Faculteit der Economische Wetenschappen en Econometric

Serie Research Memoranda

Dynamic Analysis of Multivariate Panel Data with Nonlinear Transformations

Kees van Montfort
Catrien Bijleveld

Research Memorandum 1997-54

October 1997

vrije Universiteit amsterdam

DYNAMIC ANALYSIS OF MULTIVARIATE PANEL DATA WITH NONLINEAR TRANSFORMATIONS

KEES VAN MONTFORT
DEPT. ECONOMETRICS
FREE UNIVERSITY
AMSTERDAM
e-mail KVMONTFORT@ECON.VU.NL

CATRIEN BIJLEVELD
UNIT FOR PSYCHOMETRICS & RESEARCH METHODOLOGY
DEPARTMENT OF PSYCHOLOGY
LEIDEN UNIVERSITY
e-mail BYLEVELD@RULFSW.LeidenUniv.NL

ABSTRACT

Many models for multivariate data analysis can be seen as special cases of the linear dynamic or state space model. Contrary to the classical approach to linear dynamic systems analysis, the model presented here is developed from the social science framework of approximation, data reduction and interpretation, where generalization is required not only over time points but over subjects as well. Borrowing concepts from the theory on mixture distributions, the linear dynamic model can be viewed as a multilayered regression model, in which the output variables are imprecise manifestations of an unobserved continuous process. An additional layer of mixing makes it possible to incorporate non-normal as well as ordinal variables. Using the EM-algorithm, we find estimates of the unknown model parameters, simultaneously providing stability estimates. We illustrate the applicability of the obtained procedure through an empirical example.

INTRODUCTION

Many models for multivariate data analysis can be seen as special cases of a number of general models. One such general model is the linear dynamic model, also referred to as the state space model or linear dynamic system (Ho & Kalman, 1966; Ljung, 1987; Hannan & Deistler, 1988; Aoki, 1990). The linear dynamic model specifies relations between a set of input or exogenous variables, and a set of time-dependent output or endogenous variables. The relations between input and output are channeled through latent variables. The space spanned by the latent variables splits the input and output: given this latent variables space, input and output are independent. The latent variables thus accommodate the dependence of the output on the input. At the same time, the latent variables accommodate the dependence of present measurements on past measurements. The latent state variables follow a Markov chain (although the observed output process may be much more complicated). Thus, the latent states at any time point t depend on the latent states at time point $t-1$ only. As such, the space spanned by the latent variables also splits the past and the future: given the present state, past and future are independent.

In state space analysis, the dimensionality of the state space may be high and may in fact be a lot higher than that of the dimension of the input and output. This is necessary to be able to capture the dynamic mechanisms that generate the time-dependent process, especially cyclical processes. In many common applications of state space analysis, such as are to be found in engineering or process control, the technique has been geared to find high-dimensional exact solutions. This is so because the scientific paradigm in these branches of science is directed strongly towards accurate forecasting. This paradigm prevails as well in time series analysis, which models autoregressive and moving average processses for long chains (in practice: more than 50 time points) gathered for one observation unit (Box & Jenkins, 1976). Time series analysis has strong theoretical links with state space analysis (Akaike, 1976).

Our behavioral paradigm orients us towards data reduction and approximation, description and interpretation. We will therefore strive to find low-dimensional and approximate, rather than high-dimensional and exact, solutions. For our purposes the state space has, preferably, lower dimensionality than the set of combined input and output variables. The latent variables thus reflect the notion common to factor analysis and related techniques, of a latent condition, such as a trait or ability. To distinguish our technique from the classical approach to systems analysis, we have labeled it linear dynamic analysis. A second distinction of our models to the usual application of state space analysis is that we will develop our model specifically for situations where not one, but several observation units have been measured. This is a necessary reflection of common research orientation, in which generalizability should be attained not only over timepoints, but over subjects as well. A third distinction of our model is that it can handle non-normal and ordinal measurements, which are a common occurrence in day-to-day social research practice.

The linear dynamic model can be viewed as a longitudinal extension of the well-known MIMIC model (Jöreskog & Goldberger, 1975). Each new time point links a new MIMIC model to a chain of previous MIMIC models: the linkage point is the latent state. Structural equations models may thus seem a likely candidate for analysing the type of models we are interested in, see MacCallum and Ashby (1986) and Oud, van den Bercken and Essers (1990) for examples. However, for data with a large number of replications in time (and in some cases with relatively small numbers of subjects), such techniques may be inefficient or even impossible to use. Larger numbers of time points can lead to increasingly unstable solutions and even negative estimates of variance of the error terms or disturbances. To counteract this, very large numbers of subjects would be needed. However, in social science practice the number of subjects tends to be smaller when the number of time points is large. In addition, we would like to be able to incorporate non-normal and ordinal variables in our models. Structural equations modelling may use the so-called Asymptotic Distribution Free (ADF) method to give parameter estimates for non-normally distributed variables (that may be seen as a special type of non-numerical variables), but this method uses 8-th order moments and thus needs extremely large numbers of replications over subjects, a highly uncommon occurrence in social research situations with large numbers of time points.

Recent years have seen a number of more particular applications of structural equations modelling to longitudinal data. The most notable of these are the models proposed for growth curve analysis (Rogosa & Willett, 1985; McArdle & Hamagami, 1991; Willett & Sayer, 1994; McArdle & Hamagami, 1996) that are close to random coefficient models (Bryk & Raudenbusch, 1987). Muthén (1996) presented a number of options to apply such models to data with binary outcome variables. Browne and du Toit (1991) presented other related models for growth data, notably one that incorporates concepts from ARMA modeling. These models all investigate situations in which growth on a continuous resp. dichotomous outcome variable is reduced to a number of parameters that describe the curve that can be fitted through each subject's scores. These parameters are then related to background characteristics of the subjects, such as age, gender, experimental condition and the like. This type of model is indeed hierarchical in the sense that it first reduces the longitudinal character of the data to a small number of essentially cross-sectional characteristics, such as intercept and slope, after which these are related to other cross-sectional characteristics.

Using lagged versions of variables (Molenaar, 1985; Molenaar, de Gooijer & Schmitz, 1992; Browne, 1992; Van Buuren, 1997), ARMA-type modeling can be approximated through structural equations modelling. In this manner, a great many complications are induced, however, as the assumption of independence of sample elements is violated. The issues in analysing time series data using structural equations modelling are discussed in Hershberger, Molenaar and Cornea1 (1996). In a related field, Molenaar and others (Molenaar, 1985; Molenaar ***et al.*** 1992) proposed various possibilities for the analysis of the dynamic factor model, which can be seen as a special case without input of the state space model. However, our data analytic framework includes situations in which assessment of the impact of external influences or input is the explicit research objective. In addition, we prefer to restrict ourselves to situations in which the system parameters are time-invariant, even if the developments on the latent and outcome variables themselves need not be so.

