Some Comments on LISREL and its Sensitivity Analysis

(1994年4月7日受理)

福 森 護
Mamoru Fukumori

Key words: LISREL, Sensitivity Analysis, Aptitude Test

Abstract

The present paper discusses principles and methods of LISREL and its sensitivity analysis. In numerical example, I tried to get better models by making use of convenient functions of LISREL for the evaluation and modification of models. Furthermore, in the final stage of analysis I tried to apply sensitivity analysis in LISREL models to evaluate the stability or reliability of the results of analysis. The results showed that LISREL and its sensitivity analysis were useful for the analysis of this kind of data.

1 Introduction

As for statistical method, PCA and exploratory factor analysis such as principal factor analysis (PFA) are often used in the areas of psychology and sociology to investigate complex structures of various phenomena. These methods have great effectiveness in exploring the number and rough sketch of dimensions or factors. Sometimes, however, structural equation models including LISREL (Jöreskog and Sörbom, 1986) provides more appropriate interpretation of factors. The structural equation model is used to specify the phenomemon under study in terms of putative cause and effect variables and their indicators. The models have been useful in attaching many substantive problems in the social and behavioral sciences.

Sensitivity analysis in multivariate methods has been studied by several authors to evaluate the amount of influence of small changes of data. Some of the related literatures which deal with this topic are discussed the following works. In discriminant analysis, Canpbell (1978) has derived some influence functions while Sibson (1979) has proposed perturbation analysis in classical multidimensional scaling (MDS). In the field of principal component analysis (PCA), Critchley (1985), Jolliffe (1986), Benasseni (1988), Tanaka (1988) and Romanazzi (1990) among others have utilized the perturbation theory of eigenvalue problems to study the influence on PCA. Meanwhile, some researches on sensitivity analysis in covariance structure analysis including LISREL models have been considered in the papers of Tanaka, Watadani and Moon (1991), Tanaka and Watadani (1992). The main purpose of this paper is to examine the effectiveness of LISREL and its sensitivity analysis for the analysis of this kind of data.

2 LISREL model.

Tha LISREL model (see, e. g., Jöreskog and Sörbom, 1986, Bollen, 1989) is composed of three equations (one structural equation and two measurement equations), i. e. , (1) structural equation model: $\eta = B\eta + \Gamma\xi + \zeta$, (2) measurement model for y: $y = \Lambda_y\eta + \epsilon$, and (3) measurement model for x: $x = \Lambda_x\xi + \delta$.

In the structural equation, $\eta(m \times 1)$ is a vector of latent endogenous variables; $\xi(n \times 1)$ is a vector of latent exogenous variables; $B(m \times m)$ is a coefficient matrix of the effects of endogenous on endogenous variables; $\Gamma(m \times n)$ is a coefficient matrix of the effects of exogenous on endogenous variables; and $\zeta(m \times 1)$ is a vector of residuals, or errors in equations.

In the measurement equations, $y(p \times 1)$ and $x(q \times 1)$ are observation vectors of dependent variables and independent variables, respectively; $\Lambda_y(p \times m)$ and $\Lambda_x(q \times n)$ are the coefficient or loading matrix of y on the latent dependent variable η and of x on the latent independent variable ξ, respectively; $\epsilon(p \times 1)$ and $\delta(q \times 1)$ are vectors of errors of measurement of y and of x, respectively.

In the case of confirmatory factor analysis, only the measurement equation for x is used among the three equations of the LISREL model. In this case, covariance matrix of $z = (y^T, x^T)^T$ is given as

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{xy}^T \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}$$

where $\Sigma_{yy} = O$, $\Sigma_{xy} = O$, $\Sigma_{xx} = \Lambda_x\Phi\Lambda_x^T + \Theta_\delta$, $V(\underline{\xi}) = \Phi$ and $V(\underline{\delta}) = \Theta_\delta$.

The parameter estimates are obtained using a method of unweighted least squares (ULS), generalized least squares (GLS) or maximum likelihood (ML). In the present paper we use the ML method, in which the log-likelihood function

$$\log L = const - \left(\frac{n}{2}\right)\{\log|\Sigma_*| - tr\,(\Sigma_*^{-1} S)\}$$

is maximized, where S and Σ_* indicate the sample covariance matrix and the covariance matrix reproduced from the parameters.

