SEM State Space Modeling of Panel Data and its Relationship to Traditional Single Subject State Space Modeling

Johan H.L. Oud

Department of Special Education
University of Nijmegen
The Netherlands

Johan H.L. Oud
Department of Special Education, University of Nijmegen
Montessorilaan 3
PO Box 9104, 6500 HE Nijmegen
Voice +31.24.612137/612131, Fax +31.24.612776
E-mail j.oud@ped.kun.nl

Abstract

The state space approach covers a very general class of dynamic models. Examples are ARMA and ARMAX models as well as longitudinal factor and path analysis models. Originating from modern control theory, the state space approach becomes increasingly popular in other fields as well. The paper first explains SEM state space modeling and parameter estimation. Then, using the Expectation-Maximization (EM) algorithm in combination with Kalman filtering and smoothing, the relationship to traditional $N = 1$ state space model parameter estimation is shown. Starting from the state space model in SEM form, the likelihood function is decomposed in a part for the state equation and a part for the output equation. A further decomposition using the assumption of time-invariant parameters leads to $N = 1$ state space model parameter estimation. Because of the drawbacks of both decompositions, however, the direct SEM approach without decomposition is preferred in the case of large N panel data. For the case of both N and T large (T the number of time points covered by the model), SEM overlapping cohort modeling is recommended.

Keywords: ARMA model, EM algorithm, Kalman smoother, large N modeling, longitudinal SEM model, maximum likelihood, missing data, prewhitening, SEM, state space model.

SEM State Space Modeling of Panel Data and its Relationship to Traditional Single Subject State Space Modeling

The state space approach, which is in the core of modern control theory and becomes increasingly popular in other fields as well, covers a very general class of dynamic models. In fact, all nonanticipative models (models with no causal arrows heading backward in time) can be represented in state space form. For example, both the Box-Jenkins ARMA model and the extended ARMAX model, which adds exogenous or input variables to the ARMA model, are easily formulated as special cases of the discrete-time time-invariant state space model (Caines, 1988; Deistler, 1985; Ljung, 1985). The state space model covers also longitudinal latent factor and path analysis models and allows the optimal estimation of the latent states or factor scores (Oud, van den Bercken, & Essers, 1990; Oud, van Leeuwe, & Jansen, 1993). Latent state estimation is performed by two important results of the state space approach: the Kalman filter and the Kalman smoother (Jazwinski, 1970; Lewis, 1986; Rauch, Tung, & Striebel, 1965). These results have additionally been proven to yield powerful methods for the estimation of the parameters of the state space model in the single time series ($N = 1$) case (Caines & Rissanen, 1974; Dembo & Zeitouni, 1986; Mehra, 1971; Shumway & Stoffer, 1982). The methods are based on the iterative reestimation of the unknown latent states in conjunction with stepwise improvement of parameter estimates.

Drawbacks of these parameter estimation methods stem from their restriction to the $N = 1$ case, in particular the fact that the lack of independent replications in the $N = 1$ case makes it necessary to assume time-invariant parameters and the necessity of observations from a large number of time points to obtain estimates with reasonably low variances. Also an asymptotic theory for $T \to \infty$ instead of $N \to \infty$ needs to be considered. Recently it has been shown that the state space model may also be formulated in terms of a latent variables structural equation model (SEM) and its parameters estimated by a SEM program such as LISREL (MacCallum & Ashby, 1986; Oud et al., 1990; Oud et al., 1993). Here the model is estimated on the basis of panel data, viewed as N independent replications of the random vector of all observed variables over all time points. In this approach with N large, the parameters may be chosen time-varying and the number of time points T may be arbitrarily small. The option of time-invariant parameters is retained by the possibility of specifying equality constraints between parameters.

The present paper first explains SEM state space modeling and parameter estimation. Starting with the basic formulation, next latent time-varying

ARMA modeling, the introduction of exogenous or input effects in the model, and the state-trait model is discussed. Then, using the Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977; Little & Rubin, 1987) in combination with Kalman filtering and smoothing (Jansen & Oud, 1995), the relationship to traditional $N = 1$ state space model parameter estimation is shown. Starting from the state space model in SEM form, the likelihood function is decomposed in a part for the state equation and a part for the output equation. A further decomposition using the assumption of time-invariant parameters will lead to the traditional $N = 1$ approach as given, for example, by Shumway and Stoffer (1982). The traditional approach may be extended to the $N > 1$ case (Singer, 1993). Because of the drawbacks of both decompositions, however, the direct SEM approach without decomposition is preferred in the case of large N. For the case of both N and T large, SEM overlapping cohort modeling is recommended.

State Space Modeling by Means of SEM

The state space model consists of two equations: the dynamic part or state equation (Equation 1), which describes the dependence of the latent state variables in $\mathbf{x}_t$ on their lagged values in $\mathbf{x}_{t-1}$ and the static part or output equation, which connects the latent state variables to the observables in $\mathbf{y}_t$ (Equation 2):

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{w}_{t-1} \qquad \text{with} \qquad cov(\mathbf{w}_{t-1}) = \mathbf{Q}_{t-1} \ , \tag{1}$$

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t \qquad \text{with} \qquad cov(\mathbf{v}_t) = \mathbf{R}_t \ . \tag{2}$$

The matrix $\mathbf{A}_{t-1}$ in Equation 1 contains the autoregressive and cross-lagged effects between the state variables at successive discrete time points t and $t-1$: $t, t-1 \in \{t_0, t_0+1, \ldots, t_0+T-1\}$ for integers t_0 and $T \geq 2$, with t_0 the initial time point and T the total number of time points considered. The output or measurement equation (Equation 2) is equivalent to the factor model equation in factor analysis with $\mathbf{C}_t$ the factor pattern matrix. $\mathbf{Q}_{t-1}$ and $\mathbf{R}_t$ are the process error and measurement error covariance matrix, respectively.

Instead of Equation 1, many econometric and social science models choose a so-called structural equation, which has $\mathbf{x}_t$ at its right hand side as well as its left hand side. One of the advantages of the SEM approach is that the underlying structural parameters can be estimated instead of the state equation parameters. This is possible even in combination with the Kalman filter and smoother, because before applying these devices the structural equation can be reduced to Equation 1 (Oud et al., 1990).

The process errors in successive vectors $\mathbf{w}_t$ and the measurement errors in successive vectors $\mathbf{v}_t$ are assumed to have (a) zero expectations: $E(\mathbf{w}_t) = E(\mathbf{v}_t) = \mathbf{0}$ for all t, (b) zero covariances between vectors: $E(\mathbf{w}_t\ \mathbf{v}'_{t'}) = \mathbf{0}$ for all t and t', $E(\mathbf{w}_t\ \mathbf{w}'_{t'}) = E(\mathbf{v}_t\ \mathbf{v}'_{t'}) = \mathbf{0}$ for all $t \neq t'$ (nonzero variances and covariances for errors within vectors are in $\mathbf{Q}_t$ and $\mathbf{R}_t$), and (c) zero covariances with the initial state: $E(\mathbf{w}_t\ \mathbf{x}'_{t_0}) = E(\mathbf{v}_t\ \mathbf{x}'_{t_0}) = \mathbf{0}$ for all t. Further, (d) the error vectors and the initial state are assumed to be jointly multinormally distributed. Finally, it is assumed (e) $E(\mathbf{x}_{t_0}) = E(\mathbf{y}_{t_0}) = \mathbf{0}$, implying $E(\mathbf{x}_t) = E(\mathbf{y}_t) = \mathbf{0}$ for all t. (See Meditch, 1969, pp. 168-169). Assumptions (a) through (d) are essential in state space as well as SEM modeling. Assumption (e) leads to the so-called "zero means" SEM model. Below this assumption will be dropped to obtain the flexible "structured means" SEM model which enables the estimation of the means structure in addition to the covariance structure.

By taking in the SEM model equations,

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \qquad \text{with} \qquad cov(\boldsymbol{\zeta}) = \boldsymbol{\Psi}\ , \tag{3}$$

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad \text{with} \qquad cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}\ , \tag{4}$$

$\boldsymbol{\eta} = [\mathbf{x}'_{t_0}\ \mathbf{x}'_{t_0+1} \cdots\ \mathbf{x}'_{t_0+T-1}]'$ and $\mathbf{y} = [\mathbf{y}'_{t_0}\ \mathbf{y}'_{t_0+1} \cdots\ \mathbf{y}'_{t_0+T-1}]'$ with t_0 the initial time point and T the total number of time points considered, and putting the parameter matrices of Equations 1 and 2 on the appropriate places in the parameter matrices $\boldsymbol{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Theta}$, the SEM model is easily formulated as a state space model. Notice that the initial state $\mathbf{x}_{t_0}$, being exogenous or unexplained in the state space model, has zero rows in $\boldsymbol{B}$ and its covariance matrix $\boldsymbol{\Phi}_{t_0} = E(\mathbf{x}_{t_0}\ \mathbf{x}'_{t_0})$ specified in $\boldsymbol{\Psi}$. The other nonzero elements of $\boldsymbol{\Psi}$ are the process error variances and covariances in successive matrices $\mathbf{Q}_t$ with $t = t_0, \ldots, t_0 + T - 2$. Because all and only all the assumptions of the state space model are specified in the SEM model, the SEM model becomes fully equivalent to the state space model. Several parameter estimation methods can be used in most SEM programs (e.g. Jöreskog & Sörbom, 1989, p. 16). Here the ML method is applied, which maximizes the loglikelihood function of the free parameters in parameter matrices $\boldsymbol{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$, for given data in $\mathbf{Y}$:

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{N}{2}\log|\,\boldsymbol{\Sigma}\,| - \frac{N}{2}\mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \frac{pN}{2}\log 2\pi\ . \tag{5}$$

$\boldsymbol{\theta}$ in Equation 5 contains the free parameters, $\mathbf{Y}_{p\times N}$ is the data matrix (N columns of independent replications of the p-variate vector $\mathbf{y}$, typically originating from a sample of randomly drawn subjects), $\boldsymbol{\Sigma}_{p\times p}$ is the model implied covariance matrix:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \boldsymbol{B}')^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}\ , \tag{6}$$

which is a function $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$, and $\mathbf{S}_{p\times p} = \frac{1}{N}\mathbf{Y}\ \mathbf{Y}'$ is the sample moment or covariance matrix. The ML-estimator $\hat{\boldsymbol{\theta}} = \mathrm{argmax}\ \ell(\boldsymbol{\theta}|\mathbf{Y})$ chooses that value of $\boldsymbol{\theta}$ which maximizes $\ell(\boldsymbol{\theta}|\mathbf{Y})$. However, instead of maximizing Equation 5 the SEM program minimizes fit function

$$F_{ML} = \log|\,\boldsymbol{\Sigma}\,| + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log|\,\mathbf{S}\,| - p \tag{7}$$

with the same result. Equation 7 only differs from Equation 5 in the negative multiplying constant $-\frac{2}{N}$ and an additive constant; $\mathbf{S}$ is based on the data and thus constant in the SEM fit function.

