Longitudinal Factor Score Estimation Using the Kalman Filter

Johan H. Oud, John H. van den Bercken, and Raymond J. Essers
University of Nijmegen

The advantages of the Kalman filter as a factor score estimator in the presence of longitudinal data are described. Because the Kalman filter presupposes the availability of a dynamic state space model, the state space model is reviewed first, and it is shown 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. The Kalman filter and its main properties are summarized. Relationships are shown between the Kalman filter and two well-known cross-sectional factor score estimators: the regression estimator, and the Bartlett estimator. The indeterminacy problem of factor scores is also discussed in the context of Kalman filtering, and the differences are described between Kalman filtering on the basis of a zero-means and a structured-means LISREL model. By using a structured-means LISREL model, the Kalman filter is capable of estimating absolute latent developmental curves. An educational research example is presented. *Index terms: factor score estimation, indeterminacy of factor scores, Kalman filter, LISREL, longitudinal LISREL modeling, longitudinal factor analysis, state space modeling.*

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). However, because of cross-sectional biases, 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. This procedure was shown to be inadequate by Campbell and Stanley (1963, pp. 212–213). Correct tests of the level change that account for the time structure of the data were later developed 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 to take advantage of the dynamic nature of longitudinal data. Widely-used factor score estimators, such as the regression estimator and the Bartlett estimator (Lawley & Maxwell, 1971), yield results that are less than optimal when applied to longitudinal data.

The aim of the present study was to demonstrate how longitudinal factor score estimation, which is the estimation of the evolution of factor scores for individual examinees over time, can profit from the Kalman filter, an important result of modern control theory (Gelb, 1974; Jazwinski, 1970; Kalman, 1960; Kalman & Bucy, 1961; Kwakernaak & Sivan, 1972). Assuming an appropriate dynamic state space model, 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, the state space model is first briefly reviewed. It is shown to be translatable into the well-known LISREL model, as explained by Oud (1978). The state space model is thus made estimable for behavioral science data by means of the LISREL program

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 14, No. 4, December 1990, pp. 395-418

0146-6216/90/040395-24$2.45

(Jöreskog & Sörbom, 1984). Next the Kalman filter and its properties are described. The performance of the Kalman filter is evaluated on several criteria, in comparison with conventional factor score estimation. Finally an educational research example is presented to demonstrate the use of the Kalman filter in practice.

State Space Model and LISREL Methodology

Contrary to the dynamic state space model used in Kalman filtering, most published explanatory models in behavioral science are cross-sectional. The application of the regression model cross-sectionally is very popular, for instance. Most applications of the well-known LISREL model are cross-sectional, as well. However, the state space model can be viewed as a natural extension of the regression model, as is shown below. The state space model can be formulated fairly easily, too, in terms of the LISREL model, in order to estimate its parameters.

In a cross-sectional regression model, one or more independent variables represented in vector $\mathbf{u}$ are specified to influence one or more dependent variables in vector $\mathbf{x}$ instantaneously, and unexplained errors in $\mathbf{w}$ are also allowed to enter the dependent variables: $\mathbf{x} = \mathbf{B}\mathbf{u} + \mathbf{w}$.

Matrix $\mathbf{B}$ contains coefficients indicating the influence of each of the variables in $\mathbf{u}$ on each of the variables in $\mathbf{x}$. In the longitudinal case it is convenient to label the variables in $\mathbf{u}$ and $\mathbf{x}$, as well as the influences in $\mathbf{B}$ at different time points, with the subscript t: $\mathbf{x}_t = \mathbf{B}_t\mathbf{u}_t + \mathbf{w}_t$. In this way, the cross-sectional regression model is repeated longitudinally. Further, because influences from independent variables generally need some time in order to reach the dependent variables, it is usually considered more realistic in longitudinal research to specify a time lag between dependent and independent variables: $\mathbf{x}_t = \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{w}_{t-1}$. Explaining power is almost always considerably enhanced, too, by the addition of autoregressive effects, implying that $\mathbf{x}_t$ is influenced by its lagged value $\mathbf{x}_{t-1}$:

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

Equation 1 is known as the dynamic part or state equation of the linear stochastic state space model. In Equation 1 $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ are called state, $\mathbf{u}_{t-1}$ is input, and $\mathbf{w}_{t-1}$ is process error. The state transition matrix $\mathbf{A}_{t-1}$ in Equation 1 contains the autoregression and cross-lagged effects between the state variables at $t - 1$ and t. Matrix $\mathbf{B}_{t-1}$ contains the lagged input effects on the state $\mathbf{x}_t$. The input variables in $\mathbf{u}_{t-1}$ are assumed to be observed and deterministic. A procedure for handling latent input variables is discussed below.

The state variables in $\mathbf{x}_t$ are generally assumed to be latent, and to be measurable only by fallible observed variables in output vector $\mathbf{y}_t$, containing sizeable measurement errors in $\mathbf{v}_t$:

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{v}_t \quad . \tag{2}$$

Equation 2 is the static part or output equation of the state space model. This equation is in fact completely equivalent to the model underlying factor analysis: Matrix $\mathbf{C}_t$ is the factor-pattern matrix with factor loadings as elements.

The following assumptions are made in the state space model concerning the successive process error vectors $\mathbf{w}_t$ with covariance matrices $\mathbf{Q}_t$, and the successive measurement error vectors $\mathbf{v}_t$ with covariance matrices $\mathbf{R}_t$:

1. Zero expectations,
2. Zero covariances between vectors (covariances within vectors are in $\mathbf{Q}_t$ and $\mathbf{R}_t$), and
3. Zero covariances with the initial state $\mathbf{x}_{t_0}$.

From these assumptions, a derivation is possible: $\mathbf{w}_t$ and $\mathbf{v}_t$ are uncorrelated with states $\mathbf{x}_t$ for

$t' \geq t$. The assumptions are used not only for Kalman filtering, but also for estimating the parameters of the state space model to be performed by the LISREL program. Among the different estimation methods provided for by the LISREL program, the maximum likelihood method proceeds under the additional assumption of joint multinormality of the vectors $\mathbf{w}_t$, $\mathbf{v}_t$, and $\mathbf{x}_{t_0}$. However, robustness research indicates that the maximum likelihood fitting function may be used profitably to compute parameter estimates, even if the distributions deviate moderately in specific ways from normality (Boomsma, 1983; Jöreskog & Sörbom, 1984, p. I.29). In Kalman filtering normality is desirable, but the results are optimal in a well-defined sense, even 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, pp. 69–70). The definition makes use of probability distributions, but again no normality assumption is needed. At every time point, the state contains all the information on the past of the system that is relevant for the present and the future. Knowledge of the state means that the whole past of the system can be disregarded. This is exemplified by the Kalman filter: Optimal estimation of the latent state $\mathbf{x}_t$ requires no information about the system's past prior to $t - 1$, except for an optimal estimate of $\mathbf{x}_{t-1}$.

Instead of using the general LISREL model, comprised of three equations and eight parameter matrices, Special Case 4 (Jöreskog & Sörbom, 1984, pp. I.11), comprised of only two equations and four parameter matrices, will be used to estimate the parameters of the state space model:

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

and

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

where $\boldsymbol{\eta}$ is the vector of latent variables,
$\mathbf{y}$ is the vector of observed variables,
$\boldsymbol{\zeta}$ is the vector of structural equation errors,
$\boldsymbol{\varepsilon}$ is the vector of measurement equation errors,
$\boldsymbol{B}$ is the structural equation matrix (to be distinguished from $\mathbf{B}$ in Equation 1),
$\boldsymbol{\Lambda}$ is the measurement equation matrix,
$\boldsymbol{\Psi}$ is the covariance matrix of $\boldsymbol{\zeta}$, and
$\boldsymbol{\Theta}$ is the covariance matrix of $\boldsymbol{\varepsilon}$.

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 below become possible.

In its general form, the longitudinal data matrix to be used for model estimation is of order $N \times pT$: N examinees with data on p variables for each of T time points. The pT observed variables are specified in vector $\mathbf{y}$ of the LISREL model, starting with mT input variables, followed by rT output variables: $p = m + r$. In vector $\boldsymbol{\eta}$, qT variables are specified, starting again with the mT input variables and continuing with nT state variables: $q = m + n$. When not all of the p variables are available at each of the T time points, the number of elements in vector $\mathbf{y}$, or in vector $\mathbf{y}$ and in vector $\boldsymbol{\eta}$, is diminished accordingly. For convenience it is assumed that $\mathbf{y}$ and $\boldsymbol{\eta}$ contain pT and qT elements, respectively. In the behavioral sciences, the number of latent state variables is often smaller than the number of observed output variables ($n < r$); however, state space models may also have $n \geq r$.

In Figure 1, the four vectors and four matrices of the LISREL model are shown for the case $T = 3$,

Figure 1
LISREL Model Specification of State Space Model ($T = 3$)

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

$$\underset{y}{\begin{bmatrix} u_{t_0} \\ u_{t_0+1} \\ u_{t_0+2} \\ y_{t_0} \\ y_{t_0+1} \\ y_{t_0+2} \end{bmatrix}} = \underset{\Lambda}{\begin{bmatrix} I & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{t_0} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{t_0+1} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{t_0+2} \end{bmatrix}} \underset{\eta}{\begin{bmatrix} u_{t_0} \\ u_{t_0+1} \\ u_{t_0+2} \\ x_{t_0} \\ x_{t_0+1} \\ x_{t_0+2} \end{bmatrix}} + \underset{\varepsilon}{\begin{bmatrix} 0 \\ 0 \\ 0 \\ v_{t_0} \\ v_{t_0+1} \\ v_{t_0+2} \end{bmatrix}} \qquad \underset{\Theta = E(\varepsilon\varepsilon')}{\begin{bmatrix} 0 & & & & & \\ 0 & 0 & & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & R_{t_0} & & \\ 0 & 0 & 0 & 0 & R_{t_0+1} & \\ 0 & 0 & 0 & 0 & 0 & R_{t_0+2} \end{bmatrix}}$$

but they are readily extendable to cases $T > 3$. The $\mathbf{E}$ submatrix of the $\mathbf{\Psi}$ matrix is the covariance matrix of the predetermined variables in the vectors $\mathbf{u}_{t_0}$, $\mathbf{u}_{t_0+1}$, $\mathbf{u}_{t_0+2}$, and $\mathbf{x}_{t_0}$. Because the input variables are specified as observed and deterministic, they must be considered fixed. Nevertheless, treating them as random in the LISREL program's special or general model yields the same results as the fixed case (Jöreskog & Sörbom, 1984, p. I.30).

