KM 20(1986)
pag 109-129

LONGITUDINAL FACTOR SCORE ESTIMATION USING THE KALMAN FILTER

Johan H. Oud*, John H. van den Bercken*, Raymond J. Essers*

Abstract

We expound the advantages of the Kalman filter as a factor score estimator in the presence of longitudinal data. Since the Kalman filter presupposes the availability of a dynamic state space model, first the state space model is reviewed. We show it to be translatable into the LISREL model. Several extensions of the LISREL model specification are discussed in order to enhance the applicability of the Kalman filter for behavioral research data. Next, we deal with the Kalman filter and three of its main properties in detail. The relationships are shown between the Kalman filter and two well-known cross-sectional factor score estimators: the regression estimator and the Bartlett estimator. The Bartlett estimator is recommended to be used as initial estimator in the Kalman filtering process. Finally, a worked-out educational research example is presented.

* Instituut voor Orthopedagogiek
K.U. Nijmegen
Erasmusplein 1
6500 HD Nijmegen
tel. 080-512131

The authors are grateful to Cesarius Mommers for his permission to use the Beginning Reading model results in the reseach example and to Giulio Lancioni for his comments on an earlier draft of this article.

1. Introduction

Over the past few decades, research methodology texts in the behavioral sciences have increasingly stressed the need for longitudinal research (Coleman, 1968; Harris, 1963; Nesselroade & Baltes, 1979). Because of cross-sectional biases, however, the available analytic procedures are often found to be unsuitable for analysis of longitudinal data. One example is the simple t-test applied to test the level change in interrupted time series. The procedure has already been indicated as being inadequate by Campbell and Stanley (1963, pp. 212-213). Correct tests of the level change that account for the time structure of the data could be developed afterwards using the Box-Jenkins approach (Box & Tiao, 1965; Glass, Willson & Gottman, 1975) and multivariate analysis of variance (Algina & Swaminathan, 1977, 1979; Oud, 1981).

Factor analysis constitutes another example. Typically designed to handle cross-sectional data (e.g., a set of test scores taken from different persons at a single point in time), conventional factor analytic procedures are unfit to account for and take advantage of the dynamic nature of longitudinal data. In particular, widely used factor score estimators as, for example, the regression estimator and the Bartlett estimator (Lawley & Maxwell, 1971) yield results that are inefficient when applied to longitudinal data.

The aim of the present study is to demonstrate how longitudinal factor score estimation can profit by an important result of modern control theory, the Kalman filter (Gelb, 1982; Jazwinski, 1970; Kalman, 1960; Kalman & Bucy, 1961; Kwakernaak & Sivan, 1972). Assuming an appropriate dynamic <u>state space model</u>, the changing factor scores over time, called states in control theory, are optimally estimated by the Kalman filter. Because of its central role in Kalman filtering, we first shortly review the state space model. We show it to be translatable into the well-known LISREL model as has been explained by Oud (1978). The state space model is thus made estimable for behavioral science data by means of the LISREL program (Jöreskog & Sörbom, 1981). Next, we deal with the Kalman filter and its properties in more detail. Finally, a worked-out educational research example is presented to demonstrate the use of the Kalman filter in practice and to compare the results with traditional factor score estimation. The analysis of the example is carried out by the LISKAL program,[1] computing Kalman filter and Bartlett estimates of the factor scores on the basis of LISREL program output.

[1] The LISKAL program is written in IBM-FORTRAN-77. A free listing and a copy on tape at a cost of $40 can be obtained from the first author.

2. State Space Model and LISREL Methodology

A state space model for a linear stochastic dynamical system consists of two parts: a dynamic state equation

$$x_t = A_{t-1}x_{t-1} + B_{t-1}u_{t-1} + w_{t-1} , \tag{1}$$

and a static read-out or output equation

$$y_t = C_t x_t + v_t . \tag{2}$$

The state equation describes the memory of the system: how much information at each time point is passed to the next state, x_t, by the previous state, x_{t-1}, and how much information is added by the input, u_{t-1}, from outside. The static read-out equation describes the instantaneous connections between the observable output, y_t, and the latent state, x_t. It is equivalent to the model equation of factor analysis: Matrix C_t is the factor pattern matrix with factor loadings as elements. In a more general form of the read-out equation, instantaneous input-output effects $D_t u_t$, not involving the state, are added to the instantaneous state-output effects $C_t x_t$. Presently, however, our treatment will be confined to systems without $D_t u_t$ as well as to deterministic input u_t. In the next section, both restrictions will be relaxed.

The following assumptions are made concerning the successive process error vectors w_t with covariance matrices Q_t and the successive measurement error vectors v_t with covariance matrices R_t: (a) zero expectations, (b) zero covariances between vectors (covariances within vectors are given by Q_t and R_t), and (c) zero covariances with initial state x_{t_0}. From (a), (b), and (c) one derives that $w_{t'}$ and $v_{t'}$ are uncorrelated with states x_t for $t' \geqslant t$. The assumptions are used not only in Kalman filtering but also in maximum likelihood estimation of the parameters of the state space model, to be performed by the LISREL program. The maximum likelihood method proceeds under the additional assumption of joint multinormality of the vectors w_t, v_t, and x_{t_0}. The importance of the normality assumption must not be exaggerated, however. The maximum likelihood fitting function may be used profitably to compute parameter estimates even if the distributions deviate moderately from normality (Boomsma, 1983; Jöreskog & Sörbom, 1981, p. I.29). In Kalman filtering, normality is desirable but the results are optimal in a well-defined sense also without the normality assumption (Kwakernaak & Sivan, 1972, pp. 528-531).

A precise mathematical definition of the concept of state in the stochastic case is given by Kwakernaak (1975, p. 69-70). The definition makes use of probability distributions but again no normality assumption is needed. For the purpose of the present paper, it suffices that the state contains at every time point all information of the past of the system that is relevant for the present and the future. Knowledge of the state allows one to disregard the whole past of the system. This is exemplified by the Kalman filter: Optimal estimation of the latent state x_t requires no information about the system's past prior to t-1, except for an optimal estimation of x_{t-1}.

