Chapter 7
Continuous Time Modeling of Panel Data by means of SEM

Johan H.L. Oud and Marc J.M.H. Delsing

Abstract After a brief history of continuous time modeling and its implementation in panel analysis by means of structural equation modeling (SEM), the problems of discrete time modeling are discussed in detail. This is done by means of the popular cross-lagged panel design. Next, the exact discrete model (EDM) is introduced, which accounts for the exact nonlinear relationship between the underlying continuous time model and the resulting discrete time model for data analysis. In addition, a linear approximation of the EDM is discussed: the approximate discrete model (ADM). It is recommended to apply the ADM-SEM procedure by means of a SEM program such as LISREL in the model building phase and the EDM-SEM procedure by means of Mx in the final model estimation phase. Both procedures are illustrated in detail by two empirical examples: Externalizing and Internalizing Problem Behavior in children; Individualism, Nationalism and Ethnocentrism in the Flemish electorate.

7.1 Introduction

Continuous time modeling goes back to Newton (1643-1727) and Leibniz (1646-1716), who originated the tools of differential and integral calculus. Newton's laws of motion relate the position, speed, and acceleration of physical bodies by means of differential equations. Not less than two and a half centuries later, Simon (1952) introduced the use of differential equations into social science, followed by Coleman (1968) and Blalock (1969). Blalock illustrates his discussion by means of two

Johan H.L. Oud
Behavioural Science Institute, Radboud University Nijmegen, The Netherlands
e-mail: j.oud@pwo.ru.nl

Marc J.M.H. Delsing
Praktikon, Radboud University Nijmegen, The Netherlands
e-mail: M.Delsing@acsw.ru.nl

K. van Montfort et al., *Longitudinal Research with Latent Variables*,
DOI 10.1007/978-3-642-11760-2_7,

examples. In the first example, a system of two simple linear differential equations is used to describe and explain the arms race process between opposing nations. The second example was taken from Simon (1952) and formalizes Homans's theory about the human group (Homans, 1950), relating the variables interaction, friendship, and activity by means of differential equations. It should be noted that the applications in Newtonian mechanics as well as the examples provided by Simon and Blalock are deterministic and do not allow random error to enter the system.

In contrast to its popularity in physics and natural science, the use of continuous time methods in statistically orientated sciences such as economics and social science is still rare. Undoubtedly, one reason for the slow spread has been the difficulty of handling random phenomena in continuous time, in particular the definition of the random walk process on a continuous time scale as well as the associated stochastic integral. It took a century after the discovery of Brownian motion, the random walk behavior of particles in a liquid, before Norbert Wiener in 1928 succeeded to give this motion a rigorous mathematical definition. In honor of Wiener, the motion was later called Wiener process. Wiener was also the first to define integration of the Wiener process (Wiener stochastic integral), which in 1944 was generalized by the Japanese mathematician Itô (Itô stochastic integral). Nowadays, there is no reason to avoid the specification of random error in continuous time or the use of stochastic differential equations and their solution. The mathematical problems are solved and need not concern the research practitioner as will be shown in this chapter.

Just as in natural science, most phenomena studied in economics and social science evolve in continuous time. As emphasized by Bergstrom, the pioneer of continuous time modeling in econometrics, the economy does not cease to exist in between observations nor does it function only at quarterly or annual intervals corresponding to the observations (Phillips, 1993, p. 23). Bergstrom (1988) credited the British statistician Bartlett for being the first to deal with the problem of estimating the parameters of continuous time stochastic models from discrete time series. As Bartlett (1946) put it:

> The discrete time nature of our observations in many economic and other time series does not reflect any lack of continuity in the underlying series. Thus theoretically it should often prove more fundamental to eliminate this imposed artificiality. An unemployment index does not cease to exist between readings, nor does Yule's pendulum cease to swing.

Hereby, Bartlett for the first time criticized the unfortunate identification in conventional time series analysis of the dynamically relevant interval with the observation interval. Continuous time methods put the causal mechanisms on a continuous time scale, allowing the process to proceed in infinitesimally small steps, and so distinguish the underlying dynamics clearly from the discrete time measurement time points. This is especially important in social science, where measurement almost invariably occurs in discrete time, measurement time points are chosen rather arbitrarily, and observation intervals are often large. Particularly in the case of large intervals, approximating the continuous time process by a discrete time model formulated in terms of the observation interval leads to unacceptable results.

Traditionally, the application of continuous time methods is restricted to $N = 1$ research and estimation in the stochastic case is done by $N = 1$ time series estimation methods, especially filter techniques. To solve the problem of the data points in a time series being correlated, which violates the independence assumption of sampling theory, filter techniques purge the data from the predictable correlated parts to end up with uncorrelated “innovations”. From 1990 onwards, Singer (1990, 1993, 1995, 1998) worked on the adaptation of these techniques for continuous time analysis of panel data. Singer’s (1991) program LSDE (Linear Stochastic Differential Equations) performs maximum likelihood estimation of the continuous time model on the basis of the so-called exact discrete model (EDM). The EDM, claimed to be developed in 1961-1962 by Bergstrom (Bergstrom, 1988), will also be central in the present chapter. Many alternative but approximate estimation procedures, such as the approximate discrete model ADM (Bergstrom, 1966) or the multivariate latent differential equation MLDE (Boker, Neale, & Rausch, 2004) procedure, provide more or less accurate approximations of the underlying continuous time parameters on the basis of discrete time data. The EDM has the major advantage of linking the discrete time model parameters in an exact way to the underlying continuous time model parameters by means of nonlinear constraints. The EDM and estimation procedures using the EDM make sure that the parameters estimated are exactly equal to the parameters of the underlying differential equation model.

An alternative way to estimate the continuous time parameters for panel data through the EDM is Structural Equation Modeling (SEM). This was started by Oud (1978), employing the first published version of the SEM program LISREL (Jöreskog & Sörbom, 1976) described in Jöreskog’s (1977) seminal publication about SEM. Later, Arminger (1986) and Oud, van Leeuwe, and Jansen (1993) used other SEM program versions for the same purpose. A similar approach was followed by Tuma and Hannan (1984), although they used related simultaneous equations procedures rather than SEM. Common to all these authors is that they were inspired by Coleman (1968) to employ the so-called “indirect” method in estimating the EDM. This consists of first estimating discrete time parameters by means of a SEM or similar program and then separately, in a second step, deriving the continuous time parameter values using the EDM. In general, the indirect method cannot be recommended. A simple example, where the indirect method breaks down, is in the case of unequal observation intervals (Tuma & Hannan, 1984). Here the imposition of simple equality constraints by the SEM program does not work and the direct application of the nonlinear constraints is called for.

Oud and Jansen (2000) showed how more recent nonlinear SEM software packages such as Mx (Neale, Boker, Xie, & Maes, 2006) can also be employed for maximum likelihood estimation of the continuous-time state space model parameters, but using the direct method: applying the nonlinear constraints of the EDM directly during estimation. A thorough comparison between the LSDE/EDM procedure using filter techniques and the direct SEM/EDM procedure was made by Oud and Singer (2008) in a series of Monte Carlo simulation studies. It turns out that in case the same model is analyzed in both procedures and the data are appropriate for both procedures, the estimation results from filter techniques and SEM are equal. In this chapter we will exclusively deal with SEM.

Although continuous time modeling is pertinent to an extremely broad subject field in social science, most problems with discrete time analysis and their solution by means of continuous time modeling are covered by the topic of reciprocal causal relationships. Reciprocal relationships are traditionally analyzed in discrete time by means of the cross-lagged panel design. In the next section, we will first go into this popular but, from the continuous time perspective, insidious analysis design. The motivation and basic principles of continuous time modeling will be clarified on the basis of the cross-lagged panel design. The full-fledged model and its estimation will be dealt with in the ensuing sections.

7.2 Analysis of Reciprocal Relationships in the Cross-Lagged Panel Design

The cross-lagged panel design studies and compares the effects that variables have on each other across time. Different from cross-sectional research, the causal direction in panel research is not based on instantaneous relationships between simultaneously measured variables x and y. Instead, different variables are used for opposite directions: x at time point 1 affecting y at time point 2, y at time point 1 affecting x at time point 2 (see Figure 7.1). The cross-lagged panel design is therefore supposed to be more suitable than cross-sectional research in answering, for example, whether parenting characteristics affect adolescents' adjustment or, conversely, whether adolescents' adjustment affects parenting characteristics, or whether both effects operate reciprocally (Neiderhiser, Reiss, Hetherington, & Plomin, 1999).

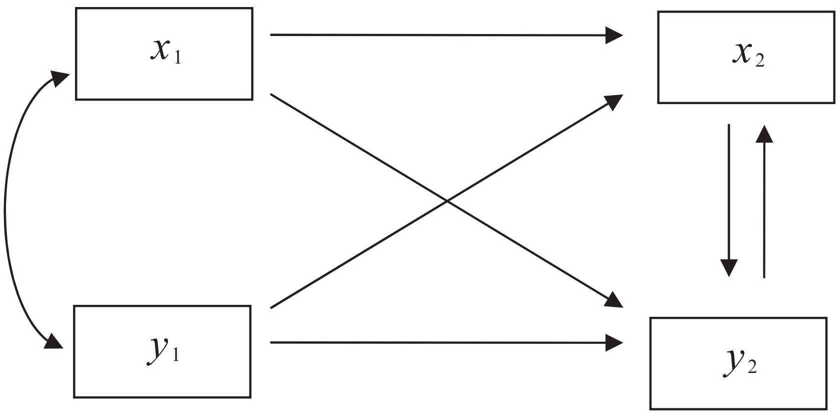

Fig. 7.1 Discrete-time cross-lagged panel design.

Having attracted attention in sociology earlier, the cross-lagged panel design is now becoming increasingly popular in psychology. Rueter and Conger (1998), for example, make clear that correlations between parental and children's behavior, which in the past were interpreted as unidirectional influences from parents to children, have in recent years assumed a reciprocal causal interpretation. This has led to a host of cross-lagged panel research to examine and test the direction of the effects. Other examples include cross-lagged reciprocal relationships between adolescent problem drug use, delinquent behavior, and emotional distress (Bui, Ellickson, & Bell, 2000), and between children's peer relations and antisocial behavior (Vuchinich, Bank, & Patterson, 1992).

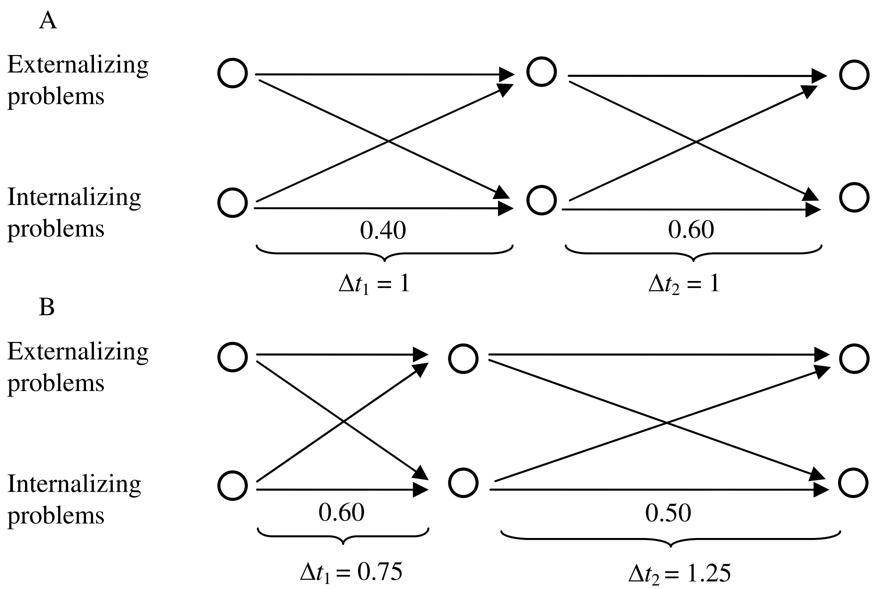

Fig. 7.2 Two cross-lagged panel designs with different spacings of the measurement time points and different values of the autoregression coefficients in the problem behavior variables.

Most cross-lagged analyses, however, are performed in discrete time. Although, for instance, parental behavior (x) and children's behavior (y), children's externalizing problem behavior (x) and children's internalizing problem behavior (y) or individualism (x) and ethnocentrism (y) influence themselves and each other continuously over time, measurements are typically taken not more than one or two times a year, resulting in a large observation interval. As a consequence, discrete time modeling becomes an oversimplification and often a distortion of reality. The path diagrams of the cross-lagged panel design in Figure 7.2 make this very clear. The oversimplification consists in the assumption that the arrows jump from one point in time to the next one and that nothing happens between measurements. In fact, the estimated cross-lagged coefficients (crossing arrows) and autoregression coefficients (horizontal arrows) over the observation interval Δt_i are complicated mixtures of the continuous time cross- and auto-effects in a constant interchange over, and heavily dependent on the length of, the chosen observation interval Δt_i. A variable with a high auto-effect, meaning that there is a strong tendency to sustain its value over time, tends also to retain the influence of 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 also be true: a relatively strong continuous time cross-effect having only small impact over a discrete time interval because of a rather low auto-effect in the dependent variable. Additionally, the result will be more strongly dependent on the auto-effect over the larger time interval ($\Delta t_2 = 1.25$ in diagram B) than over the shorter interval ($\Delta t_1 = 0.75$ in diagram B). So, the causal picture changes in discrete time, depending on the length of the chosen observation interval. Continuous time modeling is necessary to disentangle the continuous time cross-effects and auto-effects from the discrete time mixtures.

7.2.1 Relationship between Continuous and Discrete Time

The relationship between continuous and discrete time is governed by the matrix exponential

$$\mathbf{A}_{\Delta t_i} = \mathrm{e}^{\mathbf{A}\Delta t_i}. \tag{7.1}$$

Many paradoxical aspects of the relationship are explainable by the highly nonlinear character of the matrix exponential. Its power series definition will be given in (7.9) and a rather general computational form in (7.17). $\mathbf{A}_{\Delta t_i}$ is the discrete time autoregression matrix over observation interval $\Delta t_i = t_i - t_{i-1}$ $(i = 1, 2, ...)$ and $\mathbf{A}$ is the so-called drift matrix, which is the analogue of the autoregression matrix in continuous time. It is multiplied by the interval in the exponent of (7.1). Autoregression matrix $\mathbf{A}_{\Delta t_i}$ displays on the diagonal the autoregressions for each of the variables and off-diagonally the cross-lagged effects between the variables. Analogously, drift matrix $\mathbf{A}$ has the continuous time auto-effects on the diagonal and the continuous time cross-effects off-diagonally. It should be emphasized that (7.1), which specifies the exact relationship between $\mathbf{A}_{\Delta t_i}$ and $\mathbf{A}$, clearly shows that $\mathbf{A}_{\Delta t_i}$ changes as a function of the length of the observation interval, while $\mathbf{A}$ continues to be equal.

Table 7.1 gives a typical example of an $\mathbf{A}_{\Delta t_i}$ with corresponding exact $\mathbf{A}$, for $\Delta t_i = 1$ computed according to (7.1). The most conspicuous differences between the matrices are found in the diagonals of $\mathbf{A}_{\Delta t_i}$ (autoregressions 0.50, 0.40, and 0.30) and $\mathbf{A}$ (auto-effects –0.84, –1.05, and –1.60). Whereas the autoregressions in the diagonal of $\mathbf{A}_{\Delta t_i}$ are all positive, the corresponding auto-effects in $\mathbf{A}$ are all negative. This is a rather technical difference, which should be kept in mind when interpreting

Table 7.1 Discrete time autoregression matrix (left; $\Delta t_i = 1$) and corresponding continuous time drift matrix (right)

	$\mathbf{A}_{\Delta t_i}$: x_1	x_2	x_3	$\mathbf{A}$: x_1	x_2	x_3
x_1	0.50	0.30	0.22	−0.84	0.64	0.44
x_2	0.05	0.40	0.20	−0.09	−1.05	0.69
x_3	0.25	0.20	0.30	0.76	0.40	−1.60

differences. It is simply explained, however, when we take a closer look at the relationship between discrete and continuous time.

We start from the autoregression equation (7.2), specifying how by means of $\mathbf{A}_{\Delta t_i}$ each of the variables in vector $\mathbf{x}(t_i)$ is predictable by the variables in vector $\mathbf{x}(t_{i-1})$ at the previous time point:

$$\mathbf{x}(t_i) = \mathbf{A}_{\Delta t_i}\mathbf{x}(t_{i-1}). \tag{7.2}$$

For clarity, we do not yet specify an error component in (7.2), but this does not impact the relationship between autoregression matrix and drift matrix. From (7.2) we derive, dividing $\Delta\mathbf{x}(t_i) = \mathbf{x}(t_i) - \mathbf{x}(t_{i-1})$ by Δt_i:

$$\frac{\Delta\mathbf{x}(t_i)}{\Delta t_i} = \mathbf{A}_*\mathbf{x}(t_{i-1}) \quad \text{with } \mathbf{A}_* = (\mathbf{A}_{\Delta t_i} - \mathbf{I})/\Delta t_i. \tag{7.3}$$

Difference equation (7.3) in terms of $\mathbf{A}_*$ approximates differential equation (7.4) in terms of continuous time matrix $\mathbf{A}$:

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{A}\,\mathbf{x}(t). \tag{7.4}$$

We assume the rather general conditions to be satisfied, which guarantee a unique solution of (7.4) for initial value $\mathbf{x}(t_{i-1}) = \mathbf{x}(t_0) = \mathbf{x}_0$ (Zadeh & Desoer, 1963, p. 294). Note that, although the differential equation model is specified for all t in some continuous time interval and also its solution is valid for all t in the interval, the solution is observed only at the discrete time points t_i. While the solution is given in autoregression form for arbitrary discrete time points t_i in (7.2), it is made explicit in terms of continuous time drift matrix $\mathbf{A}$ by matrix exponential (7.1).