The latent Markov models proposed by Langeheine, van de Pol and others (van de Pol & De Leeuw, 1986; van de Pol & Langeheine, 1989; Langeheine & van de Pol, 1990) are appropriate for analysing research questions in which the subjects' longitudinal responses on a number of categorical outcome variables are assumed to depend upon a number of latent categorical variables. The latent categorical variables follow a Markov chain, i.e., a particular type of ARMA model. In latent transition analysis, subjects' answering profiles on a number of categorical indicators are used to analyse stage sequential models of development (Collins & Wugalter, 1992). None of these models have been equipped to incorporate the influence of exogenous variables. Recently, Mooijaart and van Montfort (1997) proposed such a state space model for categorical variables as an adaptation of the latent class model. While their model appears efficient and easy to apply, it is again limited in the sense that it is appropriate only for datasets in which all variables are categorical.

The linear dynamic model we will focus on can be seen as an extension of the DYNAMALS method proposed by Bijleveld and De Leeuw (1991), which was developed to analyse long non-numerical series gathered for one subject. In fact, Bijleveld and others (Bijleveld & Legendre, 1993; Bijleveld & Bijleveld, 1997) proposed to extend the DYNAMALS model to the analysis of long chains gathered for several subjects. However, neither the $N=1$ nor the $N>1$ DYNAMALS model provides stability information.

We will build from the general class of models proposed by De Leeuw, Bijleveld, van Montfort and Bijleveld (1997). Working from the concept of mixture distributions, they proposed to view the state space model as a multilayered regression model. The discrete output variables are then viewed as an imprecise manifestation of an unobserved continuous process, i.e. the latent state variables. A second layer of mixing makes it possible to obtain transformations of any numerical or non-numerical output variables, using for instance a Box-Cox transformation or some other useful type of transformation. As we are approximating not only the expected value of the observed data, but also its distributional aspects, we will be able to test the significance of regression coefficients. The idea of transforming non-numerical or non-normally distributed variables to normally distributed ones in a time series context is not new: Smith and Brunsdon (1989) proposed the transformation of multinomially distributed variables to normally distributed variables for the class of ARMA models, that is subsumed by our broader framework.

In the following, we will first discuss the dynamic model. Next we will deal with the derivation of maximum likelihood estimators for fitting the model, as well as our optimalization procedure. We will subsequently describe an empirical example.

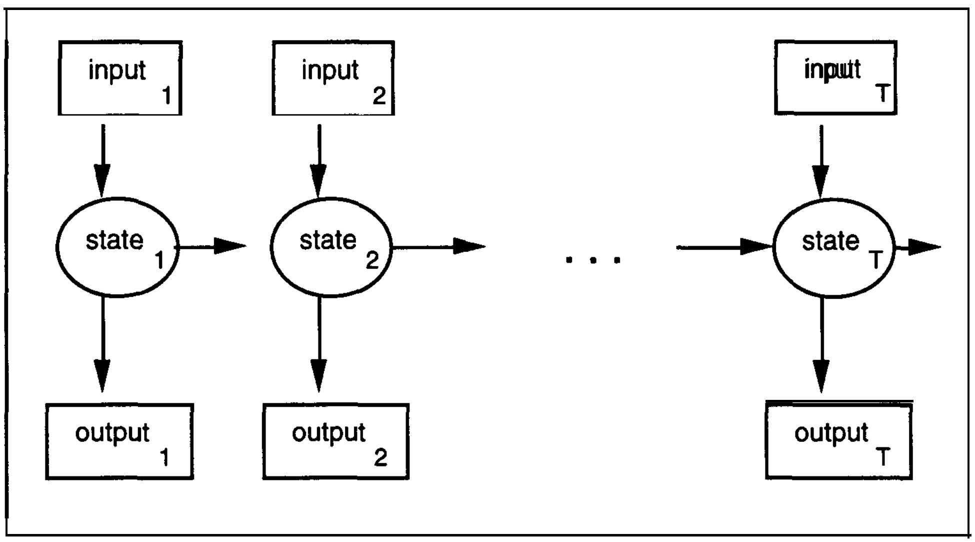

Figure 1. *Schematic Representation of the Linear Dynamic Model*

THE DYNAMIC MODEL

Suppose that, for one subject, we have measured at a total of T time points a number of input variables and a number of output variables. The output variables are time-dependent, the input variables may be time-dependent or time-independent. The impact of the input on the output is mediated by unobserved latent variables. The impact of previous measurements on future measurements is mediated by the same unobserved latent variables. A schematic representation of the model is given in Figure 1.

For each time point, we refer to the number of input variables as k, to the number of output variables as m, and to the number of latent variables as p. We assume that the transition matrices specifying the relations between the various components of the model are time-invariant, that is, we assume that the relations between the input, output and latent state variables are independent of time. We can then specify the relations between the input, output and latent state variables as follows. The p latent state space scores at any time point t depend on those of the former time point $t-1$, weighted by a transition matrix. As, however, the latent state space scores depend as well on a weighted contribution of the input variables at time point t, this means that the latent state scores should be defined as follows:

$$\mathbf{z}_t = \mathbf{F}\,\mathbf{z}_{t-1} + \mathbf{G}\,\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \tag{1a}$$

with $\mathbf{z}_t$ and $\mathbf{z}_{t-1}$ the vectors containing the p latent state space scores at time points t and $t-1$ respectively, F a $(p \times p)$ matrix of regression coefficients, $\mathbf{x}_t$ the k-dimensional vector of input variables at time point t, G a $(p \times k)$ matrix of regression coefficients, and $\boldsymbol{\varepsilon}_t$ a p-dimensional vector of disturbances.

Similarly, the output variables at any time point t are predicted from a weighted sum of the latent state space scores at that time point, which is in formula written as:

$$\mathbf{y}_t = \mathbf{H}\,\mathbf{z}_t + \boldsymbol{\delta}_t, \tag{1b}$$

with $\mathbf{y}_t$ the m-dimensional vector of output variables at time point t, $\mathbf{H}$ an $(m \times p)$ matrix of regression coefficients, and $\boldsymbol{\delta}_t$ an m-dimensional vector of measurement errors. The vectors with error terms are needed because we do not expect a perfect fit to real data. The system needs a hypothetical starting point at $t = 0$, $\mathbf{z}_0$. Note how $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{H}$ are indeed identical at all timepoints, i.e. $\mathbf{F}_t \equiv \mathbf{F}$, $\mathbf{G}_t \equiv \mathbf{G}$, and $\mathbf{H}_t \equiv \mathbf{H}$.