Several measures have been proposed for evaluating the goodness-of-fit of the assumed model. Among them we mainly refer to the goodness-of-fit likelihood ratio test statistic X^2, the goodness-of-fit index(GFI) and the root mean-square residual(RMR) which are defined by

$$X^2 = n\log[\det(\hat{\Sigma})] - n\log[\det(S)]$$

$$GFI = 1 - \frac{tr(\hat{\Sigma}^{-1}S - I)^2}{\mathrm{tr}(\hat{\Sigma}^{-1}S)^2}$$

$$RMR = 1 - 2\sum_{j=1}^{k}\sqrt{\sum_{j=1}^{k}\frac{(S_{ij} - \hat{\sigma}_{ij})^2}{k(k+1)}}$$

where S_{ij} and $\hat{\sigma}_{ij}$ are the observed and reproduced covariances, respectively, and $\hat{\Sigma} = (\hat{\sigma}_{ij})$.

In addition, in the LISREL program the so-called medification indices are available for convenience to modify the assumed model. They are defined as

$$MI_i = \left[\frac{\partial \log L}{\partial \theta_i}\right]^2 [\, I^{-1}(\hat{\theta})]_{ii} \quad i=1,\ 2,\ \cdots,\ p.$$

for the i-th fixed (or constrained) parameter, which indicates the local improvement of $\log L$ when θ_i is freed, where $\hat{\theta}$ is the estimated parameter vector and $I(\hat{\theta})$ is the estimated information matrix. These indices may be judged by means of a chi-squared distribution with 1 degree of freedom.

3 Sensitivity analysis

The aim of sensitivity analysis is to evaluate the stability or sensitivity of the results of analysis and detect influential observations if any. Sensitivity analysis has been studied by Tanaka, Watadani and Moon(1991) , Tanaka and Watadani (1992) , and Tanaka, Watadani and Inoue (1992) , in covariance structure analysis including LISREL models. They have proposed the generalized Cook's distance *CD*, *COVRATIO*-like measure *CVR*, and ΔX^2 as influence measures on the estimated parameters, the precision of the estimates and the goodness-of-fit, respectively. They are defined as follows:

$$CD = (n-1)^2 [\, IF(x_i;\hat{\theta})]^T [\, acov(\hat{\theta})]^{-1} [\, IF(x_i;\hat{\theta})]\ ,$$

$$CVR = \frac{|acov(\tilde{\theta}_{(i)})|}{|acov(\hat{\theta})|},$$

$$\Delta X^2 = \tilde{X}^2_{(i)} - X^2,$$

where $IF(x_i;\hat{\theta})$ is the empirical influence function (*EIF*) for the estimated parameters $\hat{\theta}$; $acov(\hat{\theta})$ is the estimated asymptotic covariance matrix of $\hat{\theta}$; $\tilde{\theta}_{(i)}$ is the parameter estimates based on the sample without the i-th observation; X^2 and $\tilde{X}^2_{(i)}$ are the goodness-of-fit likelihood ratio test statistics for the whole sample and for the sample without the i-th observation, respectively. Symbol (~) indicates that the quantity with it is based on the linear approximation using the *EIF*.

The general procedure of sensitivity analysis proposed by Tanaka and his coworkers can be described as follows:

1) Compute the *EIF* for $i=1,\ 2,\ \cdots,\ n$.
2) Summarize the *EIF* vector into scalar measures such as *CD*, *CVR* and ΔX^2, and search for observations which have large values of those measures, as candidates for individually influential observations.
3) Search for sets of observations, as candidates for influential subsets, which are individually relatively influential and also have similar influence patterns using principal component analysis and other multivariate techniques.

4) Confirm the influence of single or multiple observations by actually omitting them.

4 Numerical example

The aptitude test data which was presented by Fukumori (1992, 1993a, 1993b) were analyzed with LISREL and its sensitivity analysis in the following steps.

Step 1. To begin with, from the results of PFA we assumed a three-factor model with nonzero loadings in the first to sixth items of factor 1, the sixth to thirteenth items in factor 2 and the thirteenth to sixteenth items in factor 3, and then analyzed the data using the software LISREL 6. The results are given in Table 1.

Table 1. Results of confirmatory factor analysis using LISREL (Initial model)

	Factor 1			Factor 2			Factor 3			Θ_δ
Item	loading	t	MI	loading	t	MI	loading	t	MI	
Correspondence	.696	12.32				.31			.02	.515
Correction (1)	.782	14.47				2.24			6.10	.389
Correction (2)	.804	15.08				.34			.17	.353
Search (1)	.615	10.51				.37			2.60	.621
Search (2)	.709	12.62				1.12			2.64	.498
Reasoning (1)	.321	3.02		.398	3.66				2.38	.539
Diagram			9.00	.645	10.85				.40	.584
Memory			.21	.469	.7.44				.02	.780
Rule			.07	.620	10.32				3.52	.616
Discrimination			3.62	.633	10.61				.40	.599
Pattern recognition			1.13	.526	8.47				1.13	.724
Solution			.16	.610	10.13				6.78	.627
Combination			.11	.057	.26		.589	2.69		.593
Reasoning (2)			.01			.54	.667	10.51		.555
Truth or falsehood			.42			2.67	.599	9.40		.641
Syllogism			.46			5.33	.185	2.72		.966

