STRUCTURAL EQUATION MODELS WITH LATENT VARIABLES

Albert Satorra
Departament d'Economia i Empresa
Universitat Pompeu Fabra

Structural Equation Modeling (SEM) is widely used in behavioural, social and economic studies to analyse structural relations between variables, some of which may be latent (i.e., unobservable). At present, SEM encompasses a wide variety of models and methods for multivariate analysis, such as multiple and multivariate regression, errors-in-variable models, ordered-probit regression, multiple-indicator models, factor analysis, simultaneous equation models, models for panel data, growth-curve models, and so forth. Furthermore, the observed variables can be normally distributed, but also continuous non-normal, or just ordinal; in addition, the data can have a single- or multiple-group, or multilevel structure, or can be mixture-distribution.

Nowadays, there exists a variety of commercial computer programs to carry out the practice of SEM, under general conditions on the model and the type of data. The most typical programs are LISREL, of Jöreskog and Sörbom; EQS, of Bentler; Mplus, of Muthén; the procedure CALIS of SAS, among others. For a thorough historical perspective of SEM, and its present state of the art, the reader should consult the recent book of Cudeck, Du Toit & Sörbom (2001).

Catalunya has various research groups with recognized international status in the area of SEM. There is Batista and Coenders's group (Universitat Ramon Llull and Universitat de Girona) that has contributed applications of SEM in research in social and behavioural sciences, specially on the methodology of measurement. The group of Maydeu-Olivares (Universitat de Barcelona) that has contributed with fundamental work on SEM analysis of rank-ordered and general discrete-type of data. The group of Ferrando and Lorenzo-Seva (Universitat Rovira i Virgili) working on SEM analysis of discrete data and issues of rotation in Factor Analysis. Since some of these groups are also contributing to the present issue, below I will just concentrate on SEM at UPF.

Power of the test and model search: Power analysis is required when we want to assess the substantive significance of coefficients of the model, and/or the validity of the model. A procedure for computing the power of the test in SEM is developed in Satorra and Saris (1985). Goodness of fit testing and power issues are also worked out in Satorra (1989) and, more recently, in Satorra (2001). A procedure for model search and model modification is worked out in Saris, Satorra and Sörbom (1987). The so-called "expected parameter change", which was developed in this paper, is nowadays implemented in most of the software for SEM.

Asymptotic robustness for multiple-group data: Asymptotic robustness (AR) concerns the validity of inferences derived under the NT assumption when the data deviates from normality. Since multivariate data often deviates from the normality assumption, assessing the validity of an analysis based on the normality assumption is of crucial importance in practice. AR issues for statistics arising in weighted-least-square analysis of SEM are investigated in Satorra and Bentler (1990). AR for mean-and-covariance structures is investigated in Satorra (1992) and Satorra and Neudecker (1994). AR issues for multiple-group mean-covariance structures is investigated in Satorra (1993) and Satorra (2002).

Scaled and adjusted test statistics: In many applications, chi-square goodness-of-fit tests are used when data are non-normal and there is non-compliance of assumptions required for asymptotic robustness. In such situations, corrected test statistics may be useful. Scaled and adjusted test statistics (for chi-square goodness-of-fit test, difference, score and Wald type test statistics) appropriate for non-normal data are developed in Satorra and Bentler (1994), Satorra (2000) and Satorra and Bentler (2001) (the scaled statistic of Satorra and Bentler (1994) is now available in most of the software for SEM).

SEM with categorical (ordinal) data, multilevel, and complex samples: Muthén and Satorra (1989) and Muthén and Satorra (1995) investigate SEM issues for multilevel and complex-sample structure of the data. Theory for SEM for continuous and categorical (ordinal) data can be found in Muthén and Satorra (1995).

Combining data sets from different sources: Multiple-group SEM has been shown to play a central role on the problem of combining data sets from different sources (Satorra, 1997). This is an issue that is recently attracting a lot of attention from different areas of applications.

Matrix algebra aspects of SEM: Matrix algebra has become essential for multivariate analysis. Advanced matrix algebra methods are needed to implement SEM in practice (specially when using high-level computer programming such as Matlab, or R). A general matrix formulation for testing equality of moments can be found in Satorra and Neudecker (1995). A matrix equality useful for goodness of fit testing in SEM is developed in Satorra and Neudecker (2003). Expressions for best linear prediction (as in estimation of factor scores) under general conditions are obtained by Neudecker and Satorra (2003).

