Kwantitatieve Methoden 69, 2002, 1-26 http://www.vvs-or.nl/kwanmeth/km69/oud.pdf

Continuous Time Modeling of the Cross-Lagged Panel Design

Johan H.L. Oud

University of Nijmegen, Institute of Special Education,
PO Box 9104, 6500 HE Nijmegen, The Netherlands,
phone: +31243230930, email: j.oud@ped.kun.nl,
URL: http://www.socsci.kun.nl/~hano/

Abstract

Since Newton (1642-1727) continuous time modeling by means of differential equations is the standard approach of dynamic phenomena in natural science. It is argued that most processes in behavioral science also unfold in continuous time and should be analyzed accordingly. After dealing with the essentials of stochastic differential equation modeling of panel data by means of structural equation modeling (SEM) and the exact discrete model (EDM) as well as with the relevance of cross-sectional research for continuous time modeling (Coleman), several paradoxes specifically related to the cross-lagged panel design in discrete time are addressed and solved in continuous time. These paradoxes have been the cause of a lot of confusion in the interpretation of cross-lagged panel analyses in the literature. Educational research data illustrate and evaluate continuous time modeling of cross-lagged effects by means of different models and methods using SEM.

Key Words and Phrases: continuous time modeling, cross-lagged panel design, exact discrete model, stochastic differential equation, structural equation modeling.

Introduction

A quick look at the text books of longitudinal data analysis (e.g. Bijleveld & van der Kamp, 1998; Collins & Horn, 1991) learns that the continuous time approach by means of differential equations is almost nonexistent in behavioral science. Most prominent methods of longitudinal data analysis for quantitative variables are analysis of variance (ANOVA and MANOVA), multilevel analysis, and structural equation modeling.

When corresponding error covariance structures are specified, MANOVA and multilevel analysis become identical (Bijleveld & van der Kamp, 1998) and typically amount to random-coefficient curve fitting with time as pseudo-treatment or (within-subjects) "predictor" variable and possibly with measurement error (Bryk & Raudenbush, 1992; Mehta & West, 2000; McArdle & Hamagami, 1996; Miyazaki & Raudenbush, 2000). Most often polynomial curve fitting is used, before or after some suitable transformation of the time variable (Willett & Sayer, 1996). In practical applications one tries to avoid higher-order polynomial components in favor of those of the simple linear model: random intercepts and random slopes only, allowing a clear interpretation of individual curves in terms of growth or decline (Bryk & Raudenbush, 1992, p. 148; McArdle & Hamagami, 1996, p. 111). By choosing time itself as "predictor" instead of the past state of the system, the polynomial curve fitting model lacks the causality direction from past to present and therefore the ability to perform stability analysis for the model. An exception is the linear model that by definition is unstable, however. The linear model is extremely unrealistic in implying for every subject unlimited growth in one direction over time and unlimited decline in the opposite direction. Mehta and West (2000) use this to prove for the population of subjects in the linear model two other unrealistic features: (a) unlimited quadratic growth of the variance in the two time directions from a minimum variance time point in between, (b) an increasingly negative (positive) correlation between intercept and slope depending on how far the time scale origin is chosen before (after) this point. The frequently found negative correlation between initial status and growth, often presented as a remarkable empirical finding, law or paradox (Bryk & Raudenbush, 1992, pp. 137-138; Wilder, 1967; Willett & Sayer, 1996, pp. 147-149), turns out to be a simple artifact of the linear model and dependent on whether and how far the time scale origin is located before the minimum variance time point.

Structural equation modeling (SEM) is the more general one of the three methods mentioned. In addition to MANOVA and multilevel models (Willett & Sayer, 1994), difference equation models are frequently specified and estimated by means of SEM (Jöreskog, 1978; Jöreskog & Sörbom, 1985; MacCallum & Ashby, 1986). Difference equation models share the causal-dynamic character with differential equation models, both restricting the role of time to the specification of the temporal direction of effects from past to present (nonanticipative). The treatment of stochastic differential equation models, however, especially by means of SEM (Arminger, 1986; Oud & Jansen, 2000), is extremely rare and a summary will be given first in the present article.

In addition to the special arguments for continuous time modeling, to be discussed in connection with the cross-lagged panel design in behavioral science, the basic arguments expounded by its pioneers in econometrics are valid for behavioral science as well. In an interview by his pupil Phillips (1993, p. 23), Bergstrom emphasized: "The economy does not cease to exist in between observations. Nor does the economy move in regular discrete jumps, at quarterly or annual intervals corresponding to the observations.... Many economic variables, particularly financial variables, do make discrete jumps and those jumps can occur at any point in time." Gandolfo (1993, pp. 2-3) added with regard to the observation periods: "It is necessary to check that no essential result of a discrete model depends on the actual time-length of the period (the model should give the same results when such a period is, say doubled or halved). But, if the results are unvarying with respect to the period length, they should remain valid when this length tends to zero (that is, when one switches over from discrete to continuous analysis)." In a historical overview of the development and application of continuous time stochastic models in the 20th century, Bergstrom (1988) mentions crucial problems that had to be overcome. A first milestone is Wiener's (1923) rigorous mathematical description of the Brownian motion, the erratic behavior of particles in a liquid, discovered by the English botanist Brown in 1828 and later on called Wiener process. Two other milestones are the solution by Itô (1944, 1951) of the stochastic differential equation involving the Wiener process and the derivation by Bergstrom himself in 1961-1962 of the exact discrete model to solve the problem of estimating the parameters of continuous time stochastic models from discrete data.

The panel design has been proposed by Lazarsfeld (Lazarsfeld & Fiske, 1938; Lazarsfeld, 1940) to circumvent the difficult problem of assessing the direction of effects between variables x and y in cross-sectional research. The cross-lagged panel design was designed to study and compare the effects of variables on each other in opposite directions over time. Crucial for panel research is that the causal direction is not based on instantaneous relationships between simultaneously measured variables x and y but that different variables are available for opposite directions: x_{t_1} and y_{t_2} for the causal direction from x to y, y_{t_1} and x_{t_2} for the opposite direction from y to x. It is therefore supposed to be much more suitable than cross-sectional research in answering, for example, whether the advertisement of a particular commodity brand causes people to buy or, conversely, its frequent consumption causes people to remark its advertisement, or both (Lazarsfeld, 1940). Lazarsfeld's preference of the cross-lagged panel design to cross-sectional research for this type of causal analysis closely parallels Wold's (1954) later compelling argument in economics to replace interdependencies by recursive formulations in terms of lagged variables.

After discussing why continuous time modeling of cross-effects is not easily done in cross-sectional research either, several paradoxes, connected to the cross-lagged panel design in discrete time, will be explained and shown to be solvable in continuous time. First, the cross-lagged panel design offers estimates of both cross-lagged coefficients and instantaneous ones (e.g., by means of SEM). In discrete time, it is not clear how the two different sets should be combined in a unitary measure for the causal effects. Second, different researchers, studying the same causal effect in different discrete time distances, are even unable to compare the strength of the causal effects found in discrete time. Finally, simulation research shows that in discrete time the strength and order of magnitude of the cross-effects (e.g., cross-lagged effect of x on y and of y on x) heavily depend on the discrete time interval chosen by the researcher and may even reverse when appropriately studied in continuous time.

Having attracted in the past most attention in fields like sociology and economics, the cross-lagged panel design is becoming increasingly popular in psychology as well. Rueter and Conger (1998), for example, make clear that correlations between parental and children's behavior, which in the past invariably were interpreted as unidirectional influences from parents to children, in recent years typically have got a reciprocal causal interpretation and have led to a lot of cross-lagged panel research. The cross-lagged analyses, however, are performed in discrete time and their evaluation turns out to suffer precisely from incomparability and seemingly contradictory outcomes as a result of different discrete time observation intervals within and between studies (Gollob & Reichardt, 1987; Lorenz, Conger, Simons, & Whitbeck, 1995; Sher, Wood, Wood, & Raskin, 1996). The study of Vuchinich, Bank, and Patterson (1992), assessing cross-lagged as well as instantaneous effects between parental disciplinary behavior and child antisocial behavior, led to additional confusion by the fact that the results for both kinds of effects differed and significance was found only for the instantaneous effects.

In the present article, continuous time modeling of cross-lagged effects by means of SEM will be illustrated in several ways and with several models, using a longitudinal data set from education and addressing the question whether and how strongly in initial reading Decoding Skill influences Reading Comprehension, Reading Comprehension influences Decoding Skill, or both.

Continuous Time Differential Equation Modeling in Deterministic and Stochastic Form

For more than three centuries in most fields of science the standard approach of dynamic phenomena is continuous time modeling by means of differential equations. One well-known example is the application of Newton's (1642-1727) famous second law of motion to describe the pendulum motion by means of the second order differential equation:

$$\frac{d^2y(t)}{dt^2} = -\frac{1}{m}\left[\gamma\frac{dy(t)}{dt} + \omega^2 y(t)\right] . \tag{1}$$

Acceleration is inversely proportional to mass m and directly proportional to the external force exerted, which in this case depends on velocity $dy(t)/dt$ as well as location $y(t)$. Equation (1) describes a continuous-time second-order autoregressive process, called CAR(2)-process. Written as a state space model for $x_1(t) = y(t)$, $x_2(t) = dy(t)/dt$:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) \qquad \text{with} \qquad \mathbf{A} = \begin{bmatrix} 0 & 1 \\ -\omega^2/m & -\gamma/m \end{bmatrix} , \tag{2}$$

the eigenvalues of the so-called drift matrix $\mathbf{A}$,

$$\lambda_{1,2} = a \pm bj = -\frac{\gamma}{2m} \pm j\sqrt{\frac{\omega^2}{m} - \frac{\gamma^2}{4m^2}} ,$$

can be computed to give important information about the dynamic properties of the model. For small friction values γ, the eigenvalues $a \pm bj$ (j the imaginary number $\sqrt{-1}$) become complex, implying that the pendulum follows a sinusoidally oscillating movement. For zero friction ($\gamma = 0$), the pendulum oscillates constantly with angular frequency $b = \omega/\sqrt{m}$ and time period $T = 2\pi/b$, where ω inversely relates to the length of the pendulum rod. Because the real part a of the eigenvalues can never be positive ($a \leq 0$), the pendulum is inherently stable. For $a < 0$ and thus $\gamma > 0$ it is even asymptotically stable, showing a dampened oscillation converging to the equilibrium position. Finally, sufficiently large friction values result in real eigenvalues and cause the pendulum to go to the equilibrium position without oscillating.