The inclusion of input variables is not compulsory. In some longitudinal factor-analytic studies, the development of the latent variables over time is of primary interest, without consideration of the possible causal influences exerted on them from outside. The input parts of the vectors and matrices in Figure 1 may then be skipped. The implied inputless state space model has $\mathbf{B}_{t-1} = \mathbf{0}$ in Equation 1. In other cases, constant input variables are used (e.g., gender and socioeconomic status). It is easily seen that these need only inclusion in $\mathbf{u}_{t_0}$: They may still be taken to influence the states $\mathbf{x}_{t_0+2}$, $\mathbf{x}_{t_0+3}$, ..., after $\mathbf{x}_{t_0+1}$, by selecting nonzero parameters in the appropriate places in $\boldsymbol{B}$ of Figure 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 Figure 1, additional parameters must be fixed (at 0 or some different a priori known value), or equality constraints between parameters must be specified (e.g., between corresponding parameters at different time points). A necessary condition for identification requires that the number of unknown parameters (total number of distinct parameters minus the number of fixed parameters and the number of constraints) does not exceed $pT(pT + 1)/2$. Unfortunately, no general and practically useful necessary and sufficient condition for identification is available. For the particular model estimated, though, the LISREL program performs an almost fully reliable check on the identification of each unknown parameter (Jöreskog & Sörbom, 1984, pp. I.23–24).

A stepwise procedure for building an identified model is as follows. First, select a restricted model according to the model in Figure 1, but with $\boldsymbol{B} = \mathbf{0}$, and choose all factor submodels in accordance with the sufficient identification conditions given by Jöreskog (1979, pp. 40–43). The restricted model is identified. Estimate the chosen restricted model in the second step. Next, free individually all remaining parameters in the factor submodels, and all parameters in $\boldsymbol{B}$ to be included in the final model, and then reestimate for each added parameter. Before freeing, however, check that the parameter involved has a nonzero modification index (Jöreskog & Sörbom, 1984, pp. III.18–19) in the previously estimated model. As explained by Sörbom (1979, pp. 228–229), start with an identified model and free a single fixed parameter with a nonzero modification index, and an identified model is guaranteed. By contrast, freeing a fixed parameter with a zero modification index leads to a nonidentified model.

Assuming a fully identified model and proper data, the LISREL program gives maximum likelihood (ML) estimates of all parameters left free in the model. Alternatively, instead of using the iterative ML estimation procedure, one of the following can be used: the iterative generalized least-squares procedure (GLS); the iterative unweighted least-squares procedure (ULS); or IE, a non-iterative, but consistent, procedure combining instrumental variables with least squares (Jöreskog & Sörbom, 1984, pp. I.27–35, II.25–26). IE derives its name from the fact that it is used by the LISREL program to provide initial estimates for the iterative procedures.

The presence of latent variables in the model causes the relative complexity of all four procedures. This precludes the direct use of the simpler non-iterative procedures for estimating simultaneous equation models, such as the two-stage least-squares (2SLS) or the three-stage least-squares (3SLS) procedures from econometrics (Theil, 1971). ML and GLS have much the same properties in large samples, including the property of minimum variance (Browne, 1977). ULS and, in particular, IE are relatively easy to compute, but they do not attain minimum large-sample variance. An additional reason for using ML or GLS is that both provide standard errors for judging the importance of the estimated parameters. Because an overparameterized model must be avoided in Kalman filtering, the use of standard errors in the model-building process is strongly recommended. The standard errors, as well as the χ^2 goodness-of-fit measure of the model as a whole, are based on the normality assumption in both ML and GLS. Although the standard errors are rather robust, the χ^2 measure is very sensitive to departures from normality. As a final result of the LISREL analysis, the estimates of the successive matrices $\mathbf{A}_{t-1}$, $\mathbf{B}_{t-1}$, $\mathbf{Q}_{t-1}$, $\mathbf{C}_t$, and $\mathbf{R}_t$ are collected from the LISREL output to be entered into the Kalman filter.

Extensions of the LISREL Model Specification

The purpose of the extensions 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 two extensions are appropriate for behavioral research.

Instantaneous Input-Output Effects

By adding instantaneous input-output effects $\mathbf{D}_t\mathbf{u}_t$ to Equation 2, leading to the more general equation $\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{D}_t\mathbf{u}_t + \mathbf{v}_t$, a strongly causal system becomes 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 causal processes always take some time and are strongly causal, the inclusion of $\mathbf{D}_t\mathbf{u}_t$ may lead to a more accurate model. One example is the case of temporal measurement inaccuracies. Measurements $\mathbf{y}_t$ and $\mathbf{u}_t$ refer to specific points in time t, even when they are, in fact, 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 $\mathbf{D}_t\mathbf{u}_t$ should be included in the output equation. Due to the special LISREL model, however, matrices $\mathbf{D}_t$ can be easily included and estimated by replacing the $\mathbf{0}$ submatrices by $\mathbf{D}_{t_0}$, $\mathbf{D}_{t_0+1}$, $\mathbf{D}_{t_0+2}$, ..., at the appropriate places in $\mathbf{\Lambda}$ of Figure 1.

Instantaneous Intra-State Effects

The specification of instantaneous intra-state effects $\mathbf{K}_t\mathbf{x}_t$ on $\mathbf{x}_t$ simultaneously with effects of $\mathbf{x}_{t-1}$ and $\mathbf{u}_{t-1}$ on $\mathbf{x}_t$,

$$\mathbf{x}_t = \mathbf{K}_t\mathbf{x}_t + \mathbf{A}^*_{t-1}\mathbf{x}_{t-1} + \mathbf{B}^*_{t-1}\mathbf{u}_{t-1} + \mathbf{w}^*_{t-1} \quad , \tag{5}$$

defines a structural equations model. This model is widely used in econometrics and the behavioral sciences (Goldberger, 1964; Heise, 1975; Jöreskog, 1977). Depending on whether $\mathbf{K}_t$ can be chosen as a subdiagonal matrix with $\mathbf{Q}^*_{t-1}$ (covariance matrix of $\mathbf{w}^*_{t-1}$) diagonal, the model is called recursive or interdependent. Premultiplying both sides by $\mathbf{M}_t = (\mathbf{I} - \mathbf{K}_t)^{-1}$, where $\mathbf{I} - \mathbf{K}_t$ is assumed to be nonsingular, Equation 5 reduces to Equation 1: $\mathbf{A}_{t-1} = \mathbf{M}_t\mathbf{A}^*_{t-1}$, $\mathbf{B}_{t-1} = \mathbf{M}_t\mathbf{B}^*_{t-1}$, $\mathbf{w}_{t-1} = \mathbf{M}_t\mathbf{w}^*_{t-1}$, $\mathbf{Q}_{t-1} = \mathbf{M}_t\mathbf{Q}^*_{t-1}\mathbf{M}'_t$.

A structural equations model thus defines a state equation indirectly. The matrices $\mathbf{A}_{t-1}$, $\mathbf{B}_{t-1}$, and $\mathbf{Q}_{t-1}$, which are required in Kalman filtering, could be estimated directly from the state equation; consequently, the structural equations model would not be required. However, it has been argued that first estimating the matrices of Equation 5, and then deriving estimates of $\mathbf{A}_{t-1}$, $\mathbf{B}_{t-1}$, and $\mathbf{Q}_{t-1}$ in the manner shown gives more efficient estimates and can also be advantageous for several other reasons (Goldberger, 1964, pp. 364–365, 379–380; Johnston, 1972, pp. 400–404). Estimates of $\mathbf{A}^*_{t-1}$, $\mathbf{B}^*_{t-1}$, $\mathbf{Q}^*_{t-1}$, and $\mathbf{K}_t$ can be obtained by replacing $\mathbf{A}_{t-1}$, $\mathbf{B}_{t-1}$, and $\mathbf{Q}_{t-1}$ in Figure 1 by $\mathbf{A}^*_{t-1}$, $\mathbf{B}^*_{t-1}$, and $\mathbf{Q}^*_{t-1}$, respectively, and by inserting at the appropriate places in $\boldsymbol{B}$ the matrices $\mathbf{K}_{t_0+1}$, $\mathbf{K}_{t_0+2}$, ..., .

Latent Inputs

Up to now, the input has been assumed to be observed and deterministic. In many behavioral science models, however, both the state and the input are imperfectly measured, and are thus latent. The LISREL program permits the inclusion of latent inputs by means of an additional output equation, specifying how the observed input $\mathbf{u}_t$ is connected to the latent input $\tilde{\mathbf{u}}_t$: $\mathbf{u}_t = \mathbf{L}_t\tilde{\mathbf{u}}_t + \mathbf{z}_t$. In many cases, the latent inputs show some predictability over time. Otter (1985, p. 38) took advantage of this 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 $\mathring{\mathbf{x}}_t$, and specifying the following state space model for $\mathring{\mathbf{x}}_t$:

$$\begin{aligned}
\begin{bmatrix} \tilde{\mathbf{u}}_t \\ \mathbf{x}_t \end{bmatrix} &= \begin{bmatrix} \mathbf{G}_{t-1} & \mathbf{0} \\ \mathbf{B}_{t-1} & \mathbf{A}_{t-1} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{u}}_{t-1} \\ \mathbf{x}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{g}_{t-1} \\ \mathbf{w}_{t-1} \end{bmatrix} \\
\mathring{\mathbf{x}}_t &= \qquad \mathring{\mathbf{A}}_{t-1} \qquad\quad \mathring{\mathbf{x}}_{t-1} \quad + \quad \mathring{\mathbf{w}}_{t-1} \quad ,
\end{aligned} \tag{6}$$

and

$$\begin{bmatrix} \mathbf{u}_t \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \mathbf{L}_t & \mathbf{0} \\ \mathbf{D}_t & \mathbf{C}_t \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{u}}_t \\ \mathbf{x}_t \end{bmatrix} + \begin{bmatrix} \mathbf{z}_t \\ \mathbf{v}_t \end{bmatrix}$$
$$\mathring{\mathbf{y}}_t = \mathring{\mathbf{C}}_t \quad \mathring{\mathbf{x}}_t + \mathring{\mathbf{v}}_t \quad . \tag{7}$$