Basically, differential equation model (7.4) can thus be viewed as a transformation of the popular autoregression model (7.2). First in (7.3) difference quotient $\Delta\mathbf{x}(t_i)/\Delta t_i$ is placed on the left-hand side, approximation $\mathbf{A}_*$ of $\mathbf{A}$ on the right-hand side, and subsequently shifting the time interval Δt_i towards zero makes $\mathbf{A}_*$ approach $\mathbf{A}$ more and more closely. As seen in (7.3), in the transformation from $\mathbf{A}_{\Delta t_i}$ into $\mathbf{A}_*$, each autoregression value in the diagonal of $\mathbf{A}_{\Delta t_i}$ is diminished by 1 and so becomes negative for autoregressions between 0 and 1. This explains why in general positive but less than 1 autoregressions in discrete time correspond to negative auto-effects in continuous time. By equation (7.1), it is further evident that in the case of zero off-diagonals in $\mathbf{A}$, autoregressions between 0 and 1 must correspond to negative auto-effects. Zero off-diagonals cause the matrix exponential in (7.1) to reduce to scalar exponentials with negative values $-\infty < \mathrm{a} < 0$ in the diagonal of $\mathbf{A}$ leading to positive values $0 < \mathrm{e}^{\mathrm{a}\Delta t_i} < 1$ in the diagonal of $\mathbf{A}_{\Delta t_i}$ and vice versa. Note that in Table 7.1, the strength order of the positive autoregressions in the autoregression matrix (0.30 < 0.40 < 0. 50) is maintained in the negative drift matrix diagonals (–1.60 < –1.05 < –0.84). Depending on the off-diagonals, however, this is not necessarily the case.

Causally more interesting than the diagonals of the matrices in Table 7.1 are the paradoxical differences between discrete and continuous time that occur in the off-diagonal elements (effects between different variables). It turns out that the conclusions drawn in a discrete time analysis with respect to the cross-lagged coefficients in $\mathbf{A}_{\Delta t_i}$ may differ fundamentally from those to be drawn in a continuous time analysis on the basis of the corresponding cross-effects in $\mathbf{A}$.

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

For example, the two reciprocal cross-lagged coefficients with value 0.20 in the autoregression matrix – which in discrete time might lead to the conclusion that the strength of the causal effects between the variables x_2 and x_3 is equal in opposite directions – differ considerably in continuous time: 0.69 and 0.40.

- *The strength order of coefficients reverses going from discrete to continuous time.*

For example, in the autoregression matrix, the discrete time effect of x_3 on x_1 is greater than that of x_3 on x_2: 0.22 versus 0.20. However, in the corresponding drift matrix, it is the other way around: 0.44 for the first effect and 0.69 for the second effect.

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

The effect of x_1 on x_2 with positive value 0.05 in discrete time gets the negative value of –0.09 in continuous time. So, even interpreting the sign of the effect between variables is not safe for the transition from discrete to continuous time.

7.2.2 Discrete Time Problems with Unequal and Equal Observation Intervals

Continuous time analysis is needed to draw correct conclusions about causal effects. Discrete time analysis gets into extreme trouble, however, in the case of unequal observation intervals. When different discrete time distances are used in the same

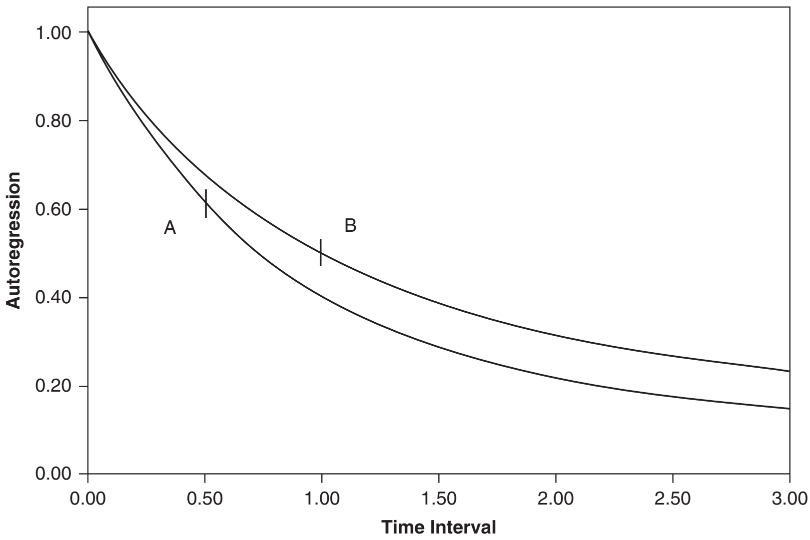

Fig. 7.3 Two different autoregression functions in two different studies A and B.

study or different researchers study the same causal effect in different distances, it becomes impossible to compare the strength of the causal effects found. This has long been recognized in discrete time analysis, notably by Gollob and Reichardt (1987). It seriously hampers cumulative progress of science, but cannot be solved in a discrete time context. This is exemplified in Figure 7.2 too. Supposing the other effects in the model to be equal over the two successive equal 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 the one over the second interval. In diagram B with unequal intervals, however, no decision can be made as to which one represents a bigger true auto-effect: 0.60 over interval $\Delta t_1 = 0.75$ or 0.50 over longer interval $\Delta t_2 = 1.25$. This is because autoregressions depend on the time interval, and, in general, the smaller the interval, the larger the autoregression, reaching 1 for $t = 0$. To find out whether or not the auto-effects over the intervals are indeed equal, again continuous time analysis is needed to relate and compare the discrete time effects on the same underlying continuous time scale.

The possibly misleading results of a discrete time analysis in case of unequal observation intervals are clearly shown by the autoregression functions A and B in Figure 7.3. By definition, autoregression functions have value 1 at an interval of length 0 (no change) and generally this value decreases, when the observation interval becomes longer. Suppose A is valid in one study and B in a second study, while in study A an observation interval Δt_1 of 0.50 year is used, and in study B an observation interval Δt_2 of 1.00 year. Because autoregression function B exceeds

A everywhere, no other conclusion should be drawn than that the autoregression in study A is lower than in study B. Nevertheless, investigator A, finding the autoregression value of 0.61 at interval $\Delta t_1 = 0.50$, could erroneously conclude that the autoregression in his study is larger than in study B, where the lower value of 0.50 was found at interval $\Delta t_2 = 1.00$. Clearly, the correct answer can only be found in continuous time analysis by comparing the auto-effects in the two studies and generating the complete autoregression functions as in Figure 7.3.

Some discrete time analysts believe that the use of unequal observation intervals is the only culprit and that all problems would be solved by using and making comparisons for equal intervals only. Equal observation intervals are hardly less problematic than unequal observation intervals, however, as will become clear from the two reciprocal cross-lagged effect functions for variables x_1 and x_3 in Figure 7.4, both based on drift matrix **A** in Table 7.1. The cross-lagged effect functions specify the cross-lagged effects, not only for one specific interval ($\Delta t_i = 1$ in Table 7.1) and even not only for all discrete time observation intervals Δt_i in the study. Like autoregression functions, cross-lagged effect functions go through all infinitesimally increasing intervals Δt in continuous time, starting from $\Delta t = 0$. Unlike autoregression functions, which start at value 1, cross-lagged effect functions have starting value 0 (different variables cannot yet have any influence on each other over a zero time interval), build up the effect more or less rapidly until a maximum is reached somewhere (in Figure 7.4 maxima 0.250 and 0.240 are reached at the quite different intervals of $\Delta t = 1.02$ and $\Delta t = 1.64$, respectively), and eventually return to 0 in a stable model. Stability is defined by the eigenvalues of drift matrix **A**. If all eigenvalues have negative real parts, the model is stable. Eigenvalues of **A** can become complex in some situations, but in this chapter only real eigenvalues will be considered .

Autoregression functions as well as cross-lagged effect functions were computed by the matrix exponential in (7.5) which differs from (7.1) merely in allowing Δt to take all values in continuous time:

$$\mathbf{A}_{\Delta t} = \mathrm{e}^{\mathbf{A}\Delta t}. \tag{7.5}$$

Crucial is that, in discrete time research, autoregression matrices $\mathbf{A}_{\Delta t_i}$ are defined and estimated for the observation intervals Δt_i in the study only and are therefore unknown for intervals that are smaller than or unequal to multiples of Δt_i ($i = 1, 2, ...T$), whereas $\mathbf{A}_{\Delta t} = \mathrm{e}^{\mathbf{A}\Delta t}$ in (7.5) is much more generally interpretable and computable for arbitrary continuous time intervals Δt. Basically, what we do in a continuous time analysis of discrete time data is first using (7.1) to find the continuous time drift matrix **A** that fits the empirical observation intervals, and next using (7.5) to generate the complete autoregression and cross-lagged effect functions on the basis of **A**.

A possible and by no means rare property of cross-lagged effect functions is shown in Figure 7.4. They are crossing at $\Delta t = 1.44$, both having the same value 0.239 at that interval. So, although according to **A** in Table 7.1 the effect of x_1 on x_3 is stronger than in the opposite direction from x_3 to x_1 (0.76 compared to 0.44) and

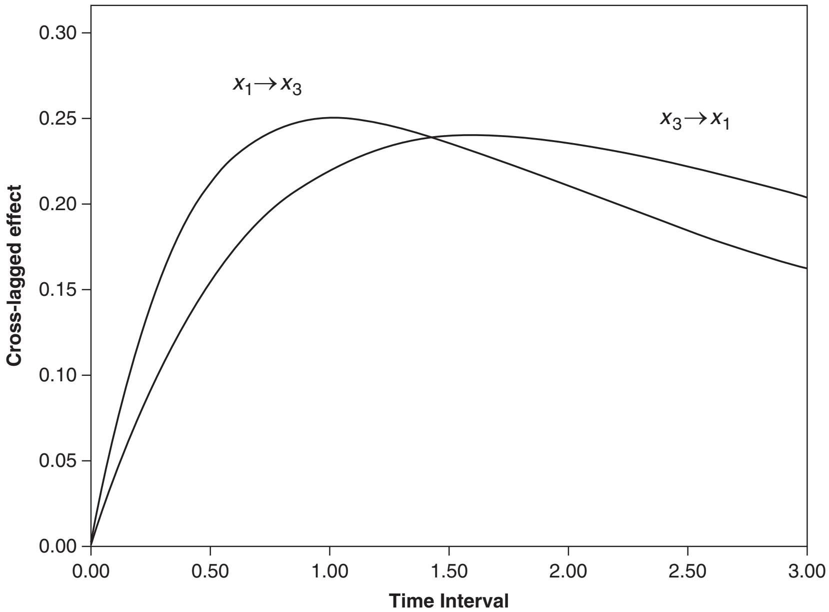

Fig. 7.4 Cross-lagged effect functions for the reciprocal effects between x_1 and x_3, based on the drift matrix in Table 7.1.

the cross-lagged effects in $\mathbf{A}_{\Delta t_i}$ of Table 7.1 indicate the same strength order at time interval $\Delta t = 1.00$ (0.25 compared to 0.22, both displayed also at $\Delta t = 1.00$ in Figure 7.4), the interplay between the variables over continuous time is such that the cross-lagged effects in Figure 7.4 become equally strong at $\Delta t = 1.44$ and even reverse the strength order for intervals $\Delta t > 1.44$. It is this possibility of crossing (non-monotone) cross-lagged effect functions (as well as non-monotone autoregression functions) that makes discrete time analysis useless for analyzing reciprocal relationships in the cross-lagged panel design. The implication of Figure 7.4 is that the relative strength of the reciprocal causal effects found between x_3 and x_1 depends on the observation interval chosen in the study. Investigators choosing their discrete time interval Δt_i between 0 and 1.44 years will come to the conclusion that x_1 has a stronger effect on x_3 (maximum difference of 0.058 reached at Δt_i = 0.46), whereas investigators choosing $\Delta t_i > 1.44$ years will arrive at the opposite conclusion (maximum difference of 0.042 reached at Δt_i = 3.26). No comparison problems would arise, at least not in the sense of contradictory results with regard to the strength order, if the cross-lagged effect functions in Figure 7.4, like the autoregression functions in Figure 7.3, were monotone (not crossing). Then it would not matter at what interval the comparison is made, because one would find the same order everywhere. However, the monotone or non-monotone character is seldom known beforehand and generally it is just the purpose of the research to find out.

What is worse is that it would not be of any help to choose and maintain the same observation interval. The cross-lagged effect functions in Figure 7.4 bring discrete time investigators using the same observation interval $\Delta t_i = 1.44$ also to a false conclusion, namely that the effect of x_1 on x_3 is equally strong as the effect of x_3 on x_1. This conclusion cannot be generalized to observation intervals $\Delta t_i \neq 1.44$, however, where different and, for $\Delta t_i > 1.44$, again false conclusions would be drawn, nor is it confirmed by the cross-effect coefficients in drift matrix **A**. Clearly, continuous time analysis, estimating the coefficients of the continuous time drift matrix **A** and displaying the consequences over the complete time axis by means of the full autoregression and cross-lagged effect functions, is the only solution to the problems of unequal as well as equal observation intervals.

7.2.3 Lagged and Instantaneous Effects Dilemma

We conclude the discussion of the cross-lagged panel design with another awkward problem often encountered by discrete time analysts, for which, again, there is no solution in discrete time. As seen in Figure 7.1, the analysis of the cross-lagged panel design allows the inclusion of two kinds of reciprocal effects between x and y: lagged reciprocal effects (i.e., x at time point 1 affecting y at time point 2, and y at time point 1 affecting x at time point 2) and instantaneous reciprocal effects (i.e., x at time point 2 affecting y at time point 2 and vice versa). One could choose the instantaneous coefficients, the lagged coefficients, or both to be present in the model, but the results are often different or even contradictory. This typically confronts the discrete time analyst with a dilemma. In the study by Vuchinich, Bank, and Patterson (1992), for example, the dilemma was whether to choose for instantaneous or lagged effects between parental disciplinary behavior and child antisocial behavior. The authors found significant instantaneous effects but no significant cross-lagged effects. The choice would become even more difficult, if these effects were to be estimated simultaneously, because then the results become highly dependent on the time interval Δt_i. In general, the longer the time interval between measurements, the higher the instantaneous coefficients become in comparison to the lagged coefficients. Most discrete time analysts feel that instantaneous and lagged effects should both be taken into consideration somehow. However, they do not and cannot know in discrete time how to connect and constrain these two types of effects to find the true underlying continuous time effects.

When analysts estimate the instantaneous and lagged effects simultaneously, autoregression equation (7.2) is in fact replaced by

$$\mathbf{x}(t_i) = \mathbf{A}_{ins}\mathbf{x}(t_i) + \mathbf{A}_{lag}\mathbf{x}(t_{i-1}). \tag{7.6}$$

Instantaneous matrix $\mathbf{A}_{ins}$ includes the instantaneous effects between the "current endogenous" variables in $\mathbf{x}(t_i)$. Lagged matrix $\mathbf{A}_{lag}$ includes the lagged effects from the "lagged endogenous" $\mathbf{x}(t_{i-1})$ on the "current endogenous" $\mathbf{x}(t_i)$. Equation (7.6) is known in econometrics as the "structural form" with (7.2) as the associated "reduced form". The clear relationships that exist between the two forms and the continuous time matrix exponential in (7.1), that is, between the coefficients in the reduced form or autoregression matrix $\mathbf{A}_{\Delta t_i}$, in the structural form matrices $\mathbf{A}_{ins}$ and $\mathbf{A}_{lag}$, and in the continuous time drift matrix $\mathbf{A}$, are shown in (7.7):

$$\mathbf{A}_{\Delta t_i} = (\mathbf{I} - \mathbf{A}_{ins})^{-1}\mathbf{A}_{lag} = \mathrm{e}^{\mathbf{A}\Delta t_i}. \tag{7.7}$$

By means of $\mathrm{e}^{\mathbf{A}\Delta t_i}$ in (7.7), which is the core of the exact discrete model EDM, nonlinear constraints are directly imposed on the coefficients in the autoregression matrix $\mathbf{A}_{\Delta t_i}$ for generating the exact drift matrix $\mathbf{A}$, skipping $\mathbf{A}_{ins}$ and $\mathbf{A}_{lag}$. In this way, the above mentioned dilemma of the choice between $\mathbf{A}_{ins}$ and $\mathbf{A}_{lag}$ is simply circumvented. The structural form is therefore not really indispensible in continuous time analysis. One might wonder, however, whether constraints could be imposed on the structural form matrices $\mathbf{A}_{ins}$ and $\mathbf{A}_{lag}$ for combining the instantaneous and lagged effects in an appropriate way to generate the underlying continuous time effects in $\mathbf{A}$ and thereby explicitly solving the discrete time dilemma. This has indeed been done in the so-called approximate discrete model ADM introduced by Bergstrom (1966; 1984, pp. 1172-1173), the same econometrician who originated the EDM. He showed, that by means of the simple linear constraints:

$$\begin{aligned} \mathbf{A}_{ins} &= \tfrac{1}{2}\mathbf{A}_{\dagger}\Delta t_i, \\ \mathbf{A}_{lag} &= \mathbf{I} + \tfrac{1}{2}\mathbf{A}_{\dagger}\Delta t_i, \end{aligned} \tag{7.8}$$

the ADM generates a quite reasonable approximation $\mathbf{A}_{\dagger}$ of exact $\mathbf{A}$. It immediately solves the dilemma of the discrete time analyst, because by means of (7.8) the two different matrices $\mathbf{A}_{ins}$ and $\mathbf{A}_{lag}$ are replaced by one and the same matrix $\mathbf{A}_{\dagger}$, which is the one to be interpreted and tested.