Equations (la) and (lb) together describe the linear dynamic model; equation (la) is often referred to as the ***system equation;*** equation (lb) as the ***measurement equation.*** We will refer to the transition matrix **F** as the ***state transition matrix,*** to G as the ***control matrix,*** and to **H** as the ***measurement matrix.***

When we have observed measurements for N subjects, we have N models such as in Model (1). We can write a combined model for all subjects and all time points as:

$$\mathbf{z}_{i,t} = \mathbf{F}\,\mathbf{z}_{i,t-1} + \mathbf{G}\,\mathbf{x}_{i,t} + \boldsymbol{\varepsilon}_{i,t}, \tag{2a}$$

$$\mathbf{y}_{i,t} = \mathbf{H}\,\mathbf{z}_{i,t} + \boldsymbol{\delta}_{i,t}, \tag{2b}$$

where $\mathbf{z}_{i,t}$ is the latent state variable for subject i at time point t, and $\mathbf{x}_{i,t}$, $\boldsymbol{\varepsilon}_{i,t}$, $\mathbf{y}_{i,t}$ and $\boldsymbol{\delta}_{i,t}$ are defined correspondingly.

Note how F, G and H are identical across subjects as well, i.e. $\mathbf{F}_i \equiv$ F, $\mathbf{G}_i \equiv$ G, and $\mathbf{H}_i \equiv$ H.

Until now we did not make any assumptions about the distribution of the random variables. Following De Leeuw *et al.* (1997) we now assume that the observed $\mathbf{y}_{i,t}$ are a function of an m-dimensional set of unobserved normally distributed latent variables $\boldsymbol{\eta}_{i,t}$. Model (2) then becomes:

$$\mathbf{z}_{i,t} = \mathbf{F}\,\mathbf{z}_{i,t-1} + \mathbf{G}\,\mathbf{x}_{i,t} + \boldsymbol{\varepsilon}_{i,t}, \tag{3a}$$

$$\boldsymbol{\eta}_{i,t} = \mathbf{H}\,\mathbf{z}_{i,t} + \boldsymbol{\delta}_{i,t}, \tag{3b}$$

$$\mathbf{y}_{i,t} = F_{\alpha}^{-1}(\boldsymbol{\eta}_{i,t}), \tag{3c}$$

where F_{α} is some kind of transformation depending on the vector a of unknown transformation parameters. Thus, $\boldsymbol{\eta}_{i,t}$ can only be observed indirectly through $\mathbf{y}_{i,t}$. Note that we set $F_{\alpha_{i\,t}} = F_{\alpha}$, i.e. F_{α} is the same for all subjects at all timepoints.

Figure 2 gives a schematic representation of model (3) for one subject. For N subjects, we have N such models stacked on top of one another.

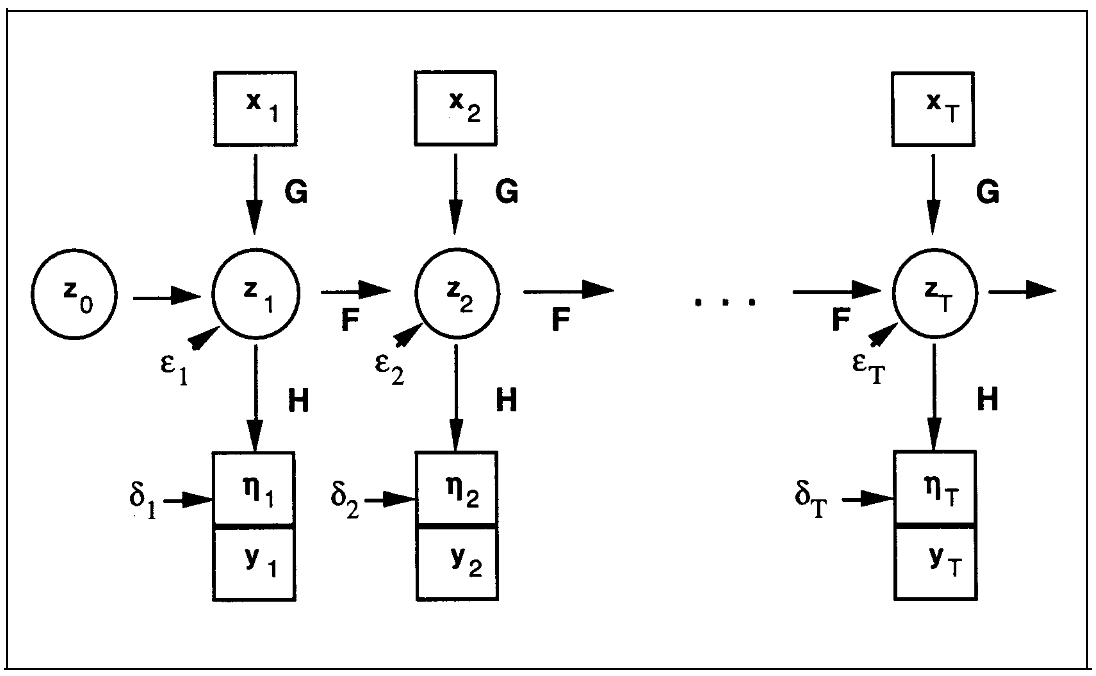

Figure 2. *Schematic Representation* ***of*** *Linear Dynamic Model with Transformation* ***of*** *the Output Variables*

In addition to the equations (3a,3b), we assume that

- $\mathbf{x}_{i,t}$, $\mathbf{y}_{i,t}$ $(i = 1, \ldots, N; t = 1, \ldots, T)$ are observed variables;
- $\mathbf{z}_{i,t}$ $(i = 1, \ldots, N; t = 1, \ldots, T)$ is a latent variable;
- $\boldsymbol{\eta}_{i,t}$ $(i = 1, \ldots, N; t = 1, \ldots, T)$ is observed indirectly through $\mathbf{y}_{i,t}$ and F;
- $\mathbf{x}_{i,t}$, $\mathbf{z}_{i,t}$ and $\boldsymbol{\eta}_{i,t}$ $(i = 1, \ldots, N; t = 1, \ldots, T)$ are normally distributed variables;
- $\boldsymbol{\varepsilon}_{i,t} \perp \mathbf{z}_{i,t-1}$; $\boldsymbol{\varepsilon}_{i,t} \perp \mathbf{x}_{i,t}$;
- $\boldsymbol{\delta}_{i,t} \perp \mathbf{z}_{i,t}$; $\boldsymbol{\delta}_{i,t} \perp \mathbf{x}_{i,t}$;
- $E(\boldsymbol{\varepsilon}_{i,t}) = E(\boldsymbol{\delta}_{i,t}) = 0$ for each $i = 1, \ldots, N; t = 1, \ldots, T$;
- the $\boldsymbol{\varepsilon}_{i,t}$ are homoscedastic over persons, but may be heteroscedastic over time points, i.e. $\mathbf{V}(\boldsymbol{\varepsilon}_{i,t}) = \boldsymbol{\Theta}_t$;
- the $\boldsymbol{\delta}_{i,t}$ are homoscedastic over persons, but may be heteroscedastic over time points, i.e. $\mathbf{V}(\boldsymbol{\delta}_{i,t}) = \boldsymbol{\Psi}_t$.