The asymptotic stability requirement of all eigenvalues of $\mathbf{A}$ having negative real part is the direct continuous time analog of the corresponding discrete time requirement of all eigenvalues being located within the unit circle. In both cases it means that the trajectory converges to the steady state trajectory. Stability analysis is of prime importance in judging the plausibility of a model. Nonstability or diverging behavior in a time-invariant model means that the system more or less quickly would explode according to the model. In most cases this is not realistic, so that a new and better model has to be found, or not desirable, so that control measures should be taken to prevent the system from diverging. For time-varying models with time-invariant matrix $\mathbf{A}$ in Equation (2) replaced by time-varying one $\mathbf{A}(t)$, stability becomes more flexible, as the eigenvalues become time-varying too and trajectories may stay within boundaries, even when eigenvalues become temporarily positive.

Examples of continuous time oscillating and non-oscillating behavior patterns are not less abundant in economics and psychology than in physics. Temper and feelings of depression, for example, are oscillating phenomena to be modeled accordingly with higher oscillation frequency in some people than in other ones, while also the amplitude (strength) as well as the phase of the oscillations, determined by initial conditions, are allowed to differ in different people. Typically, however, the description of oscillating as well as non-oscillating behavior patterns by Equation (2) turns out to be less than perfect in fields like economics and psychology as well as in special fields of physics. Continuous time noise is needed and found in the famous Wiener process $\mathbf{W}(t)$ or limiting form of the discrete time random walk process (Jazwinski, 1970, pp. 70-74):

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{G}(t)\frac{d\mathbf{W}(t)}{dt} \ . \tag{3}$$

A connection between continuous and discrete time noise is made by formally deriving from Equation (3) solution

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{A}\mathbf{x}(s)d(s) + \int_{t_0}^{t} \mathbf{G}(s)d\mathbf{W}(s) \tag{4}$$

or from simplest possible stochastic equation

$$\frac{d\mathbf{x}(t)}{dt} = \frac{d\mathbf{W}(t)}{dt} \tag{5}$$

random walk process

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} d\mathbf{W}(s) = \mathbf{x}(t_0) + [\mathbf{W}(t) - \mathbf{W}(t_0)] = \mathbf{x}(t_0) + \Delta\mathbf{W}_t, \tag{6}$$

which can be interpreted in continuous time as well as in each arbitrary subset of discrete time points in continuous time.

Assuming (a) all arbitrary (but nonoverlapping) random steps or increments $\Delta\mathbf{W}_t = [\mathbf{W}(t) - \mathbf{W}(t_0)]$ to be independent and stationary (zero mean, variance Δt) and (b) the sample trajectories $\mathbf{W}(t)$ to be continuous with initial value $\mathbf{W}(0) = \mathbf{0}$, in fact, defines the Wiener process (Arnold, 1974, p. 46) and is the basis of most of continuous time modeling results. The Wiener process thus puts the customary discrete time assumptions (independence and zero mean) on the noise in every subset of discrete time points in continuous time. In addition it requires continuity for the sample trajectories. However, Bergstrom (1983), while preserving the standard results, was able to relax the continuity assumption under rather general conditions, allowing the noise to come also in the form of “discrete jumps at random intervals of time rather than in the form of Brownian motion” (Bergstrom, 1988, p. 366). It should be further noted that the standardizing assumption (variance Δt and thus 1 for the standard discrete time interval $\Delta t = 1$) is no real restriction, because the matrix $\mathbf{G}(t)$ in Equation (3) allows transformation to arbitrary variances and covariances, nor is $\mathbf{W}(0) = \mathbf{0}$ for time-invariant models, because the stationarity of the increments allows the time scale origin to be located arbitrarily.

It is well-known that the “white noise” process $d\mathbf{W}(t)/dt$ in Equations (3) and (5) is not a stochastic process in the ordinary sense (see e.g. Arnold, 1974, p. 48). Nor can the stochastic integral $\int_{t_0}^{t} \mathbf{G}(s)d\mathbf{W}(s)$ in Equation (4) be defined as an ordinary Riemann-Stieltjes integral. An idea of the peculiar nature of continuous-time noise is obtained by realizing that the random walk steps, which occur only once per unit time interval in discrete time, are repeated over infinitesimally small intervals in continuous time. The independence of the increments over arbitrarily small intervals causes the sample trajectories to be of infinite length over any finite interval. Therefore they are not of bounded variation and nowhere differentiable in the traditional sense (Gard, 1988, p. 33). Fortunately, however, there are alternative ways to define and solve the stochastic integral $\int_{t_0}^{t} \mathbf{G}(s)d\mathbf{W}(s)$ as shown by Itô, Bergstrom and others. Itô did so for the general case of stochastic, time-varying matrices $\mathbf{G}(t)$. In this article only fixed but possibly time-varying matrices $\mathbf{G}(t)$ will be assumed, while the trajectories studied are considered to be sufficiently smooth to retain the continuity assumption.

Estimating the Continuous Time Differential Equation Model by Means of a Multiple Subject Sample

To estimate the stochastic differential equation model in the case of a sample of multiple subjects, Arminger (1986) employed the so-called “indirect” method, which consists of first estimating discrete time parameters by means of an SEM program and then in a second step deriving the continuous time parameter values using the Exact Discrete Model (EDM). The EDM, introduced in econometrics by Bergstrom (1988), links in an exact way the discrete time model parameters to the underlying continuous time model parameters by means of nonlinear restrictions. Because SEM programs like LISREL (Jöreskog & Sörbom, 1996) are not able to impose the necessary matrix exponential nonlinear constraints between the continuous time and discrete time parameters in the EDM directly during estimation, the indirect method was criticized by Hamerle, Nagl, and Singer (1991). Oud and Jansen (2000), however, show how recent standard SEM software packages like Mx (Neale, 1997) and MECOSA (Arminger, Wittenberg, & Schepers, 1996) can be employed for maximum likelihood estimation of the continuous-time state space model parameters, using the direct method: applying the nonlinear constraints of the EDM directly during estimation. Additionally, Oud and Jansen (2000) generalize the EDM to cover not only time-invariant parameters but also the cases of stepwise time-varying (piecewise time-invariant) parameters and parameters varying continuously over time according to a general polynomial scheme.

We start from the following somewhat extended version of the stochastic differential equation in Equation (3), describing the state trajectory in continuous time with possibly time-varying parameter matrices:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}(t)\mathbf{x}(t) + \boldsymbol{\gamma} + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{G}(t)\frac{d\mathbf{W}(t)}{dt} \ . \tag{7}$$

The drift matrix $\mathbf{A}(t)$ specifies how the change in the state $\mathbf{x}(t)$ depends on the state itself. Because in SEM the asymptotics is for the subject sample N instead of the number of time points T going to ∞, no stability or stationarity assumption is needed for the state. It means stability can be estimated for the model instead of being an estimation requirement. Effects $\mathbf{B}(t)\mathbf{u}(t) \neq \mathbf{0}$ of fixed input variables in $\mathbf{u}(t)$ accommodate for nonzero and nonconstant mean trajectories $E[\mathbf{x}(t)]$. By the specification of random subject effects $\boldsymbol{\gamma} \neq \mathbf{0}$ subject specific conditional mean trajectories are obtained, keeping a subject specific distance from $E[\mathbf{x}(t)]$. The zero mean normally distributed variables in $\boldsymbol{\gamma}$ can be viewed as a special kind of (unobserved and constant over time) state variables, sometimes called "trait" variables.

For nonstochastic $\mathbf{G}(t)$ the following solution of Equation (7) is obtained (Arnold, 1974, pp. 128-134; Ruymgaart & Soong, 1985, pp. 80-99):

$$\begin{aligned}\mathbf{x}(t) = \boldsymbol{\Phi}(t,t_0)\mathbf{x}(t_0) + \int_{t_0}^{t}\boldsymbol{\Phi}(t,s)\boldsymbol{\gamma}ds + \int_{t_0}^{t}\boldsymbol{\Phi}(t,s)\mathbf{B}(s)\mathbf{u}(s)ds + \int_{t_0}^{t}\boldsymbol{\Phi}(t,s)\mathbf{G}(s)d\mathbf{W}(s) \\ \text{with cov}\left[\int_{t_0}^{t}\boldsymbol{\Phi}(t,s)\mathbf{G}(s)d\mathbf{W}(s)\right] = \int_{t_0}^{t}\boldsymbol{\Phi}(t,s)\mathbf{G}(s)\mathbf{G}'(s)\boldsymbol{\Phi}'(t,s)ds \ .\end{aligned} \tag{8}$$

Central in the solution is the state transition matrix $\boldsymbol{\Phi}(t,t_0)$, defined by $\frac{d\boldsymbol{\Phi}(t,t_0)}{dt} = \mathbf{A}(t)\boldsymbol{\Phi}(t,t_0)$, $\boldsymbol{\Phi}(t_0,t_0) = \mathbf{I}$ and solving the deterministic homogeneous part $\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}(t)\mathbf{x}(t)$ of Equation (7).

Estimating the continuous time parameters in Equation (7) on the basis of discrete time panel data by means of SEM requires the EDM to be derived first. The EDM relates the discrete time parameters in the discrete time state Equation (9),

$$\mathbf{x}_t = \mathbf{A}_{t-\Delta t}\mathbf{x}_{t-\Delta t} + \boldsymbol{\kappa} + \mathbf{B}_{t-\Delta t}\mathbf{u}_{t-\Delta t} + \mathbf{w}_{t-\Delta t} \text{ with cov}(\mathbf{w}_{t-\Delta t}) = \mathbf{Q}_{t-\Delta t} \ , \tag{9}$$

to the underlying continuous time parameters via an exact expression for the solution in Equation (8). For example, for a time-invariant $\mathbf{A}(t) = \mathbf{A}$ over discrete time observation interval $\Delta t = t - t_0$, the lagged autoregression matrix $\mathbf{A}_{t-\Delta t}$ in the EDM takes on the well-known exact exponential form $\mathbf{A}_{t-\Delta t} = \boldsymbol{\Phi}(t,t_0) = e^{\mathbf{A}\Delta t}$. Detailed solutions and derivations for time-invariant as well as time-varying parameter matrices are given by Oud and Jansen (2000).

In many cases the state variables in vector $\mathbf{x}(t)$ of the EDM state Equation (9) cannot be directly observed and a so-called output or measurement equation has to be added to the EDM:

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{D}_t\mathbf{u}_t + \mathbf{v}_t \ \text{ with } \ \text{cov}\ (\mathbf{v}_t) = \mathbf{R}_t \ . \tag{10}$$

It is interesting that in those cases all state variables, at as well as between observation time points, become latent. In conjunction with the nonlinear restrictions involved, Equations (9) for successive time points are introduced as the structural equation part of the SEM model and Equations (10) as its measurement equation part.