7.2.4 ADM and EDM

It is true that by using the ADM instead of the EDM one sacrifices exactness. However, although nonlinear SEM programs such as Mx, which include the exponential and matrix algebraic functions, can implement the EDM, the linearity of the constraints in (7.8) of the ADM also holds some attraction. Less nonlinearly oriented but more user-friendly SEM programs, which lack the exponential and matrix algebraic functions and therefore the possibility to apply the EDM, mostly allow implementation of the ADM. Oud (2007b) explains in detail how to apply the ADM-SEM procedure by means of LISREL (Jöreskog & Sörbom, 1996). In addition, LISREL and similar programs are particularly valuable in the modeling process, because they provide plenty of information about model fit and about modification results of individual parameters by means of the so-called modification indices. For this reason, it could be worthwhile in practice to first apply the ADM-SEM procedure in the model building phase by means of a program such as LISREL and then the EDM-SEM procedure in the final model estimation phase by means of Mx.

It should be noted that $\mathbf{A}_{\dagger}$, as an approximation of $\mathbf{A}$, compares favorably with other well-known approximations such as the relatively crude approximation $\mathbf{A}_{*} = (\mathbf{A}_{\Delta t_i} - \mathbf{I})/\Delta t_i$ in (7.3). This is seen by putting the exact nonlinear matrix exponential form $\mathbf{A}_{\Delta t_i} = \mathrm{e}^{\mathbf{A}\Delta t_i}$ and both approximate linear constraint forms in power series expansion

$$\begin{aligned}
\mathbf{A}_{\Delta t_i} &= \mathrm{e}^{\mathbf{A}\Delta t_i} = \textstyle\sum_{k=0}^{\infty}(\mathbf{A}\Delta t_i)^k/k! \\
&= \mathbf{I} + \mathbf{A}\Delta t_i + \tfrac{1}{2}\mathbf{A}^2\Delta t_i^2 + \tfrac{1}{6}\mathbf{A}^3\Delta t_i^3 + \tfrac{1}{24}\mathbf{A}^4\Delta t_i^4 + \ldots \\
&\qquad \text{(exact)}, \\
\mathbf{A}_{\Delta t_i} &= (\mathbf{I} - \mathbf{A}_{ins})^{-1}\mathbf{A}_{lag} = (\mathbf{I} - \tfrac{1}{2}\mathbf{A}_{\dagger}\Delta t_i)^{-1}(\mathbf{I} + \tfrac{1}{2}\mathbf{A}_{\dagger}\Delta t_i) \\
&= \mathbf{I} + \mathbf{A}_{\dagger}\Delta t_i + \tfrac{1}{2}\mathbf{A}_{\dagger}^2\Delta t_i^2 + \tfrac{1}{4}\mathbf{A}_{\dagger}^2\Delta t_i^3 + \tfrac{1}{8}\mathbf{A}_{\dagger}^4\Delta t_i^4 + \ldots \\
&\qquad (\mathbf{A}_{\dagger} \text{ approximation}), \\
\mathbf{A}_{\Delta t_i} &= \mathbf{I} + \mathbf{A}_{*}\Delta t_i \quad (\mathbf{A}_{*} \text{ approximation}).
\end{aligned} \tag{7.9}$$

Whereas the $\mathbf{A}_{*}$ approximation truncates the exact infinite series, the weights of the $\mathbf{A}_{\dagger}$ approximation $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots)$ in the ADM are only seen to decrease less quickly than in the exact series $(\frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \ldots)$ used in the EDM. In a simulation study with different estimation procedures, Oud (2007a) concluded that the ADM-SEM procedure did indeed yield more biased results than the EDM-SEM procedure, but that the overall quality in terms of the root mean squared error (RMSE) was hardly lower than in the EDM-SEM procedure. The ADM-SEM procedure compared also favorably with the approximate MLDE procedure of Boker, Neale, and Rausch (2004). In the examples to be presented below, we will first apply the ADM-SEM procedure, followed by the EDM-SEM procedure.

7.3 Linear Stochastic Differential Equation Model

The full linear stochastic differential equation model used in this chapter consists of two equations: a dynamic explanatory equation and a static measurement equation. The dynamic equation, shown in (7.10), extends the basic differential equation model in (7.4) by three important elements.

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{A}\mathbf{x}(t) + \mathbf{b} + \boldsymbol{\kappa} + \mathbf{G}\frac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t}. \tag{7.10}$$

In addition to the drift matrix term $\mathbf{A}\mathbf{x}(t)$, introduced and discussed in detail in the previous section, the following new elements are found in (7.10): continuous time intercepts vector $\mathbf{b}$, continuous time "trait" variables vector $\boldsymbol{\kappa}$, and finally the continuous time error process vector $\mathbf{G}\frac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t}$, which makes the differential equation stochastic.

7.3.1 Continuous Time Intercepts

Nonzero intercepts in vector $\mathbf{b}$ aptly accommodate for the frequently observed nonzero mean trajectories $E[\mathbf{x}(t)]$. Importantly, $\mathbf{b}$ defines also the final means towards which a stable system eventually converges. In models with $\mathbf{b} = \mathbf{0}$ these final means are necessarily zero, with zero being the stable equilibrium position. An equilibrium position is the value of a constant curve satisfying the model. In models with $\mathbf{b} \neq \mathbf{0}$, it can first be proven, on the basis of (7.10), that the mean trajectories, starting from initial mean $E[\mathbf{x}(t_0)]$, follow nonzero curves (7.11) and next that convergence is towards nonzero $-\mathbf{A}^{-1}\mathbf{b}$.

$$E[\mathbf{x}(t)] = \mathrm{e}^{\mathbf{A}(t-t_0)}E[\mathbf{x}(t_0)] + \mathbf{A}^{-1}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}]\mathbf{b} \quad \text{with } E[\mathbf{x}(t \to \infty)] = -\mathbf{A}^{-1}\mathbf{b}. \tag{7.11}$$

The reason for the latter is the behavior of the matrix exponential $\mathrm{e}^{\mathbf{A}(t-t_0)}$ in a stable model. Because $\mathbf{A}$ in the exponent has negative eigenvalues and is multiplied by the interval length $t-t_0$, the matrix exponential eventually becomes zero (the concept of stability is equivalent to this property) and hence $E[\mathbf{x}(t)] \to -\mathbf{A}^{-1}\mathbf{b}$ for $t \to \infty$. The values in $-\mathbf{A}^{-1}\mathbf{b}$, in addition, are (stable) equilibrium positions. As a result of the commuting property $\mathrm{e}^{\mathbf{A}(t-t_0)}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathrm{e}^{\mathbf{A}(t-t_0)}$, choosing initial means $E[\mathbf{x}(t_0)] = -\mathbf{A}^{-1}\mathbf{b}$ in (7.11) leaves $E[\mathbf{x}(t)]$ unchanged.

So, the intercepts $\mathbf{b}$ enhance the flexibility of the model by allowing nonzero mean trajectories and nonzero final means. Flexibility is further enhanced by the possibility of subpopulation specific mean trajectories and final means within the same overall model. For this purpose (n x 1)-vector $\mathbf{b}$ is replaced by $\mathbf{B}_u\mathbf{u}$, which is also (n x 1) but the product of (n x r)-matrix $\mathbf{B}_u$ of regression coefficients and (r x 1)-vector $\mathbf{u}$ of exogenous variables. Suppose, for example, that boys and girls are assumed to follow a different development and to reach a different final position. As the first element of $\mathbf{u}$ we choose the unit variable, $u_1 = 1$ for all subjects in the population, and as the second element a dummy-variable, coded $u_2 = 0$ for boys and $u_2 = 1$ for girls. Let us call the first column of $\mathbf{B}_u$ $\mathbf{b}_1$ and the second column $\mathbf{b}_2$. By replacing $\mathbf{b}$ in (7.11) by $\mathbf{b}_1$ for boys and by $\mathbf{b}_1 + \mathbf{b}_2$ for girls, we then get two sets of n mean trajectories, $E[\mathbf{x}(t)]_{u_2=0}$ for boys and $E[\mathbf{x}(t)]_{u_2=1}$ for girls, and two sets of n final means, $E[\mathbf{x}(t \to \infty)]_{u_2=0} = -\mathbf{A}^{-1}\mathbf{b}_1$ for boys and $E[\mathbf{x}(t \to \infty)]_{u_2=1} = -\mathbf{A}^{-1}(\mathbf{b}_1 + \mathbf{b}_2)$ for girls. By using the same procedure to differentiate $E[\mathbf{x}(t_0)]_{u_2=0}$ for boys from $E[\mathbf{x}(t_0)]_{u_2=1}$ for girls, we additionally let boys and girls start from different positions. The procedure is easily extended for more than two subpopulations, more than two variables in $\mathbf{u}$ and, in addition to dummy-variables, also for metric variables in $\mathbf{u}$ such as income and, as shown by Oud and Singer (2008), even for changing exogenous variables (time-varying covariates) $\mathbf{u}(t)$. These kinds of models are often called conditional (e.g., see the chapter of Bollen and Zimmer in this volume). For metric variables u_i with many values (subpopulations) represented in the sample, which is typically the case with metric variables as, for example, income, the procedure outlined is often the only possible one. However, in the case of a limited number of subpopulations (e.g., boys and girls), an attractive alternative approach is performing a so-called multisample SEM analysis (Jöreskog & Sörbom, 1996), in which $\mathbf{b}$, $E[\mathbf{x}(t_0)]$, and possibly other parameters are allowed to vary in the subpopulations.

It should be noted that the intercepts in $\mathbf{b}$ are feeding the system continuously over time by a constant amount and therefore indeed result in different contributions from unequal intervals. In a discrete time model the intercepts contribute only at the observation time points chosen.

7.3.2 Continuous Time Trait Variables

Although, as discussed above, subpopulation intercepts allow a different mean trajectory and different final mean in each subpopulation, it is nevertheless paradoxical that a subject's current and future expected behavior should be exclusively determined by the population or subpopulation the subject happens to be modeled a member of. The flexibility of the model is further enhanced by the specification of random subject effects $\boldsymbol{\kappa}$ in (7.10): random intercept variables, called "trait" variables in the present chapter, which define for every subject an own subject-specific mean trajectory. The trait variables $\boldsymbol{\kappa}$, in distinction from the changing "state" variables $\mathbf{x}(t)$, have constant values across time as do the fixed intercepts $\mathbf{b}$. However, whereas the $\mathbf{b}$ are also constant across subjects, the normally distributed trait variables $\boldsymbol{\kappa}$ with mean $E(\boldsymbol{\kappa}) = \mathbf{0}$ and covariance matrix $\Phi_{\kappa} \neq \mathbf{0}$ have a different value for each subject and so model the subject specific deviations from the common mean defined by $\mathbf{b}$.

The constancy across time implies that $\boldsymbol{\kappa}$ already influences $\mathbf{x}(t)$ before the initial time point t_0, so that $\boldsymbol{\kappa}$ should be considered part of the initial state $\mathbf{x}(t_0)$ and, in general, $\boldsymbol{\kappa}$ and $\mathbf{x}(t_0)$ are correlated ($\Phi_{\mathbf{x}_{t_0},\kappa} \neq \mathbf{0}$). Both the variances of the trait variables in Φ_{κ} and their covariances with $\mathbf{x}(t_0)$ in $\Phi_{\mathbf{x}_{t_0},\kappa}$ are testable quantities. Both are expected to be nonzero, if subjects do indeed follow their subject-specific mean trajectory instead of coinciding with a single general mean trajectory. Supposing this is indeed the case, the distance between a subject-specific mean trajectory $E[\mathbf{x}(t)|\boldsymbol{\kappa}]$ and the (sub)population mean trajectory $E[\mathbf{x}(t)]$ is computed as

$$\begin{aligned} E[\mathbf{x}(t)|\boldsymbol{\kappa}] - E[\mathbf{x}(t)] &= \mathrm{e}^{\mathbf{A}(t-t_0)}\Phi_{\mathbf{x}_{t_0},\kappa}\Phi_{\kappa}^{-1}\boldsymbol{\kappa} + \mathbf{A}^{-1}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}]\boldsymbol{\kappa} \\ \text{with } \; E[\mathbf{x}(t \to \infty)|\boldsymbol{\kappa}] - E[\mathbf{x}(t \to \infty)] &= -\mathbf{A}^{-1}\boldsymbol{\kappa}. \end{aligned} \tag{7.12}$$

Here the first term on the right-hand side is a consequence of the regression of $\mathbf{x}(t_0)$ on $\boldsymbol{\kappa}$. As a result of the matrix exponential going again to zero for $t \to \infty$ in a stable model, the distance between the subject-specific and the (sub)population mean trajectories goes to a constant nonzero value: $-\mathbf{A}^{-1}\boldsymbol{\kappa}$.

It should be noted that the mean trajectories for (sub)populations or subjects are not only interesting as such, and in many cases even the main purpose of the study, but they also play a crucial role in the behavior of the estimated sample trajectories (conditional means $E[\mathbf{x}(t)|\mathbf{y}]$, where $\mathbf{y}$ is the total data vector of the subject). The reason is that these regress towards the mean trajectories (in a stable model) or egress from them (in an unstable model). Particularly, if a model contains trait variables, a subject's conditional mean regresses towards its own subject-specific mean trajectory (see (7.12)), whereas in a pure state model all subjects regress towards one and the same general mean trajectory (see (7.11)). Figure 7.5, taken from a youth delinquency study, shows the estimates of a mean trajectory, a subject-specific mean trajectory, and the subject's estimated sample trajectory. Outside of the measurement time points (recognizable by the kinks in the curve) the subject's sample trajectory is clearly seen to regress towards its subject-specific mean trajectory. The consequences are particularly dramatic for predictions. As the final values in the study are 2.12 for the mean trajectory and 3.88 for the subject-specific mean trajectory, the predicted final value for the subject differs no less than 1.76 from the one that would be found in a pure state model.

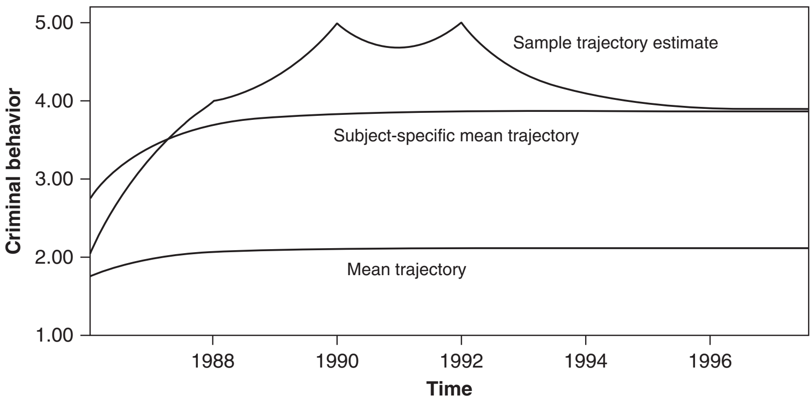

Fig. 7.5 Examples of mean trajectory, subject-specific mean trajectory and sample trajectory estimate.

7.3.3 Continuous Time Error Process

It hardly needs comment that the introduction of an error term into a longitudinal model enhances the applicability of the model. Historically, it took quite some time, however, to define continuous time error in a mathematically rigorous way. The continuous time error process $\mathbf{G}\frac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t}$ in (7.10) builds upon the famous Wiener process $\mathbf{W}(t)$, the random walk through continuous time. At first sight $\mathbf{G}\frac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t}$ looks quite complicated. Note, however, that discrete time error can be thought of analogously as the difference quotient $\frac{\Delta \mathbf{w}_t}{\Delta t}$ of a discrete time random walk $\mathbf{w}_t = \mathbf{w}_{t-\Delta t} + \mathbf{e}$. If the step sizes $\mathbf{e}$ for Δt = 1 are randomly drawn from a standard multinormal distribution $N(\mathbf{0}, \mathbf{I})$, successive non-overlapping increments $\Delta \mathbf{w}_t = \mathbf{w}_t - \mathbf{w}_{t-\Delta t}$ for $\Delta t \geq 1$ are independent with covariance matrix $\Delta t\mathbf{I}$. If one wants to model nonstandard error with larger or smaller variance than 1 for $\Delta t = 1$ and possibly correlated elements, but no change in the other properties, then the difference quotient for $\Delta t = 1$ could first be multiplied by lower triangular matrix $\mathbf{G}$, Cholesky factor of the desired covariance matrix $\mathbf{Q}$.