In the frequent case of missing data as caused, for example, by panel attrition, the state space model naturally leads to the following EM procedure (Jansen & Oud, 1995; Oud & Jansen, 1996). In addition to being missing completely at random (MCAR), the procedure allows the data to be missing at random (MAR) in the sense of Little and Rubin (1987). Not the complete data loglikelihood $\ell(\boldsymbol{\theta}|\mathbf{Y}) = \ell(\boldsymbol{\theta}|Y_{obs}, Y_{mis})$ in Equation 5 but the loglikelihood $\ell(\boldsymbol{\theta}|Y_{obs})$, given the observed data only, has to be maximized. As this cannot be done directly, the conditional loglikelihood expectation is determined and maximized repeatedly by means of the SEM program. It depends on the observed data Y_{obs} and parameter values $\hat{\boldsymbol{\theta}}_r$ of the preceding M-step: $E_{Y_{mis}}[\ell(\boldsymbol{\theta}|\mathbf{Y})|Y_{obs}, \hat{\boldsymbol{\theta}}_r]$. The expectation is taken over the distribution of the missing data Y_{mis} given the observed data Y_{obs} and the current estimate $\hat{\boldsymbol{\theta}}_r$. For implementation of the EM algorithm the conditionally expected moment or covariance matrix $\mathbf{S}_{r+1} = E_{Y_{mis}}(\mathbf{S}|Y_{obs}, \hat{\boldsymbol{\theta}}_r)$ is to be calculated in the E-step and inserted for $\mathbf{S}$ in Equation 5. This is due to the fact that the loglikelihood function in Equation 5 is linear in $\mathbf{S}$. Except for the replacement of $\mathbf{S}$ by

$\mathbf{S}_{r+1}$, $E_{Y_{mis}}[\ell(\boldsymbol{\theta}|\mathbf{Y})|Y_{obs}, \hat{\boldsymbol{\theta}}_r]$ does not differ from $\ell(\boldsymbol{\theta}|\mathbf{Y})$ in Equation 5 (which is handled in the SEM program by means of Equation 7, where again $\mathbf{S}$ is to be replaced by $\mathbf{S}_{r+1}$). In case of no missing data $E_{Y_{mis}}(\mathbf{S}|Y_{obs}, \hat{\boldsymbol{\theta}}_r) = \mathbf{S}$, and $E_{Y_{mis}}[\ell(\boldsymbol{\theta}|\mathbf{Y})|Y_{obs}, \hat{\boldsymbol{\theta}}_r]$ becomes equal to $\ell(\boldsymbol{\theta}|\mathbf{Y})$ of Equation 5. The computation of $\mathbf{S}_{r+1}$ in the E-step requires the computation of the conditional expectations for individual subjects, which are the Kalman filter and smoother values (Jazwinski, 1970; Lewis, 1986; Rauch, Tung, & Striebel, 1965). A detailed explanation of the Kalman filter and smoother as well as the computation of $\mathbf{S}_{r+1}$ can be found in Jansen and Oud (1995) and Oud and Jansen (1996). In contrast to many ad hoc missing value procedures, in the last iteration of the EM procedure the SEM program produces the correct ML parameter estimates $\hat{\boldsymbol{\theta}}$ and model implied covariance matrix $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ (see Equation 6), which are consistent and on the basis of which, under the usual independence and normality assumptions, correct asymptotic standard errors are computed by the SEM program.

Latent Time-varying ARMA Modeling by Means of SEM

In the state space model, curve shape is not only described but explained dynamically in terms of the effects of the random variables at previous time points on those at subsequent time points. The introduction of moving average parameters as well as higher order autoregressive parameters in the ARMA model further enhances dynamic explanation power. In particular, the specification of higher order effects extends the time span of past effects on the present.

Fundamental for the state space model is the Markov or state separation property. This property, in fact, defines the concept of state. The requirement is that at each point in time the state be specified to be exclusively dependent on the just preceding state but not on states further back in time. This implies that the state must follow a first-order autoregressive scheme. Because, in addition, the state space model does not contain moving average parameters, in time series terminology it seems to cover only models of the order ARMA(1,0): first-order autoregressive and zero-order moving-average. Although special techniques are known from the literature, by which the state is reformulated and extended in such a way that ARMA models of arbitrary order fit into the reformulated state space model (e.g. Akaike, 1974; Deistler, 1985; Jones, 1985), typically these reformulations (1) assume the ARMA model to be time-invariant and (2) do not fit into the SEM model. These problems are solved in the reformulation given below.

Instead of the state equation in Equation 1, the "standard noise" version in Equation 8 is chosen, which turns out to be an appropriate form for fitting in and estimating moving average parameters by means of the SEM program:

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{G}_{t-1}\mathbf{z}_{t-1} \qquad \text{with} \qquad cov(\mathbf{z}_{t-1}) = \mathbf{I} \ . \tag{8}$$

Equation 8 is more general than Equation 1, but reduces to the latter in the following way: $\mathbf{w}_{t-1} = \mathbf{G}_{t-1}\mathbf{z}_{t-1}$ and $\mathbf{Q}_{t-1} = \mathbf{G}_{t-1}\mathbf{G}'_{t-1}$. As the number of different elements in $\mathbf{G}_{t-1}$ is larger than in the symmetric matrix $\mathbf{Q}_{t-1}$, identifiability of the parameters in Equation 1 does not necessarily imply identifiability of those in Equation 8. However, by specifying (1) a diagonal matrix $\mathbf{G}_{t-1}$ when $\mathbf{Q}_{t-1}$ is diagonal or (2) a triangular matrix $\mathbf{G}_{t-1}$ when $\mathbf{Q}_{t-1}$ is nondiagonal, the number of parameters in both equations is the same and the elements of either set can be expressed in the elements of the other one. Also, the estimates of the elements in either set can be used to derive estimates for those in the other set. Estimating $\mathbf{G}_{t-1}$, however, and deriving the estimate of $\mathbf{Q}_{t-1}$ as the reduced form $\mathbf{Q}_{t-1} = \mathbf{G}_{t-1}\mathbf{G}'_{t-1}$ has the additional advantage of preventing any negative diagonals (variance estimates) showing up in the direct estimate of $\mathbf{Q}_{t-1}$ (see Jöreskog & Sörbom, 1989, pp. 239-240). Nontriangular forms of $\mathbf{G}_{t-1}$ may also be specified but require more complicated identification techniques which are outside the scope of the present article.

The next step is reformulating the state and the state space model in the following way:

$$\begin{aligned}
\begin{bmatrix} \mathbf{x}_t \\ \mathbf{z}_t \end{bmatrix} &= \begin{bmatrix} \mathbf{A}_{t-1} & \mathbf{G}_{t-1} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{z}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{z}_t \end{bmatrix} \\
\mathbf{x}^\circ_t &= \qquad \mathbf{A}^\circ_{t-1} \qquad\quad \mathbf{x}^\circ_{t-1} \quad + \quad \mathbf{w}^\circ_{t-1} \\
&\text{with } cov(\mathbf{w}^\circ_{t-1}) = \mathbf{Q}^\circ_{t-1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} ,
\end{aligned} \tag{9}$$

$$\mathbf{y}_t = [\ \mathbf{C}_t \ \ \mathbf{0}\ ]\ \mathbf{x}^\circ_t + \mathbf{v}_t \qquad \text{with} \qquad cov(\mathbf{v}_t) = \mathbf{R}_t \ . \tag{10}$$

The reformulated state space model combines the latent vectors $\mathbf{x}_t$ and $\mathbf{z}_t$ in the reformulated state $\mathbf{x}^\circ_t$. The idea behind the reformulation is that by putting the noise $\mathbf{z}_t$ at time point t in the current state $\mathbf{x}^\circ_t$, at the next point in time it becomes available as $\mathbf{z}_{t-1}$ in the lagged state $\mathbf{x}^\circ_{t-1}$ to contribute to $\mathbf{x}_t$ in the new state $\mathbf{x}^\circ_t$. Although more general than Equations 1 and 2, Equations

9 and 10 are at the same time in the form of Equations 1 and 2 and fit as easily into the SEM model.

The extension of Equation 8 and of the corresponding state space model in Equations 9-10 to arbitrary ARMA(p, q) order is now straightforward. Equation 11 shows the ARMA(2,1) example which adds to Equation 8 the second-order autoregressive term $\mathbf{A}_{t,t-2}\mathbf{x}_{t-2}$ and the first-order moving average term $\mathbf{G}_{t,t-2}\mathbf{z}_{t-2}$:

$$\mathbf{x}_t = \mathbf{A}_{t,t-2}\mathbf{x}_{t-2} + \mathbf{G}_{t,t-2}\mathbf{z}_{t-2} + \mathbf{A}_{t,t-1}\mathbf{x}_{t-1} + \mathbf{G}_{t,t-1}\mathbf{z}_{t-1} \quad . \tag{11}$$

The special identification problems associated with the introduction of moving average parameters are addressed in Oud and Jansen (1995). Equation 11 is still not in state space form, because $\mathbf{x}_t$ is specified to be dependent not only on $\mathbf{x}_{t-1}$ but also on $\mathbf{x}_{t-2}$ of lag two. However, the next reformulation in terms of new lag-one state $\mathbf{x}^\circ_{t-1}$ and corresponding current state $\mathbf{x}^\circ_t$ is in correct state space form, appropriate for use in Kalman filtering and smoothing, and fits again into the SEM model:

$$\begin{aligned} \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{z}_{t-1} \\ \mathbf{x}_t \\ \mathbf{z}_t \end{bmatrix} &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{A}_{t,t-2} & \mathbf{G}_{t,t-2} & \mathbf{A}_{t,t-1} & \mathbf{G}_{t,t-1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-2} \\ \mathbf{z}_{t-2} \\ \mathbf{x}_{t-1} \\ \mathbf{z}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{z}_t \end{bmatrix} , \\ \mathbf{x}^\circ_t &= \qquad\qquad \mathbf{A}^\circ_{t-1} \qquad\qquad\qquad \mathbf{x}^\circ_{t-1} \quad + \quad \mathbf{w}^\circ_{t-1} \end{aligned} \tag{12}$$

$$\mathbf{y}_t = [\, \mathbf{0} \;\; \mathbf{0} \;\; \mathbf{C}_t \;\; \mathbf{0} \,]\, \mathbf{x}^\circ_t + \mathbf{v}_t \quad . \tag{13}$$

Extension to higher order ARMA(p, q) models is obtained simply by adding pairs $\mathbf{A}_{t,t-i}\mathbf{x}_{t-i}$ and $\mathbf{G}_{t,t-i}\mathbf{z}_{t-i}$ of increasing lag i to Equation 11 and extending correspondingly the state and state space model in Equations 12-13 until $i = \max(p,\ q+1)$. Notice that the reformulation keeps track of the identical parts in current state $\mathbf{x}^\circ_t$ and lagged state $\mathbf{x}^\circ_{t-1}$ by inserting submatrices $\mathbf{I}$ on the second upperdiagonal of $\mathbf{A}^\circ_{t-1}$. Notice also, that in collecting the state variables in the SEM model vector $\boldsymbol{\eta}$, the identical parts in successive state vectors $\mathbf{x}^\circ_t$ need only to be specified once.

