in estimated standard errors was also found which became increasingly worse with the size of the model.

1. Introduction

A recent paper by the authors (Muthen & Kaplan, 1985) considered the problem of factor analysis with non-normal Likert variables. Estimators which assume continuous (interval scale) variables such as normal theory maximum likelihood (ML) and normal theory generalized least squares (GLS) were compared to the asymptotic distribution free (ADF) estimator of Browne (1982, 1984), which does not assume normality. These estimators were applied to both skewed ordered categorical variables and to skewed categorical variables with underlying continuous normal variables. For the latter case ML, GLS, and ADF were compared to the categorical variable methodology (CVM) estimator of Muthen (1984), which explicitly takes into account the categorical nature of the variables and assumes underlying continuous normal response variables. This note only considers the former case where our interest is in the structure for the observed variables.

†Requests for reprints.

Considering a four-variable, single-factor model, our results showed that for univariate skewness values in the range of -1.0 to 1.0 the chi-square goodness-of-fit tests obtained from the normal theory estimators were quite robust to the various degrees of non-normality studied. For univariate skewness greater than 2.0 in absolute value however, we found that ML and GLS chi-square tests became too large. Here the ADF estimator was found to reduce the values of the test statistics to an acceptable level. Our results also showed no parameter estimate bias for any of the estimators studied. A slight downward bias of estimated standard errors was found for the normal theory estimators when variables were moderately to severely skewed. No substantial downward bias in estimated standard errors was found for ADF. On the whole then, the normal theory estimators performed rather well for moderately non-normal data. These results are compared to related research (e.g. Boomsma, 1983; Browne, 1984; Tanaka, 1984) in Section 5 of Muthen & Kaplan (1985).

Since the publication of our previous paper the robustness to non-normality of ML, ADF, as well as new elliptical estimators has been studied in Harlow (1985). In particular, Harlow examined a six-variable, two-factor (8 d.f.) model for sample sizes of 200 and 400. She chose skewness values ranging from -2.0 to $+2.0$ and kurtosis values ranging from -1.0 to $+8.0$. Harlow's results agree with our previous findings in the sense that the normal theory ML estimator performed rather well for moderately non-normal data.

Results of our study and those of other researchers (Boomsma, 1983; Olsson, 1979) led us to believe that the results for normal theory estimators were 'largely independent of the number of variables' (Muthen & Kaplan, 1985, p. 187). Browne (1984) however, conjectured that the number of degrees of freedom may be an important factor in robustness studies of this kind. In the present paper we address this conjecture formally via a Monte Carlo study that expands on our previous paper in the following ways: First, we use a considerably larger number of replications than in our previous study. Second, we consider a negative kurtosis case in addition to the original five cases of our previous study. Our interest in the negative kurtosis case lies in the fact that it may give rise to an underestimation of chi-square (see Browne, 1984). Third, we examine two levels of sample size. Fourth, and most importantly, we consider four models of increasing size.

2. Design of the study

The design of this study follows closely that of our previous paper. In particular, we generate continuous random normal variables $\mathbf{y}^*$ from a known factor analysis structure $\boldsymbol{\Sigma}(\mathbf{y}^*) = \boldsymbol{\Lambda\Psi\Lambda}' + \boldsymbol{\Theta}$. The $\mathbf{y}^*$ variables are then categorized in such a way as to yield six different cases of non-normal ordered categorical variables $\mathbf{y}$ following a factor analysis model $\boldsymbol{\Sigma}(\mathbf{y})$ with the same number of factors but with different parameter values. A case is defined by all variables having the same univariate distributions. A summary of the relevant statistics related to the five original cases is given in Table 1 of our previous paper, and the distributions for all cases are graphically displayed in Fig. 1 here.