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

$$\begin{bmatrix} \tilde{\mathbf{u}}_t \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_t \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{u}}_t \\ \mathbf{x}_t \end{bmatrix} + \begin{bmatrix} \mathbf{G}_{t-1} & \mathbf{0} \\ \mathbf{B}^*_{t-1} & \mathbf{A}^*_{t-1} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{u}}_{t-1} \\ \mathbf{x}_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{g}_{t-1} \\ \mathbf{w}^*_{t-1} \end{bmatrix}$$
$$\mathring{\mathbf{x}}_t = \mathring{\mathbf{K}}_t \quad \mathring{\mathbf{x}}_t + \mathring{\mathbf{A}}^*_{t-1} \quad \mathring{\mathbf{x}}_{t-1} + \mathring{\mathbf{w}}^*_{t-1} \quad . \tag{8}$$

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

This very general state space model has a major advantage, namely that the Kalman filter becomes capable of estimating both the latent inputs and the latent states simultaneously. A still more general model results when the latent inputs $\tilde{\mathbf{u}}_t$, measured by observed inputs $\mathbf{u}_t$, are combined with observed deterministic inputs $\mathring{\mathbf{u}}_t$ in the same model. This can be done by adding deterministic input effects terms $\mathring{\mathbf{B}}_{t-1}\mathring{\mathbf{u}}_{t-1}$ and $\mathring{\mathbf{B}}^*_{t-1}\mathring{\mathbf{u}}_{t-1}$ to Equations 6 and 8, respectively. As a result, Equations 6, 7, and 8 become equal in form to Equations 1, 2, and 5, respectively, and can be handled in exactly the same way using Figure 1. The resulting estimates of the successive matrices $\mathring{\mathbf{A}}_{t-1}$, $\mathring{\mathbf{B}}_{t-1}$, $\mathring{\mathbf{Q}}_{t-1}$, $\mathring{\mathbf{C}}_t$, and $\mathring{\mathbf{R}}_t$ are entered into the Kalman filter.

Kalman Filtering

The Kalman filter originates from control theory. It is used to extract an optimal estimate of the present state of the system from knowledge about the past of a system and from present observations. For instance, ship steering requires precise knowledge of the position of the ship at successive points in time. The position is the unknown multidimensional state that has to be estimated at each point in time. In principle, estimation can proceed in two different ways that the Kalman filter combines optimally.

The first approach to estimation is prediction from the previous state estimate on the basis of a dynamic causal model. If an estimate of the previous position is known and the influences exerted on the ship in the intervening time are also known, an attempt can be made to predict the present position on the basis of the model. In behavioral science terminology the unknown state variables are called factors, and the prediction can be called a predictive factor score estimate. Because the prediction makes use of past information of the system contained in the previous state estimate, it is also called a memory estimate.

The second approach to estimating the state is an instantaneous factor score estimate, based on present observations. Of course, when no past information is available, this is the only possible way to proceed.

The Kalman filter or optimal estimator $\hat{\mathbf{x}}_t$ of the unknown latent state $\mathbf{x}_t$ combines the two estimators: memory estimator $\hat{\mathbf{x}}_{t-}$, using the previous state estimate $\hat{\mathbf{x}}_{t-1}$, and instantaneous estimator $\mathbf{H}_t\mathbf{y}_t$, using the present observed information in $\mathbf{y}_t$:

$$\hat{\mathbf{x}}_t = (\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)\hat{\mathbf{x}}_{t-} + \mathbf{H}_t\mathbf{y}_t \quad , \tag{9}$$

where $\hat{\mathbf{x}}_{t-} = \mathbf{A}_{t-1}\hat{\mathbf{x}}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1}$.

The kernel of the Kalman filter is the Kalman weighting matrix $\mathbf{H}_t$, to be explained below. The memory estimator $\hat{\mathbf{x}}_{t-}$ processes only past information of the system. Its use as the state estimator $\hat{\mathbf{x}}_t$ 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 $\mathbf{x}_{t_0}$, the state equation could be applied recursively to find the successive states $\mathbf{x}_{t_0+1}, \mathbf{x}_{t_0+2}, \ldots,$.

The memory approach does not work in the stochastic case; in general, the initial state and the process matrices $\mathbf{A}_t$ and $\mathbf{B}_t$ are not known in this case, and they also require estimation. Together with model specification errors and the process error, this case causes the successive memory estimates or forecasts to keep deteriorating, or to show increasing estimation error.

The instantaneous estimator $\mathbf{H}_t\mathbf{y}_t$ uses only the present observed output $\mathbf{y}_t$ to estimate the state $\mathbf{x}_t$. 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

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t-} - \mathbf{H}_t\hat{\mathbf{y}}_{t-} + \mathbf{H}_t\mathbf{y}_t \quad , \tag{10}$$
$$\text{with } \hat{\mathbf{y}}_{t-} = \mathbf{C}_t\hat{\mathbf{x}}_{t-}$$

makes clear that the Kalman filter corrects the memory estimate $\hat{\mathbf{x}}_{t-}$ by putting $\mathbf{H}_t\mathbf{y}_t$ in the place of its memory analogue $\mathbf{H}_t\hat{\mathbf{y}}_{t-}$ or, equivalently, by adding the linear weighting $\mathbf{H}_t$ of the output innovations $\mathbf{y}_t - \hat{\mathbf{y}}_{t-}$. No correction takes place when $\mathbf{y}_t - \hat{\mathbf{y}}_{t-} = \mathbf{0}$—that is, when there is no innovation or "surprise" in the observed $\mathbf{y}_t$.

The amount of past information used and the amount taken from the present output are defined by

$$\mathbf{H}_t = \mathbf{P}_t\mathbf{C}_t'\mathbf{R}_t^{-1} \quad . \tag{11}$$

$\mathbf{H}_t$ bilinearly transforms the factor pattern matrix $\mathbf{C}_t'$, postmultiplying it by the inverted measurement error covariance matrix $\mathbf{R}_t$, and premultiplying it by the covariance matrix $\mathbf{P}_t = \mathrm{E}(\mathbf{e}_t\mathbf{e}_t')$ of the Kalman estimation error $\mathbf{e}_t = \mathbf{x}_t - \hat{\mathbf{x}}_t$:

$$\mathbf{P}_t = (\mathbf{P}_{t-}^{-1} + \mathbf{\Gamma}_t)^{-1} \quad , \tag{12}$$
$$\text{where } \mathbf{P}_{t-} = \mathbf{A}_{t-1}\mathbf{P}_{t-1}\mathbf{A}_{t-1}' + \mathbf{Q}_{t-1}$$
$$\text{and } \quad \mathbf{\Gamma}_t = \mathbf{C}_t'\mathbf{R}_t^{-1}\mathbf{C}_t \quad .$$

The computation of the Kalman covariance matrix $\mathbf{P}_t$ does not need knowledge of the output $\mathbf{y}_t$, so its time path can be evaluated before the filtering process starts. The quality of the filtering results is thus known in advance. There are many alternative formulations for $\mathbf{H}_t$ and $\mathbf{P}_t$, for example:

$$\mathbf{H}_t = \mathbf{P}_{t-}\mathbf{C}_t'(\mathbf{C}_t\mathbf{P}_{t-}\mathbf{C}_t' + \mathbf{R}_t)^{-1} \tag{13}$$

and

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)\mathbf{P}_{t-}(\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)' + \mathbf{H}_t\mathbf{R}_t\mathbf{H}_t' \quad . \tag{14}$$

These formulations have the advantage that only one matrix, the innovations covariance matrix $\mathbf{C}_t\mathbf{P}_{t-}\mathbf{C}_t' + \mathbf{R}_t$, must be inverted. This matrix may be invertible, even when $\mathbf{P}_{t-}$ or $\mathbf{R}_t$ is singular.

Thus far, the treatment of the Kalman filter has been restricted to strongly causal systems, or systems without the instantaneous input-output matrix $\mathbf{D}_t$. For systems with latent inputs $\tilde{\mathbf{u}}_t$ that instantaneously influence $\mathbf{y}_t$, no problem arises if Equation 7 is chosen. The $\mathbf{D}_t$ matrix is then handled as part of $\mathring{\mathbf{C}}_t$ and is entered that way into the Kalman filter. For systems with deterministic inputs $\mathbf{u}_t$ that instantaneously influence $\mathbf{y}_t$, the only change required is to replace $\mathbf{y}_t$ in Equation 9 with $\mathbf{y}_t - \mathbf{D}_t\mathbf{u}_t$, thus treating this new quantity as the output $\mathbf{y}_t$.

As can be seen intuitively from $\mathbf{H}_t$ in Equation 11, the Kalman filter reduces to the memory estimator (1) when the model approaches the deterministic, perfect knowledge case ($\mathbf{P}_t \rightarrow \mathbf{0}$), and (2) when the measurement errors in the observed output become very large ($\mathbf{R}_t^{-1} \rightarrow \mathbf{0}$). Because $\mathbf{\Gamma}_t \rightarrow \mathbf{0}$ in the latter case, $\mathbf{P}_t$ will reduce to the forecast error covariance matrix $\mathbf{P}_{t-}$. On the other hand, Equation 9 shows that the Kalman filter becomes equal to the instantaneous estimator when memory effects are absent ($\mathbf{A}_{t-1} = \mathbf{B}_{t-1} = \mathbf{0}$) or when $\mathbf{H}_t\mathbf{C}_t = \mathbf{I}$.