Postponing a detailed exposition of the way the Kalman filter combines past and present information, we now explain the fitting in of the state space model with LISREL methodology for estimating its parameters. Instead of the general LISREL model, comprising three equations and eight parameter matrices, we will use special case 4 (Jöreskog & Sörbom, 1981, pp. I.11) with two equations:

$$\eta = \mathcal{B}\eta + \zeta, \tag{3}$$

$$y = \Lambda\eta + \varepsilon, \tag{4}$$

in four vectors: η (vector of latent variables), y (vector of observed variables), ζ (vector of structural equation errors), ε (vector of measurement equation errors), and four parameter matrices: $\mathcal{B}$ (structural equation matrix, to be distinguished from B in Equation 1), Λ (measurement equation matrix), Ψ (covariance matrix of ζ), Θ (covariance matrix of ε). Somewhat paradoxically, the special case is more flexible than the general model: Although the state space model presented above could be formulated within the general model, it is only by means of the special case that several of the extensions discussed in the next section become possible.

In its general form, the longitudinal data matrix to be used for model estimation is of order N by pT: N subjects with data on p variables for each of T time points. The pT observed variables are specified in vector y of the LISREL model, first the mT input-variables, followed by rT output-variables: $p = m + r$. In vector η, qT variables are specified, starting again with the mT input-variables and continuing with nT state variables: $q = m + n$. Although, especially in behavioral science, the number of latent state-variables is often smaller than the number of observed output-variables: $n < r$, state space models may have also: $n \geqslant r$. In Table 1 the four vectors and four matrices of the LISREL model are shown for the case $T = 3$, but they are readily extendable for cases $T > 3$. E in Ψ is the covariance matrix of the predetermined variables in the vectors u_{t_0}, u_{t_0+1}, u_{t_0+2}, and x_{t_0}. Since the input-variables are specified as observed and deterministic, they have, in fact, to be considered fixed. Nevertheless, treating them as random in the LISREL program's special or general model yields the same results as in the fixed case (Jöreskog & Sörbom, 1981, p. I.30).

Table 1. LISREL model specification of state space model (T = 3)

$$\underset{\eta}{\begin{bmatrix} u_{t_0} \\ u_{t_0+1} \\ u_{t_0+2} \\ x_{t_0} \\ x_{t_0+1} \\ x_{t_0+2} \end{bmatrix}} \underset{B}{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ B_{t_0} & 0 & 0 & A_{t_0} & 0 & 0 \\ 0 & B_{t_0+1} & 0 & 0 & A_{t_0+1} & 0 \end{bmatrix}} \underset{\zeta}{\begin{bmatrix} u_{t_0} \\ u_{t_0+1} \\ u_{t_0+2} \\ x_{t_0} \\ w_{t_0} \\ w_{t_0+1} \end{bmatrix}} \underset{\Psi}{\begin{bmatrix} \begin{bmatrix} & & & \\ & & E & \\ & & & \end{bmatrix} & & \\ 0 \quad 0 \quad 0 \quad 0 & Q_{t_0} & \\ 0 \quad 0 \quad 0 \quad 0 & 0 & Q_{t_0+1} \end{bmatrix}}$$

$$\underset{y}{\begin{bmatrix} u_{t_0} \\ u_{t_0+1} \\ u_{t_0+2} \\ y_{t_0} \\ y_{t_0+1} \\ y_{t_0+2} \end{bmatrix}} \underset{\Lambda}{\begin{bmatrix} I & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{t_0} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{t_0+1} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{t_0+2} \end{bmatrix}} \underset{\varepsilon}{\begin{bmatrix} 0 \\ 0 \\ 0 \\ v_{t_0} \\ v_{t_0+1} \\ v_{t_0+2} \end{bmatrix}} \underset{\Theta}{\begin{bmatrix} 0 & & & & & \\ 0 & 0 & & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & R_{t_0} & & \\ 0 & 0 & 0 & 0 & R_{t_0+1} & \\ 0 & 0 & 0 & 0 & 0 & R_{t_0+2} \end{bmatrix}}$$

It should be noted that the inclusion of input-variables is not compulsory. In some longitudinal factor analytic studies, one is only interested in the development of the latent variables over time without consideration of possible causal influences exerted on them. The input parts of the vectors and matrices in Table 1 may then be skipped. The implied inputless state space model, having $B_{t-1} = 0$ in Equation 1, is called autonomous. In other cases, constant input-variables are used (e.g., sex and socioeconomic status). It is easily seen that these need only inclusion in u_{t_0}: They may still be taken to influence the states x_{t_0+2}, x_{t_0+3}, etc. after x_{t_0+1} by choosing nonzero parameters on the appropriate places in B of Table 1. This treatment of constant input-variables has been previously suggested by Jöreskog (1978) for background variables.

To obtain an identified model, besides the parameters already fixed at 0 in Table 1, one mostly has to fix additional parameters (at 0 or some different a priori known value) or to specify equality constraints between parameters (e.g. between corresponding parameters at different time points). A necessary condition for identification requires the number of unknown parameters (total number of distinct parameters minus the number of fixed ones and the number of constraints) not to exceed $\frac{1}{2}pT(pT + 1)$. Unfortunately, no general and practically useful necessary and sufficient condition for identification is available. For the particular model, however, the LISREL program performs a nearly fully reliable check on the identification of each unknown parameter (Jöreskog & Sörbom, 1981, p. I.24). Assuming a fully identified model and proper data, the LISREL program gives maximum likelihood estimates of all parameters left free in the model. The resulting estimates of the successive matrices A_t, B_t, C_t, Q_t, and R_t are entered into the Kalman filter.

3. Extensions of the LISREL model specification

The purpose of the extensions in this section is to enhance the accessibility of the state space model and, hence, of the Kalman filter for behavioral research data. Although the first extension is encountered mainly in systems and control theory, it is potentially useful in behavioral science. The other extensions are directly geared to behavioral research practice.