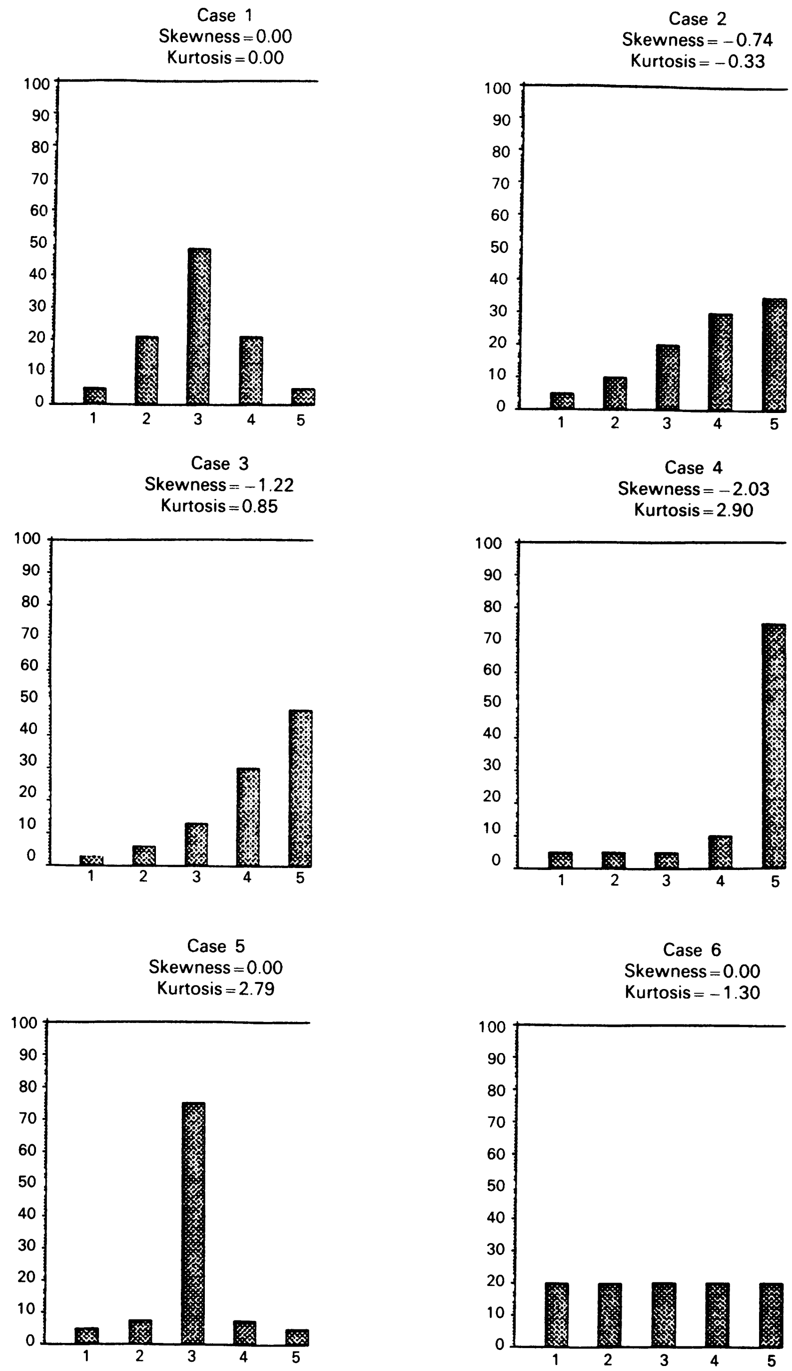

Figure 1. Histograms for six cases of non-normality including univariate skewness and kurtosis.

In addition to the five original cases and the additional negative kurtosis case, four levels of model size were chosen. The first is a six-variable, two-factor (8 d.f., 13 parameters) model. The second is a nine-variable, three-factor (24 d.f., 21 parameters) model. The third is a twelve-variable, three-factor (51 d.f., 27 parameters) model, and the fourth is a fifteen-variable, three-factor (87 d.f., 33 parameters) model. We consider robustness to non-normality for two levels of sample size, $N = 500$ and $N = 1000$, the latter previously studied in Muthen & Kaplan (1985). The number of replications within each cell of the six (cases) by four (levels) by two (sample sizes) is 1000. The number of replications is considerably larger than that which has been previously used in related studies, and was chosen particularly to be able to adequately assess standard error behaviour and the rejection frequency of chi-square. From our previous study we found no difference between normal theory ML and GLS, therefore in this study we will only utilize the GLS estimator. In addition to the normal theory estimators we will also study the behaviour of the ADF estimator.

The data generation and estimation were carried out by the LISCOMP program (Muthen, 1987). Multivariate normal data were generated and categorized according to the six cases. The multivariate normal data were created by generating uniformly distributed random numbers as in Kennedy & Gentle (1980, p. 147), normal numbers as in Box & Mueller (1958), and multivariate normal variates by Cholesky decomposition as in Kennedy and Gentle (1980, pp. 294–301). The ADF estimator uses the calculations suggested by Browne (1982, 1984), without correction for bias.