(MI:modification index, Θ_δ: unique variance)

Goodness-of-fit indices:

$GFI = 0.940$, $RMR = 0.041$, $X^2 = 112.86$ ($d.\ f. = 99$)

It is noted that the model fits the data rather well as shown by the value 0. 940 of *GFI*. However, there are two cells in the loading matrix, i. e., "combination" in factor 2 and "syllogism" in factor 3, which have small values of loadings. Hence, as a candidate for better models, we modified the model with the additional constraints of those two loadings being zero.

Step 2. As the second step we reanalyzed the data assuming the model mentioned above. The results are given in Table 2. The value 0.945 of *GFI* is slightly better than that of step 1. There are no loadings which are particularly small. But looking at the values of the modification indices, it seems that the loading of "diagram" on factor 1 may be better to be freed. Hence, for the next step we modified this part.

Step 3. As the third step we reanalyzed the data with the modified model. As expected the value of 0. 950 of

Table 2. Results of confirmatory factor analysis using LISREL (Modified model)

	Factor 1			Factor 2			Factor 3			Θ_δ
Item	loading	t	MI	loading	t	MI	loading	t	MI	
Correspondence	.697	12.34				.27			.01	.515
Correction (1)	.783	14.50				2.31			5.18	.387
Correction (2)	.803	15.06				.32			.09	.355
Search (1)	.615	10.50				.37			2.11	.622
Search (2)	.709	12.62				1.15			2.55	.498
Reasoning (1)	.323	3.04		.395	3.64				2.60	.540
Diagram			9.01	.645	10.86				.30	.584
Memory			.23	.469	7.44				.06	.780
Rule			.10	.618	10.29				3.49	.618
Discrimination			3.53	.635	10.64				.65	.597
Pattern recognition			1.17	.525	8.46				.97	.725
Solution			.11	.612	10.17				6.43	.626
Combination			.27			.01	.640	10.39		.590
Reasoning (2)			.00			1.05	.656	10.69		.570
Truth or falsehood			.31			1.32	.592	9.48		.650

(MI:modification index, Θ_δ: unique variance)

Goodness-of-fit indices:

$GFI=0.945$, $RMR=0.039$, $X^2=95.51$ ($d.f.=86$)

GFI was better than that of step 2. However, the loading of "diagram" on factor 1 became negative, and this fact caused the difficulty of interpretation. So we went back to the previous step, and looked again the loadings and modification indices. However, as it seemed that there were no parts to be modified, we decided to adopt this model as the best one.

Step 4. We applied this general procedure of sensitivity analysis to the final results of the LISREL analysis in the previous section. In this sensitivity analysis we calculated the *EIF* for *GFI* and *RMR*, which were denoted by $GFI^{(1)}$ and $RMR^{(1)}$, respectively, in addition to the measures *CD*, *CVR* and ΔX^2. Figure 1 shows the index plots of $GFI^{(1)}$ *CD* and *CVR*.

From the plot of *CD*, it seems that there are some observations, such as numbers 17, 55, 59 and 253 which may be influential to the parameter estimates. Also, the plots of the *CVR* and *GFI* reveal that there may be a few observations which are influential to the precision or the goodness-of-fit.

To search for influential subsets we applied principal component analysis to the set of the *EFI* vectors for all the parameters. The eigenvalues and their proportions in parenthesis of the first six principal components were 34.167(0.410), 8.601(0.103), 5.583(0.067), 5.333(0.064), 3.163(0.038) and 2.681(0.032).

As an illustration we show the scatter plot of the first versus the second principal components. Considering up to the fifth principal components we found the following five candidates for influential subsets: (96,194), (17,55,73,213), (39,53,68,89), (59), and (253,122).