Instantaneous input-output effects

By adding instantaneous input-output effects $D_t u_t$ to Equation 2, leading to the more general read-out equation $y_t = C_t x_t + D_t u_t + v_t$, one passes from a strongly causal system to a weakly causal one. In a strongly causal system, "the output lags, at least infinitesimally, the input" (Willems, 1975, p.26). Even if it is assumed that in reality causal processes always take some time and are, in fact, strongly causal, the inclusion of $D_t u_t$ may lead to a more precise model. One example is the case of temporal measurement inaccuracies. Measurements y_t and u_t refer to specific points in time t, even when they, in fact, are measured over longer periods (e.g., income measured as a sum or average over one year periods). Whenever the measurement period overlaps with the time required for the causal processes between input and output, a term $D_t u_t$ should be included in the read-out equation. Due to the special LISREL model, however, the inclusion and estimation of matrices D_t are easily carried out (see Table 2).

Instantaneous intra-state effects

The specification of instantaneous intra-state effects $K_t x_t$ on x_t simultaneously with effects of x_{t-1} and u_{t-1} on x_t:

$$x_t = K_t x_t + A^*_{t-1} x_{t-1} + B^*_{t-1} u_{t-1} + w^*_{t-1} , \qquad (5)$$

defines a structural equation model. This model is widely used in econometrics and behavioral science (Jöreskog, 1977; Heise, 1975; Theil, 1971). Depending on whether K_t can be chosen as a subdiagonal matrix or not, the model is called recursive or interdependent. Premultiplying both sides by $M_t = (I - K_t)^{-1}$, where matrix $I - K_t$ is assumed nonsingular, Equation 5 reduces to Equation 1: $A_{t-1} = M_t A^*_{t-1}$, $B_{t-1} = M_t B^*_{t-1}$, $w_{t-1} = M_t w^*_{t-1}$, $Q_{t-1} = M_t Q^*_{t-1} M'_t$.

A structural equation model thus defines a state equation indirectly. One could estimate the matrices A_{t-1}, B_{t-1}, and Q_{t-1}, which are required in Kalman filtering, directly from the state equation, thereby skipping the structural equation model. It has been argued, however, that first estimating the matrices of Equation 5 and then deriving estimates of A_{t-1}, B_{t-1}, and Q_{t-1} in the manner shown gives more efficient estimates (Johnston, 1972, pp. 400-404). Maximum likelihood estimates of the matrices K_t, A^*_{t-1}, B^*_{t-1}, and Q^*_{t-1} can be obtained by including them in the LISREL model as indicated in Table 2.

Instantaneous and lagged input-state effects

Instantaneous input-state effects, that is, effects from u_t instead of u_{t-1} on x_t do not need any changes in Table 1. It is possible to cope with them simply by filling in for u_{t_0} , u_{t_0+1}, ... input-variables which are, in fact, one time point ahead (time anticipating input-variables). Not only time-anticipating but also time-lagged input-variables (one or more time points behind the nominal time point) can easily be inserted. When using time-anticipating input-variables, caution is required that only zero coefficients are assigned to them in the matrices D_{t_0} , D_{t_0+1}, ... of a weakly causal system; nonzero coefficients would imply the present to be influenced by the future.

Instantaneous input-state effects are as popular in econometrics and behavioral science as instantaneous intra-state effects. Theil (1971, p. 463), for instance, gives an example with input-variables (called exogenous variables in econometrics) appearing in both anticipating and nonanticipating form in the same model. Models with differently lagged input-variables get considerable attention in econometrics too.

Latent inputs

Up to now, the input has been assumed to be observed and deterministic. In many behavioral science models, however, not only the state but also the input is imperfectly measured and thus latent. The LISREL program permits the inclusion of latent inputs by means of an additional read-out equation, specifying how the observed input u_t is connected to the latent input $\tilde{u}_t$: $u_t = L_t \tilde{u}_t + z_t$. Table 2 adds at the appropriate places the successive latent input-vectors $\tilde{u}_t$, input factor-pattern matrices L_t, input measurement error vectors z_t, and the covariance matrices of the latter F_t.

Table 2. Extended LISREL model specification of state space model (T = 3)

$$
\underset{\eta}{\begin{bmatrix} \tilde{u}_{t_0} \\ \tilde{u}_{t_0+1} \\ \tilde{u}_{t_0+2} \\ x_{t_0} \\ x_{t_0+1} \\ x_{t_0+2} \end{bmatrix}}
\underset{B}{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ B^*_{t_0} & 0 & 0 & A^*_{t_0} & K_{t_0+1} & 0 \\ 0 & B^*_{t_0+1} & 0 & 0 & A^*_{t_0+1} & K_{t_0+2} \end{bmatrix}}
\underset{\zeta}{\begin{bmatrix} \tilde{u}_{t_0} \\ \tilde{u}_{t_0+1} \\ \tilde{u}_{t_0+2} \\ x_{t_0} \\ w_{t_0+1} \\ w_{t_0+2} \end{bmatrix}}
\underset{\Psi}{\begin{bmatrix} \begin{bmatrix} & & & \\ & E & & \\ & & & \end{bmatrix} & & \\ 0 \quad 0 \quad 0 \quad 0 & Q^*_{t_0} & \\ 0 \quad 0 \quad 0 \quad 0 & 0 & Q^*_{t_0+1} \end{bmatrix}}
$$

$$
\underset{y}{\begin{bmatrix} u_{t_0} \\ u_{t_0+1} \\ u_{t_0+2} \\ y_{t_0} \\ y_{t_0+1} \\ y_{t_0+2} \end{bmatrix}}
\underset{\Lambda}{\begin{bmatrix} L_{t_0} & & & & & \\ 0 & L_{t_0+1} & & & & \\ 0 & 0 & L_{t_0+2} & & & \\ D_{t_0} & 0 & 0 & C_{t_0} & & \\ 0 & D_{t_0+1} & 0 & 0 & C_{t_0+1} & \\ 0 & 0 & D_{t_0+2} & 0 & 0 & C_{t_0+2} \end{bmatrix}}
\underset{\varepsilon}{\begin{bmatrix} z_{t_0} \\ z_{t_0+1} \\ z_{t_0+2} \\ v_{t_0} \\ v_{t_0+1} \\ v_{t_0+2} \end{bmatrix}}
\underset{\Theta}{\begin{bmatrix} F_{t_0} & & & & & \\ 0 & F_{t_0+1} & & & & \\ 0 & 0 & F_{t_0+2} & & & \\ 0 & 0 & 0 & R_{t_0} & & \\ 0 & 0 & 0 & 0 & R_{t_0+1} & \\ 0 & 0 & 0 & 0 & 0 & R_{t_0+2} \end{bmatrix}}
$$