3. Results

The results for normal theory GLS and ADF chi-square are shown in Tables 1 and 2, respectively. The 95 per cent prediction interval around the expected value of 50 for the rejection frequency is 36 to 64. It should be noted here that owing to the computationally heavy features of ADF we only report ADF results for Models 1, 2 and 3.

3.1 Performance of small model

As stated earlier, our previous study examined a four-variable, single-factor (2 d.f.) model. If we compare that model to Model 1 for $N = 1000$ in this study, we find similar results—namely that normal theory GLS performs rather well for data with Case 2 non-normality. Unlike our previous study however, we see that the normal theory GLS estimator obtains too large chi-square values for data already with Case 3-type non-normality. The ADF chi-square tests are seen to perform quite well throughout for Model 1. In addition, we replicate our previous findings in that no parameter estimate bias is found for normal theory GLS or ADF.

Results for GLS and ADF standard error bias are given in Tables 3 and 4, respectively. It should be noted that throughout we find the estimated standard errors always very close to the population standard errors. The bias is calculated as per cent over- or underestimation of the mean estimated standard error relative to the empirical variation in the estimates over the 1000 replications. With respect to

Table 1. Normal theory GLS chi-square for all cases[a]

Case	Model 1 6-var/2f (8 d.f.)		Model 2 9-var/3f (24 d.f.)		Model 3 12var/3f (51 d.f.)		Model 4 15-var/3f (87 d.f.)	
	$N=500$	$N=1000$	$N=500$	$N=1000$	$N=500$	$N=1000$	$N=500$	$N=1000$
y^*								
Mean	7.797	7.945	24.046	24.034	50.837	51.062	85.857	85.955
Variance	15.880	15.840	46.961	48.843	96.590	103.072	175.717	171.327
Reject freq.	46	41	56	49	49	58	46	44
1								
Mean	7.530	7.853	23.755	24.045	50.866	50.821	86.420	86.442
Variance	14.684	15.461	45.324	46.330	94.265	103.592	165.123	178.089
Reject freq.	44	39	39	52	49	53	40	53
2								
Mean	8.372	8.410	26.085	25.805	55.090	55.050	93.039	93.488
Variance	17.978	16.687	51.921	51.160	121.506	126.473	211.295	215.767
Reject freq.	61	61	88	76	137	138	132	139
3								
Mean	9.060	9.343	27.959	27.731	59.221	59.597	101.387	102.224
Variance	21.354	21.818	58.870	63.441	135.440	120.690	245.643	250.684
Reject freq.	93	96	138	133	229	244	294	300
4								
Mean	11.015	11.433	33.795	33.679	72.679	73.540	126.640	126.417
Variance	32.593	31.293	92.300	93.370	201.186	206.468	354.569	364.81
Reject freq.	190	214	364	347	618	651	777	814
5								
Mean	8.443	8.490	25.298	25.820	54.351	54.398	93.324	94.646
Variance	16.983	17.699	54.155	60.393	112.976	111.831	198.444	205.263
Reject freq.	59	65	79	88	111	108	120	149
6								
Mean	7.788	7.959	24.157	24.538	51.818	51.856	88.596	88.886
Variance	14.387	15.516	45.684	53.189	100.782	106.069	168.790	186.268
Reject freq.	44	50	50	67	75	76	58	85

[a] Reject freq. denotes the frequency of samples with chi-squares greater than the 5 per cent critical value (expected number is 50).

Model 1, we find as in our previous study that normal theory GLS estimated standard errors are consistently underestimated for Case 3 and Case 4-type non-normality. For ADF we observe a downward bias of estimated standard errors—a result not found in our previous study. For both GLS and ADF the bias appears worse for $N=500$ than for $N=1000$.

We conclude that for small models (8 d.f. or less) the normal theory GLS estimator performs well for Cases 1, 2, 5 and 6. ADF is preferred for Case 3 or Case 4 non-normality.