Connections Between the Kalman Filter and Two Cross-Sectional Factor Score Estimators

There are close connections between the instantaneous estimator $\mathbf{H}_t\mathbf{y}_t$ and two popular cross-sectional factor score estimators: the regression estimator and the Bartlett estimator. In fact, these estimators are equal to $\mathbf{H}_t\mathbf{y}_t$ when the regression covariance matrix (Lawley & Maxwell, 1971, pp. 109–110),

$$\mathbf{P}_t = \mathbf{\Phi}_t(\mathbf{I} + \mathbf{\Gamma}_t\mathbf{\Phi}_t)^{-1} = (\mathbf{\Phi}_t^{-1} + \mathbf{\Gamma}_t)^{-1} \quad , \tag{15}$$

and the Bartlett covariance matrix

$$\mathbf{P}_t = \mathbf{\Gamma}_t^{-1} \quad , \tag{16}$$

are substituted for the Kalman $\mathbf{P}_t$ in $\mathbf{H}_t$; $\mathbf{\Phi}_t$ in the regression $\mathbf{P}_t$ is the factor or state covariance matrix $\mathrm{E}(\mathbf{x}_t\mathbf{x}_t')$. Moreover, by setting $\mathbf{A}_{t-1} = \mathbf{B}_{t-1} = \mathbf{0}$ in the cross-sectional case, the Kalman $\mathbf{P}_t$ reduces to the regression $\mathbf{P}_t$. From $\mathbf{A}_{t-1} = \mathbf{B}_{t-1} = \mathbf{0}$,

$$\mathbf{P}_{t-} = \mathbf{A}_{t-1}\mathbf{P}_{t-1}\mathbf{A}_{t-1}' + \mathbf{Q}_{t-1} = \mathbf{Q}_{t-1} \tag{17}$$

and

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{w}_{t-1} = \mathbf{w}_{t-1} \text{ or } \mathbf{\Phi}_t = \mathbf{Q}_{t-1} \tag{18}$$

can be derived. Thus $\mathbf{P}_{t-}^{-1}$ in the Kalman $\mathbf{P}_t$ (Equation 12) becomes equal to $\mathbf{\Phi}_t^{-1}$ in the regression $\mathbf{P}_t$ (Equation 15). A condition under which the Kalman $\mathbf{P}_t$ reduces to the Bartlett $\mathbf{P}_t$ is $\mathbf{Q}_{t-1} \rightarrow \infty$, and thus $\mathbf{P}_{t-}^{-1} \rightarrow \mathbf{0}$, implying also $\mathbf{H}_t\mathbf{C}_t \rightarrow \mathbf{I}$. This condition is discussed more extensively below.

Because of the term $\mathbf{\Phi}_t^{-1}$, which appears in the regression $\mathbf{P}_t$ but not in the Bartlett $\mathbf{P}_t$, 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, p. 107). On the other hand, the Bartlett estimator has the minimum variance property within the class of unbiased linear estimators (Lawley & Maxwell, 1971, pp. 110–111). Therefore, when unbiasedness is a desirable property in the cross-sectional case, the Bartlett estimator is to be preferred.

Four Properties of the Kalman Filter

In the longitudinal case, if the initial estimator $\hat{\mathbf{x}}_{t_0}$ is unbiased and its covariance matrix $\mathbf{P}_{t_0}$ is assumed minimum, the Kalman filter is unbiased and has minimum variance in the class of all linear estimators. These properties result in the Kalman filter being referred to as "best linear" in control theory (Kwakernaak & Sivan, 1972, pp. 528–530; Otter, 1985, pp. 60–63). "Best" or "optimal" is used in the sense that $\mathbf{P}_t$ of any other linear estimator exceeds that of the Kalman filter by a non-negative definite matrix; thus $\mathbf{P}_t$ of any other linear estimator minus $\mathbf{P}_t$ of the Kalman filter is always non-negative definite. Because $\mathbf{P}_{t-}^{-1}$ exceeds $\mathbf{\Phi}_t^{-1}$ and $\mathbf{\Phi}_t^{-1}$ exceeds $\mathbf{0}$ by non-negative definite matrices, it is immediately clear from Equations 12, 15, and 16 that the Kalman filter is better than the regression estimator and the Bartlett estimator. The linearity restriction can be dropped when $\mathbf{x}_{t_0}$, successive $\mathbf{w}_t$, and successive $\mathbf{v}_t$ are jointly multinormally distributed, and the Kalman filter becomes the best of all estimators, both linear and nonlinear (Kwakernaak & Sivan, 1972, pp. 528–531; Otter, 1985, p. 64).

Although the optimality of the Kalman filter holds only if the initial estimator $\hat{\mathbf{x}}_{t_0}$ is optimal (i.e., unbiased with $\mathbf{P}_{t_0}$ minimum), control engineers are often not greatly concerned with the initial value problem. They typically insert some plausible estimates for $\hat{\mathbf{x}}_{t_0}$ and $\mathbf{P}_{t_0}$. This is because of a third property of the Kalman filter: Under rather mild conditions after sufficient time points, the Kalman filter estimates become independent of both $\hat{\mathbf{x}}_{t_0}$ and $\mathbf{P}_{t_0}$ (Jazwinski, 1970, pp. 239–243). As more and more data are processed, the Kalman filter "forgets" the initial values $\hat{\mathbf{x}}_{t_0}$ and $\mathbf{P}_{t_0}$. This property insures that biases stemming from the chosen initial values become smaller and smaller as time proceeds. Regardless of how valuable this result may be, because of the small numbers of time points typically used in behavioral research the initial values do matter and must be chosen carefully. Therefore, the Bartlett estimator is proposed as the standard initial estimator. It is well known in cross-sectional factor analysis, and has the property of minimum variance unbiasedness when multinormality is assumed.

The fourth property of the Kalman filter is time-variance, which appears in the subscript t added to 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 the same causal mechanisms are still working at the end of an extended period of time that were working at the beginning. The Kalman filter allows different matrices to be inserted as time proceeds, and thus seems especially suitable for longitudinal behavioral research studies. (See, however, the typical advantages of time-invariant $\mathbf{C}_t$ matrices for interpretation of the latent factors discussed below.)

Time-invariance can be specified and tested with the LISREL program by the introduction of equality constraints between corresponding parameters in successive matrices. For example, when the influences of each first state variable on the second one are hypothesized to be equal over time, a_{21} in $\mathbf{A}_{t_0}$ should be specified to be equal to a_{21} in $\mathbf{A}_{t_0+1}$, a_{21} in $\mathbf{A}_{t_0+1}$ should be specified to be equal to a_{21} in $\mathbf{A}_{t_0+2}$, and so on. In order to avoid identification problems, it is probably a good strategy to start with time-invariances wherever possible, and to change to time-varying parameters only when the relaxing of equality constraints leads to a significant improvement of model fit.

Special attention should be drawn to longitudinal models in which the factor pattern matrices are different over time—that is, with time-varying $\mathbf{C}_t$. Examples have been given and analyzed by Sörbom (1975, p. 148; 1979, pp. 228–230), and by Jöreskog and Sörbom (1977, pp. 300–301). The import of time-varying $\mathbf{C}_t$ for Kalman filtering differs, however, according to whether the LISREL model analyzed has zero means or structured means.

Kalman Filtering on the Basis of a Zero-Means LISREL Model

A zero-means LISREL model, $E(\mathbf{y}) = E(\boldsymbol{\eta}) = \mathbf{0}$ in Equations 3 and 4, is appropriate in a considerable number of factor-analytic and related studies in behavioral science. This is because the origin, as well as the measurement unit of the observed variables, are often considered arbitrary. As a consequence, differences in means and standard deviations between variables are not informative, and the interpretation of analysis results can be made more convenient by transforming to zero means and standardizing all variables (setting additionally all standard deviations at 1). Analysis of the data covariance matrix by the LISREL program implies that all observed and latent means are set automatically at 0. Furthermore, the so-called standardized solution of the LISREL program provides the analysis results with all latent variables standardized. Finally, the LISREL solution can be transformed to standardized observed variables after it has been obtained (analyzing the data correlation matrix directly instead of the data covariance matrix is generally not permitted; Cudeck, 1989; Jöreskog & Sörbom, 1984, p. I.39).

The application of the Kalman filter on the basis of a LISREL model with standardized latent variables (and thus zero latent means) enables the evaluation of an examinee's position in the group for which the model has been estimated. In a similar way, the use of a standardized psychological test with a normative group does not require the examinee to 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 examinee decrease over time from 2 to 0, it can be concluded that this examinee's position changed from 2 (unknown) standard deviations above the (unknown) mean in the first latent distribution to the (unknown) mean in the second latent distribution.

This is a legitimate use of the Kalman filter, and it gives the appropriate answer for a considerable number of practical problems in behavioral science, particularly those concerned with the normative assessment of people at consecutive time points. Moreover, this use does not presuppose time-invariant $\mathbf{C}_t$. Time-invariant $\mathbf{C}_t$ are attractive because they result in latent variables that can firmly be assumed to keep the same meaning over time, but time-varying $\mathbf{C}_t$ are allowed in Kalman filtering and cannot be avoided in practice. It is often not possible in behavioral research to use the same measurement instruments over the whole age range, due to an insufficient ceiling for some persons and an insufficient bottom for others. In addition, many important psychological and educational constructs, such as intelligence and reading comprehension, are known to change content across developmental stages, and thereby to manifest themselves differently in the observables over time. Whether the latent variables keep exactly the same meaning over time, or keep their $\mathbf{C}_t$ time-invariant, it is always meaningful to consider the development of an individual in terms of his or her changing position within successive latent distributions of a well-defined group of individuals. As already explained, this is done optimally with the Kalman filter, based on a LISREL model with zero means.

Kalman Filtering on the Basis of a Structured-Means LISREL Model

In many situations, however, interest is focused not only on an examinee's relative position in a group, but also on the estimation of an examinee's or a group's absolute developmental curve. The zero-means LISREL model is not able to solve this problem, but the LISREL model with structured means is (Jöreskog & Sörbom, 1984, pp. V.14–16). It adds intercept-coefficient vectors $\boldsymbol{\alpha}$ and $\mathbf{v}$ to Equations 3 and 4:

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \quad , \tag{19}$$

and

$$\mathbf{y} = \mathbf{v} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad . \tag{20}$$

The coefficients in $\boldsymbol{\alpha}$ and in $\mathbf{v}$ can easily be made part of matrices $\boldsymbol{B}$ and $\boldsymbol{\Lambda}$ and estimated in the form of β and λ coefficients, if a constant unit input variable is added to the vectors $\boldsymbol{\eta}$ and $\mathbf{y}$ in Figure 1. The structured-means model yields latent structured means $\mathrm{E}(\boldsymbol{\eta}) = (\mathbf{I} - \boldsymbol{B})^{-1}\boldsymbol{\alpha}$ and observed structured means $\mathrm{E}(\mathbf{y}) = \mathbf{v} + \boldsymbol{\Lambda}(\mathbf{I} - \boldsymbol{B})^{-1}\boldsymbol{\alpha}$, which are both possibly different from 0, and also yields standard deviations, possibly different from 1.

A strategy for identifying, as well as interpreting, the $\boldsymbol{\alpha}$ and $\mathbf{v}$ coefficients is (1) to have equality constraints between the $\mathbf{v}$ coefficients of observed variables at different time points that come from the same measuring instrument (setting origins of observed variables), and (2) to fix at 0 $\boldsymbol{\alpha}$ coefficients for the latent variables at initial time point t_0 (setting origins of latent variables). The $\boldsymbol{\alpha}$ coefficients at later time points can then be estimated, and they result in a mean latent developmental curve, possibly being nonzero. Estimation of the structured-means model through the LISREL program proceeds by analyzing the data moment matrix about 0, instead of the data covariance matrix. The moment matrix must include the unit input variable, with 1 in the diagonal of the moment matrix and the means of all the other variables as off-diagonal elements. For the application of the Kalman filter to estimating the latent absolute developmental curve of an individual examinee, the $\boldsymbol{\alpha}$ and $\mathbf{v}$ coefficients are made part of the $\mathbf{B}_t$ and $\mathbf{D}_t$ matrices, respectively, and are multiplied by the constant unit input added to each input vector $\mathbf{u}_t$.

