A COMPARISON OF THE APPLICABILITY OF TWO APPROACHES TO LINEAR DYNAMIC SYSTEMS ANALYSIS FOR N SUBJECTS

CATRIEN C.J.H. BIJLEVELD*
DEPARTMENT OF PSYCHOMETRICS & RESEARCH METHODOLOGY
LEIDEN UNIVERSITY, THE NETHERLANDS

FRITS D. BIJLEVELD
DEPARTMENT OF PSYCHOMETRICS & RESEARCH METHODOLOGY
LEIDEN UNIVERSITY, THE NETHERLANDS

ABSTRACT

We compare two approaches for linear dynamic systems analysis. In the first approach (Bijleveld & De Leeuw, 1991) estimates for parameters, latent state variables and optimal quantifications for the categories of any non-numerical input and output variables are obtained through an alternating least squares algorithm that fits in the Gifi (1990) approach to non-linear data analysis. In the second approach a quasi- Newton type of algorithm is used for estimating parameters and latent state variables (Bijleveld *et al.*, 1994).

Both approaches can analyse data sets with either one or several subjects. In all cases, the number of time points should be large. The two approaches differ in the constraints they use to reduce the indeterminacy of solutions. We compare the two methods in these respects, both in terms of the model specifications as well as through an empirical example.

We identify a number of important areas for further research to facilitate the applicability of either technique. These are: confidence intervals for the relevant entities in the solutions, guidelines for the choice on the number of latent state variables as well as guidelines for interpretation.

* dr. C.C.J.H. Bijleveld
Dept. Psychometrics & Research Methodology
Faculty of Social Sciences, Leiden University
P.O. Box 9555
2300 RB Leiden, the Netherlands

e-mail BYLEVELD@RULFSW.LeidenUniv.NL

INTRODUCTION

We consider situations where we have measured a number of input variables and a number of time-dependent output variables. Neither of the two types of variables are necessarily numerical, that is, they may be a mixture of variables measured at interval, ordinal or nominal measurement level. What is special about our variables is that the successive measurements are not independent. The output and input variables have been measured at time points t (t = 1,...,T), so that there is reason to assume that present observations depend to some extent on past observations. The output and input variables have been measured for N subjects. N may be large as is usual in many types of behavioural research. However, N may just as well be 1. However, in all cases we will assume that T is large; if T is small, other models provide better alternatives for our type of problem. In addition, we would like T to be large because we expect that the accuracy of our estimates improves using sufficiently large numbers of replications. T need not be the same for different subjects, that is, subjects may have been measured at varying numbers of time points.

Models such as these are useful in situations where one or more criterion or dependent variables have been measured over a successive number of time points. We want to predict the development of these dependent variables from a number of predictor or independent variables. Similar to a factor analytic or reductionist approach, we want to achieve our prediction in the lowest possible dimensionality. Such data-analytic situations occur frequently in developmental and clinical psychology, history, econometrics, in fact in all areas of science in which development or growth - and the prediction thereof - is the explicit research objective. Examples are the prediction of maternal behaviour from infant's behaviour in developmental psychology, the prediction of therapeutic alliance from therapeutic content in clinical psychology, the prediction of various cardiovascular indices from mood states in biomedical psychology, the prediction of maize prices from various combinations of economic and demographic indicators in pre-industrialist societies, and the like. In all such cases, a model is warranted that explicitly models temporal dependence, that incorporates successive measurements on one or more predictor variables, and that reduces the dimensionality of the problem.

In developing our models, we will build from and extend the classical state space models made famous by Rudolf Kalman (Kalman, 1960; Ho & Kalman, 1966). However, state space modelling or Kalman filtering has been developed for the $N = 1$ case from the engineering framework of exact, if necessary high-dimensional, forecasting. The behavioural framework focuses instead on approximation and interpretation. Therefore, classical state space analysis is less widely applicable in a behavioural context and the same applies to the closely related ARIMA models described by Box and Jenkins (Box & Jenkins, 1976; Akaike, 1976). It is therefore not surprising that the terminology 'state space modelling' for the type of models presented below has sometimes created unnecessary confusion and controversy, generated more by the differing paradigm and application of the models than by differences in model formulation. To avoid such confusion, we prefer to use the term 'linear dynamic systems analysis' in the following.

A number of authors have presented models that resemble our type of model. For data-analytic situations with a limited number of time points and relatively large numbers of subjects, the structural equations approach, first exemplified in Jöreskog & Goldberger's MIMIC model

(1975), provides an attractive approach to linear dynamic systems analysis (cf. MacCallum & Ashby, 1986; Oud *et al.*, 1990). However, whenever the number of time points becomes large, and is not accompanied by accordingly increased numbers of subjects, structural equations models have been reported to suffer from improper solutions (with for instance negative variance estimates, Bijleveld & Mooijaart (1993)). When only categorical variables are present, latent Markov chain analysis provides a family of models (van de Pol & Langeheine, 1989). *Ad hoc* adaptations of existing methodology may at times provide a suitable approximation to linear dynamic systems modelling. Examples can be given from multilevel modelling (Van der Leeden, 1997), growth curve models (Pothoff & Roy, 1964) as well as from various forms of non-linear multi-set analysis (Bijleveld & van der Burg, 1997). However, all such adaptations are in varying respects suboptimal.

In the following, we will compare two approaches for linear dynamic systems analysis that have been advanced recently, the first by Bijleveld and De Leeuw (1991), the second by Bijleveld *et al.* (1994). Both estimate parameters and latent state variables from the observed data, using a non-recursive procedure for the latter. Either technique can be applied in case more than one series has been observed. Both approaches can in principle handle mixed measurement level variables, although this has as yet been implemented in the first approach only. We will first outline our model. Next we will compare the two approaches with respect to particular model specifications. We will compare the practical applicability of the two techniques through an empirical example, and conclude with recommendations for further research.

LINEAR DYNAMIC SYSTEM FOR ONE SUBJECT

Suppose thus that we have observed k input variables $\mathbf{x}$ and m output variables $\mathbf{y}$, each at T consecutive occasions. We know that there is time-dependence in the measurements, which is modelled by supposing that the $\mathbf{x}$ influence the $\mathbf{y}$ through p-dimensional latent variables $\mathbf{z}$. The state at time t, $\mathbf{z}_t$, depends linearly on the input at time t, $\mathbf{x}_t$. The output at time t depends linearly on the state at time t, $\mathbf{z}_t$. The $\mathbf{z}$ accommodate the time dependence in the measurements by following a Markov type of dependency: the state at time t, $\mathbf{z}_t$, depends linearly on the state at time t-1: $\mathbf{z}_{t-1}$. As the state at time point t depends on the state at the previous time point t-1 only, this implies that all information relevant for forecasting the present is contained in the previous state. As such, the latent states serve as the memory of the system. Although the dimensionality of the state may have to be quite high for capturing the dynamic nature of the latent state values, we want - given our behavioural paradigm - the dimensionality of the $\mathbf{z}$ to be small and preferably lower than that of the smallest of the dimensionalities of $\mathbf{x}$ and $\mathbf{y}$. In that case, the $\mathbf{z}$ also filter the dependence of output on input. A visual representation of the linear dynamic model for one subject is in Figure 1.