Introducing Exogenous or Input Effects into the State Space and SEM Model

Because the traditional SEM model specified zero means, it did not allow to study and predict nonzero mean development over time. Zero means state space and SEM modeling, $E(\mathbf{y}_t) = E(\mathbf{x}_t) = \mathbf{0}$ for all t, is appropriate, if the data are generated indeed by zero means processes. If this is not valid, as is typically the case for real life data, a popular way of getting rid of nonzero means in the sample, while keeping the assumption $E(\mathbf{y}_t) = E(\mathbf{x}_t) = \mathbf{0}$ for all t, is subtracting the sample means from the data and dividing by $N-1$ in the computation of the sample variances and covariances. However, if instead of this deviation score transformation of the data, one simply wants to model and estimate explicitly the means process parameters in addition to those of the covariance structure, the state space model and corresponding SEM model need to be extended with fixed so-called input-effects $\mathbf{B}_{t-1}\mathbf{u}_{t-1} \neq \mathbf{0}$ and $\mathbf{D}_t\mathbf{u}_t \neq \mathbf{0}$ and also the assumption $E(\mathbf{x}_{t_0}) = \mathbf{0}$ may be dropped:

$$
\begin{aligned}
\mathbf{x}_t &= \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{w}_{t-1}, && (14)\\
\mathbf{y}_t &= \mathbf{C}_t\mathbf{x}_t + \mathbf{D}_t\mathbf{u}_t + \mathbf{v}_t \;. && (15)
\end{aligned}
$$

This realizes great flexibility in the specification of mean trajectories for latent and observed variables:

$$
\begin{aligned}
E(\mathbf{x}_t) &= \mathcal{A}_{t,t_0}E(\mathbf{x}_{t_0}) + \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k \;, && (16)\\
E(\mathbf{y}_t) &= \mathbf{C}_t\mathcal{A}_{t,t_0}E(\mathbf{x}_{t_0}) + \mathbf{C}_t\sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k + \mathbf{D}_t\mathbf{u}_t \;. && (17)
\end{aligned}
$$

Here $\mathcal{A}_{t,t_0} = \prod_{k=1}^{t-t_0} \mathbf{A}_{t-k}$ is the well-known state transition matrix, also defined for $t = t_0$: $\mathcal{A}_{t_0,t_0} = \mathcal{A}_{t,t} \equiv \mathbf{I}$ (Desoer, 1970, p. 71).

In one special case only a single, so-called unit input-variable is specified ($u_t = 1$ for all t), which is constant over time points as well as over subjects in the sample (Jöreskog & Sörbom, 1989, pp. 273-275). Here the vectors $\mathbf{b}_{t-1}$ represent latent growth intercepts and the vectors $\mathbf{d}_t$ location parameters (origins) of the measurement instruments. The model implies a means process which is common to all subjects in the sample. In another special case the input-variables are all constant over time ($\mathbf{u}_t = \mathbf{u}_{t-k}$ for all t and $k > 0$) but, apart from the unit input-variable, varying over subjects. This gives rise to the longitudinal SEM model with background variables (gender, socioeconomic status, etc.). In the general case, considered here, additional input-variables are specified that vary over time points as well as over subjects.

Whether $E(\mathbf{x}_{t_0}) = \mathbf{0}$ or $E(\mathbf{x}_{t_0}) \neq \mathbf{0}$, in both cases Equations 16-17 can be written more succinctly by

$$E(\mathbf{x}_t) = \sum_{k=t_0}^{t} \mathcal{A}_{t,k}\mathbf{B}_{k-1}\mathbf{u}_{k-1} \quad , \tag{18}$$

$$E(\mathbf{y}_t) = \mathbf{C}_t \sum_{k=t_0}^{t} \mathcal{A}_{t,k}\mathbf{B}_{k-1}\mathbf{u}_{k-1} + \mathbf{D}_t\mathbf{u}_t \quad , \tag{19}$$

where because of

$$E(\mathbf{x}_{t_0}) = \mathbf{B}_{t_0-1}\mathbf{u}_{t_0-1} \quad , \tag{20}$$

$$E(\mathbf{y}_{t_0}) = \mathbf{C}_{t_0}E(\mathbf{x}_{t_0}) + \mathbf{D}_{t_0}\mathbf{u}_{t_0} \quad , \tag{21}$$

the initial state mean $E(\mathbf{x}_{t_0})$ is modeled by means of an additional matrix $\mathbf{B}_{t_0-1}$, to be specified zero except, in the case of $E(\mathbf{x}_{t_0}) \neq \mathbf{0}$, for the elements corresponding to the unit input variable in $\mathbf{u}_{t_0-1}$. The value and identifiability of $E(\mathbf{x}_{t_0})$ depend on the choice of $\mathbf{D}_{t_0}$ as well as of the factor loading matrix $\mathbf{C}_{t_0}$. The choice of the latter additionally determines the value and identifiability of the initial state covariance matrix $\mathbf{\Phi}_{t_0} = E([\mathbf{x}_{t_0} - E(\mathbf{x}_{t_0})]\,[\mathbf{x}_{t_0} - E(\mathbf{x}_{t_0})]')$. In fact, these choices determine the measurement scales (origins and measurement units) of the latent state variables. For example, by specifying values 0 and 1 on specific places of, respectively, $\mathbf{D}_{t_0}$ and $\mathbf{C}_{t_0}$, the latent measurement scales are chosen equal to those of specific observed variables in $\mathbf{y}_{t_0}$. Special identification techniques are needed, however, to guarantee that the latent measurement scales maintain the same origins and measurement units across the whole time range (Oud et al., 1993, pp. 15-16).

For deriving the SEM model first write Equations 14-15 in the following form:

$$\begin{bmatrix} \mathbf{u}_t \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{t-1} & \mathbf{A}_{t-1} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{t-1} \\ \mathbf{x}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_t \\ \mathbf{w}_{t-1} \end{bmatrix} \quad , \tag{22}$$

$$\begin{bmatrix} \mathbf{u}_t \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D}_t & \mathbf{C}_t \end{bmatrix} \begin{bmatrix} \mathbf{u}_t \\ \mathbf{x}_t \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{v}_t \end{bmatrix} . \tag{23}$$

Next, collecting all input-variables in the input-vector $\mathbf{u}$ but specifying the constant input-variables (e.g. the unit input-variable) and other exactly linearly related input-variables only once in $\mathbf{u}$, and defining

$$\begin{aligned}
&\boldsymbol{\eta} = [\mathbf{u}'\,\mathbf{x}']' \text{ with } \mathbf{x} = [\mathbf{x}'_{t_0}\,\mathbf{x}'_{t_0+1}\,\ldots\,\mathbf{x}'_{t_0+T-1}]', \\
&\boldsymbol{\zeta} = [\mathbf{u}'\,\mathbf{w}']' \text{ with } \mathbf{w} = [[\mathbf{x}_{t_0} - E(\mathbf{x}_{t_0})]'\,\mathbf{w}'_{t_0}\,\ldots\,\mathbf{w}'_{t_0+T-2}]', \\
&\mathbf{y} = [\mathbf{u}'\,\mathbf{y}'_0]' \text{ with } \mathbf{y}_0 = [\mathbf{y}'_{t_0}\,\mathbf{y}'_{t_0+1}\,\cdots\,\mathbf{y}'_{t_0+T-1}]', \\
&\boldsymbol{\varepsilon} = [\mathbf{0}'\,\mathbf{v}']' \text{ with } \mathbf{v} = [\mathbf{v}'_{t_0}\,\mathbf{v}'_{t_0+1}\,\cdots\,\mathbf{v}'_{t_0+T-1}]'
\end{aligned}$$

the SEM model is derived as follows:

$$\begin{aligned}
\begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix} &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix} + \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \end{bmatrix} , \\
\boldsymbol{\eta} &= \quad B \quad\; \boldsymbol{\eta} \;+\; \boldsymbol{\zeta}
\end{aligned} \tag{24}$$

$$\begin{aligned}
\begin{bmatrix} \mathbf{u} \\ \mathbf{y}_0 \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{v} \end{bmatrix} , \\
\mathbf{y} &= \quad \Lambda \quad\; \boldsymbol{\eta} \;+\; \boldsymbol{\varepsilon}
\end{aligned} \tag{25}$$

where all parameter matrices $\mathbf{A}_{t-1}$, $\mathbf{B}_{t-1}$, $\mathbf{C}_t$, $\mathbf{D}_t$ are put on the appropriate places in $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, respectively. Notice that in $\mathbf{x}$ the initial state $\mathbf{x}_{t_0}$ gets zero rows in $\mathbf{A}$ but $\mathbf{B}_{t_0-1}$ in $\mathbf{B}$ for modeling its mean $E(\mathbf{x}_{t_0})$, which therefore is subtracted from $\mathbf{x}_{t_0}$ in $\mathbf{w}$. From Equations 24-25 one derives

$$\begin{aligned}
\begin{bmatrix} \mathbf{u} \\ \mathbf{y}_0 \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}+\mathbf{D} & \mathbf{C}(\mathbf{I}-\mathbf{A})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{v} \end{bmatrix} . \\
\mathbf{y} &= \qquad \Lambda(\mathbf{I}-B)^{-1} \qquad \boldsymbol{\zeta} \;+\; \boldsymbol{\varepsilon}
\end{aligned} \tag{26}$$

Because

$$
(\mathbf{I}-\mathbf{A})^{-1} = \begin{bmatrix}
\mathbf{I} & & & & \mathbf{0} \\
\mathbf{A}_{t_0} & \mathbf{I} & & & \\
\mathbf{A}_{t_0+1}\mathbf{A}_{t_0} & \mathbf{A}_{t_0+1} & \ddots & & \\
\vdots & \vdots & \ddots & \mathbf{I} & \\
\prod_{k=t_0+T-2}^{t_0} \mathbf{A}_k & \prod_{k=t_0+T-2}^{t_0+1} \mathbf{A}_k & & \mathbf{A}_{t_0+T-2} & \mathbf{I}
\end{bmatrix} \tag{27}
$$

and $E(\mathbf{w}) = \mathbf{0}$ it is easily seen that the means of the $\mathbf{y}_t$ in $\mathbf{y}_0$ of the SEM model equal those given by Equation 19 or 17.