When estimating an absolute developmental curve on the basis of a structured-means model, it is crucial—in sharp contrast to the discussed use of the zero-means model—that the latent variables keep exactly the same content over time, or equivalently that the observed variables at different time points measure exactly the same latent traits. If not, speaking about growth or decay in the latent variables and trying to estimate these changes would be meaningless. Time-invariance of corresponding λ coefficients in successive $\mathbf{C}_t$ matrices (and of corresponding $\mathbf{v}$ coefficients in successive $\mathbf{D}_t$ matrices) would provide very convincing evidence for this sameness in content, and hence for the appropriateness of the use of the structured-means model. Time-invariance is testable by the LISREL program through the introduction of equality constraints between the corresponding parameters.

However, when the $\mathbf{C}_t$ matrices are time-varying (as may happen, for example, when different instruments are used at different time points), the latent variables could measure different things and thus the meaningfulness of estimating an absolute developmental curve becomes questionable. Fortunately, this situation need not be hopeless, because different observed variables may measure the same latent trait or be "congeneric" (Jöreskog, 1974, pp. 5–6), even though their factor loadings are not equal. For congeneric observed variables that originate from the same time point, it is essential that the correlations between their underlying latent variables equal 1. This, too, can be tested by the LISREL program, as explained by Jöreskog (1974, pp. 9–12). Longitudinally, however, this test for congenericity does not help, because the latent values underlying congeneric observed variables from different time points may have changed over time, and as a result, correlations may be less than 1.

An indirect test of longitudinal congenericity is possible when measuring instruments used at different time points overlap in time—for example, when instrument A is used at time points 1 and 2, B (being a little more difficult and thus appropriate at somewhat higher age levels) at time points 2 and 3, C at time points 3 and 4, and so forth. Overlap of measuring instruments, with equality constraints specified between λ and $\mathbf{v}$ coefficients of the same instrument at different time points, is also essential for obtaining identification of the $\boldsymbol{\alpha}$ coefficients. The procedure proposed aims at keeping constant the latent origin and measurement unit: Between pairs of consecutive time points, at least one instrument should remain the same, with its λ and $\mathbf{v}$ coefficients constrained to be equal over time. At each time point, different instruments are tested for congenericity in order for their λ and $\mathbf{v}$ coefficients to be linkable by a common underlying latent variable. Suppose that congenericity could be demonstrated between instruments A and B at time point 2, between B and C at time point 3, and so forth. It is then indirectly implied that congenericity exists between A and C at different time points, and that they, too, may be considered linear transformations of the same underlying variable. Although perhaps less convincing than testing for time-invariance of the $\mathbf{C}_t$ matrices, this indirect congenericity test gives some evidence that the latent variables keep the same content over time, and therefore that the structured-means model would be applicable as a basis for estimating an individual's absolute developmental curve by means of Kalman filtering.

Two Additional Properties of the Kalman Filter, and the Indeterminacy Problem

In the literature of traditional orthogonal factor analysis, factor score estimators are evaluated in terms of (1) unbiasedness, (2) size of estimation error variance, (3) univocality, and (4) orthogonality. Orthogonal factor analysis assumes that the factors or true factor scores are uncorrelated—that is, $E(\mathbf{x}_t\mathbf{x}_t') = \mathbf{I}$ for standardized $\mathbf{x}_t$. Univocality, $E(\hat{\mathbf{x}}_t\mathbf{x}_t') = \mathbf{I}$, means that the covariances between estimator and true factor scores are equal to the true factor scores' variances and covariances. Orthogonality, $E(\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t') = \mathbf{I}$, means that the variances and covariances of the estimator are equal to those of the true factor scores.

These two additional criteria can easily be generalized to the oblique case and to factors not being standardized. Ideally, it should be the case that

$$E(\hat{\mathbf{x}}_t\mathbf{x}_t') = E(\mathbf{x}_t\mathbf{x}_t') = \mathbf{\Phi}_t \tag{21}$$

and

$$E(\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t') = E(\mathbf{x}_t\mathbf{x}_t') = \mathbf{\Phi}_t \quad . \tag{22}$$

In practice, though, most estimators deviate from the ideal. Lawley and Maxwell (1971, pp. 109–110) show for the regression estimator that

$$E(\hat{\mathbf{x}}_t\mathbf{x}_t') = E(\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t') = \mathbf{\Phi}_t - (\mathbf{\Phi}_t^{-1} + \mathbf{\Gamma}_t)^{-1} \quad , \tag{23}$$

and for the Bartlett estimator that,

$$E(\hat{\mathbf{x}}_t\mathbf{x}_t') = \mathbf{\Phi}_t,\ E(\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t') = \mathbf{\Phi}_t + \mathbf{\Gamma}_t^{-1} \quad . \tag{24}$$

Hence only $E(\hat{\mathbf{x}}_t\mathbf{x}_t')$ of the Bartlett estimator meets the criterion entirely, but at the cost of having its $E(\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t')$ deviating more than that of the regression estimator: $\mathbf{\Gamma}_t^{-1}$ exceeds $(\mathbf{\Phi}_t^{-1} + \mathbf{\Gamma}_t)^{-1}$ by a nonnegative definite matrix.

How does the Kalman filter behave on the two additional criteria? The covariance matrices involved can be derived as follows:

$$\begin{aligned} E(\hat{\mathbf{x}}_t\mathbf{x}_t') = E[(\mathbf{x}_t - \mathbf{e}_t)\mathbf{x}_t'] &= \mathbf{\Phi}_t - E(\mathbf{e}_t\mathbf{x}_t') \\ &= \mathbf{\Phi}_t - E[\mathbf{e}_t(\mathbf{e}_t' + \hat{\mathbf{x}}_t')] \\ &= \mathbf{\Phi}_t - \mathbf{P}_t - E(\mathbf{e}_t\hat{\mathbf{x}}_t') \end{aligned} \tag{25}$$

and

$$\begin{aligned} E(\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t') = E[(\mathbf{x}_t - \mathbf{e}_t)\ (\mathbf{x}_t' - \mathbf{e}_t')] &= \mathbf{\Phi}_t - E(\mathbf{e}_t\mathbf{x}_t') - E(\mathbf{x}_t\mathbf{e}_t') + \mathbf{P}_t \\ &= \mathbf{\Phi}_t - \mathbf{P}_t - E(\mathbf{e}_t\hat{\mathbf{x}}_t') - E(\hat{\mathbf{x}}_t\mathbf{e}_t') \quad . \end{aligned} \tag{26}$$

It is known for the regression estimator that $E(\mathbf{e}_t\hat{\mathbf{x}}_t') = \mathbf{0}$ (estimation error uncorrelated with estimator—see McDonald, 1985, p. 163) and $\mathbf{P}_t = (\mathbf{\Phi}_t^{-1} + \mathbf{\Gamma}_t)^{-1}$ (see Equation 15). Therefore, the covariance matrices in Equation 23 are seen to be special cases of those of the Kalman filter. Similarly, for the Bartlett estimator, the Kalman covariance matrices specialize to those in Equation 24, because the Bartlett estimator has $E(\mathbf{e}_t\mathbf{x}_t') = \mathbf{0}$ (estimation error uncorrelated with true factor scores—see McDonald, 1985, p. 163) and $\mathbf{P}_t = \mathbf{\Gamma}_t^{-1}$ (see Equation 15). Gelb (1974, pp. 112–113) proved the following recursive relation for the Kalman filter:

$$E(\mathbf{e}_t\hat{\mathbf{x}}_t') = (\mathbf{I} - \mathbf{H}_t\mathbf{C}_t)\,\mathbf{A}_{t-1}E(\mathbf{e}_{t-1}\hat{\mathbf{x}}_{t-1}')\,\mathbf{A}_{t-1}' \quad , \tag{27}$$

and hence

$$E(\mathbf{e}_t\hat{\mathbf{x}}_t') = \mathbf{0} \text{ for } E(\mathbf{e}_{t_0}\hat{\mathbf{x}}_{t_0}') = \mathbf{0} \quad . \tag{28}$$

Non-optimal initial estimators could have $E(\mathbf{e}_{t_0}\hat{\mathbf{x}}_{t_0}') \neq \mathbf{0}$, but even then it can be proven, under rather mild conditions, that $E(\mathbf{e}_t\hat{\mathbf{x}}_t')$ goes to $\mathbf{0}$ as time proceeds. It thus holds for the Kalman filter, exactly or with arbitrary accuracy after sufficient time points, that

$$\text{Kalman } E(\hat{\mathbf{x}}_t\mathbf{x}_t') = E(\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t') = \mathbf{\Phi}_t - \mathbf{P}_t = \mathbf{\Phi}_t - (\mathbf{P}_{t-}^{-1} + \mathbf{\Gamma}_t)^{-1} \quad . \tag{29}$$

Because $\mathbf{P}_{t-}^{-1}$ exceeds $\mathbf{\Phi}_t^{-1}$, and $\mathbf{\Phi}_t^{-1}$ exceeds $\mathbf{0}$ by non-negative definite matrices, Equations 23, 24, and 29 show the Kalman $E(\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t')$ to be closer to $\mathbf{\Phi}_t$ than either the regression or the Bartlett $E(\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t')$ is, and the Kalman $E(\hat{\mathbf{x}}_t\mathbf{x}_t')$ to be closer to $\mathbf{\Phi}_t$ than the regression $E(\hat{\mathbf{x}}_t\mathbf{x}_t')$. In summary, the Kalman filter is as good or better than the two other estimators on all four criteria mentioned, except that the Bartlett estimator is better on $E(\hat{\mathbf{x}}_t\mathbf{x}_t')$, but at the cost of being worse on $E(\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t')$ and estimation error variance.

A final point that has received much attention in traditional factor analysis is the indeterminacy of factor scores (Elffers, Bethlehem, & Gill, 1978; Guttman, 1955; McDonald, 1974; Mulaik, 1976). Indeterminacy refers to the same factor model being consistent with different factor score assignments for the examinees in the population, even when their values on the observed variables are assumed to remain the same. This indicates that what has to be estimated by a factor score estimator is, in a certain sense, indeterminate: For a given model and observed value assignment, there exist different but equally good factor candidates (consistent factor score assignments).