Continuous Time Modeling by Means of Cross-Sectional Research

Before going into details about continuous time modeling in the cross-lagged panel design, we will first examine whether and which contribution cross-sectional research could make to causal inference from a continuous time perspective, including the assessment of reciprocal causal influence. Starting from a simplified time-invariant and deterministic version of Equation (7):

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \ , \tag{11}$$

Coleman (1968) tried to make sense out of cross-sectional research by deriving under the assumption of equilibrium, $\frac{d\mathbf{x}(t)}{dt} = \mathbf{0}$ and thus $\mathbf{0} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$, the cross-sectional reduced form equation

$$\mathbf{x}(t) = \mathbf{B}^{*}\mathbf{u}(t) \quad \text{with} \quad \mathbf{B}^{*} = -\mathbf{A}^{-1}\mathbf{B} \ . \tag{12}$$

Coleman (1968, p. 444) is very clear about the meaning of Equation (12): "The cross-section analysis assumes, either implicitly or explicitly, that the causal processes have resulted in an equilibrium state."

The cross-sectional regression coefficients in $\mathbf{B}^{*}$ of Equation (12) do not have a simple and informative relationship to the underlying dynamic coefficients in $\mathbf{A}$ and $\mathbf{B}$. However, the coefficients in Equation (13), which is easily seen to be the structural form of a structural equation model with Equation (12) as the reduced form, do:

$$\mathbf{x}(t) = \mathbf{A}^{*}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad \text{with} \quad \mathbf{A}^{*} = \mathbf{I} + \mathbf{A} \ . \tag{13}$$

The reduced form coefficients in $\mathbf{B}^{*}$ (Equation (12)) and the structural form coefficients in endogenous or instantaneous effects matrix $\mathbf{A}^{*}$ and exogenous effects matrix $\mathbf{B}$ (Equation (13)) have the well-known relationship $\mathbf{B}^{*} = (\mathbf{I} - \mathbf{A}^{*})^{-1}\mathbf{B}$. So, after estimating for obvious stochastic versions of Equations (11)-(13) the structural and therefore immediately the differential form, the reduced form can be derived in the customary way.

For a nonrecursive model, matrix $\mathbf{A}^{*} = \mathbf{I} + \mathbf{A}$ in Equation (13) contains the famous interdependencies (instantaneous reciprocities), extensively discussed in econometrics and heavily criticized by people like Wold (1954) and Meissner (1971). Their criticism involved exactly the assumption Coleman formulated for enabling cross-sectional coefficients to be interpreted causally-dynamically. They too made clear that subjects should be in a (stable) equilibrium state for the interdependencies to make sense and to be interpretable dynamically. The problem is that the equilibrium assumption as well as the implied stationarity and time invariance of the model seldom apply in practice. In their view the model assumes what has to be proven dynamically and what in most cases turns out to be false.

Additionally, as a result of the lacking longitudinal information, the structural form matrix $\mathbf{A}^{*}$ is often non-identified, especially with regard to the reciprocal causal effects one is interested in. It means that in practice it is often impossible to disentangle the true causal-dynamic coefficients of $\mathbf{A}$ and $\mathbf{B}$ in the structural cross-sectional form (Equation (13)) from the mixtures $\mathbf{B}^* = -\mathbf{A}^{-1}\mathbf{B}$ in the reduced cross-sectional form (Equation (12)).

Finally, for identification reasons it is customary in cross-sectional analysis to fix the values of the self-loop coefficients or diagonals of $\mathbf{A}^*$ at 0 and therefore the diagonals of $\mathbf{A} = \mathbf{A}^* - \mathbf{I}$ at -1. Although these negative feedback values of -1 are consistent with the assumption of the subjects being in a stable equilibrium state, it is a highly restrictive specification as infinitely many other negative values would also be consistent with this assumption. If therefore the diagonal elements of $\mathbf{A}$ should be said to be only rough approximations of the true ones, the interconnectedness between the elements of $\mathbf{A}$ means that also the other elements are only approximate in cross-sectional research. It is possible in SEM to fix the diagonals of $\mathbf{A}^*$ at other values than 0 and thus the diagonals of $\mathbf{A}$ at other values values than -1. However, because the values to fix at are usually unknown and in need of estimation themselves, this usually is of no help in practice.

Because of these reasons, in spite of the illuminating connection Coleman made between cross-sectional and longitudinal models, cross-sectional research does not seem to be more suitable for causal-dynamic inference from a continuous time perspective than from a discrete time perspective. Nevertheless, Equation (13) is one of the ways proposed to estimate the continuous time differential equation parameters and will be called Approximate Cross-sectional Model (ACM) in the sequel.

How to Choose from Two Sets of Coefficients in Discrete Time Modeling

Instead of EDM Equation (9) or its time-invariant version

$$\mathbf{x}_t = \mathbf{A}_{\Delta}\mathbf{x}_{t-\Delta t} + \boldsymbol{\kappa} + \mathbf{B}_{\Delta}\mathbf{u}_{t-\Delta t} + \mathbf{w}_{t-\Delta t} \ , \tag{14}$$

analysts in discrete time modeling often choose the structural form, which for Equation (14) will be written

$$\mathbf{x}_t = \mathbf{A}_0\mathbf{x}_t + \mathbf{A}_\ell\mathbf{x}_{t-\Delta t} + \boldsymbol{\kappa}_\ell + \mathbf{B}_\ell\mathbf{u}_{t-\Delta t} + \mathbf{w}_{\ell,t-\Delta t} \ . \tag{15}$$

Equation (15) contains simultaneously two kinds of effect coefficients between the state variables: instantaneous coefficients in matrix $\mathbf{A}_0$ and lagged coefficients in matrix $\mathbf{A}_\ell$. For each of the effects the discrete time analyst could choose the instantaneous coefficient, the lagged coefficient or both to be present in the model. Estimating both and deciding on the basis of statistical testing which one to retain, is often prevented by identification problems. The choice is made the more difficult for the discrete time analyst, as it is evident that estimating both without constraints in a model with both identified, would give results that are highly dependent on the time interval Δt. In general, the larger the distance Δt of lagged value $\mathbf{x}_{t-\Delta t}$ from the current $\mathbf{x}_t$, the higher the instantaneous coefficients become at the expense of the lagged coefficients. Distinguishing the causally-dynamically relevant interval δt from the observation interval Δt and realizing that typically the former is somewhere between 0 and Δt, most analysts feel that $\mathbf{A}_0$ and $\mathbf{A}_\ell$ should both be taken into consideration somehow. The problem is how to connect and constrain the elements in $\mathbf{A}_0$ and $\mathbf{A}_\ell$ to find the true underlying coefficients in $\mathbf{A}$.

The problems described are easily solved, when the true δt is identified with the infinitesimal dt of continuous time. From Equation (15) (structural form) Equation (14) (reduced form) turns out to be

$$\mathbf{x}_t = \left[(\mathbf{I}-\mathbf{A}_0)^{-1}\mathbf{A}_\ell\right]\mathbf{x}_{t-\Delta t} + \boldsymbol{\kappa} + \left[(\mathbf{I}-\mathbf{A}_0)^{-1}\mathbf{B}_\ell\right]\mathbf{u}_{t-\Delta t} + \mathbf{w}_{t-\Delta t} \tag{16}$$

and, assuming the input to be constant over the observation interval in Equation (8), one derives the EDM parameter matrices $\mathbf{A}_\vartriangle$ and $\mathbf{B}_\vartriangle$ to be:

$$\begin{aligned} \mathbf{A}_\vartriangle &= (\mathbf{I}-\mathbf{A}_0)^{-1}\mathbf{A}_\ell \;=\; e^{\mathbf{A}\Delta t}\,, \\ \mathbf{B}_\vartriangle &= (\mathbf{I}-\mathbf{A}_0)^{-1}\mathbf{B}_\ell \;=\; \mathbf{A}^{-1}(e^{\mathbf{A}\Delta t}-\mathbf{I})\mathbf{B}\,. \end{aligned} \tag{17}$$

This shows that there are clear, though highly nonlinear, relationships between the coefficients in reduced form matrices $\mathbf{A}_\vartriangle, \mathbf{B}_\vartriangle$ as well as structural form matrices $\mathbf{A}_0, \mathbf{A}_\ell, \mathbf{B}_\ell$ on the one hand, and those in the underlying continuous time matrices $\mathbf{A}, \mathbf{B}$ on the other hand. From this standpoint also, there is no need to put the constraints explicitly on the structural form, in view of the possibility to put them directly and exactly on the reduced form by means of the EDM.

A reason to use nevertheless the structural form model is provided and explained by Bergstrom (1984, pp. 1172-1173). He shows that the following simple linear constraints,

$$\begin{aligned} \mathbf{A}_0 &= \tfrac{1}{2}\,\tilde{\mathbf{A}}\Delta t\,, \\ \mathbf{A}_\ell &= \mathbf{I} + \tfrac{1}{2}\tilde{\mathbf{A}}\Delta t\,, \\ \mathbf{B}_\ell &= \tilde{\mathbf{B}}\Delta t\,, \end{aligned} \tag{18}$$

on the structural form matrices lead to reasonable "trapezoid" approximations $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ (method III in Gard, 1988, p. 192) of $\mathbf{A}$ and $\mathbf{B}$. These compare favorably with the extremely rough "rectangle" approximations $\tilde{\tilde{\mathbf{A}}}$ and $\tilde{\tilde{\mathbf{B}}}$ in $\mathbf{A}_\vartriangle = \mathbf{I} + \tilde{\tilde{\mathbf{A}}}\Delta t$ and $\mathbf{B}_\vartriangle = \tilde{\tilde{\mathbf{B}}}\Delta t$. This can easily be seen by putting both the exact nonlinear matrix exponential form of $\mathbf{A}_\vartriangle$ and the approximate linear constraint forms in power series expansion:

$$\begin{aligned} \mathbf{A}_\vartriangle &= e^{\mathbf{A}\Delta t} \\ &= \mathbf{I} + \mathbf{A}\Delta t + \tfrac{1}{2}\mathbf{A}^2\Delta t^2 + \tfrac{1}{6}\mathbf{A}^3\Delta t^3 + \tfrac{1}{24}\mathbf{A}^4\Delta t^4 + \ldots \quad \text{(exact)} \\ \mathbf{A}_\vartriangle &= (\mathbf{I} - \tfrac{1}{2}\,\tilde{\mathbf{A}}\Delta t\,)^{-1}(\mathbf{I} + \tfrac{1}{2}\tilde{\mathbf{A}}\Delta t\,) \\ &= \mathbf{I} + \tilde{\mathbf{A}}\Delta t + \tfrac{1}{2}\tilde{\mathbf{A}}^2\Delta t^2 + \tfrac{1}{4}\tilde{\mathbf{A}}^3\Delta t^3 + \tfrac{1}{8}\tilde{\mathbf{A}}^4\Delta t^4 + \ldots \quad \text{(trapezoid)} \\ \mathbf{A}_\vartriangle &= \mathbf{I} + \tilde{\tilde{\mathbf{A}}}\Delta t \quad \text{(rectangle)}\,. \end{aligned} \tag{19}$$

While the rectangle approximation truncates the exact series, the weights in the trapezoid approximation are seen to go down only less quickly than in the exact series. In the approximation of the homogeneous part $\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t)$ of the differential equation the rectangle approximation uses only the value at the starting point of the interval: $\frac{\Delta \mathbf{x}_t}{\Delta t} = \tilde{\tilde{\mathbf{A}}}\mathbf{x}_{t-\Delta t}$ or $\mathbf{x}_t = (\mathbf{I} + \tilde{\tilde{\mathbf{A}}}\Delta t)\mathbf{x}_{t-\Delta t}$, while the trapezoid approximation averages the values at the starting and end point of the interval: $\frac{\Delta \mathbf{x}_t}{\Delta t} = \tilde{\mathbf{A}}(\mathbf{x}_t + \mathbf{x}_{t-\Delta t})/2$ or $\mathbf{x}_t = \mathbf{A}_0\mathbf{x}_t + \mathbf{A}_\ell\mathbf{x}_{t-\Delta t}$ (see $\mathbf{A}_0$ and $\mathbf{A}_\ell$ in Equation (18)).