3.2 Model size effects

We now turn to the influence of model size on GLS and ADF estimation for non-normal data. As a quality check for the number of replications, we first analyse

Table 2. ADF chi-square for all cases

Case	Model 1 6-var/2f (8 d.f.)		Model 2 9-var/3f (24 d.f.)		Model 3 12-var/3f (51 d.f.)	
	$N=500$	$N=1000$	$N=500$	$N=1000$	$N=500$	$N=1000$
y*						
Mean	8.092	8.085	26.108	25.109	59.149	54.953
Variance	18.171	16.478	61.411	56.446	148.499	126.465
Reject freq.	58	47	101	75	246	119
1						
Mean	7.770	7.974	25.638	25.032	59.080	54.715
Variance	16.103	15.983	58.301	52.474	152.066	127.530
Reject freq.	54	45	84	73	218	138
2						
Mean	8.012	7.916	26.417	24.956	59.051	54.704
Variance	17.289	15.099	59.168	50.524	159.585	132.220
Reject freq.	53	46	113	60	241	143
3						
Mean	8.010	8.078	26.091	24.602	58.490	54.491
Variance	19.516	16.449	56.478	49.687	144.870	128.680
Reject freq.	58	51	107	66	221	119
4						
Mean	8.066	8.146	26.082	24.751	58.583	54.893
Variance	17.362	16.018	53.290	49.055	115.606	113.776
Reject freq.	51	49	91	63	209	126
5						
Mean	8.235	8.108	25.715	25.314	58.556	54.590
Variance	15.455	16.254	56.138	56.516	133.673	118.124
Reject freq.	50	59	91	77	217	123
6						
Mean	7.904	7.944	25.685	25.135	59.238	54.763
Variance	15.478	15.788	56.644	58.363	156.140	126.698
Reject freq.	52	52	80	85	228	138

continuous multivariate normal **y***s before introducing any categorization. The **y*** results for GLS are shown in Table 1. We find that GLS on **y*** performs quite well for both sample sizes and all four model sizes. Inspection of the categorized Cases 1 through 6 in Table 1 reveals that GLS is quite sensitive to both non-normality and size of model. In fact, GLS seems to behave well only for a limited set of cases—either for normally distributed variables (continuous or categorical) or for Models 1 and 2 with non-normality exhibited in Cases 2, 5 or 6. Differences due to sample size appear minor.

Our findings regarding parameter estimates will only be reported in the text. On the whole, we find very little GLS parameter estimate bias regardless of the degree of non-normality or size of model considered here. This finding is in agreement with the GLS results of our previous study.

Standard error results for GLS are reported in Table 3. When comparing estimated standard errors to empirical standard deviations the results show a clear downward bias in estimated standard errors that becomes increasingly worse for Case 3 or Case

Table 3. Normal theory GLS standard error bias percentage[a]

Case	Par	Model 1 6-var/2f (8 d.f.)		Model 2 9-var/3f (24 d.f.)		Model 3 12-var/3f (51 d.f.)		Model 4 15-var/3f (87 d.f.)	
		$N=500$	$N=1000$	$N=500$	$N=1000$	$N=500$	$N=1000$	$N=500$	$N=1000$
1	λ	−2	0	−2	0	−2	0	0	0
	θ	−2	0	0	0	−3	0	−3	0
	ψ	2	3	−2	−3	−6	−6	0	−3
2	λ	−2	0	−3	0	0	−3	−2	0
	θ	−6	−7	−9	−8	−7	−8	−9	−7
	ψ	−2	−3	−7	−7	−8	−5	−6	−6
3	λ	−15	−12	−13	−12	−13	−13	−13	−13
	θ	−18	−16	−17	−16	−17	−18	−21	−19
	ψ	−11	−9	−11	−11	−14	−12	−11	−11
4	λ	−28	−29	−29	−28	−29	−29	−30	−28
	θ	−32	−30	−32	−32	−34	−33	−35	−35
	ψ	−23	−22	−23	−23	−24	−23	−23	−24
5	λ	−18	−17	−18	−18	−22	−19	−22	−19
	θ	−20	−17	−20	−18	−24	−23	−23	−24
	ψ	−6	−4	−7	−6	−7	−9	−6	−5
6	λ	7	8	6	9	10	11	10	15
	θ	1	3	0	1	2	3	2	3
	ψ	0	3	−3	−2	−6	−5	−2	−3

[a]In parentheses is given percentage under- or overestimation of the empirical standard deviations of the estimates.

4 non-normality. The results also show that the bias is roughly constant across sample size and model size.

