QÜESTIIÓ, Vol 18, 1 pp. 75-84, 1994

THE USE OF THIRD-ORDER MOMENTS IN STRUCTURAL MODELS

ERIK MEIJER and AB MOOIJAART

Leiden University

Structural models are usually estimated using only second order moments (covariances or correlations). When variables are not multivariate normally distributed, however, methods that also fit higher order moments, such as skewnesses, are theoretically asymptotically preferable. This article reports results from a Monte Carlo simulation study in which estimators that fit both second-order moments and third-order moments are compared with estimators that fit only second-order moments.

Key words: Structural models, robustness, elliptical distribution, asymptotically distribution-free, generalized least squares.

1. INTRODUCTION

Structural models, as most other statistical and data-analytical models and techniques, are usually estimated by fitting only covariances or correlations (e.g. Jöreskog, 1967; Jöreskog & Goldberger, 1972; Bentler, 1983; Browne, 1982,

*Erik Meijer & Ab Mooijaart. Department of Psychometrics and Research Methodology. Leiden University. P.O. Box 9555. NL-2300 RB Leiden. The Netherlands.

–Article presentat al Seventh International Conference on Multivariate Analysis, setembre 1992.

–Acceptat el maig de 1993.

1984). The reason for this stems from the classical normal theory: if the variables are multivariate normally distributed then all information from the sample —assuming that the means are of no interest— is contained in its covariance matrix.

In practice, however, variables are often not normally distributed and more information can be extracted from the higher-order moments. This implies, for example, that otherwise unidentified models may be identified by fitting higher order moments (see, e.g., Van Montfort, Mooijaart, & De Leeuw, 1987; Mooijaart & Bentler, 1986), and it implies that estimators that use higher order moments are generally asymptotically more efficient (see Kano, Bentler, & Mooijaart, 1993). In particular, Mooijaart (1985) showed that the exploratory factor analysis model is identified if the factors are independent, whereas the usual factor analysis estimators all have the well-known rotation problem.

On the other hand, it is well-known that higher order moments are estimated with considerably less precision than covariances, and it may be questionable, particularly in small samples, whether the use of higher order moments leads to better estimates because of the extra information that is used, or to worse estimates because sample statisticis are used that may not be good estimates of their population counterparts.

Besides identification and precision, correctness of inference is a third aspect. If one tests whether a model must be rejected by using a test statistic that is based on strong assumptions regarding the distribution of the variables then the test statistic may reject the (structural) model too often or not often enough, depending on the specific (wrong) assumptions that are made and the specific (correct) distribution of the variables. This applies also to confidence intervals that are used to test whether a parameter may be equal to a certain value (typically 0 or 1). This problem of *robustness* of statistical inference is discussed by several authors; see, for example, Browne & Shapiro (1988), Mooijaart & Bentler (1991), and the references therein. Their results are quite positive in this respect: test statistics of normal theory based estimators are asymptotically correct within a broad range of models, and confidence intervals concerning some classes of parameters in structural models are asymptotically correct, but other confidence intervals are not.

In this article, the focus of attention is on the precision of estimators and correctness of statistical inference, rather than on the identification problem. Estimators that fit third-order moments as well as second-order moments are compared with estimators that fit only second-order moments in a Monte-Carlo simulation study.

2. METHOD

To compare the properties of estimators that use higher-order moments with the properties of estimators that use only second-order moments, a Monte Carlo simulation study was conducted. Samples were generated from a 4-variable, 1-factor, factor analysis model, with loadings 0.8, 0.8, 0.6, and 0.6, respectively. Sample sizes varied from 300 to 10,000. The nonnormalities in the data were generated by drawing the factor from a lognormal distribution and then correcting it to have mean zero, variance one, and prespecified skewness, which was varied between 0.1 and 4.0. The errors were generated from a normal distribution with mean zero and variances 0.36, 0.36, 0.64, and 0.64, respectively. Consequently, the variables all had expectations zero and variances one, and the skewnesses varied —between variables and samples— from 0.02 to 2.0. With each combination of sample size and factor skewness, 300 independent samples were drawn.

The parameters were estimated by *Generalized Least Squares* (GLS) estimators, and by *Linearized Generalized Least Squares* (LGLS) estimators. LGLS estimators are asymptotically equivalent to GLS estimators, but are computationally more efficient. See Bentler & Dijkstra (1985) for a discussion of GLS and LGLS. Maximum Likelihood estimators were not used, because these estimators apply only when normality is assumed, and in this study the focus is on methods that fit higher-order moments, which implies that nonnormality is assumed.