Then, we applied LISREL analysis five times assuming the final model in the previous section to the sample

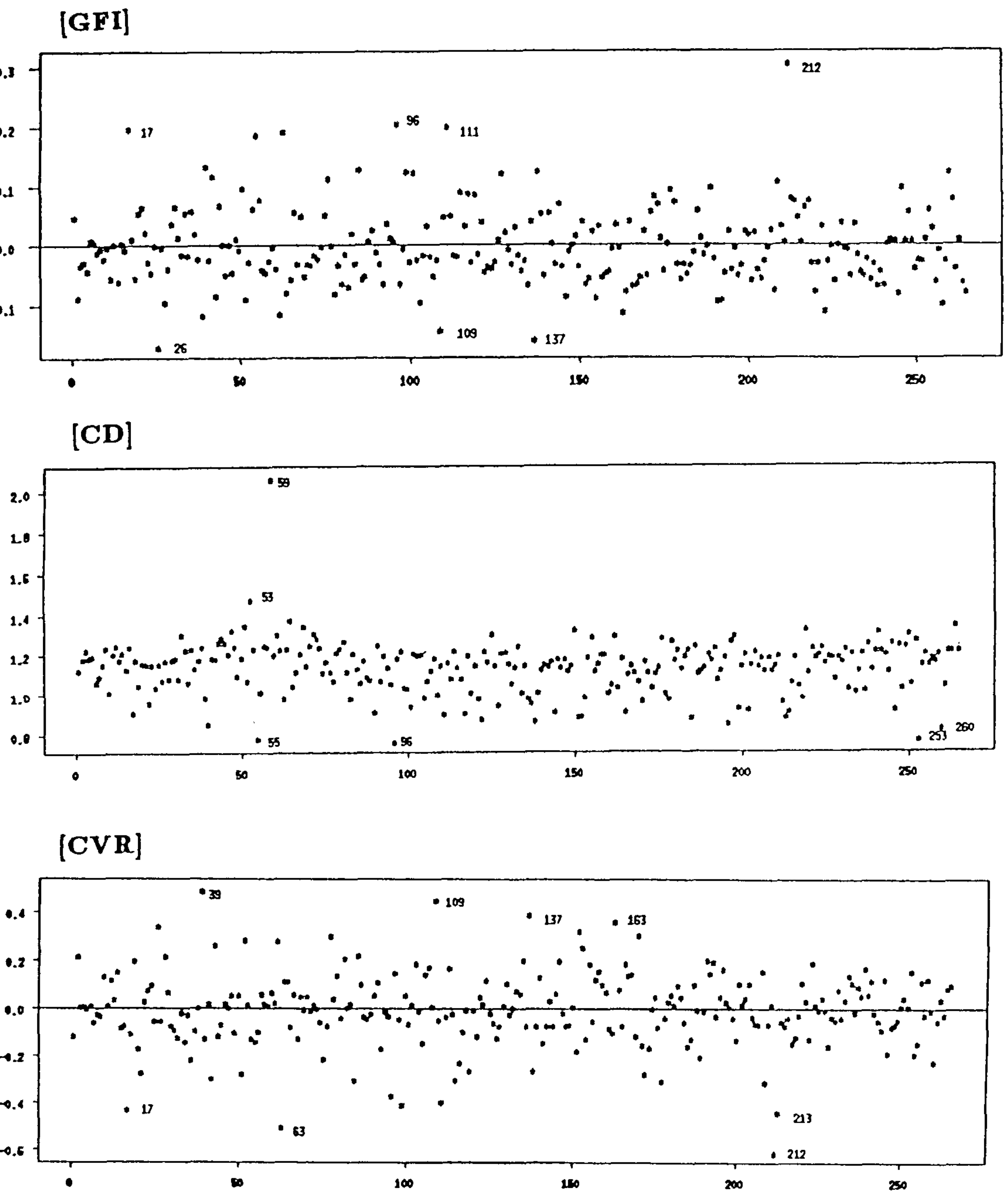

Figure 1: Index plots of $GFI^{(1)}$, CD and CVR

with one of the five subsets removed in turn. The values of GFI varied from 0.947 (the largest value when subset (17,55,73,213) was deleted) to 0.943 (the smallest when subset (39,59,68,89) or (253,122) was deleted). These values are not so different from 0. 945 for the whole sample. Also we could not find any great change in the parameter estimates. Thus it can be said that the results of LISREL analysis with the final model are stable for small changes of the data, and this fact can be considered to support the validity of the model.