The structural model in Equation (15) with the linear constraints given in Equation (18) will be called Approximate Discrete Model (ADM). Although SEM programs like Mx do not encounter any difficulty in implementing the exact nonlinear constraints of the EDM, in the less nonlinearly oriented SEM programs the ADM might be useful. The simple linear constraints in Equation (18) are easily implemented by specifying the corresponding off-diagonal elements in instantaneous $\mathbf{A}_0$ and lagged $\mathbf{A}_\ell$ to be equal, and each diagonal element in $\mathbf{A}_\ell$ to be 1 plus the corresponding diagonal element in $\mathbf{A}_0$. Multiplying estimated $\mathbf{A}_0$ by $\frac{2}{\Delta t}$ yields estimated (approximate) $\mathbf{A}$, multiplying estimated $\mathbf{B}_\ell$ by $\frac{1}{\Delta t}$ yields estimated (approximate) $\mathbf{B}$ (see Equation (18)). Interestingly, this is one of the few cases in SEM that the self-loop coefficients in the endogenous structural form matrix are estimated instead of specified to be identically zero. For this reason, the ADM can be expected to yield much better approximations of continuous time matrices $\mathbf{A}$ and $\mathbf{B}$ than the cross-sectional ACM in the previous section, where the 0 self-loop coefficients forced the diagonals of $\mathbf{A}$ to be -1.

The Cross-Lagged Panel Design Reconsidered in Continuous Time

Apart perhaps from some special cases like true automata, real life processes typically evolve in continuous time and are not restricted to the discrete observation time points the researcher happens to choose. Therefore, in most cases, discrete time models applied to real life processes are used as approximations or should be considered as such. This need not be a problem, as long as the discrete time interval Δt approximating the infinitesimal dt is small. However, in behavioral science with, for example, Verbal Ability and Quantitative Ability influencing themselves and each other continuously over the school year but with measurements taken not more often than one or two times a year, Δt is typically taken quite large. As a result discrete time modeling becomes a simplification and in many cases a distortion of reality.

This is made clear in the path diagrams of the cross-lagged panel design in Figure 1. The simplification of a true discrete time model consists in the assumption that the arrows jump from one point in time to the next one and that between measurements nothing exists or can happen. Instead, however, the estimated cross-lagged coefficients over the observation interval Δt are complicated mixtures of the continuous time cross- and auto-effects in a constant interchange over and heavily dependent on the length of the observation interval Δt chosen. It is clear, for example, that the cross-lagged coefficient for the effect of Verbal Ability on Quantitative Ability is in fact also dependent on the auto-effect of Quantitative Ability with a larger auto-effect resulting in a larger estimated cross-lagged coefficient, while additionally, over the larger time interval $\Delta t = 1.25$ in diagram B, the result will be more heavily dependent on the auto-effect than over the shorter interval $\Delta t = 0.75$. Continuous time modeling is necessary to disentangle the continuous time cross- and auto-effects from the discrete time mixtures.

It is a paradoxical consequence of discrete time modeling that different researchers, studying the same causal effect but with different discrete time distances, are unable to even compare the strength of the causal effects they found. Continuous time modeling is needed to provide the common base for an accurate comparison of differently spaced models of the same real life

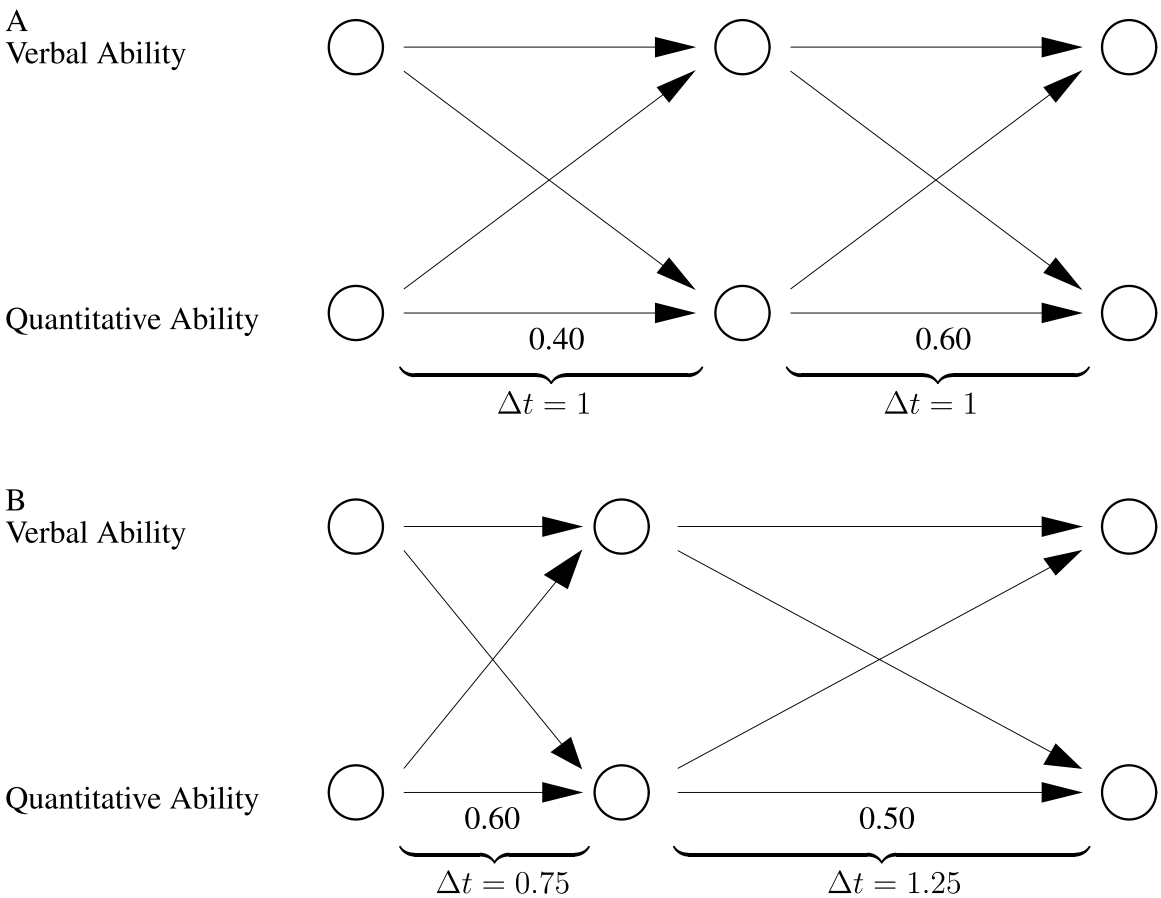

Figure 1: Two cross-lagged panel designs with different spacings of the measurement time points and different values of the autoregression coefficients in the ability variables

process. This also is exemplified in Figure 1. Supposing the other effects in the model to be equal over the two successive intervals in diagram A, one would not need continuous time modeling to conclude on the basis of the autoregressions (0.40 and 0.60) that the auto-effect over the first interval is smaller than over the second interval. In diagram B, however, because it is known that autoregressions depend on the time interval (in general, the smaller the interval, the larger the autoregression, reaching 1 for $\Delta t = 0$), it cannot be decided on the basis of the respective discrete time models in the two different time units, which one implies a bigger auto-effect: 0.60 over interval $\Delta t = 0.75$ or 0.50 over interval $\Delta t = 1.25$. Using the matrix exponential $\mathbf{A}_{\Delta} = e^{\mathbf{A}\Delta t}$ and for convenience assuming the cross-effects to be zero, it is easily calculated that the higher autoregression value of 0.60 over $\Delta t = 0.75$ in $\mathbf{A}_{\Delta}$ implies a lower auto-effect in drift matrix $\mathbf{A}$ than the autoregression of 0.50 over $\Delta t = 1.25$, respectively -0.68 and -0.55 in $\mathbf{A}$. It should be noted that the exponential function is such that discrete time autoregression values ranging from 1 to 0 correspond to continuous time drift coefficients from 0 to $-\infty$. Using the matrix exponential in two directions to transform 0.60 and 0.50 to the same time interval $\Delta t = 1.00$ would lead to autoregression values of 0.51 and 0.57, again indicating that 0.60 over $\Delta t = 0.75$

		x_1	x_2	x_3	x_1	x_2	x_3
pair I	x_1	0.50	0.30	0.21	−0.93	0.68	0.43
	x_2	0.20	0.40	0.20	0.39	−1.22	0.59
	x_3	0.20	0.20	0.30	0.50	0.50	−1.52
pair II	x_1	0.50	0.30	0.21	−0.94	0.69	0.43
	x_2	0.10	0.40	0.20	−0.11	−1.02	0.70
	x_3	0.40	0.20	0.30	1.30	0.19	−1.70
			$\mathbf{A}_\vartriangle$			$\mathbf{A}$	

Table 1: Two pairs of a discrete time autoregression matrix $\mathbf{A}_\vartriangle$ and corresponding continuous time drift matrix $\mathbf{A}$

represents in fact a smaller auto-effect than 0.50 over $\Delta t = 1.25$.