We can write this model for one subject as:

$$\mathbf{z}_t = \mathbf{F}_t\, \mathbf{z}_{t-1} + \mathbf{G}_t\, \mathbf{x}_t + \boldsymbol{\delta}_t \qquad \text{system equation} \qquad (1a)$$

$$\mathbf{y}_t = \mathbf{H}_t\, \mathbf{z}_t + \boldsymbol{\varepsilon}_t \qquad \text{measurement equation} \qquad (1b)$$

where $\mathbf{F}_t$ is the $p \times p$ state transition matrix at time point t, $\mathbf{G}_t$ the $p \times k$ control matrix at time point t and $\mathbf{H}_t$ the $m \times p$ measurement matrix at time point t. The vector of errors $\boldsymbol{\delta}_t$ is $p \times 1$; $\boldsymbol{\varepsilon}_t$ is $m \times 1$. The usual properties apply to $\boldsymbol{\delta}_t$ and $\boldsymbol{\varepsilon}_t$, that is:

- $\boldsymbol{\varepsilon}_t \perp \mathbf{z}_{t-1}$; $\boldsymbol{\varepsilon}_t \perp \mathbf{x}_t$;
- $\boldsymbol{\delta}_t \perp \mathbf{z}_t$; $\boldsymbol{\delta}_t \perp \mathbf{x}_t$;
- $E(\boldsymbol{\varepsilon}_t) = E(\boldsymbol{\delta}_t) = 0$ for each $t = 1, ..., T$;
- $E(\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t)$ is finite; $E(\boldsymbol{\delta}_t'\boldsymbol{\delta}_t)$ is finite.

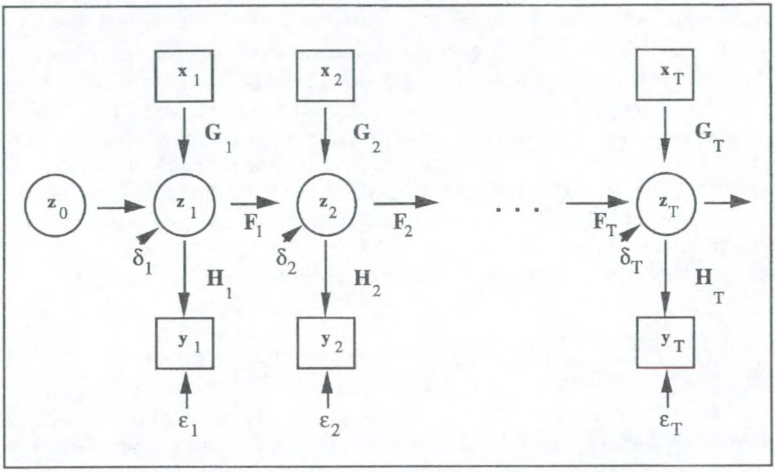

Figure 1. Linear dynamic system for one subject.

Note the inherent asymmetry in the model: the output depends on the input, and the present depends on the past. Model (1) is known under various names: linear dynamic system or model, state space model, longitudinal reduced rank regression model (cf. Ho & Kalman, 1966; Otter, 1986; Ljung, 1987; Hannan & Deistler, 1988; Aoki, 1990; Bijleveld & De Leeuw, 1991).

We want to investigate the impact of the input on the output, and to describe the relationships between the two sets of variables in as parsimonious a way as possible. As we (usually) have several input variables and (usually) several output variables, we are looking for a latent space for explaining the relation between $\mathbf{X}$ and $\mathbf{Y}$ that is of lower dimensionality than that of the $\mathbf{X}$ and the $\mathbf{Y}$.

We will first introduce a number of assumptions. We assume that the relations between the variables are time-invariant, that is, the transition matrices are the same across time points: $\mathbf{F}_t = \mathbf{F}$, $\mathbf{G}_t = \mathbf{G}$, $\mathbf{H}_t = \mathbf{H}$. While simplifying our model, these assumptions are also necessary in the sense that only when the transitions are time-invariant, we can be assured that the state variable is indeed the same variable across time points. In matrix notation for one subject, for all time points simultaneously, the model can then be written as:

$$\mathbf{Z} = \mathbf{BZF}' + \mathbf{XG}' + \mathbf{D} \tag{2a}$$

$$\mathbf{Y} = \mathbf{ZH}' + \mathbf{E}, \tag{2b}$$

with $\mathbf{Z}$ the $T \times p$ matrix containing the latent states from time 1 until time T, $\mathbf{X}$ the $T \times k$ matrix containing the input variables from time 1 until time T, $\mathbf{Y}$ the $T \times m$ matrix containing the output variables from time 1 until time T, $\mathbf{D}$ the $T \times p$ matrix containing the errors in the latent states from time 1 until time T, $\mathbf{E}$ the $T \times m$ matrix containing the errors in the output variables from time 1 until time T, and $\mathbf{B}$ the $T \times T$ shift matrix such that $\mathbf{B}(\mathbf{z}_1',\ldots,\mathbf{z}_t') = (\mathbf{z}_0',\ldots,\mathbf{z}_{t-1}')$. Several options are available for estimating , e.g. $\hat{\mathbf{z}}_0 = \mathbf{z}_1$, $\hat{\mathbf{z}}_0 = \mathbf{z}_T$ or $\hat{\mathbf{z}}_0 = \bar{\mathbf{z}}$. Note that, in general, $\mathbf{B}$ is identical to the well-known backshift operator from time series modelling.

In (2), the unknown entities are the transition matrices $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{H}$ as well as the matrix of latent states $\mathbf{Z}$. In practice, $\mathbf{D}$ and $\mathbf{E}$ are estimated as the prediction errors in $\mathbf{Y}$ and $\mathbf{Z}$.

There are several options for estimating the unknowns in (2) from the data. We follow Bijleveld and De Leeuw (1991) who proposed the following loss function:

$$\sigma_\omega(\mathbf{Z},\mathbf{F},\mathbf{G},\mathbf{H}) = \omega^2 \text{ SSQ } (\mathbf{Z} - \mathbf{BZF}' - \mathbf{XG}') + \text{SSQ } (\mathbf{Y} - \mathbf{ZH}'), \tag{3}$$

where SSQ stands for the sum of squares over all the arguments. The weight ω determines the relative importance of the input and output, and the possible values for ω define a class of solutions for $(\mathbf{Z},\mathbf{F},\mathbf{G},\mathbf{H})$. As such, the solutions generated by minimisation of (3) are dependent upon the choice for ω, and are thus in a sense arbitrary. In principle, (3) can be specified without the weight ω, but this amounts to specifying $\omega = 1$. And even though this might appear rather unequivocal, also the choice $\omega = 1$ is - strictly speaking - not without some arbitrariness, as it gives a solution that is still dependent upon the normalisation of the variables. There are only two choices for ω which are non-arbitrary: $\omega = 0$ and $\omega \to \infty$.