Defining

$$
\mathbf{D}_0 \equiv \mathbf{C}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}+\mathbf{D} \quad \text{and} \quad \mathbf{C}_0 \equiv \mathbf{C}(\mathbf{I}-\mathbf{A})^{-1}
$$

Equation 26 becomes

$$
\begin{array}{ccccccc}
\begin{bmatrix} \mathbf{u} \\ \mathbf{y}_0 \end{bmatrix} & = & \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D}_0 & \mathbf{C}_0 \end{bmatrix} & \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \end{bmatrix} & + & \begin{bmatrix} \mathbf{0} \\ \mathbf{v} \end{bmatrix} & . \\
\mathbf{y} & = & \Lambda(\mathbf{I}-B)^{-1} & \zeta & + & \varepsilon &
\end{array} \tag{28}
$$

Writing the random vector $\mathbf{y}_0$ of the SEM model in terms of $\mathbf{D}_0$ and $\mathbf{C}_0$:

$$
\mathbf{y}_0 = \mathbf{D}_0\mathbf{u} + \mathbf{C}_0\mathbf{w} + \mathbf{v} \quad ,
$$

its mean and covariance matrix are found to be:

$$
\boldsymbol{\mu}_0 = E(\mathbf{y}_0) = \mathbf{D}_0\mathbf{u} \quad , \tag{29}
$$

$$
\begin{aligned}
\boldsymbol{\Sigma}_0 = cov(\mathbf{y}_0) &= E[(\mathbf{y}_0-\boldsymbol{\mu}_0)(\mathbf{y}_0-\boldsymbol{\mu}_0)'] \\
&= E[(\mathbf{C}_0\mathbf{w}+\mathbf{v})(\mathbf{C}_0\mathbf{w}+\mathbf{v})'] \\
&= \mathbf{C}_0\boldsymbol{\Psi}_0\mathbf{C}_0' + \boldsymbol{\Theta}_0
\end{aligned} \tag{30}
$$

where

$$
\boldsymbol{\Psi}_0 \equiv E(\mathbf{w}\mathbf{w}') \quad \text{and} \quad \boldsymbol{\Theta}_0 \equiv E(\mathbf{v}\mathbf{v}') \quad .
$$

The loglikelihood function then becomes

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{N}{2}\log \mid \boldsymbol{\Sigma}_0 \mid -\frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_{0i} - \boldsymbol{\mu}_{0i})'\boldsymbol{\Sigma}_0^{-1}(\mathbf{y}_{0i} - \boldsymbol{\mu}_{0i}) - \frac{p_0 N}{2}\log 2\pi \ , \quad (31)$$

where both $\mathbf{y}_{0i}$ and $\boldsymbol{\mu}_{0i}$ have subscript i because $\mathbf{u}$ may vary over subjects. In contrast to $\boldsymbol{\mu}_{0i}$, however, $\boldsymbol{\Sigma}_0$ is assumed common to all subjects. In fact, each subject in the sample is considered to be drawn from one of $N' \leq N$ distinct but, apart from the specific $\mathbf{u}$, equally distributed populations, having in particular $E(\mathbf{x}_{t_0})$, $\boldsymbol{\Phi}_{t_0}$ and all other parameter values equal. If q is the number of (fixed) elements in $\mathbf{u}$, the set of N' fixed values $\mathbf{u}_i$ in the sample must contain at least q linearly independent ones. An important advantage of $\mathbf{u}$ being fixed is that no distribution needs to be specified for its elements, which even need not be interval scale variables (e.g. income) but may also be dummy variables representing simple group membership (e.g. gender). Stochastic input variables can be handled too by adding them as extra elements to the state vector $\mathbf{x}_t$, while adding their observables as extra elements to $\mathbf{y}_t$ (Oud et al., 1990, pp. 400-401).

Most SEM programs do not maximize the loglikelihood function as given in Equation 31. However, it can be proven (see Appendix A) that minimizing the SEM fit function 7 on the basis of the sample moment matrix, augmented with the input variables, gives the same result as maximizing Equation 31.

State-trait model

An important aspect of the model is whether and towards which value or trajectory the state regresses. While in the zero means model (Equations 1-2) stabilizing feedback (all eigenvalues of $\mathbf{A}_{t-1}$ within the unit circle in the complex plane) implies that the state regresses towards the common mean $\mathbf{0}$, in the nonzero means model (Equations 14-15)

$$E(\mathbf{x}_t|\mathbf{x}_{t_0}) = \boldsymbol{\mathcal{A}}_{t,t_0}\mathbf{x}_{t_0} + \sum_{k=t_0}^{t-1}\boldsymbol{\mathcal{A}}_{t,k+1}\mathbf{B}_k\mathbf{u}_k \quad (32)$$

or

$$E(\mathbf{x}_t|\mathbf{x}_{t_0}) - \sum_{k=t_0}^{t-1}\boldsymbol{\mathcal{A}}_{t,k+1}\mathbf{B}_k\mathbf{u}_k = \boldsymbol{\mathcal{A}}_{t,t_0}\mathbf{x}_{t_0}$$

and stabilizing feedback from $t-1$ to t implies $||\mathcal{A}_{t,t_0}\mathbf{x}_{t_0}|| < ||\mathcal{A}_{t-1,t_0}\mathbf{x}_{t_0}||$ for all $\mathcal{A}_{t-1,t_0}\mathbf{x}_{t_0} \neq \mathbf{0}$ and

$$||E(\mathbf{x}_t|\mathbf{x}_{t_0}) - \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k|| < ||E(\mathbf{x}_{t-1}|\mathbf{x}_{t_0}) - \sum_{k=t_0}^{t-2} \mathcal{A}_{t-1,k+1}\mathbf{B}_k\mathbf{u}_k|| \quad \text{for all } \mathcal{A}_{t-1,t_0}\mathbf{x}_{t_0} \neq \mathbf{0} \ . \tag{33}$$

Starting from arbitrary initial state $\mathbf{x}_{t_0} \neq \mathbf{0}$ the expected state decreases between $t-1$ and t its Euclidian distance to the trajectory $\sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k$. Thus regression is towards

$$E(\mathbf{x}_t|\mathbf{x}_{t_0} = \mathbf{0}) = \sum_{k=t_0}^{t-1} \mathcal{A}_{t,k+1}\mathbf{B}_k\mathbf{u}_k \ , \tag{34}$$

which is the conditional zero-initial-state mean $E(\mathbf{x}_t|\mathbf{x}_{t_0} = \mathbf{0})$ or the unconditional mean $E(\mathbf{x}_t)$ for $E(\mathbf{x}_{t_0}) = \mathbf{0}$. This mean towards which the state regresses, is common again to all subjects, if the only input-variable is the unit input-variable. It may also be subpopulation specific, however. If one extra input-variable is, for example, gender, regression is towards the mean of the subpopulation of males or subpopulation of females. More generally, a subject's state regresses towards (egresses from) the mean of the subpopulation of subjects sharing the same input history $\mathbf{u}_{[t_0,t)}$. Notice that in the sample only one subject may happen to be present from such subpopulation.

The question arises whether one could specify a model which makes a subject regress to (egress from) its own mean instead of that of some (sub)population it shares the input history with. This is possible indeed by adding to the state equation constant (over time) random subjects effects $\boldsymbol{\xi}$:

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \boldsymbol{\xi} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{w}_{t-1} \ . \tag{35}$$

These constant random subject effects became very popular in econometric panel analysis (see e.g. Baltagi, 1995) and are sometimes called "trait" variables. A trait variable may be specified for each of the state variables. It can be characterized as a random (but constant over time) intercept term, to be contrasted to the fixed (but possibly time-varying) intercept, associated with the unit input-variable. Because of the specification $E(\boldsymbol{\xi}) = \mathbf{0}$, $\boldsymbol{\xi}$ may be viewed as the subject specific deviation from the common fixed intercept. It is assumed to be normally distributed over subjects.

In the state-trait model

$$E(\mathbf{x}_t|\mathbf{x}_{t_0},\boldsymbol{\xi}) = \boldsymbol{\mathcal{A}}_{t,t_0}\mathbf{x}_{t_0} + \sum_{k=t_0}^{t-1}\boldsymbol{\mathcal{A}}_{t,k+1}\boldsymbol{\xi} + \sum_{k=t_0}^{t-1}\boldsymbol{\mathcal{A}}_{t,k+1}\mathbf{B}_k\mathbf{u}_k \quad , \tag{36}$$

and a subject's state regresses towards the subject specific mean

$$E(\mathbf{x}_t|\mathbf{x}_{t_0}=\mathbf{0},\boldsymbol{\xi}) = \sum_{k=t_0}^{t-1}\boldsymbol{\mathcal{A}}_{t,k+1}\boldsymbol{\xi} + \sum_{k=t_0}^{t-1}\boldsymbol{\mathcal{A}}_{t,k+1}\mathbf{B}_k\mathbf{u}_k \quad , \tag{37}$$

which keeps a subject specific distance $\sum_{k=t_0}^{t-1}\boldsymbol{\mathcal{A}}_{t,k+1}\boldsymbol{\xi}$ from the (sub)population mean $\sum_{k=t_0}^{t-1}\boldsymbol{\mathcal{A}}_{t,k+1}\mathbf{B}_k\mathbf{u}_k$. So, while keeping the advantages of large N modeling, in a sense the state-trait model specifies for each of the subjects in each (sub)population a subject specific $N = 1$ model, causing regression or egression to be to or from this mean trajectory instead of the one of some arbitrary (sub)population which happens to be chosen by the researcher.