Not less paradoxical differences between the discrete time models studied in behavioral science on the one hand and the underlying continuous time models on the other hand can be observed in Table 1. On the basis of simple simulated autoregression matrices $\mathbf{A}_\vartriangle$ (both typical in the sense of having higher diagonal than nondiagonal elements), it is shown that the conclusions drawn with respect to the cross-lagged coefficients in $\mathbf{A}_\vartriangle$ may differ quite fundamentally from those drawn on the basis of the corresponding cross-effects in drift matrices $\mathbf{A}$. In addition to the autoregression matrices $\mathbf{A}_\vartriangle$ also the corresponding drift matrices $\mathbf{A}$, determined according to the exponential form $\mathbf{A}_\vartriangle = e^{\mathbf{A}}$ with time interval $\Delta t = 1$, are shown.

- *Equal discrete time coefficients may become different in continuous time.*

For example, the two reciprocal cross-lagged coefficients with value 0.20 in the first autoregression matrix, which could lead to the conclusion that the strength of the causal effects between the variables x_2 and x_3 is equal in opposite directions, are different in continuous time: 0.59 and 0.50.

- *The order of magnitude of coefficients may reverse going from discrete to continuous time.*

For example, in the first autoregression matrix, the discrete time effect from x_3 on x_1 is greater than in the opposite direction from x_1 on x_3: 0.21 versus 0.20. However, in the corresponding drift matrix it is the other way around: 0.43 for the first effect and 0.50 for the second effect.

- *Discrete time nonzero coefficients may vanish or even change sign in continuous time.*

An example is shown by pair II. The positive effect from x_1 on x_3 of 0.10 in discrete time gets the negative value of -0.11 in continuous time.

All these paradoxical differences in causal inference on the basis of corresponding discrete time and continuous time effects are easily explained. In contrast to the continuous time cross-effects, the discrete time cross-lagged coefficients are more or less complicated mixtures of continuous time cross-effects and auto-effects. A variable with a high auto-effect, meaning that there is a strong tendency to keep its value over time, tends also to retain the influence from other variables over a longer time interval than a variable with a low auto-effect. So, even a relatively small continuous time cross-effect can result in a relatively high cross-lagged effect in discrete time, if the variable influenced has a high auto-effect. But the converse can be true too: a relatively strong continuous time cross-effect nevertheless corresponding to little impact over a discrete time interval because of a quite low auto-effect in the dependent variable.

All this clearly depends on the time interval. Therefore, statements about direction and strength of a causal effect in discrete time are meaningless without indicating the exact time interval Δt the statement refers to. This is the clear message of Figures 2 and 3, where for several of the cross-lagged coefficients in Table 1 not only the value at $\Delta t = 1$ but the development over the whole period from $\Delta t = 0$ until $\Delta t = 2$ years according to exponential form $\mathbf{A}_{\Delta} = e^{\mathbf{A}\Delta t}$ is given. Figures 2 and 3 give the continuous time impulse-response, that is the effects of an isolated unit-impulse in a single independent variable over continuously increasing time intervals on the dependent variable.

The implication of Figure 2 is that the conclusion about the relative strength of the reciprocal causal effects between x_3 and x_1 (pair I of Table 1) on the basis of the discrete time model depends on the time interval chosen in the model. Researchers choosing the discrete time interval Δt between 0 and 0.66 year will come to the conclusion that x_1 has a larger effect on x_3 (maximum difference reached at $\Delta t = 0.27$), while researchers choosing $\Delta t > 0.66$ come to the opposite conclusion (maximum difference reached at $\Delta t = 2.74$). The cross-lagged coeffi-

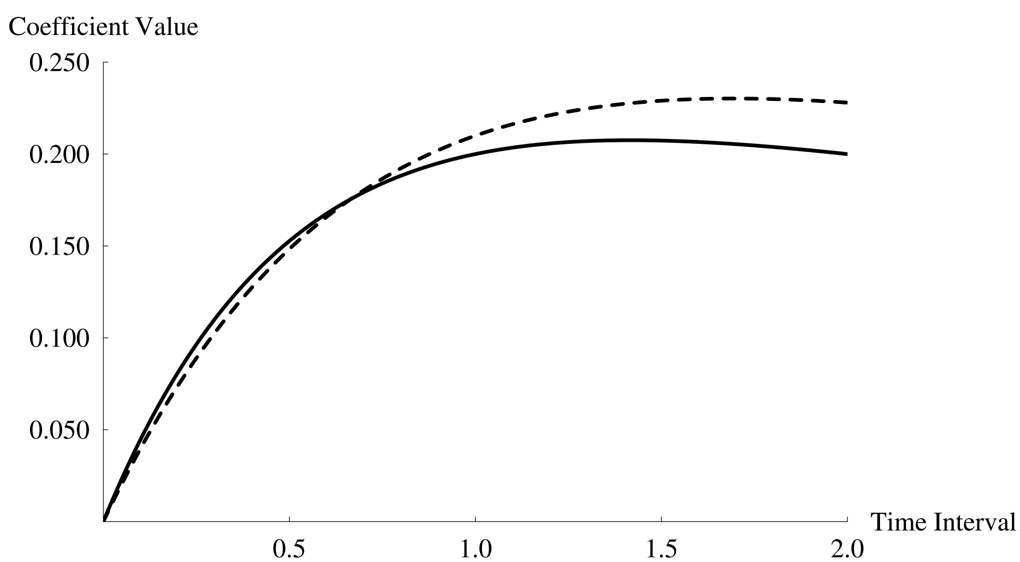

Figure 2: Cross-lagged coefficients $a_{\Delta,31}$ (solid line) and $a_{\Delta,13}$ (dotted line) in autoregression matrix $\mathbf{A}_{\Delta}$ of pair I in Table 1 for corresponding continuous time coefficients $a_{31} = 0.50$ and $a_{13} = 0.43$ in $\mathbf{A}$ as functions of the time interval $\Delta t \in [0, 2]$

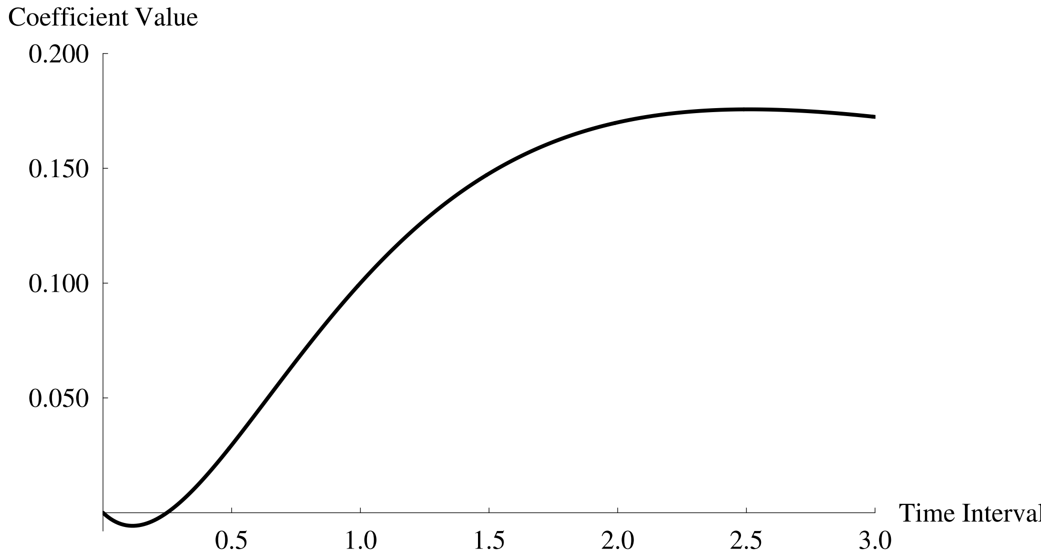

Figure 3: Cross-lagged coefficient $a_{\Delta,21}$ in autoregression matrix $\mathbf{A}_\Delta$ of pair II in Table 1 for corresponding continuous time coefficient $a_{21} = -0.11$ in $\mathbf{A}$ as a function of the time interval $\Delta t \in [0, 2]$

cients, by definition having value 0 over $\Delta t = 0$, will first once or repeatedly go up or down but eventually go to 0 for this and other asymptotically stable models (all eigenvalues of $\mathbf{A}$ strictly negative). Such stable models also imply a maximum value for the cross-lagged effect to be reached after some finite time interval Δt. This time interval is different for different variables and different models. For $a_{\Delta,31}$ the maximum values of 0.208 is reached at $\Delta t = 1.42$, for $a_{\Delta,13}$ it is 0.230, reached at $\Delta t = 1.70$.

Figure 3 describes the discrete time effect from x_1 on x_2 (pair II of Table 1) in models with different Δt. Its clear implication is that even the sign of the cross-lagged coefficient need not be the same as the one of the true underlying continuous time effect. Although the true underlying continuous time effect is negative, only researchers using discrete time models with Δt from 0 until 0.25 year will conclude to a negative effect. For all other time intervals $\Delta t > 0.25$ the interplay between auto-effects and cross-effects leads to a positive effect in discrete time and would therefore lead to an erroneous conclusion about the sign of the true underlying cross-effect in continuous time. Because also the model in pair II is asymptotically stable, there is a minimum and maximum value for the cross-lagged coefficient (-0.006 and 0.176), reached by researchers using observation interval $\Delta t = 0.11$ and $\Delta t = 2.51$ years, respectively. Figure 3 clearly gives valuable information in addition to the continuous time coefficients. The interplay of the three variables is apparently such that the long term effect of x_1 on x_2 may not be interpreted as a negative one, in spite of an undisputable negative direct effect of x_1 on x_2 and obviously due indirectly to the positive intervening role of x_3 which eventually is dominating.

Continuous Time Modeling of the Cross-Lagged Panel Design: An Example

We will now give analysis results of a cross-lagged panel design where the dynamic differential model parameters were estimated in different ways by means of SEM programs Mx and LISREL. The data were taken from a more comprehensive study of school achievement in the Netherlands, conducted with several cohorts over the period 1991-1997 (Aarnoutse, van Leeuwe, Voeten, van Kan, & Oud, 1996). The data consisted of two decoding tests (One-Minute-Test Form A and B) and different pairs out of five reading comprehension tests (Cito Reading Comprehension Tests 2, M3 and M4, Aarnoutse Reading Comprehension Tests 4 and 5) administered at four time points to 638 Dutch primary school pupils during first and second grade. Pairs of tests were used to construct latent variables corrected for measurement error in the observed variables. The purpose was to examine in continuous time the question that has occupied reading investigators for a long time in discrete time: whether and how strongly latent Decoding Skill influences latent Reading Comprehension, latent Reading Comprehension influences latent Decoding Skill, or both latent variables influence each other reciprocally.