Looking at equation (3), one can easily see that, when ω becomes smaller, the first (input) part of the loss function is weighted by a smaller number, and any loss accrued in the system equation makes a smaller contribution to the total loss σ_ω. The smallest value ω can assume is 0. Then, any loss in the system equation has become totally irrelevant, as it is nullified by ω. In that case, the total loss is minimised by minimising only the second part: SSQ $(\mathbf{Y} - \mathbf{ZH}')$. Minimising (3) then amounts to a principal components analysis of the output variables $\mathbf{Y}$. With the $\mathbf{Z}$ constructed as the principal components of $\mathbf{Y}$, and given that $p < m$, this means that the linear space spanned by the latent state variables is a subspace of the linear space spanned by $\mathbf{Y}$. In other words: the latent state variables are a linear combination of the basis of $\mathbf{Y}$.

Conversely, when ω becomes larger, any loss accrued in the input part of the loss function will be weighted more heavily, and will thus makes a bigger contribution to the total loss σ_ω. In the limiting case where $\omega \to \infty$, the total loss can be minimised only, if loss from the system equation is zero, that is if $\mathbf{Z}$ satisfy $\mathbf{Z} = \mathbf{BZF}' + \mathbf{XG}'$. In that case, the latent state variables are a linear combination of any basis of $\mathbf{X}$ and the initial state $\mathbf{z}_0$. The loss accrued from the system equation is then 0, and the total loss equal to the remaining loss in the measurement equation. The space spanned by the latent state variables is then a subspace of the space spanned by the input variables and the initial state $\mathbf{z}_0$.

The reader may wonder why all the fuss is necessary, when in classical state space analysis methods such as those advanced by Kalman (1960) disregard ω (thereby effectively varying ω per time point, depending on the update scheme for the Kalman gain respectively the error terms $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\delta}_t$), and simply solve (1) for $\mathbf{Z}$? Such a comparison is flawed however, as in these methods assume that $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{H}$ are known, and that specific information on the error in the system and measurement equations is available. This makes it possible to arrive at a non-arbitrary solution. As we have no such information at our disposal in a behavioural context, we have to eliminate the arbitrariness in another manner. As stated above, the only two unequivocal solutions are $\omega = 0$ and $\omega \rightarrow \infty$. As the option $\omega = 0$ is not relevant in a dynamic context, the only remaining interesting option is $\omega \rightarrow \infty$. Note that this is the straightforward longitudinal analogue of reduced rank regression analysis (Van den Wollenberg, 1977).

The option $\omega \rightarrow \infty$ thus amounts to finding $\mathbf{Z}$ that satisfy:

$$\mathbf{Z} = \mathbf{BZF}' + \mathbf{XG}' \tag{4a}$$

that simultaneously minimise

$$\text{SSQ } (\mathbf{Y} - \mathbf{ZH}'). \tag{4b}$$

Note that now $\mathbf{D} = \mathbf{0}$. Another way to view model (4) is as a principal component analysis of the output with restrictions $\mathbf{Z} = \mathbf{BZF}' + \mathbf{XG}'$ on the component scores $\mathbf{Z}$.

EXTENSION OF THE LINEAR DYNAMIC MODEL TO N SUBJECTS

The classical state space model, like many time series models, has been developed for the analysis of single subject designs. The model outlined above can however easily be extended to situations where more than one subject has been measured, or $N > 1$. Several possibilities exist for the extension of the model to the simultaneous analysis of $N > 1$ designs (Bijleveld & Legendre, 1993; Bijleveld et al., 1991; Verboon & Heiser, 1990). The simplest of these seems to be through the construction of a super-data matrix, super-latent state matrix and a super-shift matrix.

This can be envisaged as follows. We start by stacking the input and output data for the N subjects in an $(NT \times k)$ matrix $\mathbf{X}_{\text{sup}}$ and an $(NT \times m)$ matrix $\mathbf{Y}_{\text{sup}}$ respectively, where we define NT as the sum of the respective numbers of time points T_i at which the N subjects have been measured:

$$NT \equiv \sum_{i=1}^{N} T_i.$$

The $\mathbf{X}_{\text{sup}}$ and $\mathbf{Y}_{\text{sup}}$ can be envisaged as follows:

$$\mathbf{X}_{\text{sup}} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \\ \mathbf{X}_4 \\ \cdot \\ \cdot \\ \mathbf{X}_N \end{bmatrix}, \quad \mathbf{Y}_{\text{sup}} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \mathbf{Y}_4 \\ \cdot \\ \cdot \\ \mathbf{Y}_N \end{bmatrix}.$$

If we similarly stack the latent state scores into an $(NT \times p)$ matrix of latent state scores $\mathbf{Z}_{\text{sup}}$, we can then construct an $(NT \times NT)$ block diagonal super-shift matrix $\mathbf{B}_{\text{sup}}$ using for each of the N blocks the shift operator $\mathbf{B}_i$. See Figure 2.

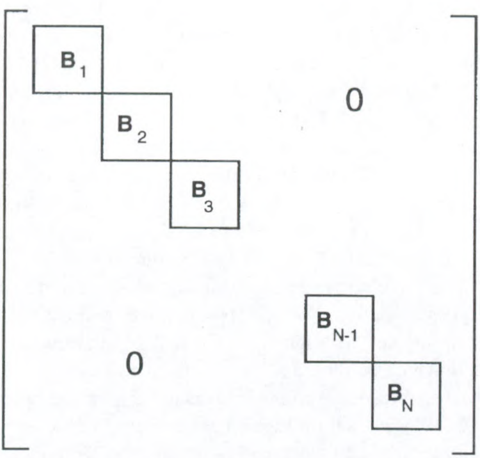

Figure 2. Block diagonal super-shift matrix $\mathbf{B}_{\text{sup}}$.

As we want to fit one model for all subjects simultaneously, we assume that the transition matrices $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{H}$ are the same across subjects: $\mathbf{F}_i = \mathbf{F}$, $\mathbf{G}_i = \mathbf{G}$, $\mathbf{H}_i = \mathbf{H}$. For N subjects, for all time points simultaneously, model (2) can then be written as:

$$\mathbf{Z}_{\text{sup}} = \mathbf{B}_{\text{sup}}\mathbf{Z}_{\text{sup}}\mathbf{F}' + \mathbf{X}_{\text{sup}}\mathbf{G}' + \mathbf{D}_{\text{sup}} \tag{5a}$$

$$\mathbf{Y}_{\text{sup}} = \mathbf{Z}_{\text{sup}}\mathbf{H}' + \mathbf{E}_{\text{sup}}. \tag{5b}$$

Note that, as all measurements for the various time points are stacked vertically, the number of time points may very well differ per subject: $T_i \neq T$. Thus: persons may have been measured during series of unequal length.