Thus, the observed random output variables $\mathbf{y}_{i,t}$ are transformed to unobserved normally distributed output variables $\boldsymbol{\eta}_{i,t}$. No assumptions are made on the distribution of the $\mathbf{y}_{i,t}$. In principle, the same could be done for the input variables. However, for reasons of simplicity and overview, we will restrict ourselves in the next sections to the case where the input variables have been completely observed.

FITTING THE LINEAR DYNAMIC MODEL

Before deriving a procedure for fitting the linear dynamic model, we first introduce some simplifying assumptions. This causes no loss of generality as the procedure we present can be derived for the more complicated case in a straightforward manner.

We assume that the dimensionality of z is 1 and that $z_{i,0} = 0$. In that case the latent state vector for subject i at time point t reduces to the scalar $z_{i,t}$, the error vector $\boldsymbol{\varepsilon}_{i,t}$ for subject i at time point t reduces to the scalar $\varepsilon_{i,t}$, and the state transition matrix F reduces to a scalar, that we will in the following write as 'f. Equations (3a) and (3b) can be rewritten as:

$$z_{i,t} = f\, z_{i,t-1} + \mathbf{g}'\mathbf{x}_{i,t} + \varepsilon_{i,t} = \sum_{s=1}^{t} p_{ts}\mathbf{g}'\mathbf{x}_{i,s} + \sum_{s=1}^{t} p_{ts}\varepsilon_{i,s} \,, \tag{4a}$$

with $p_{ts} = f^{t-s}$, and with g the k-dimensional vector containing the elements of the $(1 \times k)$ control matrix G, and

$$\boldsymbol{\eta}_{i,t} = \mathbf{h}\, z_{i,t} + \boldsymbol{\delta}_{i,t} = \sum_{s=1}^{t} \mathbf{h}\, p_{ts}\mathbf{g}'\mathbf{x}_{i,s} + \sum_{s=1}^{t} \mathbf{h}\, p_{ts}\varepsilon_{i,s} + \boldsymbol{\delta}_{i,t} \,, \tag{4b}$$

with h the m-dimensional vector containing the elements of the $(m \times 1)$ measurement matrix H.

The variance covariance matrices of the error terms in (4a) and (4b) can be written as:

$$\mathbf{V}\Big(\sum_{s=1}^{t} p_{ts}\varepsilon_{i,s}\Big) = \Theta_{*t}; \tag{4c}$$

$$\mathbf{V}\Big(\sum_{s=1}^{t} \mathbf{h}' p_{ts}\varepsilon_{i,s} + \boldsymbol{\delta}_{i,s}\Big) = \Psi_{*t}\,, \tag{4d}$$

where also Θ_{*t} and Ψ_{*t} are equal for all *i*. (Note that also Θ_{*t} is now a scalar.)
We switch to matrix notation, defining:

$$\mathbf{B} \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ f & 1 & 0 & 0 & \cdots & 0 & 0 \\ f^2 & f & 1 & 0 & \cdots & 0 & 0 \\ f^3 & f^2 & f & 1 & \cdots & 0 & 0 \\ . & . & . & . & \cdots & . & . \\ . & . & . & . & & . & . \\ f^{T-1} & f^{T-2} & f^{T-3} & f^{T-4} & \cdots & f & 1 \end{bmatrix}, \quad \boldsymbol{\varepsilon}_i \equiv \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ . \\ \varepsilon_{iT} \end{bmatrix},$$

with $\boldsymbol{\varepsilon}_i$ a T - dimensional vector and cov $\boldsymbol{\varepsilon}_i \equiv \boldsymbol{\Theta}$ (with $\boldsymbol{\Theta}$ a $T \times T$ diagonal matrix).
Next, the T - dimensional vector $\mathbf{z}_i$ is defined as:

$$\mathbf{z}_i \equiv \begin{bmatrix} z_{i1} \\ z_{i2} \\ z_{i3} \\ z_{i4} \\ \\ z_{iT} \end{bmatrix}, \text{ and the } kT \text{ - dimensional vector } \mathbf{x}_i \text{ as: } \mathbf{x}_i \equiv \begin{bmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \\ \mathbf{x}_{i3} \\ \mathbf{x}_{i4} \\ . \\ \mathbf{x}_{iT} \end{bmatrix}.$$

The mT - dimensional vectors $\boldsymbol{\delta}_i$ and $\boldsymbol{\eta}_i$ are defined as:

$$\boldsymbol{\delta}_i \equiv \begin{bmatrix} \boldsymbol{\delta}_{i1} \\ \boldsymbol{\delta}_{i2} \\ \boldsymbol{\delta}_{i3} \\ \boldsymbol{\delta}_{i4} \\ \\ \boldsymbol{\delta}_{iT} \end{bmatrix}, \qquad \boldsymbol{\eta}_i \equiv \begin{bmatrix} \boldsymbol{\eta}_{i1} \\ \boldsymbol{\eta}_{i2} \\ \boldsymbol{\eta}_{i3} \\ \boldsymbol{\eta}_{i4} \\ . \\ . \\ \boldsymbol{\eta}_{iT} \end{bmatrix}$$

respectively, and cov $\delta_i \equiv \Psi$(with Ψ an $mT \times mT$ (block)diagonal matrix containing the Ψ_t for the various time points)

Then we can write (4a), (4b), (4c) and (4d) in matrix notation as:

$$z_i = (B \otimes g')\, x_i + B\varepsilon_i ,$$

$$\eta_i = (B \otimes g')\, x_i \otimes h + (B\varepsilon_i)\ 6\ h + \delta_i ,$$

$$\Theta_* = V\,(B\varepsilon_i) = B\Theta B' ,$$

$$\Psi_* = V\,((B\varepsilon_i) \otimes H + \delta_i) = (BQB') \otimes (hh') + \Psi ,$$

Θ_* thereby being a function of f and Θ, with Θ a $T \times T$ (block)diagonal matrix containing the Θ_t for the various time points, and Ψ_* thereby being a function of f, Θ and Ψ, with Ydefined as above.

It follows that, after stacking the η_{it} for all timepoints into a $(T \times m)$ matrix η_i, we may decompose the complete likelihood into two parts:

$$L = L_1\,(h,\Psi,F_\alpha,Y,Z,\eta).\ L_2(f,\ \Theta,g,z,X),$$

where Y, **Z** and X contain the realizations for the output, latent state and input variables for all subjects at all timepoints.

$$L_1 = (2\pi)^{-N/2} .\ I\ B\Theta B' I^{-N/2}.$$

$$\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{N}(z_i - (B \otimes g')\ Xi)\right]' (B\Theta B')^{-1} \left[\sum_{i=1}^{N}(z_i - (B \otimes g')\ x_i)\right]\right\} ;(5a)$$

$$L_2 = (2\pi)^{-N/2} .\ I\ (BOB') \otimes (hh') + \Psi I^{-N/2}.$$

$$\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{N}(F_\alpha\,(y_i)-((B\otimes g')x_i)\otimes h)\right]' \left[\ (B\Theta B')\otimes(hh')+\Psi\right]^{-1}\right.$$

$$\left.\left[\sum_{i=1}^{N}(F_\alpha\,(y_i)-((B\otimes g')x_i)\otimes h)\right]\right\} . \qquad (5b)$$

As the complete likelihood function can be decomposed into two simpler parts, this opens up the possibility to optimize the two parts separately.