Model specification

While different models were specified, all had the same measurement equation part, which will be addressed first:

$$\mathbf{y}_{t_i} = \mathbf{C}_{t_i}\mathbf{x}_{t_i} + \mathbf{d}_{t_i} + \mathbf{v}_{t_i} \quad \text{with} \quad \text{cov}(\mathbf{v}_{t_i}) = \mathbf{R}_{t_i} \ . \tag{20}$$

The parameter matrices for successive observation time points t_0, t_1, t_2, t_3 are shown in Table 2. By fixing the factor loading λ of the One-Minute-Test Form A at value 1 and its measurement origin ν at value 0 (in the first row of $\mathbf{C}_{t_0}$ and $\mathbf{d}_{t_0}$ in Table 2) we equalled the variance and mean of the latent Decoding Skill (DS) variable at the initial time point t_0 to the true variance (total variance minus measurement error variance) and mean of the One-Minute-Test Form A at that time point. In the same manner, the true variance and mean of Cito Reading Comprehension Test 2 defined the variance and mean of the latent Reading Comprehension (RC) variable at t_0. Connecting the latent scales in this way to familiar test scales made the former more easily interpretable by teachers.

For DS we used the same pair of tests (One-Minute-Test Form A and B) at all four time points. By constraining the λs and νs of each of both decoding speed tests to be equal over the whole time period, the measurement unit and origin of the latent DS-variable also were kept constant and latent DS-growth became meaningfully interpretable. Because no reading comprehension tests were available that could be used over the entire time period, for RC we had to proceed differently to guarantee constant latent measurement characteristics. Here a pattern was chosen with pair of tests A and B at time point t_0 followed by pair of tests B and C at t_1, C and D at t_2 and finally D and E at t_3. So, going from one time point to the next, one test was always used twice and by constraining its λs and νs to be equal, indirectly the latent measurement unit and origin also were kept constant over successive pairs of time points. The pairs of tests at successive time

Measurement equation

$$\mathbf{C}_{t_0} = \begin{bmatrix} 1 & 0 \\ \lambda_1 & 0 \\ 0 & 1 \\ 0 & \lambda_2 \end{bmatrix} \mathbf{d}_{t_0} = \begin{bmatrix} 0 \\ \nu_1 \\ 0 \\ \nu_2 \end{bmatrix} \mathbf{R}_{t_0} = \begin{bmatrix} \theta_1 & & & \\ 0 & \theta_2 & & \\ 0 & 0 & \theta_3 & \\ 0 & 0 & 0 & \theta_4 \end{bmatrix}$$

$$\mathbf{C}_{t_1} = \begin{bmatrix} 1 & 0 \\ \lambda_1 & 0 \\ 0 & \lambda_3 \\ 0 & \lambda_2 \end{bmatrix} \mathbf{d}_{t_1} = \begin{bmatrix} 0 \\ \nu_1 \\ \nu_3 \\ \nu_2 \end{bmatrix} \mathbf{R}_{t_1} = \begin{bmatrix} \theta_1 & & & \\ 0 & \theta_2 & & \\ 0 & 0 & \theta_5 & \\ 0 & 0 & 0 & \theta_4 \end{bmatrix}$$

$$\mathbf{C}_{t_2} = \begin{bmatrix} 1 & 0 \\ \lambda_1 & 0 \\ 0 & \lambda_3 \\ 0 & \lambda_4 \end{bmatrix} \mathbf{d}_{t_2} = \begin{bmatrix} 0 \\ \nu_1 \\ \nu_3 \\ \nu_4 \end{bmatrix} \mathbf{R}_{t_2} = \begin{bmatrix} \theta_1 & & & \\ 0 & \theta_2 & & \\ 0 & 0 & \theta_5 & \\ 0 & 0 & 0 & \theta_6 \end{bmatrix}$$

$$\mathbf{C}_{t_3} = \begin{bmatrix} 1 & 0 \\ \lambda_1 & 0 \\ 0 & \lambda_5 \\ 0 & \lambda_4 \end{bmatrix} \mathbf{d}_{t_3} = \begin{bmatrix} 0 \\ \nu_1 \\ \nu_5 \\ \nu_4 \end{bmatrix} \mathbf{R}_{t_3} = \begin{bmatrix} \theta_1 & & & \\ 0 & \theta_2 & & \\ 0 & 0 & \theta_7 & \\ 0 & 0 & 0 & \theta_6 \end{bmatrix}$$

State equation

$$\mathbf{A}_{\Delta} = \begin{bmatrix} a_{\Delta 11} & a_{\Delta 12} \\ a_{\Delta 21} & a_{\Delta 22} \end{bmatrix} \mathbf{b}_{\Delta} = \begin{bmatrix} b_{\Delta 1} \\ b_{\Delta 2} \end{bmatrix} \mathbf{Q}_{\Delta} = \begin{bmatrix} q_{\Delta 11} & \\ q_{\Delta 21} & q_{\Delta 22} \end{bmatrix}$$

$$\boldsymbol{\mu}_{t_0} = \begin{bmatrix} E(x_{1,t_0}) \\ E(x_{2,t_0}) \\ 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\Phi}_{t_0} = \begin{bmatrix} \phi_{x_{1,t_0}} & & & \\ \phi_{x_{2,t_0},x_{1,t_0}} & \phi_{x_{2,t_0}} & & \\ \phi_{\kappa_1,x_{1,t_0}} & \phi_{\kappa_1,x_{2,t_0}} & \phi_{\kappa_1} & \\ \phi_{\kappa_2,x_{1,t_0}} & \phi_{\kappa_2,x_{2,t_0}} & \phi_{\kappa_2,\kappa_1} & \phi_{\kappa_2} \end{bmatrix}$$

Table 2: Free and fixed elements in the discrete time parameter matrices of the EDM

points turned out to be pairwise congeneric in the sense of Jöreskog (1974, pp. 5-12). Although therefore measuring one and the same latent variable, their measurement unit and origin did not need to be equal. The measurement unit and origin of each newly introduced test (λ_3 and ν_3 at t_1, λ_4 and ν_4 at t_2, λ_5 and ν_5 at t_3) could be estimated, however, and so did not distort the latent measurement characteristics chosen at the initial time point. It should be noted finally, that the measurement variances θ of each same test were constrained to be equal across time, resulting in 7 θs to be estimated. Together with the 5 nonfixed λs and the 5 nonfixed νs there were 17 measurement parameters to be estimated.

The parameter matrices of the discrete time state equation

$$\mathbf{x}_t \;=\; \mathbf{A}_\Delta \mathbf{x}_{t-\Delta t} + \boldsymbol{\kappa} + \mathbf{b}_\Delta + \mathbf{w}_{t-\Delta t} \quad \text{with} \quad \text{cov}(\mathbf{w}_{t-\Delta t}) = \mathbf{Q}_\Delta \tag{21}$$

are also shown in Table 2. As they contain 21 unknown parameters, the total number of parameters to be estimated is 38. There are 16 observed variables or 16 observed means and 136 (distinct) elements in the observed covariance matrix, resulting in 114 degrees of freedom for the SEM model as a whole. The model contains the latent state variables DS and RC or x_1 and x_2 and corresponding constant trait variables κ_1 and κ_2, which because of the number of time points being 4 leads to a total of 10 latent variables in the structural equation model. $\boldsymbol{\mu}_{t_0}$ specifies the initial latent means which by definition are 0 for the trait variables. Trait variables are constant over time but vary over subjects to accommodate for deviations of subject specific developmental curves from the mean curve. $\boldsymbol{\Phi}_{t_0}$ contains the trait variances and covariances ($\boldsymbol{\Phi}_\kappa$), the initial state variances and covariances ($\boldsymbol{\Phi}_{x_{t_0}}$) as well as the covariances between trait and initial states ($\boldsymbol{\Phi}_{\kappa,x_{t_0}}$). While the constants in $\mathbf{b}_\Delta$ contribute to changes in mean development, both the process error covariance matrix $\mathbf{Q}_\Delta$ and the trait covariance matrix $\boldsymbol{\Phi}_\kappa$ feed the state covariance matrix.

The stochastic differential equation

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \boldsymbol{\gamma} + \mathbf{b} + \mathbf{G}\frac{d\mathbf{W}(t)}{dt}\,, \tag{22}$$

describes the development of the latent variables in continuous time, containing in particular continuously contributing traits $\boldsymbol{\gamma}$ and constants $\mathbf{b}$. The EDM relates the continuous time parameter matrices in Equation (22) as follows to the discrete time parameter matrices in Table 2 (Oud & Jansen, 2000):

$$\begin{aligned}
\mathbf{A}_\Delta &= e^{\mathbf{A}\Delta t}\,,\\
\mathbf{b}_\Delta &= \mathbf{A}^{-1}[\mathbf{A}_\Delta - \mathbf{I}]\,\mathbf{b}\,,\\
\mathbf{Q}_\Delta &= \text{irow}[(\mathbf{A}\otimes\mathbf{I} + \mathbf{I}\otimes\mathbf{A})^{-1}(\mathbf{A}_\Delta\otimes\mathbf{A}_\Delta - \mathbf{I}\otimes\mathbf{I})\,\text{row}(\mathbf{G}\mathbf{G}')]\,,\\
\boldsymbol{\Phi}_\kappa &= \mathbf{A}^{-1}[\mathbf{A}_\Delta - \mathbf{I}]\boldsymbol{\Phi}_\gamma\,[\mathbf{A}'_\Delta - \mathbf{I}]\mathbf{A}'^{\,-1}\,,\\
\boldsymbol{\Phi}_{\kappa,x_{t_0}} &= \mathbf{A}^{-1}[\mathbf{A}_\Delta - \mathbf{I}]\boldsymbol{\Phi}_{\gamma,x_{t_0}}\,.
\end{aligned} \tag{23}$$

Here, $\otimes$ is the Kronecker product, "row" is the rowvec operation, putting the elements of a matrix rowwise in a column vector, "irow" the inverse operation. Because the time intervals between the measurements were approximately half a year, we started by fixing Δt for the intervals $t_1 - t_0, t_2 - t_1, t_3 - t_2$ at $\Delta = 0.50$.

If, in addition to time-invariance for all parameters (as in Equation (22)) and equal observation intervals, the number of free discrete-time parameters in the EDM is equal to the number of free continuous-time parameters and a one-to-one relationship exists between the values in the discrete-time and continuous-time parameter space, estimating the EDM by means of the indirect method gives exactly the same results as the direct method. The indirect method consists of first estimating the discrete time parameters in Equation (21) by means of an SEM program as, for example, LISREL and next applying Equation (23) on the estimates to derive the continuous time parameter estimates. The direct method (applying the nonlinear restrictions in Equation (23) during estimation by means of an SEM program as, e.g., Mx), however, is much more general. Except for the basic EDM, the direct method is indeed applicable and will be applied in the sequel for a model with unequal intervals (different Δs leading to a different set of discrete time parameters for each Δ) and also for two models with, additionally, the growth constants in $\mathbf{b}$ replaced by linearly and quadratically time-varying parameters.