Results for ADF chi-square tests are displayed in Table 2. We find that ADF chi-square is quite sensitive to model size even for the multivariate normal **y*** case. For larger models, the chi-square behaviour is considerably worse for $N=500$ than for $N=1000$. There is a tendency, however, for ADF to give somewhat lower chi-square values than GLS for data with Case 3 or Case 4 non-normality. Nevertheless, the results do not support the use of ADF for improvement of chi-square for Model 3 unless very large samples are available. It seems likely that the behaviour of the ADF chi-square tests would deteriorate even further for larger models than those considered here.

Regarding ADF parameter estimates, the results show a tendency for estimates to be slightly biased for large models with Case 4 non-normality. For smaller models and less severe non-normality, very little ADF parameter estimate bias was found.

The results for ADF standard errors are displayed in Table 4. We find ADF estimated standard errors to be biased downward, becoming worse with degree of non-normality. Unlike GLS, ADF standard error bias also becomes worse as the size of the model increases. The size of the bias is reduced when going from $N=500$ to $N=1000$.

Since the models considered here are scale-free, we may choose to analyse the correlation matrix instead of the covariance matrix. The asymptotic covariance

Table 4. ADF standard error bias percentage

		Model 1 6-var/2f (8 d.f.)		Model 2 9-var/3f (24 d.f.)		Model 3 12-var/3f (51 d.f.)	
Case	Par	$N=500$	$N=1000$	$N=500$	$N=1000$	$N=500$	$N=1000$
1	λ	−5	−1	−9	−6	−14	−7
	θ	−4	−2	−9	−3	−15	−7
	ψ	−3	1	−9	−8	−17	−11
2	λ	−5	−3	−7	−3	−16	−8
	θ	−5	−3	−10	−4	−15	−8
	ψ	−11	2	−10	−7	−16	−8
3	λ	−7	−5	−10	−5	−20	−10
	θ	−5	−2	−10	−4	−16	−9
	ψ	−5	0	−10	−7	−18	−9
4	λ	−9	−5	−16	−6	−24	−13
	θ	−6	−4	−14	−6	−18	−8
	ψ	−9	−7	−17	−10	−23	−14
5	λ	−5	−3	−13	−6	−19	−9
	θ	−3	0	−10	−4	−15	−7
	ψ	−7	−4	−12	−8	−19	−11
6	λ	−4	−2	−9	−4	−15	−7
	θ	−5	−1	−9	−5	−16	−8
	ψ	−2	0	−8	−5	−16	−10

matrix for correlations under both normality and non-normality is given in, for example, Steiger & Hakstian (1982). For computations in the normal case, see also Jennrich (1970). Regarding computing the ADF type asymptotic covariance matrix for correlations, see Mooijaart (1985). However, when studying a selected set of cases, no improvement was found for either chi-square or standard errors when using a correlation matrix instead of a covariance matrix. In fact, the use of the correlation matrix gave slightly worse results.

3.3 Chi-square distributions

In line with suggestions from a reviewer, we also decided to augment the Monte Carlo information presented above with Q–Q plots of the chi-square test values for a limited set of scenarios. For each of these scenarios, percentile values for the 1000 observed chi-square values were calculated and plotted against the corresponding theoretical chi-square values. If the observed chi-square values are well behaved, the plot would approximate a 45 degree line through the origin. If not, a transformation of the observed chi-squares might be suggested. We chose the scenarios Case 2, Model 2; Case 4, Model 2; Case 2, Model 4; and Case 4, Model 4. Only NTGLS was studied, using a sample size of 500. The choice of NTGLS rather than ADF was made to see if this simpler estimator could be given a chi-square correction. The plots of the four scenarios are given below in Figures 2–5.