ALGORITHM

A natural class of algorithms for maximizing L is the Expectation-Maximization (EM) algorithm (Dempster, Laird & Rubin, 1977). The EM-algorithm consists of two alternating steps: the E(xpectation)-step and the M(aximization)-step. In the E-step of the EM-algorithm the expectation of the sufficient statistics of the conditional distribution has to be formulated. For our case, these sufficient statistics are defined as:

$$\mathbf{m}_{xx} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i \, x_i' \; ;$$

$$\mathbf{m}_{\eta\eta} \equiv \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\eta}_i \, \boldsymbol{\eta}_i' \; ;$$

$$\mathbf{m}_{\eta x} \equiv \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\eta}_i \, \mathbf{x}_i' \; ;$$

$$\mathbf{m}_{zz} \equiv \frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_i \, \mathbf{z}_i' \; ;$$

$$\mathbf{m}_{xz} \equiv \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \, \mathbf{z}_i' \; ;$$

$$\mathbf{m}_{\eta z} \equiv \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\eta}_i \, \mathbf{z}_i' \; .$$

We can formulate the expectations of these matrices, conditional on x and $\boldsymbol{\eta}$ as functions of sample moments $\hat{m}_{xx}$, $\hat{\mathbf{m}}_{\eta x}$ and $\hat{\mathbf{m}}_{\eta\eta}$:

$$m^*_{xx} \equiv \frac{1}{N} E\left(\sum_{i=1}^{N} \mathbf{x}_i \, x_i' \mid \mathbf{x}_i, \boldsymbol{\eta}_i \right) = \hat{\mathbf{m}}_{xx},$$

$$\mathbf{m}^*_{\eta\eta} \equiv \frac{1}{N} E\left(\sum_{i=1}^{N} \boldsymbol{\eta}_i \, \boldsymbol{\eta}_i' \mid \mathbf{x}_i, \boldsymbol{\eta}_i \right) = \hat{\mathbf{m}}_{\eta\eta};$$

$$\mathbf{m}^*_{\eta x} \equiv \frac{1}{N} E\left(\sum_{i=1}^{N} \boldsymbol{\eta}_i \, \mathbf{x}_i' \mid \mathbf{x}_i, \boldsymbol{\eta}_i \right) = \hat{\mathbf{m}}_{\eta x};$$

showing that $\mathbf{m}^*_{xx}$, $\mathbf{m}^*_{\eta\eta}$ and $\mathbf{m}^*_{\eta x}$ need to be estimated once only in the optimalization procedure;

$$\mathbf{m}^*_{zz} \equiv \frac{1}{N} E\left(\sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i' \mid \mathbf{x}_i, \boldsymbol{\eta}_i\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \operatorname{cov}(\mathbf{z}_i \mid \mathbf{x}_i, \boldsymbol{\eta}_i) + \frac{1}{N}\sum_{i=1}^{N} \{E(\mathbf{z}_i \mid \mathbf{x}_i, \boldsymbol{\eta}_i)\}^2$$

$$= (\mathbf{B} \otimes \mathbf{g}')\, \mathbf{m}^*_{xx}\, (\mathbf{B} \otimes \mathbf{g}')' + \mathbf{B}\Theta\mathbf{B}'$$

$$- \; E\left(\mathbf{z}_i \begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}'\right) \left[E\begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}\begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}'\right]^{-1} E\left(\mathbf{z}_i \begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}'\right)$$

$$+ \; \left\{\frac{1}{N}\sum_{i=1}^{N} E\left(\mathbf{z}_i \begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}'\right) \left[E\begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}\begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}'\right]^{-1} \begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}\right\}^2.$$

which reduces asymptotically for large N to its first part only.

$$\mathbf{m}^*_{\binom{x}{\eta} z} \equiv \frac{1}{N} E\left(\sum_{i=1} \begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix} \mathbf{z}_i' \mid \mathbf{x}_i, \boldsymbol{\eta}_i\right) =$$

$$= \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix} E'(\mathbf{z}_i \mid \mathbf{x}_i, \boldsymbol{\eta}_i) =$$

$$= \frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix} \left\{E\left(\mathbf{z}_i \begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}'\right) \left[E\begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}\begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}'\right]^{-1} \begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix}\right\}',$$

which reduces asymptotically for large N to:

$$E\left(\begin{pmatrix}\mathbf{x}_i\\ \boldsymbol{\eta}_i\end{pmatrix} \mathbf{z}_i'\right),$$

with i ($i = 1, \ldots, N$) some arbitrary integer, and $E(\mathbf{x}_i\, \mathbf{z}_i') = \hat{\mathbf{m}}_{xx}\, (\mathbf{B} \otimes \mathbf{g}')$. Note that the conditional likelihood (5b), given the realizations of x and $\boldsymbol{\eta}$, is a function of the unknown model and transformation parameters, and of a number of observed sample moments. Because $\boldsymbol{\eta}_{it} = F_\alpha(\mathbf{y}_{it})$ for every i and t, the sample moments are thus also a function of the transformation parameters a.

In the second step of the EM-algorithm, the M-step, we optimize the conditional likelihood function and find expressions for the unknown model and transformation parameters. We introduce a further simplification by assuming that $m = 1$, implying that the number of output variables at a certain time point equals 1. This simplification is no reduction of generality as again all expressions can be extended to the case where $m > 1$. We write the following conditional likelihood function given x, $\boldsymbol{\eta}$ and a:

$$L^* = L_1^* \cdot L_2^* ;$$

$$\begin{aligned} L_1^* &= E(\log L_1 \mid \mathbf{x}, \boldsymbol{\eta}, \alpha) \\ &= \text{const} - \frac{N}{2} \log \mid \mathbf{B\Theta B}' \mid \\ &\quad - \frac{N}{2} \operatorname{tr} [\mathbf{m}_{zz}^* (\mathbf{B\Theta B}')^{-1} - 2\, \mathbf{m}_{xz}^* (\mathbf{B\Theta B}')^{-1} (\mathbf{B} \otimes \mathbf{g}') \\ &\qquad + \mathbf{m}_{xx}^* (\mathbf{B} \otimes \mathbf{g}')' (\mathbf{B\Theta B}')^{-1} (\mathbf{B} \otimes \mathbf{g}')] ; \end{aligned} \tag{6}$$

$$\begin{aligned} L_2^* &= E(\log L_2 \mid \mathbf{x}, \boldsymbol{\eta}, \alpha) \\ &= \text{const} - \frac{N}{2} \log \mid Q \mid \\ &\quad - \frac{N}{2} \operatorname{tr} [\mathbf{m}_{\eta\eta}^* \mathbf{Q}^{-1} - 2\, \mathbf{m}_{\eta x}^* h \mathbf{Q}^{-1} (\mathbf{B} \otimes \mathbf{g}') \\ &\qquad + h^2 \mathbf{m}_{xx}^* (\mathbf{B} \otimes \mathbf{g}')' \mathbf{Q}^{-1} (\mathbf{B} \otimes \mathbf{g}')], \end{aligned} \tag{7}$$

where $Q = ((\mathbf{B\Theta B}') \otimes hh') + \boldsymbol{\Psi}$.

