State space analysis for the behavioural sciences.

This issue of 'Kwantitatieve Methoden' contains a number of contributions that were presented during the seminar 'State space analysis for the behavioural sciences', which was held on July 5th 1995 at the occasion of the 9th European Meeting of the Psychometric Society, Leiden, the Netherlands.

One reason for organising this seminar was that state space analysis is a relatively little used technique in the behavioural sciences. State space analysis, popularly often referred to only as 'Kalman filtering', originated in the 1960's from the quite different science branch of engineering. It is a technique for estimating the unknown values of a latent state variable from a number of input and output variables, observed for one system or observation unit under conditions in which both information on the transition between input, output and state variables as well as on the noise is presumed available. While such information may be available from physical properties of systems in the physical and natural sciences, clearly such information is mostly unavailable in the social and behavioural sciences. Many authors have presented methods to estimate the transition parameters from noisy data, see e.g. Aoki (1990), Hannan & Deistler (1988) to name but a few. However, the focus of most of these models is on high-dimensional, accurate forecasting for one system or observation unit. In the social and behavioural sciences the focus is generally, instead of on such a 'black-box' approach to modelling, on low-dimensional inexact and thus suboptimal approximation for many observation units. It is therefore not surprising that classical state space modelling as such has found little application there. Instead, quite a number of adaptations of or approximations to state space modelling have been developed. It is the purpose of this issue to bring together a number of these approaches.

In this volume, Han Oud elaborates on the relation between traditional $N = 1$ (one subject or series) state space modelling and an approach to state space modeling that uses structural equations modeling (Oud et al, 1990). This approach is thus useful in situations where large samples have been measured. Using LISREL or EQS, a state space model can be analyzed and transition matrices estimated. Using these and the estimates of the error terms produced by these techniques, the Kalman filter can be applied for estimating the latent state space scores. The fact that the Kalman filter already after a very small number of time points produces estimates that are insensitive to the estimate for the starting value(s) for the state, comes in handy here. When the observed chains become long the model may become inefficient, in which case it is recommended that overlapping 'chunks' of the data matrix be analyzed separately and their parameters constrained to provide one solution.

Catrien Bijleveld and Frits Bijleveld extend the model for $N = 1$ state space analysis of mixed measurement level data that was proposed by Bijleveld and De Leeuw (1991) to situations where N is large. They compare two models for estimating the transition matrices and latent state values: one that uses Alternating Least Squares and thus fits in the Gifi-system and one that uses quasi-Newton methods. Both are suited especially for data analytic situations in which long chains have been measured. They recommend that both be extended so as to provide information on the statistical properties of the solution. They raise a number of issues in the interpretation of data gathered for many subjects as well as many time points.

Linda Collins and Alison Tracy evaluate the performance of longitudinal latent class models with multiple indicators. Such models are versions of the latent class models proposed by Van der Pol and De Leeuw (1986) in which time dependence is modelled by assuming that subjects progress through a number of categorical latent stages. The latent stages or latent classes have categorical indicators. If multiple items have been administered over multiple time points, the number of cells may increase very rapidly and a large part of them may thus be empty. Collins and Tracy report that the estimation of parameters is robust even for small sample sizes, even though caution is recommended in using second order models. They recommend that more research be conducted with respect to issues in model selection as the likelihood ratio statistic G^2 is not chi-square distributed.

Ab Mooijaart and Kees van Montfort also investigate longitudinal latent class models with categorical latent variables, assuming a first order Markov property for the latent variables. They create the possibility for group memberschip to change over time, effectively adding categorical input variables to the model. They use the EM algorithm to estimate the transition parameters, which they assume to be time-invariant for interpretation purposes. They show how this leads to an efficient algorithm which gives results that are well interpretable.

Jan de Leeuw, Catrien Bijleveld, Kees van Montfort and Frits Bijleveld propose a general framework for connecting all the above and other models for the analysis of multipe input-output series. Using theory from mixture distributions, they show how many models that use latent variables can be fitted in this encompassing framework, which is also applicable in situations where (mixtures of) continuous, interval, ordinal and nominal are present. They propose the EM-algorithm for estimating the unknowns. They recommend that the performance of their overall model be compared with existing methods for submodels.

We thank the reviewers for their contribution to the present collection of articles. We thank the editors of Kwantitatieve Methoden for giving us the opportunity to publish the articles in this volume.

Catrien Bijleveld
Ab Mooijaart

Leiden, January 14th 1997

Aoki, M. (1990). *State space modeling of time series*. Berlin: Springer.

Bijleveld, C.C.J.H. and De Leeuw, J. (1991). Fitting longitudinal reduced rank regression models by alternating least squares. *Psychometrika, 56*, 433-447.

Hannan, E.J. and Deistler, M. (1988). *The statistical theory of linear systems*. New York: Wiley.

Oud, J.H.L., van den Bercken, J.H. and Essers, R.J. (1990). Longitudinal factor score estimation using the Kalman filter. *Applied Psychological Measurement, 14*, 395-418.

Van der Pol, F. and De Leeuw, J. (1986). A latent Markov model to correct for measurement error. *Sociological Methods & Research, 15*, 118-141.