The estimators that fit higher-order moments that were used in this study were introduced by Mooijaart (1985). These estimators fit second-order and third-order moments, which implies that they are based on the assumption that the variables are nonsymmetrically (skewly) distributed, which is not a very restrictive assumption. The discrepancy function that is minimized is based on the *Asymptotically Distribution-free* (ADF) theory of Browne (1984). These estimators will therefore be called ADF3. The ADF3 estimators were compared with estimators that fit only second-order moments, namely Normal theory (NORM), Elliptical theory (ELL), and Asymptotically Distribution-free (ADF2) estimators. The NORM estimators were described by Jöreskog & Goldberger (1972), and these estimators are asymptotically equivalent to normal theory maximum likelihood estimators. The ELL estimators were described by Bentler (1983), and use the assumption that the variables are elliptically distributed, which is a relaxation of the normality assumption. The ADF2 estimators were described by Browne (1982, 1984), and use no specific assumptions about the distributions of the variables. Their asymptotic properties are independent of the distribution of the variables. The NORM, ELL, and ADF2 estimators are implemented in the program EQS (Bentler, 1989).

The discrepancy function (loss function) that had to be minimized has the following form:

$$F(\theta) = (s - \sigma(\theta))' W (s - \sigma(\theta)),$$

where s contains sample statistics, $\sigma(\theta)$ is the (asymptotic) expectation of s according to the model, θ is the parameter vector that has to be estimated, and W is the *weight matrix*. For the second-order estimators, s consists of the nonduplicated elements of the sample covariance matrix, and for the ADF3 estimators, s consists of all nonduplicated second and third-order moments. The vector $\sigma(\theta)$ reflects the structural model that has to be estimated. It is well-known (e.g., Bentler & Dijkstra, 1985) that W is optimal if $\text{plim}(W) = \boldsymbol{\Gamma}^{-1}$, where $\boldsymbol{\Gamma}$ is the asymptotic covariance matrix of $\sqrt{N}(s-\sigma)$, and N is the sample size. Therefore, W was chosen in such a way that W is a consistent estimator of $\boldsymbol{\Gamma}^{-1}$. But the elements of $\boldsymbol{\Gamma}$, and consequently the elements of W, depend on the distributions of the variables. Therefore, W reflects the distributional assumptions. In LGLS, $\sigma(\theta)$ is linearized around an initial consistent estimate θ_1 of θ.

The structural model $\sigma(\theta)$ can be tested by the chi-square test statistic $NF(\hat{\theta})$, where $F(\hat{\theta})$ is the minimum of $F(\theta)$, and $\hat{\theta}$ is the argument for which the minimum is attained. This test statistic follows an asymptotic chi-square distribution if σ and W are correctly specified.

The asymptotic covariance matrix of $\sqrt{N}(\hat{\theta}-\theta)$ can be consistently estimated by $\left(\Delta(\hat{\theta})'W\Delta(\hat{\theta})\right)^{-1}$, where

$$\Delta(\theta) = \frac{\partial \sigma}{\partial \theta'}(\theta),$$

and confidence intervals for the elements of θ can be easily obtained from the (diagonal) elements hereof and the (consistent) estimates $\hat{\theta}$ of θ.

The model that had to be estimated was equal to the model with which the data were generated, that is, a one-factor model with a skewly distributed factor and symmetrically distributed errors. The variance of the factor was fixed to one and the factor loadings were free parameters.

Based on the results of robustness studies (e.g., Browne, 1987), it was expected that the NORM, ADF2, and ADF3 test statistics would perform well, but that the ELL test statistic would not perform well. Furthermore, it was expected that the confidence intervals of the ADF3 and ADF2 methods would be approximately correct and that the confidence intervals of the error standard deviations of the NORM estimators would be approximately correct, but that the confidence intervals of the loadings of the NORM estimators would be incorrect, and that the confidence intervals of the ELL estimators would be incorrect. If the model had been specified with one factor loading fixed to one and unspecified factor variance, which would have rendered a model that is statistically indistinguishable from the one used, the confidence intervals of the loadings would have been consistently estimated by the NORM estimators. On the other hand, this would have implied that the factor variance is a free parameter that had to be estimated, and the confidence interval of the factor variance would be inconsistently estimated, so the problem of inconsistency is only shifted from one set of parameters to another. We have chosen the first parametrization, because we think that this is the most exploratory one. In confirmatory factor analysis, the second parametrization is often used.

From the results of Kano, Bentler, & Mooijaart (1993), it was expected that the ADF3 estimators would be more efficient than the second-order estimators, and from the results of robustness studies (e.g., Mooijaart & Bentler, 1991), it was expected that the ADF2, NORM, and ELL estimators would be equally efficient.