Results for EDM, ADM, and ACM: indirect method and approximate models

In the first column of Table 3, estimation results for the basic EDM are given. As expected in view of the equivalence in this case, the indirect method by means of LISREL gave exactly the same results as the direct method by means of Mx. With regard to the four coefficients in drift matrix $\mathbf{A}$, an important difference in interpretability should be noted between the auto-effects a_{11} (DS) and a_{22} (RC) on the one hand and the cross-effects a_{12} (effect from RC to DS) and a_{21} (effect from DS to RC) on the other hand. The auto-effects are scale free in the sense that they do not change under arbitrary linear transformations of DS and RC and so are directly interpretable. In particular, both DS and RC show negative feedback (-1.381 and -1.564), implying stability or a rather strong tendency to converge to the subject specific mean trajectories $E[\mathbf{x}(t)|\boldsymbol{\gamma}]$.

Although their products $a_{12}a_{21}$ and $a_{21}a_{12}$ are scale free, the cross-effects themselves are not and therefore need first standardization to become comparable. The coefficients a_{12} and a_{21} shown in Table 3 have been standardized through multiplication by the ratios of the initial standard deviations: $\sqrt{\phi_{x_{2,t_0}}}/\sqrt{\phi_{x_{1,t_0}}}$ and $\sqrt{\phi_{x_{1,t_0}}}/\sqrt{\phi_{x_{2,t_0}}}$, respectively. The standardized values of 0.735 and 0.690 seem to reveal the existence of a reciprocal effect between DS and RC with the effect from RC to DS ($p < .001$) at least as strong as in the opposite direction from DS to RC ($p < .01$) and with the likelihood ratio test of equality ($a_{12} - a_{21} = 0$) not leading to rejection ($p > .05$).

In addition to the basic EDM, for comparison, corresponding approximate models ADM and ACM were estimated by means of LISREL too. The specification of the discrete time parameters in terms of LISREL is the same in the three models except for those related to $\mathbf{A}$. All three are identified, and because the number of free parameters in the approximate models is the same again as in the EDM with one-to-one correspondence between the values in the parameter spaces, the three models are observationally equivalent and lead to exactly the same fit again. While the advantage of the indirect method is that no additional constraints between discrete and continuous time parameters during estimation are needed at all, in contrast to the EDM with its highly nonlinear constraints the ADM needs only few simple extra linear constraints and the ACM still simpler fixation constraints. While for the EDM the four coefficients in (reduced form) autoregression matrix $\mathbf{A}_\Delta$ were specified and $\mathbf{A}$ as well as the other continuous time parameters were derived according to Equation (23), in the ADM two (structural form) matrices $\mathbf{A}_0$ and $\mathbf{A}_\ell$ containing eight coefficients were specified. However, this was done under four constraints between the instantaneous coefficients in $\mathbf{A}_0$ and the lagged coefficients in $\mathbf{A}_\ell$ (see Equation (18)), leading to the same number of degrees of freedom for the LISREL model as a whole ($df = 114$ in Table 3) as in the EDM but to the simple approximation $\tilde{\mathbf{A}} = 2\Delta^{-1}\mathbf{A}_0$ of $\mathbf{A}$. Also the remaining continuous time parameter approximations for the ADM in Table 3 are easily derived using $\tilde{\mathbf{b}} = \Delta^{-1}\mathbf{b}_\ell$, $\tilde{\boldsymbol{\Phi}}_\gamma = \Delta^{-2}\boldsymbol{\Phi}_\kappa$, $\tilde{\boldsymbol{\Phi}}_{\gamma,x_{t_0}} = \Delta^{-1}\boldsymbol{\Phi}_{\kappa,x_{t_0}}$. ACM got the same specification and han-

parameter	EDM	ADM	ACM
a_{11}	-1.381***	-1.290***	-1.000
a_{12}	0.735***	0.639***	0.369***
a_{21}	0.690**	0.600**	0.317**
a_{22}	-1.564***	-1.450***	-1.000
b_1	37.0***	38.6***	13.7
b_2	36.2***	34.8***	22.2***
$E(x_{1,t_0})$	32.0***	32.0***	32.0***
$E(x_{2,t_0})$	21.6***	21.6***	21.6***
$\phi_{x_{1,t_0}}$	191.3***	191.3***	191.3***
$\phi_{x_{2,t_0},x_{1,t_0}}$	57.0***	57.0***	57.0***
$\phi_{x_{2,t_0}}$	36.4***	36.4***	36.4***
ϕ_{γ_1}	221.0***	193.8***	130.3***
ϕ_{γ_2,γ_1}	-44.4*	-34.1*	-36.2**
ϕ_{γ_2}	29.0***	24.5**	16.9**
χ^2	488.6	488.6	488.6
df	114	114	114

* $p < .05$
** $p < .01$
***$p < .001$

Table 3: Selected parameter estimates and model fit information for the basic EDM and corresponding approximate models ADM and ACM

dling as ADM with again $df = 114$, except for the following modification. As a cross-sectional simulation ACM simply used the instantaneous cross-coefficients in $\mathbf{A}_0$ as approximations of the continuous time cross-coefficients in $\mathbf{A}$, without constraints between $\mathbf{A}_0$ and $\mathbf{A}_\ell$, specifically with the lagged cross-coefficients in $\mathbf{A}_\ell$ fixed at 0, the diagonals of $\mathbf{A}_0$ at 0 and correspondingly those of $\mathbf{A}$ at -1 (see Equation (13) where $\mathbf{A}_0$ is called $\mathbf{A}^*$).

Comparing the approximate models ADM and ACM with the exact model EDM learns that the estimates of the initial parameters in $\boldsymbol{\mu}_{t_0}$ and $\boldsymbol{\Phi}_{t_0}$ are exactly the same as in the EDM, while other parameters, with few exceptions, result in less pronounced estimates (lower or less negative) than in the EDM. Also, as expected, ADM approximates the estimates in the EDM closer than ACM, the latter clearly suffering a lot from the diagonal feedback coefficients in $\mathbf{A}$ forced to be -1.

Results for EDM0, EDM1, and best fitting EDM2: direct method

The poor fit of all three models in Table 3 ($\chi^2 = 488.6$ for only $df = 114$) motivated the examination of the more complicated models in Table 4, estimable only by means of the direct method with a nonlinear constraint SEM program like the program Mx used in this study. EDM0, apart from giving the observation intervals the slightly differing true values ($t_1 - t_0 = 0.50, t_2 - t_1 = 0.45, t_3 - t_2 = 0.55$) being equal to the basic EDM in Table 3, led to an impressive improvement in fit ($\chi^2 = 391.3$ for the same $df = 114$) but nonetheless to almost the same parameter estimates. Also, likelihood ratio testing of the coefficients a_{12} and a_{21} proved their estimates to be significant again at the levels .001 and .01. Because a systematic lack of fit was observed in the observed means for the basic EDM and EDM0 (the fitted means being too low at t_1 and too high at t_2), we tried to obtain further improvements by the models EDM1 and EDM2, having $\mathbf{b}$ in Equation (22) replaced by linearly and quadratically time-varying $\mathbf{b}_t = \mathbf{b}_0 + \mathbf{b}_1(t - t_0)$ and $\mathbf{b}_t = \mathbf{b}_0 + \mathbf{b}_1(t - t_0) + \mathbf{b}_2(t - t_0)^2$, respectively.

Compared with EDM0, most conspicuous differences of the best fitting model EDM2 ($\chi^2 = 274.5$ for $df = 110$ and therefore an impressive improvement again over EDM0) were, besides a lower value of -2.418 for a_{22} and thus a more stable development of RC, a somewhat higher effect from RC to DS ($a_{12} = 0.795; p < .001$) and lower effect in the opposite direction from DS to RC ($a_{21} = 0.418; p = .12$). Unfortunately the results for a_{21} were rather ambiguous, because simultaneously the estimated a_{21} was not significantly different from 0 ($p = .12$ for $\chi^2 = 2.4$ with $df = 1$) and the likelihood ratio test of $a_{12} - a_{21} = 0$ did not reach significance ($p = .24$ for $\chi^2 = 1.4$ with $df = 1$). We conclude that the study confirmed a clear effect from RC to DS and did not disconfirm the strength of the effect in the opposite direction from DS to RC to be similar, although its estimate was considerably lower.

With regard to the dynamic properties of the best fitting model EDM2 it is important to note that both latent state variables show again negative feedback. This implies stability or a rather strong tendency for the state to converge to the subject specific mean trajectory $E[\mathbf{x}(t)|\boldsymbol{\gamma}]$. This latter curve is easily derived from Equation (8) and should be distinguished clearly from the overall mean trajectory $E[\mathbf{x}(t)]$ as well as from the Kalman smoother $E[\mathbf{x}(t)|\mathbf{y}]$ or optimal estimate of the subject's sample trajectory. An example of all three curves for DS is shown in Figure 4, where for γ a subject's estimate of -11.46 was taken and for the model parameters the estimates of EDM2 in Table 4. In a stable model deviations of a subject's sample trajectory from the subject specific mean trajectory are expected to go to zero and the more rapidly, the more negative the feedback. While in space technology it is of utmost importance to be able to know and control just the sample trajectory, in education the subject specific mean trajectory is perhaps as important as its sample trajectory. A big advantage of the state-trait model is that it distinguishes trait variance (unobserved heterogeneity between subjects) clearly from stability. Because in a pure state model ($\boldsymbol{\gamma} = \mathbf{0}$) all subject specific mean trajectories coincide with the overall mean trajectory, trait variance and stability are confounded in the sense that an actually nonzero trait variance would spuriously lead to a less stable model. The bigger the actual trait variance, the less negative the feedback coefficients become in a pure state model in order to keep the sample trajectories at their subject specific distances from the overall mean. In a state-trait model trait variance and stability are unrelated in this way. For example, Table 4 shows that RC

parameter	EDM0	EDM1	EDM2
a_{11}	-1.376	-1.787	-1.468
a_{12}	0.702	0.503	0.795
a_{21}	0.621	0.133	0.418
a_{22}	-1.626	-2.872	-2.418
b_{01}	39.3	83.7	43.7
b_{02}	39.3	75.5	66.5
b_{11}		10.9	-26.3
b_{12}		10.8	-15.2
b_{21}			18.9
b_{22}			12.8
$E(x_{1,t_0})$	32.0	32.0	32.0
$E(x_{2,t_0})$	21.6	21.6	21.6
$\phi_{x_{1,t_0}}$	191.4	192.2	191.6
$\phi_{x_{2,t_0},x_{1,t_0}}$	57.0	57.3	57.3
$\phi_{x_{2,t_0}}$	36.4	36.6	36.6
ϕ_{γ_1}	225.9	440.9	257.8
ϕ_{γ_2,γ_1}	-35.9	77.0	-9.0
ϕ_{γ_2}	29.0	125.2	71.8
χ^2	391.3	351.4	274.5
df	114	112	110

Table 4: Selected parameter estimates and model fit information for EDM0, having unequal observation intervals, and EDM1 and EDM2 having, in addition, linearly and quadratically increasing mean growth parameters

behaves more stable than DS (the estimated values of a_{22} is more negative than of a_{11}) but at the same time the RC trait variance relative to its state variance ($\phi_{\gamma_2}/\phi_{x_{2,t_0}} = 1.96$) is greater than for DS ($\phi_{\gamma_1}/\phi_{x_{1,t_0}} = 1.35$).