In many cases, inputs show some predictability over time. Otter (1985, p. 35) takes advantage thereof for latent inputs by modeling effects between them. This is done most easily by combining the latent input variables with the state variables in a new state vector $\mathring{x}_t$ and specifying the following autonomous state space model for $\mathring{x}_t$:

$$
\begin{bmatrix} \tilde{u}_t \\ x_t \end{bmatrix} = \begin{bmatrix} G_{t-1} & 0 \\ B_{t-1} & A_{t-1} \end{bmatrix} \begin{bmatrix} \tilde{u}_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} g_{t-1} \\ w_{t-1} \end{bmatrix}
$$

$$
\mathring{x}_t = \mathring{A}_{t-1} \quad \mathring{x}_{t-1} + \mathring{w}_{t-1} , \tag{6}
$$

$$\begin{bmatrix} u_t \\ y_t \end{bmatrix} = \begin{bmatrix} L_t & 0 \\ D_t & C_t \end{bmatrix} \begin{bmatrix} \tilde{u}_t \\ x_t \end{bmatrix} + \begin{bmatrix} z_t \\ v_t \end{bmatrix}$$

$$\mathring{y}_t = \mathring{C}_t \quad \mathring{x}_t + \mathring{v}_t \; . \tag{7}$$

As in the case of deterministic input, instead of the state equation (Equation 6), a structural model equation may be chosen:

$$\begin{bmatrix} \tilde{u}_t \\ x_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & K_t \end{bmatrix} \begin{bmatrix} \tilde{u}_t \\ x_t \end{bmatrix} + \begin{bmatrix} G_{t-1} & 0 \\ B^*_{t-1} & A^*_{t-1} \end{bmatrix} \begin{bmatrix} \tilde{u}_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} g_{t-1} \\ w^*_{t-1} \end{bmatrix}$$

$$\mathring{x}_t = \mathring{K}_t \quad \mathring{x}_t + \mathring{A}^*_{t-1} \quad \mathring{x}_{t-1} + \mathring{w}^*_{t-1} \; . \tag{8}$$

Equation 6 is a special case of Equation 8 for $K_t = 0$. Also, for $K_t \neq 0$, Equation 8 may be reduced to Equation 6 as explained previously.

This very general state space model has a major advantage, that is, the Kalman filter becomes suitable to estimate both the latent inputs and the latent states simultaneously. Due to the special LISREL model, the translation in LISREL form proceeds without problems, using either Table 1 (Equation 6 and 7) or Table 2 (Equation 8 and 7). The resulting estimates of the successive matrices $\mathring{A}_t$, $\mathring{C}_t$, $\mathring{Q}_t$, and $\mathring{R}_t$ are entered into the Kalman filter.

Correlated errors over time

Situations may be encountered in longitudinal research where, contrary to the assumptions, the process errors w_t (Jöreskog, 1978) or the measurement errors v_t (Sörbom, 1975) or both are correlated over time. The LISREL program is able to detect such correlations by large modification indices (Jöreskog & Sörbom, 1981, pp. III.18-19) on the places of zero-matrices $Q_{t,t'}$, and $R_{t,t'}$ $(t \neq t')$ in Table 1 or 2. This problem of $Q_{t,t'} \neq 0$ or $R_{t,t'} \neq 0$ can often be solved by adding one or more new state variables to the model and estimating the most effective new parameters in the augmented matrices A_t or A^*_t, and C_t (Kwakernaak & Sivan, 1972, p. 356). In deciding which parameters are most effective in obtaining $Q_{t,t'} = 0$ or $R_{t,t'} = 0$ for the new model, the modification indices of the LISREL program may be helpful again.

4. Kalman Filter

The Kalman filter or optimal estimator $\hat{x}_t$ of the unknown latent state x_t:

$$\hat{x}_t = (I - H_tC_t)\hat{x}_{t-} + H_ty_t$$
$$\text{where } \hat{x}_{t-} = A_{t-1}\hat{x}_{t-1} + B_{t-1}u_{t-1} \quad , \tag{9}$$

combines, in fact, two state estimators: the memory estimator $\hat{x}_{t-}$ and the instantaneous estimator H_ty_t. Its kernel is constituted by the Kalman weighting matrix H_t, to be explained below. The memory estimator processes exclusively past information of the system. Its use would be appropriate in the deterministic case (no process error), provided the state equation and the initial state are perfectly known. Starting with the exact initial state x_{t_0}, the state equation could be applied recursively to find the successive states x_{t_0+1}, x_{t_0+2}, ... The memory approach does not work in the stochastic case, because, in general, neither the initial state nor the process matrices A_t and B_t are known. They also require estimation. Together with model specification errors and the process error, this causes the successive memory estimates or forecasts $\hat{x}_t = \hat{x}_{t-}$ to keep deteriorating, that is, to show increasing estimation error $e_t = x_t - \hat{x}_t$.

The instantaneous estimator H_ty_t uses only the present observed output y_t to estimate the state x_t. This is the approach chosen in cross-sectional factor analytic studies where latent factor scores are estimated by means of observables at the same time point. Writing Equation 9 in the form $\hat{x}_t = \hat{x}_{t-} - H_t\hat{y}_{t-} + H_ty_t$ with $\hat{y}_{t-} = C_t\hat{x}_{t-}$ makes clear that the Kalman filter corrects the memory estimate $\hat{x}_{t-}$ by putting H_ty_t in the place of its memory analogue $H_t\hat{y}_{t-}$ or, equivalently, by adding the linear weighting H_t of the output innovations $y_t - \hat{y}_{t-}$.