From (6) and (7) one can see that L^* is a function of h, g, f, $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$. Optimalization can be simplified considerably by formulating a **bijection** between (h, g, **f**, $\boldsymbol{\Theta}$, $\boldsymbol{\Psi}$) and (h, **g**, f, $\mathbf{B\Theta B}'$, $\mathbf{Q}$). We will thus write L^* as a function of (h, g, f, $\mathbf{B\Theta B}'$, Q), and subsequently optimize the new expression.

Within the M-step, we will optimize with respect to the model parameters h, g, f, $\mathbf{B\Theta B}'$, Q and the transformation parameters a altematingly.

Setting the derivative of the likelihood function equal to zero with respect to the various unknowns, it can be shown that:

$$h = \frac{\operatorname{tr} [\mathbf{m}_{\eta x}^* \mathbf{Q}^{-1} (\mathbf{B} \otimes \mathbf{g}')]}{\operatorname{tr} [\mathbf{m}_{xx}^* (\mathbf{B} \otimes \mathbf{g}')' \mathbf{Q}^{-1} (\mathbf{B} \otimes \mathbf{g}')]}, \tag{8}$$

$$f = \frac{\operatorname{tr} [\mathbf{m}_{xz}^* (\mathbf{B\Theta B}')^{-1} (\mathbf{B}_* \otimes \mathbf{g}') + \mathbf{m}_{\eta x}^* h \mathbf{Q}^{-1} (\mathbf{B}_* \otimes \mathbf{g}')]}{\operatorname{tr} [\mathbf{m}_{xx}^* (\mathbf{B} \otimes \mathbf{g}')' (\mathbf{B\Theta B}')^{-1} (\mathbf{B}_* \otimes \mathbf{g}') + h^2 \mathbf{m}_{xx}^* (\mathbf{B} \otimes \mathbf{g}')' \mathbf{Q}^{-1} (\mathbf{B}_* \otimes \mathbf{g}')]}, \tag{9}$$

with $\mathbf{B}_* = \frac{(\mathbf{B} - \mathbf{I})}{f}$.

$$(\mathbf{B\Theta B'}) = \mathbf{m}_{zz}^* - 2\mathbf{m}_{zx}^* (\mathbf{B}\otimes\mathbf{g}')' + (\mathbf{B}\otimes\mathbf{g}')\, \mathbf{m}_{xx}^* (\mathbf{B}\otimes\mathbf{g}')' , \qquad (10)$$

and

$$Q = \mathbf{m}_{\eta\eta}^* - 2\mathbf{m}_{\eta x}^*\, h\, (\mathbf{B}\otimes\mathbf{g}')' + h^2(\mathbf{B}\otimes\mathbf{g}')'\, \mathbf{m}_{xx}^* (\mathbf{B}\otimes\mathbf{g}') , \qquad (11)$$

and

$$\mathbf{g} = \mathbf{M}^{-1}\mathbf{m}, \qquad (12)$$

where

$$\mathbf{M}_{ij} = \mathrm{tr}\,[\mathbf{m}_{xx}^*(\mathbf{B}\otimes\mathbf{I}_i')'(\mathbf{B\Theta B}')^{-1})(\mathbf{B}\otimes\mathbf{I}_j') + h^2\mathbf{m}_{xx}^*(\mathbf{B}\otimes\mathbf{I}_i')'\mathbf{Q}^{-1}(\mathbf{B}\otimes\mathbf{I}_j')] ,$$

$$\mathbf{m}_i = \mathrm{tr}\,[\mathbf{m}_{xz}^*(\mathbf{B\Theta B}')^{-1})(\mathbf{B}\otimes\mathbf{I}_i') + \mathbf{m}_{\eta x}^*\, h\, \mathbf{Q}^{-1}(\mathbf{B}\otimes\mathbf{I}_i')] ,$$

with $(i, j = 1, \ldots. k)$ and $\mathbf{I}_i$ a vector with the i - th element equalling 1, and all other elements equalling zero.

However, before we can optimize L* with respect to a, we have to specify F_α. Several options are available regarding the choice of F_α. It is possible to choose some specific family of distributions. If F_α is a Box-Cox transformation (with only one parameter), L^* can be optimized easily with respect to a. Such a Box-Cox transformation can be written as:

$$Y_{it} = F_\alpha^{-1}(\eta_{it}),$$

where $F_\alpha = \dfrac{(\eta_{it})^\alpha - 1}{a}$, with a $\neq$ 0.

The estimation procedure for the unknown model parameters h, g, f, Ψ, Θ and a is thus as follows:

step 0. Initialization
Choose starting values for h, g, f, Ψ, Θ and a. From these compute starting values for $\mathbf{B\Theta B'}$ and Q.

step la. E-step
Compute $\mathbf{m}_{xx}^*$, $\mathbf{m}_{\eta\eta}^*$, $\mathbf{m}_{\eta x}^*$, $\mathbf{m}_{zz}^*$ and $\mathbf{m}_{xz}^*$.

step lb. M-step
Compute h, g, f, $\mathbf{B\Theta B'}$, $\mathbf{Q}$ and a alternatingly.

step 2. Repeat steps la and lb until convergence.

step 3. Compute Θ and Ψ using h, g, f, $\mathbf{B\Theta B'}$, and Q.

IDENTIFIABILITY

Solutions for the model defined above in (2) have rotational freedom. This can be seen easily if we define:

$$y_{i,t} = \mathbf{H}\,(\mathbf{F}\,\mathbf{z}_{i,t-1} + \mathbf{G}\,\mathbf{x}_{i,t} + \boldsymbol{\varepsilon}_{i,t}) + \boldsymbol{\delta}_{i,t} =$$

$$\mathbf{HF}\mathbf{z}_{i,t-1} + \mathbf{HG}\mathbf{x}_{i,t} + \mathbf{H}\boldsymbol{\varepsilon}_{i,t} + \boldsymbol{\delta}_{i,t}.$$

Any solution for H can be replaced by the solution HR, with R some rotation matrix, and $\mathbf{z}_{i,t-1}$, $\boldsymbol{\varepsilon}_{i,t}$, F, and G replaced by $\mathbf{R}^{-1}\mathbf{z}_{i,t-1}$, $\mathbf{R}^{-1}\boldsymbol{\varepsilon}_{i,t}$, $\mathbf{R}^{-1}\mathbf{FR}$ and $\mathbf{R}^{-1}\mathbf{G}$, respectively. Thus, the model is not identified without additional restrictions.