Reformulating the state-trait model as follows:

$$\begin{array}{ccccccccc} \begin{bmatrix} \mathbf{x}_t \\ \boldsymbol{\xi} \end{bmatrix} & = & \begin{bmatrix} \mathbf{A}_{t-1} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} & \begin{bmatrix} \mathbf{x}_{t-1} \\ \boldsymbol{\xi} \end{bmatrix} & + & \begin{bmatrix} \mathbf{B}_{t-1} \\ \mathbf{0} \end{bmatrix} & \mathbf{u}_{t-1} & + & \begin{bmatrix} \mathbf{w}_{t-1} \\ \mathbf{0} \end{bmatrix} \\ \mathbf{x}^\circ_t & = & \mathbf{A}^\circ_{t-1} & \mathbf{x}^\circ_{t-1} & + & \mathbf{B}^\circ_{t-1} & \mathbf{u}_{t-1} & + & \mathbf{w}^\circ_{t-1} \end{array} \tag{38}$$

$$\mathbf{y}_t = [\ \mathbf{C}_t \quad \mathbf{0}\ ]\ \mathbf{x}^\circ_t + \mathbf{D}_t\mathbf{u}_t + \mathbf{v}_t \quad , \tag{39}$$

makes clear that it is nothing more than a special case of the state space model and may be fitted accordingly into the SEM model and its parameters estimated by means of the SEM program. The constancy over time makes that $\boldsymbol{\xi}$ should also be considered part of the initial state $\mathbf{x}_{t_0}$

$$\mathbf{x}_{t_0} = \mathbf{A}_{t_0-1}\mathbf{x}_{t_0-1} + \boldsymbol{\xi} + \mathbf{B}_{t_0-1}\mathbf{u}_{t_0-1} + \mathbf{w}_{t_0-1} \quad ,$$

so that $\boldsymbol{\xi}$ and $\mathbf{x}_{t_0}$ in

$$\mathbf{x}_{t_0+1} = \mathbf{A}_{t_0}\mathbf{x}_{t_0} + \boldsymbol{\xi} + \mathbf{B}_{t_0}\mathbf{u}_{t_0} + \mathbf{w}_{t_0}$$

cannot be assumed to be uncorrelated and the initial state covariance matrix in the SEM model becomes:

$$\boldsymbol{\Phi}^\circ_{t_0} = \begin{bmatrix} \boldsymbol{\Phi}_{x_{t_0}} & \boldsymbol{\Phi}_{\xi,x_{t_0}} \\ \boldsymbol{\Phi}_{x_{t_0},\xi} & \boldsymbol{\Phi}_{\xi} \end{bmatrix} .$$

Significance tests on the existence of constant random subject effects can easily be performed by the SEM model: both the variances of the trait variables in $\mathbf{\Phi}_{\xi}$ and their covariances with the initial state variables in $\mathbf{\Phi}_{x_{t_0},\xi}$ are testable quantities in the SEM model, required to be different from $\mathbf{0}$ (Baltagi, 1995, p. 125). Being constant variables, the trait variables need not be repeated for successive time points in the SEM model but may be specified once, while trait variables with nonsignificant variances may be deleted.

The benefit of the recursive state space specification in Equations 38-39 is that the Kalman filter can be used to optimally estimate for each subject the constant value of each trait variable in addition to the changing values of each state variable. The (changing) Kalman filter estimates of the constant trait value become more precise as time proceeds (because of the constancy, however, the Kalman smoother does not improve upon the results of the Kalman filter). So, like the polynomial random effects model (Bryk & Raudenbush, 1987; Willett & Sayer, 1994), the state-trait model allows different kinds of random subject effects to be specified, tested and their variances and covariances as well as their values estimated. In contrast to polynomial curve fitting, however, the state-trait model gives a causal dynamic description of the curves.

Linking the Large N SEM Parameter Estimation Procedure to the Traditional $N = 1$ Procedure

In fact, different procedures for estimating the state space model parameters exist in the $N = 1$ case. Well-known are the prediction error decomposition procedure (Caines & Rissanen, 1974; Mehra, 1971; Schweppe, 1965), using the Kalman filter and the first procedure for which complete consistency results have been obtained (Caines, 1988, p. 411), and the more recently developed EM-prewhitening decomposition procedure (Dembo & Zeitouni, 1986; Shumway & Stoffer, 1982; Singer, 1990), which in addition to two decompositions uses the Kalman smoother. As important computational advantages are claimed for the latter procedure (Shumway & Stoffer, 1982, p. 255; Singer, 1990, p. 82) and the use of EM additionally solves the missing value problem, the EM-prewhitening decomposition procedure will be concentrated upon here. Starting from the SEM model loglikelihood $\ell(\boldsymbol{\theta}|\mathbf{Y})$ the procedure is derived in two steps. The first step introduces the EM decomposition and the second step the prewhitening decomposition, which is based on the extra assumption of time-invariance.

EM decomposition

As an alternative for the maximization of $\ell(\boldsymbol{\theta}|\mathbf{Y})$ in Equation 5 or 31 (minimization of F_{ML} in Equation 7) EM can be used even when $\mathbf{Y}$ contains complete data. Combining all latent states in the $mT \times N$ matrix $\mathbf{X}$ (m the number of state variables in state vector $\mathbf{x}_t$ and mT the total number of state variables in $\boldsymbol{\eta}$), it consists of considering the latent unknown $\mathbf{X}$ as the missing data. Bayes' formula enables one to decompose the complete data loglikelihood $\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})$ into separate loglikelihoods for the structural (state equation) and measurement (output equation) parts of the SEM model:

$$\begin{aligned}\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X}) &= \ell(\boldsymbol{\theta}|\mathbf{X}) + \ell(\boldsymbol{\theta}|\mathbf{Y}|\mathbf{X}) \\ &= \ell(\boldsymbol{\theta}_{B,\Psi}|\mathbf{X}) + \ell(\boldsymbol{\theta}_{\Lambda,\Theta}|\mathbf{Y}|\mathbf{X}) \qquad (40)\end{aligned}$$

Writing $\boldsymbol{\Sigma}_\eta = (\mathbf{I}-B)^{-1}\boldsymbol{\Psi}(\mathbf{I}-B')^{-1}$, $\mathbf{S}_\eta = \frac{1}{N}\mathbf{X}\,\mathbf{X}'$, and $\mathbf{S}_\varepsilon = \frac{1}{N}(\mathbf{Y}-\boldsymbol{\Lambda}\mathbf{X})(\mathbf{Y}-\boldsymbol{\Lambda}\mathbf{X})' = \mathbf{S} + \boldsymbol{\Lambda}\mathbf{S}_\eta\boldsymbol{\Lambda}' - \mathbf{S}_{y,x}\boldsymbol{\Lambda}' - \boldsymbol{\Lambda}\mathbf{S}_{x,y}$ one derives the separate loglikelihoods:

$$\ell(\boldsymbol{\theta}_{B,\Psi}|\mathbf{X}) = -\frac{N}{2}\log|\,\boldsymbol{\Sigma}_\eta\,| - \frac{N}{2}\mathrm{tr}(\mathbf{S}_\eta\boldsymbol{\Sigma}_\eta^{-1}) - \frac{mT.N}{2}\log 2\pi\ , \qquad (41)$$

$$\ell(\boldsymbol{\theta}_{\Lambda,\Theta}|\mathbf{Y}|\mathbf{X}) = -\frac{N}{2}\log|\,\boldsymbol{\Theta}\,| - \frac{N}{2}\mathrm{tr}(\mathbf{S}_\varepsilon\boldsymbol{\Theta}^{-1}) - \frac{pN}{2}\log 2\pi\ . \qquad (42)$$

The measurement part in Equation 42 now takes the form of a restricted regression analysis problem.

The conditional loglikelihood expectation to be maximized iteratively in EM decomposes accordingly:

$$\begin{aligned}E_{\mathbf{X}}[\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})|\mathbf{Y},\hat{\boldsymbol{\theta}}_r] = \;& E_{\mathbf{X}}[\ell(\boldsymbol{\theta}_{B,\Psi}|\mathbf{X})|\mathbf{Y},\hat{\boldsymbol{\theta}}_r] + \\ & E_{\mathbf{X}}[\ell(\boldsymbol{\theta}_{\Lambda,\Theta}|\mathbf{Y}|\mathbf{X})|\mathbf{Y},\hat{\boldsymbol{\theta}}_r] \qquad (43)\end{aligned}$$

Implementation of the EM algorithm by maximizing Equation 43 in the M-step requires substituting in Equations 41 and 42 the unknown $\mathbf{S}_\eta$ and the unknown $\mathbf{X}$ in $\mathbf{S}_\varepsilon$ by, respectively, $\mathbf{S}_{\eta,r+1} = E_{\mathbf{X}}(\mathbf{S}_\eta|\mathbf{Y},\hat{\boldsymbol{\theta}}_r)$ and $\mathbf{X}_{r+1} = E_{\mathbf{X}}(\mathbf{X}|\mathbf{Y},\hat{\boldsymbol{\theta}}_r)$, leading to the Kalman smoother in the E-step (Jansen & Oud, 1995). It can be proven that (Singer, 1993) the successive EM estimates $\hat{\boldsymbol{\theta}}_{r+1} = [\hat{\boldsymbol{\theta}}'_{B,\Psi}\ \hat{\boldsymbol{\theta}}'_{\Lambda,\Theta}]'_{r+1} =$ argmax $E_{\mathbf{X}}[\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})|\mathbf{Y},\hat{\boldsymbol{\theta}}_r]$ converge to the maximum likelihood estimate argmax $\ell(\boldsymbol{\theta}|\mathbf{Y})$. This implies that the separate maximization of Equations 41 and 42 (e.g. by means of the SEM program but also by means of any observed variables structural equation modeling program), using and iteratively inserting Kalman smoother estimates, leads to the same parameter estimates as the direct maximization of Equation 5 or 31. Disadvantages of the indirect decomposition procedure are that no cross-restrictions between the parameters in $\boldsymbol{\theta}_{B,\Psi}$ and $\boldsymbol{\theta}_{\Lambda,\Theta}$ are possible, that no correct standard errors are provided which must be computed separately (using $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ according to Equation 6 and computing the information matrix as explained e.g. in Jöreskog, 1973), and that the Kalman smoother estimates need to be computed for all subjects, not only for subjects with missing data in $\mathbf{Y}$. In the case of missing data in $\mathbf{Y}$, $E_{\mathbf{X}}[\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})|\mathbf{Y},\hat{\boldsymbol{\theta}}_r]$ in Equation 43 is replaced by $E_{Y_{mis},\mathbf{X}}[\ell(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})|Y_{obs},\hat{\boldsymbol{\theta}}_r]$, and therefore $\mathbf{S}$ and $\mathbf{S}_{y,x}$ in Equation 42 by, respectively, $\mathbf{S}_{r+1} = E_{Y_{mis}}(\mathbf{S}|Y_{obs},\hat{\boldsymbol{\theta}}_r)$ mentioned previously and $\mathbf{S}_{y,x,r+1} = E_{Y_{mis},\mathbf{X}}(\mathbf{S}_{y,x}|Y_{obs},\hat{\boldsymbol{\theta}}_r)$.