MODEL I: LINEAR DYNAMIC MODELLING USING MAJORISATION

As can be seen easily, there are far too many unknowns in either (2) or (5) to solve the minimisation problem uniquely. Bijleveld and De Leeuw (1991) opted to minimise the loss function using Alternating Least Squares. Their choice was guided by the fact that the minimisation problem can be divided into two parts: on the one hand solving for **Z**, which is possible if **F**, **G** and **H** are known, and on the other hand solving for **F**, **G** and **H**, which is possible if **Z** is known.

Thus, the minimisation problem was divided into two subproblems, that could each be solved by least squares minimisation procedures. Starting with an initial solution for **Z**, one obtains a sequence of solutions in which the loss is each time decreased, and which always converges to a solution. For details we refer to Bijleveld and De Leeuw (1991). To avoid trivial solutions, **Z** is standardised to $\mathbf{Z}'\mathbf{Z} = T\,\mathbf{I}$ and $\mathbf{u}'\mathbf{Z} = 0$ (with $\mathbf{u}$ a $T \times 1$ vector with all elements equal to 1). Because of the restriction $\mathbf{Z}'\mathbf{Z} = T\,\mathbf{I}$, estimates for **Z** cannot be obtained by ordinary least squares. Instead, the Bijleveld and De Leeuw algorithm uses majorisation to arrive at a solution for **Z**. Majorisation (De Leeuw, 1988) is a minimisation procedure in which a complicated loss function is minimised through the aid of a simpler loss function, which is an approximation of the complicated loss function in the sense that its values are always greater than or equal to the complicated loss function. Minimising the simple function, the complicated function is also decreased. The effectiveness and efficiency of the majorisation procedure depends of course on the appropriateness of the simpler 'majorising' function: the closer that function is to the original complicated loss function, the better performance is. In general, however, majorisation slows down convergence.

One major advantage of the procedure proposed by Bijleveld and De Leeuw is that it can handle non-numerical variables. Adding a third substep to each iteration step, the categories of the variables can be rescaled according to their measurement level. Similarly, least squares optimal estimates of missing observations can be obtained.

Implementation of the option $\omega \rightarrow \infty$ induces numerical problems, as it may lead to an ill-conditioned Hessian; this means slow minimisation and inaccurate solutions, a problem well known from penalty methods in minimisation theory (see e.g. Murray, 1969). In practice, the option $\omega = 1$ is thus used, with both input and output variables standardised to zero mean and standard deviation 1 to minimise arbitrariness. This standardisation is comparable to standardisations generally used in similar data analytic situations, such as factor analysis.

MODEL II: LINEAR DYNAMIC MODELLING USING QUASI-NEWTON METHODS

The second approach to linear dynamic modelling presented here was proposed by Bijleveld *et al.* (1994). It differs from the former approach in a number of respects. It has a somewhat different model formulation in which, firstly, variables are not standardised anymore, as is usual in many instances of behavioural research. Instead, an intercept is added to the model equations. For $N = 1$, the model can thus be written as follows:

$$\mathbf{z}_t = \mathbf{F}\ \mathbf{z}_{t-1} + \mathbf{G}\mathbf{x}_t + \boldsymbol{\delta}_t + \mathbf{v} \tag{6a}$$
$$\mathbf{y}_t = \mathbf{H}\ \mathbf{z}_t + \boldsymbol{\varepsilon}_t + \mathbf{w}. \tag{6b}$$

The vectors $\mathbf{v}$ and $\mathbf{w}$ have properties similar to the properties of intercepts in common regression models. In the following, the model equations have been written for one subject, but again can be generalised easily to situations where $N > 1$.

We suppose that the latent state variables are a linear combination of the input variables and the initial state, in other words: $\boldsymbol{\delta}_t = 0$. This means that we effectively implement $\omega \rightarrow \infty$. Equation (6a) then becomes:

$$\mathbf{z}_t = \mathbf{F}\ \mathbf{z}_{t-1} + \mathbf{G}\mathbf{x}_t + \mathbf{v}, \tag{7}$$

We can write the equations for the successive time points as:

$$\mathbf{z}_1 = \mathbf{F}\ \mathbf{z}_0 + \mathbf{G}\mathbf{x}_1 + \mathbf{v}$$
$$\mathbf{z}_2 = \mathbf{F}\ \mathbf{z}_1 + \mathbf{G}\mathbf{x}_2 + \mathbf{v} = \mathbf{F}\ (\mathbf{F}\ \mathbf{z}_0 + \mathbf{G}\mathbf{x}_1 + \mathbf{v}) + \mathbf{G}\mathbf{x}_2 + \mathbf{v}$$
$$\mathbf{z}_3 = \mathbf{F}\ \mathbf{z}_2 + \mathbf{G}\mathbf{x}_3 + \mathbf{v} = \mathbf{F}\ (\mathbf{F}\ (\mathbf{F}\ \mathbf{z}_0 + \mathbf{G}\mathbf{x}_1 + \mathbf{v}) + \mathbf{G}\mathbf{x}_2 + \mathbf{v}) + \mathbf{G}\mathbf{x}_3 + \mathbf{v}, \text{ etc.}$$

The state at time *t* can then be written more concisely as:

$$\mathbf{z}_t = \mathbf{F}^k\ \mathbf{z}_0 + \sum_{k=0}^{t-1} \mathbf{F}^k\ (\mathbf{G}\ \mathbf{x}_{t-k} + \mathbf{v}). \tag{8}$$

Equation (8) shows that the state at time t, $\mathbf{z}_t$, is thus completely determined by $\mathbf{z}_0$, $\mathbf{x}_1$, ..., $\mathbf{x}_t$, $\mathbf{F}$, $\mathbf{G}$ and the shift $\mathbf{v}$. Proof is in Bijleveld *et al.* (1994). Writing the system equations in this manner $\mathbf{Z}$ effectively becomes a linear combination of $\mathbf{z}_0$ and $\mathbf{X}$ plus an intercept term.

Equation (8) also shows that the terms $\mathbf{z}_t$ can thus be computed recursively once $\mathbf{F}$, $\mathbf{G}$, $\mathbf{v}$ and $\mathbf{z}_0$ have been estimated. Building on, and using the property that the latent state variables are a function of $\mathbf{X}$ and $\mathbf{z}_0$, the loss function becomes:

$$\sigma = \sum_{t=1}^{T} \sum_{j=1}^{m} (\mathbf{y}_{jt} - \sum_{k=1}^{p} \mathbf{H}_{jk}\mathbf{z}_{kt} - \mathbf{w}_j)^2, \tag{9}$$

where $\mathbf{H}_{jk}$ are the elements of the matrix $\mathbf{H}$ and $\mathbf{z}_{kt}$ is the *k*-th element of the *p*-dimensional latent state variable at time point *t*. The matrix $\mathbf{H}$ and the vector $\mathbf{w}$ are determined by $\mathbf{F}$, $\mathbf{G}$, $\mathbf{z}_t$ $\mathbf{v}$ and the input variables. For each combination of these, $\mathbf{H}$ and $\mathbf{w}$ can be determined.