In terms of the particular model to which our derivations apply; if $\{g^*_1, \ldots g^*_k, h^*, f^*, \mathrm{var}(z)^*\}$ is a feasible parameter solution, then $\{cg^*_1, \ldots cg^*_k, \frac{1}{c}h, f^*, c^2\mathrm{var}(\mathbf{z})^*\}$ is an equivalent parameter solution in terms of model fit (with c any arbitrary real value unequal to zero). To solve this scaling or rotational problem, we have to introduce a constraint. The options for doing so are to set either $h \equiv 1$, $\mathbf{g}_1 \equiv 1$ or $\mathrm{var}(\mathbf{z}_1) \equiv 1$. Compare Bijleveld and De Leeuw (199 1).

THE CHOICE OF INITIAL STATE

We have assumed the value of the state at timepoint 0 to be zero for each observation unit. In principle, it is possible to model $\mathbf{z}_0$ as a model parameter, which can be estimated along with the other parameters using the EM-algorithm. However, if one would restrict the class of feasible models in the sense that the choice is made to center the input and output variables to zero (that is: expectations are all zero), then $\mathbf{z}_0$ has to be equal to zero for each subject (proof by contradiction). The importance of a proper choice for the values of $\mathbf{z}_0$ depends on the value of f. The smaller f, the smaller the impact of $\mathbf{z}_0$ and the smaller the disturbance created by an improper choice of $\mathbf{z}_0$. Much more can be said about the initial state, but this is outside the scope of this paper.

EXAMPLE USING REAL DATA

To illustrate our model, we have analysed a data set on the relationship between mood and urge to smoke[1] (Olmstead, 1996). Data had been gathered for 35 subjects on two consecutive days. Approximately every 20 minutes, subjects had filled in a

[1] The authors wish to thank dr. L. Jamner for his kind permission to use the data. Data collection had been supported by a grant from the National Institute of Neurological Disorders and Stroke Research (grant no NS-34143) as well as by UC-TRDRP grant no 1RT0205.

questionnaire, registering their mood states of the moment. They had also registered their urge to smoke. As the data had been gathered over a period of two days, every series contained at least one big gap (when the subjects were sleeping); however, many series contained other smaller gaps as well (for instance when subjects were driving, engaged in sports, in meetings etc.). For our purposes, we selected all uninterrupted series that had at least 10 time points, retaining the first 10 time points of each. As such, a number of subjects appear several times in the data set. A total number of 59 series remained in this way.

The mood variables that had been measured were: energy, tiredness, anxiety, hostility, happiness and sadness. The mood variables had been measured on 7 point scales, 'urge to smoke' had been recorded on a 7 point scale as well. All variables were moderately to heavily skewed.

We ran our model on this 59 x 7 x 10 dataset. To solve for indeterminacy of solutions, h was fixed at 1. This choice was based on the considerations that the dimensionality of h equalled that of z, and that the envisaged Box-Cox transformation gave subsequent liberty in the choice of transformation. We used Box-Cox transformations for the input and output variables. The transformation functions were assumed to be time-invariant. We gave all the Box-Cox parameters starting values equal to one. This implies that y and η are equal to each other in the first iteration step of the EM-algorithm only.

The overall model fit can be assessed objectively using, amongst others, the Goodness of Fit Index (GFI) and the Adjusted Goodness of Fit Index (AGFI) (cf. Bollen & Long, 1993). For calculating these values, we use the transformed input- and output variables, which are normally distributed.

Table 1. Parameter Estimates, Variances and Box-Cox Parameters for Mood and Urge to Smoke Data

Parameter		Value	(Variance)	Box-Cox Parameter Values
f		.49	(.18)	(-)
g	energy	.01	(.01)	.9
	tiredness	.56	(.14)	.4
	anxiety	.11	(.08)	1.0
	hostility	.09	(.08)	1.0
	happiness	.30	(.09)	.6
	sadness	.01	(.01)	1.0
h		1	(-)	.8

The algorithm converged rapidly to fit values of .95 (GFI) and .92 (AGFI), both of which, being larger than .9, are quite acceptable values. We also ran the algorithm without the Box-Cox transformations (i.e. with all the Box-Cox parameters values fixed at one). We used the loglikelihood ratio statistic to compare the model with transformed variables to the model without transformed variables. The value of this loglikelihood ratio statistic turned out to be 17.9 with seven degrees of freedom, so that we may conclude that the null hypothesis with all Box-Cox parameters equal to one can be rejected on the basis of the data (5% significance level). Using the Box-Cox transformation (or another non-linear transformation) therefore appears justified.

The parameter estimates, with variances given between brackets and Box-Cox transformation parameters are given in Table 1. The algorithm converged to a technically satisfactory solution for the linear dynamic system. The value of f is between 0 and 1, indicating that a stable system has been modeled in which the values of the latent state when unperturbed by outward influences converge to zero. The values of the Box-Cox parameters show that for each timepoint the first, second and fifth input variable are transformed non-linearly by the Box-Cox functions. The values of the transformation furthermore show that the output variable 'urge to smoke' did indeed behave nonlinearly.

Once the estimates of the model parameters have been computed, their variances may be found from the information matrix. This is the matrix of second order derivatives with respect to the unknown parameters of the loglikelihood function.

When we tested the hypothesis $f = 0$, the likelihood ratio statistic and the t-statistic showed that f is significant on a 5% level. Thus, there is a significant impact of the previous measurements on the present ones. Next, we tested the significance of each of the six input variables, using loglikelihood ratio statistics and t-statistics for each regression coefficient. From these tests, it emerged that the first, third, fourth and sixth variable could be dropped out of the model (probability levels of 0.7, 0.8, 0.7 and 1.0 respectively, and t-values smaller than 1.96). Only the second and fifth input variable are relevant statistically speaking (with probability levels of 0.04 and 0.05 respectively and t-values larger than 1.96). Both of these variables had been transformed nonlinearly as can be seen from the Box-Cox parameter values.

From the estimated parameter values one can deduce that tiredness and previous urge to smoke are the most important predictors for urge to smoke. Happiness is a less important predictor. Urge to smoke is high when subjects are tired, when their urge to smoke was high at the previous time point, and (though less particularly so) when they are happy. Energy, anxiety, hostility and sadness play hardly any role at all.