A much debated topic in traditional factor analysis is how to measure the amount of indeterminacy or determinacy. McDonald (1974) argues that the determinacy measure should be the squared multiple correlation between the true factor and the observed variables, or equivalently, the squared correlation between the true factor and its least-squares estimator, based on the observed variables. This measure is shown by Mulaik (1976) to give a lower bound to the average correlation between pairs of factor candidates independently drawn from the total set of factor candidates. Guttman (1955) took the correlation between maximally different factor candidates in the total set as a measure of factor determinacy. Both measures are simply related, however: Guttman's correlation is equal to $2\rho^2 - 1$ in terms of McDonald's squared multiple correlation ρ^2. The debate seems motivated primarily by the interpretation of levels: A multiple correlation ρ as high as .707 with corresponding ρ^2 as high as .50 implies a zero determinacy in terms of Guttman's $2\rho^2 - 1$.

The factors or latent variables in $\mathbf{x}_t$ can be standardized with the LISREL program by taking the factor model matrices $\mathbf{C}_t$, $\mathbf{R}_t$, $\mathbf{\Phi}_t$, as well as (in the longitudinal case) $\mathbf{A}_{t-1}$ and $\mathbf{Q}_{t-1}$ from the standardized solution. When this occurs, and $\hat{\mathbf{x}}_t$ is the least-squares estimator of $\mathbf{x}_t$ with $E(\hat{\mathbf{x}}_t\mathbf{x}_t') = E(\hat{\mathbf{x}}_t\,\hat{\mathbf{x}}_t')$, ρ^2 simply equals the corresponding factor's diagonal element in $E(\hat{\mathbf{x}}_t\mathbf{x}_t')$.

In cross-sectional factor analysis, the regression estimator is the least-squares estimator and $E(\hat{\mathbf{x}}_t\mathbf{x}_t')$ in Equation 23 is thus appropriate. In the longitudinal case, $E(\hat{\mathbf{x}}_t\mathbf{x}_t')$ of the Kalman filter in Equation 29 should be chosen: In the longitudinal case, the Kalman filter is the least-squares estimator that minimizes each of the diagonal elements of $\mathbf{P}_t$ (Jazwinski, 1970, p. 203), and thus yields diagonal elements of $E(\hat{\mathbf{x}}_t\mathbf{x}_t')$ as great or greater than any other estimator. In particular, the determinacies in terms of the Kalman filter must be at least as great as those in terms of the regression estimator. This means that extending a cross-sectional factor model longitudinally by making it part of a state space model cannot diminish the determinacy of the latent variables, and actually always increases it. The indeterminacy problem is therefore lessened in the longitudinal case, and depending on the quality of the state space model chosen, it may be lessened considerably in practice.

Educational Research Example

Reading disabilities in primary school children were assessed by means of Kalman filtering, based on a dynamic LISREL model for Beginning Reading. Because the LISREL analysis was performed on the data covariance matrix (computed by the LISREL program from the correlations in Table 1 and the standard deviations in Table 2), and standardized latent variables were chosen using the standardized LISREL solution, the ensuing Kalman filtering is an example of Kalman filtering on the basis of a zero-means LISREL model. The objective was not to estimate absolute increase or decrease in reading ability of individual children, but only the relative change in comparison to a normative group. As indicated above, such a use of the Kalman filter does not require that the latent variables retain exactly the same content over time.

The LISREL Model

The model (see Figure 2) was a longitudinally extended version of the Beginning Reading model described by Mommers and Oud (1984), and by Mommers (1987), and a simplified version of the one in Mommers, van Leeuwe, Oud, and Janssens (1986). The model was estimated by means of ML from a group of 225 Dutch primary school children in first through third grade. It contained three latent state variables: Decoding Speed (DS), Reading Comprehension (RC), and Spelling (SP). These latent state variables were measured at five successive time points by means of 26 observed variables. The intervals between the successive states in Figure 2 were 6-month periods, with the first or initial state $\mathbf{x}_{t_0}$ (comprising DS_1, RC_1, and SP_1) after 7 months of reading instruction, and the fifth state $\mathbf{x}_{t_0+4}$ (comprising DS_5, RC_5, and SP_5) after 31 months of reading instruction. The model shows instantaneous intra-state effects, represented in matrices $\mathbf{K}_t$ of Equation 5, in addition to autoregressive or memory effects between states, represented in matrices $\mathbf{A}^*_{t-1}$ of Equation 5. The model does not contain any deterministic input-effects, and thus $\mathbf{B}^*_{t-1}\mathbf{u}_{t-1} = \mathbf{0}$ in Equation 5.

Model identification and estimation. The LISREL model was identified by means of the following fixed and constrained parameters in the four parameter matrices $\boldsymbol{B}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Theta}$, and $\boldsymbol{\Lambda}$. Except for the 22 free parameters in $\boldsymbol{B}$, which correspond to the arrows between circles in Figure 2, all parameters in $\boldsymbol{B}$ were fixed at 0. Matrix $\boldsymbol{\Psi}$ was specified to be diagonal with free diagonal elements, except for the submatrix of the predetermined variables DS_1, RC_1, and SP_1, which had all elements freed. $\boldsymbol{\Theta}$ was specified to be diagonal. In the eight cases where only a single observed variable was available to measure the latent variable (y_1, y_2, y_5, y_6, y_{10}, y_{16}, y_{17}, y_{21} in Figure 2), the corresponding measurement error variance (diagonal element of $\boldsymbol{\Theta}$) was fixed at $s^2(1 - r)$ (see Table 2). For the remaining 18 observed variables, $s^2(1 - r)$ was used as a starting value for estimating the measurement error variance in the diagonal of $\boldsymbol{\Theta}$.

Table 2 also shows all nonzero fixed, constrained, and free factor loadings in $\boldsymbol{\Lambda}$. The factor loadings correspond to the arrows connecting circles and squares in Figure 2. The usual procedure of fixing columnwise one factor loading at the arbitrary value 1 was used for the five latent DS variables, as well as for the five latent RC variables. For the latent SP variables, however, only one of the factor loadings in the SP_1 column was fixed at 1. The latent variances of the four remaining SP variables could all be identified in the SP_1 metric by specifying equality constraints between the factor loadings of five pairs of observed variables (see Table 2). Each pair came from the same measuring instrument used at two consecutive time points, but the pairs also showed overlap in time (i.e., y_4 measuring SP_1 and y_8 measuring SP_2 came from the same instrument, y_9 measuring SP_2 and y_{14} measuring SP_3 came from the same instrument, and so forth). The five equality constraints were one more than is necessary for identification, because the latent variables SP_3 and SP_4 were connected by two

Table 1
Correlations Between Observed Variables (Decimal Points Omitted)

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}	y_{18}	y_{19}	y_{20}	y_{21}	y_{22}	y_{23}	y_{24}	y_{25}
y_2	502																								
y_3	447	435																							
y_4	378	465	610																						
y_5	834	492	487	358																					
y_6	450	686	383	388	489																				
y_7	461	509	552	507	474	534																			
y_8	430	471	555	567	429	513	603																		
y_9	496	459	603	553	469	505	660	666																	
y_{10}	771	507	487	372	867	500	469	424	477																
y_{11}	349	526	274	288	407	581	454	306	381	461															
y_{12}	433	519	330	355	460	533	443	410	385	531	613														
y_{13}	473	441	583	483	468	415	568	510	605	494	371	385													
y_{14}	419	489	512	534	433	516	527	584	651	508	402	454	642												
y_{15}	461	419	467	424	494	437	542	519	543	529	383	394	603	601											
y_{16}	694	484	423	356	798	473	465	412	466	888	494	494	488	486	510										
y_{17}	500	601	372	415	534	580	510	446	548	585	653	632	487	511	514	616									
y_{18}	400	447	502	536	417	420	524	452	554	503	464	426	674	588	534	527	531								
y_{19}	405	381	387	502	443	412	533	490	527	457	355	384	532	516	598	501	497	673							
y_{20}	418	475	474	520	425	483	520	537	581	489	375	402	591	608	587	497	517	723	750						
y_{21}	612	435	380	297	724	430	381	352	408	830	439	457	416	464	483	881	596	504	459	465					
y_{22}	474	572	349	405	470	543	404	366	401	525	598	601	362	381	447	497	721	461	353	431	466				
y_{23}	405	528	293	375	421	544	358	363	312	484	658	596	325	328	355	496	648	454	354	398	471	717			
y_{24}	404	384	517	469	437	371	510	497	579	518	381	330	586	568	516	554	445	671	576	631	498	383	347		
y_{25}	384	370	420	452	375	369	422	444	484	435	333	357	519	536	522	459	441	587	573	642	420	362	405	670	
y_{26}	483	388	424	348	495	334	502	392	495	572	398	369	617	540	552	568	442	526	513	548	558	394	307	638	578

Table 2
Standard Deviations of Observed Variables s, Reliability Coefficients r From Previous Research, Unstandardized Factor Loadings $\hat{\lambda}$, and One Child's Observed Standard Scores z