The properties defining the standard Wiener process $\mathbf{W}(t)$ are, in addition to its sample trajectories being continuous and starting at $\mathbf{W}(0) = \mathbf{0}$ (both with probability 1), precisely the conditions of independently and normally distributed increments $\Delta\mathbf{W}(t) = \mathbf{W}(t) - \mathbf{W}(t - \Delta t)$ with mean $\mathbf{0}$ and covariance matrix $\Delta t\mathbf{I}$ (Arnold, 1974, p. 46; Kuo, 2006, p. 7). The lower triangular matrix $\mathbf{G}$ in the continuous time error process $\mathbf{G}\frac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t}$ is just there to allow increment variances to become lower or higher than 1 for $\Delta t = 1$ and to get nonzero correlations between elements. Product $\mathbf{Q} = \mathbf{G}\mathbf{G}'$ is the continuous time error covariance matrix, called "diffusion" matrix, and $\mathbf{G}$ the Cholesky factor of $\mathbf{Q}$. So, $\mathbf{Q}$ and $\mathbf{G}$ provide the same information and are easily expressed into each other.

The fame of the Wiener process is undoubtedly due to two peculiar facts that have given rise to a host of mathematical research. Its derivative $\frac{\mathrm{d}\mathbf{W}(t)}{\mathrm{d}t}$ or "white noise" cannot be defined as a derivative in the normal sense nor can the stochastic integral $\int_{t_0}^{t}\mathbf{F}(s)\mathrm{d}\mathbf{W}(s)$ in terms of a possibly time-varying function $\mathbf{F}(t)$ be defined as an ordinary integral. Solution (7.13) of stochastic differential equation (7.10) (see e.g., Arnold, 1974, pp. 128-134) is nevertheless seen to contain this type of integral for the error component. Defined in a proper way, however, its correct covariance matrix can be derived as given in (7.13):

$$
\begin{aligned}
&\mathbf{x}(t) = \mathrm{e}^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \mathbf{A}^{-1}[\mathrm{e}^{\mathbf{A}(t-t_0)} - \mathbf{I}](\mathbf{b}+\boldsymbol{\kappa}) + \int_{t_0}^{t}\mathrm{e}^{\mathbf{A}(t-s)}\mathbf{G}\mathrm{d}\mathbf{W}(s)\\
&\quad\text{with } \operatorname{cov}\int_{t_0}^{t}\mathrm{e}^{\mathbf{A}(t-s)}\mathbf{G}\mathrm{d}\mathbf{W}(s) = \int_{t_0}^{t}\mathrm{e}^{\mathbf{A}(t-s)}\mathbf{Q}\mathrm{e}^{\mathbf{A}'(t-s)}\mathrm{d}s\\
&\qquad\qquad = \operatorname{irow}\{\mathbf{A}_{\#}^{-1}[\mathrm{e}^{\mathbf{A}_{\#}(t-t_0)} - \mathbf{I}]\operatorname{row}\mathbf{Q}\}\\
&\qquad\qquad \text{for } \mathbf{Q} = \mathbf{G}\mathbf{G}' \text{ and } \mathbf{A}_{\#} = \mathbf{A}\otimes\mathbf{I} + \mathbf{I}\otimes\mathbf{A}.
\end{aligned}
\tag{7.13}
$$

Here $\otimes$ is the Kronecker product (postmultiplying each element of the former matrix by the latter matrix), "row" the rowvec operation (putting the elements of a matrix row-wise in a column vector), and “irow” the inverse operation (putting the elements back into the matrix again).

Note in solution (7.13) the predominant role of the matrix exponential having drift matrix $\mathbf{A}$ in the exponent, which appears in all three terms in the solution. Note also the similar structure of the integral expression $\mathbf{A}^{-1}[e^{\mathbf{A}(t-t_0)}-\mathbf{I}]$ in the second term and $\mathbf{A}_{\#}^{-1}[e^{\mathbf{A}_{\#}(t-t_0)}-\mathbf{I}]$ in the covariance matrix of the error term, replacing (nxn) matrix $\mathbf{A}$ in the second term by $(n^2 x\, n^2)$ matrix $\mathbf{A}_{\#}$ in the covariance matrix. $\mathbf{A}_{\#}$ is based again on drift matrix $\mathbf{A}$ and has all eigenvalues negative, if $\mathbf{A}$ has all eigenvalues negative. Therefore, analogously to final mean $-\mathbf{A}^{-1}\mathbf{b}$, the final error covariance matrix in a stable model is given by $\text{irow}(-\mathbf{A}_{\#}^{-1}\text{row}\,\mathbf{Q})$. If, for $\boldsymbol{\kappa}=\mathbf{0}$, the process starts with mean $-\mathbf{A}^{-1}\mathbf{b}$ and covariance matrix $\text{irow}(-\mathbf{A}_{\#}^{-1}\text{row}\,\mathbf{Q})$, the process is stationary, keeping the same mean and covariance matrix. Details about solution (7.13) can be found in Singer (1990) and Oud and Jansen (2000). It has the form that allows all parameters of the model to be estimated by means of the EDM as will be shown in the next section. Observe also, that the mean trajectory (7.11) is an immediate derivation from solution (7.13).

The rationale behind the approximate ADM procedure is different. The ADM is not based on the differential equation solution (7.13), where $\mathbf{x}(t)$ appears only on one side of the equation, but puts differential equation (7.10) first in integral form:

$$\begin{aligned}&\textstyle\int_{t-\Delta t}^{t}\mathrm{d}\mathbf{x}(s)=\mathbf{A}\int_{t-\Delta t}^{t}\mathbf{x}(s)\mathrm{d}(s)+(\mathbf{b}+\boldsymbol{\kappa})\Delta t+\mathbf{G}[\mathbf{W}(t)-\mathbf{W}(t-\Delta t)].\\ &\text{or}\\ &\mathbf{x}(t)=\mathbf{x}(t-\Delta t)+\mathbf{A}\textstyle\int_{t-\Delta t}^{t}\mathbf{x}(s)\mathrm{d}(s)+(\mathbf{b}+\boldsymbol{\kappa})\Delta t+\mathbf{G}\Delta\mathbf{W}(t).\end{aligned}\tag{7.14}$$

It next replaces the integral $\int_{t-\Delta t}^{t}\mathbf{x}(s)\mathrm{d}(s)$ on the right-hand side, having $\mathbf{x}(t)$ still inside of the integral, by the so-called trapezoid approximation $\frac{1}{2}[\mathbf{x}(t)+\mathbf{x}(t-\Delta t)]\Delta t$, which multiplies the length Δt of the integration interval by the average value at the end points. This gives rise to the approximate solution

$$\begin{aligned}\mathbf{x}(t)\approx[\tfrac{1}{2}\mathbf{A}\Delta t]\mathbf{x}(t)+[\mathbf{I}+\tfrac{1}{2}\mathbf{A}\Delta t]\mathbf{x}(t-\Delta t)+(\mathbf{b}+\boldsymbol{\kappa})\Delta t+\mathbf{G}\Delta\mathbf{W}(t)\\ \text{with }\operatorname{cov}[\mathbf{G}\Delta\mathbf{W}(t)]=\mathbf{G}\mathbf{G}'\Delta t.\end{aligned}\tag{7.15}$$

It explains the constraints imposed in (7.8) on the instantaneous and lagged coefficients for obtaining approximation $\mathbf{A}_{\dagger}$ of $\mathbf{A}$.

7.3.4 Measurement Equation

Latent variables abound in social science. It is probably no exaggeration to claim that the greater part of psychology and sociology draws on latent variables. Some of the latent variables, such as the trait variables in $\boldsymbol{\kappa}$ of (7.10) or the state variables $\mathbf{x}(t)$ in between measurement time points t_i, have no direct connection at all to the observed variables. For the latent state variables at the measurement points

t_i, however, we need to extend the model with a measurement equation, specifying how each of them is connected to the directly observed variables in $\mathbf{y}_{t_i}$:

$$\mathbf{y}_{t_i} = \mathbf{C}\mathbf{x}(t_i) + \mathbf{d} + \mathbf{v}_{t_i} \quad \text{with} \quad \text{cov}(\mathbf{v}_{t_i}) = \mathbf{R}. \tag{7.16}$$

Parameter matrix $\mathbf{C}$ specifies the loadings of the observed variables on the latent variables, parameter vector $\mathbf{d}$ the measurement intercepts or origins, and $\mathbf{R}$ the measurement error variances of the observed variables. If the state variables at the measurement time points are all observed, $\mathbf{y}_{t_i} = \mathbf{x}(t_i)$ and we specify $\mathbf{C} = \mathbf{I}$, $\mathbf{d} = \mathbf{0}$, $\mathbf{R} = \mathbf{0}$.

For identification reasons it is customary to fix, for each latent variable, one of the loadings in $\mathbf{C}$ at 1 and one of the measurement origins in $\mathbf{d}$ at 0. These values fix the measurement scale of the latent variable, 1 in $\mathbf{C}$ giving the latent variable, apart from measurement error variance, the same variance as the observed variable involved and 0 in $\mathbf{d}$ giving the same mean. Observe that the absence of a time index for the measurement parameter matrices and vector $\mathbf{C}$, $\mathbf{d}$, and $\mathbf{R}$ makes them time-invariant. Although time-invariance of the measurement model (measurement invariance) is no strict requirement, it is nevertheless extremely important for making sure that the latent variables keep the same meaning over time. One should have convincing reasons to deviate from measurement invariance for specific variables. Therefore, in the further development of the model, we will assume time-invariance.

7.4 Model Estimation by Means of SEM

In this section, based on the exact differential equation solution (7.13) and the approximate integral form (7.15), respectively, the full EDM and the full ADM will be formulated. Next, for estimation by means of a SEM program, all EDM or ADM parameter matrices will be put into inclusive SEM parameter matrices.

7.4.1 Full EDM

As will be clear from the subscripts t_i and Δt_i in the full EDM (7.17), the EDM is a discrete time model. The matrices with subscript Δt_i in (7.17) are defined for the discrete-time measurement time points only. Simultaneously, however, the EDM covers the continuous time model because of the nonlinear constraints imposed on the discrete time matrices in terms of the continuous time matrices from differential equation (7.10). It means that by applying the constraints on the discrete time matrices $\mathbf{A}_{\Delta t_i}, \mathbf{b}_{\Delta t_i}, \mathbf{H}_{\Delta t_i}, \mathbf{Q}_{\Delta t_i}$ during estimation, we simultaneously estimate the underlying continuous time parameter matrices $\mathbf{A}$, $\mathbf{b}$, Φ_{κ},$\Phi_{\mathbf{x}_{t_0},\kappa}$, $\mathbf{Q} = \mathbf{G}\mathbf{G}'$ (for convenience, vectors $\mathbf{b}_{\Delta t_i}$ and $\mathbf{b}$ are called matrices). The connection between the differential equation and the EDM is made by the exact solution (7.13) (choosing for

$t-t_0$ the observation intervals $\Delta t_i = t_i - t_{i-1}$, starting with $\Delta t_1 = t_1 - t_0$). All constraints on the discrete time matrices in EDM (7.17) are directly taken from exact solution (7.13).

For the computation of the matrix exponential $e^{\mathbf{A}\Delta t_i}$ in (7.17) the diagonalization method is used, which reduces the computation to scalar exponentials. After first diagonalizing $\mathbf{A} = \mathbf{MVM}^{-1}$ ($\mathbf{M}$ eigenvector matrix and $\mathbf{V}$ diagonal eigenvalue matrix of $\mathbf{A}$), next the scalar exponentials in diagonal matrix $e^{\mathbf{V}\Delta t_i}$ are computed, which finally is premultiplied by $\mathbf{M}$ and postmultiplied by $\mathbf{M}^{-1}$. SEM programs such as Mx do not allow to compute the matrix exponential directly, but allow matrix diagonalization and provide the scalar exponential function.

There are two options with regard to the trait covariance matrices Φ_κ and $\Phi_{\mathbf{x}_{t_0},\kappa}$: either you impose constraints on the discrete time analogues $\Phi_{\kappa\Delta t_i}$ and $\Phi_{\mathbf{x}_{t_0},\kappa\Delta t_i}$ separately in the forms shown in (7.17) or you constrain the coefficient matrix $\mathbf{H}_{\Delta t_i}$ of $\boldsymbol{\kappa}$ once, as a result of which both Φ_κ and $\Phi_{\mathbf{x}_{t_0},\kappa}$ come out in the right form automatically. The latter option is easiest and used here.

$$
\begin{aligned}
\mathbf{x}_{t_i} &= \mathbf{A}_{\Delta_{t_i}} \mathbf{x}_{t_i-\Delta_{t_i}} + \mathbf{b}_{\Delta_{t_i}} + \mathbf{H}_{\Delta_{t_i}} \boldsymbol{\kappa} + \mathbf{w}_{t_i-\Delta_{t_i}} \\
&\qquad\qquad\qquad \text{with } \operatorname{cov}(\mathbf{w}_{t_i-\Delta_{t_i}}) = \mathbf{Q}_{\Delta_{t_i}}, \\
\mathbf{A}_{\Delta t_i} &= e^{\mathbf{A}\Delta t_i} = \mathbf{M} e^{\mathbf{V}\Delta t_i} \mathbf{M}^{-1}, \\
\mathbf{b}_{\Delta t_i} &= \mathbf{A}^{-1}(e^{\mathbf{A}\Delta t_i} - \mathbf{I})\mathbf{b}, \\
\mathbf{H}_{\Delta t_i} &= \mathbf{A}^{-1}(e^{\mathbf{A}\Delta t_i} - \mathbf{I}), \\
\Phi_{\kappa\Delta t_i} &= \mathbf{H}_{\Delta t_i} \Phi_\kappa \mathbf{H}'_{\Delta t_i}, \\
\Phi_{\mathbf{x}_{t_0},\kappa\Delta t_i} &= \Phi_{\mathbf{x}_{t_0},\kappa} \mathbf{H}'_{\Delta t_i}, \\
\mathbf{Q}_{\Delta t_i} &= \operatorname{irow}\{\mathbf{A}_{\#}^{-1}[e^{\mathbf{A}_{\#}\Delta t_i} - \mathbf{I}]\operatorname{row}\mathbf{Q}\} \\
&\qquad \text{with } \mathbf{Q} = \mathbf{GG}' \text{ and } \mathbf{A}_{\#} = \mathbf{A} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{A}.
\end{aligned}
\tag{7.17}
$$

Evidently, the EDM repeats equation (7.17) for successive observation intervals $\Delta t_i = \Delta t_1, \ldots, \Delta t_{T-1}$ (T the total number of observation time points). If the observation intervals are unequal, the discrete time matrices with subscript Δt_i are different across time but relate nonlinearly in terms of the common time-invariant continuous time matrices $\mathbf{A}$, $\mathbf{b}$, Φ_κ, $\Phi_{\mathbf{x}_{t_0},\kappa}$, $\mathbf{G}$. If the observation intervals are equal, simple equality constraints between the discrete time matrices of successive observation intervals would suffice too. In addition to the direct estimation method, therefore, the indirect method would become applicable: computing the estimates of the continuous time matrices on the basis of the five previously estimated discrete time matrices by applying the constraints in (7.17) in inverse direction. In particular, starting from

$$
\mathbf{A} = \frac{1}{\Delta t_i} \ln \mathbf{A}_{\Delta t_i} = \frac{1}{\Delta t_i} \mathbf{M} \ln(\mathbf{V}_{\Delta t_i}) \mathbf{M}^{-1} \tag{7.18}
$$

by diagonalizing $\mathbf{A}_{\Delta t_i} = \mathbf{MV}_{\Delta t_i}\mathbf{M}^{-1}$ ($\mathbf{M}$ eigenvector matrix and $\mathbf{V}_{\Delta t_i} = e^{\mathbf{V}\Delta t_i}$ diagonal eigenvalue matrix of $\mathbf{A}_{\Delta t_i}$), $\mathbf{A}$ is found and then the other continuous time matrices easily follow. It cannot be emphasized enough that the indirect method is only applicable in the case of equal observation intervals. As in the case of unequal intervals no equality constraints can be imposed, each interval yields a different set of discrete time matrices, and, as a result of sampling fluctuations, each interval would also yield a different set of continuous time parameter matrices. So, in the case of unequal observation intervals, the direct method is the only suitable one.

In addition to the continuous time parameter matrices and the $T-1$ times repeated discrete time matrices in (7.17), the EDM as well as the ADM need one more parameter vector and one more parameter matrix for the initial time point t_0: initial means vector $\boldsymbol{\mu}_{\mathbf{x}_{t_0}}$ and initial covariance matrix $\Phi_{\mathbf{x}_{t_0}}$.