We note the plots do suggest linearity throughout most of the range. To more

Model 3 12-var/3f (51 d.f.)	
$N=500$	$N=1000$
−14	−7
−15	−7
−17	−11
−16	−8
−15	−8
−16	−8
−20	−10
−16	−9
−18	−9
−24	−13
−18	−8
−23	−14
−19	−9
−15	−7
−19	−11
−15	−7
−16	−8
−16	−10

ɔrmality is given in, for
he normal case, see also
ptotic covariance matrix
ng a selected set of cases,
ırd errors when using a
he use of the correlation

to augment the Monte
-square test values for a
tile values for the 1000
ıinst the corresponding
ɜs are well behaved, the
If not, a transformation
the scenarios Case 2,
.el 4. Only NTGLS was
rather than ADF was
re correction. The plots

ɔf the range. To more

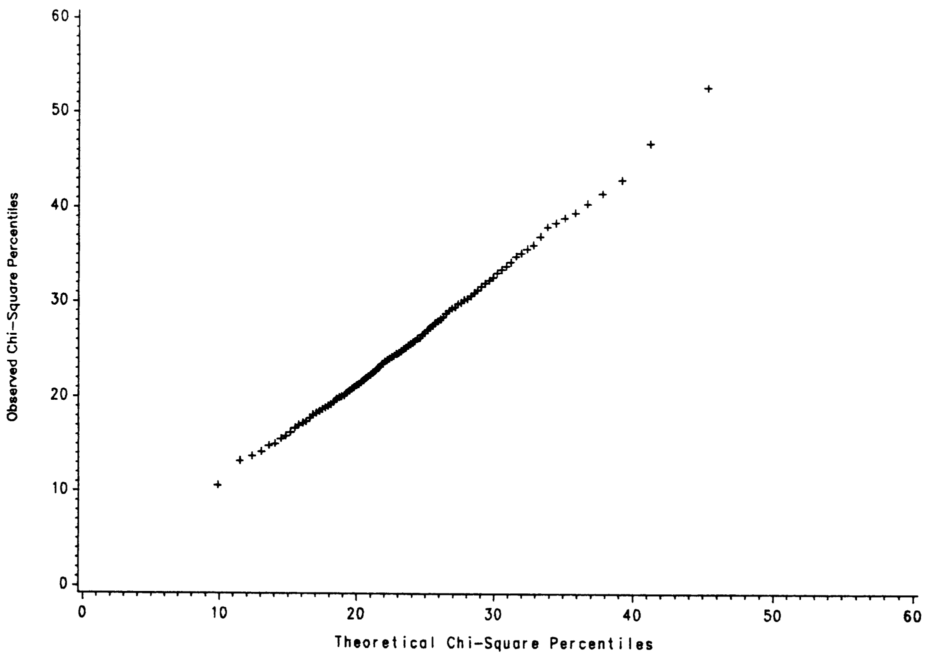

Figure 2. Q–Q plot for normal theory GLS, Case 2, Model 2.

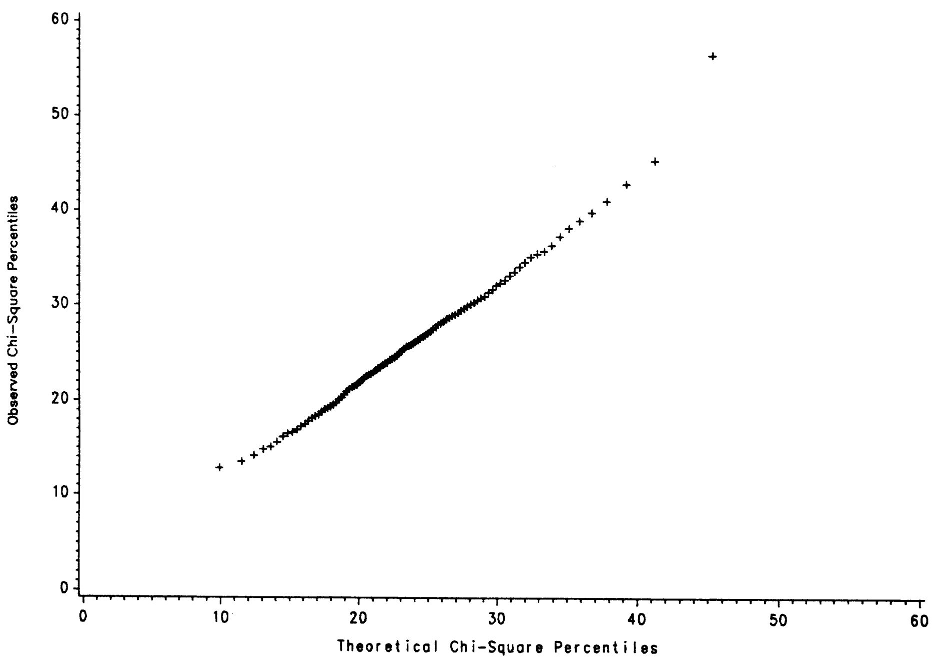

Figure 3. Q–Q plot for normal theory GLS, Case 4, Model 2.