How much past information is used and how much is taken from the present output, is defined by

$$H_t = P_tC_t'R_t^{-1} \quad . \tag{10}$$

H_t bilinearly transforms the read-out or factor pattern matrix C_t', postmultiplying it with the inverted measurement error covariance matrix R_t and premultiplying it with the covariance matrix P_t of the Kalman estimation error $e_t = x_t - \hat{x}_t$:

$$P_t = (P_{t-}^{-1} + \Gamma_t)^{-1}$$

$$\text{where } P_{t-} = A_{t-1}P_{t-1}A'_{t-1} + Q_{t-1}$$

$$\text{and} \quad \Gamma_t = C'_t R_t^{-1} C_t . \tag{11}$$

The computation of the Kalman covariance matrix P_t does not need knowledge of the output y_t and so its time path can be evaluated before the filtering process starts. The quality of the filtering results is thus known in advance. Also, it becomes clear from P_t and H_t that Kalman filtering is precluded for state space models yielding singular matrices P_{t-} and R_t.

As can be seen intuitively from H_t, the Kalman filter reduces to the memory estimator, (a) when the model approaches the deterministic, perfect knowledge case ($P_t \rightarrow 0$) and (b) when the measurement errors in the observed output become very large ($R_t^{-1} \rightarrow 0$). Because in the latter case also $\Gamma_t \rightarrow 0$, P_t is seen to reduce to the forecast error covariance matrix P_{t-}. On the other hand, as it can be observed in Equation 9, the Kalman filter becomes equal to the instantaneous estimator when memory effects are absent ($A_{t-1} = B_{t-1} = 0$).

It is interesting to note the close connections that exist between the instantaneous estimator $H_t y_t$ and two popular cross-sectional factor score estimators: the regression estimator and the Bartlett estimator. In fact, these estimators become equal to $H_t y_t$ when their covariance matrices (Lawley & Maxwell, 1971, p. 109-110):

$$\text{Regression } P_t = \Phi_t(I + \Gamma_t\Phi_t)^{-1} = (\Phi_t^{-1} + \Gamma_t)^{-1} ,$$

$$\text{Bartlett} \quad P_t = \Gamma_t^{-1} , \tag{12}$$

are substituted for the Kalman P_t in H_t (Φ_t in the regression P_t is the factor or state covariance matrix). Moreover, by setting $A_{t-1} = B_{t-1} = 0$ in the cross-sectional case, the Kalman P_t reduces to the regression P_t. From $A_{t-1} = B_{t-1} = 0$ one derives:

$$P_{t-} = A_{t-1}P_{t-1}A'_{t-1} + Q_{t-1} = Q_{t-1} ,$$

$$x_t = A_{t-1}x_{t-1} + B_{t-1}u_{t-1} + w_{t-1} = w_{t-1} \quad \text{or} \quad \Phi_t = Q_{t-1}$$

Thus P_{t-}^{-1} in the Kalman P_t (Equation 11) becomes equal to Φ_t^{-1} in the regression P_t (Equation 12). A condition, for which the Kalman P_t reduces to the

Bartlett P_t, is mentioned at the end of the present article in section 6.

Because of the term Φ_t^{-1}, appearing in the regression P_t but not in the Bartlett P_t, the regression estimator has smaller variance than the Bartlett estimator. In fact, the regression estimator has minimum variance among all linear cross-sectional estimators (Lawley & Maxwell, 1971, pp. 107). On the other hand, the Bartlett estimator has the minimum variance property under the additional restriction of being unbiased (Lawley & Maxwell, 1971, pp. 110-111). So, when in the cross-sectional case unbiasedness is a desirable property, the Bartlett estimator is the preferred one.

In the longitudinal case, $A_{t-1} \neq 0$ and/or $B_{t-1} \neq 0$, if the initial estimator $\hat{x}_{t_0}$ is unbiased and the initial covariance matrix P_{t_0} assumed minimum, the Kalman filter can be proven to be minimum variance linear unbiased. In control theory, this important property of the Kalman filter is more often called best linear unbiased (Kwakernaak & Sivan, 1972, pp. 528-530; Otter, 1985, pp. 60-63). "Best" or "optimal" in the sense that P_t of any other linear unbiased estimator exceeds the one of the Kalman filter by a positive semidefinitive matrix. The linearity restriction can be dropped when x_{t_0}, w_t, and v_t are multinormally distributed, the Kalman filter becoming the best of all unbiased estimators, linear and nonlinear (Kwakernaak & Sivan, 1972, pp. 528-531; Otter, 1985, p. 64).

Although strictly the optimality of the Kalman filter holds only, if the initial estimator $\hat{x}_{t_0}$ is minimum variance unbiased, control engineers often are not greatly concerned with the initial value problem. Typically, some more or less realistic guesses are inserted for $\hat{x}_{t_0}$ and P_{t_0}. The reason is a second property which states that - under rather mild conditions - the Kalman filter estimates become after sufficient time points independent of both $\hat{x}_{t_0}$ and P_{t_0} (Jazwinski, 1970, pp. 239-243). As more and more data are processed the Kalman filter forgets, so to speak, the initial values $\hat{x}_{t_0}$ and P_{t_0}. This property makes sure that biases stemming from the chosen initial values become smaller as time proceeds. Despite of how valuable this result may be, for the typically small numbers of time points in behavioral resarch, the initial values do matter and must be chosen carefully. Therefore, we propose to use as the standard initial estimator the Bartlett estimator. The Bartlett estimator is well-known in factor analysis and has, indeed, for multinormality the required property of minimum variance unbiasedness. It is together with the Kalman filter included in the LISKAL program.

The third property of the Kalman filter to be stressed is time-variance, showing up in the subscript t of all matrices involved. Unlike many other results in systems and control theory, the Kalman filter has the advantage that time-invariance is not needed anywhere. In behavioral science, it is often unrealistic to assume that at the end of an extended period of time the same causal mechanisms are still working as in the beginning. Because the Kalman filter allows different matrices to be inserted as time proceeds, it is very suitable for longitudinal behavioral research as will be illustrated by the example in the next section.

Heretofore, the treatment of the Kalman filter has been restricted to strongly causal systems, that is, systems without the instantaneous input-output matrix D_t. For systems with latent inputs $\tilde{u}_t$ that instantaneously influence y_t, no problem arises if Equation 7 is chosen. The D_t-matrix in issue is then handled as part of $\overset{o}{C}_t$ and that way entered into the Kalman filter. For systems with deterministic inputs u_t that instantaneously influence y_t, the only change needed is replacing y_t in Equation 9 by $y_t - D_t u_t$, thus treating this new quantity as the output y_t.