7.4.2 Full ADM

With regard to (7.19), the analogue of (7.17) for the ADM, which is directly taken from approximate integral form (7.15), the following observations apply. First, whereas the EDM (7.17) is formulated as a reduced form equation, the ADM (7.19) is in structural form. It means that the single autoregression matrix $\mathbf{A}_{\Delta t_i}$ in the EDM is replaced by two matrices in the ADM: instantaneous $\mathbf{A}^*_{\Delta t_i}$ and lagged $\mathbf{A}^{**}_{\Delta t_i}$. Both have been discussed earlier in Subsection 7.2.3, called there $\mathbf{A}_{ins}$ and $\mathbf{A}_{lag}$. The move from reduced form to structural form in combination with the replacement of exact drift matrix $\mathbf{A}$ by approximation $\mathbf{A}_\dagger$ leads to a dramatic simplification of the constraints on the discrete time matrices. The complicated nonlinear constraints in EDM (7.17) are replaced in ADM (7.19) by extremely simple linear expressions in terms of just the observation interval Δt_i or $\frac{1}{2}\Delta t_i$. Whereas the EDM constraints can only be applied by SEM programs such as Mx, which provide the exponential and matrix algebraic functions needed, the ADM constraints are applicable by almost any SEM program, in particular also by LISREL.

$$
\begin{aligned}
\mathbf{x}_{t_i} &= \mathbf{A}^*_{\Delta t_i}\mathbf{x}_{t_i} + \mathbf{A}^{**}_{\Delta t_i}\mathbf{x}_{t_i-\Delta_{t_i}} + \mathbf{b}^*_{\Delta t_i} + \mathbf{H}^*_{\Delta t_i}\boldsymbol{\kappa} + \mathbf{w}^*_{t_i-\Delta t_i} \\
&\qquad\qquad\qquad\qquad \text{with } \operatorname{cov}(\mathbf{w}^*_{t_i-\Delta t_i}) = \mathbf{Q}^*_{\Delta t_i}, \\
\mathbf{A}^*_{\Delta t_i} &= \tfrac{1}{2}\Delta t_i \mathbf{A}_\dagger, \\
\mathbf{A}^{**}_{\Delta t_i} &= \mathbf{I} + \tfrac{1}{2}\Delta t_i \mathbf{A}_\dagger, \\
\mathbf{b}^*_{\Delta t_i} &= \Delta t_i \mathbf{b}_\dagger, \\
\mathbf{H}^*_{\Delta t_i} &= \Delta t_i \mathbf{I}, \\
\Phi^*_{\kappa\Delta t_i} &= \Delta t_i^2 \Phi_{\dagger\kappa}, \\
\Phi^*_{\mathbf{x}_{t_0},\kappa\Delta t_i} &= \Delta t_i \Phi_{\dagger\mathbf{x}_{t_0},\kappa}, \\
\mathbf{Q}^*_{\Delta t_i} &= \Delta t_i \mathbf{Q}_\dagger = \Delta t_i \mathbf{G}_\dagger \mathbf{G}'_\dagger .
\end{aligned}
\tag{7.19}
$$

Again, if the observation intervals are equal, it becomes possible to extract the approximate continuous parameter matrices $\mathbf{A}_\dagger$, $\mathbf{b}_\dagger$, $\Phi_{\dagger\kappa}$, $\Phi_{\dagger\mathbf{x}_{t_0},\kappa}$, $\mathbf{G}_\dagger$ from the previously estimated set of structural form matrices by applying the simple constraints in

(7.19) in inverse direction. In fact, after applying during estimation, in addition to the equality constraints, the simple and only ADM constraints:

$$\mathbf{A}^{**}_{\Delta t_i} = \mathbf{I} + \mathbf{A}^{*}_{\Delta t_i}, \tag{7.20}$$

it comes down to dividing the left hand sides in (7.19) by $\frac{1}{2}\Delta t_i$, Δt_i, Δt_i^2, respectively, and finally computing the Cholesky factor $\mathbf{G}_{\dagger}$ of $\mathbf{Q}_{\dagger}$.

There is more, however, in the case of equal observation intervals. From structural form (7.19), by means of transformation matrix $\mathbf{D} = (\mathbf{I} - \mathbf{A}^{*}_{\Delta t_i})^{-1}$, we obtain the reduced form (7.21) which is in the form of (7.17). Because, as a result of the applied equality constraints, the reduced form matrices are equal across the successive time observation intervals, one-to-one relationships can be built between the ADM solutions, the reduced form solutions, and the EDM solutions, every reduced form solution giving rise to just one EDM solution and just one ADM solution and vice versa. As the confrontation with the data takes place via the common reduced form solution, the corresponding ADM and EDM solutions are equivalent, giving exactly the same model fit.

$$\begin{aligned}\mathbf{x}_{t_i} = \mathbf{D}\mathbf{A}^{**}_{\Delta t_i}\mathbf{x}_{t_i-\Delta_{t_i}} + \mathbf{D}\mathbf{b}^{*}_{\Delta t_i} + \mathbf{D}\mathbf{H}^{*}_{\Delta t_i}\boldsymbol{\kappa} + \mathbf{D}\mathbf{w}^{*}_{t_i-\Delta t_i} \\ \text{with } \operatorname{cov}(\mathbf{D}\mathbf{w}^{*}_{t_i-\Delta t_i}) = \mathbf{D}\mathbf{Q}^{*}_{\Delta t_i}\mathbf{D}'.\end{aligned} \tag{7.21}$$

In practice it means that one could start with the relatively simple ADM solution (7.19) by means of LISREL or some other user-friendly SEM program. Next, one could derive its reduced form using (7.21) and finally compute the corresponding EDM solution by means of the inverse constraints in (7.17) without any new SEM analysis. It should be noted that this is not possible in the case of unequal observation intervals, because then the indirect method is no option nor are the ADM and EDM solutions equivalent. Even then, however, it is often profitable to start with the relatively simple ADM solution to explore and evaluate the model and then use it as a reasonable initial solution for the final EDM analysis by means of the Mx program.

7.4.3 Putting ADM and EDM into SEM

A SEM model often can be specified in quite different ways and by different numbers of parameter matrices. Here we will put the ADM and EDM each into two equations with four parameter matrices: measurement parameter matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Theta}$, and structural parameter matrices $\mathbf{B}$ and $\boldsymbol{\Psi}$:

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad \text{with } \operatorname{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}, \tag{7.22}$$

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \qquad \text{with } \operatorname{cov}(\boldsymbol{\zeta}) = \boldsymbol{\Psi}. \tag{7.23}$$

For clarity, we limit the presentation to the case of four time points ($T = 4$: t_0, t_1, t_2, t_3) but this is easily reduced to three or extended to more than four time points. The model implied moment matrix $\mathbf{\Sigma} = \mathrm{f}(\mathbf{\Lambda}, \mathbf{\Theta}, \mathbf{B}, \mathbf{\Psi})$ is a function of the parameter matrices, the likelihood in turn is a function of $\mathbf{\Sigma}$ and sample moment matrix $\mathbf{S}$, and the maximum likelihood solution minimizes the discrepancy between $\mathbf{\Sigma}$ and $\mathbf{S}$ in the ML sense. Hence, for obtaining the maximum likelihood estimate of the ADM or EDM by means of a SEM program, it suffices to show how ADM and EDM are put into SEM parameter matrices $\mathbf{\Lambda}$, $\mathbf{\Theta}$, $\mathbf{B}$, and $\mathbf{\Psi}$. As the ADM is slightly simpler than the EDM, we start with the ADM.

For four time points the vectors $\mathbf{y}, \boldsymbol{\varepsilon}, \boldsymbol{\eta}, \boldsymbol{\zeta}$ in (7.22) and (7.23) look like:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_{t_0} \\ \mathbf{y}_{t_1} \\ \mathbf{y}_{t_2} \\ \mathbf{y}_{t_3} \\ 1 \end{bmatrix}, \ \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_{t_0} \\ \boldsymbol{\varepsilon}_{t_1} \\ \boldsymbol{\varepsilon}_{t_2} \\ \boldsymbol{\varepsilon}_{t_3} \\ 0 \end{bmatrix}, \ \boldsymbol{\eta} = \begin{bmatrix} \mathbf{x}_{t_0} \\ \mathbf{x}_{t_1} \\ \mathbf{x}_{t_2} \\ \mathbf{x}_{t_3} \\ 1 \\ \boldsymbol{\kappa} \end{bmatrix}, \ \boldsymbol{\zeta} = \begin{bmatrix} \mathbf{x}_{t_0} - \boldsymbol{\mu}_{\mathbf{x}_{t_0}} \\ \mathbf{w}_{t_1 - \Delta_{t_1}} \\ \mathbf{w}_{t_2 - \Delta_{t_2}} \\ \mathbf{w}_{t_3 - \Delta_{t_3}} \\ 1 \\ \boldsymbol{\kappa} \end{bmatrix}. \tag{7.24}$$

If the total number of variables in $\mathbf{y}$, the vector of observed variables including as the last variable the unit variable (1 for every subject in the sample), is $Tm + 1$, the total number of variables in $\boldsymbol{\eta}$, the vector of latent variables, is $(T + 1)n + 1$ with n the number of state variables as well as trait variables. Hence, case $m = n$ (e.g., when all state variables are observed) is one example in which the total number of latent variables may exceed the total number of observed variables.

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mu}_{\mathbf{x}_{t_0}} & \mathbf{0} \\ \mathbf{A}^{**}_{\Delta t_1} & \mathbf{A}^{*}_{\Delta t_1} & \mathbf{0} & \mathbf{0} & \mathbf{b}^{*}_{\Delta t_1} & \mathbf{H}^{*}_{\Delta t_1} \\ \mathbf{0} & \mathbf{A}^{**}_{\Delta t_2} & \mathbf{A}^{*}_{\Delta t_2} & \mathbf{0} & \mathbf{b}^{*}_{\Delta t_2} & \mathbf{H}^{*}_{\Delta t_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^{**}_{\Delta t_3} & \mathbf{A}^{*}_{\Delta t_3} & \mathbf{b}^{*}_{\Delta t_3} & \mathbf{H}^{*}_{\Delta t_3} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$$\mathbf{\Psi} = \begin{bmatrix} \Phi_{\mathbf{x}_{t_0}} & & & & & \\ \mathbf{0} & \mathbf{Q}^{*}_{\Delta t_1} & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}^{*}_{\Delta t_2} & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}^{*}_{\Delta t_3} & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & \\ \Phi_{\dagger \mathbf{x}_{t_0}, \kappa} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \Phi_{\dagger \kappa} \end{bmatrix}, \tag{7.25}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{C} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{d} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & \mathbf{0} & \mathbf{0} & \mathbf{d} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} & \mathbf{0} & \mathbf{d} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C} & \mathbf{d} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & \mathbf{0} \end{bmatrix}, \ \mathbf{\Theta} = \begin{bmatrix} \mathbf{R} & & & & \\ \mathbf{0} & \mathbf{R} & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{R} & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \end{bmatrix}.$$

The ADM (7.19) is put into the SEM model matrices $\mathbf{B}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Theta}$ in the way shown in (7.25). Observe that, due to the specification of the matrices $\mathbf{H}^{*}_{\Delta t_i}$ in SEM matrix $\mathbf{B}$, the trait covariance matrices $\Phi_{\dagger\kappa}$ and $\Phi_{\dagger\mathbf{x}_{t_0},\kappa}$ appear directly in SEM matrix $\boldsymbol{\Psi}$.

For the EDM, an alternative specification of the trait variables in the latent vector $\boldsymbol{\eta}$ and its error vector $\boldsymbol{\zeta}$ is employed.

$$\boldsymbol{\eta} = \begin{bmatrix} \begin{bmatrix} \mathbf{x}_{t_0} \\ \boldsymbol{\kappa} \end{bmatrix} \\ \begin{bmatrix} \mathbf{x}_{t_1} \\ \boldsymbol{\kappa} \end{bmatrix} \\ \begin{bmatrix} \mathbf{x}_{t_2} \\ \boldsymbol{\kappa} \end{bmatrix} \\ \begin{bmatrix} \mathbf{x}_{t_3} \\ \boldsymbol{\kappa} \end{bmatrix} \\ 1 \end{bmatrix}, \quad \boldsymbol{\zeta} = \begin{bmatrix} \begin{bmatrix} \mathbf{x}_{t_0} - \boldsymbol{\mu}_{\mathbf{x}_{t_0}} \\ \boldsymbol{\kappa} \end{bmatrix} \\ \begin{bmatrix} \mathbf{w}_{t_1 - \Delta_{t_1}} \\ \mathbf{0} \end{bmatrix} \\ \begin{bmatrix} \mathbf{w}_{t_2 - \Delta_{t_2}} \\ \mathbf{0} \end{bmatrix} \\ \begin{bmatrix} \mathbf{w}_{t_3 - \Delta_{t_3}} \\ \mathbf{0} \end{bmatrix} \\ 1 \end{bmatrix}. \tag{7.26}$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\mu}_{\mathbf{x}_{t_0}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{A}_{\Delta t_1} & \mathbf{H}_{\Delta t_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{b}_{\Delta t_1} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{\Delta t_2} & \mathbf{H}_{\Delta t_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{b}_{\Delta t_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{\Delta t_3} & \mathbf{H}_{\Delta t_3} & \mathbf{0} & \mathbf{0} & \mathbf{b}_{\Delta t_3} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \end{bmatrix},$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \Phi_{\mathbf{x}_{t_0}} & & & & & & & & \\ \Phi_{\mathbf{x}_{t_0},\kappa} & \Phi_{\kappa} & & & & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_{\Delta t_1} & & & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{\Delta t_2} & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_{\Delta t_3} & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 \end{bmatrix}, \tag{7.27}$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mathbf{C} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{d} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{d} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{d} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C} & \mathbf{0} & \mathbf{d} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 \end{bmatrix}, \quad \boldsymbol{\Theta} = \begin{bmatrix} \mathbf{R} & & & & \\ \mathbf{0} & \mathbf{R} & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{R} & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \end{bmatrix}.$$

It could also be used for the ADM as the previous specification could be used for the EDM. The alternative specification highlights that the trait variables $\boldsymbol{\kappa}$ are just a special kind of state variables, namely constant across time. Hence, as shown in (7.26), the trait variables are added to the state vector $\mathbf{x}_{t_i}$ at each time point, leading to the larger total number of $2Tn+1$ variables in latent vector $\boldsymbol{\eta}$. The specification puts the model directly in a suitable form for the application of state space techniques, in particular the Kalman smoother or conditional mean $E[\mathbf{x}(t)|\mathbf{y}]$, for optimally estimating a subject's sample trajectory (Oud & Jansen, 1996; Oud, Jansen, van Leeuwe, Aarnoutse, & Voeten, 1999). Based on (7.17), the SEM specification of the EDM in (7.27), follows, apart from the trait variables, the same pattern as in the case of the ADM in (7.25). Whereas the state variables develop across time according to $\mathbf{A}_{\Delta t_i}$ and are influenced by the trait variables according to $\mathbf{H}_{\Delta t_i}$, the trait variables themselves, remaining constant over time, develop according to the identity matrix $\mathbf{I}$. The trait covariance matrix Φ_κ and state-trait covariance matrix $\Phi_{\mathbf{x}_{t_0},\kappa}$ are found in the second row of $\boldsymbol{\Psi}$. The measurement model matrices do not differ from the ones in (7.25), except that $\boldsymbol{\Lambda}$ has extra zero columns at the places of the unobserved trait variables.

7.4.4 Relating Models on Different Time Scales

Researchers in the same or different subject fields often use different observation intervals. As argued in Subsection 7.2.2, comparing longitudinal models with different observation intervals, a clear condition for cumulative progress in science, requires continuous time analysis. This is not all, however. Meaningful comparison also requires the results to be put on the same time scale. Because both the ADM and EDM are time-invariant, time scale shifts, $t' = t + d$, do not change the results. The ADM and EDM in (7.17) and (7.19) show, however, that changing the time scale unit, $t' = ct$ (for example, going from years t to months t': $t' = 12t$), indeed affects the parameter matrices, but in a quite simple way, not requiring any re-estimation of the parameter matrices. If one wants to compare one's results with another researcher, who used time scale $t' = ct$ ($\Delta t_i' = c\Delta t_i$) instead of one's own scale t, simply multiply four of the five parameter matrices by $1/c$ ($\mathbf{A}$, $\mathbf{b}$, $\Phi_{\mathbf{x}_{t_0},\kappa}$, $\mathbf{G}$ in EDM; $\mathbf{A}_\dagger$, $\mathbf{b}_\dagger$, $\Phi_{\dagger\mathbf{x}_{t_0},\kappa}$, $\mathbf{G}_\dagger$ in ADM) and one by $1/c^2$ (Φ_κ in EDM ; $\Phi_{\dagger\kappa}$ in ADM). The reason is that the discrete time matrices on the left hand side of the constraints equations in (7.17) and (7.19) keep the same value, while the change from Δt_i to $c\Delta t_i$ or from $\mathbf{A}^{-1}$ to $c\mathbf{A}^{-1}$ on the right hand side needs compensation by multiplying by $1/c$.

7.5 Relationships between Externalizing and Internalizing Problem Behavior

In this and the next section applications will be presented. The LISREL program will be used for the ADM-SEM procedure (see the commented LISREL script in the Appendix of Oud, 2007b) and the Mx program for the EDM-SEM procedure. All input and output files of the analyses performed in both sections are available in Chapter 7 at the book website `http://www.econ.upf.edu/~satorra/longitudinallatent/readme.html`. In this section, four LISREL maximum likelihood analyses and two Mx maximum likelihood analyses examine the relationships between externalizing and internalizing problem behavior in adolescents. It will be studied in continuous time whether and how strongly externalizing problem behavior influences internalizing problem behavior (failure perspective: Burke, Loeber, Lahey, & Rathouz, 2005; Capaldi, 1992), internalizing problem behavior influences externalizing problem behavior (acting out perspective: Carlson & Cantwell, 1980; Gold, Mattlin, & Osgood, 1989), or both variables influence each other reciprocally (mutual influence perspective: Overbeek, Vollebergh, Meeus, Luypers, & Engels, 2001). The two state variables in the model (externalizing problem behavior and internalizing problem behavior) are observed. So, the measurement model part includes only loadings 1, intercepts 0, and measurement error variances 0. In the next section, a model with an elaborate measurement model will be presented for relationships between three latent state variables.