Prewhitening decomposition

The next step is the prewhitening decomposition, which is motivated by the fact that in the $N = 1$ case no replications are left in the loglikelihood functions of Equations 41-42. Replications must be introduced in an alternative way. Instead of replicating over subjects one could replicate over time points, replacing N by T in Equations 41-42. Evidently, this cannot be done without the assumption that the data at successive time points obey the same parameter values, that is, are produced by a time-invariant model. This solves only part of the problem, however. The loglikelihood function also assumes replications to be independently distributed. Some authors (e.g. Molenaar, 1985; Molenaar, de Gooijer, & Schmitz, 1992) have performed LISREL analyses, using Equation 7 instead of EM, by filling out the N columns of the data matrix with stretches from one and the same time series, the stretches showing overlap in time (so-called windowing technique). This procedure, however, introduces strongly dependent data in the data matrix (different columns containing exactly the same data), the dependence remaining unaccounted for in the formulation of the loglikelihood function as used by LISREL and other SEM programs. It causes the maximum likelihood property to be lost and produces invalid standard errors and significance tests (Jöreskog, 1973, p. 88; Jöreskog & Sörbom, 1989, p. 162). There is, however, a solution outside of SEM, called prewhitening, which yields correct maximum likelihood results in the $N = 1$ case. The term prewhitening stresses that, in contrast to data produced by random sampling, time-dependent data need first transformation to independence before being usable as replications in the loglikelihood function. It should be noted that no new or hidden assumptions are involved in the derivation of the transformation except those of the time-invariant state space model.

The prewhitening decomposition of Equation 41 for a single subject, derived by introducing time-invariance ($\mathbf{A}_{t_0} = \mathbf{A}_{t_0+1} = \cdots = \mathbf{A}_{t_0+T-2} = \mathbf{A}$ and $\mathbf{Q}_{t_0} = \mathbf{Q}_{t_0+1} = \cdots = \mathbf{Q}_{t_0+T-2} = \mathbf{Q}$), is as follows (see Appendix B):

$$
\begin{aligned}
\ell(\boldsymbol{\theta}_{B,\Psi}|\mathbf{x}) \quad &= \\
&-\frac{1}{2}\log|\boldsymbol{\Phi}_{t_0}| - \frac{1}{2}\mathbf{x}'_{t_0}\boldsymbol{\Phi}^{-1}_{t_0}\mathbf{x}_{t_0} \\
&-\frac{T-1}{2}\log|\mathbf{Q}| - \frac{1}{2}\sum_{j=1}^{T-1}(\mathbf{x}_{t_0+j} - \mathbf{A}\mathbf{x}_{t_0+j-1})'\mathbf{Q}^{-1}(\mathbf{x}_{t_0+j} - \mathbf{A}\mathbf{x}_{t_0+j-1}) \\
&- constant \quad . \qquad (44)
\end{aligned}
$$

In an analogous but somewhat simpler manner than in Appendix B one finds that the regression analysis problem over subjects in Equation 42 becomes in decomposed form over time points for a single subject, assuming time-invariance ($\mathbf{C}_{t_0} = \mathbf{C}_{t_0+1} = \cdots = \mathbf{C}_{t_0+T-1} = \mathbf{C}$ and $\mathbf{R}_{t_0} = \mathbf{R}_{t_0+1} = \cdots = \mathbf{R}_{t_0+T-1} = \mathbf{R}$),

$$
\begin{aligned}
\ell(\boldsymbol{\theta}_{\Lambda,\Theta}|\mathbf{y}|\mathbf{x}) \quad &= \quad -\frac{T}{2}\log|\mathbf{R}| - \frac{1}{2}\sum_{j=0}^{T-1}(\mathbf{y}_{t_0+j} - \mathbf{C}\mathbf{x}_{t_0+j})'\mathbf{R}^{-1}(\mathbf{y}_{t_0+j} - \mathbf{C}\mathbf{x}_{t_0+j}) \\
&\quad - constant \quad . \qquad (45)
\end{aligned}
$$

The data transformations are seen to consist in the subtraction of $\mathbf{A}\mathbf{x}_{t_0+j-1}$ and $\mathbf{C}\mathbf{x}_{t_0+j}$ from $\mathbf{x}_{t_0+j}$ and $\mathbf{y}_{t_0+j}$, respectively. Together Equations 44-45 form the complete data loglikelihood as given, for example, by Shumway and Stoffer (1982, p. 256), to be maximized by EM. Note that because of the input-effects in the nonzero means case (Equations 14-15), instead of $\mathbf{A}\mathbf{x}_{t_0+j-1}$ and $\mathbf{C}\mathbf{x}_{t_0+j}$, $\mathbf{A}\mathbf{x}_{t_0+j-1} + \mathbf{B}\mathbf{u}_{t_0+j-1}$ and $\mathbf{C}\mathbf{x}_{t_0+j} + \mathbf{D}\mathbf{u}_{t_0+j}$ have to be subtracted, respectively.

As indicated, the SEM program does not work in the $N = 1$ case and so for this case cannot be proven to give the same result. However, Equations 44-45 generalize to arbitrary N as follows:

$$
\begin{aligned}
&\ell(\boldsymbol{\theta}_{B,\Psi}|\mathbf{X}) = \\
&-\frac{N}{2}\log|\boldsymbol{\Phi}_{t_0}| - \frac{1}{2}\sum_{i=1}^{N}\mathbf{x}'_{i,t_0}\boldsymbol{\Phi}^{-1}_{t_0}\mathbf{x}_{i,t_0} \\
&-\frac{N(T-1)}{2}\log|\mathbf{Q}| - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{T-1}(\mathbf{x}_{i,t_0+j} - \mathbf{A}\mathbf{x}_{i,t_0+j-1})'\mathbf{Q}^{-1}(\mathbf{x}_{i,t_0+j} - \mathbf{A}\mathbf{x}_{i,t_0+j-1}) \\
&- constant \quad , \qquad (46)
\end{aligned}
$$

$$\ell(\boldsymbol{\theta}_{\Lambda,\Theta}|\mathbf{Y}|\mathbf{X}) =$$
$$-\frac{NT}{2}\log|\mathbf{R}| - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=0}^{T-1}(\mathbf{y}_{i,t_0+j} - \mathbf{C}\mathbf{x}_{i,t_0+j})'\mathbf{R}^{-1}(\mathbf{y}_{i,t_0+j} - \mathbf{C}\mathbf{x}_{i,t_0+j})$$
$$- \mathit{constant} \quad . \tag{47}$$

or written in terms of the sample covariance matrices $\mathbf{S}_{t_0} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{i,t_0}\mathbf{x}'_{i,t_0}$, $\mathbf{S}_w = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{j=1}^{T-1}\mathbf{w}_{i,t_0+j}\mathbf{w}'_{i,t_0+j}$, and $\mathbf{S}_v = \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=0}^{T-1}\mathbf{v}_{i,t_0+j}\mathbf{v}'_{i,t_0+j}$ for initial state $\mathbf{x}_{i,t_0}$ and transformed $\mathbf{w}_{i,t_0+j} = \mathbf{x}_{i,t_0+j} - \mathbf{A}\mathbf{x}_{i,t_0+j-1}$ and $\mathbf{v}_{i,t_0+j} = \mathbf{y}_{i,t_0+j} - \mathbf{C}\mathbf{x}_{i,t_0+j}$:

$$\begin{aligned}\ell(\boldsymbol{\theta}_{B,\Psi}|\mathbf{X}) &= -\frac{N}{2}\log|\boldsymbol{\Phi}_{t_0}| - \frac{N}{2}\mathrm{tr}(\mathbf{S}_{t_0}\boldsymbol{\Phi}_{t_0}^{-1}) \\ &\quad -\frac{N(T-1)}{2}\log|\mathbf{Q}| - \frac{N(T-1)}{2}\mathrm{tr}(\mathbf{S}_w\mathbf{Q}^{-1}) - \mathit{constant} \quad ,\end{aligned} \tag{48}$$

$$\ell(\boldsymbol{\theta}_{\Lambda,\Theta}|\mathbf{Y}|\mathbf{X}) = -\frac{NT}{2}\log|\mathbf{R}| - \frac{NT}{2}\mathrm{tr}(\mathbf{S}_v\mathbf{R}^{-1}) - \mathit{constant} \quad . \tag{49}$$

Putting Equations 48-49 instead of Equations 41-42 in Equation 43 and maximizing the result by EM constitutes the EM-prewhitening decomposition procedure for arbitrary N. A computer program applying this procedure is described by Singer (1990, 1993). It differs only from the EM decomposition procedure (Equations 41-42 in Equation 43) in the M step. The derivation of both as alternative maximization procedures of the SEM model loglikelihood proves that, for a time-invariant SEM model and a sample size N large enough to make $\mathbf{S}$ in the SEM procedure positive definite, the parameter estimates at the maximum in both procedures coincide with those of the SEM procedure (Equation 5 or 31 as handled by minimization of Equation 7).

Comparison of the prewhitening Equations 48-49 with the non-prewhitening Equations 41-42 and with the original SEM Equation 5 or 31 shows that (1) the number of replications increases from N (non-prewhitening) to NT in Equation 49 and $N(T-1)$ in Equation 48, both over subjects and time points simultaneously, (2) the matrices of order $mT \times mT$ or $rT \times rT$ (m the number of state variables and r the number of output variables per time point), especially $rT \times rT$ matrix $\mathbf{S}$, vanish, collapsing into much smaller matrices of order $m \times m$ (Equation 48) and $r \times r$ (Equation 49), (3) as a result of the initial state $\mathbf{x}_{i,t_0}$ remaining necessarily untransformed, its covariance matrix $\mathbf{\Phi}_{t_0}$ keeps a separate part in Equation 48, being estimated on the basis of N replications only.

The last point implies a typical problem for the $N = 1$ case. Although the initial state can be estimated by means of the Kalman filter and smoother in the $N = 1$ case, by lack of replications its mean and covariance matrix cannot be estimated or only extremely unreliably: fixing the mean leaves only a single replication for estimating the covariance matrix and vice versa. In practice $N = 1$ research is forced to fix both the initial mean and the initial covariance matrix at some arbitrarily chosen values (Shumway & Stoffer, 1982, p. 257). This is one of the reasons an asymptotically stable model has to be assumed in the $N = 1$ case, to guarantee that the influence of the initial value choices dies out over time, requiring also the time series available for analysis to be quite long to get reasonably reliable estimates (T large).