For the interpretation of parameter values it is important to note that some are and some are not dependent on the choice of the time scale. For example, multiplying time by factor c does not influence the state variances $\phi_{x_{1,t}}$, $\phi_{x_{2,t}}$, but it multiplies a_{11}, a_{22} by $1/c$ and the trait variances $\phi_{\gamma_1}, \phi_{\gamma_2}$ by $1/c^2$. For this reason it would not make sense to compare trait variances for different time scales or trait variances with state variances, although it makes sense to compare ratios of trait and state variance for the same time scale.

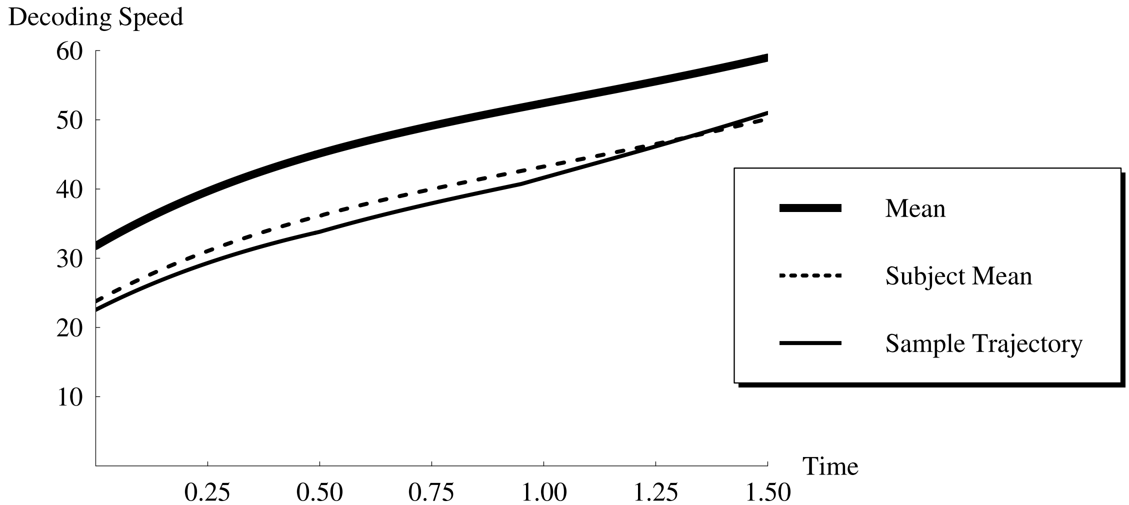

Figure 4: Mean trajectory $E[\mathbf{x}(t)]$, subject specific mean trajectory $E[\mathbf{x}(t)|\boldsymbol{\gamma}]$, and Kalman smoother $E[\mathbf{x}(t)|\mathbf{y}]$

Conclusions

The educational research example discussed illustrates and confirms the main arguments made earlier in the article. Especially, the reciprocal influence between Decoding Skill and Reading Comprehension typically is a process that unfolds in continuous time and needs continuous time modeling. Accounting for the true small differences in continuous time intervals between the measurement time points turned out to lead to a dramatic improvement in model fit without any additional loss of degrees of freedom (cfr. the basic EDM in Table 3 and the EDM0 in Table 4).

The usefulness of the indirect method in stochastic differential equation model estimation has been criticized in the literature, especially by Singer (1990, 1992, 1998, 1999). Although for a special class of models the indirect and direct method give the same result (e.g., for the basic EDM in Table 3), the direct method is much more generally applicable and should therefore be preferred. Examples of simple models not accessible for the indirect method are the EDM0, EDM1 and EDM2 displayed in Table 4. The indirect method also necessarily fails, when input variables, instead of by step functions, are described or approximated by piecewise linear functions between measurements or by more complicated higher-order polynomial functions (Singer, 1992). Not surprisingly, linear and higher-order polynomial input functions have almost the same effect on the EDM as linearly and higher-order polynomially time-varying parameters in input-effects matrix $\mathbf{B}(t)$ (Oud & Jansen, 2000). In all these and many more cases the direct method is the only feasible one and applicable also by means of SEM as explained in Oud and Jansen (2000).

The reason for using the approximate models ADM and ACM is, apart from the lack of longitudinal data in the case of ACM, the possibility to use simple linear constraints in estimating the stochastic differential equation model. In this respect, however, the ADM and the ACM suffer from the same limitations as the EDM estimated by the indirect method, which in the first step also avoids nonlinear constraints. An extra drawback of the ADM and ACM is the approximate character, which for the ACM in the example led to quite unacceptable results in comparison to the exact results of the EDM.

References

Aarnoutse, C.A.J., van Leeuwe, J.F.J., Voeten, M.J.M., van Kan, P.L.M. & Oud, J.H.L. (1996). *Longitudinaal onderzoek schoolvorderingen in het basisonderwijs* [Longitudinal study of school achievement in primary education]. Nijmegen, The Netherlands: Vakgroep Onderwijskunde Nijmegen.

Arminger, G. (1986). Linear stochastic differential equations for panel data with unobserved variables. In N.B. Tuma (Ed.), *Sociological Methodology* (pp. 187-212). Washington: Jossey-Bass.

Arminger, G., Wittenberg, J., & Schepers, A. (1996). *MECOSA 3: Mean and covariance structure analysis*. Friedrichsdorf, Germany: Additive.

Arnold, L. (1974). *Stochastic differential equations*. New York: Wiley.

Bergstrom, A.R. (1983). Gaussian estimation of structural parameters in higher order continuous time dynamic models. *Econometrica, 51*, 117-152.

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

Bergstrom, A.R. (1988). The history of continuous-time econometric models. *Econometric Theory, 4*, 365-383.

Bijleveld, C.C.J.H. & van der Kamp, L.J.Th. (1998). *Longitudinal data analysis: designs, models and methods.* London: Sage.

Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models.* Sage: Newbury Park.

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

Collins, L.M. & Horn, J.L. (Eds.) (1991). *Best methods for the analysis of change.* Washington DC: American Psychological Association.

Gandolfo, G. (1993). Continuous-time econometrics has come of age. In G. Gandolfo (Ed.), *Continuous time econometrics* (pp. 1-11). London: Chapman & Hall.

Gard, T.C. (1988). *Introduction to stochastic differential equations.* New York: Marcel Dekker.

Hamerle, A., Nagl, W., & Singer, H. (1991). Problems with the estimation of stochastic differential equations using structural equations models. *Journal of Mathematical Sociology, 16*, 201-220.

Itô, K. (1944). Stochastic integral. *Proceedings of Imperial Academy of Tokyo, 20,* 519-524.

Itô, K. (1951). *On stochastic differential equations* (American Mathematical Society Memoirs, No. 4). New York: American Mathematical Society.

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

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: Freeman.

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

Jöreskog, K.G., & Sörbom, D. (1985). Simultaneous analysis of longitudinal data from several cohorts. In W.M. Mason & S.E. Fienberg (Eds.), *Cohort analysis in social research: Beyond the identification problem* (pp. 323-341). New York: Springer.

Jöreskog, K.G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.

Lazarsfeld, P.F., (1940). "Panel" studies. *Public Opinion Quarterly, 4*, 122-128.

Lazarsfeld P.F., & Fiske, M. (1938). The panel as a new tool for measuring opinion. *Public Opinion Quarterly, 2,* 596-612.

Lorenz, F.O., Conger, R.D., Simons, R.L., & Whitbeck, L.B. (1995). The effects of unequal variances and covariances in simultaneous inference: The case of hostility and marital quality. *Journal of Marriage and the Family, 574*, 1049-1064.

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

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

Mehta, P.D., & West, St.G. (2000). Putting the individual back into individual growth curves. *Psychological methods, 5,* 23-43.

Meissner, W. (1997) *Ökonometrische Modelle: Rekursivität und Interdependenz aus der Sicht der Kybernetik.* Berlin: Duncker und Humblot.

Miyazaki, Y., & Raudenbusch, W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological methods, 5,* 44-63.

Neale, M.C. (1997) *Mx: Statistical Modeling* (4th ed.). Richmond VA: Department of Psychiatry.

Oud, J.H.L., & Jansen, R.A.R.G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika, 65*, 199-215.

Phillips, P.C.B. (1993). The ET Interview: A.R. Bergstrom. In P.C.B. Phillips (Ed.), *Models, methods, and applications of econometrics* (pp. 12-31). Cambridge MA: Blackwell.

Rueter, M.A., & Conger, R.D. (1998). Reciprocal influences between parenting and adolescent problem-solving behavior. *Developmental Psychology, 34*, 1470-1482.

Ruymgaart, P.A. & Soong, T.T. (1985). *Mathematics of Kalman-Bucy filtering.* Berlin: Springer.

Sher, K.J., Wood, M.D., Wood, P.K., & Raskin, G. (1996). Alcohol outcome expectancies and alcohol use: A latent variable cross-lagged panel study. *Journal of Abnormal Child Psychology, 105*, 561-574.

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

Singer, H. (1992). The aliasing-phenomenon in visual terms. *Journal of Mathematical Sociology, 17*, 39-49.

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

Singer, H. (1999). *Finanzmarktökonometrie.* Heidelberg: Physica.

Vuchinich, S., Bank, L., & Patterson, G.R. (1992). Parenting, peers, and the stability of antisocial behavior in preadolescent boys. *Developmental Psychology, 38*, 510-521.

Wiener, J. (1923). Differential space. *Journal of Mathematical Physics, 38*, 131-174.

Wilder, J. (1999). *Stimulus and response: The law of initial value.* Bristol: Wright.

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

Willett, J.B. & Sayer, A.G. (1996). Cross-domain analysis of change over time: Combining growth modeling and covariance structure analysis. In G.A. Marcoulides & R.E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 125-157). Mahwah NJ: Lawrence Erlbaum.

Wold, H. (1954). Causality and econometrics. *Econometrica, 22,* 162-177.