5. Educational Research Example

In the research example to be presented the Kalman filter is used for diagnosing reading disabilities in primary school children, based upon a dynamic LISREL model for Beginning Reading. The model (see Figure 1 and Table 3) is a longitudinally extended version of the Beginning Reading model described by Mommers and Oud (1984) and has been estimated in a group of 225 Dutch primary school children, 1st to 3rd grade. It contains two state-variables: Reading Comprehension and Spelling, and three input-variables: School Readiness, Phonemic Awareness and Decoding Speed. The intervals between the successive states in Figure 1 are six month periods, the first state, x_{t_0}, occuring after 7 months of reading instruction and the fifth state, x_{t_0+4}, after 31 months of reading instruction. School Readiness and Phonemic Awareness have been measured only once: just before reading instruction. They are assumed to be constant input-variables. Originally, Decoding Speed was considered a state variable too and it had been measured as often as Reading Comprehension and Spelling. In preliminary analyses, however, Decoding Speed turned out only to influence but not to be influenced by other variables in the model and, in addition, to correlate almost perfectly with itself over time. For these reasons, in the final version of the LISREL model, it came to

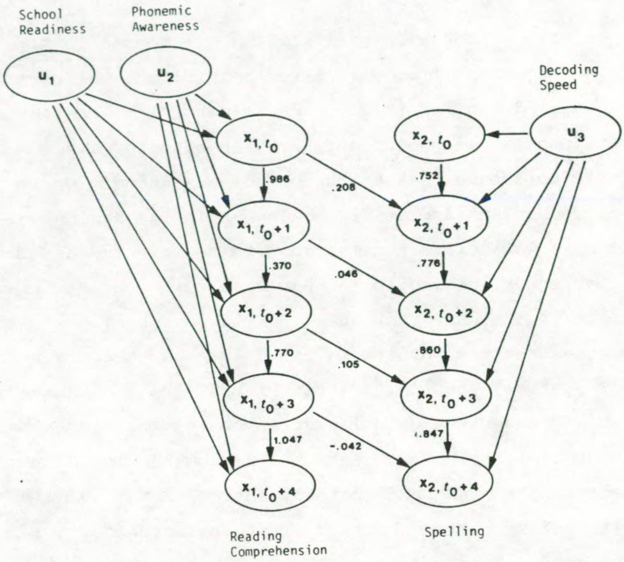

Figure 1. Dynamic LISREL model for Beginning Reading (all variables standardized)

Table 3. Nonzero elements in read-out or factor pattern matrices of the Beginning Reading model (all variables standardized)

$$
\begin{array}{c}
C_{t_0} \\
\begin{array}{l} \\ y_1 \\ \cdot \\ y_9 \\ y_{10} \\ \cdot \end{array}
\begin{array}{cc} x_1 & x_2 \\ \left[\begin{array}{cc} .876 & \\ & \\ & .809 \\ & .774 \\ & \end{array}\right] \end{array}
\end{array}
\quad
\begin{array}{c}
C_{t_0+1} \\
\begin{array}{l} \\ \cdot \\ y_2 \\ \cdot \\ y_{10} \\ y_{11} \\ y_{12} \\ \cdot \end{array}
\begin{array}{cc} x_1 & x_2 \\ \left[\begin{array}{cc} & \\ .999 & \\ & \\ & .797 \\ & .794 \\ & .858 \\ & \end{array}\right] \end{array}
\end{array}
\quad
\begin{array}{c}
C_{t_0+2} \\
\begin{array}{l} \\ \cdot \\ y_3 \\ y_4 \\ \cdot \\ y_{12} \\ y_{13} \\ y_{14} \\ \cdot \end{array}
\begin{array}{cc} x_1 & x_2 \\ \left[\begin{array}{cc} & \\ .818 & \\ .784 & \\ & \\ & .814 \\ & .815 \\ & .766 \\ & \end{array}\right] \end{array}
\end{array}
\quad
\begin{array}{c}
C_{t_0+3} \\
\begin{array}{l} \\ \cdot \\ y_5 \\ \cdot \\ y_{14} \\ y_{15} \\ y_{16} \\ \cdot \end{array}
\begin{array}{cc} x_1 & x_2 \\ \left[\begin{array}{cc} & \\ .865 & \\ & \\ & .815 \\ & .849 \\ & .878 \\ & \end{array}\right] \end{array}
\end{array}
\quad
\begin{array}{c}
C_{t_0+4} \\
\begin{array}{l} \\ \cdot \\ y_6 \\ y_7 \\ y_8 \\ \cdot \\ y_{16} \\ y_{17} \\ y_{18} \end{array}
\begin{array}{cc} x_1 & x_2 \\ \left[\begin{array}{cc} & \\ .876 & \\ .846 & \\ .851 & \\ & \\ & .799 \\ & .853 \\ & .758 \end{array}\right] \end{array}
\end{array}
$$

be considered a constant input-variable like School Readiness and Phonemic Awareness. The measurements of Decoding Speed on time point t_0 have been used for estimation.

Figure 1 contains also the estimated coefficients of the inter-state effects in matrices A_{t_0} through A_{t_0+3}. Very conspicious are the large memory effects of each state variable on itself. In comparison, the influences between Reading Comprehension and Spelling are low and decrease still as time proceeds. After the very first steps in beginning reading, the development of both appears to become practically independent. Because of the large memory effects found, the application of the Kalman filter for estimating the latent states promises to be advantageous. One compelling argument for longitudinal research and the use of the Kalman filter is, in fact, that strong causal effects found cross-sectionally often turn out to decrease or disappear in favor of memory effects when estimated in dynamic models (Oud, 1982).

The read-out or factor pattern matrices in Table 3 demonstrate clearly the usefulness of the time-variance property of the Kalman filter. These matrices show large differences over time, the main reason being the use of different spelling and reading comprehension tests at different didactical ages. For example, test y_1 is only used after 7 months of reading instruction (time point t_0) and test y_{16} only after 25 and 31 months (time points t_0+3 and t_0+4). In behavioral research, it is seldom possible to use the same measurement instruments over the whole age range because of insufficient ceiling for older persons and insufficient bottom for younger persons. Apart from this, time variance also accounts for latent variables manifesting themselves differently, that is, by different coefficients in the observables over time (cf. in Table 3 the estimated coefficients for the same tests on different time points).