The data analyzed are taken from a comprehensive Dutch study of family relationships and adolescent problem behavior (Nijmegen Family and Personality Study; Haselager & van Aken, 1999). Participants were 280 adolescents (139 boys, 141 girls) who were 14.5 years old on average (ranging from 11.4 to 16.0) at the first measurement wave. To assess adolescents' externalizing and internalizing problem behavior, participants completed the Nijmegen Problem Behavior List (NPBL; De Bruyn, Scholte, & Vermulst, 2005) at each of the three annual measurement waves. Further details regarding sample characteristics, measures, and procedure can be found in Delsing, Oud, van Aken, De Bruyn, and Scholte (2005).

Although the aim of the ADM is the estimation of the (approximate) underlying continuous time parameters, it is nevertheless clarifying to view in Figure 7.6 the discrete time part of the ADM in SEM form. The model contains the state variables *Ext* and *Int* and corresponding constant trait variables *Trait-Ext* and *Trait-Int*, for three time points leading to a total of eight variables in the SEM model (apart from the ninth unit variable, which is not depicted in Figure 7.6). The figure clearly shows the discrete time part of the ADM with instantaneous coefficients ($\mathbf{A}^{*}_{\Delta t_i}$) as well as lagged coefficients ($\mathbf{A}^{**}_{\Delta t_i}$). The ADM in SEM form is one of the rare SEM models with self-loop coefficients specified and estimated (diagonals in the instantaneous matrices $\mathbf{A}^{*}_{\Delta t_i}$, indicated in the figure by self-referencing arrows). In total, the continuous time part of the model contains 21 parameters to be estimated:

4 drift coefficients in $\mathbf{A}_{\dagger}$,
2 intercepts feeding mean development in $\mathbf{b}_{\dagger}$,

3 trait variances and covariance in $\Phi_{\dagger\kappa}$,
4 covariances between traits and initial states in $\Phi_{\dagger \mathbf{x}_{t_0},\kappa}$,
3 state variable diffusion coefficients in $\mathbf{G}_{\dagger}$,
2 initial state means in $\boldsymbol{\mu}_{\mathbf{x}_{t_0}}$,
3 initial state variances and covariance in $\Phi_{\mathbf{x}_{t_0}}$.

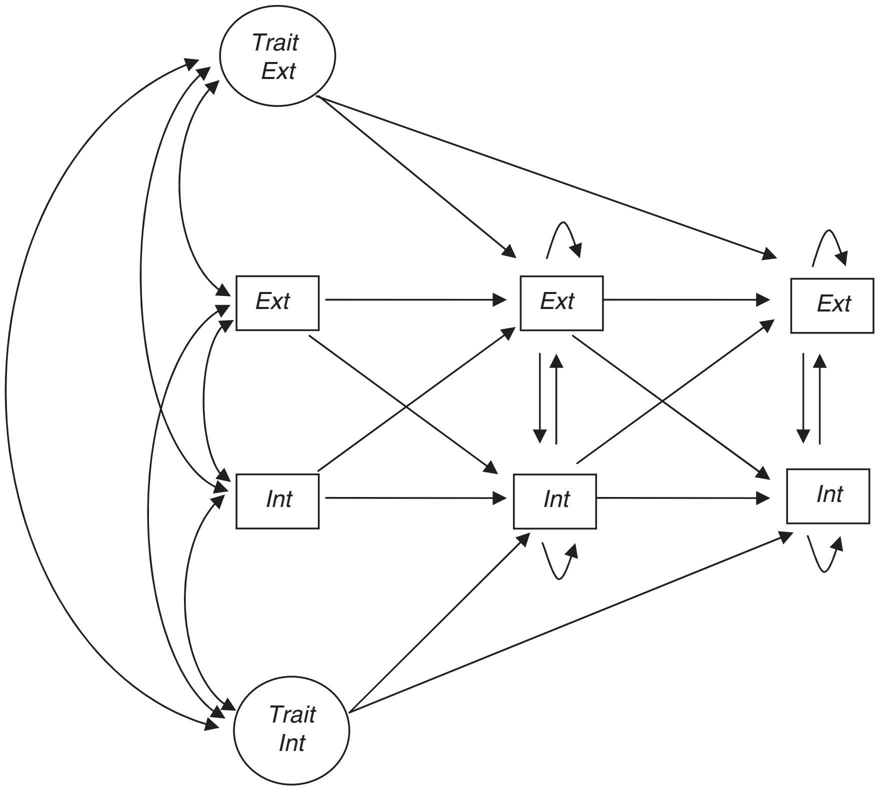

Fig. 7.6 The three-wave ADM-SEM cross-lagged panel model for adolescents' externalizing and internalizing problem behavior, including corresponding trait variables (subject-specific intercepts).

A full ADM as well as a full EDM can be proven to be identified for $T \geq 3$, assuming the measurement model part is identified. As the ADMs and EDMs in this section have only observed state variables, the model does not have free measurement parameters and so this part is identified automatically. Column 1 of Table 7.2 displays the estimate of the full ADM model (input file *ADM1.ls8* and output file *ADM1.out*). For convenience, the subscript † in the ADM parameter names is suppressed in Table 7.2. For equal observation intervals of length $\Delta t_i = 1$, many of the parameter estimates are immediately found in the LISREL parameter matrices $\mathbf{B}$ (BETA) and $\boldsymbol{\Psi}$ (PSI). Parameters not immediately found there but estimated as

Table 7.2 Estimates and model fit information for ADM1 (Full ADM), ADM2 (No Trait 1), ADM3 (No Trait 1, No *Int* → *Ext*) and EDM3 (No Trait 1, No *Int* → *Ext*); standardized drift coefficients

Parameter	ADM1	ADM2	ADM3	EDM3
$a_{11}(Ext)$	-0.790**	-0.302**	-0.317**	-0.320**
$a_{12}(Int \rightarrow Ext)$	0.347	-0.039		
$a_{21}(Ext \rightarrow Int)$	0.788**	0.616**	0.605**	0.704**
$a_{22}(Int)$	-1.056**	-1.134**	-1.110**	-1.251**
$\mu_{x_{1t_0}}$	17.943**	17.943**	17.943**	17.943**
$\mu_{x_{2t_0}}$	21.106**	21.106**	21.106**	21.106**
$\phi_{x_{1t_0}}$	27.323**	27.323**	27.261**	27.323**
$\phi_{x_{2t_0}}$	38.212**	38.212**	37.462**	38.212**
$\phi_{x_{21t_0}}$	10.990**	10.990**	10.774**	10.990**
b_1	7.817	5.979**	5.559**	5.606**
b_2	5.598	10.890*	10.730*	11.505*
g_{11}	4.922**	4.761**	4.762**	4.782**
g_{22}	6.228**	6.226**	6.219**	6.562**
g_{21}	1.347**	1.648**	1.557**	1.315**
ϕ_{κ_1}	10.965			
ϕ_{κ_2}	26.014	27.646	26.554	33.958
$\phi_{\kappa_{21}}$	-15.059			
$\phi_{x_{1t_0},\kappa_1}$	7.235			
$\phi_{x_{2t_0},\kappa_1}$	-5.364			
$\phi_{x_{1t_0},\kappa_2}$	-12.731*	-8.063*	-7.988*	-8.887*
$\phi_{x_{2t_0},\kappa_2}$	14.731	18.619*	17.999	20.694
χ^2	5.4	9.9	10.6	10.6
df	6	10	11	11
RMSEA	0.0	0.0	0.0	0.0

$*p \leq .05$; $**p \leq .01$.

so-called additional parameters are the four drift coefficients: auto-effects a_{11} and a_{22} (called PA(1) and PA(2), respectively, in the LISREL output) and cross-effects a_{12} and a_{21} (PA(3) and PA(4), respectively), and the three diffusion coefficients g_{11}, g_{22}, and g_{21} (PA(7), PA(9), and PA(8), respectively).

With regard to the main purpose of the study, assessing the existence and strength of the cross-effects a_{12} $(Int \rightarrow Ext)$ and a_{21} $(Ext \rightarrow Int)$ between internalizing and externalizing problem behavior, one should realize an important difference in interpretability between them and the auto-effects a_{11}, a_{22}. The auto-effects are scale free in the sense that they do not change under arbitrary linear transformations of the variables *Ext* and *Int* and so are directly interpretable. In particular, both *Ext* and *Int* show negative feedback ($-.790$ and -1.056), indicating stability or a quite strong tendency for an individual to converge to its subject-specific mean trajectory. The negative eigenvalues of the drift matrix confirm that the model as a whole is stable. The cross-effects are not scale free, however. Their value depends on the standard deviations of the independent and dependent variable involved. The cross-effects in Table 7.2 have therefore been standardized (PA(5) and (PA(6) in the LISREL output) through multiplication by the ratios of the initial standard deviations. However, as t-values are scale free, testing can best be done in terms of the unstandardized values of 0.293 for a_{12} (not significant) and 0.932 for a_{21} $(p < .01)$ (the t-values computed by LISREL for the standardized values inappropriately also include the sampling variability of the standard deviations). The standardized values of 0.347 for a_{12} and 0.788 for a_{21} in combination with the testing results reported in Table 7.2 seem to reveal the existence of a strong unidirectional effect from externalizing to internalizing problem behavior with no or little effect in the opposite direction.

The specification of traits in a model has, in general, high impact on the estimates of the other parameters. Because in the full ADM no significant variances and covariances were found for the traits κ_1 and κ_2 except for the covariance between κ_2 and x_{1t_0}, we decided to retain only κ_2 in the next model ADM2 (files *ADM2.ls8* and *ADM2.out*), so that this model has subject-specific mean trajectories for internalizing problem behavior but only a single general mean trajectory for externalizing problem behavior. It is interesting that the exclusion of κ_1 from ADM2 led to the non-significant effect a_{12} $(Int \rightarrow Ext)$ in ADM1 turning slightly negative in ADM2 but with a non-significant t-value again that was even lower than in the ADM1. This was reason to next delete a_{12} from the model. The resulting ADM3 (files *ADM3.ls8* and *ADM3.out*) has all parameters significant except two which are related to κ_2. It retains in particular an impressively strong effect a_{21} $(Ext \rightarrow Int)$.

Having found a clear and particularly well fitting ADM (the extra constraints introduced into ADM2 and ADM3 do not deteriorate the fit shown by χ^2 and RMSEA), the obvious next step is to replace the approximate ADM by the exact EDM (see EDM3 in Table 7.2). As explained above for the case of equal observation intervals, one possibility would be to apply the indirect method by computing the reduced form matrices according to (7.21) and deriving the EDM from the ADM3 by (7.17) instead of using the direct method by running the Mx program. Note, that the (estimated) reduced form autoregression matrix $\mathbf{A}_{\Delta t_i}$ is already computed by LISREL in the first part of the matrix "Total effects of ETA on ETA" in *ADM3.out*. This indeed turns out to be exactly equal to the (estimated) autoregression matrix $\mathbf{A}_{\Delta t_i}$ computed by Mx in the first part of its BETA matrix (called "A" in GROUP 7 of Mx output file *EDM3.mxo*). Autoregressions for *Ext* and *Int* in both are equal to 0.72627 and 0.28614 and the cross-lagged effect in both is equal to 0.39352. One reason to apply the direct method by running the Mx program could be, however, that in addition to the EDM solution itself one gets also the correct standard errors.

The Mx analysis has therefore indeed been done (files *EDM3.mx* and *EDM3.mxo*) and the results are displayed in the last column of Table 7.2.

Although the reduced forms of ADM3 and EDM3 should be and are indeed equal in this case of equal observation intervals as is the fit of both models (the models are equivalent via their reduced form), the solutions themselves are close to each other but not equal; compare the last two columns in Table 7.2 (the EDM3 solution is found in GROUP 9 in the Mx output file *EDM3.mxo,* displaying all estimated parameter matrices; the standardized value of a_{21} is computed and found in GROUP 10). Our experience is, that the EDM often yields a somewhat more pronounced solution with the parameter estimates showing higher absolute values. This is clearly also the case here. For example, the standardized value of 0.704 for a_{21} $(Ext \rightarrow Int)$ in EDM3 points to an even stronger effect of externalizing problem behavior on internalizing problem behavior than found in ADM3. Our analyses leave little doubt that the failure perspective is the one confirmed by the data in this section and not the acting out or mutual influence perspectives discussed in the literature.

As has been stressed several times, the equivalence of the ADM and EDM is based on the equality of the observation intervals. If the observation intervals are unequal, ADM and EDM can give quite different reduced forms and a quite different fit. So, then no other choice is left than estimating the EDM independently from the ADM. To show how analyses with unequal observation intervals are performed

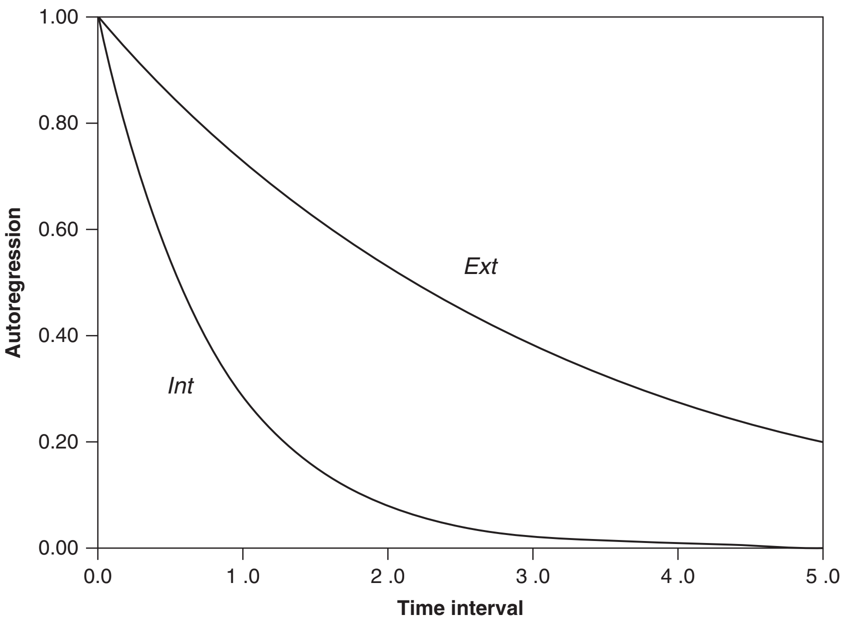

Fig. 7.7 Autoregression functions of *Ext* and *Int*, based on model EDM3 in Table 7.2.

and what could happen in the example with unequal intervals, instead of the true equal intervals $\Delta t_1 = \Delta t_2 = 1$, we simulated the second interval to be slightly larger than the first interval: $\Delta t_1 = 1$, $\Delta t_2 = 1.2$. So, the interval between the second and third wave is taken 1.2 years instead of the true 1 year. The analyses are performed by input files *ADM4.ls8* and *EDM4.mx* for LISREL and Mx, respectively, and the output is found in files *ADM4.out* and *EDM4.mxo.* Instead of the same χ^2-value of 10.6 for both ADM3 and EDM3, we now find $\chi^2 = 17.3$ for ADM4 and $\chi^2 = 11.2$ for EDM4. In both the fit deteriorates but ADM4 turns out to be much more sensitive to the wrong specification of the second observation interval than EDM4. Of course, autoregressions and cross-lagged effect are different for the unequal intervals in each analysis (whereas, as expected, the autoregressions were lower over the longer second interval, the cross-lagged effect turned out to be higher), but they also differ now between ADM4 and EDM4. Over the first interval the autoregressions in the ADM4 were 0.748 and 0.253 for *Ext* and *Int*, respectively, and in the EDM4 0.747 and 0.261; the cross-lagged effect in the ADM4 was 0.405 and in the EDM4 0.410. This clearly illustrates the necessity to estimate the EDM independently from the ADM in the case of unequal intervals.

We conclude the example with some of the consequences of the model estimated in continuous time: autoregression functions for *Ext* and *Int*, cross-lagged effect function for a_{21} $(Ext \rightarrow Int)$ and mean trajectories for *Ext* and *Int*, all based on the final EDM3 in Table 7.2. Autoregression and cross-lagged effect functions are

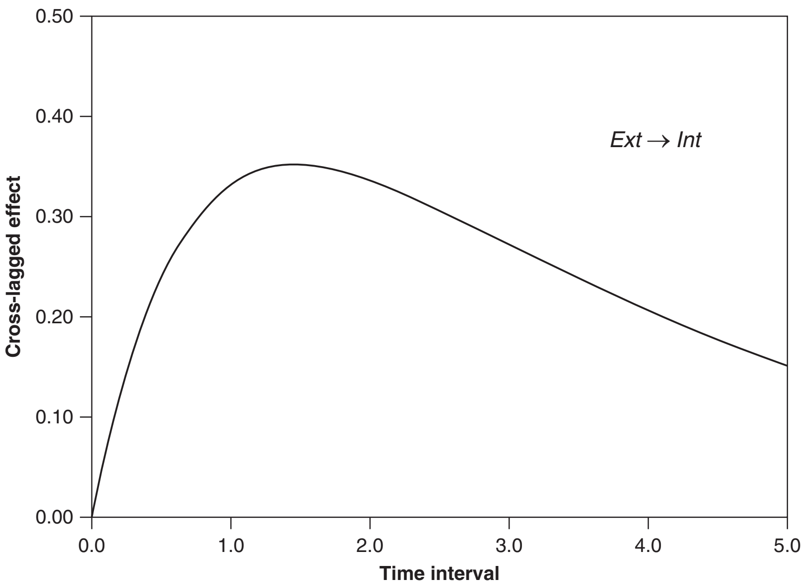

Fig. 7.8 Standardized cross-lagged effect function for *Ext* $\rightarrow$ *Int*, based on model EDM3 in Table 7.2.

computed using (7.5). An autoregression function describes the autonomous development of a variable, telling which proportion of its value present at the start predictably persists after increasing intervals. In discrete time analysis the autoregression is only computable on the basis of the discrete time intervals in the study. The continuous time analysis in Figure 7.7 reveals that the part of *Ext* predictable from its value at the start is everywhere higher than for *Int*, after an interval of $\Delta t =$ 2.2 years the predictable *Ext* part is still not less than half the original *Ext*, whereas only 7% of the original *Int* is left, and already after 6 months half of the original *Int* is lost.