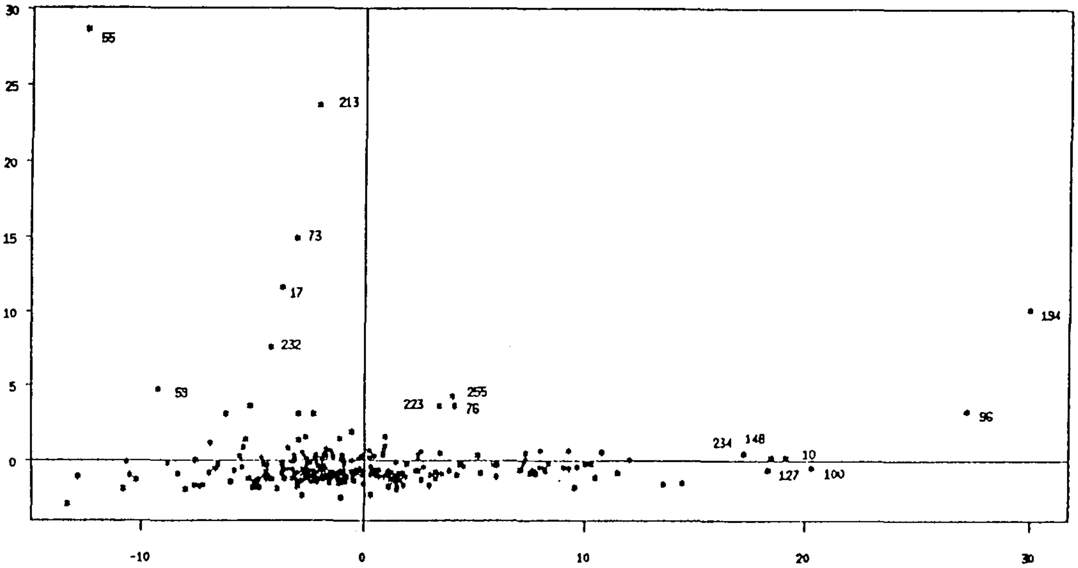

Figure 2: Scatter plots of the first two principal components

5 Concluding remarks

We found that LISREL and its sensitivity analysis are effective to get a good model in the sense that it not only fits the data well but also is stable and easy to interpret. As tools to search for better models the LISREL program has convenient functions. In particular, t-values for the estimated parameters and modification indices are very useful. They show the importance of each parameter which is contained in the model and of each parameter which is not contained in the model, respectivery. Referring to those values we can remove unimportant parameters and choose important ones. As for sensitivity analysis it plays just the same role in the LISREL analysis as the regression diagnostics in regression analysis. Using the function of sensitivity analysis we can investigate whether the results of analysis are stable and whether there are influential observations which have strong effects on the results. In our study we could confirm that the obtained results were not unstable.

References

1) Benasseni, J. (1988) Sensitivity of principal component analysis to data perturbation. *Data Analysis and Informatics, V.* (Edited by Diday, E.), 303-310, Elsevier.

2) Bollen, K. A. (1989). Structural Equations with Latent Variables. Wiley.

3) Critchley, F. (1985). Influence in principal component analysis. *Biometrika*, 72, 627-636.

4) Fukumori. M. (1992). An Analysis of the Information Processing Ability Using the LISREL Model. Journal of Chugoku Junior College, Vol. 23,165-171.

5) Fukumori. M. (1993a). Application of LISREL and its Sensitivity Analysis to the Data Obtained with Aptitude Test. Journal of Chugoku Junior College, Vol. 24, 257-263.

6) Fukumori. M. , Inoue. K. and Tanaka. Y. (1993b). Statistical Analysis of the Data Obtained with an Aptitude Test: Application of LISREL and its Sensitivity Analysis. Journal of the Japanese Society of Computational Statistics.

7) Hirai, Y. (1990). On some part of abilities in information processing (1) . Bulletin of School of Education, Okayama University No. 84, 179-186. (In Japanese)

8) Hirai, Y. and Fukumori. M. (1990). On some part of abilities in information processing (2) . Bulletin of School of Education, Okayama University No. 85, 61-67. (In Japanese)

9) Jolliffe, I. T. (1986). Principal Component Analysis. Springer Verlag.

10) Jöreskog, K. G. and Sörbom, D. (1986). LISREL 6-Analysis of Linear Structural Relationship by Maximum Likelihood, Instrumental Variables, and Least Squares Methods. Scientific Software, Inc.

11) Sibson, R. (1979). Studies in the robustness of multidimensional scaling: perturbation analysis of classical scaling. *J. Roy. Statist. Soc.* B41, 217-229.

12) Tanaka, Y. and Watadani, S. (1992). Sensitivity analysis in covariance structure analysis with equality constraints. *Comm. Statist. Soc.* A21, 1501-1515.

13) Tanaka, Y. , Watadani, S. and Inoue, K. (1992). Sensitivity analysis in structural equation models. In *"Computational Statistics Vol. 1"* (Edited by Y. Dodge and J. Whittaker), Physica-Verlag, 493-498.

14) Tanaka, Y. , Watadani, S. and Moon, S. H. (1991) Influence in covariance structure analysis: with an application to confirmatory factor analysis. *Comm. Statist.,* A20, 3805-3821.