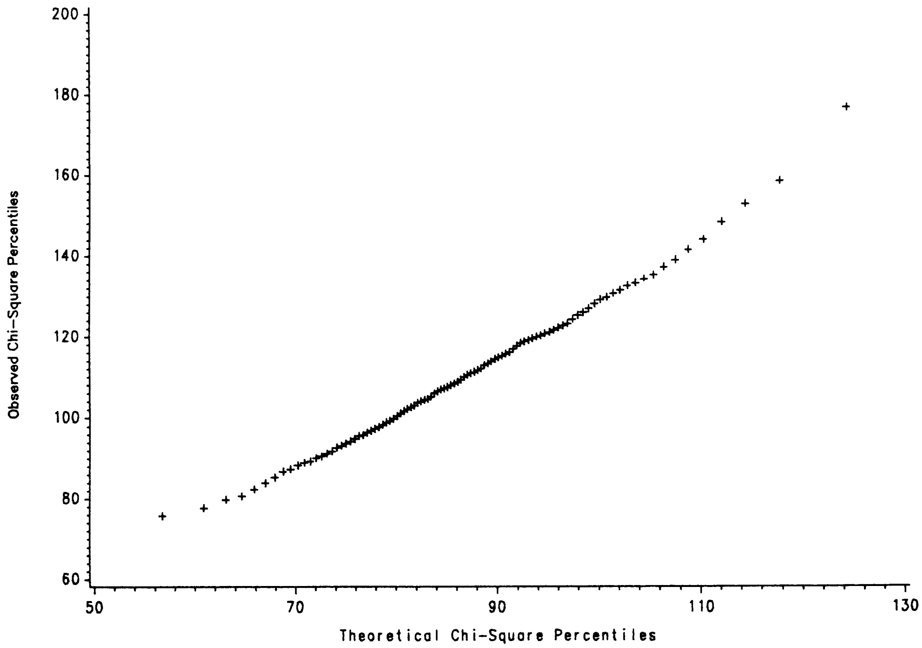

Figure 4. Q–Q plot for normal theory GLS, Case 2, Model 4.

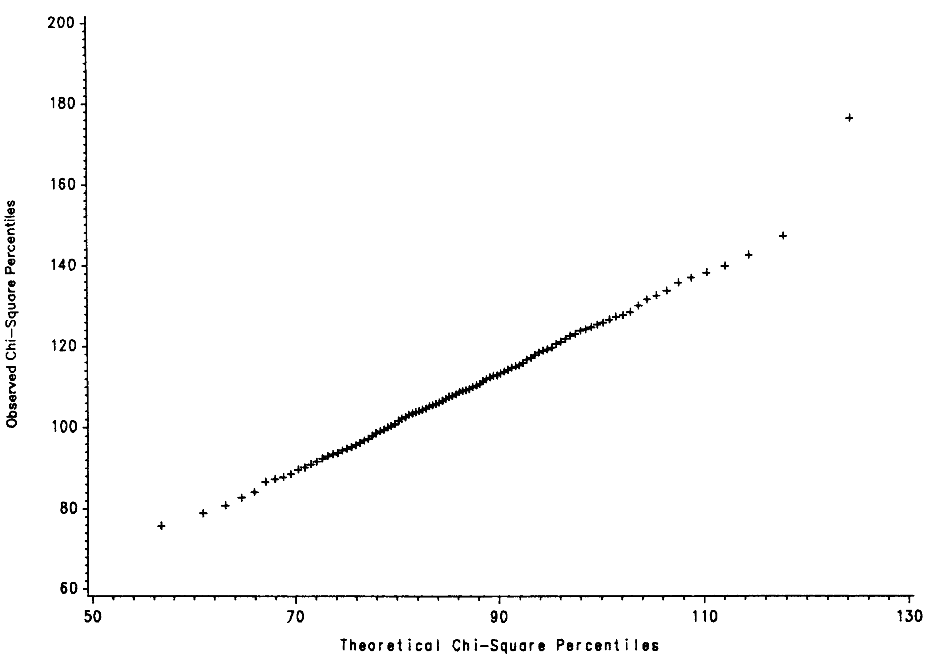

Figure 5. Q–Q plot for normal theory GLS, Case 4, Model 4.