The cross-lagged effect function in Figure 7.8 reveals that a unit (standard deviation) increase in *Ext* has its maximum impact of 35.2% of a standard deviation in *Int* after 1.5 years and that after 5 years still 15% is left. So, the cross-lagged effect function nicely clarifies what the meaning and impact across all intervals in continuous time is of the impressive cross-effect of 0.704 in model EDM3.

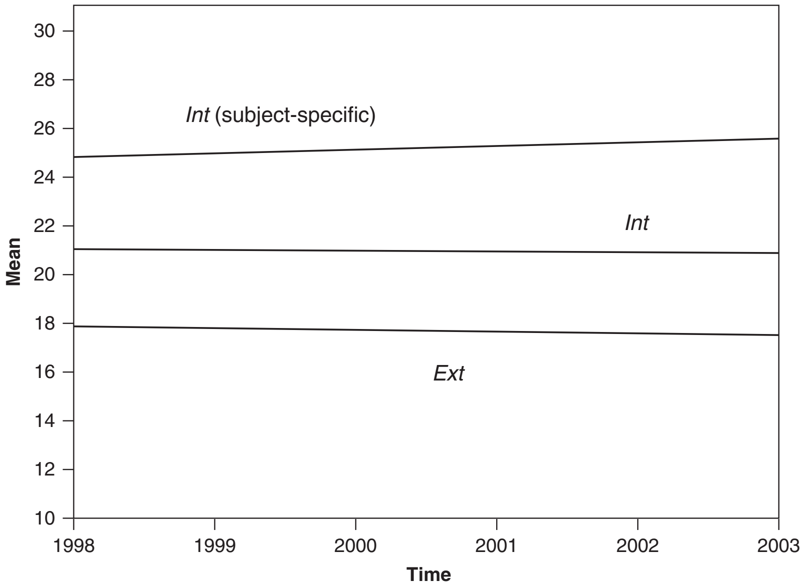

Fig. 7.9 Mean trajectories for *Ext* and *Int* and a subject-specific mean trajectory for *Int*, based on model EDM3 in Table 7.2.

General mean trajectories for *Ext* and *Int* in Figure 7.9 were computed using (7.11) and the subject-specific mean trajectory for *Int* using (7.12). As an illustration, the subject-specific mean trajectory was computed for a subject at one standard deviation above trait mean for *Int*, that is for $\kappa_2 = 6.18$. Because κ_1 was deleted from model EDM3 and therefore Φ_κ is not positive definite, (7.12) was applied by filling out in regression matrix $\Phi_{\mathbf{x}_{t_0},\kappa}\Phi_\kappa^{-1}$ the regressions of x_{1t_0} and x_{2t_0} on κ_2 only.

It turns out that the means hardly change. The general means, starting from 17.9 and 21.1 in 1998, going down only a tiny fraction over the data collection period between 1998 and 2000 towards 17.7 and 21.0, are predicted to be 17.6 and 20.9 in 2003 and converge for $t \to \infty$ to final values of 17.53 and 20.86. The subject-specific mean trajectory for *Int* of the subject with subject-specific intercept value of $b_2 + \kappa_2 = 17.7$, however, increases: from 24.9 in 1998 to 25.2 in 2000 and then in the prediction period, after reaching 25.6 in the middle of 2003, to final value 25.80. This final value is not much higher than the value in 2000 at the end of the data collection period.

7.6 Relationships between Individualism, Nationalism and Ethnocentrism in Flandres

The example in this section, taken from Toharudin, Oud, and Billiet (2008), is more comprehensive than the one in the previous section for two reasons. First, the state variables are latent and based on an elaborate measurement model for measuring the theoretical constructs Individualism (I), Nationalism (N) and Ethnocentrism (E). The constructs were repeatedly measured in three waves (1991, 1995, and 1999) in a panel of $N = 1274$ Flemish respondents and Dutch-speaking respondents in Brussels. Second, whereas the number of state variables was two in the previous section, in this example it is three, leading to a 3 x 3 drift matrix with six different causal connections between the latent variables. The purpose of the study was to find out, how the constructs develop and influence each other across time. On the basis of previous research a recursive causal structure was hypothesized: $I \to N, N \to E, I \to E$. Thus, in addition to the auto-effects, only three of the six possible cross-effects were hypothesized to be nonzero. In previous research, causal connections between the constructs were analyzed cross-sectionally only, or, if longitudinally, solely in correlational form without taking care of the causal direction of the effects, and never in continuous time. Again, the continuous time analysis of the data set started with an ADM analysis (LISREL input file *ADM-INE.ls8* and output file *ADM-INE.out*), followed by the corresponding EDM analysis (Mx input file *EDM-INE.mx* and output file *EDM-INE.mxo*).

First, attention will be paid to the measurement model. Individualism (I) or "unrestrained striving for personal interests", Nationalism (N) or "identification with the Flemish community in Belgium", and Ethnocentrism (E) or "negative attitude toward outgroups" were measured by 5, 4, and 8 items, respectively. Most of the items were 5-point-scale items, the answers consisting of different degrees of agreement/disagreement. Two item examples for each of the constructs are:

Individualism (I)
-Everybody has to take care of himself first.
-What counts is money and power.

Nationalism (N)
-Flanders must decide.
-Belgium has to disappear.

Ethnocentrism (N)
-Belgium should not have allowed in guest workers.
-Immigrants cannot be trusted.

Several items under Nationalism had slight differences in formulation between the first wave year 1991 and the two subsequent wave years 1995 and 1999. Measurement invariance (time invariance) analyses were performed, in which the loadings, measurement intercepts, and measurement error variances (**C**, **d**, and **R** in (7.16), respectively) of these N-items were compared between 1991 on the one hand and 1995 and 1999 on the other hand. It yielded that time invariance of N-item 4 for 1991 (called "4na91" in *ADM-INE.out)* in comparison with 1995 and 1999 had to be rejected. Consequently, the three measurement parameters of this item were allowed to deviate in 1991. Only one more deviation from time-invariance was allowed in the measurement model by freeing the measurement intercept of I-item 5 in 1999 (called "5in99" in *ADM-INE.out*). Freeing this single parameter, thereby increasing its value from 2.456 to 3.286, had the effect of decreasing the model χ^2 for ADM and EDM with the huge amount of 1209, implying a considerable improvement in model fit. It prevents the increase in this single item from unduly influencing the latent mean development of I between 1995 and 1999. All information about the loadings of the items can be found in the main body of matrix "LAMBDA-Y" in output file *ADM-INE.out*, about the measurement intercepts in the last column of this matrix, and about the measurement error variances in "THETA-EPS".

With regard to the dynamic model part and the initial state variances and means, first the ADM and EDM estimates will be presented and then details about the way the input files were formulated to obtain the estimates. Both solutions are given in Tables 7.3 and 7.4. In Table 7.3 the ADM solution is on the left hand side and the EDM solution on the right hand side. Because the estimates of the initial state variances and means are equal in both solutions, they are given only once in Table 7.4. Both tables also give t-values, providing precise information about the significance of the parameter estimates as well as about the standard errors (t = estimate/standard error). Although in both ADM and EDM, trait variables were specified (in the form of extra state variables as in (7.26) and (7.27)), all three trait variances were fixed at zero in the final analysis, because no positive estimates were found or expected to be found (see in *ADM-INE.out* the negative values under "Expected change for Psi" for the variances of "04w1TrI", "05w1TrN", and "06w1TrE"). We conclude that the initial variances are sufficient to differentiate trajectories for individual subjects from the mean trajectory and no extra trait variances are warranted. Diffusion coefficient matrix **G** and diffusion matrix $\mathbf{Q} = \mathbf{GG}'$ were specified diagonal because of the rather low modification indices and expected changes in *ADM-INE.out* for the off-diagonal elements.

Comparing the ADM and EDM solutions in Tables 7.3-7.4, it is striking how similar both solutions are with only very small differences in the third decimal of the parameter estimates. Also the differences in t-values are small, proving that the standard errors are very similar too. These almost equal results, obtained on the basis of different models by quite diverse programs, clearly confirm each other. It means also that the EDM can safely be evaluated by means of the ADM estimates, standard errors and other information given by the LISREL program. As expected, also the fit information is equal within precision limits. $\chi^2 = 7881$ in LISREL and $\chi^2 = 7880$ in Mx with $df = 1304$ do not seem to imply a particularly good fit. It should be noted, however, that the model with 51 observed variables is huge, the sample of $N = 1274$ big, and the (almost) strict time-invariance of the continuous-time model puts a lot of heavy constraints on the model, on the measurement part as well as on the dynamic part of the model. It is therefore no surprise that the popular

Table 7.3 ADM and EDM estimates; standardized coefficients in drift matrices $\mathbf{A}_\dagger$ and $\mathbf{A}$, t-values between parentheses

	$\mathbf{A}_\dagger$: I	$\mathbf{A}_\dagger$: N	$\mathbf{A}_\dagger$: E	$\mathbf{A}$: I	$\mathbf{A}$: N	$\mathbf{A}$: E
I	−0.069** (−9.41)	−0.007 (−1.39)	0.033** (5.15)	−0.070** (−9.23)	−0.008 (−1.44)	0.033** (5.13)
N	0.013* (2.31)	−0.061** (−10.79)	0.011* (2.09)	0.013* (2.33)	−0.061** (−10.17)	0.012* (2.11)
E	0.039** (7.48)	0.003 (0.80)	−0.062** (−11.98)	0.040** (7.53)	0.003 (0.82)	−0.063** (−11.94)

	$\mathbf{b}_\dagger$	$\mathbf{b}$
I	0.061** (3.28)	0.061** (3.32)
N	0.105* (2.21)	0.105* (2.19)
E	0.094** (7.58)	0.095** (7.62)

	$\mathbf{G}_\dagger$: I	$\mathbf{G}_\dagger$: N	$\mathbf{G}_\dagger$: E	$\mathbf{G}$: I	$\mathbf{G}$: N	$\mathbf{G}$: E
I	0.280** (11.37)			0.281** (10.83)		
N		0.683** (11.06)			0.685** (10.70)	
E			0.210** (15.39)			0.211** (14.92)

$*p \leq .05$; $**p \leq .01$.

Table 7.4 Estimates of initial state (co)variances and means; t-values between parentheses, first line of ADM and second line of EDM

	I	N	E	
I	0.778**			2.460**
	(17.45)			(83.18)
	(16.57)			(82.96)
N	−0.183*	5.999**		4.231**
	(−2.41)	(17.38)		(51.64)
	(−2.41)	(16.51)		(50.49)
E	0.321**	0.106	0.508**	2.899**
	(13.44)	(1.87)	(17.10)	(116.82)
	(13.34)	(1.86)	(17.11)	(116.73)
		$\Phi_{\mathbf{x}_{t_0}}$		$\boldsymbol{\mu}_{\mathbf{x}_{t_0}}$

*$p \leq$ .05; **$p \leq$ .01.

fit measure RMSEA (Browne and Cudeck, 1993) with value 0.068 indicates that the model fits reasonably.

Turning to the drift matrix **A**, which should give the answers to the main questions in the study, we first observe that the auto-effects are all three negative $(-0.069, -0.061, -0.062)$, indicating stability or a long-term tendency for the trajectories to converge to the mean trajectory. Stability is confirmed by the negative eigenvalues of the drift matrix. Interestingly, by accounting appropriately for the 4 year observation interval, the auto-effects are correctly comparable to the auto-effects of -0.320 and -1.251 in the previous example (Section 7.5) with a 1 year interval. Individualism, Nationalism and Ethnocentrism in the present example have a much weaker tendency to converge to their mean trajectory than externalizing problem behavior and internalizing problem behavior in the previous example (Section 7.5).

The cross-effects do not confirm the hypothesized recursive structure $I \rightarrow N$, $N \rightarrow E$, $I \rightarrow E$. In the place of non-significant and almost zero effect $N \rightarrow E$ come significant effects $E \rightarrow N$ and $E \rightarrow I$. The role of Nationalism is therefore quite different from what was expected. N turns out not to influence E, but, in contrast, to undergo a weak influence from E. So, N comes out as the dependent variable in the structure, weakly and nearly equally influenced by both other constructs (standardized coefficients of 0.013 and 0.011). In addition, a clear reciprocal relationship shows up between Individualism and Ethnocentrism: $I \rightarrow E$ but also $E \rightarrow I$ with standardized coefficients of 0.039 and 0.033, respectively. All standardized effects are small in strength, though, and, although significant, much smaller than the standardized effect of 0.704 found in the previous example (Section 7.5).

As mentioned above, in both the ADM and the EDM trait variables were specified as extra state variables. This can be seen by SEM matrix **B** (called "BETA" in

LISREL output *ADM-INE.out* and "A" in GROUP 42 of Mx output *EDM-INE.mxo*) containing 19 variables: in addition to the last variable, the unit variable, at each time point 3 state variables are followed by the extra 3 trait variables. Just as $\mathbf{B}$, $\boldsymbol{\Psi}$ is also a 19 x 19 matrix (called "PSI" in *ADM-INE.out* and "P" in GROUP 42 of *EDM-INE.mxo),* showing the initial covariance matrix $\Phi_{\mathbf{x}_{t_0}}$ in the first 3 x 3 diagonal block and the trait covariance matrix Φ_{κ} (in the final analysis fixed at zero) in the next 3 x 3 diagonal block.

For the ADM analysis, the equal observation intervals of $\Delta t_1 = \Delta t_2 = 4$ were reason to apply the simple ADM constraints in (7.20) between lagged and instantaneous matrices $\mathbf{A}^{**}_{\Delta t_1}$ and $\mathbf{A}^{*}_{\Delta t_1}$ in addition to the equality constraints between time points. In the LISREL output file *ADM-INE.out* one finds the lagged matrix $\mathbf{A}^{**}_{\Delta t_1}$ in BETA at variables 7-9 (dependent) and 1-3 (lagged independent) and the instantaneous matrix $\mathbf{A}^{*}_{\Delta t_1}$ at variables 7-9 (dependent and independent). The ADM constraints (7.20) are formulated in LISREL input file *ADM-INE.ls8* following "!ADM equality (auto)" and "!ADM equality constraints (cross)". As explained below formula (7.20), the estimates of drift coefficients in $\mathbf{A}_{\dagger}$, intercepts in $\mathbf{b}_{\dagger}$, and diffusion coefficients in $\mathbf{G}_{\dagger}$ can easily be obtained by hand from the estimated discrete time matrices using (7.19). For convenience, these simple computations have also been done by means of the LISREL program: division of $\mathbf{A}^{*}_{\Delta t_1}$ by $\frac{1}{2}\Delta t_1 = 2$ (multiplication by 0.5, yielding the drift coefficients in additional LISREL parameters PA(1)-PA(9)), division of $\mathbf{b}^{*}_{\Delta t_1}$ by $\Delta t_i = 4$ (multiplication by 0.25, yielding the drift coefficients in additional LISREL parameters PA(16)-PA(18)), and division of $\mathbf{Q}^{*}_{\Delta t_1}$ by $\Delta t_i = 4$, followed by the square-root of the result ($\sqrt{\frac{1}{4}q^{*}_{ii,\Delta t_1}} = 0.5\text{x}(q^{*}_{ii,\Delta t_1})^{0.5}$ yielding the diffusion coefficients in additional LISREL parameters PA(19)-PA(21)). All elements in trait matrices $\Phi_{\dagger\kappa}$ and $\Phi_{\dagger\mathbf{x}_{t_0},\kappa}$ were specified zero in the final analysis, but, if nonzero, they could have been computed by dividing $\Phi^{*}_{\kappa\Delta t_i}$ and $\Phi^{*}_{\mathbf{x}_{t_0},\kappa\Delta t_i}$ by $\Delta t_1^2 = 16$ and $\Delta t_1 = 4$, respectively.