It is now possible to estimate the values of the parameters for which the derivative of (9) is closest to zero with, for instance, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, an algorithm generally known to be fast, efficient and stable (Fletcher, 1981). For details, see Bijleveld *et al.* (1994).

EXAMPLE

For illustrative purposes, we will now compare the two models by analysing an example from psychosomatic medicine using each of the two models. We had at our disposal daily measures of well-being and distress that had been gathered in 59 healthy men and women in a study on menstrual symptoms (Bijleveld *et al.*, 1980). The measures had been collected for 35 consecutive days; this experimental time-span had been chosen to include at least one menstrual cycle for the female subjects. Men had been included in the sample to serve as control subjects. A self-administered questionnaire contained 53 items; it had been constructed from Moos's menstrual distress questionnaire (Moos, 1968) extended with 6 additional items reflecting positive moods that had emerged as meaningful in a pilot study. Subjects had indicated every day on a rating scale ranging from 1 to 6 whether they had 'no complaints regarding this item' (1) to 'many complaints regarding this item' (6). Factor analysis using the P-technique on the $59 \times 53 \times 35$ data box had produced nine factors (Van den Boogaard & Bijleveld, 1988). In the example here, we will further analyse the relations between the variables that constituted three of these factors.

More particularly, we will investigate to what extent variables typically associated with pre-menstrual and menstrual distress such as water retention and abdominal pain influence concentration. For this purpose, we selected four variables reflecting symptoms of water retention and two variables reflecting symptoms of abdominal pain. These variables serve as input variables in our example. Five variables had been used as indicators of concentration; for reasons of simplicity, we summed these into one variable, which we label 'concentration'. This variable serves as our the output variable. The three-dimensional data for our example now have dimensionalities $59 \times 6 \times 35$ for the input variables, and $59 \times 1 \times 35$ for the output variable. See Table 1 for an overview.

Table 1. Overview of Input and Output Variables used in the Example

Input Variables	Output Variable
ABDOMINAL PAIN	CONCENTRATION
CRAMPS	
PAINFUL BREASTS	
SWELLING	
SKIN DISORDERS	
WEIGHT GAIN	

The research question we will analyse is to what extent the above mentioned factors typically associated with pre-menstrual and menstrual distress influence concentration. Male subjects serve as controls in the sample. Female subjects could be divided into those using oral contraception (OC) and those using no oral contraception (NOC). The answers to all items had been poled such, that high scores are unfavourable, reflecting high levels of complaints. For either model, we modelled one dimension for the latent state. For the first model, we set $\omega = 1$. For the second model, we set $\mathbf{v} = \mathbf{w} = 0$. For reasons of comparability, we did not optimally rescale the essentially ordinal rating scales of the input variables and/or output variable. For Model I, we used the default option for estimating $\mathbf{z}_0$, that is, we set $\hat{\mathbf{z}}_0 = \mathbf{z}_1$. We modelled one dimension for the latent state.

Model I

The model converged to a solution with a standardised fit of .922. **F** was estimated at .934, meaning that there is a strong influence of the previous state on the present state. Because of possible multicollinearity, we do not interpret the coefficients in **G**, but the correlations of the input variables and output variable with the latent state values, computed over subjects over time points. These are given in Table 2.

Table 2. Overview of Correlations of Input and Output Variables with the Latent State

Input Variables		Output Variable	
ABDOMINAL PAIN	.285	CONCENTRATION	.954
CRAMPS	.330		
PAINFUL BREASTS	.111		
SWELLING	.344		
SKIN DISORDERS	.209		
WEIGHT GAIN	.364		

From Table 2, it appears that the latent state variable strongly resembles the output variable. Thus, we may infer that the latent state variable is best interpreted as reflecting concentration. The correlations of the input variables with this latent state then show that concentration is most strongly influenced by Weight Gain, Swelling, Cramps and Abdominal Pain, these variables constituting the most prominent menstrual distress aspects. Skin Disorders plays a lesser role, and Painful Breasts has a very low correlation, which is not surprising given that most male subjects will have given the lowest possible entries on this variable.

The development of the latent variables over time has been summarised in Figure 3. To facilitate interpretation, we have averaged the latent scores for each of the three groups in the study, i.e. men, women using oral contraception (OC) and women using no oral contraception (NOC). For the women, the latent scores have been synchronised such that for each female subject, the score at day 1 in the figure is the first day of her pre-menstrual phase. As the pre-menstrual phase had been defined as the week before the onset of menstruation, menstruation starts for all women at day 8. As women's cycles were of unequal length, the number of subjects decreases rapidly after day 28, so that we do not give the average latent scores after this day anymore, given that they are bound to become heavily influenced by attrition. For men, scores were synchronised such that for all men, day 1 was the first Monday in the research period, this being the day that was most often shared by the men as the first day they participated in the research.

Looking at the *level* of the curves in Figure 3, we see that women using oral contraception occupy on average the highest level; as the variables were all poled such that high values mean many complaints, this means that women using oral contraception generally report the highest level of complaints regarding concentration. Men assume an intermediate position. Women using no oral contraception report on average the lowest level of complaints. In a sense, this is surprising as one would expect women using no hormonal contraceptives and men to be on a somewhat equal footing, both being in a 'natural state'. However, as 'treatments' (contraception) were not assigned randomly, it is quite likely that the groups that were formed by natural selection differ with regard to other relevant aspects (e.g. the women using oral contraception may more often have had a partner; the women not using oral contraception may have been younger). Because of privacy reasons, demographic information for investigating such presumptions had not been made available.

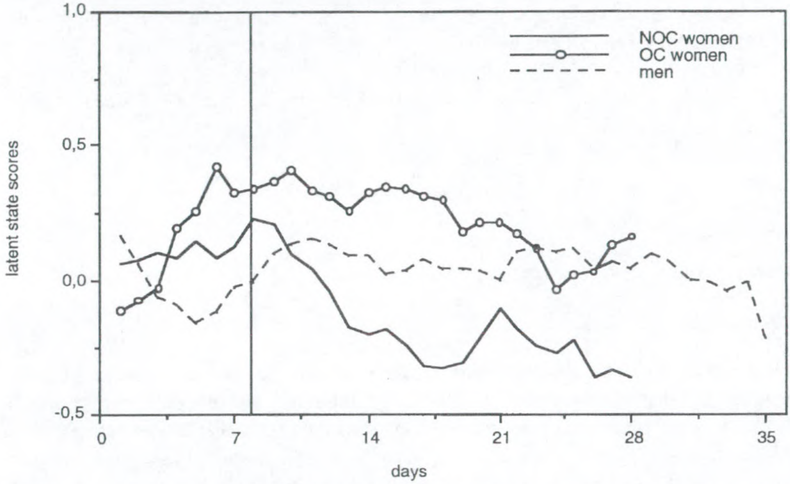

Figure 3. Synchronised latent state scores of men (extending up to day 35), women using oral contraception and women not using oral contraception (both extending up to day 28).