3. RESULTS

Test statistic

Some results on the properties of the test statistics are given in Tables 1a and 1b. The ADF3 test statistic performed well with sample sizes 1000 and larger, and its performance was independent of factor skewness. The ADF2 and NORM test statistics accepted the model in all cases about 95% of the replications, independent of sample size or skewness. The ELL test statistic nearly always accepted the model if the factor skewness was 2.0 or larger, where the skewness of the variables was 0.4 or larger. The GLS and LGLS test statistics generally performed equally well, except in the ADF3 case: at relatively small sample sizes the GLS test statistic performed better.

Parameter estimates

From Table 2 it can be seen that the ADF3 estimators were slightly biased downward. The GLS estimators were more biased than the LGLS estimators. This bias diminished at larger sample sizes, as expected because of the consistency of the estimators. The second order estimators were practically unbiased.

Table 1.

Percentage of times the model was accepted ($\alpha = .05$).

(a) Factor skewness = 2.0.

sample	ADF3		ADF2		NORM		ELL	
size	GLS	LGLS	GLS	LGLS	GLS	LGLS	GLS	LGLS
300	84	76	91	91	93	93	100	100
500	89	87	95	95	95	95	99	99
700	89	86	93	93	94	94	98	98
1000	94	93	95	95	95	95	99	99
2000	95	94	95	95	95	95	100	100
10000	90	90	95	95	95	95	100	100

(b) Sample size = 10000.

factor	ADF3		ADF2		NORM		ELL	
skewness	GLS	LGLS	GLS	LGLS	GLS	LGLS	GLS	LGLS
0.1	95	95	97	97	97	97	97	97
0.25	97	97	95	95	95	95	95	95
0.5	95	95	96	96	96	96	96	96
1.0	97	97	95	95	95	95	96	96
2.0	90	90	95	95	95	95	100	100
3.0	96	96	97	97	97	97	100	100
4.0	96	94	92	92	92	92	100	100

Table 2.

Average bias of the loadings estimates. (Factor skewness = 2.0.)

sample	ADF3		ADF2		NORM		ELL	
size	GLS	LGLS	GLS	LGLS	GLS	LGLS	GLS	LGLS
300	−0.04	−0.02	−0.01					
500	−0.04	−0.03	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
700	−0.02	−0.01						
1000	−0.02	−0.01						
2000	−0.01	−0.01						
10000								

Note. Blank means absolute value of bias was less than 0.005.

From Table 3 it can be seen that the ADF3 estimators had indeed slightly smaller standard deviations (were more efficient) than the second order estimators. The differences, however, were very small. At smaller sample sizes LGLS estimators were better than GLS estimators for the ADF methods.

Table 3.

Average standard deviation of the loadings estimates. (Factor skewness = 2.0.)

sample	ADF3		ADF2		NORM		ELL	
size	GLS	LGLS	GLS	LGLS	GLS	LGLS	GLS	LGLS
300	0.085	0.078	0.127	0.082	0.081	0.081	0.081	0.081
500	0.058	0.058	0.069	0.063	0.063	0.063	0.063	0.063
700	0.050	0.050	0.065	0.052	0.052	0.052	0.052	0.052
1000	0.042	0.041	0.065	0.044	0.043	0.043	0.043	0.043
2000	0.030	0.030	0.031	0.031	0.031	0.031	0.031	0.031
10000	0.013	0.013	0.014	0.014	0.014	0.014	0.014	0.014

Note. Figures were calculated as follows: In each of 300 replications with the same skewness and sample size, factor loadings were estimated by each of the eight methods. For each of the four loadings, the standard deviation was calculated per method. These four standard deviations were averaged.

Table 4.

Percentage of times the confidence intervals for the loadings covered the true value. (Factor skewness = 2.0; $\alpha = .05$)

sample	ADF3		ADF2		NORM		ELL	
size	GLS	LGLS	GLS	LGLS	GLS	LGLS	GLS	LGLS
300	74	82	91	93	81	81	94	94
500	81	85	91	92	83	83	93	93
700	85	88	94	94	85	85	95	95
1000	88	90	93	94	84	84	94	94
2000	88	90	94	94	82	82	95	95
10000	94	94	94	94	83	83	95	95

Confidence intervals

The ADF3 confidence intervals were correct only in very large samples (see Table 4). This result was independent of factor skewness (Table 5), but strongly related to bias (cf. Table 2). The ADF2 confidence intervals were generally satisfactory. At larger skewnesses (larger departure from symmetry), the NORM confidence intervals for the loadings and the ELL confidence intervals for the error standard deviations were unreliable (see Tables 4 and 5), a result that was independent of sample size, but due to the inconsistency of the estimated asymptotic covariance matrix (cf. Browne, 1987, for the normal case). For the ADF methods the LGLS confidence intervals were better than the GLS ones, most probably because of the smaller bias of the LGLS estimators.