The scale free character of t-values can be observed by comparison of the t-values for $\mathbf{A}^{*}_{\Delta t_1}$ and $\mathbf{A}_{\dagger}$ as well as for $\mathbf{b}^{*}_{\Delta t_1}$ and $\mathbf{b}_{\dagger}$. In both cases, the estimated values are different but the t-values are indeed equal. One would expect the t-values also to be equal for $\mathbf{Q}^{*}_{\Delta t_1}$ and $\mathbf{G}_{\dagger}$ as well as for the six unstandardized cross-effects in PA(2), PA(3), PA(4), PA(6), PA(7), PA(8) and their standardized values computed in PA(10)-PA(15). It turns out that the t-values of the three standard deviations in $\mathbf{G}_{\dagger}$ are not one time but exactly two times those of the variances in $\mathbf{Q}^{*}_{\Delta t_1}$. This is a consequence of the square root transformation covering negative as well as positive values. Because negative values have to be excluded for standard deviations, one should stick to half the values found for the diagonals in $\mathbf{G}_{\dagger}$ as reported in Table 7.3. The t-values computed by LISREL for the standardized coefficient values inappropriately also take into account the variability of the standard deviations used in standardization. The scale of the variables can be chosen arbitrarily, however. If the standard deviations would have been specified in PA(10)-PA(15) as fixed quantities, the t-values would have been equal to those computed for $\mathbf{A}^{*}_{\Delta t_1}$ and $\mathbf{A}_{\dagger}$ as reported in Table 7.3.

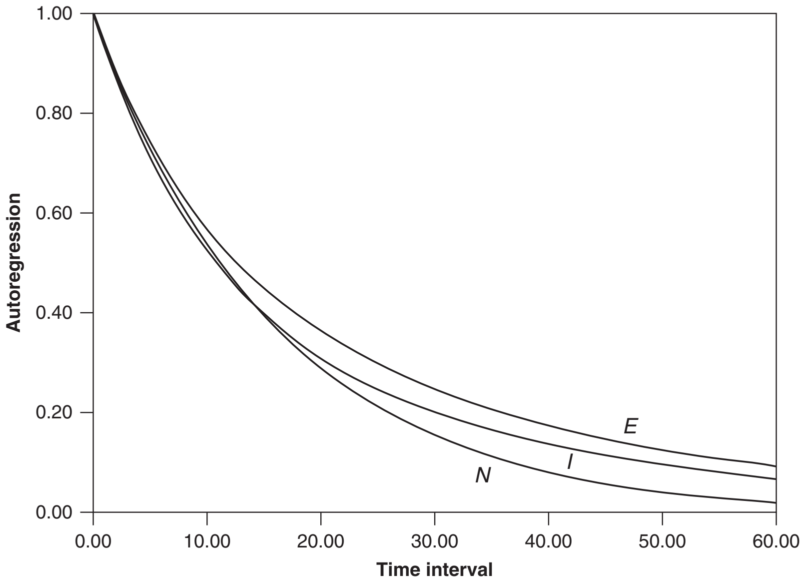

Fig. 7.10 Autoregression functions of I, N, and E, based on model EDM in Table 7.3.

For the EDM analysis a much more elaborate Mx input file was used than in the previous example (Section 7.5). The 45 GROUPS enable the estimation of the more general, time-varying model discussed in Oud and Jansen (2000). For estimating the present time-invariant model, only a subset of the GROUPS is needed, however. At the start of the input file *EDM-INE.mx*, it is clearly indicated in which GROUPS the time-invariant model matrices are specified. GROUP 44 in output file *EDM-INE.mxo* displays all the estimated matrices, reported in Tables 7.3-7.4. In GROUP 45 the standardized drift matrix is computed. Standard errors, on the basis of which the t-values reported in Tables 7.3-7.4 were computed, are displayed by Mx at the start of the output before GROUP 1. As in the case of the ADM, the t-values for the diffusion coefficients (standard deviations) in **G** have been halved. All discrete time matrices (see (7.17)) are found at the positions specified by (7.27) in **B** ("A" in GROUP 42) and in $\mathbf{\Psi}$ ("P" in GROUP 42).

To depict in continuous time the short-run and long-run implications of the model for Flandres, we conclude the example with the autoregression functions (Figure 7.10), cross-lagged effect functions (Figure 7.11), and mean trajectories (Figure 7.12), all based on the EDM estimates in Tables 7.3-7.4. Autoregression and cross-lagged effect functions have been computed using (7.5). The autoregressions and cross-lagged effects are computed not only for the actual observation intervals of 4 and 8 years, but interpolated and predicted for any interval over a rather extended prediction period. Although the differences between the autoregression functions

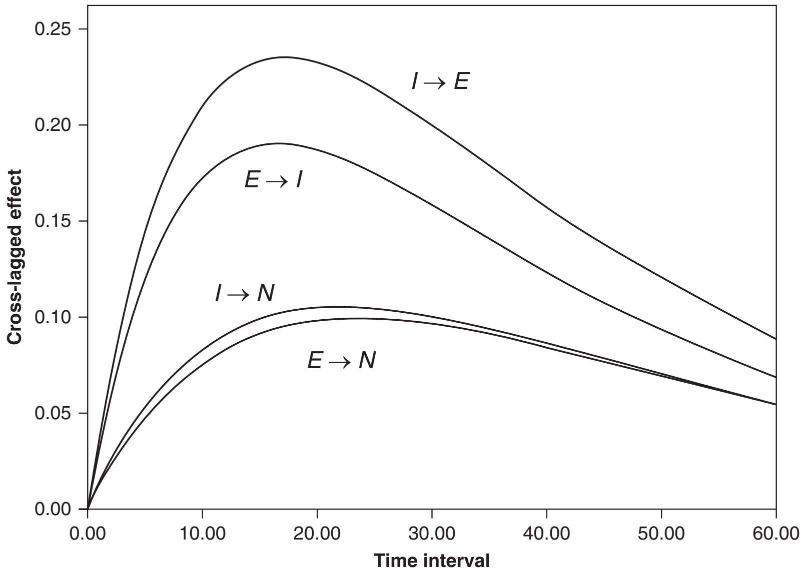

Fig. 7.11 Standardized cross-lagged effect functions for significant cross-effects of EDM in Table 7.3.

in Figure 7.10 are rather small (all three state variables have a strong tendency to persist over time, much stronger than, for instance, externalizing and internalizing problem behavior in Figure 7.7), the non-monotone character of the autoregression functions nicely illustrates the need for analyzing in continuous time. The strength order between N and I reverses after interval 14.2, N becoming the variable with the lowest autoregression. It means that a discrete-time analyst confronted, for example, with observation interval 16 or 20 would come to different conclusions than his colleague, working with interval 4 or 8. Continuous time analysis, however, prevents such erroneous conclusions by showing the complete picture.

In contrast to the the autoregression functions in Figure 7.10, the standardized cross-lagged effect functions in Figure 7.11 show monotonicity. For all intervals, at least over 0 to 60 years, the strength order between the four cross-lagged effect functions turns out to be the same as for the actual observation intervals of 4 and 8. Particularly, in the feedback loop between I and E, $I \rightarrow E$ exceeds $E \rightarrow I$ everywhere. With regard to the two relatively smaller effects, we observe that everywhere $I \rightarrow N$ is slightly stronger than $E \rightarrow N$. An interesting result of the continuous time analysis is further that all four cross-lagged effects reach their maximum quite some time after the empirical observation intervals of 4 and 8 years. The maximum of $I \rightarrow E$ (0.235) is reached somewhat later, at interval 17.0, than the lower maximum of $E \rightarrow I$ in the opposite direction (0.190), reached at 16.4. The maximum of $I \rightarrow N$

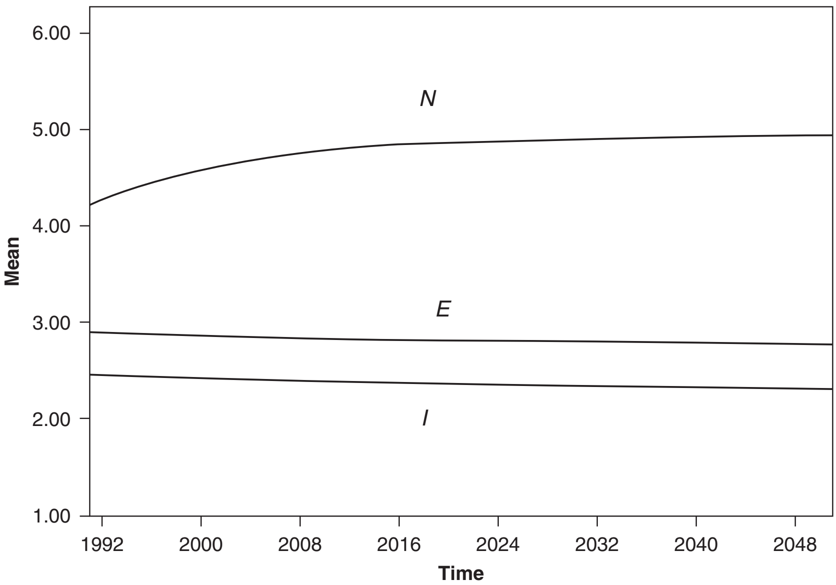

Fig. 7.12 Latent mean trajectories for I, N and E, based on Tables 7.3–7.4.

(0.105) at interval 22 is reached 1.2 years before the maximum of $E \rightarrow N$ (0.099) at 23.2.

Finally, Figure 7.12 shows that the mean values of Individualism and Ethnocentrism hardly changed in the data collection period 1991-1999 and are hardly expected to change in the prediction period. The mean of Individualism, starting at 2.46 in 1991 and decreasing to 2.43 in 1999, is expected to converge to final value 2.30. The mean of Ethnocentrism, starting at 2.90 in 1991 and decreasing to 2.87 in 1999, is expected to converge to final value 2.77. The mean of Nationalism, however, increased in the data collection period from 4.23 to 4.55, and a further limited increase for the near future is expected, but then the prediction levels off towards final value 4.91.

7.7 Conclusions

The development of externalizing and internalizing problem behavior in children or the attitude change in the Flemish electorate with regard to individualism, nationalism, and ethnocentrism are continuously evolving processes, rather than processes that show isolated, sudden changes at discrete points in time. The analyst, however, only observes at discrete points in time (for example, biennial, yearly or monthly observations). The typical approach in conventional (that is, discrete) time series modeling and panel data analysis is to ignore the continuous time nature of the processes underlying discrete time observations. Consequently, discrete time series and discrete panel data analyses are simplifications and often distortions of reality.

Comparability of results between different studies is the key to cumulative progress in science. Just because of the frequent model formulation and estimation in terms of the observation interval at hand, comparability is low in social and behavioral science. By means of the continuous-time approach using the exact discrete model (EDM) or approximate discrete model (ADM), explained in this chapter, the model parameters are made independent of the observation interval, and thus provide a common basis for accurate comparison of differently time-spaced models of the same or similar processes. As shown also in this chapter, if analysis results for the EDM or ADM from different authors use time scales in different units, they are easily translated into each other. Thus, results are made comparable without re-estimation being necessary.

Not all topics in continuous time analysis could be covered by the EDM-SEM and ADM-SEM procedures as expounded in the present chapter. We mention, in particular, time-varying models (Oud & Jansen, 2000) and models for oscillating movements (Oud, 2007a). In our conviction, however, the models presented in this chapter give a continuous time formulation to the typical kind of problems current longitudinal and panel research in social and behavioral science is involved with. A final but important topic not dealt with in the present chapter is the handling of incomplete data. In a longitudinal SEM context this can be solved in most cases by the expectation-maximization (EM) procedure using the Kalman smoother, explained in Oud and Jansen (1996) and applied in continuous-time modeling by Oud and Jansen (2000), or the individual likelihood procedure (Neale, 2000; Wothke, 2000) as implemented, for instance, in Mx.

References

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

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.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Journal of the Royal Statistical Society (Supplement), 7*, 27-41.

Blalock, H. M., Jr. (1969). *Theory construction: From verbal to mathematical formulations.* Englewood Cliffs, NJ: Prentice-Hall.

Bergstrom, A. R. (1966). Nonrecursive models as discrete approximations to systems of stochastic differential equations. *Econometrica, 34,* 173-182.

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.

Boker, S., Neale, M., & Rausch, J. (2004). Latent differential equation modeling with multivariate multi-occasion indicators. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Recent developments* *on structural equations models: Theory and applications* (pp. 151-174). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Browne, M. W. & R. Cudeck (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. Scott Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park: Sage.

Bui, K. V. T., Ellickson, P. L., & Bell, R. M. (2000). Cross-lagged relationships among adolescent problem drug use, delinquent behavior, and emotional distress. *Journal of Drug Issues, 30,* 283-303.

Burke, J. D., Loeber, R., Lahey, B. B., & Rathouz, P. J. (2005). Developmental transitions among affective and behavioral disorders in adolescent boys. *Journal of Child Psychology and Psychiatry, 46,* 1200-1210.

Capaldi, D. M. (1992). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: II. A 2-year follow-up at grade 8. *Development and Psychopathology, 4,* 125-144.

Carlson, G. A., & Cantwell, D. P. (1980). Unmasking masked depression in children and adolescents. *American Journal of Psychiatry, 137,* 445-449.

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

De Bruyn, E. E. J., Scholte, R. H. J., & Vermulst, A. A. (2005). *Psychometric analyses of the Nijmegen problem behavior list (NPBL): A research instrument for assessing problem behavior in community samples using self- and other reports of adolescents and parents.* Nijmegen, The Netherlands: Institute of Family and Child Studies, Radboud University Nijmegen

Delsing, M. J. M. H., Oud, J. H. L., van Aken, M. A. G., De Bruyn, E. E. J., & Scholte, R. J. H. (2005). Family loyalty and adolescent problem behavior: The validity of the family group effect. *Journal of Research on Adolescence, 15,* 127-150.

Gold, M., Mattlin, M., & Osgood, D. W. (1989). Background characteristics and response to treatment of two types of institutionalized delinquent boys. *Criminal Justice and Behaviour, 16,* 5-33.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58,* 80-92.

Haselager, G. J. T., & van Aken, M. A. G. (1999). *Codebook of the research project Family and Personality: Vol. 1. First measurement wave.* Nijmegen, The Netherlands: University of Nijmegen, Faculty of Social Science.

Homans, G. C. (1950). *The human group.* New York: Harcourt, Brace & World.

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

Jöreskog, K. G., & Sörbom, D. (1976). *LISREL III: Estimation of linear structural equation systems by maximum likelihood methods: A FORTRAN IV program.* Chicago: National Educational Resources.

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

Kuo, H.-H. (2006). *Introduction to stochastic integration.* New York: Springer.

Neale, M.C. (2000). Individual fit, heterogeneity, and missing data in multigroup structural equation modeling. In T.D. Little, K.U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data* (pp. 219-240). Mahwah NJ: Lawrence Erlbaum.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2006). *Mx: Statistical Modeling* (7th ed.). Richmond, VA: Department of Psychiatry.

Neiderhiser, J. M., Reiss, D., Hetherington, E. M., & Plomin, R. (1999). Relationships between parenting and adolescent adjustment over time: Genetic and environmental contributions. *Developmental Psychology, 35,* 680-692.

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

Oud, J. H. L. (2007a). Comparison of four procedures to estimate the damped linear differential oscillator for panel data. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 19-39). Mahwah, NJ: Lawrence Erlbaum Associates.

Oud, J. H. L. (2007b). Continuous time modeling of reciprocal effects in the cross-lagged panel design. In S.M. Boker & M.J. Wenger (Eds.), *Data analytic techniques for dynamical systems in the social and behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

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.

Oud, J. H. L., Jansen, R. A. R. G., van Leeuwe, J.F.J., Aarnoutse, C. A. J., & Voeten, M. J. M. (1999). Monitoring pupil development by means of the Kalman filter and smoother based upon SEM state space modeling. *Learning and Individual Differences, 10,* 121-136.

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

Oud, J. H. L., & Singer, H. (2008). Continuous time modeling of panel data: SEM versus filter techniques. *Statistica Neerlandica, 62*, 4-28.

Overbeek, G. J., Vollebergh, W. A. M., Meeus, W. H. J., Luijpers, E. T. H., & Engels, R. C. M. E. (2001). Course, co-occurence and longitudinal associations between emotional disturbance and delinquency from adolescence to young adulthood: A six-year three-wave study. *Journal of Youth and Adolescence, 30*, 401-426.

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.

Simon, H. A. (1952). A formal theory of interaction in small groups. *American Sociological Review, 17,* 202-211.

Singer, H. (1990). *Parameterschätzung in zeitkontinuierlichen dynamischen Systemen [Parameter estimation in continuous time dynamic systems]*. Konstanz: Hartung-Gorre.

Singer, H. (1991). *LSDE - A program package for the simulation, graphical display, optimal filtering and maximum likelihood estimation of linear stochastic differential equations: User's guide*. Meersburg: Author.

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. (1995). Analytical score function for irregularly sampled continuous time stochastic processes with control variables and missing values. *Econometric Theory, 11*, 721–735.

Singer, H. (1998). Continuous panel models with time dependent parameters. *Journal of Mathematical Sociology*, 23, 77-98.

Toharudin, T., Oud, J. H. L., & Billiet, J. B. (2008). Assessing the relationships between Nationalism, Ethnocentrism, and Individualism in Flanders using Bergstrom's approximate discrete model. *Statistica Neerlandica, 62*, 83-103.

Tuma, N. B., & Hannan, M. (1984), *Social dynamics: Models and methods*. New York:Academic Press.

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

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T.D. Little, K.U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data* (pp. 219-240). Mahwah NJ: Lawrence Erlbaum.

Zadeh, L. A., & Desoer, C. A. (1963*). Linear system theory: The state space approach.* New York: McGraw-Hill.