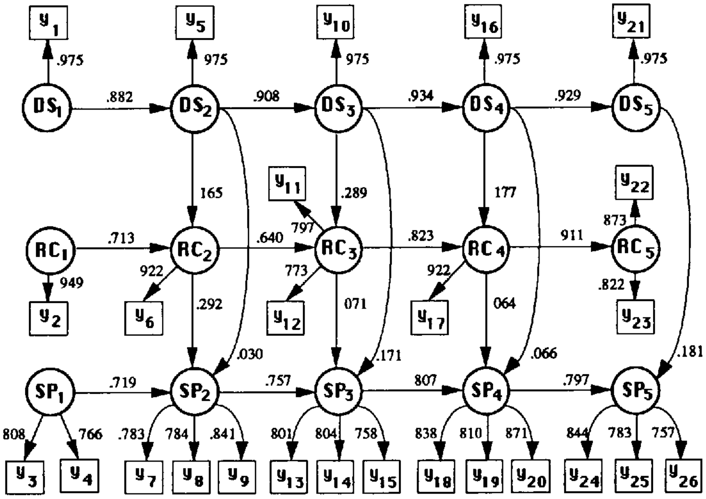
Variable	s	r	$\hat{\lambda}$	z
y_1	13.634	.95[a]	1.0[b]	-.510
y_2	4.601	.90[a]	1.0[b]	-1.709
y_3	4.888	.91	1.0[b]	-.358
y_4	2.907	.89	.563[c]	-.754
y_5	12.985	.95[a]	1.0[b]	-1.190
y_6	3.835	.85[a]	1.0[b]	-.431
y_7	2.391	.65	.526	-.379
y_8	2.554	.82	.563[c]	.050
y_9	3.860	.83	.913[d]	.391
y_{10}	13.131	.95[a]	1.0[b]	-.330
y_{11}	4.736	.90	1.0[b]	-1.109
y_{12}	4.549	.82	.930	-1.200
y_{13}	2.886	.77	1.141[e]	.949
y_{14}	2.274	.75	.913[d]	-.414
y_{15}	3.250	.80	1.251[f]	-.233
y_{16}	13.396	.95[a]	1.0[b]	-.635
y_{17}	5.570	.85[a]	1.0[b]	-.561
y_{18}	3.099	.78	1.141[e]	.470
y_{19}	3.594	.79	1.251[f]	.004
y_{20}	4.350	.83	1.647[g]	-.013
y_{21}	12.888	.95[a]	1.0[b]	-.346
y_{22}	6.347	.88	1.0[b]	-.305
y_{23}	4.281	.83	.635	.337
y_{24}	3.126	.81	2.033	-1.453
y_{25}	2.729	.77	1.647[g]	.188
y_{26}	2.892	.81	1.687	1.099

[a] Measurement error variance fixed at $s^2(1 - r)$.
[b] Fixed value.
[c] Equality constraint between λs of y_4 and y_8.
[d] Equality constraint between λs of y_9 and y_{14}.
[e] Equality constraint between λs of y_{13} and y_{18}.
[f] Equality constraint between λs of y_{15} and y_{19}.
[g] Equality constraint between λs of y_{20} and y_{25}.

equality constraints: one between the factor loadings of y_{13} and y_{18}, and one between those of y_{15} and y_{19}. Because the decrease in χ^2 proved to be less than 1 when one of the two restrictions causing the overidentification was relaxed, time invariance of the factor loadings involved could be confirmed. This gave some evidence that at least the latent variables SP_3 and SP_4 had the same content.

Although the overall fit of the model ($\chi^2 = 445$, $df = 283$, adjusted goodness-of-fit index $= .840$) is perhaps insufficient from a strictly statistical standpoint (formal χ^2 testing of the overall fit should be avoided for several reasons—see Jöreskog & Sörbom, 1984, pp. I.38–39), it may be judged reasonable in view of the large size of the model and the many restrictions placed on it by the state space model. The state space model allows only relatively few parameters to be freed. For example, parameters referring to effects between latent variables backward in time, or to effects overlapping more than one time interval are not allowed. In general, overfitting is not as much of a problem in state space

Figure 2
Dynamic LISREL Model for Beginning Reading
(All Variables Standardized)



modeling using LISREL as it is in cross-sectional LISREL modeling, and the problem decreases still more as the number of time points increases.

In addition to the causal arrows, Figure 2 gives the estimated values of the factor loadings, as well as estimates of the coefficients in the matrices $\mathbf{K}_t$ and $\mathbf{A}^*_{t-1}$. All indicated values were taken from the standardized LISREL solution with the factor loadings being additionally transformed to standardized observed variables. Because each observed variable is only influenced by a single latent variable, the square of each indicated factor loading equals the observed variable's reliability coefficient. The large memory effects of each state variable on itself are very conspicuous. In comparison, the influences between Decoding Speed, Reading Comprehension, and Spelling are low and tend to decrease even more as time proceeds. The application of the Kalman filter for estimating the latent states promises to be advantageous, because of the large memory effects found. One compelling argument for longitudinal research and the use of the Kalman filter is that strong causal effects found cross-sectionally often turn out to decrease or disappear, in favor of memory effects, when dynamic models are used (Oud, 1982).

Comparison of Kalman Filter and Bartlett Estimates

Kalman filter estimates of the latent standard scores for Decoding Speed, Reading Comprehension, and Spelling were determined and compared to the estimates of its main cross-sectional competitor, the Bartlett estimator applied to each time point separately. The computation of the Kalman filter and Bartlett estimates proceeded as follows. First, from the unstandardized LISREL solution, the estimated matrices $\mathbf{K}_t$, $\mathbf{A}^*_{t-1}$, $\mathbf{Q}^*_{t-1}$, $\mathbf{C}_t$, and $\mathbf{R}_t$ were collected. Next, all $\mathbf{K}_t$, $\mathbf{A}^*_{t-1}$, $\mathbf{Q}^*_{t-1}$, and $\mathbf{C}_t$ were transformed to standardized latent variables, as in the standardized LISREL solution. The resulting triples $\mathbf{K}_t$, $\mathbf{A}^*_{t-1}$, and $\mathbf{Q}^*_{t-1}$ for each time point t were reduced to $\mathbf{A}_{t-1}$ and $\mathbf{Q}_{t-1}$. Finally, the Bartlett estimator (using $\mathbf{C}_t$ and $\mathbf{R}_t$) and the Kalman filter (using $\mathbf{A}_{t-1}$, $\mathbf{Q}_{t-1}$, $\mathbf{C}_t$, and $\mathbf{R}_t$) were applied to the children's observed deviation scores (raw observed scores minus mean raw observed scores).

All of the computations were carried out by the LISKAL program, which uses as input LISREL program output (unstandardized LISREL solution) and individual examinees' observed deviation scores (observed raw scores in the case of a structured-means LISREL model).

Because the Bartlett estimator is taken as the initial estimator for the Kalman filter, the Kalman and Bartlett estimates in Figure 3, as well as the corresponding estimation error variances in Table 3, coincide at the initial time point t_0. As expected, in view of its use of past information, the Kalman filter estimates in Figure 3 for an arbitrarily chosen child (see Table 2 for this child's observed standard scores z; the child's observed deviation scores, entering the Kalman filter, can be computed as sz) exhibit more memory and change more cautiously over time than those of the Bartlett estimator. The Bartlett estimates vary much more unpredictably, and because the Kalman filter uses more and more information over time, its estimation error variances in Table 3 (diagonals of $\mathbf{P}_{t_0+1}, \ldots, \mathbf{P}_{t_0+4}$) are considerably smaller than those of the Bartlett estimator. As mentioned in the previous

Figure 3
Latent State Estimates of Decoding Speed, Reading Comprehension, and Spelling for One of the Children by Two Estimators

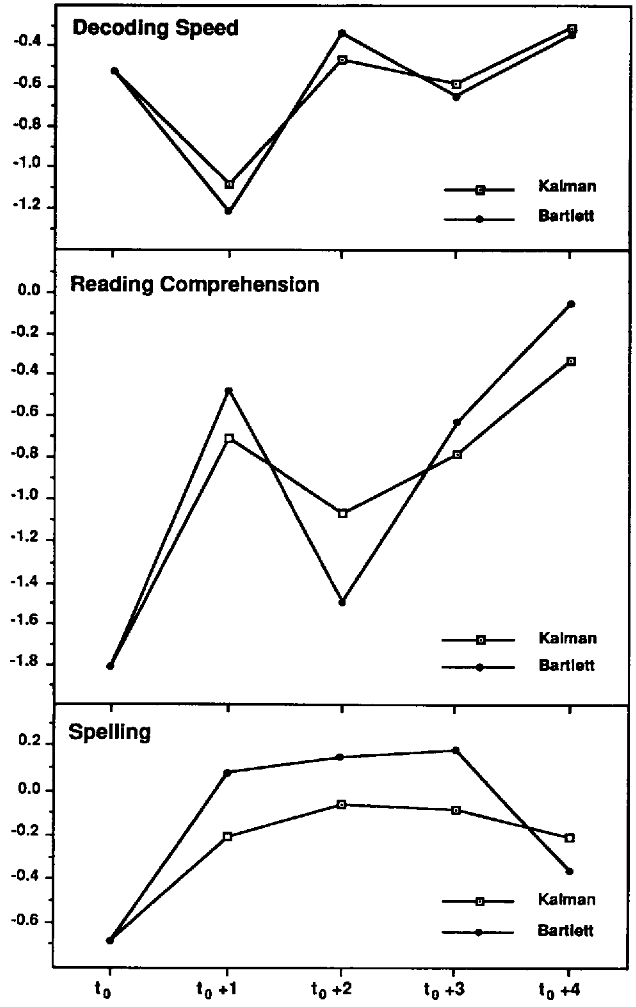


Table 3
Estimation Error Variances

Estimator	Time Point t_0	$t_0 + 1$	$t_0 + 2$	$t_0 + 3$	$t_0 + 4$
Decoding Speed					
Bartlett	.053	.053	.053	.053	.053
Kalman	.053	.044	.042	.040	.040
Kalman from initial value 5.0	5.0	.052	.042	.040	.040
Reading Comprehension					
Bartlett	.111	.180	.320	.188	.197
Kalman	.111	.122	.175	.108	.112
Kalman from initial value 5.0	5.0	.168	.179	.108	.112
Spelling					
Bartlett	.304	.181	.209	.141	.195
Kalman	.304	.116	.116	.095	.109
Kalman from initial value 5.0	5.0	.170	.122	.096	.109

section, the Kalman filter gives the smallest estimation error among all linear estimators. Because the latent variables are standardized in this research example, each diagonal in $\mathbf{I} - \mathbf{P}_t$ may be interpreted as the proportion of latent variance accounted for by the Kalman filter (the proportions in Table 3 are at least .825, and often considerably higher); the square roots of the $\mathbf{P}_t$ diagonals may be used as standard errors for computing interval estimates of an individual child's latent scores.

To test the effect of different initial values on the series of Kalman estimation error variances, the extremely deviant initial value 5 was inserted for all three state variables, in addition to the Bartlett initial values .053, .111, and .304. As Table 3 shows, the convergence proved to be very fast. By the second estimate at time point $t_0 + 2$, none of the differences in estimation error variance was larger than .006. For Decoding Speed, the difference at $t_0 + 2$ was virtually 0. In this latter case, the small measurement error variances of the observed variables for Decoding Speed must have been responsible. Small measurement errors, in fact, have a dual effect on the Kalman filter: First, as mentioned in the previous section, the instantaneous part $\mathbf{H}_t\mathbf{y}_t$ becomes more important, in comparison to the memory part; second, as shown in Equation 12, small values in $\mathbf{R}_t$ cause $\mathbf{P}_t$ to decrease more quickly.