When we inspect the *development* of the average latent curves over time, we see that the women using no oral contraception start out at a high group-specific level of complaints in the pre-menstrual phase. During day 1 and day 2 of the menstrual phase, concentration complaints are at their highest. From day 3 onwards, complaints decrease rapidly, remaining at a low level from day 13 onwards, i.e. after most women will have stopped menstruating. We see an increase and subsequent decrease between days 21 and 25, which might be related to mid-phase distress sometimes reported around the time of ovulation; however, given general levels of variability, it is hard to say whether this is a chance phenomenon or not.

For the women using oral contraception, we see a steady increase of concentration complaints, from low levels at day 1 to the highest group-specific levels at day 6, which is the day the women in the sample - who all used a combination-type of containing pill containing oestrogen and progesterone - terminated their strip of pills. After two days, all women in this group started their menstrual period, this menstruation being a withdrawal bleeding actually. After 6 days, concentration complaints start picking up again from day 14, levelling off in a step-wise fashion, and rising once more from day 25 onward.

For the men, we see something of a swing during the first 7 days of reporting, from higher to lower levels and up again. After day 9, levels remain fairly constant, with some fluctuation. Even though the initial swing is not dramatic in the sense that it spans the range of values exhibited later on in the curve, it is remarkable. It is hard to say what may have caused it; one likely explanation for the initial decrease may be offered by the fact that in menstrual research subjects are well known to report generally higher levels of complaints during the beginning of the investigation period; as the men's average initial values are actually located at the beginning of the investigation period, this may thus emerge much clearer for the men than for the women, the sequencing in whose series has become jumbled because of the synchronisation to menstrual periods.

While the developments of the women in the group using no oral contraception can thus be linked in an interpretable fashion to the menstrual cycle and pre-menstrual and menstrual periods therein, it is hard to associate the fluctuations in the group of women using oral contraception with the menstrual phase or with possible other hormonal changes through the period of pill use. Men can be shown to be variable in their reported levels of concentration as well.

Model II

The model first converged to a degenerate solution. This is a common though by no means sure occurrence in using Model II for $N > 1$ data. Given that the technique is free to estimate $\mathbf{z}_{i0}$, the initial state value for each subject i, it can be seen from Equation (8) that if $\mathbf{z}_{i0}$, $\mathbf{v}$, $\mathbf{F}$ and $\mathbf{G}$ are estimated as:

$$\begin{aligned}
\mathbf{z}_{i0} &\approx \text{average } (\mathbf{z}_{it}) \\
\mathbf{v} &\approx \mathbf{0}, \\
\mathbf{F} &\approx \mathbf{I}, \text{ and} \\
\mathbf{G} &\approx \mathbf{0},
\end{aligned}$$

the model achieved will be

$$\mathbf{z}_{it} \approx \mathbf{1} \,.\, \mathbf{z}_{i0} + \sum_{k=0}^{t-1} \mathbf{1} \,.\, (\mathbf{0}\; \mathbf{x}_{i,t-k} + \mathbf{0}). \qquad (8)$$

and thus:

$$\hat{\mathbf{z}}_{it} \approx \mathbf{z}_{i0}.$$

This, in a meta-sense, provides a sensible approximation to the observed data structure: the latent state scores for each subject are estimated as his or her average output score, this being the simplest model to explain the observed variability in scores. However, as we are interested in intraindividual as well as interindividual change, this is not a satisfactory solution. Given the fact that the dimensions are not necessarily orthogonal, it therefore seems most pragmatic to circumvent the occurrence of this type of solution by replacing the scores $\mathbf{y}_{it}$ by scores $\mathbf{y}^*_{it}$ defined as:

$$\mathbf{y}^*_{it} = \mathbf{y}_{it} - \bar{\mathbf{y}}_{i.}\ ,$$

in words: to center each subject's scores around his or her mean. For forecasting purposes, the standardisation can always be made undone afterwards.

After this standardisation, the model converged to a solution with a standardised fit of .787. It is to be expected that the fit be lower than that in Model I, as Model II is more restrictive. **F** was estimated at .993, meaning that there is now an even stronger influence of the previous state on the present state. Because of possible multicollinearity, also here we interpret the correlations of the input variables and output variable with the latent state values. These have been summarised in Table 3.

Table 3. Overview of Correlations of Input and Output Variables with the Latent State

Input Variables		Output Variable	
ABDOMINAL PAIN	.248	CONCENTRATION	.757
CRAMPS	.299		
PAINFUL BREASTS	.076		
SWELLING	.313		
SKIN DISORDERS	.168		
WEIGHT GAIN	.319		

The pattern of correlations is fairly similar to that arrived at through Model I. The latent state variable again most strongly resembles the output variable, although the correlation coefficient is sizeably lower. This is to be expected as the latent state variable is now restricted to be a subspace of the space spanned by the input and initial state variable. Thus, we infer that the latent state variable is now best interpreted as a mixture of concentration and abdominal pain/water retention complaints. The correlations of the input variables show that this latent state is most strongly influenced by Weight Gain, Swelling, Cramps and Abdominal Pain. Skin Disorders plays a lesser role, and Painful Breasts has again a quite low correlation.

The development of the latent variables over time has been summarised in Figure 4. To facilitate interpretation, we have again averaged the latent scores for each of the three groups in the study, i.e. men, women using oral contraception and women using no oral contraception. The cycles have been synchronised as described above for Model I.

The developments in the curves in Figure 4 are different from those given in Figure 3 for Model I. This is not surprising as the meaning of the latent variable is different now. We see how the From Figure 4 we see that the average development of the men is now very smooth. We also see that again men started at higher levels of concentration complaints. Similarly to the results from Model I, we see that the women using oral contraception again report the highest levels of complaints; they also vary most. (This is irrespective of the sudden jagged development around day 3 which is attributable to one starting value for $\mathbf{z}_0$ that was extremely low and should perhaps be deleted; without the synchronisation of the curves, this odd value would remain at the starting point of the series and thus have much less of a distorting influence.) Similarly to their development in the curve derived from Model I, their levels of complaints increase once more towards the end of the menstrual cycle. The women not using oral contraception show a similar development as they did in the curve derived from Model I, with somewhat lower group-specific levels of complaints in the pre-menstrual phase, towards a peak in the first two days of menstruation, decreasing from day 9 onwards, and somewhat increasing and decreasing once more around days 21-25, again possibly reflecting mid-phase distress around ovulation.

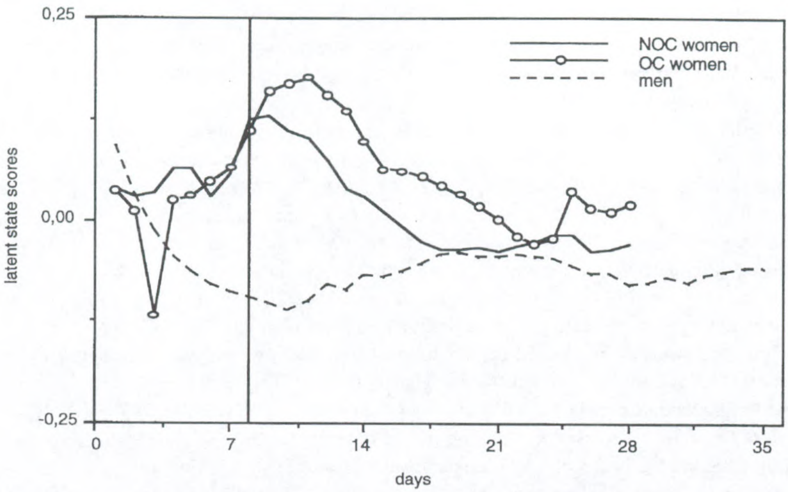

Figure 4. Synchronised latent state scores of men (extending up to day 35), women using oral contraception and women not using oral contraception (both extending up to day 28).