The Kalman filter estimates of the latent Reading Comprehension and Spelling scores will now be compared with those of its main cross-sectional competitor, the Bartlett estimator, applied for each time point separately. Because the latter estimator is also taken as initial estimator for the Kalman filter, the Kalman and Bartlett estimates in Figure 2 coincide on the initial time point t_0. As expected in view of its use of past information, the Kalman filter gives estimates that exhibit more memory and change more cautiously over time than those of the Bartlett estimator. Moreover, because the Kalman filter uses increasing information as time proceeds, its estimation error variances in Table 4 (diagonals of P_{t_0}, P_{t_0+1}, ..., P_{t_0+4}) are seen to be

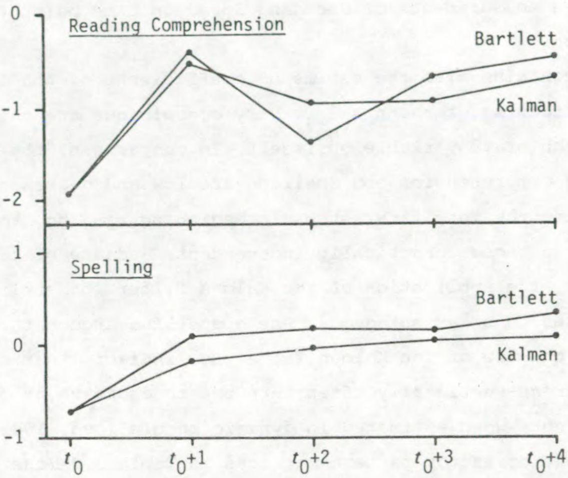

Figure 2. Latent state estimates of Reading Comprehension and Spelling for one of the children by two estimators

Table 4. Estimation error variances

Estimator	Time point t_0	t_0+1	t_0+2	t_0+3	t_0+4
	Reading Comprehension				
Bartlett	.336	.001	.290	.370	.128
Kalman	.336	.001	.159	.094	.079
Kalman from initial value 5	5	.001	.159	.094	.079
	Spelling				
Bartlett	.310	.173	.196	.132	.181
Kalman	.310	.111	.111	.090	.104
Kalman from initial value 5	5	.163	.117	.091	.104

smaller. As explained in the previous section, the Kalman filter gives, in fact, the smallest estimation error among all linear unbiased estimators.

To test the effect of different initial values on the series of Kalman estimation error variances, besides the Bartlett initial values .336 and .310, the extremely deviating initial value 5 was inserted for both state variables. The convergence proved, indeed, to be very fast. Already with the second estimate on time point t_0+2, the difference for Spelling is as small as .006 and for Reading Comprehension, immediately with the first estimate on t_0+1, it is virtually 0. In the latter case, the very small measurement error variance .001 of the single observed Reading Comprehension variable y_2 on that time point must have been responsible. Small measurement errors have, in fact, a dual effect on the Kalman filter: First, as mentioned in the previous section, the instantaneous part $H_t y_t$ becomes more important in comparison to the memory part; second, via Equation 11, small values in R_t cause P_t to decrease more quickly.

Next, we computed for the sample of 225 children the correlations between the Bartlett estimates as well as between the Kalman estimates and compared both of them with the state correlations as given by the LISREL program. Since the Bartlett and Kalman correlations are only indirect estimates based on LISREL solution matrices, while the LISREL solution uses directly all information on the sample level, the LISREL correlations must be considered closer to the true ones. The differences of the Bartlett and the Kalman correlations with the LISREL correlations are given in Table 5. The Kalman correlations turn out to be almost everywhere very close to the LISREL correlations, while the Bartlett correlations show larger differences and err also systematically on the low side. For illustrative reasons, Table 6 gives the computed correlations between the estimates on t_0+1 and the estimates of the same state-variable at later time points. As expected in view of their memory component, the Kalman estimates show higher intercorrelations than the Bartlett estimates.

6. Concluding remarks

In this article, several advantages of the Kalman filter for estimating latent scores have been pointed out. Comparing for an example the Kalman filter results with those of its main cross-sectional competitor, the Bartlett estimator, we showed the Kalman estimates to change more cautiously over time, to have lower estimation error variances, and to reproduce more precisely the

Table 5. Bartlett estimate correlations minus LISREL correlations (above diagonal) and Kalman estimate correlations minus LISREL correlations (below diagonal)

State	x_{1,t_0+1}	x_{2,t_0+1}	x_{1,t_0+2}	x_{2,t_0+2}	x_{1,t_0+3}	x_{2,t_0+3}	x_{1,t_0+4}	x_{2,t_0+4}
x_{1,t_0+1}		-.10	-.15	-.13	-.20	-.13	-.09	-.16
x_{2,t_0+1}	-.01		-.10	-.15	-.08	-.12	-.09	-.10
x_{1,t_0+2}	.11	.10		-.07	-.26	-.08	-.10	-.07
x_{2,t_0+2}	.03	.06	.05		-.10	-.13	-.11	-.09
x_{1,t_0+3}	.04	.08	-.03	.05		-.01	-.12	-.05
x_{2,t_0+3}	.00	.03	.01	-.02	.05		-.08	-.15
x_{1,t_0+4}	.04	.05	-.06	.03	-.03	.04		-.10
x_{2,t_0+4}	-.01	.06	.02	.03	.05	.02	.05	

Table 6. Correlations between estimate at t_0+1 and those at later time points

	Time point		
Estimator	t_0+2	t_0+3	t_0+4
	Reading Comprehension		
Bartlett	.550	.497	.563
Kalman	.805	.739	.699
	Spelling		
Bartlett	.742	.675	.638
Kalman	.948	.824	.799

LISREL program state correlations. We conclude by stressing two points concerning the applicability of the Kalman filter in practice.