Applications: On the side of applications, Ventura and Satorra (2001) use SEM on Spanish ECPF survey data to investigate variation of expenditure-patterns across various family groups, while Rivera and Satorra (2002) use multiple group SEM on ISSP-1993 data (a survey on twenty-two countries) to investigate the variation across countries of attitudes toward environmental issues.

Finally, we would like to mention other research areas in which statisticians at the UPF are involved, areas that relate with statistical modeling but that go beyond the scope of SEM. Work on the **description and visualization of multivariate data** is represented elsewhere under «Data Analysis and Data Mining» [Greenacre, at UPF]. Research on **information theory and non-parametric methods** is represented by Delicado (2000) [Delicado, now at UPC], Devroye, Lugosi and Udina (2001) Lugosi and Nobel (1999), and Udina (1999) [Lugosi and Udina, at UPF]. **Multilevel analysis** applied to data in education is represented by Cuxart-Jardí and Longford (2000) [Cuxart-Jardí, at UPF]. **Small are estimation** applied to regional data is represented by Costa, Satorra and Ventura (2002) [Ventura, at UPF]. **Analysis of preference data** by Van de Velden and Neudecker (2000) [Van de Velden, at UPF]. **Modeling human birth data and environmental statistics** is represented elsewhere under «Data Analysis and Data Mining» [Graffelman, now at UPC]. **Survival analysis with measurement error on covariates** is represented by Espinal-Berenguer and Satorra (1996) [Espinal, now at UAB]. Research on **subsampling and improved estimation** is represented by Romano and Wolf (2001) and Ledoit and Wolf (2002) [Wolf, at UPF]. Work on **stochastic processes** is represented elsewhere under «Research group on Stochastic Processes» [Kohatsu and Alós, at UPF]. There is also a vast amount of research on Operations Research and in Econometrics published by other colleagues at the UPF.

References

1. Costa, A., Satorra, A. and Ventura, E. (2002). «Estimadores Compuestos en Estadística Regional: una Aplicación a la Estimación de la Tasa de Variación de la Ocupación en la Industria». *Qüestiió*, 26, 213-243.
2. Cuxart-Jardí, A. and Longford, N. J. (1998). «Monitoring the University Admission Process in Spain». *Higher Education in Europe*, 23.
3. Cudeck, R., Du Toit, S. and Sörbom, D. (eds.). (2001). *Structural Equation Modeling: Present and Future*, SSI Scientific Software International, Lincolnwood, IL.
4. Devroye, L., Lugosi, G. and Udina, F. (2000). «Inequalities for a new data-based method for selecting nonparametric density estimates», in *Asymptotics in Statistics and Probability. Papers in Honor of George Gregory Roussas*, M. L. Puri (ed.), 133-154, International Science Publishers, Amsterdam.
5. Espinal-Berenguer, A. and Satorra, A. (1996). «Survival Analysis with Measurement Error on Covariates», in *Proceedings in Computational Statistics*, A. Prat, (ed.), 253-263, Springer Verlag, Heidelberg.
6. Ledoit, O. and Wolf, M. (2002). «Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size». *Annals of Statistics*, 30, 1081-1102.

7. Lugosi, G. and Nobel, A. B. (1999). «Adaptive model selection using empirical complexities». *The Annals of Statistics*, 27, 1830-1864.

8. Muthén, B. and Satorra, A. (1989). «Multilevel Aspects of Varying Parameters in Structural Models», in R. D. Bock (ed.), *Multilevel Analysis of Educational Data*, 87-99, New York: Academic Press.

9. Muthén, B. and Satorra, A. (1995*a*). «Complex Sample Data in Structural Equation Modeling». *Sociological Methodology*, 25, 267-316.

10. Muthén, B. and Satorra, A. (1995*b*). «Technical aspects of Muthen's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model». *Psychometrika*, 60, 489-503.

11. Neudecker, H. and Satorra, A. (1991). «Linear structural relations: Gradient and Hessian of the fitting function». *Statistics & Probability Letters*, 11, 57-61.