Pros and Cons of SEM State Space Modeling

Considering the pros and cons of SEM state space modeling, it is clear that the availability of data from a single or few subjects excludes SEM as an option. Many research problems in the social and behavioral sciences require generalization of the results to a well-defined population, however, and thus use samples with large N. For large N, the drawbacks of the decompositions as used in the EM-prewhitening decomposition procedure are the advantages of SEM. First of all, SEM offers extreme freedom in the choice of time-varying (nonstationary) models. This also is of crucial importance in social and behavioral science, because the causal mechanisms governing human development as well as the available measurement instruments and their characteristics typically change over the life span. For example, there may be increase or decrease in the autoregressive and cross-lagged effects of the state variables over time and there may be change in the strength of the input effects as well. In contrast to SEM, the time-invariance requirement of the prewhitening decomposition rules out the possibility of estimating and testing the existence of such changing parameter values over time.

Because SEM replicates exclusively over subjects, the number of time points may be taken arbitrarily small. Also, there is no need considering an asymptotic theory for $T \to \infty$. Especially the asymptotic stability assumption, needed additionally to the time-invariance in consistency proofs for the $N = 1$

case (Caines, 1988, p. 412), becomes superfluous. Instead the standard theory for $N \to \infty$ is applicable. This does not take away the merits of a stable model. However, in SEM state space modeling asymptotic stability does not enter as a prerequisite for consistent parameter estimation, nor is a large T required.

Another important advantage of the large N characteristic of SEM is the estimability of the initial state mean and covariance matrix as a natural part of the procedure, thus bypassing the unsatisfactory fixation at arbitrarily chosen values.

One of the consequences of the prewhitening decomposition is the disappearance of $\mathbf{S}$. However, it is just by means of $\mathbf{S}$ that the specified model (e.g the state space model with or without the time-invariance assumptions) is tested in SEM. $\mathbf{S}$ is used as the estimated $\mathbf{\Sigma}$ of a saturated model, to which the one of the specified model is compared. The residuals or difference between $\mathbf{S}$ and the estimated $\mathbf{\Sigma}$ of the specified model give a detailed insight into the nature of misfit, if it occurs. High off-diagonal residuals, for example, may give indications for choosing a more appropriate ARMA-structure. High residuals at specific time points may suggest relaxing the time-invariance assumption for those time points. In fact, the SEM fit function F_{ML} (see Equation 7) is a discrepancy measure, measuring the difference (times $-\frac{2}{N}$) between the specified model loglikelihood and the saturated model loglikelihood, the latter having $|\mathbf{\Sigma}| = |\mathbf{S}|$ and $\mathrm{tr}(\mathbf{S\Sigma^{-1}}) = p$ at its maximum. Most testing and fit measures are based on this discrepancy measure. In contrast, the prewhitening decomposition presupposes the correctness of the state space model and time-invariance assumptions, but the collapsing of $\mathbf{S}$ as a result of the decomposition makes it impossible to test these and alternative assumptions against the general saturated model.

A special feature of SEM is the optional specification of structural parameters or instantaneous effects between state variables. These became very popular in econometric and social science models. From the structural form the state space or so-called reduced form can be derived. In addition to the importance of the structural form itself, however, the reduced form estimate with the structural restrictions incorporated leads typically to a considerable precision gain over the direct reduced form estimate (Bergstrom, 1984, pp. 1146-1147).

Three more advantages of the SEM procedure relate specifically to the EM decomposition. This splits and handles the loglikelihood in two separate parts linked only by the $mT \times N$ state matrix $\mathbf{X}$. By deducing the explicit form of $\mathbf{X}$ out of the model and integrating the two parts in a single loglikelihood function, SEM enables to skip the repeated, time consuming estimation of $\mathbf{X}$ (except for subjects with missing data), and additionally allows cross-restrictions to be imposed between parameters in the two parts and standard errors to be computed. Standard errors computed within the separate parts are incorrect.

Finally, although in practice most panel data sets combine large N with a relatively small number of observation time points T, a possible problem with SEM should be considered for large T. While the prewhitening decomposition leads to $m \times m$ and $r \times r$ matrices which do not grow for increasing T, the SEM matrices do. Depending on the hardware used, the SEM program could run out of memory space for large T. If this occurs, one of the possible solutions is dividing the time axis in parts and analyzing each corresponding part of the total data set separately: $t_0 \ldots t_{\tau_1}$, $t_{\tau_1} \ldots t_{\tau_2}, t_{\tau_2} \ldots t_{\tau_3}$, etc. The connection between the model parts is made by analyzing the last time point of each part (t_{τ_1}, t_{τ_2}, etc.) again as the initial time point of the next part. Not much is lost in this procedure, if there are no cross-restrictions between the parameters of the model parts and the parts are chosen not too small. The result for each part analysis, taken separately, is full information maximum likelihood, although the result of all parts taken together is only limited information maximum likelihood. The impossibility of applying cross-restrictions between the model parts is a serious limitation in practice, however.

The same results would be obtained, if the model parts and corresponding data sets are analyzed simultaneously in a multi-sample SEM analysis as performed, for example, by the LISREL program (Jöreskog & Sörbom, 1989, pp. 255-272). The use of the multi-sample analysis is strongly recommended, because it has the additional advantage of allowing cross-restrictions between the parameters of the part models, so that from the modeling perspective nothing is lost in comparison to the analysis of the total time axis and corresponding data set in a one-sample analysis. The gain of the multi-sample analysis in memory space is considerable. For example, supposing there are k equally sized parts, the total number of elements in the k part matrices of overall $\mathbf{S}$ is only $\frac{1}{k}$ the number of elements in $\mathbf{S}$. Again information is left out in the computation of the loglikelihood, but mainly by neglecting distant off-diagonal elements in overall $\mathbf{S}$, which are often known to be virtually zero a priori.

Evidently, the loss of information in the multi-sample analysis solution is just a consequence of the fact that the data sets for the model parts originate from one and the same group of subjects and not from independently drawn samples. This suggests another, statistically more appropriate solution, which turns out to have other important advantages as well. It consists in substituting the data sets from the same group of subjects by data sets from truly independent samples of subjects, either drawn from the same population at different points in time or from differently aged populations at the same point in time. The latter case gives rise to the so-called overlapping cohort design, in which the age intervals of the samples in the investigation are chosen overlapping. In both cases the multi-sample analysis gives full information maximum likelihood results, thereby solving the problem of building a large T model by SEM in a statistically appropriate way. A perhaps more important reason for choosing the overlapping cohort design is the possibility to collect the data sets for the model parts simultaneously, thereby shortening considerably the data collection period needed for state space model parameter estimation (Oud et al., 1993, pp. 18-19). It builds the large T model in only a fraction of the time period covered by the model. Moreover, by giving contiguous pairs of cohorts more overlap in time than just the last time point of each first and the initial time point of each second cohort, cohort effects become testable and thereby the assumption of the model parts originating from equal populations except for age (Jansen & Oud, 1996).

Discussion

In addition to the SEM procedure, several alternative procedures for estimating the state space model parameters were explained. In some cases, as when N is small or the model is time-varying, the choice is limited. For the cases that several procedures can be used, we stressed the advantages of the SEM procedure. However, if possible, it may be wise to apply different procedures to the same data set and compare the results. One reason for this is the possibility of solutions in nonadmissable regions of the parameter space. It is well-known, for example, that LISREL and other SEM programs may produce negative variance estimates, especially in the case of a badly fitting model. Some of these nonadmissable solutions are avoided in the EM decomposition procedure, because Kalman smoother values are actually inserted for the latent state variables and therefore no negative variance estimate nor, for example, a covariance estimate exceeding the estimates of both variances involved may show up. So, applying different procedures and checking whether the results are the same diminishes the risk of finding and taking seriously nonadmissible solutions, local maxima, and other abnormal results.

APPENDIX A: PROOF THAT MINIMIZATION OF SEM FIT FUNCTION 7 USING THE AUGMENTED MOMENT MATRIX EQUALS MAXIMIZATION OF LOGLIKELIHOOD FUNCTION 31

Write the augmented sample moment matrix

$$\mathbf{S}_{(p_0+q)\times(p_0+q)} = \tfrac{1}{N}\mathbf{Y}\ \mathbf{Y}' = \begin{bmatrix} \frac{1}{N}\sum\limits_{i=1}^{N}\mathbf{u}_i\mathbf{u}_i' & \frac{1}{N}\sum\limits_{i=1}^{N}\mathbf{u}_i\mathbf{y}_{0i}' \\ \frac{1}{N}\sum\limits_{i=1}^{N}\mathbf{y}_{0i}\mathbf{u}_i' & \frac{1}{N}\sum\limits_{i=1}^{N}\mathbf{y}_{0i}\mathbf{y}_{0i}' \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_u & \mathbf{S}_{u,y_0} \\ \mathbf{S}_{y_0,u} & \mathbf{S}_{y_0} \end{bmatrix},$$

and derive for $\mathbf{\Sigma}$ in Equation 7 using Equations 28-30

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{\Lambda}(\mathbf{I}-B)^{-1}\mathbf{\Psi}(\mathbf{I}-B')^{-1}\mathbf{\Lambda}' + \mathbf{\Theta} \\ &= \begin{bmatrix} \mathbf{\Phi}_u & \mathbf{\Phi}_u\mathbf{D}_0' \\ \mathbf{D}_0\mathbf{\Phi}_u & \mathbf{D}_0\mathbf{\Phi}_u\mathbf{D}_0' + \mathbf{\Sigma}_0 \end{bmatrix}, \end{aligned}$$

$$|\,\mathbf{\Sigma}\,| = |\,\mathbf{\Phi}_u\,|\ |\,\mathbf{\Sigma}_0\,| \ ;$$

$$\log|\,\mathbf{\Sigma}\,| = \log|\,\mathbf{\Phi}_u\,| + \log|\,\mathbf{\Sigma}_0\,| \ .$$

Hence $\log|\,\mathbf{\Sigma}\,|$ in Equation 7 differs only a constant $\log|\,\mathbf{\Phi}_u\,|$ from $\log|\,\mathbf{\Sigma}_0\,|$ in Equation 31, while

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \mathbf{\Phi}_u^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{D}_0' \\ \mathbf{I} \end{bmatrix} \mathbf{\Sigma}_0^{-1}\,[-\mathbf{D}_0\ \ \mathbf{I}] \ ,$$

$$\begin{aligned} \mathrm{tr}(\mathbf{S}\mathbf{\Sigma}^{-1}) &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_i'\mathbf{\Sigma}^{-1}\mathbf{y}_i) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left([\mathbf{u}_i'\ \ \mathbf{y}_{0i}']\ \mathbf{\Sigma}^{-1}\begin{bmatrix}\mathbf{u}_i \\ \mathbf{y}_{0i}\end{bmatrix}\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{u}_i'\mathbf{\Phi}_u^{-1}\mathbf{u}_i) + \frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_{0i}-\mathbf{D}_0\mathbf{u}_i)'\mathbf{\Sigma}_0^{-1}(\mathbf{y}_{0i}-\mathbf{D}_0\mathbf{u}_i) \\ &= \mathrm{tr}(\mathbf{\Phi}_u\mathbf{\Phi}_u^{-1}) + \frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_{0i}-\mathbf{D}_0\mathbf{u}_i)'\mathbf{\Sigma}_0^{-1}(\mathbf{y}_{0i}-\mathbf{D}_0\mathbf{u}_i) \\ &= q + \frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_{0i}-\boldsymbol{\mu}_{0i})'\mathbf{\Sigma}_0^{-1}(\mathbf{y}_{0i}-\boldsymbol{\mu}_{0i}) \ . \end{aligned}$$

Therefore, $-\frac{N}{2}$ times the SEM fit function F_{ML} (Equation 7) is equal to the loglikelihood function (Equation 31) plus a constant.