Compared to the curves derived from Model I, women using no oral contraception now do not report the lowest levels of complaints any more. This is not surprising, as in Model I, the latent variable could be interpreted mainly as a concentration variable, whereas for Model II it incorporates menstrual distress to a larger extent. As men will generally report much lower levels on any of these menstrual distress variables, their scores on the latent variable are thus bound to be the lowest now.

DISCUSSION

One of the main advantages over existing methodology of the model proposed by Bijleveld and De Leeuw (1991) is that it can incorporate optimal scaling of nominal and ordinal variables.

Where traditional state space models allow the subsequent dimensions of the latent state to be non-orthogonal, the orthogonality of the latent state variables is thus in principle a restrictive feature of the first model. However, as the general solution is underdetermined, some restrictions are needed anyway; in addition, orthogonality has practical advantages as it simplifies interpretation which is our main focus. A second disadvantage of the first model is that the majorisation algorithm can at times be slow. Then, solutions are affected by standardisation, which means that some types of development (e.g. exponential) cannot be retrieved by the algorithm. Lastly, it is difficult to ensure that $\omega \rightarrow \infty$. This means that solutions are determined by an essentially arbitrary choice of normalisation, although the arbitrariness can be minimised by setting $\omega = 1$ and standardising $\mathbf{X}$ and $\mathbf{Y}$ to zero mean and standard deviation one.

In the second model, solutions for the latent state variables are always in the space of the input variables and the initial state $\mathbf{z}_0$ because of the way in which the model has been formulated. Variables may be left unstandardized. The estimation method is fast and reliable, but if the largest eigenvalue of the state transition matrix $\mathbf{F}$ is larger than 1 or even close to 1, the algorithm may run into numerical problems because of the exponential role of $\mathbf{F}$ in (8). A practical problem is that the nonorthogonality of solutions may make interpretation very hard. The technique has a lot of freedom in the choice of $\mathbf{z}_{i0}$. As our example showed, this may give rise to degenerate solutions, which can be circumvented by standardisation. And although it is a general property of linear dynamic models that latent state values can be interpreted reliably only after stabilisation after a number of time points from the starting point of the series, our example also showed that $\mathbf{z}_{i0}$ may sometimes assume eccentric values. The model should definitely be extended with an option for scaling of non-numerical variables.

Both techniques can be used for situations with $N > 1$, which is an eminently useful feature for a behavioural environment. At first glance, this gain comes at the price of a possibly problematic huge super-matrix $\mathbf{B}_{sup}$ (e.g. for moderate values of $N = 100$ and $T = 20$, $\mathbf{B}_{sup}$ is already 2000×2000). However, by making use of the property that it is very sparsely filled and simply structured, this aspect is of little practical relevance.

A number of practical aspects need to be resolved before either of the two techniques can be applied widely. The first of these is the formulation of guidelines for the choice of dimensionality of the latent state. Contrary to the forecasting focus of classical state space models and time series analysis in general, our main aim is interpretation. Thus, we do not want to find an exhaustive but a parsimonious model, and we are content with approximate, not necessarily exact, solutions. This means also that we want to find only a small, and actually very small, number of dimensions, for most real life applications 3 to 4 at the most. However, in neither of the two approaches do we have criteria to decide on the number of dimensions for $\mathbf{Z}$. As dimensions are not nested, it is hard to apply an elbow-type criterion. In either of the two approaches, the dimensions of $\mathbf{Z}$ are in practice - though not exactly nested - somewhat comparable across solutions with varying dimensionality. One can in practice thus make do with an elbow-type of criterion. However, in the second approach the dimensions are not orthogonal and subsequent dimensions, while rendering the solution increasingly complicated, may add very little in terms of interpretation. Both approaches are thus in need of an objective criterion for choice of dimensionality.

The applicability of both techniques would be enhanced by guidelines for interpretation of the latent state variables. As in many comparable situations in correlational analysis, e.g. canonical correlation analysis, there may be multicollinearity, and instead of looking at the regression weights, we prefer to look at the correlations of the input and output variables with the state. However, when there is more than one subject, there are several ways to compute these correlations. To start with, they can be computed per subject over time points. This gives separate interpretations of the latent state variables for each subject. That would make it hard to compare subjects, but makes it easy to compare time points within subjects. A second method is to compute correlations per time point over subjects. That makes it easy to compare subjects at each respective time point, but comparison over time points is now hard. Thus, for the investigation of development, and thus for many of the areas in which one would be interested if one did linear dynamic systems analysis, this method seems inappropriate. It is most pragmatic (and this is what we did in the example above) to compute correlations over subjects over time points. At the cost of maybe overaveraging, this gives a general solution, where the latent state variables have identical interpretations over time points and over subjects, so that comparisons can be made across subjects as well as within subjects (that is, interindividually as well as intraindividually).

A third practical consideration is the interpretation of subjects' development on the latent state dimensions. While a lot can already be gained from simple visual inspection, it would be nice if objective criteria for interpretation were at hand. One possibility to approach this is by implementing concepts from individual differences modelling (Carroll & Chang, 1970) In that case, the individual latent state scores are decomposed into two components: $\mathbf{z}_t$ and $\mathbf{w}_i$:

$$\mathbf{z}_{it} = \mathbf{w}_i \, \mathbf{z}_t,$$

in which the $\mathbf{z}_t$ stand for the common latent state space, and the weights $\mathbf{w}_i$ stand for the weights each subject attaches to each dimension of the common state space $\mathbf{Z}$. The decomposition can be computed through some optimisation procedure similar to the ones used in individual differences modelling. As the weights reflect the importance subjects attach to the respective dimensions, the weights can be used to look for similarities and differences between subjects. Subjects share the same common solution of latent scores, but they differ in the importance they assign to the respective dimensions. In that case, we could compare subjects' positions in the individual differences space, and subjects' clustering on the dimensions can be interpreted. In all instances, efforts to understand subjects' developments on the latent state variables, will necessarily incorporate some type of simplification, as the wealth of analysis output necessitates this.

Further research is also needed to establish the links of either of the two models with existing methodology for state space modelling. Most notable candidates for comparison are provided by the models as described in Caines (1988), Hannan and Deistler (1988), Aoki (1990). A likely outcome of such a comparison is firstly the non-recursive property of the estimation method for $\mathbf{Z}$ in Model I, the loss criterion that focuses on approximation of the observed series rather than on optimisation of forecasts through the model, and lastly the focuses of either model on low-dimensional approximation, where an overview with all its limitations prevails over the complexity of high-dimensional accuracy.