12. Rivera, P. and Satorra, A. (2001). «Analysing Group Differences: A Comparison of SEM Approaches», in *Latent Variable and Latent Structure Models* (G. Marcoulides and I. Moustaki, eds.) Lawrence Erlbaum Associates, Inc, Publishers, NJ.

13. Romano, J. P. and Wolf, M. (2001). «Subsampling intervals in autoregressive models with linear time trend». *Econometrica*, 69, 1283-1314.

14. Saris, W. E., Satorra, A. and Sörbom, D. (1987). «Detection and Correction of Specification on Errors in Structural Equation Models». *Sociological Methodology*, 17, 105-129.

15. Satorra, A. (1997). «Fusion of Data Sets in Multivaraite Linear Regression with Errors-in-Variables». *Classification and Knowledge Organization*, 195-207, R. Klar and O. Opitz (eds.), Springer Verlag: London.

16. Satorra, A. (1989). «Alternative Test Criteria in Covariance Structure Analysis: A Unified Approach». *Psychometrika*, 54, 131-151.

17. Satorra A. (1992). «Asymptotic robust inferences in the analysis of mean and covariance structures». *Sociological Methodology*, 22, 249-278.

18. Satorra, A. (1993). «Asymptotic Robust Inferences in Multi-Sample Analysis of Augmented-Moment Matrices», in *Multivariate Analysis: Future Directions*, 2, 211-229, (C. R. Rao, C. M. Cuadras, eds.), North-Holland, New York.

19. Satorra, A. (2000). «Scaled and Adjusted Restricted Tests in Multi-Sample Analysis of Moment Structures», in *Innovations in Multivariate Statistical Analysis: A Festschrift for Heinz Neudecker*. Heijmans, D. D. H., Pollock, D. S. G. and Satorra, A. (eds.), 233-247, Kluwer Academic Publishers, Dordrecht.

20. Satorra, A. (2001). «Goodness of fit testing of structural equation models with multiple group data and nonnormality», in *Structural Equation Modeling: Present and Future*, R. Cudeck, S. Du Toit and D. Sörbom (eds.), SSI Scientific Software International, Lincolnwood, IL.

21. Satorra, A. (2002). «Asymptotic robustness in multiple group linear-latent variable models». *Econometric Theory*, 18, 297-312.

22. Satorra, A. and Bentler, P. M. (1990). «Model Conditions for Asymptotic Robustness in the Analysis of Linear Relations». *Computational Statistics & Data Analysis*, 10, 235-249.

23. Satorra, A. and Bentler, P. M. (1994). «Correctios to test statistics and standard errors in covariance structure analysis», in *Latent variable Analysis in Developmental Research*, 285-305 (A. van Eye and C. C. Clogg, eds.), SAGE Publications, Inc., Thousand Oaks, CA.

24. Satorra, A. and Bentler, P. M. (2001). «A scaled difference chi-square test statistic for moment structure analysis». *Psychometrika*, 66, 507-514.

25. Satorra A. and Saris, W. E. (1985). «Power of the likelihood ratio test in covariance structure analysis». *Psychometrika*, 50, 83-89.

26. Satorra, A. and Neudecker, H. (1994). «On the asymptotic optimality of alternative minimum-distance estimators in linear latent-variable models». *Econometric Theory*, 10, 867-883.

27. Satorra, A. and Neudecker, H. (1997). «Compact Matrix Expressions for Generalized Wald Tests of Equality of Moment Vectors». *Journal of Multivariate Analysis*, 63, 259-276.

28. Neudecker, H. and Satorra, A. (2003). «On Best Affine Prediction». *Statistical Papers*, (forthcoming).

29. Satorra, A. and Neudecker, H. (2003). «A matrix equality useful in goodness-of-fit testing of structural equation models». *Journal of Statistical Planning and Inference*, (forthcoming).

30. Ventura, E. and Satorra, A. (2001). «Product expenditure patterns n the ECPF survey: An analysis using multiple group latent-variables models». *Qüestiió*, 25, 93-117.

31. Udina, F. (1999). «Implementing interactive computing in an object-oriented environment». *Journal of Statistical Software*, 5, 1-20.

32. Van de Velden, M. and Neudecker, H. (2000). «On an eigenvalue property relevant in correspondence analysis and related methods». *Linear Algebra and Its Applications*, 324, 347-364.