easily comp.
was perforn
were used i
the estimate
1.22. It app
scaling corr
change as
Further rese

In conclusi
model with
models of r
not as robu
appear to
unless the
conclusions
studied no
only affecte
ADF stanc
warrant fur

The researcl
Science Fou
and Tammy

Boomsma, A
small sa
Groninge
Box, G. E.
Mathemat
Browne, M.
Multivari
Browne, M.
structures
Harlow, L.
covarianc
Universit
Harlow, L.
ML, ADI
meeting c
Jennrich, R
Journal o

easily compare the lines, a linear regression of the observed on theoretical chi-squares was performed. Only the middle values corresponding to the middle 90 percentiles were used in these regressions to obtain more stable results. In order of the figures, the estimated intercepts and slopes were: −1.21, 1.13; 0.82, 1.05; −8.61, 1.37; 3.90, 1.22. It appears that although the plots show a large degree of linearity, a simple scaling correction of the observed chi-squares is elusive. The parameters of the lines change as a function of the degree of non-normality and the size of the model. Further research is needed in this area.

4. Conclusions

In conclusion, this study has added to previous work by considering the size of the model within a considerably larger simulation framework. The results show that, for models of realistic size, the chi-square and standard errors of normal theory GLS are not as robust to non-normality as previously believed. In addition, ADF does not appear to work well as a means of compensating for the effects of non-normality unless the model is small and N is large. For continuous variables, similar conclusions have recently been reached by Harlow, Chou & Bentler (1986) who studied normal theory ML and ADF. Our findings that GLS standard errors are only affected by degree of non-normality, while GLS chi-square, ADF chi-square and ADF standard errors are affected by both non-normality and model size may warrant further theoretical work.

Acknowledgements

The research of the first author was supported by grant no. SES-8312583 from the National Science Foundation. We wish to thank Shinn-Tzong Wu for valuable computing assistance and Tammy Tam for assisting with the figures.

References

Boomsma, A. (1983). On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality. Unpublished PhD dissertation, University of Groningen, Groningen, The Netherlands.

Box, G. E. P. & Mueller, M. E. (1958). A note on generation of normal deviates. *Annals of Mathematical Statistics*, **28**, 610–611.

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in Applied Multivariate Analysis*. London: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62–83.

Harlow, L. L. (1985). Behavior of some elliptical theory estimators with nonnormal data in a covariance structure framework: A Monte Carlo study. Unpublished PhD dissertation, University of California, Los Angeles.

Harlow, L. L., Chou, C.-P. & Bentler, P. M. (1986). Performance of chi-square statistic with ML, ADF, and elliptical estimators for covariance structures. Paper presented at the annual meeting of the Psychometric Society, Toronto, Canada.

Jennrich, R. I. (1970). An asymptotic χ^2 test for the equality of two correlation matrices. *Journal of the American Statistical Association*, **65**, 904–912.

Kennedy, W. J. & Gentle, J. E. (1980). *Statistical Computing*. New York: Dekker.

Mooijaart, A. (1985). The weight matrix in asymptotic distribution-free methods. *British Journal of Mathematical and Statistical Psychology*, **38**, 190–196.

Muthen, B. (1984). A general structural equation model with dichotomous ordered categorical, and continuous latent variables. *Psychometrika*, **49**, 115–132.

Muthen, B. (1987). LISCOMP. Analysis of linear structural equations with a comprehensive measurement model. Mooresville, IN: Scientific Software.

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

Steiger, J. H. & Hakstian, A. R. (1982). The asymptotic distribution of elements of a correlation matrix: Theory and application. *British Journal of Mathematical and Statistical Psychology*, **35**, 208–215.

Tanaka, J. S. (1984). Some results on the estimation of covariance structure models. Unpublished Ph.D. dissertation, University of California, Los Angeles.

British Journal of Mathe

Thurstc

Department of l

Thurstc
because
the effe
present
partial
As a r
imposir
utilities
probab
procedu
approa

Thurstone's (19
highly influenti
the concept of
of objects can
This idea prove
the recent dev
(Böckenholt &
& DeSarbo, 19
based on a Th
were found to
Wise, 1978; K
describing disc
models for ana
Mullen, 1986;
MacKay, 1983)
This paper
comparison da
the method of
there are man

†Requests for re