First, although Kalman filtering requires the availability of a dynamic model and the better the model the smaller the estimation error, no perfect model is needed. Even when not all relevant variables can be included or some relationships are, in fact, moderately nonlinear, applying the Kalman filter instead of the Bartlett estimator usually does pay off. In contrast to a faulty initial estimator, modeling errors in the state equation (Equation 6) do not jeopardize the unbiasedness of the Kalman filter but, as explained by Poulisse (1980, p. 70), only amplify its estimation error. In addition, modeling errors in the state equation are expected to increase the process error covariance matrix Q_{t-1} as estimated by the LISREL program, thus leading to an increased forecast error covariance matrix P_{t-} (see Equation 11). Only in the end, $Q_{t-1} \rightarrow \infty$ and $P_{t-} \rightarrow \infty$, the Kalman P_t becomes equal to the Bartlett P_t and the Kalman filter becomes equal to the Bartlett estimator. So there is a built-in mechanism, correcting the Kalman filter in the direction of the Bartlett estimator for dynamic modeling errors.

The second point regards the frequent use of standardized variables in behavioral research practice, especially in the field of factor analysis. The model of our research example has been formulated in terms of standardized variables too. The application of the Kalman filter on the basis of a standardized variables model or, equivalently, of a correlational model enables one to evaluate a subject's position in the group for which the model has been estimated. (As with the use of a standardized psychological test, the subject need not be a member of the group but only of the population from which it is drawn.) When, for example, the Kalman filter estimates of a particular subject decrease from 2 to 0, nothing more can be concluded than that this subject's position in the group went down from two unknown standard deviations above the unknown mean to the unknown mean. If, however, one wants to estimate a subject's deviation score developmental curve or absolute developmental curve, the LISREL model can, respectively, be couched in variance-covariance terms or additionally be provided with so-called structured means (Jöreskog & Sörbom, 1981).

References

Algina, J., & Swaminathan, H. (1977). A procedure for the analysis of time-series designs. Journal of Experimental Education, 45, 56-60

Algina, J. & Swaminathan, H. (1979). Alternatives to Simonton's analysis of the interrupted and multiple-group time-series designs. Psychological Bulletin, 86, 919-926.

Boomsma, A. (1983). On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality. Amsterdam: Sociometric Research Foundation.

Box, G.E.P., & Tiao, G.C. (1965). A change in level of a non-stationary time-series. Biometrika, 52, 181-192.

Campbell, D.T., & Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), Handbook of research on teaching (pp. 171-246). Chicago: Rand McNally.

Coleman, J.S. (1968). The mathematical study of change. In H.M. Blalock, Jr., & A.B. Blalock (Eds.), Methodology in social research (pp. 428-478). New York: McGraw-Hill.

Gelb, A. (Ed.). (1974). Applied optimal estimation. Cambridge, MA: M.I.T.

Glass, G.V., Willson, V.L., & Gottman, J.M. (1975). Design and analysis of time-series experiments. Boulder, CO: Colorado Associated University Press.

Harris, C.W. (Ed.). (1963). Problems in measuring change. Madison: University of Wisconsin Press.

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

Johnston, J. (1972). Econometric methods (2nd ed.). New York: McGraw-Hill.

Jöreskog, K. (1977). Structural equation models in the social sciences: specification, estimation and testing. In P.R. Krishnaiah (Ed.), Applications of Statistics (pp. 265-287). Amsterdam: North Holland.

Jöreskog (1978). An econometric model for multivariate panel data. Annales de l'Insée, 30-31, 355-366.

Jöreskog, K.G., & Sörbom, D. (1981). LISREL V: Analysis of linear structural relationships by the method of maximum likelihood: User's guide. Chicago: International Educational Services.

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

Kalman, R.E., & Bucy, R.S. (1961). New results in linear filtering and prediction theory. Journal of Basic Engineering (Trans. ASME, Ser. D.), 83, 95-108.

Kwakernaak, H. (1975). Probabilistische systemen. In H.F.J.M. Buffart & J.H.L. Oud (Eds.), Systeemleer (pp. 62-94). Leiden: Stenfert Kroese.

Kwakernaak, H., & Sivan, R. (1972). Linear optimal control systems. New York: Wiley-Interscience.

Lawley, D.N., & Maxwell, A.E. (1971). Factor analysis as a statistical method. London: Butterworth.

Mommers, M.J.C., & Oud, J.H.L. (1984). Op zoek naar de samenhang tussen technisch lezen, begrijpend lezen en spellingsvaardigheid in de eerste twee leerjaren van de lagere school. In L.F.W. de Klerk & A.M.P. Knoers (Eds.), Onderwijspsychologisch onderzoek (pp. 193-202). Lisse: Swets & Zeitlinger.

Nesselroade, J.R., & Baltes, P.B. (Eds.). (1979). Longitudinal research in the study of behavior and development. New York: Academic Press.

Otter, P.W. (1985). Dynamic feature space modelling, filtering and self-tuning control of stochastic systems. New York: Springer Verlag.

Oud, J.H.L. (1978). Systeem-methodologie in sociaal-wetenschappelijk onderzoek. Nijmegen, The Netherlands: Alfa.

Oud, J.H.L. (1981). Onderzoek van orthopedagogische en onderwijskundige interventies aan de hand van tijdreeksen: een MANOVA-procedure. Tijdschrift voor Onderwijsresearch, 6, 267-291.

Oud, J.H.L. (1982). De stap van input-output naar toestand in sociaal-wetenschappelijk onderzoek. In A.F.G. Hanken & J.H.L. Oud (Eds.), Onderneming en overheid in systeemdynamisch perspectief (pp. 17-33). 's-Gravenhage: Academic Service.

Poulisse, H.N.J. (1983). State- and parameters estimation as chemometrical concepts. Unpublished doctoral dissertation, University of Nijmegen, Nijmegen.

Sörbom, P. (1975). Detection of correlated errors in longitudinal data. British Journal of Mathematical and Statistical Psychology, 28, 138-151.

Willems, J.C. (1975). Minimal realization in state space form from input/output data. In H.F.J.M. Buffert & J.H.L. Oud (Eds.), Systeemleer (pp. 22-61). Leiden: Stenfert Kroese.