Both techniques should be supplied with features to provide information on the reliability of the estimates, e.g. the correlation coefficients. At present, in both approaches, the data are approximated and no distributional information is taken along. Given the ordered character of the values $\mathbf{z}_t$, the bootstrap-type moving blocks approach (Efron & Tibshirani, Chapter 8, 1993; Shao & Tu, Chapter 9, 1995) is an option for obtaining information on the reliability in time of the estimates. This approach cuts out 'chunks' of the time series, treating them as new subjects. The chunks, however, should be relatively long, which means that also the original time series must be pretty long indeed, and maybe very long for behavioural practice. If the series are shorter, another option is to randomly leave out measurements, estimating them on the basis of the model. The first model appears to be more suited for this given the possibility for least squares optimal estimation of missing values. A second possibility is to resample the residuals (Efron & Tibshirani, Chapter 8, 1993; Shao & Tu, Chapter 9, 1995). However, many time series models for behavioural data are non-stationary by nature, as in behavioural applications we are usually investigating developmental processes in which there is growth of some kind. Any changes in the sequence of measurements give an unnatural and ill-fitting disruption of the flow of the series. Of course, bootstrapping over subjects is always possible when $N > 1$. This appears to be definitely one natural choice, as we are given our paradigm, definitely interested in stability over the independent replications.

This leads the way to a quite useful feature that the two models share. Where cross-sectional methods derive stability from a suitably large number of replications over subjects, and where time series models derive stability from a suitably large number of replications over time, the two methods discussed here analyse over time points as well as over subjects. Thus, every subject adds T additional measurements to the data, or conversely: every time point adds N additional measurements to the data. Contrary to the rivalling - essentially cross-sectional - methods like structural equations modelling that derive stability from sufficient (and for time series applications usually unrealistically large) numbers of replications over subjects, it is likely that under hopefully mild conditions the essentially three-way methods proposed here can derive sufficient stability of the estimates at lower N and/or T than conventional time series or cross-sectional two way methods. An other way to say this is to say that N and T boost each other in terms of stability. In that case, the methods discussed here will be applicable for a much wider range of research situations, requiring fewer replications over subjects for a certain number of time points. In practice, this will mean that they will be an especially useful extension to the existing methodology for longer series obtained for smaller numbers of subjects.

REFERENCES

Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion. In: R. Mehra and D.G. Lainotis (Eds.). *Advances and case studies in system identification.* New York: Academic Press, 27-96.

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

Bijleveld, C.C.J.H., Bogaards, M., Bredero, A. et al. (1980). *Het Rooie Kruis: een verslag over menstruatie*. Vakgroep Klinische Psychologie, Rijksuniversiteit Leiden.

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

Bijleveld, C.C.J.H. and Legendre, P. (1993). Spatio-temporal regression analysis using state space models. In: R. Steyer, K.F. Wender & K.F. Widaman (Eds.), *Proceedings of the 7th European Meeting of the Psychometric Society in Trier*, Stuttgart, New York: Gustav Fisher Verlag, pp. 27-32.

Bijleveld, F.D., Bijleveld, C.C.J.H., De Leeuw, J. and Oppe, S. (1991). Least squares optimization of linear dynamic models using majorization, quasi-Newton and multiplier methods. Los Angeles, *UCLA Statistics Series #84*.

Bijleveld, C.C.J.H. and Mooijaart, A. (1993). Structural equations modelling of state space models. Leiden: Dept. of Psychometrics, PRM-93-03.

Bijleveld, F.D., Bijleveld, C.C.J.H. and De Leeuw, J. (1994). On least squares optimization of linear dynamic systems. Los Angeles, *UCLA Statistics Series #145*.

Bijleveld, C.C.J.H. and van der Burg, E. (1997). Exploratory multivariate analysis of categorical longitudinal data using optimal scaling techniques. In: Van der Kamp, L.J. Th. et al. (Eds.). *Longitudinal data analysis for the behavioral sciences*. London: Sage, in press.

Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis: forecasting and control*. San Francisco: Holden-Day.

Caines, P.E. (1988). *Linear stochastic systems*. New York: Wiley.

Carroll, J.D. and Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika, 35*, 283-319.

De Leeuw, J. (1988). Convergence of the majorization method for multidimensional scaling. *Journal of Classification, 5*, 163-180.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Fletcher, R. (1981). *Practical methods of optimization*. Vol I. Chichester: Wiley.

Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester: Wiley.

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

Ho, B.L. and Kalman, R.E. (1966). Effective construction of linear state variable models from input/output functions. *Reglungstechnik, 12*, 545-592.

Jöreskog, K.G. and Goldberger, A.S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*, 631-639.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems *ASME Transactions, Series D, Journal of Basic Engineering, 82*, 35-45.

Ljung, L. (1987). *System identification. Theory for the user*. Englewood cliffs, NJ: Prentice-Hall.

MacCallum, R. and Ashby, F.G. (1986). Relationships between linear systems theory and covariance structure modelling. *Journal of Mathematical Psychology, 30*, 1-27.

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.

Moos, R.H. (1968). The development of a Menstrual Distress Questionnaire. *Psychosomatic Medicine, 30*, 853-867.

Murray, W. (1969). *An algorithm for constrained minimization*. London: Academic Press.

Otter, P.W. (1986). Dynamic structural systems under indirect observation: Identifiability and estimation aspects from a system theoretic perspective. *Psychometrika, 51*, 415-428.

Pothoff, R.F. and Roy, S.N. (1964). A generalized multivariatre analysis of variance model useful especially for growth curve problems. *Biometrika, 51*, 313-326.

Shao, J. and Tu, D. (1995). *The jackknife and bootstrap*. New York: Springer.

Van den Boogaard, T.G.H.M. and Bijleveld, C.C.J.H. (1988). Daily menstrual symptoms measures in women and men using an extended version of Moos's instrument. *Journal of Psychosomatic Obstetrics and Gynaecology, 9*, 103-110.

Van den Wollenberg, A.L. (1977). Redundancy analysis: an alternative for canonical correlation analysis. *Psychometrika, 42*, 207-219.

Van de Pol, F. and Langeheine, R. (1989). Mixed Markov latent class models. In: C.C.Clogg (Ed.). *Sociological methodology*. Oxford: American Sociological Association, pp. 213-247.

Van der Leeden, M. (1997). Multilevel analysis multivariate analysis of categorical longitudinal data using optimal scaling techniques. In: Van der Kamp, L.J. Th. *et al.* (Eds.). *Longitudinal data analysis for the behavioural sciences*. London: Sage publications, in press.

Verboon, P. and Heiser, W.J. (1990). *Some possibilities for the analysis of dynamic three-way data*. Leiden: Department of Data Theory RR 90-01.

ontvangen: 4-12-95
geaccepteerd: 17-12-96