APPENDIX B: DERIVING EM-PREWHITENING DECOMPOSED LOGLIKELIHOOD FUNCTION 44 FROM EM DECOMPOSED FORM 41 FOR A SINGLE SUBJECT UNDER TIME-INVARIANCE

As in Equation 41

$$\mathbf{\Sigma}_\eta = (\mathbf{I} - B)^{-1}\mathbf{\Psi}(\mathbf{I} - B')^{-1}$$

with $(\mathbf{I} - B)^{-1}$ triangular (see the matrix in Equation 27) and

$$\mathbf{\Psi} = \begin{bmatrix} \mathbf{\Phi}_{t_0} & & & & \mathbf{0} \\ & \mathbf{Q}_{t_0} & & & \\ & & \mathbf{Q}_{t_0+1} & & \\ & & & \ddots & \\ \mathbf{0} & & & & \mathbf{Q}_{t_0+T-2} \end{bmatrix},$$

by introducing time invariance for the parameters in Equation 41 ($\mathbf{A}_{t_0} = \mathbf{A}_{t_0+1} = \cdots = \mathbf{A}_{t_0+T-2} = \mathbf{A}$ and $\mathbf{Q}_{t_0} = \mathbf{Q}_{t_0+1} = \cdots = \mathbf{Q}_{t_0+T-2} = \mathbf{Q}$) one finds that $\log|\mathbf{\Sigma}_\eta|$ in Equation 41 decomposes in

$$\log|\mathbf{\Sigma}_\eta| = \log|\mathbf{\Psi}| = \log|\mathbf{\Phi}_{t_0}| + (T-1)\log|\mathbf{Q}| \ .$$

Next, as in

$$\mathbf{\Sigma}_\eta^{-1} = (\mathbf{I} - B')\mathbf{\Psi}^{-1}(\mathbf{I} - B)$$

$$\mathbf{I} - B' = \begin{bmatrix} \mathbf{I} & -\mathbf{A}'_{t_0} & & & \mathbf{0} \\ & \mathbf{I} & -\mathbf{A}'_{t_0+1} & & \\ & & \mathbf{I} & \ddots & \\ & & & \ddots & -\mathbf{A}'_{t_0+T-2} \\ \mathbf{0} & & & & \mathbf{I} \end{bmatrix}$$

and

$$\mathbf{\Psi}^{-1} = \begin{bmatrix} \mathbf{\Phi}_{t_0}^{-1} & & & & \mathbf{0} \\ & \mathbf{Q}_{t_0}^{-1} & & & \\ & & \mathbf{Q}_{t_0+1}^{-1} & & \\ & & & \ddots & \\ \mathbf{0} & & & & \mathbf{Q}_{t_0+T-2}^{-1} \end{bmatrix},$$

while in the $N = 1$ case

$$\mathbf{S}_\eta = \mathbf{x}\mathbf{x}' = [\mathbf{x}'_{t_0}\ \mathbf{x}'_{t_0+1}\ \dots\ \mathbf{x}'_{t_0+T-1}]'\ [\mathbf{x}'_{t_0}\ \mathbf{x}'_{t_0+1}\ \dots\ \mathbf{x}'_{t_0+T-1}]\ ,$$

$$\mathrm{tr}(\mathbf{S}_\eta \boldsymbol{\Sigma}_\eta^{-1}) = [\mathbf{x}'_{t_0}\ \mathbf{x}'_{t_0+1}\ \dots\ \mathbf{x}'_{t_0+T-1}](\mathbf{I} - \boldsymbol{B}')\boldsymbol{\Psi}^{-1}(\mathbf{I} - \boldsymbol{B})[\mathbf{x}'_{t_0}\ \mathbf{x}'_{t_0+1}\ \dots\ \mathbf{x}'_{t_0+T-1}]'$$

becomes

$$\mathrm{tr}(\mathbf{S}_\eta \boldsymbol{\Sigma}_\eta^{-1}) = \mathbf{x}'_{t_0} \boldsymbol{\Phi}_{t_0}^{-1} \mathbf{x}_{t_0} + \sum_{j=1}^{T-1} (\mathbf{x}_{t_0+j} - \mathbf{A}\mathbf{x}_{t_0+j-1})' \mathbf{Q}^{-1} (\mathbf{x}_{t_0+j} - \mathbf{A}\mathbf{x}_{t_0+j-1})\ .$$

Substituting $\log |\boldsymbol{\Sigma}_\eta|$ and $\mathrm{tr}(\mathbf{S}_\eta \boldsymbol{\Sigma}_\eta^{-1})$ in 41 one finds the decomposed form of the latter in Equation 44.

References

Akaike, H. (1974). Stochastic theory of minimal realization. *IEEE Transactions on Automatic Control, 19*, 667-674.

Baltagi, B.H. (1995). *Econometric analysis of panel data.* Chichester: Wiley.

Bergstrom, A.R. (1984). Continuous time stochastic models and issues of aggregation over time. In Z. Griliches & M.D. Intriligator (Eds.), *Handbook of econometrics: Vol. 2.* (pp. 1145-1212). Amsterdam: North-Holland.

Bryk, A.S., & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147-158.

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

Caines, P.E., & Rissanen, J. (1974). Maximum likelihood estimation of parameters in multivariate Gaussian processes. *IEEE Transactions on Information Theory, 20*, 102-104.

Deistler, M. (1985). General structure and parametrization of ARMA and state-space systems and its relation to statistical problems. In E.J. Hannan, P.R. Krishnaiah, & M.M. Rao (Eds.), *Handbook of statistics: Vol. 5. Time series in the time domain* (pp. 257-277). Amsterdam: North-Holland.

Dembo, A., & Zeitouni, O. (1986). Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. *Stochastic Processes and their Applications, 23*, 91-113.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Ser. B, 39*, 1-38.

Desoer, C.A. (1970). *Notes for a second course on linear systems.* New York: Van Nostrand Reinhold.

Jansen, R.A.R.G., & Oud, J.H.L. (1995). Longitudinal LISREL model estimation from incomplete panel data using the EM algorithm and the Kalman smoother. *Statistica Neerlandica, 49*, 362-377.

Jansen, R.A.R.G., & Oud, J.H.L. (1996). *Nonstandard constraints in SEM state space modeling of panel data: Modeling stationarity and overlapping cohort commonality.* Manuscript submitted for publication.

Jazwinski, A.H. (1970). *Stochastic processes and filtering theory.* New York: Academic Press.

Jones, R.H. (1985). Time series analysis with unequally spaced data. In E.J. Hannan, P.R. Krishnaiah, & M.M. Rao (Eds.), *Handbook of statistics: Vol. 5. Time series in the time domain* (pp. 157-177). Amsterdam: North-Holland.

Jöreskog, K.G., & Sörbom, D. (1989). *LISREL 7: User's reference guide.* Mooresville IN: Scientific Software.

Jöreskog, K.G. (1973). A general method for estimating a structural equation system. In A.S. Goldberger & O.D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York: Seminar.

Lewis, F.L. (1986). *Optimal estimation with an introduction to stochastic control theory.* New York: Wiley.

Little, D.J.A., & Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Ljung, L. (1985). Estimation of parameters in dynamical systems. In E.J. Hannan, P.R. Krishnaiah, & M.M. Rao (Eds.), *Handbook of statistics: Vol. 5. Time series in the time domain* (pp. 189-211). Amsterdam: North-Holland.

MacCallum, R., & Ashby, F.G. (1986). Relationships between Linear Systems Theory and Covariance Structure Modeling. *Journal of Mathematical Psychology, 30*, 1-27.

Meditch, J.S. (1969). *Stochastic optimal linear estimation and control.* New York: McGraw-Hill.

Mehra, R.K. (1971). Identification of stochastic linear dynamic systems using Kalman filter representation. *AIAA Journal, 23*, 28-31.

Molenaar, P.C.M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychmetrika, 50*, 181-202.

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

Oud, J.H.L., van den Bercken, J.H.L., & Essers, R.J. (1990). Longitudinal factor scores 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 behavioral sciences* (pp. 3-26). Nijmegen: ITS.

Oud, J.H.L., & Jansen, R.A.R.G. (1995). An ARMA extension of the longitudinal LISREL model for LISKAL. In I. Parchev (Ed.), *Multivariate analysis in the behavioral sciences: philosophic to technical*, (pp. 49-69). Sofia: "Prof. Marin Drinov" Academic Publishing House.

Oud, J. H. L., & Jansen, R. A. R. G. (1996). Nonstationary longitudinal LISREL model estimation from incomplete panel data using EM and the Kalman smoother. In U. Engel & J. Reinecke (Eds.), *Analysis of change: advanced techniques in panel data analysis* (pp. 135-159). Berlin: Walter de Gruyter.

Rauch, H.E., Tung, F., & Striebel, C.T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal, 3*, 1445-1450.

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

Schweppe, F. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory, 11*, 61-70.

Shumway, R.H., & Stoffer, D.S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis, 3*, 253-264.

Singer, H. (1990). *Parameterschätzung in zeitkontinuierlichen dynamischen Systemen.* Konstanz: Hartung-Gorre.

Singer, H. (1993). Continuous-time dynamical systems with sampled data, errors of measurement and unobserved components. *Journal of Time Series Analysis, 14*, 527-545.

Willet, J.B. & Sayer, A.G. (1993). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*, 363-381.

ontvangen: 12-10-95
geaccepteerd: 3- 6-96