Table 5.

Percentage of times the confidence intervals for the error standard deviations covered the true value. (Sample size = 1000; α = .05.)

factor	ADF3		ADF2		NORM		ELL	
skewness	GLS	LGLS	GLS	LGLS	GLS	LGLS	GLS	LGLS
0.1	90	90	96	96	97	97	97	97
0.25	88	89	95	95	95	96	96	96
0.5	89	90	96	96	96	96	96	96
1.0	88	89	94	94	94	94	97	97
2.0	90	89	95	95	96	96	99	99
3.0	88	88	94	94	95	95	100	100
4.0	89	88	96	96	96	96	100	100

4. DISCUSSION

The ADF3 method generally performed a little worse than the ADF2 and NORM methods, although confidence intervals of the NORM estimators should be corrected, for example by using the "robust covariance matrix of the estimator" in EQS (Bentler, 1989). However, the differences in performance were small in larger samples, and the advantages of the ADF3 method, such as identification and information, may outweigh its somewhat worse performance. This may particularly be the case if the ADF3 estimates are corrected for bias, for example by using bootstrap or jackknife estimators (see, e.g., Efron, 1982).

The LGLS estimators generally did not perform worse than their GLS counterparts, and they may be preferred because of their computational efficiency.

Although not the main focus of attention, the ELL estimators were also studied, and because the test statistic and the confidence intervals performed so badly, they cannot be recommended.

The results obtained here apply only to the relatively simple 4-variable, 1-factor, factor analysis model with normal errors and a skewly distributed factor. It is hoped that the conclusions generalize to more complicated models with more variables and factors, but this remains to be investigated.

REFERENCES

[1] **Bentler, P.M.** (1983). "Some contributions to efficient statistics in structural models: specification and estimation of moment structures". *Psychometrika*, **48**, 493–517.

[2] **Bentler, P.M.** (1989). *EQS structural equations program manual.* Los Angeles: BMDP Statistical Software.

[3] **Bentler, P.M. & T. Dijkstra** (1985). "Efficient estimation via linearization in structural models". In: P.R. Krishnaiah (ed.), *Multivariate Analysis VI* (pp. 9–42). Amsterdam: Elsevier Science.

[4] **Browne, M.W.** (1982). "Covariance structures". In: D.M. Hawkins (ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, UK: Cambridge University Press.

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

[6] **Browne, M.W.** (1987). "Robustness of statistical inference in factor analysis and related models". *Biometrika*, **74**, 375–384.

[7] **Browne, M.W. & A. Shapiro** (1988). "Robustness of normal theory methods in the analysis of linear latent variate models". *British Journal of Mathematical and Statistical Psychology*, **41**, 193–208.

[8] **Efron, B.** (1982). *The jackknife, the bootstrap and other resampling plans.* Philadelphia: SIAM.

[9] **Jöreskog, K.G.** (1967). "Some contributions to maximum likelihood factor analysis". *Psychometrika,* **32**, 443–482.

[10] **Jöreskog, K.G. & A.S. Goldberger** (1972). "Factor analysis by generalized least squares". *Psychometrika*, **37**, 243–260.

[11] **Kano, Y., P.M. Bentler & A. Mooijaart** (1993). "Additional information and precision of estimators in multivariate structural models". In: K. Matusita, M.L. Puri, T. Hayakawa (eds.), *Statistical sciences and data analysis, Proceedings of the Third Pacific Area Statistical Conference* (pp. 187–196). Utrecht, The Netherlands: VSP.

[12] **Mooijaart, A.** (1985). "Factor analysis for non-normal variables". *Psychometrika,* **50**, 323–342.

[13] **Mooijaart, A. & P.M. Bentler** (1986). "Random polynomial factor analysis". In: E. Diday, Y. Escoufier, L. Lebart, J.P. Pagès, Y. Schektman, & R. Tomassone (eds.), *Data analysis and informatics, IV* (pp. 241–250). Amsterdam: North Holland.

[14] **Mooijaart, A. & P.M. Bentler** (1991). "Robustness of normal theory statistics in structural equation models". *Statistica Neerlandica,* **45**, 159–171.

[15] **Van Monfort, K., A. Mooijaart & J. De Leeuw** (1987). "Regression with errors in variables: estimators based on third order moments". *Statistica Neerlandica*, **41**, 223–239.