DISCUSSION

In the above, we presented an extension to the existing methodology for multivariate panel data that is both useful and efficient. The model is useful as it can handle situations with several subjects and more than just a few time points or waves. The model can easily incorporate non-normal and ordinal output variables. An advantage of our method over existing methodology for such data analytic situations, is that our technique provides stability information on the parameter estimates. While this was not presented in the example give above, the variances of the parameters estimates can be used to compute confidence intervals, which may be even more easily interpretable. An additional advantage, not immediately obvious from our formulas, is that it is not necessary for all subjects to have series of equal length: $T_i \neq T$. This means that subjects who drop out of the study before the end of the longitudinal research phase can be incorporated in the analysis up to the last time point they participated. Dropout is one of the most serious validity problems in longitudinal research, which becomes progressively worse the longer the observation period and is a generally non-random process. The possibility to retain subjects for as long as they partook in the study, is thus a true advantage over techniques such as LISREL or MANOVA, that need complete series. Hierarchical models also allow subjects to have series of unequal length. The type of temporal developments captured by our type of model is, however, far more general than those featuring in hierarchical models, such as linear or quadratic. A second advantage of our models over structural equations models, is that, whereas in LISREL the inverse of a $(kT+mT) \times (kT+mT)$ matrix has to be computed, in our method the inverse has to be computed of one matrix that has dimensionality $pT \times pT$ and one matrix that has dimensional@ $mT \times mT$. This makes a considerable difference in computing time, computer memory and precision.

From our admittedly limited experience thus far, our method appears to converge swiftly. This is probably due to the bijection between $\mathbf{B\Theta B'}$ and $\mathbf{\Theta}$ on the one hand and Q and Ψ on the other hand. Our method can be sped up even further if it can be safely assumed that the variance covariance matrices of the error terms are equal not only across subjects, but over time points as well. In that case, the $mT \times mT$ matrix Q reduces to an $m \times m$ matrix, which considerably simplifies the inversion problem in (8), (9), (10), (11) and (12). We could also have used a step-function for non-linearly transforming the output variables. However, this would have introduced many more parameters to our model, and would thus have implied a less parsimoneous model.

All the model simplifications introduced in the course of this paper can be gotten rid of again, if necessary. This will make the derivations, and computations, more complicated. In that case, the algorithm will take longer to converge. However, given our experiences thus far with exceptionally speedy convergence, we do not expect this to prove a constraint of any practical relevance.

In our example, we fixed the value of h at 1 to solve the identification problem. In general, fixing the variance of the latent state space values may be a more appropriate policy, as it constitutes a fairly unequivocal choice, and is similar to other standardizations in comparable data analytic situations (e.g. factor analysis).

For behavioral applications, it may be useful to be able to compare subjects with respect to their latent state scores. Our technique does not produce such scores. Following Oud *et al.* (1990) and Oud, van Leeuwe and Jansen (1993), latent state scores may be found using the Kalman filter (Kalman, 1960). Assigning an arbitrary starting value, values for the latent state scores can be computed using the transmission parameters, expected values and (co)variances of the error terms. The Kalman filter is known to become independent of the starting values after a very small number of time points, in practice this may occur as swiftly as after three to five time points from t_0.

The practical relevance of our model will have to be tested by extensive application on generated as well as empirical data. The addition of transformations for input variables as well as categorical variables would further increase this relevance.

REFERENCES

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

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

De Leeuw, J., Bijleveld, C.C.J.H., van Montfort, K., & Bijleveld, F.D. (1997). Latent variables, state spaces and mixing. *Kwantitatieve Methoden,* **55**, **95**- 111.

Bijleveld, C.C.J.H., & Bijleveld, F.D. (1997). A comparison of two approaches to state space modelling. *Kwantitatieve Methoden,* **55**, **37-56**.

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

Bijleveld, C.C.J.H. & 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,* pp. 27-32. Stuttgart, New York: Gustav Fisher Verlag.

Bollen, K.A., & Long, J. Scott (1993). *Testing structural equations models.* London: Sage.

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

Browne, M.W. (1992). Circumplex models for correlation matrices. *Psychometrika,* 57, 469-497.

Browne, M.W., & du Toit, S.H.C. (1991). Models for learning data. In L.M. Collins & J. L. Horn (Eds.), *Best models for the analysis of change,* pp. 47-68. Washington: American Psychological Association.

Bryk, A.S., & Raudenbusch, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Collins, L.M., & Wugalter, S.E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research, 27,* 13 1- 157.

De Leeuw, J., Bijleveld, C.C.J.H., van Montfort, K., & Bijleveld, F.D. (1997). Latent variables, state spaces and mixing. *Kwantitatieve Methoden,* **55**, **95**- 111.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B, 39,* l-38.

'Jöreskog, K.G., & 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**, **63** l-639.

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

Hershberger, S.L., Molenaar, P.C.M., & Corneal, S.E. (1996). A hierarchy of univariate and multivariate structural time series models. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling: issues and techniques*, pp. 159- 194. Mahwah, NJ: Lawrence Erlbaum.

Ho, B.L., & Kalman, R.E. (1966). Effective construction of linear state-variables models from input/output functions. *Regelungstechnik,* **12, 545-548.**

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

Langeheine, R., & van de Pol, F. (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods and Research,* **18,** 416-441.

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

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

McArdle, J.J., & Hamagami, F. (1991). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. In L. M. Collins & J. L. Horn (Eds.), *Best models for the analysis of change,* pp. 276-298. Washington: American Psychological Association.

McArdle, J.J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: issues and techniques,* pp. 89-124. Mahwah, NJ: Lawrence Erlbaum.

Molenaar, P.C.M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika, 50,* 18 l-202.

Molenaar, P. C. M., de Gooijer, J.G., & Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika, 57, 333-349.*

Mooijaart, A., & van Montfort, K. (1997). State space analysis of categorical variables. *Kwantitatieve Methoden,* **55**, 73-94.

Muthén, B. (1996). Growth modeling with binary responses. In A. von Eye & C.C. Clogg (Eds.), *Categorical variables in developmental research,* pp. 37-54. San Diego, CA: Academic Press.

Olmstead, R. E. (1996). Diurnal associations of smoking and nicotine with mood. UMI number 9704609, Ann Arbor, MI: UMI.

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

Oud, J.H.L, van Leeuwe, J.F.J., & Jansen, R.A.R.G. (1993). Kalman filtering in discrete and continuous time based on longitudinal LISREL models. In J.H.L. Oud & A.W. van Blokland-Vogelesang (Eds.), *Advances in longitudinal and* *multivariate analysis in the behavioural sciences*, pp. 3-26. Nijmegen, the Netherlands: ITS.

Rogosa, D.R., & Willett, J.B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika,* **50**, 203-228.

Smith, T.M.F., & Brunsdon, T.M. (1989). The time series analysis of compositional data. *American Statistical Association. Proceedings of Survey Research Methods Section,*. pp. 26-32.

Van Buuren, S. (1997). Fitting ARMA time series by structural equation models. *Psychometrika, 62,* 215236.

Van de Pol, F., & De Leeuw, J. (1986). A latent Markov model to correct for measurement error. *Sociological Methods and Research,* **15,** 118- 14 1.

Van de Pol, F., & Langeheine, R. (1989). Mixed Markov latent class models. In C. C. Clogg (Ed.), *Sociological methodology,* pp. 2 13-247. Oxford: American Sociological Association.

Willett, J.B., & Sayer, A.G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin,* **116, 363-381.**