The Kalman filter is specifically designed to give optimal estimates of latent values at the individual level. Its aim is not to estimate latent group means, standard deviations, or correlations; the estimation of such latent group characteristics is more appropriately performed in other ways—for example, by means of the LISREL program. The correlations between the Bartlett estimates and those between the Kalman estimates for the entire sample of 225 children were computed for comparison. They were then compared with the latent state correlations, as estimated by the LISREL program. The LISREL correlations must be considered closer to the true values because the Bartlett and Kalman correlations are only indirect estimates based on LISREL solution matrices, whereas the LISREL solution uses all information on the sample level directly. The differences between the Bartlett and Kalman correlations and the LISREL correlations are given in Table 4. On the average, the Kalman correlations were closer to the LISREL correlations than the Bartlett correlations. The Bartlett correlations also showed the most extreme differences: –.15 between RC_2 and RC_3, –.20 between RC_3 and RC_4, and –.17 between RC_4 and RC_5. The Kalman correlations were particularly good where there were large memory effects in the model: successive correlations between DS_2, DS_3, DS_4, DS_5, between RC_2, RC_3, RC_4, RC_5, and between SP_2, SP_3, SP_4, SP_5. The reasons why the Bartlett correlations were systematical-

Table 4
Kalman Estimate Correlations Minus LISREL Correlations (Above Diagonal) and Bartlett Estimate Correlations Minus LISREL Correlations (Below Diagonal)

	x_{t_0+1}			x_{t_0+2}			x_{t_0+3}			x_{t_0+4}		
State	DS_2	RC_2	SP_2	DS_3	RC_3	SP_3	DS_4	RC_4	SP_4	DS_5	RC_5	SP_5
DS_2		.06	.05	.00	.04	.04	−.01	.00	−.02	−.03	.00	−.02
RC_2	−.02		.03	.11	.05	.04	.12	.03	.02	.10	.06	.01
SP_2	−.01	−.07		.10	.09	.02	.11	.11	.02	.08	.10	.05
DS_3	−.04	.04	.04		.11	.10	.00	.05	.05	.00	.05	.05
RC_3	−.11	−.15	−.06	−.04		.11	.13	−.02	.08	.12	.02	.08
SP_3	−.07	−.11	−.13	.01	−.08		.12	.13	.00	.10	.10	.05
DS_4	−.05	.04	.06	−.05	.00	.03		.10	.10	.00	.08	.09
RC_4	−.10	−.13	.03	−.06	−.20	−.01	−.01		.12	.11	.01	.10
SP_4	−.11	−.10	−.11	−.04	−.08	−.13	.02	−.01		.10	.10	.02
DS_5	−.06	.03	.02	−.04	−.01	.02	−.05	.01	.02		.09	.09
RC_5	−.09	−.06	−.04	−.04	−.10	−.08	−.04	−.17	−.04	−.03		.09
SP_5	−.11	−.13	−.05	−.03	−.09	−.06	.00	−.06	−.13	−.02	−.06	

ly lower than the Kalman correlations are, first, that the Bartlett standard deviations are larger than those of the Kalman filter (the latter deviate in the opposite direction from the true standard deviations; compare Equations 24 and 29), and, second, that the memory component incorporated in the Kalman filter is absent from the Bartlett estimator. For illustrative reasons, Table 5 gives the computed correlations between the estimates at $t_0 + 1$, and the estimates of the same state variable at later time points.

Table 5
Correlation Between Estimates at $t_0 + 1$ and Those at Later Time Points

	Time Point		
Estimator	$t_0 + 2$	$t_0 + 3$	$t_0 + 4$
Decoding Speed			
Bartlett	.867	.798	.724
Kalman	.910	.840	.763
Reading Comprehension			
Bartlett	.622	.580	.585
Kalman	.828	.743	.706
Spelling			
Bartlett	.750	.670	.643
Kalman	.904	.792	.749

Conclusions

These results show that Kalman estimates change more cautiously over time, have lower estimation error variances, and reproduce the LISREL program latent state correlations more precisely. Two points should be stressed, however, concerning the applicability of the Kalman filter in practice.

First, although Kalman filtering requires the availability of a dynamic model, and a more precise model yields smaller estimation errors, no perfect model is needed. Even when all relevant state variables cannot be included, or when some relationships are moderately nonlinear, application of the

Kalman filter instead of the Bartlett estimator usually results in better estimates of the factor scores. As has been pointed out by Lewis (1986, p. 205), the Kalman filter is rather robust for dynamic modeling errors, provided that the process errors are sufficiently large. To prevent the process error variances from being spuriously low, the LISREL model on which the Kalman filter is based should not be overparameterized. This would result, for example, when too many effects of too many state variables are specified. Modeling errors in the state equation, which do not result from overparameterization, would be expected to increase the process error covariance matrix $\mathbf{Q}_{t-1}$, as estimated by the LISREL program, and this would lead to an increased forecast error covariance matrix $\dot{\mathbf{P}}_{t-}$ (see Equation 12). The Kalman $\mathbf{P}_t$ becomes equal to the Bartlett $\mathbf{P}_t$ in the limit $\mathbf{Q}_{t-1} \rightarrow \infty$, and $\mathbf{P}_{t-}^{-1} \rightarrow \mathbf{0}$, when the Kalman filter becomes equal to the Bartlett estimator. There is a built-in mechanism, changing the Kalman filter in the direction of the Bartlett estimator for dynamic modeling errors.

The second point concerns the kind of information that is available when applying the Kalman filter. It has tacitly been assumed here that for the estimation of the state at time point t, no future output information after t is available. This is true when the goal is to obtain an optimal estimate of the most recent state. Then no future information is available, and the Kalman filter estimates cannot be improved upon. However, an improvement of the state estimates at *past* time points t can occur by exploiting output information from subsequent time points $t' > t$, as well as from time points $t' \leq t$. So-called smoothing techniques, instead of the Kalman filter, achieve this effect (Gelb, 1974, pp. 156–159; Lewis, 1986, pp. 127–134). Optimal smoothing—combining (forward) Kalman filtering with a specific type of backward filtering—is outside the scope of this article, but it should be considered when optimal estimates of all values of a developmental curve, including past ones, are desirable.

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.

Browne, M. W. (1977). Generalized least squares estimators in the analysis of covariance structures. In D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socio-economic models* (pp. 205–226). Amsterdam: North-Holland.

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.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105,* 317–327.

Elffers, H., Bethlehem, J., & Gill, R. (1978). Indeterminacy problems and the interpretation of factor analysis results. *Statistica Neerlandica, 32,* 181–199.

Gelb, A. (Ed.). (1974). *Applied optimal estimation.* Cambridge MA: MIT Press.

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

Goldberger, A. S. (1964). *Econometric theory.* New York: Wiley.

Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common factor theory. *British Journal of Statistical Psychology, 8,* 65–82.

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

Heise, D. R. (1975). *Causal analysis.* New York: Wiley.

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. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. II, pp. 1-56). San Francisco CA: Freeman.

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

Jöreskog, K. G. (1979). A general approach to confirmatory maximum likelihood factor analysis (with addendum). In K. G. Jöreskog & D. Sörbom, *Advances in factor analysis and structural equation models* (pp. 21-43). Cambridge MA: Abt Books.

Jöreskog, K. G., & Sörbom, D. (1977). Statistical models and methods for analysis of longitudinal data. In D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socio-economic models* (pp. 285-325). Amsterdam: North-Holland.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood: User's guide.* Mooresville IN: Scientific Software, Inc..

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, The Netherlands: 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: Butterworths.

Lewis, F. L. (1986). *Optimal estimation.* New York: Wiley.

McDonald, R. P. (1974). The measurement of factor indeterminacy. *Psychometrika, 39,* 203-222.

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale NJ: Erlbaum.

Mommers, M. J. C. (1987). An investigation into the relation between word recognition skills, reading comprehension and spelling skills in the first two years of primary school. *Journal of Research in Reading, 10,* 122-143.

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 search of the relation between decoding speed, reading comprehension, and spelling in the first two years of primary school]. In L. F. W. de Klerk & A. M. P. Knoers (Eds.), *Onderwijspsychologisch onderzoek* (pp. 193-202). Lisse: Swets & Zeitlinger.

Mommers, M. J. C., van Leeuwe, J. F. J., Oud, J. H. L., & Janssens, J. M. A. M. (1986). Decoding skills, reading comprehension and spelling: A longitudinal investigation. *Tijdschrift voor Onderwijsresearch, 11,* 97-113.

Mulaik, S. A. (1976). Comments on "The measurement of factor indeterminacy." *Psychometrika, 41,* 249-262.

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 modeling, filtering and self-tuning control of stochastic systems.* New York: Springer-Verlag.

Oud, J. H. L. (1978). *Systeem-methodologie in sociaal-wetenschappelijk onderzoek [Systems methodology in social science research].* Nijmegen, The Netherlands: Alfa.

Oud, J. H. L. (1981). Onderzoek van orthopedagogische en onderwijskundige interventies aan de hand van tijdreeksen: een MANOVA-procedure [Evaluation of orthopedagogical and educational interventions by means of time-series research: A MANOVA procedure]. *Tijdschrift voor Onderwijsresearch, 6,* 267-291.

Oud, J. H. L. (1982). De stap van input-output naar toestand in sociaal-wetenschappelijk onderzoek [The step from input-output to state in social science research]. In A. F. G. Hanken & J. H. L. Oud (Eds.), *Onderneming en overheid in systeemdynamisch perspectief* (pp. 17-33). 's-Gravenhage: Academic Service.

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

Sörbom, D. (1979). An alternative to the methodology for analysis of covariance. In K. G. Jöreskog & D. Sörbom, *Advances in factor analysis and structural equation models* (pp. 219-234). Cambridge MA: Abt Books.

Theil, H. (1971). *Principles of econometrics.* New York: Wiley.

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

Acknowledgments

The authors thank Caesarius Mommers for his permission to use the Beginning Reading model results in the research example, Jaap Molenaar for his solutions to several statistical problems associated with the Kalman filter, Melissa Bowerman, Giulio Lancioni, Jan van Leeuwe, and two anonymous reviewers for their helpful comments on earlier versions of this article.

Author's Address

Send requests for reprints, further information, or a free listing of the LISKAL program to Johan H. Oud, Instituut voor Orthopedagogiek, KU Nijmegen, Erasmusplein 1, 6500 HD Nijmegen, The Netherlands.